Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

# Technological advances to make single-cell and spatial proteomics biologically relevant

Marvin Christoffer Thielert

aus

Sigmaringen, Deutschland

2024

## Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Matthias Mann betreut.

## Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

| | |
|---|---|
| München, 20.03.2025 | |
| Ort, Datum | Marvin Christoffer Thielert |

| | |
|---|---|
| Dissertation eingereicht am | 20.03.2024 |
| 1. Gutachter/in: | Hon.-Prof. Dr. Matthias Mann |
| 2. Gutachter/in: | Prof. Dr. Stephan Sieber |
| Mündliche Prüfung am | 16.04.2024 |

# Summary

The cell, as the smallest foundational unit of life, plays an important role in our understanding of biology. It is the basis for studying molecular mechanisms and function of our holistic organism in health and disease. This also enables the unravelling of disease origins and devising potential novel treatment strategies. To study single cells, a broad portfolio of technologies and methods have been developed over the years. Mass spectrometry-based proteomics is one of the most recent ones, and it stands out as an unbiased technique enabling a global analysis of proteins – the cell's principal functional drivers – at the single-cell level. Despite its obvious attractions, limitations in sensitivity have hampered single-cell proteomics until recently. In a larger perspective, the human body is an intricate spatial arrangement of these functional cellular units, and these functions are tied to the surrounding context. Spatial proteomics offers vast potential to disentangle this complex cellular machinery within its native environment. Although we are only at the beginning of single-cell and spatial proteomics, a multitude of biological questions can already be addressed from this new experimental angle.

This PhD thesis brings together my efforts in advancing single-cell and spatial proteomics technologies and leverages them to address pertinent biological questions. To this end, I first contributed to the development of novel scan modes for acquisition of proteomic data by mass spectrometry. I highlight our advances in single-cell proteomics to enable label-free analysis and also to increase proteomic depth and throughput by multiplexed-data independent acquisition. The thesis then transitions from the isolated single cell to complex spatial arrangements. Here, I directed my efforts to improve and streamline the workflow and to enhance spatial proteomics to accommodate single-cell and three-dimensional analysis. With these technologies in hand, I contributed to the investigations in ovarian cancer of serous borderline tumors, and in a separate project to the discovery of a novel therapy for drug-induced skin reaction diseases. Finally, I contributed principles of ultra-high single-cell proteomics to immunopeptidomics to characterize peptides bound by circulating MHC receptors.

Together, this thesis presents technological advances in single-cell and spatial proteomics and their application in clinically and biologically relevant contexts, demonstrating the power of these tools on their own or in combination. In a future perspective, this work boosts system-wide biological and clinical studies for understanding the functional mechanisms in health and disease, and helps to pave the way towards precision medicine.

# Zusammenfassung

Die Zelle, als kleinste grundlegendste Einheit des Lebens, ist essentiell für unser Verständnis von Biologie. Sie bildet die Basis für das Erforschen molekularer Mechanismen und Funktionen unseres gesamtheitlichen Organismus im Kontext von Gesundheit und Krankheit. Ein tiefgründiges Wissen ermöglicht nicht nur das Verständnis grundlegender biologischer Mechanismen, sondern eröffnet auch Perspektiven zur Aufdeckung von Krankheitsursachen und potenziellen neuen Behandlungsstrategien. Zur Analyse einzelner Zellen wurden viele verschiedene Techniken entwickelt. Die Massenspektrometrie-basierte Proteomik ist eine der neuesten und ermöglicht als unvoreingenommene Methode die globale Analyse von Proteinen – den hauptsächlichen funktionalen Treibern der Zelle – auf Einzelzellenebene. Trotz der naheliegenden Vorzüge war die Einzelzell-Proteomik aufgrund ihrer begrenzten Sensitivität bis vor kurzem in ihrer Anwendung limitiert. Betrachtet im Gesamtbild ist der menschliche Körper eine komplexe räumliche Anordnung dieser funktionalen zellulären Einheiten, dessen Funktion von der Umgebung beeinflusst ist. Räumliche Proteomik bietet enormes Potenzial, diese komplexe zelluläre Maschinerie in ihrer natürlichen Umgebung aufzulösen. Obwohl wir erst am Anfang der Einzelzell- und räumlicher Proteomik stehen, können bereits mehrere biologischen Fragen aus dieser neuen experimentellen Perspektive beantwortet werden.

Diese Doktorarbeit umfasst die technische Entwicklung von Einzelzell- und räumlicher Proteomik und nutzt diese, um wichtige biologische Fragen zu beantworten. Zunächst habe ich zur Entwicklung neuer Scan Modi für die Datenaufnahme in der Proteomik mittels Massenspektrometrie beigetragen. Danach führe ich unsere Fortschritte auf dem Gebiet der Einzelzell-Proteomik auf, die die Label-freie Analyse sowie die Steigerung der Proteintiefe und des Durchsatzes ermöglichen. Danach erweitere ich den Fokus von der isolierten Einzelzelle zur komplexen räumlichen Anordnung der Zellen. Hier habe ich mich darauf konzentriert, die Probenvorbereitung zu optimieren und die räumliche Proteomik auf Einzelzell- und dreidimensionale Analysen zu erweitern. Mit diesen Entwicklungen trug ich wesentlich zu den Untersuchungen zu Eierstockkrebs bei, sowie zur Entdeckung einer neuen Therapie für medikamenteninduzierte Hautreaktionen. Zuletzt wirkte ich mit die Prinzipen der Einzelzell-Proteomik auf die Immunopeptidomik zu übertragen, um die gebunden Peptide auf zirkulierenden MHC-Rezeptoren zu charakterisieren.

Zusammenfassend stellt diese Arbeit die technologischen Fortschritte in der Einzelzell-Proteomik und räumlichen Proteomik und ihre Anwendung in klinisch und biologisch relevanten Bereichen vor. Diese zeigen das Potential sowohl einzeln als auch in Kombination. Aus zukünftiger Perspektive fördert diese Arbeit das Potenzial weiterer Entwicklungen, um systemweite biologische und klinische Studien zur Erforschung der funktionalen Mechanismen in Gesundheit und Krankheit zu ermöglichen, und ebnet den Weg zur personalisierten Medizin.

# Abbreviations

| Abbreviation | Explanation |
|---|---|
| ACN | acetonitrile |
| AGC | automatic gain control |
| AI | artificial intelligence |
| Astral | asymmetric track lossless |
| CID | collision induced dissociation |
| CV | coefficient of variation / compensation voltage / central vein |
| DDA | data-dependent analysis |
| DIA | data-independent analysis |
| DISCO-MS | 3D imaging of solvent-cleared organs profiled by MS |
| DNA | deoxyribonucleic acid |
| DRESS | drug reaction with eosinophilia and systemic symptoms |
| DVP | Deep Visual Proteomics |
| ESI | electrospray ionization |
| FACS | fluorescent-activated cell sorting |
| FAIMS | field-asymmetric ion mobility spectrometry |
| FDR | false discovery rate |
| FFPE | formalin-fixed, paraffin-embedding |
| HIER | heat-induced epitope retrieval |
| HLA | human leukocyte antigen |
| IM | ion mobility |
| LC | liquid chromatography |
| LCM | laser-capture microdissection |
| LFQ | label-free quantification |
| LGSC | low-grad serous cancer |
| m/z | mass-to-charge |
| mDIA | multiplexed data-independent analysis |
| MHC | major histocompatibility complex |
| MS | mass spectrometry |
| MS2 | tandem mass spectrometry, also MS/MS |
| NGS | next generation sequencing |
| PASEF | parallel accumulation – serial fragmentation |
| PTM | post-translational modification |
| PV | portal vein |
| (m)RNA | (messenger) ribonucleic acid |
| SBT(-MP) | serous borderline tumors (micropapillary) |
| SCoPE-MS | single cell proteomics by mass spectrometry |
| sc-proteomics | single-cell proteomics |
| scRNA-seq | single-cell RNA sequencing |
| TEN | toxic epidermal necrolysis |
| TIMS | trapped ion mobility spectrometry |
| TMT | tandem mass tag |
| TOF | time-of-flight |

## Table of contents page

**Summary** ..... **V**

**Zusammenfassung** ..... **VI**

**Abbreviations** ..... **IX**

**1. Introduction** ..... **1**

1.1. Biological importance of the cell and their proteins ..... 1
1.1.1. History of cell theory ..... 1
1.1.1. Proteomic diversity and complexity as functional unit in the cell ..... 2

1.2. Single-cell omics ..... 4
1.2.1. Single-cell oligonucleotide sequencing ..... 4
1.2.2. Single-cell proteomics ..... 6

1.3. Mass spectrometry-based proteomics for single-cell proteomics ..... 7
1.3.1. Bottom-up proteomics ..... 8
1.3.2. Sample preparation ..... 9
1.3.3. Liquid chromatography ..... 11
1.3.4. Mass spectrometry ..... 14
1.3.5. Data acquisition strategies ..... 21
1.3.6. Computational proteomics ..... 23
1.3.7. Summary of the challenges in single-cell proteomics ..... 26

1.4. Spatial proteomics ..... 27
1.4.1. Deep Visual Proteomics (DVP) ..... 28

1.5. Clinical proteomics ..... 30
1.5.1. Proteome in health and disease ..... 30
1.5.2. Post-translational modifications of proteins ..... 30
1.5.3. Cancer biology ..... 31
1.5.4. Pancreatic islet biology ..... 32
1.5.5. Immunopeptidomics ..... 33

**2. Aims of the thesis** ..... **34**

**3. Publications** ..... **36**

3.1. Method development for mass spectrometry ..... 36
Article 1. Rapid and In-Depth Coverage of the (Phospho-)Proteome with Deep Libraries and Optimal Window Design for dia-PASEF. ..... 36
Article 2. Synchro-PASEF allows precursor-specific fragment ion extraction and interference removal in data-independent acquisition. ..... 54

3.2. Ultra-high sensitivity mass spectrometry and single-cell proteomics ..... 70
Article 3. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. ..... 70
Article 4. Robust dimethyl-based multiplex-DIA workflow doubles single-cell proteome depth via a reference channel ..... 87
Article 5. Making single-cell proteomics biologically relevant. ..... 112

3.3. Spatial proteomics ..... 119
Article 6. A standardized and reproducible workflow for membrane glass slides in routine histology and spatial proteomics. ..... 120
Article 7. Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome. ..... 132
Article 8. Spatial proteomics in three-dimensional intact specimens. ..... 145
Article 9. Spatial proteo-transcriptomic profiling reveals the molecular landscape of borderline ovarian tumors and their invasive progression. ..... 179
Article 10. Single-cell spatial proteomics identifies the JAK/STAT pathway as an actionable therapeutic target in lethal cutaneous drug reactions. ..... 207
3.4. Transfer of technology on the example of immunopeptidomics ..... 240
Article 11. IMBAS-MS discovers organ-specific HLA peptide patterns in plasma. ..... 240
3.5. Pancreatic islet biology ..... 256
Article 12. Cell-type resolved proteome dissection of mouse endocrine pancreatic cell types ..... 256
**4. Discussion and Outlook ..... 276**
4.1. Improvements in data-acquisition ..... 276
4.2. Single-cell proteomics ..... 277
4.3. Multiplexed-DIA and the reference channel concept ..... 278
4.4. The promise of spatial proteomics ..... 279
4.5. A vision for single-cell and spatial proteomics ..... 281
**5. Bibliography ..... 282**
5.1. References ..... 282
5.2. List of figures ..... 292
**6. Supplements ..... 294**
Article 13. Large-scale proteomics across neurological conditions uncovers biomarker panels and targets in multiple sclerosis ..... 294
Article 14. AlphaViz: Visualization and validation of critical proteomics data directly at the raw data level. ..... 295
**7. Acknowledgement ..... 296**

‘We cannot solve our problems with the same thinking we used when we created them.’

**Albert Einstein**

‘If you only do what you can do
you will never be more than you are now.’

**Master Shifu, Kong Fu Panda 3**

# 1. Introduction

## 1.1. Biological importance of the cell and their proteins

### 1.1.1. History of cell theory

" The cell, as the simplest form of life-manifestation that nevertheless fully represents the idea of life, is the organic unit, the indivisible living one."

Rudolf Virchow (1821-1902), 1849[1]

This statement of Rudolf Virchow in 1849 suggests that by studying the function of individual cells, we will be able to explain the essence of life. This statement concluded the cell theory defined by Matthias Jakob Schleiden[2] and Theodor Schwann[3] in the middle of the nineteenth century. Originally however, the term 'cell' had already been introduced in 1665. In that year, Robert Hooke first discovered a multitude of tiny pores that represented the microscopic units of cork[4]. Inspired by these minute compartments, Hooke minted the term from the Latin word *Cella*, meaning 'a small chamber' in an ancient temple, or *Cellulae*, a 'small compartment'. Building on this observation, Anton van Leeuwenhoek, known for his development of the microscope, described motile cells which he already referred to as living cells. In 1674, van Leeuwenhoek also discovered different types of cells – the red blood cells and sperm cells[5]. At that time, the idea of a fundamental unit of life already existed, but its precise nature was unknown until Henri Dutrochet proposed in 1821 that 'the cell is the fundamental element of organization'[6]. Eventually in 1838, Matthias Jakob Schleiden, who focused on the origin of cells, projected that every plant structure originates from cells or is a result of cells[2]. With this, he started the cell theory for plant cells. His postulate was then generalized to all living organisms, including animals, by Theodor Schwann[3]. Together their ideas constituted the first two parts of cell theory: (i) living organisms are composed of one or more cells and (ii) the cell is the most basic unit of life. Rudolf Virchow, the father of pathology, further extended the cell theory. He published his idea that all cells arise from other pre-existing cells: 'omnis cellula e cellula' in 1855. This idea took cues from Robert Remak's proposal in 1852 that all embryonic cells multiply by cell division[7]. Further, the statement of Rudolf Virchow formed the basis for the idea of evolution presented by Charles Darwin[8] and paved the way for advances in cytology and medicine. It also concluded the classical cell theory.

Modern cell theory is one of the main principles of biology, encompassing the three historical fundamentals. Since then, a few more principles have been added to the original postulated ones, such as: DNA is the information passed on from cell to cell, all cells have the same chemical composition and metabolic activity, which is carried out inside the cell, as well as that the cell's activity is dependent on activities of structures inside the cell.

### 1.1.1. Proteomic diversity and complexity as functional unit in the cell

Having established the individual unit of a cell with the modern cell theory, one can dive deeper into its molecular function. Another essential principle in biology is the 'central dogma of biology', which was first introduced by Francis Crick three years after their foundation of the double-helical DNA structure[9]. It states that genetic information flows from desoxyribonucleic acid (DNA, genome) by transcription to ribonucleic acid (RNA, transcriptome) and translation to proteins, a fundamental process known as gene expression (Figure 1).

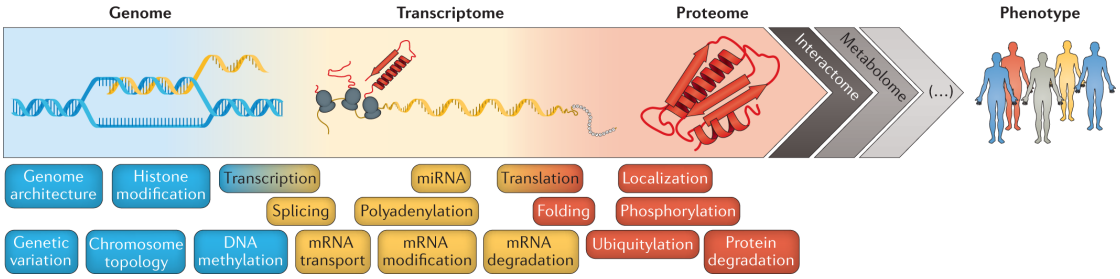



**Figure 1: Overview of the gene expression pathway.**
Visualization of the information flow from DNA (genome) to RNA (transcriptome) to proteins (proteome) in the 'central dogma' and the connections between genotype and phenotype. Across the gene expression cascade the complexity increases with many different possibilities of modifications and regulations on all three levels. Reproduced with permission from Springer Nature[10], under license 5752111506281.

Once the genetic information (DNA) is expressed in the form of proteins, it cannot be recast into nucleic acids[11,12]. A central factor of biological research is to understand the regulation of expression of specific genes and thus the fate of a cell and how gene expression adapts on internal and external factors. As the central dogma was established, novel technologies were developed to access its hierarchical levels as a whole, called *omics* technologies. Pioneering efforts to reveal base pair sequences of nucleic acids by Frederick Sanger in 1977[13], was followed by sequencing of whole organisms like the bacteria *Haemophilus influenzae*[14], yeast *Saccharomyces cerevisiae*[15], *Caenorhabditis elegans*[16], and the fruit fly *Drosophila melanogaster*[17]. This developed into massive parallelization and arrayed optical image detection, termed next-generation sequencing (NGS)[18–20] to study the genome at population-wide scale[21] but also in single cells[22]. The Human Genome Project[23] in part led by Francis Collins

solved the sequence of the whole human genome at the beginning of the millennium, revealing more than 20,000 protein coding genes[24,25] – although the last 8% of genome sequence were only recently completed[26]. The decoding of the human genome initiated the era of *genomics*, enabling to study organisms at an entirely new level of detail, which revolutionized biology, medicine and our understanding of evolution. However, as long as the functional assignments of encoded elements – the genes – remain elusive, the genome as the genetic 'blueprint', alone does not provide the translation into a biological phenotype[27]. Therefore, researchers started to analyze the next level of gene expression, the transcriptome[28], which became much more accessible by NGS. The transcribed molecules include the protein-coding messenger RNA (mRNA), the link between genome and proteins, but also a majority of non-coding RNAs which are mainly involved in gene regulation at different levels and various mechanisms[29]. However, transcripts do not represent the final product in the gene expression pathway, but rather an intermediate link to proteins. Further, gene expression can be regulated on many levels (Figure 1). Typically, genome regulation happens most prominently by histone modifications, which modulate chromatin accessibility to transcription factors, as part of the epigenome[30,31]. At the RNA level, many processes can influence stability by modification, degradation[32], or alternative splicing to form new transcript variants[33]. but other forms of RNA such as miRNA can also regulated gene expression. Importantly, all these types of regulation need to eventually influence the proteome level to have functional effects. This is because proteins are thought to be the component in the gene expression cascade that most directly determine the cellular phenotype and function as proteins are key actors in any living cell in the functional perspective – 'the workhorses'. Hence, it is highly desirable to measure proteins.

Apart from its 20,000 protein coding genes (only 3% of the whole genome) and over 10,000 different expressed genes, a single mammalian cell comprises about $10^9$ to $10^{11}$ protein molecules (about 250 pg)[34,35], while the abundance of individual proteins ranges from tens to over ten million copies per cell[36]. Interestingly, complexity vastly increases along the path of gene expression[37]. Genome alteration, alternative RNA processing and post-translational modifications (PTMs) as well as (proteolytic) cleavages of proteins increase the number of potential distinct proteins of a (controversial discussed) estimation of several million molecules[38], so called 'proteoforms'[39,40]. For illustration, a protein with five potential acetylation and six phosphorylation sites ($x2^5$ and $x2^6$) and a N-terminal acetylation (x2) could theoretically result in 4,000 different proteoforms. Importantly, many PTMs like phosphorylation are reversible and fast modifications of molecular switches driving many signaling pathways[41,42]. Although most studies focus

on individual proteins, it is important to keep in mind that most proteins do not act in an isolated manner. They are part of larger dynamic protein complexes or networks in which subtle changes influence other proteins directly or indirectly[43].

In summary, the central dogma of biology is in principle straightforward and still holds true, but has numerous control points of immense complexity with proteins being the functional units in the cell[10,44]. Therefore, analyzing a single cell – 'the fundamental unit of life', while capturing the proteome system in its entire complexity – including protein expression, abundance, localization, modification state and interaction with other biomolecules – is a central goal of proteomics and an enormous technological challenge.

## 1.2. Single-cell omics

As the cell is the fundamental unit of life, it is desirable to characterize its integrity on multiple levels to fully understand its function and phenotype. Even though the analysis of bulk samples – compromising a pool of diverse cells – reveals insights into health and disease, information about the individual contribution of cells is lost[45]. Therefore, many discoveries often only hold true in the chosen experimental conditions. Methodological and technical advancements have enabled the analysis of single cells in several *omics* modalities like genome, epigenome, transcriptome and – very recently – also the proteome. Many imaging-based techniques have been developed, like fluorescent-activated cell sorting (FACS) and fluorescence in situ hybridization (FISH) that aid single-cell analysis. Most of these generally applied methods use antibodies and fluorophores which limit them to a fixed number of targets. However, there are also broadly employed methods, such as genomics and transcriptomics, that enable a global and unbiased analysis. These methods rely on downstream amplification and sequencing.

### 1.2.1. Single-cell oligonucleotide sequencing

Single-cell genomic sequencing addresses the first layer of biological information and is applied in many research areas. An example is the application in genetic mosaicism which describes a condition in which a multicellular organism possesses more than one distinct genotype because of a genetic mutation[46]. This has mainly been studied in cancer to characterize tumor initiation, maintenance and evolution[47,48]. Applying single-cell genomic sequencing, studies have determined the clonal structure and evolutionary histories of the cancer[49]. The field is highly advanced but some challenges remain with the differentiation of true biological variation to technical noise and artifacts, as well as

throughput of cell isolation techniques and selection biases in amplification. Additionally, many modified RNA-seq techniques have been developed to study, for instance, the modalities of chromatin structure and epigenome[50]. One method called ATAC-seq (assay for transposase-accessible chromatin using sequencing) uses a hyperactive transposase to insert primers into open chromatin[51,52]. This allows the amplification and read out of accessible genome stretches. Looking at dynamic changes of gene expression, single-cell RNA sequencing (scRNA-seq) is one of the most broadly applied methods[53] which was first demonstrated in 2009[54]. Tremendous efforts in advancing RNA sequencing have enabled robust and fast measurement of cellular transcriptomes of thousands of cells to reveal biological insights into cellular heterogeneity[55,56]. scRNA-seq was designated to the 'Method of the Year' by Nature Methods in 2013 and has developed into a more and more routine technology due to accessibility and commercialization[57]. The field can be divided into microfluidic droplet-based[58–60] and plate-based transcriptomics[61,62]. Droplet-based methods are much higher in throughput but have the limitation that they characterize the RNA from the 5' to 3' end of the transcript, neglecting the detection of isoforms at the 3' side. scRNA-seq has proven itself in many studies relevant in health and disease[63,64] and has created a plethora of cell atlases to characterize cell types and cell states in different organs[65–67]. These large-scale projects have driven bioinformatic tool creation for *omic* technologies in general, such as normalization techniques to mitigate systematic biases by batch effects[68].

Although transcriptomics is a global measurement, it only classifies a system at a given time. However, the temporal dynamics and trajectory of gene expression change, called RNA velocity, can be reconstructed by analyzing the ratio of the spliced and unspliced mRNA transcript[69–71].

Single-cell characterization has nearly been synonymous with scRNA-seq, since RNA is easy to isolate and work with. However, copy numbers of RNA vastly vary across cell type and biological status. Many transcripts have low-level expression (5-20 transcript copies per cell)[72,73], which can be less than one transcript on average in a steady state condition[74]. A cell contains less than 1 pg mRNA, which has to be amplified after extraction[75]. This has the result that RNA has high stochastic fluctuation in the number of transcripts per gene and per cell (so called Poisson or shot noise). One other interesting aspect is the comparison of rapid proliferation and quiescence on gene expression. In an example of fission yeast mRNA numbers decreased by 70%, whereas protein numbers were only reduced by 9.5%[76]. As a general conclusion the proteome tends to be stable while the transcriptome is more variable[77,78]. This is one of many

reasons to complement single-cell omics by investigating biological function also by single-cell proteomics.

### 1.2.2. Single-cell proteomics

As mentioned above, since there is no direct relationship and interaction between abundance of transcripts and proteins, it is highly desirable to measure the proteins as the functional units in a cell. Additionally, biological changes are usually subtle and not on-off switches, which is why a quantitative approach is needed. Many different methods have evolved to quantify proteins. Most are based on affinity probes against proteins, mainly by antibodies. The classical approach is the direct readout of the antibody signal with fluorescence staining and FACS. Although it is a widely applicable technique, this approach is limited by multiplexing for up to 60 proteins in practice[79–81]. CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing)[82] or also less known as REAP-seq[83] are extensions of the droplet-based scRNA-seq analysis that enable the measurement of surface proteins with the use of antibodies conjugated to combinatorial nucleotide tags, which are then sequenced. Since any excess of antibody will result in signal, extensive wash steps and background controls are needed, which limit the use of this method to surface proteins. There have been major efforts to develop methods for intranuclear[77,84] and intracellular cytoplasmic compartments[85]. However, analysis of intracellular protein targets remains challenging. Companies like Olink[86] and SomaLogic[87] apply similar methods to clinically relevant samples like plasma in a high-throughput manner and variants of their technologies are also used in single cells[88,89].

In mass cytometry, proteins are imaged and ionized by a multiplexed ion beam[90], which may even achieve subcellular resolution[91]. A specific application of mass cytometry is called cytometry by time of flight (CyTOF). Here, targeted proteins are labeled by antibodies coupled to heavy metals – usually lanthanides – which are then ionized in an argon plasma and quantified by time-of-flight mass spectrometry[92,93]. CyTOF is, however, limited by its availability of both antibodies and heavy metal isotopes to 70 targets[92,94–96].

Although those techniques have a relatively high throughput, one limitation of these affinity probe-based single-cell technologies is their potential cross-reactivity to different targets and between each other. Therefore, many controls need to be validated and multiple antibodies for one protein target should ideally be included to ensure quantitative accuracy. Additionally, the limited availability of affinity probes and their need of extensive characterization may add expense and be time-consuming[97]. Finally,

using affinity probes in itself restricts the method to a targeted set of specific proteins, which limits the attraction as a discovery approach.

Therefore, single-cell proteomic analysis would clearly benefit from an unbiased, untargeted and proteome-wide approach. Mass spectrometry (MS) is a unbiased method as it measures all proteins within its range of detection[98] and would therefore be an ideal candidate. In tissues and cell lines, 10,000 proteins can now be quantified in a single LC run and even more after initial fractionation[99–101]. The field of single-cell MS-based proteomics (sc-proteomics) has been on the rise over the last years and driven by several major developments. Its major challenge is sensitivity since in contrast to other methods, MS measures the proteins directly without any amplification step. Therefore, all steps of the MS-based workflow need to be optimized to retrieve the highest possible signal. Of the total protein amount in a single cell (about 150 to 250 pg in mammalian HeLa cells[34,102]), as little as possible should be lost in the sample processing. This has been addressed early on in the field by mixing in a hundred-fold more abundant 'carrier channel' in a method termed single cell proteomics by mass spectrometry (SCoPE-MS)[103]. Despite the conceptional attractiveness, it has many intrinsic issues, which are discussed in the section quantification strategies (1.3.6.1) and as a result, few biological applications have been published. Most technologies that achieve better quantification accuracy have used label-free approaches and are therefore limited to one cell at a time. Note that even with multiplexing MS-based proteomics does not have the throughput of scRNA-seq. Below, I will discuss the principal steps in MS-based proteomics in relation to the development of sc-proteomics.

## 1.3. Mass spectrometry-based proteomics for single-cell proteomics

Since the crucial invention of electrospray ionization (ESI) in the 1980s, the field of liquid chromatography-mass spectrometry (LC-MS) has been developing at a tremendous speed[104–106]. ESI enables the transfer of biomolecules from the liquid into the gas phase in their intact form. Ions are dragged by the gas flow from ambient conditions into the vacuum of a MS which measures mass-to-charge (m/z) ratios of ions. This results in high specificity and sensitivity – even down to the detection of single ions[98,107]. In principle, MS-based proteomic can be applied to intact proteins (termed 'top-down'), to larger parts of them ('middle-down') and finally and most commonly to peptides ('bottom-up' proteomics). In top-down, analyzing the protein as a whole directly reveals information on proteoforms and their native constitution. While this promises complete sequence coverage of proteins, top-down approaches are challenged by chromatographic separation due to the complexity of proteoforms as well as sensitivity

due to poor ionization of full proteins and signal loss into highly complex charge and isotope states[39,40,108]. Because of these reasons, middle-down approaches have been deployed to mitigate the limitations of top-down proteomics[109]. Despite these attractive principles, by far the most common method is bottom-up proteomics because of its sensitivity, proteome coverage, throughput and unbiased analysis capabilities for discovery proteomics.

### 1.3.1. Bottom-up proteomics

Bottom-up proteomics has developed into the most prominent approach in proteomics due to many breakthroughs and has now become nearly universally applicable to all kinds of proteomic samples and research questions[98,110]. In this approach, proteins are enzymatically cleaved into peptides by sequence-specific endopeptidases. Observed mass spectral signatures are compared to theoretical peptide masses and the fragment spectra of *in silico* proteolytically digested proteins of reference databases to identify spectral matches of peptides (see 1.3.6 Computational proteomics). Finally, protein identity is inferred and quantified by reassembling identified peptides onto a reference proteome to enable biological interpretation in downstream bioinformatic analyses[111].

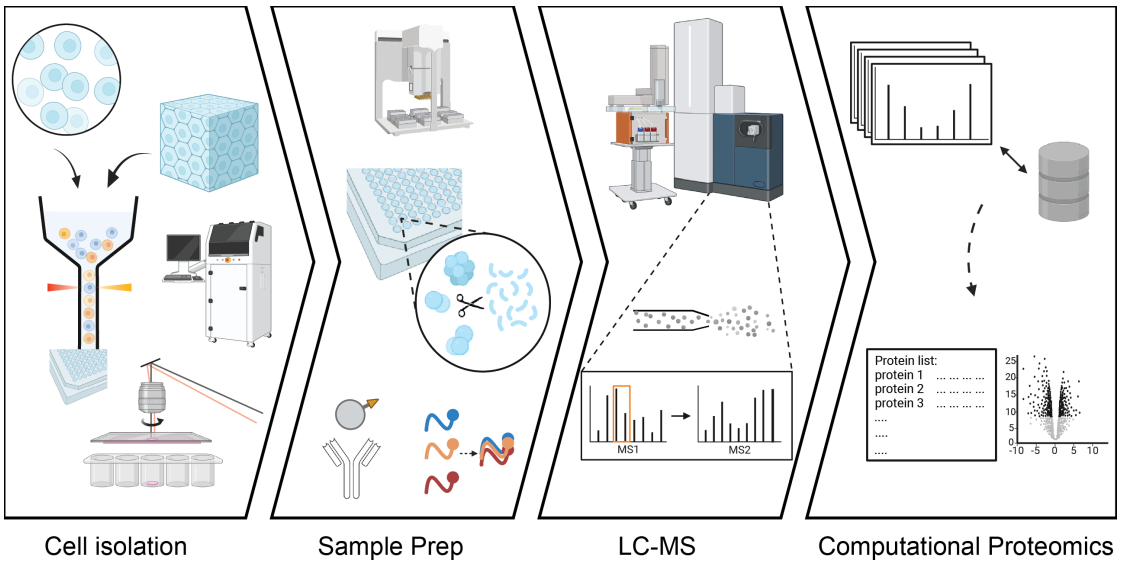



**Figure 2: Schematic of bottom-up MS-based proteomics workflow.**
Single cells, exemplary shown as biological sample, need to be isolated. Methods can include FACS, laser-capture microdissection or other single-cell producing methods. Single cells are lysed, enzymatically digested and processed in a miniaturized sample preparation format. Automated liquid handling systems are used to improve robustness and throughput. Digested peptides can be subjected to multiplexing mass tag reagents or any additional enrichment approaches such as phosphopeptide enrichment for PTM analysis. Processed peptides are transferred into the mass spectrometer (MS) using electrospray ionization (ESI) after complexity reduction and concentration via liquid chromatography (LC). Spectral information of MS1 and MS2 level is searched against a reference database for peptide identification and quantification. Subsequently, peptides are reassembled into proteins, whose quantification is used for bioinformatic analysis and data visualization. Created with BioRender.com.

The bottom-up proteomics workflow consists of sample preparation for peptide generation, liquid chromatography to decrease the complexity of the sample and MS-based detection, followed by computational analysis of the intensive and complex raw data as well as the downstream bioinformatic analysis for biological interpretation (Figure 2). Sc-proteomics requires overcoming key technological challenges and achieving best performance in all of those areas: lossless sample preparation, high-performance chromatography, ultra-high sensitivity MS measurements and optimal analysis of the low-level signals for quantification.

### 1.3.2. Sample preparation

The sample preparation is a crucial step in the bottom-up proteomic analysis. It is sample dependent and has to be optimized for sample type, initial amount, use case and the experimentally addressed question. Many different sample preparation protocols have been developed like filter-aided sample preparation (FASP)[112], protein aggregation capture (PAC)[113], suspension trapping (S-trap)[114] and solvent precipitation (SP3)[115]. In all cases the major steps include: cell lysis, protein solubilization, potential reduction of disulfide bonds and alkylation of free cysteines, specific proteolytic digestion by proteases and a potential peptide purification for MS-based analysis. Thermal treatments achieve cell lysis with temperatures between 70 and 95 °C or freeze-thaw cycles, but mechanical treatments like sonication are applied as well. Protein solubilization is accomplished by detergents like sodium dodecyl sulfate (SDS) or sodium deoxycholate (SDC), chaotropic reagents like urea or guanidinium chloride (GdmCl), or organic solvents like acetonitrile (ACN). This is achieved in combination with a buffer system to maintain a constant and favorable pH, usually Tris (tris(hydroxymethyl)aminomethane), TEA (tris, acetic acid and EDTA) or TEAB (triethyl ammonium bicarbonate). Typically, samples are reduced and alkylated by the reducing agents TCEP (Tris(2-carboxyethyl)phosphine hydrochloride) or DTT (dithiothreitol) and IAA (iodo-) or CAA (chloroacetamide) for alkylation of free cysteine to prevent disulfide bridge reformation. After successful protein solubilization and denaturation, specific proteases, mainly trypsin and LysC, but also others like LysN are used to cut the proteins into peptides. The typical length of tryptic peptides are 7 to 40 amino acids, with an average of 12 observed in the MS. LysC and trypsin are endopeptidases that hydrolyse proteins specifically at the carboxyl side of lysine and additionally arginine, respectively. In contrast, LysN cuts at the amino terminal side of lysine. The digestion step is important since missed cleavages might entail the loss of signal by the formation of a non-tryptic peptide and thus the signal reduction of the correctly cut peptide. Depending on the reagents used in the sample preparation, digested peptides may need to be

purified to remove salts, detergents, chaotropic agents and insoluble aggregates to prevent ionization suppression and the build-up of debris on hardware components in the MS. This is typically achieved by solid phase extraction (SPE) with various materials[116], for instance C18 and it is the primary point where sample loss can occur during sample preparation.

In sc-proteomics, the sample preparation is especially crucial since all proteins of a cell need to be retrieved in the best way to reach the highest signal possible in the MS. Therefore, the purification step is usually minimized by using MS-compatible reagents in the sample preparation. The main buffer system is TEAB, which is volatile. Additionally, its buffer system has a pH of 8.5 where trypsin and LysC have their optimal activity. Different varieties have been described for MS-compatible detergent reagents like MaSDeS[117], although most prominently dodecyl-β-D-maltoside (DDM)[118] and ACN are used. A further advantage of DDM at low concentrations is that it prevents protein losses by reducing plastic binding[119]. Another crucial feature is the miniaturization of the sample preparation. Many studies have used 384-well plates[78,120–122], but microfluidic devices have also been developed[123,124]. A pioneering approach in miniaturization and reduction of sample losses was nanodroplet processing in one pot for trace samples (nanoPOTS)[125]. Protein losses to plastic can also be circumvented by using inert materials like Teflon and glass instead of plastic. A general strategy in sc-proteomics is to simplify the sample preparation to one-pot reaction protocols[119,126]. This enables automated liquid handling systems to increase robustness and throughput in sample preparation. Similarly, transfer steps need to be avoided or reduced to a minimum to limit the duration and surface area of plastic contacts[126]. Sc-proteomics as a field has only recently developed, but there are already some commercial products for sample preparation available. One example is the cellenONE (Cellenion) which uses a piezo liquid handling technique to dispense picoliters of volume[127]. It also has a humidified chamber to ensure that low volume droplets of several hundred nanoliters do not evaporate, leading to higher reproducibility. Building on this product, there have been developments to decrease transfer steps to a bare minimum using a chip (‘proteoCHIP EVO 96’) to spin the samples onto C18 tips[128,129], which are then able to be transferred and measured on the LC-MS.

The final peptide mixture can also be subjected to additional modifications dependent on the experimental question and MS acquisition approach. One possibility is chemical labeling with specific mass tags that are attached to peptides. Using mass tags with different isotopic composition enables the labeling of peptides by different defined masses. This allows the combination of multiple samples in one MS run, referred to as

multiplexing. This is used in sc-proteomics to increase the throughput. However, proteomics does not reach the massive multiplexing of NGS, which employs oligonucleotide barcoding. For further details about multiplexing in proteomics see 1.3.6.1 Quantification strategies.

Additionally, different enrichment approaches can be applied to the processed peptide mixture. Enrichment ensures that the analyte is less complex and that the MS measurement time is focused on the peptides of interests. One example is the enrichment of PTM-modified peptides. Most prominent is the analysis of phosphopeptides[130] by immobilized metal affinity chromatography (IMAC)[131] of iron-NTA or TiO2[132] which can be automated with the AssayMap[133]. Another example is immunoprecipitation (also known as affinity pulldowns) by specific antibodies to analyze protein-protein interactions or the HLA peptidome. Although this has not been applied to sc-proteomics yet because of sensitivity issues, this might become an interesting analysis in the future.

### 1.3.3. Liquid chromatography

Liquid chromatography (LC) serves as powerful tool in proteomics to separate the complex mixture of peptides. They are most often separated by an increasing gradient of a non-polar solvent (typically ACN) and a nonpolar stationary phase (e.g. porous silica beads modified with C18 alkyl chains), also called reverse-phase chromatography[134]. Dependent on their polarity, the peptides are differently partitioned between the column material and the solvent ACN.

In sc-proteomics, the LC is challenged on different levels to achieve highest resolution, accuracy and sensitivity. This is important since efficient chemical separation reduces the complexity of the sample composition – specifically separating chemical background from single cell derived peptides, which expands the dynamic range of the measurement and reduces ionization of background. One key aspect is to maximize the peptide concentration and avoid dilution of the sample, while maintaining sharp peaks for high peptide signal in the MS. One factor is the liquid flow speed during peptide separation. At the other extreme, in high sample amount applications in large scale industrial settings, high flow rates are used to improve chromatic performance according to the van Deemter equation[135,136] for reproducibility, throughput and stability[137,138]. However, sample amounts are extremely limited in sc-proteomics. At high flow rates, more total volume of solvent would be used, which results in diluting the peptide concentration of the sample. Overall, reducing the flow rate therefore maximizes the peptide concentration and signal. Many studies have used low flow gradients with less than

100 nL/min[78,139,140] or even 20 to 50 nL/min[125,141]. Flow rates lower than 1 nL/min with custom-made setups and small-bore emitters of 2 µm are also possible[142], however, this comes at the price of reproducibility and throughput. For instance, in this case, sample introduction into the LC system is not straightforward[143]. Another fundamental aspect of low flow rates is the transfer of peptides from the LC to the MS. Here, peptides need to enter the gas phase via electrospray ionization (ESI), which is flow rate dependent. Low flow approaches have a higher ionization efficiency resulting in an improved number of peptides that enter the MS as charged ions. Together, low flow rates lead to a maximized peptide concentration in the chromatographic system and increase the ESI efficiency, resulting in higher ion signal and thus an increased sensitivity. One disadvantage of low flow gradients is the limited throughput since higher flowrates achieve better separation in shorter time and the overhead time of equilibration and sample loading is prolonged. Different approaches of introducing trap columns[144] or multiple analytical columns[140] for parallel equilibration and sample loading have been developed to increase the throughput by decreasing the overhead time. Another approach to enhance peptide signal has been to use solvent additives to improve electrospray ionization, like DMSO[145] or 0.5% acetic acid[146,147].

The main liquid system instruments used in sc-proteomics are the Evosep One (EvoSep Biosystems), Vanquish Neo (ThermoFisher Scientific) and nanoElute 2 (Bruker Daltonics). The Vanquish Neo is a highly flexible system in which all parameters of the gradient can be adjusted. It works in a wide range of flow rates from 100 nL/min to 100 µL/min. Additionally, it has very robust high-pressure pumps that allow for fast equilibration of the system and sample loading at high flow rates before separating the peptides with a low flow rate, keeping the overhead time at a minimum. The Evosep One is a standardized high-performance instrument and comes with fixed standard microflow gradients, but also low flow applications[148]. Peptides are stored on single-use C18 StageTips[116,149,150], termed Evotips, from which they are eluted into the system with up to 40% ACN[148,151]. This approach ensures minimal carryover and it retains contaminations as well as insoluble material on the tip. The Evosep One has been crucial to many single-cell studies because of the robust low flow gradients called ‘Whisper’ and the elution of peptides in ‘nano-packages’ of 20 nL into a preformed gradient. This gradient allows for *in situ* gradient admixture retaining the sample concentration at a maximum while maintaining the full control of the flow rate across the gradient[148]. Additionally, the Evotip solves the issue of peptide transfer from sample preparation to the LC system by efficiently storing the peptides until measurement on the C18 material of the tip. In contrast, peptide mixtures stored in autosamplers adhere

to sample vials and are exposed to evaporation and thus undergo losses or even disappeared before analysis[152]. Some approaches have used direct transfer on capillaries, but those suffer in the ability to be automated and in throughput[125].

Different columns can be used for sc-proteomics. Parameters like length, inner diameter, backpressure, price, robustness, carrier over and low-dead volume need to be considered. Overall, the longer the column the better the separation, but this reduces the throughput, necessitating a compromise. In the last years, some commercially available columns have pushed the field forward such as the Aurora (IonOpticks) and µPAC (ThermoFisher) products. Most common columns are bead-packed silica columns such as the Aurora. Standard inner diameters (ID) are 75 µm for normal nanoliter proteomics and for low flow applications. However, very narrow columns have already shown great sensitivity, though being difficult to manufacture with high reproducibility and low-dead volume[141]. Low-dead volume columns and connections are crucial to avoid peak broadening[153]. Given the difficulties with extreme miniaturization of bead-based packed columns, multiple alternatives have been explored. One example are porous layer open tubular (PLOT) columns, which utilize in situ-polymerization as stationary phase[154,155]. Another column type are microfluidic pillar array columns (µPAC, Thermo Fisher Scientific). µPAC columns have lithographically etched micro pillars, which are made porous and coated with C18 for peptide separation[156–158]. This reduces dispersion and provides highly efficient separation at lower operation pressures, which leads to less variability since it is less affected by pressure changes. Additionally, the µPAC column has proven to have low carryover between sample runs, which is a key aspect in sc-proteomics[159,160].

Capillary electrophoresis (CE) is an alternative to LC for peptide separation. It is based on differential migration rates of ions in a liquid in an electric field dependent on the charge, viscosity and atom's radius[161]. Detection limits have been reported to be close to one zeptomol ($10^3$ attomol = $10^{-21}$ mol)[162], while LC systems are around hundreds zeptomol[163]. This can be achieved since CE does not have adsorptive losses due to a stationary phase. Although major challenges are being addressed, like ionization[164], CE still faces issues with less natural cleanup of the sample compared to LC and the need for preconcentration prior to gradient elution.

### 1.3.4. Mass spectrometry

Conceptually, the MS consists of an ion source, mass analyzer and a detector. The ion source is at ambient pressure while the MS operates under vacuum which is needed to keep the ions on stable trajectories enabling high transmission efficiency and resolving power.

As previously described, ESI is used as the ion source in proteomics. Following the separated biomolecules in liquid phase continuously eluting from the LC, the peptides are continuously ionized and transferred into the gas phase without damage. At the interface between LC and MS, a thin needle capillary (also called emitter) is subjected to an electrical field of high voltage. The excess of the same charge on the liquid at the emitter causes it to form a Taylor-cone which disperses the peptide-containing droplets towards the MS. While the solvent rapidly evaporates ('desolvation'), also facilitated by a high-temperature carrier gas, ESI forms multi-charged intact molecular ions that migrate into the vacuum chamber of the MS. A smaller initial droplet size increases the efficiency of ESI[165], which is the case for low flow rates and compounds increasing the conductivity, e.g., acetic acid, as discussed before. While a higher ESI efficiency means more charged ions entering the MS, this eventually improves sensitivity. The same is true when opening the diameter of the orifice. Additionally, closed sources increase the desolvation efficacy and are thus beneficial for sc-proteomics applications[166]. At this state, the peptides are called precursors, split into multiple charge states and modifications, in contrast to the fragmentation products. The tryptic precursors are predominantly double positive charged.

The MS efficiently focuses and transfers the precursor through vacuum via static or dynamic electric fields, usually in radio frequencies (RF). The ion transfer has improved through many contributions over the years, which has greatly contributed to the sensitivity of the instruments[167]. Next, precursors are analyzed by the mass analyzer and quantified by the detector by their m/z (termed MS1 scans). The precursor mass alone does not explicitly identify a peptide sequence; thus, fragmentation is needed, so called tandem-MS (MS/MS)[168]. Precursors of different m/z are sequentially selected by a quadrupole and fragmented by collision-induced dissociation (CID). The quadrupole transmits selected ions on oscillating trajectories through periodically alternating frequencies on four cylindrical rods, while others are not stable and collide with the rods[169,170]. The fragments are read out in MS2 scans. The type of generated fragment ions depends on the fragmentation method (Figure 3). In CID of tryptic peptides b- and predominantly y-ions are generated by collision with accelerated inert gas molecules

such as nitrogen. Different acceleration energies are used for efficient fragmentation depending on mass, charge state and sequence. Besides CID, many other fragmentation methods have been developed with various principles, like high collision dissociation (HCD), which is very similar to CID, electron transfer dissociation (ETD) and ultraviolet photodissociation (UVPD)[171]. ETD forms c- and z-ions by the transfer of electrons, which is advantageous for intact protein and PTM analysis as it leaves labile modifications intact such as phosphorylation[172–174]. Ideally, this results in a full ion series of the whole peptide, increasing the chances for confidently matching modification sites. For the MS1 and MS2 measurements, several concepts and hardware have been established. In sc-proteomics, the aim is to make use of as many of the ions entering the MS as possible. The efficiency in collecting spectra during an acquisition cycle is called duty cycle and it determines the percentage of ions used.

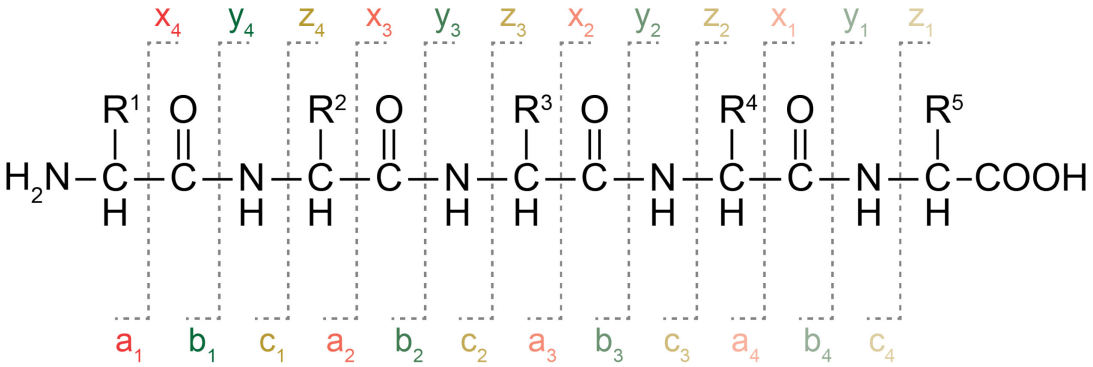
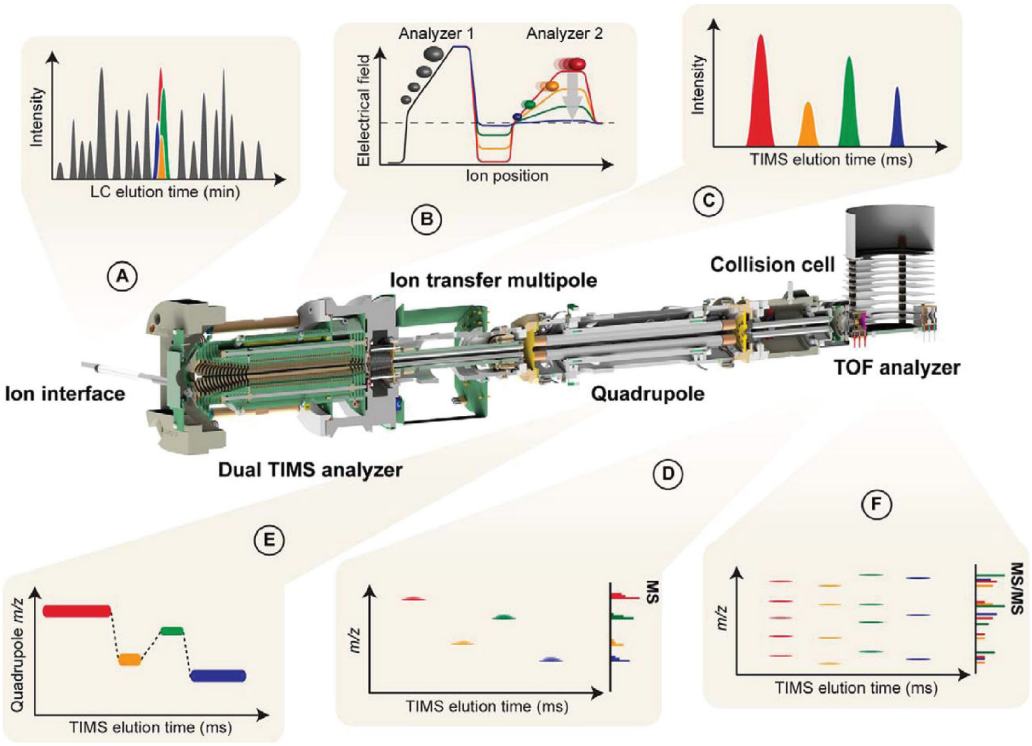


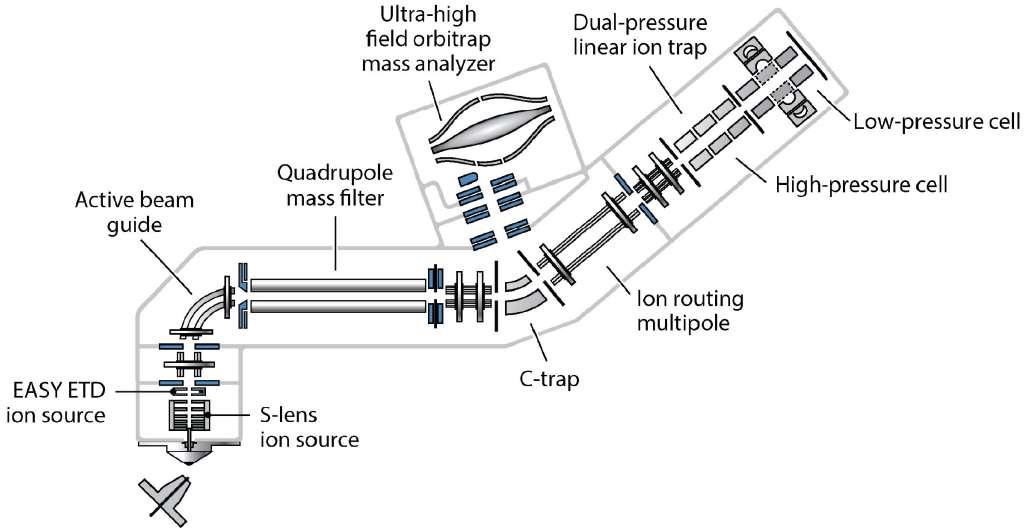

**Figure 3: Different fragmentation pattern of the peptide backbone for MS2 spectra.** Fragmentation occurs at the peptide backbone between different chemical groups of amino acids. The fragments can be classified several different ion series according to the Roepstoff-Fohlman nomenclature[175]. N-terminal charged fragments are called a-, b-, c-ions while C-terminal ones are x-,y-,z-ions. The fragmentation at the peptide bond generates b- and y-ions, while the breakage within an amino acid on the amino group site produces a- and x- ions and on the carboxy group site c- and z-ions. Different fragmentation methods produce specific ion series dependent on their fragmentation principles. For example, CID forms b- and predominantly y-ions. Figure was adapted from Roepstorff and Fohlman[175].

Many developments have been made in mass spectrometers that can be used in proteomics, constantly pushing the boundaries[167]. The main instruments for the development of sc-proteomics have been the timsTOF platform (Bruker Daltonics) with the timsTOF Pro2, SCP and Ultra (Figure 4a)[78] as well as Exploris[176], Orbitrap Tribid MS like the Fusion (Figure 4b) and – very recently – the newly introduced Orbitrap Astral[177–179] (all from Thermo Fisher Scientific)) (Figure 4c). These instruments use the orbitrap analyzer and time-of-flight (TOF) analyzer as main mass analyzer, respectively.

**a**


Intensity
LC elution time (min)
Analyzer 1
Analyzer 2
Elelectrical field
Ion position
Intensity
TIMS elution time (ms)
A
B
C
Ion transfer multipole
Collision cell
Ion interface
Quadrupole
TOF analyzer
Dual TIMS analyzer
D
E
F
Quadrupole m/z
TIMS elution time (ms)
m/z
MS
TIMS elution time (ms)
m/z
MS/MS
TIMS elution time (ms)

**b**


Ultra-high field orbitrap mass analyzer
Dual-pressure linear ion trap
Low-pressure cell
High-pressure cell
Quadrupole mass filter
Active beam guide
Ion routing multipole
C-trap
EASY ETD ion source
S-lens ion source

**c**

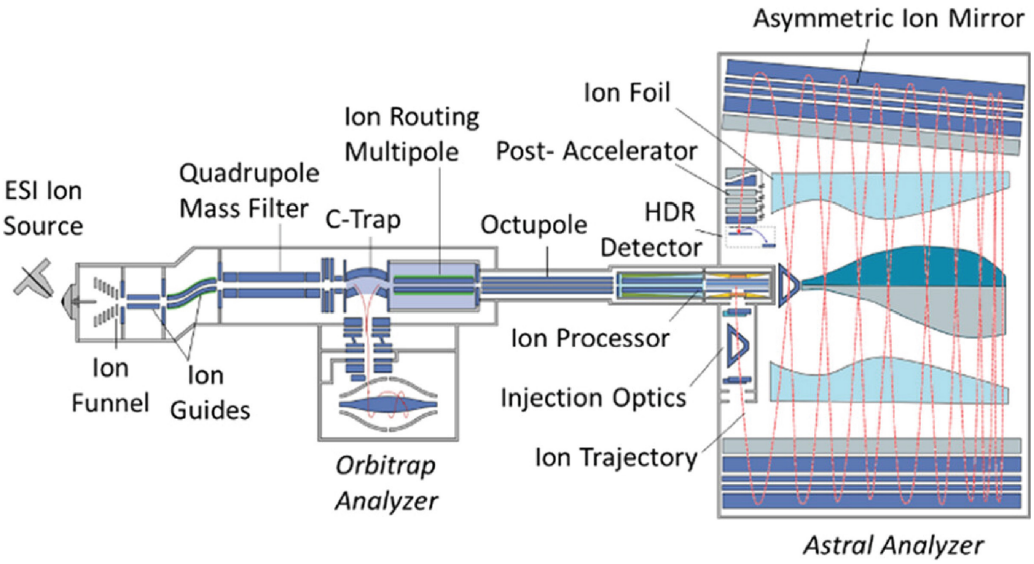

Asymmetric Ion Mirror
Ion Foil
Ion Routing Multipole
Post- Accelerator
ESI Ion Source
Quadrupole Mass Filter
C-Trap
Octupole
HDR Detector
Ion Processor
Ion Funnel
Ion Guides
Injection Optics
Ion Trajectory
Orbitrap Analyzer
Astral Analyzer

**Figure 4: Ion path of different mass spectrometers and their MS analyzer principles.**
The detection of ions in both MS1 and MS2 scans is achieved through different methods. After electrospray ionization, ions are directed into the MS in vacuum, where they are transmitted and focused by ion guides and undergo mass selection by the quadrupole.
a. The timsTOF platform from Bruker Daltonics includes an additional separation of ion. Through the combination of trapped ion mobility separation (TIMS) and TOF, it incorporates a principle known as Parallel Accumulation – Serial Fragmentation (PASEF). Ions are separated and sequentially eluted based on their ion mobility in the TIMS analyzer, which is synchronized with the quadrupole. Ion detection is facilitated by a fast TOF analyzer which records the ions as a function of their time-of-flight after acceleration in an electric field. The schematic was reprinted as permitted by the CC-BY 4.0 international license in Meier *et al*.[180].
b. In the Orbitrap Fusion Tribid MS (Thermo Fisher Scientific), ions are accumulated in the C-trap, before being transferred into the Orbitrap analyzer. This principle decouples the Orbitrap analyzer from the continuous ion beam. The orbitrap measures the ion mass as their frequency of oscillation around the main central electrode. Additionally, the Tribrid instrument series include a dual-pressure linear ion trap which enables MSn ion detection. It facilitates high acquisition rates and experimental flexibility in MS scheduling in combination with multiple fragmentation modes (including CID, HCD, ETD and UVPD). Reproduced with permission from Annual Reviews[167], under license 1462612-1.
c. In the Orbitrap Astral (Thermo Fisher Scientific), the orbitrap is similarly used to record MS1 scans. For MS2 scans, ions are transitioned from an ion routing multipole to the ion processor, similar to the C-trap. Fragments, produced through higher-energy collisional dissociation (HCD) are separated based on their m/z ratio in the Astral analyzer. This separation happens in a multi-pass reflector flight path between two asymmetrical ion mirrors. The ions are then detected by the High Dynamic Range detector. The schematic was reprinted as permitted by the CC-BY 4.0 CC-BY-NC-ND 4.0 international license in Stewart *et al*.[160].

#### 1.3.4.1. Orbitrap mass analyzer

The orbitrap analyzer, developed by Alexander Makarov[181] and commercialized in 2005, traps ions around a central spindle electrode which confines them oscillating along its longitudinal axis[182]. The frequency of oscillations is proportional to the m/z of the ions and is retrieved by Fourier transformation. The resolution of the analyzer improves with longer oscillation times of the ions around the spindle – also called 'transient' – until de-phasing due to subtle ion collisions limits it[183]. For example, the current orbitrap needs 256 ms transient time for a mass spectrum with a resolution of 120,000 [184]. In modern instruments, the orbitrap analyzer is combined with the C-trap, also an ion trap, which decouples the orbitrap from the continuous ion stream. Ions are accumulated and stored in the C-trap and then transferred into the orbitrap analyzer in nanosecond pulses. While the analyzer requires a certain number of ions to obtain an adequate signal, performance is decreased by overloading the analyzer with too many ions. Therefore, an automated gain control (AGC) parameter is calculated on short pre-scans or estimated on earlier acquired scans, which is used to determine the appropriate fill time to obtain an optimal ion number.

In routine proteomics applications, achieving the set AGC target is generally limited by the maximum injection time. This maximum injection time is the upper time limit allowed for ion accumulation before subsequent injection into the mass analyzer, even when the AGC target has not been reached. In sc-proteomics, the maximum injection time is another critical parameter, which is usually set fairly high to compensate for the low

sample amounts and ensure sufficient ion collection for meaningful detection. This requires adaption of the acquisition method for the sample and the injection amount. By its nature sc-proteomics deals with very limited material and therefore decreased ion fluxes, which results in the need of long MS2 ion injection times to achieve a quantifiable signal. Longer injection times using the same size of quadrupole isolation windows reduces the duty cycle and thus productive MS2 spectra. This can be overcome by using wider windows for MS2 acquisition. While this usually increases spectral complexity and consequently decreases the identification of peptide fragment ions, this is not a major issue in sc-proteomics due to the significantly lower complexity of single-cell spectra compared to higher input samples. In the special case of SCoPE-MS (see below), acquisition times have to be extended beyond an adequate signal for the carrier channel, so that there is a sufficient number of peptides from the sample channels[122]. The orbitrap analyzer with its fairly long transition times is limited by speed in many applications even in sc-proteomics. Therefore, researchers have used various techniques to increase the duty cycle by using the linear ion trap as MS2 detector[185]. In those approaches, the orbitrap is used as mass analyzer for high-resolution MS1 survey scans and the linear ion trap for sensitive lower-resolved MS2 scans.

#### 1.3.4.1. Time-of-flight mass analyzer

Time-of-flight (TOF) analyzers measure the time ions need to reach the detector after their m/z-dependent acceleration in an electric field. The resolution of TOFs increases proportionally with the length of the flight tube and is independent of m/z, in contrast to an orbitrap, where it is inversely proportional to the square root of m/z[181]. TOFs have high frequencies allowing for high scanning speeds to sample the incoming ions, but are usually limited in the percentage of ions that they can use. Typically, individual acquired mass spectra are accumulated to reach sufficient signal-to-noise ratios[186,187].

A new mass spectrometer, called Orbitrap Astral, has just been released in 2023[160,177,179]. It consists of an orbitrap mass analyzer and a TOF, a so-called Asymmetric Track Lossless (Astral) analyzer[178]. In the Astral, asymmetric ion mirrors and ion foil guide and focus the ions along their 30-meter multi-reflector path, which features reflections for increased spectral resolution and minimal ion losses. A high dynamic range (HDR) detector with single-ion sensitivity enables extremely sensitive and low-noise detection. The addition of the Astral analyzer allows for the MS2 acquisition of high-resolution accurate mass (HRAM) scans at a scanning speed of up to 200 Hz in parallel to the MS1 scans in the orbitrap. While the Astral analyzer is currently only used for MS2, there is promising potential that it may also be implemented

for MS1 scans, allowing for very high sensitivity. Together, this new instrument opens up new avenues for sc-proteomics[128,188–190].

The timsTOF platform uses only a TOF analyzer. It is connected with an upstream trapped ion mobility separation (TIMS) tunnel (see next section) which perfectly aligns with LC and TOF analyzer timings. To make this instrument suitable for sc-proteomics, several developments were necessary, culminating in the timsTOF SCP mass spectrometer[78]. This instrument was co-developed by our laboratory with Bruker Daltonics and enabled a higher ion transmission and other optimizations, increasing sensitivity more than four-fold. Together with a low flow chromatographic system of 100 nL/min, the sensitivity increased by another ten-fold, making label-free sc-proteomics possible. Further improvements let to even more increased sensitivity, which included scanning speed up to 300 Hz for MS2 with the current model, the timsTOF Ultra.

#### 1.3.4.2. Ion mobility spectrometry

There are different techniques to separate ions by their mobility of which trapped ion mobility spectrometry (TIMS) is of special interest. In TIMS, ions are trapped with an electric field against a constant nitrogen flow. According to their mobility, ions can move more or less with the electric force against the nitrogen flow and less mobile ions are stacked closer to the exit of the tunnel (Figure 5a). According to this principle, ions are separated by inverse ion mobility in the TIMS tunnel along the electric force gradient and can then be gradually eluted by lowering the electric field. One important feature is that the TIMS can be in line with a TOF analyzer in a compact set up. Note that ion mobility and m/z is closely related. The TIMS separates the ions in milliseconds which is a great fit after the LC in which peaks elute in the second range, while the TOF acquires scans at microseconds time scale. Together, the time scale of the TIMS elution is aligned with the TOF in an acquisition strategy called Parallel Accumulation–Serial Fragmentation (PASEF)[180,191]. In PASEF, precursors are accumulated in a first TIMS, sequentially released from the second TIMS tunnel and then selected for quadrupole mass selection, fragmentation and TOF acquisition. This parallelization of TIMS accumulator and analyzer as well as the TOF, enhances the duty cycle to nearly 100%, thus efficiently using all ions entering the MS. This is very advantageous for sc-proteomics. By adding another dimension of separation, TIMS helps to resolve the complexity of the proteome.

Ion mobility can also separate ions with different charge state. Note that single charged ions are primarily contamination and solvent clusters (except in certain cases as in

peptidomics)[192]. At low sample concentrations, typical for single cells, these interfering signals can become predominant. This increases spectral complexity and occupies finite charge capacity, limiting the storage and accumulation of sample ions. Therefore, it is desirable to remove them.

Field asymmetric ion mobility spectrometry (FAIMS) is an ion mobility filter device, which is often used in sc-proteomics together with trap-based instrument. FAIMS is operated at atmospheric pressures and separates ions based on their differential mobility in alternating strong and weak electric fields (Figure 5b). It uses a parallel gas flow to move ions along the alternating electric fields while the strong field has shorter time intervals than the weak field. When the ion mobility in both fields is the same, the ions pass through the device. A compensation voltage (CV) is applied to the planar electrodes to separate ions and transmit specific ions. Differential mobility depends on a number of properties; thus, it is more orthogonal to the m/z dimension than TIMS. FAIMS does not selectively filter out single charged ions but it can enrich for multiple charged ions and remove interfering ion species. Therefore, FAIMS improves dynamic range and detection limit despite being a filtering device unlike the TIMS.

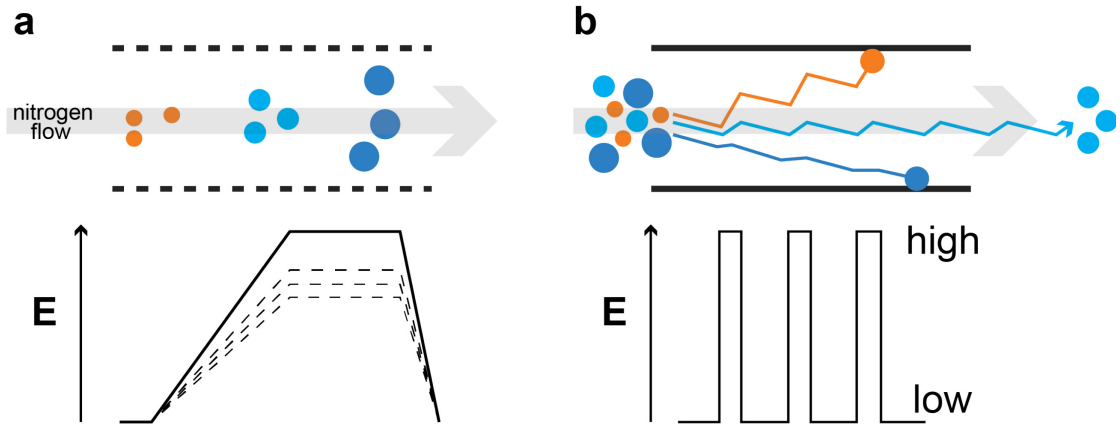



**Figure 5: Ion mobility principles of TIMS and FAIMS.**
a. Trapped ion mobility spectrometry (TIMS) separates ions based on their ion mobility in an electric field against a constant carrier gas flow. Mobile ions are dragged along the electric force, while low mobile ions not, therefore aligning the ions along their inverse mobility. By lowering the electric field, ions are released.
b. Field asymmetric ion mobility spectrometry (FAIMS) separates ions based on their differential ion mobility in alternating high and low electric fields. Along the path a gas flow drags the ions through. Ions pass, if their ion mobility is the same in both fields. By applying a compensation voltage (CV), ions are separated and selected differential ion mobility ions are transmitted. Figure was adapted from Winter *et al.*[193].

### 1.3.5. Data acquisition strategies

Over the years, many different data acquisition strategies have been developed. Dependent on the experimental approach, aim of the study, samples and the MS instrument, the optimal data acquisition strategy is chosen. The main three strategies are: data-dependent acquisition (DDA), targeted acquisition and data-independent acquisition (DIA) (Figure 6). One important metric in data acquisition is the cycle time. It is the time until the next MS1 survey is measured. This time determines when a precursor eluting as a peak from the LC is sampled again. A major factor in method design is the compromise between identification and quantification. The shorter the cycle time the more data points across the chromatographic peptide peak and thus the more accurate the quantification. However, this also means less time for fragmentation and therefore less identifications. For good quantitative accuracy, researches have set a standard of 5 to 8 data points per peak although in DIA, as few as three or four are commonly achieved.

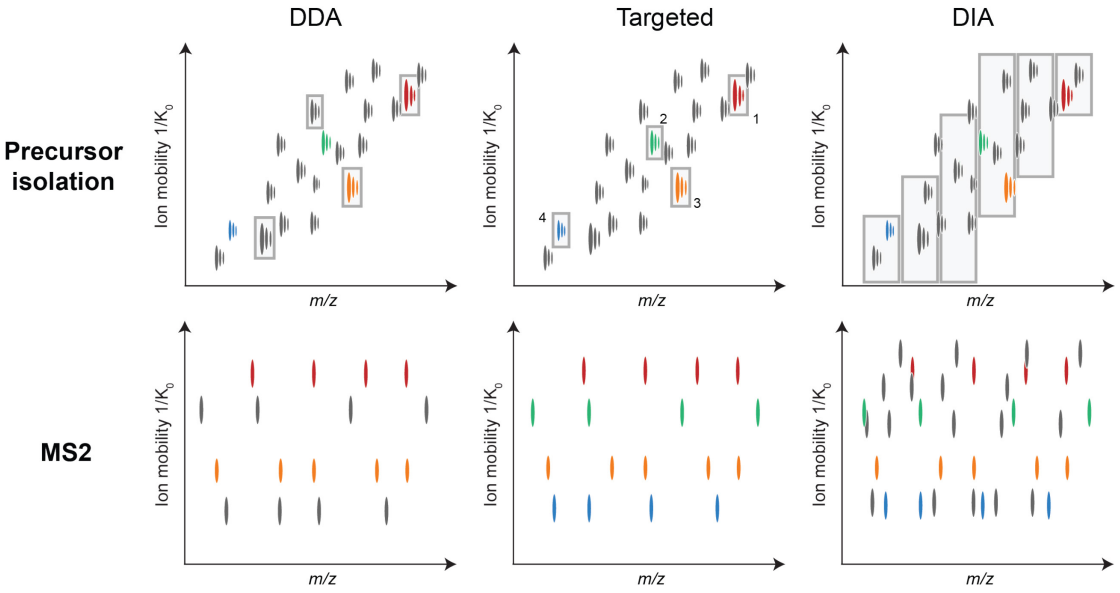



**Figure 6: Different data acquisition strategies as an overview.**
In all acquisition strategies, the MS consecutive acquires an MS1 and multiple MS2 scans. The MS1 acquisition is similar between the different strategies, while the principles for selection of precursors from the MS1 scan for fragmentation and MS2 acquisition differ.
In DDA (left), the top-N most intense precursors are selected for fragmentation and MS2 scans utilize narrow quadrupole isolation. Thus, the MS2 scans are dependent on the MS1 survey scan.
The targeted approach is similar (middle), but precursors are already predefined. Dependent on the strategy, fragments or transitions may also be specified.
In DIA (right), fragment scans are independent of the acquired MS1 scans. In this strategy, the quadrupole employs wide isolation windows that are predefined to methodically cover the entire m/z range of interest. Adapted from Meier *et al.*[194].

#### 1.3.5.1. Data-dependent acquisition

In data-dependent acquisition (DDA), the MS2 scan depends on the survey MS1 scan. A certain number of top-N abundant precursors are picked and subjected to fragmentation. The positive effects of DDA are relatively clean MS2 spectra which can be easily interpreted in that measured precursor m/z can be assigned to their fragments. However, DDA is prone to stochastic issues, since it will sample according to the MS1 scan in a finite time. This may result in many 'missing values' and a low data completeness, which is especially the case in sc-proteomics.

Many sc-proteomics benchmarks have still been done at the DDA level. However, due to the stochastic issues and missing value issues just mentioned, many researches have switched to DIA. For sc-proteomics, DDA is mainly used in combination with TMT multiplexing. Many different methods have been employed to alleviate the stochastic issues of DDA and to enhance the acquisition of mass spectra of relevant information with for example real-time search[195] and advanced inclusion lists[196].

#### 1.3.5.2. Targeted proteomics

In targeted proteomics, the analyzed precursors are already predefined, and this information is used to schedule MS2 scans. In parallel reaction monitoring (PRM), a full MS1 survey scan is acquired and then specific predefined precursors of the protein of interest are sent for fragmentation and MS2 acquisition. A faster one is PRM-PASEF on the timsTOF platform, which uses the alignment of TIMS and TOF for selection of multiple precursors. It achieves several hundred targets in practice. The original targeted approach is single or multiple reaction monitoring (SRM or MRM) which is mainly used with triple quadrupoles. The first quadrupole selects a precursor while the third one selects specific fragments which are predefined (the second one is the collision cell). Therefore, MRM is slower than PRM but may have more accurate quantification, since quantification is performed on predefined fragments with minimal interference.

Targeted proteomics is a the most prevalent tool for clinical setups to do standard and routine measurements of given biomarkers. Additionally, if known standard concentrations are spiked in, absolute quantification is possible. However, discovery proteomics mainly uses DIA acquisition.

#### 1.3.5.3. Data-independent acquisition

Although DDA is still a widely used acquisition strategy today, advances on mass spectrometers and search algorithms are rapidly shifting the field towards data-independent acquisition (DIA). In DIA, precursors are not selected individually based on

the MS1 survey scan, but the quadrupole selects sections of a certain mass range for MS2 fragmentation scans, so called isolation windows[197,198]. According to this principle, MS2 scans are sampled independent of the abundance of individual peptides or prior knowledge. In more detail, the full mass range is divided into multiple windows which are iteratively scanned after an MS1 precursor scan, a so-called DIA cycle. This DIA cycle is repeated throughout the entire LC run. To keep the cycle in a reasonable time, isolation windows are typically 25 Th wide. By selecting a mass range for fragmentation, MS2 scans are inherently much more complex than in DDA and they need to be deconvoluted. Search algorithms have been developed to deal with this issue and over the years they have been improved on many levels to make the analysis of these complex spectra possible and reliable. Inherent in the design of the scan mode is the advantage that all information of an acquired mass range is recorded in the spectra and all peptides are sent for fragmentation. Thus, DIA is unbiased and complete in producing fragment ions of peptides and their modifications. Although scheduling the iterative DIA cycle is straightforward, the placement of isolation windows is a challenge and needs to be optimized in order to cover an appropriate mass range within a particular time interval without generating overly complex spectra.

In sc-proteomics, DIA has also proven itself to be crucial in the acquisition of more complete data[78,199]. In order to enhance coverage, speed and sensitivity many scan modes have been developed. One of them is dia-PASEF, which combines the PASEF principle with DIA[200]. Single cells have lower spectral complexity which allows therefore to increase the window size, in some cases even the entire mass range is selected together, like in slice-PASEF[201].

### 1.3.6. Computational proteomics

A core part of computational proteomics are proteomic search algorithms. These attempt to identify peptides from fragmentation spectra and assemble several such peptides into a protein identification and quantification for downstream analysis. By now many different software packages have been developed in order to process the acquired signal intensity peaks of measured m/z in the dimensions of LC and potential ion mobility back to protein quantities. They use different search algorithms which include preprocessing of the mass spectra, peptide spectral matching (PSM) via a database search against *in silico* predicted fragments, PSM scoring and FDR control against a decoy database as well as the inference of proteins. In the preprocessing, recalibration of retention time, m/z and potential other dimensions such as ion mobility is a crucial step for reliable identification and quantification.

As a reminder, the MS1 spectra include information on intensities of peptide precursors, while MS2 spectra contain their fragments. In the most common search strategy, called library-based search, fragmentation mass spectra are compared to database libraries, either just calculated from the peptide sequences for DDA or experimental acquired or fully predicted in DIA, to obtain PSMs. To confidently identify a peptide, an adequate number of fragments needs to be present to distinguish it from other sequences. In order to avoid false spectra matching, observable target peptides compete with non-existing decoy peptides – randomly shuffled or reversed sequences of peptides[202]. Today, machine learning is applied to stratify targets ('true positive') and decoys ('false positive') matches to assure a generally accepted false discovery rate (FDR) of 1%. This FDR incorporates additional scoring of peptide properties to achieve a (posterior error probability (PEP)). Subsequently, identified peptide matches are assembled into protein identifications against a reference database using inference principles such as Occam's razor[203]. All observed peptides are explained using as few proteins as possible, minimizing the number of low confidence proteins. Afterwards, peptide quantifications are converted into protein quantifications.

Currently, the most prominent software packages for DIA are DIA-NN[204,205] and Spectronaut[206], which mainly allow to process data with established scan modes and experimental setups. To address this limitation, more modular and open-source alternatives are needed. Our laboratory is developing alphaDIA (Wallmann *et al.*, in preparation), a highly modular and transparent open-source pipeline in Python, which includes a DIA module that demonstrates competitive performance compared to the above software. Across all these platforms, machine learning has drastically improved DIA results through advanced scoring of PSMs by deep learning[204,206], highlighting the field's rapid technological evolution.

Recently the need for an empirical DIA library has also been superseded by predicted libraries, generated by deep learning[204,207,208]. While many still rely on experimental acquired libraries, their acquisition is time-consuming and often infeasible for single-cell and clinical samples as it requires much input material. Yet, in sc-proteomics, it is crucial to retrieve low signals with high confidence. Therefore, enhances of predicted libraries by artificial intelligence (AI) will be the future for sc-proteomics in biological application[209].

Subsequently, peptide or protein output tables are used for bioinformatic analysis in various software packages such Perseus[210] or AlphaPeptStats from our laboratory[211]. Today, most of them are implemented on top of R and increasingly Python[212–214]. In sc-

proteomics, any downstream bioinformatic pipeline implemented by transcriptomics can be used, however also proteomic specific ones have been developed and adapted, e.g. SCeptre[121].

### 1.3.6.1. Quantification strategies

Many different strategies can be applied for quantification in proteomics (Figure 7a). The simplest and most common is a label-free approach (LFQ). LFQ directly uses the integrated MS signal over all separation dimensions and compares signals from the same peptide across samples. This relative quantification is challenged by sample input variability, especially in sc-proteomics, which requires robust normalization strategies. An established algorithm, MaxLFQ, calculates pairwise peptide ratios across all samples to retrieve normalization factors after minimizing total variance. Although MaxLFQ is widely applied it comes with computational expenses when applied to large data sets. This has recently been addressed in the directLFQ approach by performing median normalization on 'intensity traces', while maintaining quantitative accuracy. In sc-proteomics, MS1 quantification (peptide signal) is an important way to achieve accurate quantification because fragment intensities may be extremely low, and this should be considered in DIA method design[216,217].

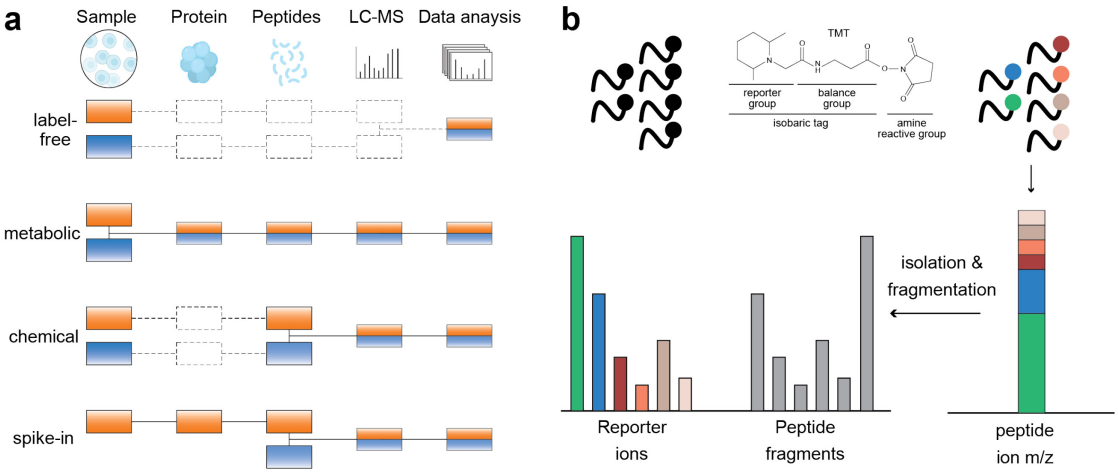



**Figure 7: Schematics of different quantification strategies in proteomic data.**
a. Label-free quantification measure each sample individually and retrieves its relative quantification by pairwise comparisons of peptide ratios. Spike-ins: heavy-labeled peptides are added to each sample. By comparing the intensity signal of the standard with the sample, absolute quantification is achieved. Metabolic labeling: the incorporation of different isotopes into the amino acids allows for ratio quantification between multiplexed samples. Chemical labeling: Mass tags are be added to peptides. Isotopic labels have the same mass, while non-isobaric labels have different masses. Adapted from Bantscheff *et al.*[215].
b. Schematic of isobaric labeling overlaying its signal in MS1 and quantification in MS2 upon fragmentation of the tandem mass tag (TMT) label. Adapted from Specht *et al.*[122].

MS-based multiplexing by introducing isotopic-different mass labels can be applied to use different quantification strategies. One of them uses synthetized heavy labelled spike-in samples introduced into the sample preparation. If the added amount is known

absolutely, the measured intensity for the sample can be converted to absolute quantities. This is mainly used in targeted acquisition strategies where absolute quantification of biomarkers or specific peptides of interest is important. Further, metabolic labeling can be achieved by heavy isotope labeled amino acids (N15 and C13). An example is stable isotope labeling by amino acids in cell culture (SILAC) where the proteomes are already differentially labeled before the sample preparation. In this strategy, both samples undergo the same treatment and are therefore less prone to analytical and technical variations. Finally, peptides can be labeled chemically with isotopically different mass tags. Peptides are then relatively quantified based on ratios between the mass tags. Chemical labels can be differentiated into isobaric and non-isobaric. Isobaric mass labels have the same precursor mass between different labels and composite a reporter group and a balancing group. Therefore, the complexity of the MS1 spectra is not increased by multiplexing. The mass label is labile upon fragmentation and the different isotopic states can be differentiate and quantified only in the MS2. Isobaric labels are only applicable in DDA and thus share the limitations of that approach. The most prominent example of an isobaric tag is tandem mass tag (TMT)[218,219]. In sc-proteomics, TMT is used in a pioneering approach termed SCoPE-MS which introduces the concept of a 'carrier channel'[103]. In this case one of the channels consists of a much higher number of cells boosting the MS1 signal since it overlaps with the other signals from single cell proteins. However, a limitation is that it is prone to ratio distortion due to the crosstalk between reporter groups of the fragmented peptide and more abundant co-isolated background peptides with the same reporter ions and the carrier peptide itself[220,221].

Non-isobaric labels are an attractive alternative that could maintain the advantages of DIA. However, non-isobaric labels increase the complexity with each additional multiplexed sample since each channel introduces another feature for each peptide in the spectra. With improving data analysis algorithms, this approach has become possible, especially in sc-proteomics with its lower spectral complexity. Thus, apart from using label-free approaches[78,120], multiplexed DIA ('mDIA' or 'plexDIA')[216,222] promises to increase throughput and proteomic depth in the future.

### 1.3.7. Summary of the challenges in single-cell proteomics

In summary, sc-proteomics is a rising field which challenges MS on many levels, which have been addressed in different ways. First of all, efficient sample preparation workflows have been established featuring miniaturization, one-pot reactions and automated liquid handling to retrieve the highest possible peptide amount. Many

approaches aim to deepen the proteome coverage, reduce missing values and improve the throughput by faster gradients or multiplexing. All these strategies require sensitive LC-MS instrumentations and sensitive data acquisition methods. In data analysis sc-proteomics will be able to learn from scRNA-seq processing such as batch effect, but will also need adaptions to proteomics workflows[223].

## 1.4. Spatial proteomics

The complexity of organism can be addressed by the simplicity of the cell theory, as discussed above. Sc-proteomics promises to analyze molecular mechanisms at the level of proteins as the functional unit of life, while capturing their heterogeneity. However, organisms are also a spatial arrangement of different cell types and cell states, in which the localization of a cell influences its activity and function[224]. Therefore, it is as important to understand their behavior and function in a spatial arrangement. As usually applied, sc-proteomics by itself loses the spatial context which is important to fully understand the function of cells in relation to each other.

Eventually the goal is to understand the complexity of life which arises by intricate functional and spatial adaption of its cell types and cell states. Cells are clustered into groups of cell types which generate specialized functionalities namely in tissues and organs[225,226]. Generally, tissue architecture is conserved by their different types of epithelial, connective, muscle and nerve tissue. These tissues are based on multiple cell types. The human body consists of several hundred cell types which together form various organs with different functions.

To conduct successful characterizations of cellular heterogeneity, native tissue needs to be stabilized for molecular analyses. Different preservation techniques have been used to make tissue accessible. The gold standard in pathology is formalin-fixed, paraffin-embedding (FFPE), in which format huge biobanks store human clinical samples. Formalin preserves the integrity of proteins and molecular structures of tissue by preventing enzymatic degradation. Formalin induces cross-links in proteins by forming robust methylene bridges[227]. Extensive heat steps – one hour at 95°C – need to be introduced in sample preparation protocols to break up those cross-links for proteomic analysis. It is crucial to preserve the tissue as fast as possible to avoid biases through degradation although autopsy-derived samples are stable for most purposes for a post-mortem interval of at least 24 hours[228]. Additionally, fresh frozen tissue can be embedded using optimum cutting temperature (OCT) medium, although this is less widespread because of its availability due to long-term storage. FFPE and fresh frozen

sections can achieve similar proteomic quality[229] with minimal influence on measured protein abundance[230–233].

A portfolio of methods has been developed to access the intricate functional and spatial adaption of its cell types and cell states[234], but each come with its caveats in spatial information, depth of coverage, molecular or cellular throughput and the choice of data acquisition[235,236]. Many methods have already been discussed in relation to single-cell proteomics (see 1.2.2) with multiplexed fluorescence imaging like CODEX or mass cytometry approaches[93,235], like CyTOF[94,237]. Some methods can even resolve at the subcellular level[91,238,239]. In a community effort termed the human protein atlas well-characterized antibodies were developed and used for imaging of proteins throughout the human body by antibody-based imaging[238,240]. Matrix-assisted laser desorption/ionization (MALDI)-MS uses a different principle for biomolecular ionization[241,242]. It is coupled to fast TOF analyzers to identify various molecular classes like metabolites, lipids and small proteins on grid positions in the tissue.

MS-based proteomics approaches have used grid-based laser-capture microdissection (LCM), so called tissue voxels, to retrieve spatial information[243], similar to spatial transcription methods. This approach cannot achieve single-cell and cell type resolved characteristics, and thus results may be biased by different cell type and state compositions.

The promise of spatial omic analysis has been demonstrated by spatial transcriptomics, which has been named the ‘Method of the year 2020’[244], with many technological variants such as multiplexed error-robust fluorescence in situ hybridization (MERFISH)[245] and GeoMx technology (‘nanostring’)[246].

### 1.4.1. Deep Visual Proteomics (DVP)

As describe above, it is important to integrate single-cell and spatially resolved molecular data to retrieve the complexity of a biological system. Our laboratory has developed a technology called ‘Deep Visual Proteomics’, which combines artificial intelligence (AI)-driven cell segmentation, high-content imaging, laser microdissection of single cells and ultra-sensitive MS (Figure 8). By doing so, DVP retrieves the spatial component of the tissue.

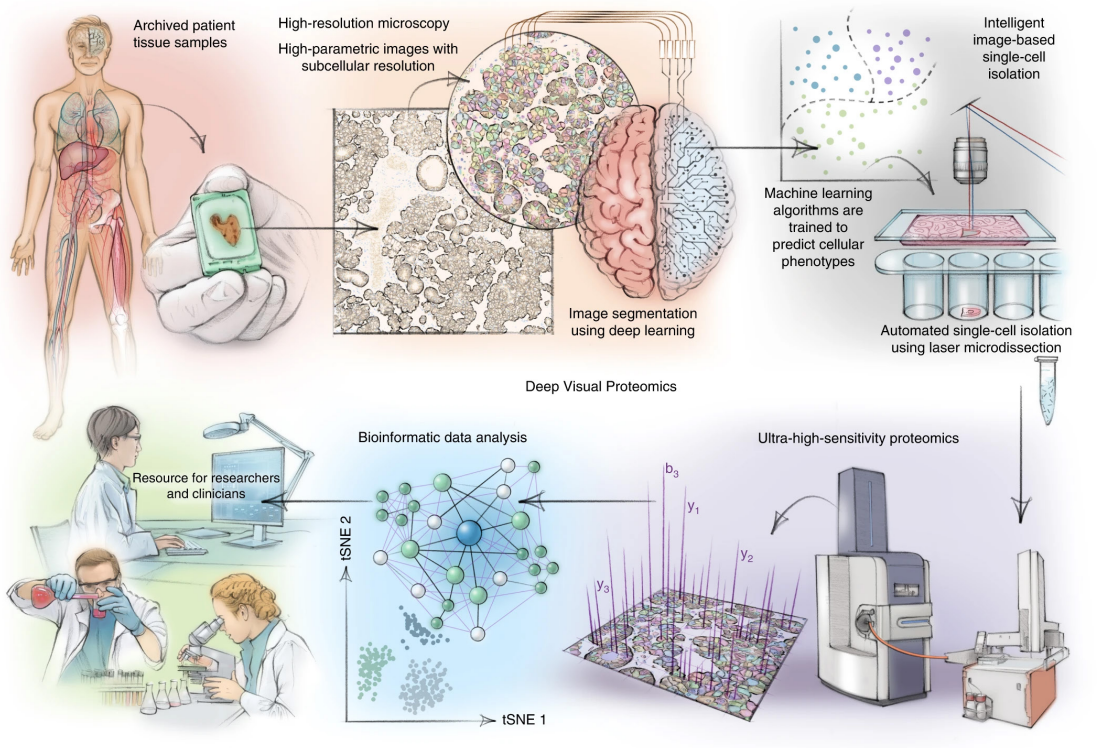



**Figure 8: Overview of the Deep Visual Proteomics (DVP) technology.**
DVP uses archived patient tissue in thin sections placed on membrane slides. Tissue slides are stained by fluorescence for AI-driven cell segmentation and their phenotypical classification. This information is used for laser-capture guided microdissection to collect single cell types, collected by gravity. Samples are processed by proteomic sample preparation workflows and measured by ultra-high sensitive MS workflows similar to those in sc-proteomics. Proteomic information and analysis retrieve biological interpretation for researchers and clinicians. The figure was reprinted as permitted by the CC-BY 4.0 international license in Mund *et al.*[247].

A thin tissue section of 2.5 to 3 µm is placed on a membrane slide of polyethylene naphthalate (PEN) and ‘standard’ fluorescence staining is applied. The membrane allows for later applied assisted extraction of tissue by LCM. High-content imaging is performed on slide scanner instruments (e.g., Zeiss Axio 7) for fast and high-resolution acquisition of the fluorescence signal. Several processing techniques have made it possible to image full tissue sections such as stitching to assemble single tile images and extended depth of focus (EDF), an algorithm to increase the depth of focus from z-stack images. Deep-learning image analysis segments single cells from high-content images. Cellpose is an algorithm that builds on large and heterogenous training datasets, which allows its application to many different cell and tissue types without extensive retraining. Another feature of AI-image analysis is that it recognizes morphological features and phenotypically classifies single cells on the basis of their fluorescence imaging signal. This allows DVP to classify the segmented cells into cell types based on established markers. Subsequently, DVP cuts single cells from the same slide by laser capture microdissection (LCM, e. g. Leica LMD7) and collects them by gravity. These samples are subjected to ultra-high sensitive MS as developed for sc-

proteomics. To achieve a greater proteomic depth, single cells of the same classification can be pooled to biologically enrich and retrieve spatial information on robust cellular phenotypes. Finally, measure protein abundances are analyzed for biological interpretation.

In summary, DVP connects the visual information by fluorescence imaging and spatial arrangement of cellular neighborhoods with molecular function and mechanisms derived from proteomic analysis, thereby measuring the cell type and state in a spatial context[247,248].

## 1.5. Clinical proteomics

### 1.5.1. Proteome in health and disease

Clinical proteomics aims to systematically profile health and disease states to achieve a holistic understanding of disease mechanisms at the protein level. MS-based proteomics is a powerful tool to unravel dynamic changes of proteins upon disease. Many model systems *in vitro* cell culture or patient-derived models have helped to understand these mechanisms. However, it is advantageous to analyze the disease phenotype directly at the molecular level of clinical samples to understand the disease in all its complexity. Recent approaches have shown that the analysis of patient tissue and the spatial context can lead to better treatment recommendations[249,250]. Milestones in clinical proteomics were the creation of atlases of specific organs (liver, skin and heart)[251,252] and of the entire body[253,254] from healthy to disease states[255].

Apart from tissue (fresh frozen[256], FFPE[257,258]), classical clinical samples are body fluids, grouped into those that can be obtained non-invasively like urine[259], stool[260] and serum/plasma[261] or invasive ones like CSF[262]. Body fluid proteomics faces its own analytical challenges[263,264], but once these are overcome one could uncover novel protein biomarkers for prognostic and predictive purposes. Large studies have been performed to discover biomarkers or biomarker panels instead of specific biomarkers. In this process AI is set to will help in biomarker discovery and support disease diagnostic[209].

Together, clinical proteomics has great potential in personalized medicine, especially when combined with artificial intelligence[265].

### 1.5.2. Post-translational modifications of proteins

Post-translational modifications (PTMs) modulate protein activity and function at different levels, and their dysregulation is associated with various disease such as

cancer and diabetes. Most PTMs are reversible and dynamic regulating molecular interactions, protein localization and stability. There is a great diversity with >300 known modifications, but most studied PTMs are phosphorylation, glycosylation, ubiquitination and acylation[266]. Each PTM has its own biological aspects: while glycosylation modulates protein stability and structure, ubiquitination targets proteins for degradation and acetylation is involved in metabolism and regulation of transcription.

Phosphorylation amplifies intra- and trans-cellular signaling pathways and is crucial for biological processes[267]. It is regulated by kinases and phosphates that attach and remove a phosphoryl group to the hydroxyl group of serin, threonine and tyrosine residues, respectively. The site localization is important for activity.

MS-based proteomics can detect a multitude of phosphopeptides with site-localization information. As mentioned in the sample preparation section (see 1.3.2), PTMs such as phosphorylation require an enrichment step to analyze modified peptides without the background of the non-modified proteome. Accurate and in-depth quantification can today be accomplished with library-free DIA[268]. Since PTMs usually have low stoichiometry, high-sensitive MS is needed for analysis of low input amounts.

Of note, (albeit critically discussed) FFPE blocks preserve PTMs when samples handled such that there is a minor influence until and during tissue fixation[269–271].

### 1.5.3. Cancer biology

Cancer is an abnormal growth of cells and has been evolving alarmingly over the last century, although survival rates have improved in the last decades. It is one of the most studied diseases, but mechanisms are still incompletely understood[272]. Although research has brought many new insights and some treatments are encouraging, overall it remains a dreaded disease. Although there are many types of cancers which behave very differently on the basis of malignancy, invasion and mechanisms, major hallmarks of cancer have been described[273]. The six main hallmarks are sustaining proliferative signals, escaping growth suppressors, resistance to cell death, enabling replicative immortality, induction of angiogenesis and activation of invasion and metastasis.

Genome instability and mutation underly all these hallmarks; while clonal evolution contributes to its vast heterogeneity until cellular subclones outgrowth the cancer[274]. Although cancers are fundamentally genetic diseases, the analysis of proteins is crucial to understand functional relationships to develop new therapies. Proteogenomic studies, the integration of proteomics with genomics, has provided new biological insights to identify therapeutic targets[275]. This combination can assign tumor subtypes which share

the same molecular mechanisms and respond to targeted therapies, supporting the vision of precision cancer medicine[276].

An emerging topic in cancer research is the tumor microenvironment with its hallmarks of reprogramming of energy metabolism and evasion of immune destruction[277]. The tumor microenvironment involves an interplay between many different cell types, including individual tumor cells, immune cells, fibroblasts and the extra cellular matrix (ECM).

In summary, intra-tumor heterogeneity and tumor microenvironment still remain a major challenge for tissue-centered proteomics and requires cell-type resolved, spatial proteomics approaches to answer fundamental questions that are still open.

### 1.5.4. Pancreatic islet biology

Pancreatic islets are clusters in the pancreas that consist of endocrine cells which produce hormones. The pancreatic islets are sparse with 1 to 2% of the volume of the whole organ and present an intricate spatial arrangement of different cell types: alpha cells produce glucagon, beta cells insulin, delta cells somatostatin and gamma cells pancreatic polypeptide. While the tissue architecture and cell proportions differ between species, beta cells are the most abundant ones at 70%, followed by alpha and delta cells. Clinically, the misfunctioning of pancreatic beta cells result in diabetes, primarily diagnosed by abnormally high blood glucose levels. Diabetes is a chronic metabolic disorder with increasing prevalence worldwide[278]. The disease is divided into two major groups, type 1 (T1D) and type 2 diabetes (T2D) with hyperglycemia as a common symptom. However, the underlying molecular mechanisms differ. T1D is an autoimmune disease cause by beta cell depletion, resulting in absolute insulin deficiency while T2D involves insufficient insulin production due to insulin resistance, resulting in progressive loss of functional beta cells by cell death or de-differentiation[279,280]. Still, mechanistic heterogeneity at the molecular level is observed in those groups making further subgrouping relevant for patient treatment[281]. Functional beta cell loss is a well-established phenomenon which manifests in pancreatic islet cell remodeling. In the past decades, pancreatic islet research has focused to a large degree on the analysis of beta cells. Therefore cell-type resolved, single-cell and spatial proteomics across all cell types can help the field to find new avenues to characterize the disease and unravel the underlying molecular mechanisms in patho- and physiological context. Researchers have successfully differentiated human stem cells into pancreatic beta cells[282], which may address scarcity in cadaveric donor tissues for islet transplantation. However, how these cells can nucleate functional islets in humans is not yet clear. This is, in part, due

to our lack of knowledge about proteomic changes and hormone production states in cells differentiating into pancreatic islet cell colonies. During my thesis, I have been involved in numerous projects studying the proteomics make up and functions of the different cell types of the islets *in vitro* and *in vivo*.

### 1.5.5. Immunopeptidomics

Immunopeptides are produced by degradation of most cellular proteins via the proteasome, specifically the immune proteasome. These peptides, in this context also called antigens, are shuttled onto major histocompatibility complex (MHC) proteins and eventually presented at the cell's surface[283]. There are two major classes: MHC class I, which predominantly presents peptides from intracellular proteins and is present on all nucleated cells; and MHC class II, which is mostly expressed on immune cells and upon inflammation on professional antigen presenting cells (APCs), presenting processed peptides from extracellular sources. MHC or human leukocyte antigen (HLA; the human version of MHC proteins) peptide presentation allows the immune system to distinguish between non-self and self to eliminate infected or abnormal cells and plays therefore a fundamental role in immune surveillance. Immunopeptides mediate disease-specific immune responses via disease-specific antigens evident in cancer and inflammatory diseases. One example of how these mechanisms can be harnessed is the development of cancer immunotherapies such as chimeric antigen receptors (CAR) T cells against tumor antigens. Immunopeptides can be studied with MS-based proteomics tools. In contrast to proteomics, MHC peptides introduce additional challenges like their non-tryptic nature and exceedingly low abundance. Millions of cells and many grams of tissue have been necessary to their analysis, but advances in sample preparation, sensitivity and specificity now allow much more streamlined workflows[284,285].

# 2. Aims of the thesis

My PhD thesis aims to push the boundaries of mass spectrometry (MS) to lift single-cell proteomics (sc-proteomics) to the next level and to produce relevant studies in which sc-proteomics significantly enhances our biological understanding. Beyond the aim of studying cellular heterogeneity, my goal was to promote sc-proteomics to the context of tissues to investigate spatial biological changes at the protein level. In particular, this required technological and methodological improvements of MS to enhance sensitivity and reproducibility for proteome depth, while retaining throughput and quantitative accuracy. Overall, the advancements in my PhD thesis are divided into four major parts: the development of scan modes for MS to increase coverage and quantitative data points, enhancements on sensitivity for sc-proteomics, biological investigation on spatial proteomics and the transfer of technological knowledge to the field of immunopeptidomics and pancreatic islet biology.

Sc-proteomics requires the use of all available ions and information. To this end, we created dia-PASEF scan modes which allowed the highest possible coverage of the ion cloud in dia-PASEF windows and we also transferred this to phospho proteomics (**Article 1**). In particular, the diagonally movement of the quadrupole improved the speed of dia-PASEF methods (**Article 2**). Next, we co-developed a new mass spectrometer with Bruker that had increased sensitivity to enable the measurement of single cells in a label-free manner (**Article 3**). Bringing those technological advances on the MS and dia-PASEF scan modes together with multiplexing on DIA (mDIA), I established a reference channel concept to double the proteomic depth in sc-proteomics while also increasing the throughput (**Article 4**). Next, I shared my knowledge and perspective on sc-proteomics and biological application in a Comment in Nature Methods, that has been quite influential (**Article 5**). Having established technological advancements for sc-proteomics, I transferred them to the spatial level and to biological investigations. First, I helped to transfer the 'Deep Visual Proteomics' (DVP) technology from our research group in Copenhagen to the one in Munich. While establishing the technology, we optimized the staining procedures on PEN membrane glass slides to improve the reproductivity (**Article 6**). Together, this paved the way for elevating tissue proteomics to the level of single cells (scDVP), by studying the spatial zonation of the liver proteome one cell shape at a time (**Article 7**). Not stopping at a two-dimensional level of spatial proteomics, we combined our workflow and technology with tissue clearing and co-developed DISCO-MS to analyze the spatial proteome three-dimensionally and in intact organs (**Article 8**). Proving biological impact of the spatial proteomics technology, we

investigated two severe diseases: different types of ovarian cancer by finding a potential new switch in malignancy (**Article 9**) and finding the first treatment strategy for a form of cutaneous adverse drug reaction (**Article 10**). Finally, I collaborated on the transfer of technological advancement in sc-proteomics to immunopeptidomics to enable deep coverage of immunopeptides in plasma (**Article 11**). During all these studies, I managed the proteomics part of a European consortium called ISLET, which aims to innovate advanced cell therapy for diabetes. Here, I helped to improve the therapeutic product in advanced cell-based therapies by for example characterizing the proteome of pancreatic islet cell types of alpha, beta and delta cells (**Article 12**). Many projects in this network are also part of my PhD studies but not part of this thesis.

# 3. Publications

## 3.1. Method development for mass spectrometry

### Article 1. Rapid and In-Depth Coverage of the (Phospho-)Proteome with Deep Libraries and Optimal Window Design for dia-PASEF.

*Mol Cell Proteomics*, 2022 (21, 9, 100279); DOI: 10.1016/j.mcpro.2022.100279

Patricia Skowronek[1], Marvin Thielert[1], Eugenia Voytik[1], Maria C. Tanzer[1], Fynn M. Hansen[1], Sander Willems[1], Ozge Karayel[1], Andreas-David Brunner[1], Florian Meier[1,2], Matthias Mann[1,3]

*[1]Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [2]Department of Functional Proteomics, Jena University Hospital, Jena, Germany. [3]Faculty of Health Sciences, NNF Center for Protein Research, University of Copenhagen, Copenhagen, Denmark.*

As discussed in the section on data acquisition strategies (1.3.5.3), data-independent acquisition (DIA) methods are attractive for sc-proteomics since they achieve high data completeness and a high dynamic range. However, the originally developed dia-PASEF methods were hampered by constraints which restricted the method design to equidistant quadrupole window sizes. As a result, a compromise needed to be made between high coverage of a wide m/z range and quantitative accuracy because of a long cycle time. Additionally, the quadruple isolation window placements were imbalanced in their spectral complexity due to differences in precursor densities across the m/z-ion mobility (IM) space. In this study, the aim was to develop and design dia-PASEF methods for several different gradients with highest coverage of the ion cloud. To create these methods, we developed an algorithm called py_diAID (Python package for DIA with an automated isolation design) which optimally places two-dimensional dia-PASEF windows into the m/z-IM plane. The algorithm adjusts the variable quadrupole isolation according to the precursor density so that precursors are best covered by a given cycle time and each window isolation enclosed the same precursor density and thus comparable spectral complexity. Deep fractionated libraries served as reference libraries, but single shot measurements proved to be sufficient as well. With the variable isolation width, we enabled short acquisition cycles while covering the full m/z-IM plane, an important feature for sc-proteomics. py_diAID is freely available and allows for fast and easy adaption of the dia-PASEF method tailored to specific needs such as sc-proteomics.

We also developed an optimal method for phosphoproteomics. Due to the mass shift of 79.966 Da (monoisotopic) resulting from the phospho group, the phosphopeptide precursor cloud is shifted towards higher m/z and IM values. The coverage of the precursor cloud was increased more than 50% compared to the original dia-PASEF method. With this, we were able to reach 10,000 phospho sites and close to 20,000 phosphopeptides in 21 minutes gradient. The analysis of epidermal growth factor (EGF)-stimulated HeLa cells intensively mapped the EGFR signaling pathway with over 35,000 phospho sites.

All in all, this algorithm was essential to create methods with fast cycle times which still covered the maximum of precursors in the m/z-IM space. This was key for sc-proteomics to sample sufficient data point for quantitative accuracy while still covering most of the precursors for a maximized coverage.

***Contribution:***

Co-authorship, second-author. The study was conceptualized by Patricia Skowronek, Andreas-David Brunner, Florian Meier and Matthias Mann. Patricia Skowronek conducted this study as first author. I contributed in the design and analysis of the proteomic benchmark experiment. I designed the phospho proteomic experiments together with Patricia Skowronek, Fynn M. Hansen and Ozge Karayel, as well as performed the phospho proteomic experiments together with Patricia Skowronek. My contribution on the phospho proteomics part led to the deep analysis of the phospho proteome in a short analysis time. Together with the other co-authors, I revised and edited the manuscript.

MCP TECHNOLOGICAL INNOVATION AND RESOURCES

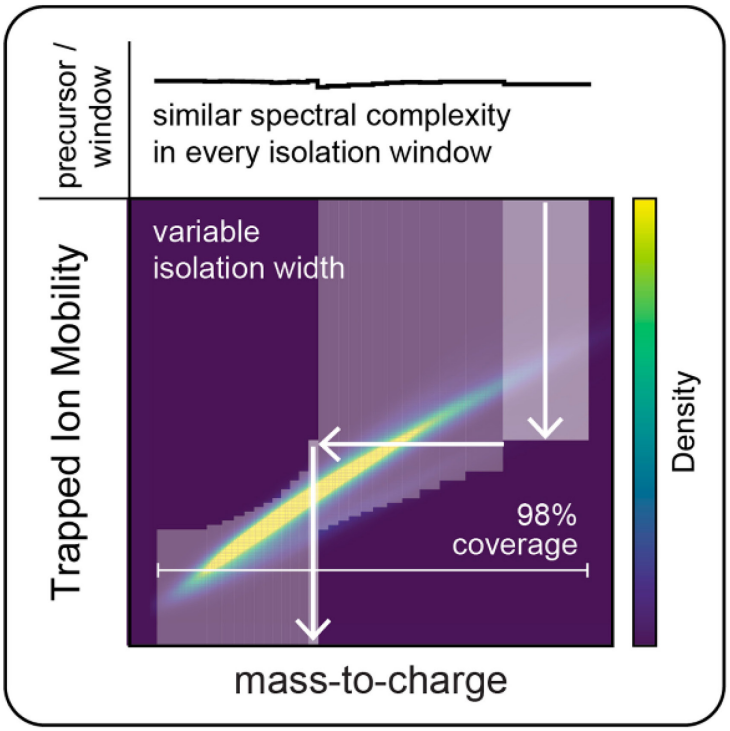
# Rapid and In-Depth Coverage of the (Phospho-) Proteome With Deep Libraries and Optimal Window Design for dia-PASEF

**Authors**
Patricia Skowronek, Marvin Thielert, Eugenia Voytik, Maria C. Tanzer, Fynn M. Hansen, Sander Willems, Ozge Karayel, Andreas-David Brunner, Florian Meier, and Matthias Mann

**Correspondence**
mmann@biochem.mpg.de

**In Brief**
dia-PASEF uses the correlation of molecular weight and ion mobility in a trapped ion mobility mass spectrometer for enhanced sensitivity. However, previous methods only covered a part of the precursor space, especially for phosphopeptides. Here, we developed py_diAID for optimal placement of dia-PASEF isolation windows for nearly complete precursor coverage. Combined with short liquid chromatography gradients (100 samples per day) and project-specific in-depth libraries, we achieve substantial coverage of the proteome and phosphoproteome.

**Graphical Abstract**




**Highlights**

- Optimal dia-PASEF window design with py_diAID combined with deep libraries.
- Quantification of the HeLa cell proteome to a depth of >7700 in only 44 min.
- Ion mobility–resolved phosphoproteomics determines >35,000 class I phosphosites.
- py_diAID is freely available as GUI, CLI, and Python modules.

2022, Mol Cell Proteomics *21*(9), 100279
© 2022 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
https://doi.org/10.1016/j.mcpro.2022.100279

MCP TECHNOLOGICAL INNOVATION AND RESOURCES

# Rapid and In-Depth Coverage of the (Phospho-) Proteome With Deep Libraries and Optimal Window Design for dia-PASEF

Patricia Skowronek[1], Marvin Thielert[1], Eugenia Voytik[1], Maria C. Tanzer[1], Fynn M. Hansen[1], Sander Willems[1], Ozge Karayel[1], Andreas-David Brunner[1], Florian Meier[1,2], and Matthias Mann[1,3,*]

**Data-independent acquisition (DIA) methods have become increasingly attractive in mass spectrometry–based proteomics because they enable high data completeness and a wide dynamic range. Recently, we combined DIA with parallel accumulation–serial fragmentation (dia-PASEF) on a Bruker trapped ion mobility (IM) separated quadrupole time-of-flight mass spectrometer. This requires alignment of the IM separation with the downstream mass selective quadrupole, leading to a more complex scheme for dia-PASEF window placement compared with DIA. To achieve high data completeness and deep proteome coverage, here we employ variable isolation windows that are placed optimally depending on precursor density in the *m/z* and IM plane. This is implemented in the freely available py_diAID (Python package for DIA with an automated isolation design) package. In combination with in-depth project-specific proteomics libraries and the Evosep liquid chromatography system, we reproducibly identified over 7700 proteins in a human cancer cell line in 44 min with quadruplicate single-shot injections at high sensitivity. Even at a throughput of 100 samples per day (11 min liquid chromatography gradients), we consistently quantified more than 6000 proteins in mammalian cell lysates by injecting four replicates. We found that optimal dia-PASEF window placement facilitates in-depth phosphoproteomics with very high sensitivity, quantifying more than 35,000 phosphosites in a human cancer cell line stimulated with an epidermal growth factor in triplicate 21 min runs. This covers a substantial part of the regulated phosphoproteome with high sensitivity, opening up for extensive systems-biological studies.**

Mass spectrometry (MS)–based proteomics has become a powerful tool to study proteomes in a systematic and an unbiased manner (1). In recent years, this development has been accelerated by data-independent acquisition (DIA) (2), where predefined isolation windows cycle through the *m/z* range of interest and regularly subject the covered peptide precursors to fragmentation (3–6). Although the concept of DIA was established more than a decade ago (4, 7), only the most recent DIA implementations and hardware advancements in MS and data analysis are at par or even exceeding data-dependent acquisition (DDA) with regard to sensitivity, reproducibility, and dynamic range coverage (2, 6, 8) and surpass targeted approaches in throughput and ease of use (9, 10). This holds also true for studying post-translational modifications (PTMs) (11–13).

DIA has recently shown promise in combination with trapped ion mobility (IM) spectrometry (TIMS) mass spectrometers, as demonstrated with single-cell analysis (14, 15). The TIMS tunnel is a compact and high-performance implementation of IM separation. It captures the peptides from the incoming ion beam discretizing the continuous liquid chromatography (LC) elution. Within the TIMS tunnel, each ion reaches an equilibrium position based on the opposing forces of a gas flow and an electric field gradient. Decreasing the electric field gradient elutes the peptide ions as a function of their IM (16–19). In the Bruker timsTOF instruments, the TIMS device is placed upstream of mass-selective quadrupole and high-resolution time-of-flight (TOF) mass analyzer and is itself divided into two parts (20–22). The mobility separation can be synchronized with the quadrupole isolation, leading to high ion beam utilization, increased sensitivity, and decreased spectral complexity because of the additional IM dimension (6, 20, 23). This principle is termed PASEF for parallel accumulation–serial fragmentation (21, 24).

When combined with DIA (dia-PASEF), peptide precursors separate not only in the *m/z* but also in the IM dimension, in contrast to standard DIA modes (2, 6). We have observed that dia-PASEF is particularly beneficial for acquiring a wide range of proteomics data while maintaining a high sequence coverage and very high sensitivity (6, 15). Furthermore, ions

From the [1]Department Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany; [2]Department of Functional Proteomics, Jena University Hospital, Jena, Germany; [3]Faculty of Health Sciences, NNF Center for Protein Research, University of Copenhagen, Copenhagen, Denmark
*For correspondence: Matthias Mann, mmann@biochem.mpg.de.



© 2022 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.1016/j.mcpro.2022.100279

are detected by inherently fast TOF analysis allowing fast DIA cycle times, which is particularly advantageous for short LC gradients (6, 25). The resulting complex spectra can be efficiently analyzed by machine learning or deep learning–based algorithms such as DIA-NN (26, 27).

Here, we set out to explore the potential of dia-PASEF to further increase coverage and quantitative accuracy on the fast and sensitive IM-MS platform. In dia-PASEF, two-dimensional precursor isolation schemes are defined in the *m/z*–IM plane. We used a Bayesian optimization algorithm ensuring optimal placement of the acquisition scheme in both dimensions. Single runs acquired with these optimal dia-PASEF methods were searched against in-depth project-specific libraries. Furthermore, we combined dia-PASEF with the Evosep One LC system, which features a preformed gradient particularly designed for high throughput by eliminating inter-run overhead (6, 28). Together, our optimized dia-PASEF workflow for high-throughput proteomics quantified more than 7000 proteins in only 21 min from quadruplicate injections of a tryptic HeLa digest.

Motivated by these proteomic results, we also investigated py_diAID (Python package for DIA with an automated isolation design) for phosphorylation analysis. On the Orbitrap MS platform, Olsen *et al.* (11) recently demonstrated an efficient combination of fast chromatography runs with DIA, quantifying more than 13,000 phosphopeptides in very short (15 min) LC–MS runs from HeLa cells using the Spectronaut software (Biognosys AG). In a small-scale study, Ishihama *et al.* (29) showed that phosphopeptide analysis benefits from the additional IM dimension in PASEF. For large-scale PTM studies, our optimized py_diAID acquisition schemes cover nearly all theoretical phosphopeptide precursors and quantified expected changes in the well-studied epidermal growth factor (EGF) receptor (EGFR) signaling pathway with minimal time and sample consumption.

## EXPERIMENTAL PROCEDURES

### *Experimental Design and Statistical Rationale*

All experiments were done using HeLa cell lysate obtained from HeLa S3 cells (American Type Culture Collection) and routinely used for proteomics method development and benchmark experiments (supplemental Fig. S1). Altogether, the dataset includes 322 raw data files (uploaded to PRIDE, see later). We used the same HeLa batch for generating libraries and single-run data of both proteome and phosphoproteome measurements. In brief, proteome measurements with different gradient lengths and the technical comparisons of the original and optimal dia-PASEF methods for phosphoproteomics were acquired in quadruplicates. Unless otherwise mentioned, 200 ng HeLa lysate was used for single-run proteome and 100 µg for the single-run phosphopeptide enrichment experiments. The libraries were acquired as described later. The experimental design and statistical rationale are described in the respective figure legends. The EGF experiment was performed in biological triplicates to determine significantly different phosphosite levels between the EGF-treated and control samples. Technical quadruplicates were acquired to evaluate reproducibility and quantitative accuracy by calculating CVs and mean of the replicate injections. Moreover, we alternated the MS run order to avoid potential carryover effects or any similar biases.

### *Sample Preparation*

HeLa S3 cells (American Type Culture Collection) were cultured in Dulbecco's modified Eagle's medium (Life Technologies Ltd) containing 20 mM glutamine, 10% fetal bovine serum, and 1% penicillin–streptomycin. Sample preparation was essentially performed as previously described in the in-stage tip protocol (30). In brief, the cells were washed with PBS and lysed. Protein reduction and alkylation and digestion with trypsin (Sigma–Aldrich) and LysC (WAKO) (1:100 dilution, enzyme/protein, w/w) were performed in one step. Resulting peptides were dried and reconstituted in a solution A* (0.1% TFA/2% acetonitrile [ACN]). Peptide concentrations were measured optically at 280 nm (Nanodrop 2000; Thermo Fisher Scientific), and 200 ng peptides were loaded onto Evotips for LC–MS/MS analysis as described previously (15). The Evotips were washed with 0.1% formic acid (FA)/99.9% ACN, equilibrated with 0.1% FA, loaded with the sample dissolved in 0.1% FA, and washed with 0.1% FA.

For phosphoproteomics, HeLa cells at a plate confluence of 80% were treated for 10 min with 100 ng/ml animal-free recombinant human EGF (PeproTech) or Gibco distilled water (Thermo Fisher Scientific) and washed three times with ice-cold TBS before lysis in 2% sodium deoxycholate in 100 mM Tris–HCl (pH 8.5) at 95 °C. Protein concentrations were determined using the bicinchoninic acid assay, and samples were then reduced and alkylated with 10 mM Tris(2-carboxyethyl)phosphine and 40 mM chloroacetamide, respectively. Altogether, 25 mg protein material of sample was used for the library generation, 8 mg for EGF-treated experiments including method benchmarking, and 4 mg for untreated experiments. The sample was digested with trypsin (Sigma–Aldrich) and LysC (WAKO) (1:100 dilution, enzyme/protein, w/w) overnight and subsequently desalted using Sepax Extraction columns (Generik DBX). Each cartridge was prepared with 100% MeOH and 99% MeOH/1% TFA. After equilibration with 0.2% TFA, the samples were loaded with a protein concentration of 1 mg/ml, washed with 99% isopropylamine/1% TFA, 0.2% TFA/5% ACN, and 0.2% TFA solutions. The peptides were eluted with 5% $NH_4OH$/80% ACN. Lyophilized peptides were reconstituted in equilibration solution (1% TFA/80% ACN), and 100 µg peptide material per sample/AssayMAP cartridge, each containing 5 µl Fe(III)–nitrilotriacetic acid, was enriched for phosphopeptide with the AssayMAP bravo robot (Agilent) (31). Phosphopeptides were dried in a SpeedVac for 20 min at 45 °C and loaded onto Evotips as described previously.

### *High-pH Reverse-Phase Fractionation for Library Generation*

To generate proteome libraries, 10 and 60 µg peptides were separated with high-pH reverse-phase chromatography into 24 and 48 fractions, respectively, on a 30 cm $C_{18}$ column with an inner diameter of 250 µm at a flow rate of 2 µl/min using the spider sample fractionator (32). The gradient consisted of the binary buffer system (PreOmics GmbH). The buffer B concentration of 3% was increased to 30% in 45 min, 40% in 12 min, 60% in 5 min, and 95% in 10 min. After washing at 95% for 10 min, buffer B concentration was re-equilibrated to 3% in 10 min. The exit valve concatenated the eluted peptides automatically by switching after a defined collection time (80 s for 24 and 60 s for 48 fractions). The fractions were dried in a SpeedVac and reconstituted in solution A*. A quarter of each fraction was loaded onto Evotips for LC–MS/MS analysis. Later, we will refer to "the reference proteome library" that represents a 24 high-pH fraction and DDA–PASEF spectral library of a tryptic HeLa digest acquired with a 21 min Evosep gradient.

To generate a phosphoproteome library, peptides obtained from the EGF-stimulated cells were separated using an UFLC system

(Shimadzu). About 6 mg peptide material was fractionated with a binary buffer system: A (2.5 mM ammonium bicarbonate) and B (2.5 mM ammonium bicarbonate/80% ACN). The peptides were loaded onto a reversed-phase column (ZORBAX 300Extend-$C_{18}$; Agilent) and separated at a 1 ml/min flow rate at 40 °C. The buffer B concentration of 2.5% was increased to 38% in 82.5 min, 75% in 2 min, and 100% in 8 min. It stayed at 100% for 2 min and was reduced to 2.5% in 2 min. In total, 95 fractions were collected, and fractions with low peptide yield, as determined using Nanodrop, were pooled (supplemental Table S1) and dried in a SpeedVac. Next, 76 fractions were enriched for phosphopeptide, which were subsequently loaded onto Evotips.

### *LC–MS/MS Analysis*

The Evosep One LC system coupled with a timsTOF Pro mass spectrometer (Bruker) was used to measure all samples. The 60 and 100 SPD (samples per day) methods required an 8 cm × 150 μm reverse-phase column packed with 1.5 μm $C_{18}$-beads (PepSep), and the 30 SPD method a 15 cm × 150 μm column with 1.9 μm $C_{18}$-beads (PepSep) at 40 °C. The analytical columns were connected with a fused silica ID emitter (10 μm ID; Bruker Daltonics) inside a nano-electrospray ion source (Captive spray source; Bruker). The mobile phases comprised 0.1% FA as solution A and 0.1% FA/99.9% ACN as solution B.

The library samples were acquired in DDA–PASEF mode with four PASEF/MSMS scans at a throughput of 60 and 100 SPDs and 10 PASEF/MSMS scans at 30 SPD per topN acquisition cycle. Singly charged precursors were filtered out by their position in the *m/z*–IM plane, and only precursor signals over an intensity threshold of 2500 arbitrary units were picked for fragmentation. While precursors over the target value of 20,000 arbitrary units were dynamically excluded for 0.4 min, ones below 700 Da were isolated with a 2 Th window and ones above with 3 Th. All spectra were acquired within an *m/z* range of 100 to 1700 and an IM range from 1.51 to 0.6 V $cm^{-2}$.

We described the original dia-PASEF method in the study by Meier *et al.* (6). The dia-PASEF methods optimized here with py_diAID cover an *m/z* range from 300 to 1200 for proteome and from 400 to 1400 for phosphoproteome measurements. Each method includes two IM windows per dia-PASEF scan with variable isolation window widths adjusted to the precursor densities. Eight, 12, and 25 dia-PASEF scans were deployed at a throughput of 100 (cycle time: 1.0 s), 60 (cycle time: 1.4 s), and 30 SPDs (cycle time: 2.7 s), respectively. We created dia-PASEF methods with equidistant window widths (supplemental Fig. S5) with the software "Compass DataAnalysis" (Bruker Daltonics). These acquisition schemes are plotted on top on a kernel density estimation of precursors from a reference library in supplemental Figs. S2–S4. The IM range was set to 1.5 and 0.6 V $cm^{-2}$. The accumulation and ramp times were specified as 100 ms for all experiments. As a result, each MS1 scan and each MS2/dia-PASEF scan last 100 ms plus additional transfer time, and a dia-PASEF method with 12 dia-PASEF scans has a cycle time of 1.38 s. The collision energy was decreased as a function of the IM from 59 eV at $1/K_0$ = 1.6 V $cm^{-2}$ to 20 eV at $1/K_0$ = 0.6 V $cm^{-2}$, and the IM dimension was calibrated with three Agilent ESI Tuning Mix ions (*m/z*, $1/K_0$: 622.02, 0.98 V $cm^{-2}$, 922.01, 1.19 V $cm^{-2}$, 1221.99, and 1.38 V $cm^{-2}$). For phosphoproteomics experiments, the collision energy was decreased from 60 eV at 1.5 Vs $cm^{-2}$ to 54 eV at 1.17 Vs $cm^{-2}$ to 25 eV at 0.85 Vs $cm^{-2}$ and end at 20 eV at 0.6 Vs $cm^{-2}$.

### *Raw Data Analysis*

We employed DIA-NN, MSFragger, and Spectronaut for transforming raw data into precursor and fragment identifications based on 3D peak position (retention time [RT], *m/z* precursor, and IM). In each case, all data were searched against the reviewed human proteome (UniProt, November 2021, 20,360 entries without isoforms) with trypsin/LysC as digestion enzymes. Cysteine carbamidomethylation was set as fixed modification. Methionine oxidation, methionine excision at the N terminus, and in the case of the phosphoproteome searches, phosphorylation (STY) was selected as variable modifications. A maximum of two missed cleavages and up to three variable modifications were allowed.

The project-specific libraries for DIA-NN analyses were generated with FragPipe (27) (FragPipe 16.2, MSFragger 3.4 (33–35), Philosopher 4.0.0 (36), Python 3.8, EasyPQP 0.1.25 (https://github.com/grosenberger/easypqp)). The default settings were kept except that the precursor mass tolerance was set from −20 to 20 ppm and the fragment mass tolerance to 20 ppm. In addition, Pyro-Glu or ammonia loss at the peptide N terminus and water loss on N-terminal glutamic acid were selected as variable modification. The output tables were filtered for a 1% false discovery rate (FDR) using the Percolator (37, 38) and ProteinProphet (39) option in FragPipe (supplemental Table S2).

DIA-NN 1.8 was used to analyze the single-shot experiments against the project-specific libraries generated with FragPipe (27). The default settings were kept except that we changed the charge state to 2 to 4. The precursor's *m/z* range was restricted from 300 to 1200 for proteome and 400 to 1400 for phosphoproteome analysis. The fragment *m/z* range was set from 100 to 1700, and the mass and MS1 accuracy was set to 15 ppm. "Match between run" was enabled, whereas "protein inference" was disabled. We also enabled "robust LC (high precision)" as the quantification strategy. The proteomics output tables were filtered for a maximum of 1% of *q* value at both precursor and global protein levels. For phosphoproteomics, the PTM *q* value also had to be a maximum of 1%. The "PG.MaxLFQ" column integrated in the DIA-NN output tables reports normalized quantity employing the MaxLFQ principle (40) and was used for quantitative analysis on the protein level. For our phosphoproteomics analysis, we used the scoring of post-translational sites implemented in DIA-NN with "PTM.Site.Confidence" indicating the localization probability (13).

Spectronaut (version 16) (3) was used for comparative analysis, and we used the same search settings as described previously if not stated differently. The FDR cutoff was set to 1%. The precursor peptide and *q* value cutoffs were 0.2 and 0.01, respectively. The protein *q* value experiment and run wide cutoffs were 0.01 and 0.05, respectively. The dataset was analyzed with a sparse *q* value, and no imputation was performed. For phosphoproteomics experiments, the PTM localization cutoff was set to 0. The results were filtered for the best N fragments per peptide between 3 and 25.

Peptide collapse (version 1.4.1), a plug-in tool for Perseus (41), collapsed peptide output tables from DIA-NN or Spectronaut to phosphosite tables using default settings and a localization cutoff of 0.75 (class I sites) (11). The DIA-NN output table was reformatted by renaming all columns and entries calculating peptide positions to conform to the format required for the plug-in tool. For collapsing, Perseus took only phosphorylation into account. During collapsing phosphopeptide ions to phosphosites, each phosphosite corresponding to the same peptide obtains the same intensity; however, imputation may lead to differences in fold changes. If the same phosphosite was identified on different peptides, which may also have modifications other than phosphorylation or different charge states, the intensities were summed up.

### *Statistical Analysis*

Visualization and statistical analyses were performed using the output tables of DIA-NN or Spectronaut with Python (version 3.8, Jupyter Notebook; Project Jupyter) and the packages pandas (1.4.2) and pyfaidx (0.6.1) for data accession and py_diAID (0.0.16), AlphaMap (0.1.10), matplotlib (3.4.3), and seaborn (0.11.2) for visualization.

## Optimal Design of dia-PASEF Windows

The statistical analysis of the EGF experiment was performed in Perseus (version 1.6.2.2). $Log_2$-transformed intensities were filtered for 100% valid values in at least one condition. The missing values were replaced drawing from a normal distribution (width 0.3 and downshift 1.8). Next, we applied the two-sided Student's *t* test ($S_0$ = 0.1, FDR = 0.05) to obtain the significantly changing phosphorylated peptides. A Fisher's exact test was performed for Gene Ontology term and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis ($p < 0.002$).

RESULTS

### *Principle and Limitations of the Original dia-PASEF Window Design*

In the timsTOF mass spectrometer (Bruker Daltonics), a dual TIMS tunnel releases the captured peptide ion species individually as a function of their mobility. In a PASEF MS–MS scan, a quadrupole transmits part of the ion beam where the precursor *m/z* values fall into a predefined isolation window (Fig. 1*A*). These precursors are subsequently fragmented by applying a particular collision energy. A downstream TOF analyzer acquires high-resolution mass spectra. In dia-PASEF, changing the quadrupole position is synchronized to the IM elution, increasing the MS efficiency because the isolation window is placed on top of the precursor cloud (6). This movement happens in distinct steps and thereby divides one PASEF scan into multiple IM windows. The quadrupole isolation window is first placed at high *m/z* for a certain amount of time, after which it jumps to a position in the lower *m/z* range. This transition point corresponds to a particular IM value for each dia-PASEF scan. In each subsequent dia-PASEF scan, the starting *m/z* window is offset to lower values (Fig. 1, *B* and *C*). Together, these isolation windows cover a large proportion of the *m/z* and the IM dimensions, constituting a two-dimensional acquisition scheme (Fig. 1*B*).

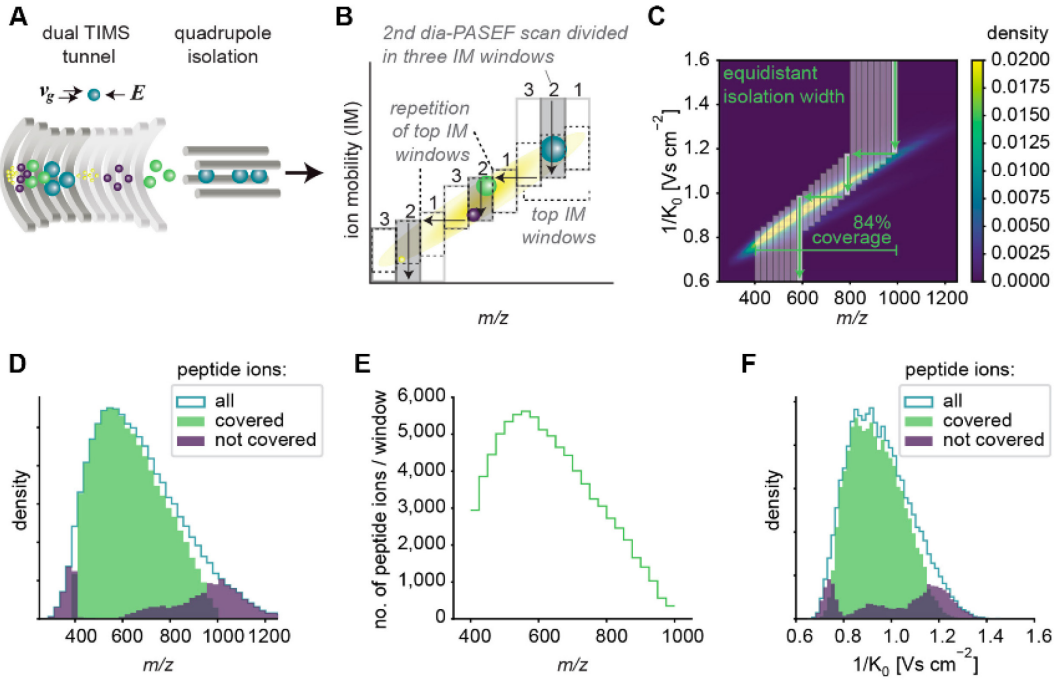



FIG. 1. **Principle of dia-PASEF on a timsTOF with equidistant two-dimensional isolation windows.** *A*, schematic of a TIMS tunnel followed by quadrupole isolation. *B*, dia-PASEF acquisition scheme depicting three dia-PASEF scans divided into three ion mobility (IM) windows. *Vertical arrows* indicate the elution of the ions with decreasing electrical field, and *horizontal arrows* indicate the movement of the quadrupole. The pattern of the *top* IM windows is repeated, and the *top* and *bottom* IM windows are extended to the upper and lower IM range, respectively. *C*, original dia-PASEF acquisition scheme (6) plotted on a kernel density distribution of all precursors. One dia-PASEF scan is divided into three IM windows by three distinct movements of quadrupole isolation. This scheme comprises eight dia-PASEF scans with equidistant isolation width covering in total 84% of the peptide ion population. *D*, histogram of *m/z* of all peptides covered by the acquisition method in (*C*), and peptides not covered by the method but identified in a separately recorded spectral library. *E*, number of peptide ions per isolation window. *F*, histogram of IMs of all peptides covered by the acquisition method, and peptides not covered by the method but identified in a separately recorded spectral library. The subfigures *C–F* are based on a reference proteome library (see the Experimental Procedures section). DIA, data-independent acquisition; PASEF, parallel accumulation–serial fragmentation; TIMS, trapped ion mobility spectrometry.

Because of software constraints, the original dia-PASEF methods (6) comprise a repeating pattern of the top IM windows per dia-PASEF scan. This leads to a configuration with equidistant quadrupole isolation widths (Fig. 1*B*). As a result, covering a wide *m/z* range comes at the cost of a high cycle time and reduced quantitative accuracy because of lower elution peak coverage. Alternatively, many peptide ions outside the *m/z* range would not be included in the acquisition scheme (Fig. 1, *C* and *D*).

Moreover, when using equidistant isolation windows, the distribution of peptide ions per window is imbalanced, resulting in a high spectral complexity in highly dense regions (Fig. 1*E*). Finally, this scheme for acquisition window setting is also suboptimal in the IM dimension (Fig. 1*F*).

### *Establishing an Optimal dia-PASEF Window Design*

We first investigated the optimal balance between the number of dia-PASEF scans and IM windows per dia-PASEF scan to obtain a deep proteome coverage and quantitative accuracy. As described previously, the original dia-PASEF method included three IM windows per dia-PASEF scan. Having more IM windows per dia-PASEF scan reduces cycle time but also diminishes precursor coverage because of smaller isolation windows in the IM dimension (supplemental Fig. S5, *A* and *B*). For instance, splitting the isolation width into two parts halved the complexity per spectrum and thereby increased identifications. However, doubling the number of dia-PASEF scans increases cycle time, which worsens the quantitative accuracy since only half as many data points are collected over one elution peak (supplemental Fig. S5*A*). We tested the impact of increasing the number of IM windows per dia-PASEF scan and found that two IM windows per dia-PASEF scan are optimal (supplemental Fig. S5*C*). While optimizing the cycle time of DIA methods, one has to prioritize for the desired trade-off between identification and quantitative accuracy. The Olsen and Reiter groups achieving an average of for four and eight data points per peak, respectively (11, 42), and here we aimed at six points per elution peak. In the case of 21 min gradients (60 SPD), we empirically found an average peak width of 8.3 s (base to base, as reported by DIA-NN, see supplemental Fig. S5*D*). Each individual dia-PASEF scan takes around 100 ms plus one 100 ms MS1 scan per cycle and overhead time. Hence, 12 dia-PASEF scans amount to a cycle time of 1.4 s, representing an optimal trade-off between the adequate quantitative representation of the LC elution peak and proteomics depth (see the Experimental Procedures section, supplemental Fig. S5*E*). If a study requires a stronger focus on quantitative accuracy, a lower number of dia-PASEF scans and hence shorter cycle time may be beneficial.

Given the limitations of our previous two-dimensional acquisition scheme, we needed to place and adjust *m/z* and IM isolation windows flexibly. Existing tools such as "Define dia-PASEF Region" in Compass DataAnalysis (Bruker) or the "dia-PASEF window Editor" in TimsControl (Bruker) require the manual fitting of the scan area onto the peptide ion population and only generate isolation windows with equidistant widths. Therefore, we developed py_diAID. It places two-dimensional dia-PASEF acquisition schemes in the *m/z*–IM plane based on desired parameters (number of dia-PASEF scans, covered *m/z* and IM range, and cycle time) and the empirical acquired reference data, which can be a proteomics library containing precursor ion information. The algorithms in py_diAID optimally adjust the variable quadrupole isolation widths according to the precursor density, aiming for an equal number of precursors fragmented per isolation window. Our simulations show that variable isolation widths enable short acquisition cycles covering essentially the entire *m/z*–IM range (Fig. 2*A*, *right panel*).

Our algorithm first bins the precursor ion populations equally along the *m/z* dimension. A trapezoid defines the extent of scan area and the position of the acquisition scheme in the *m/z*–IM plane (Fig. 2*A*, *left panel*). Based on this, py_diAID calculates the optimal dimensions of each isolation window (Fig. 2*A*, *middle panel*) and extends the top and bottom IM windows to the limits of the measured IM range to maximize the covered peptide ion population (Fig. 2*A*, *right panel* and supplemental Fig. S6). The selected mass window of the quadrupole jumps at the determined transition point of each IM window within each dia-PASEF scan. In each subsequent dia-PASEF scan, the starting *m/z* window is offset to lower values based on the individual width of the previous window (Fig. 2*A*). Next, py_diAID evaluates the generated acquisition scheme based on the covered precursor ions of an experimentally acquired library or subset thereof, for example one filtered by a charge state or by a population of modified peptides. This is a multivariant nonlinear optimization problem, and we used the gp_minimize module provided by the Scikit-Optimize (skopt) library in Python to perform this task that is highly used in machine and deep learning for the hyperparameter optimization (see the Experimental Procedures section). Its inputs are the trapezoid corners, and it iteratively decides which parameters should be tested next based on the aforementioned evaluation. This process is repeated for many iterations (about 200 in practice, supplemental Fig. S7) until it converges to the best window placement. py_diAID is available as a Python module, a command-line interface, and a graphical user interface on all major operating systems under an Apache 2.0 license (supplemental Fig. S8). The source code is freely available on GitHub (https://github.com/MannLabs/pydiAID).

We first benchmarked the optimal dia-PASEF methods designed with py_diAID against the original dia-PASEF method, which we termed "high speed" in our original dia-PASEF publication (6). That method covered 88% of all doubly and 71% of all triply charged precursors in the "reference library," which was generated with FragPipe. In contrast, the optimal dia-PASEF method calculated by

**Optimal Design of dia-PASEF Windows**

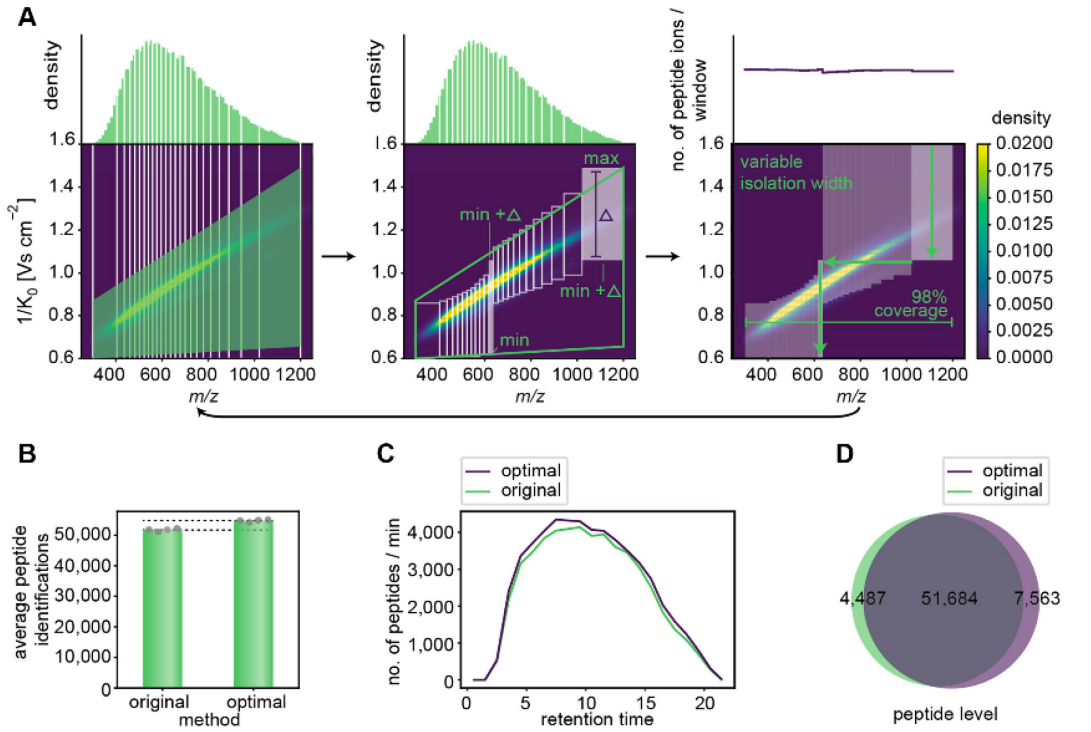



FIG. 2. **py_diAID algorithm and evaluation.** *A*, py_diAID design of the optimal acquisition scheme and window placement for a 21 min gradient (60 SPD, Evosep) with variable widths to balance the distribution of peptide ions, providing nearly complete peptide ion coverage. The *left panel* illustrates the first steps of the py_diAID algorithm: defining the *m/z* range of interest, binning the peptide ions in the *m/z* dimension and definition of the scan area in the IM dimension. *Middle panel*, calculation of the isolation window dimensions and coordinates based on the scan area. *Right panel*, extension of the isolation windows to the limits of the IM ranges. The *arrow* at the *bottom* indicates that the py_diAID algorithm evaluates the new acquisition scheme, defines the following test set of scan area parameters by Bayesian optimization, and resumes with the steps in the *left panel*. This is repeated for a user-defined number of iterations (more details in supplemental Fig. S6). *A* is plotted on top of a kernel density distribution based on the reference proteome library. *B*, average peptide identifications by the original and optimal dia-PASEF methods. *C*, number of peptides identified per minute over the entire retention time. *D*, Venn diagram showing the shared and unique peptides identified by both methods. Data in *B–D* are from quadruplicate injections of 200 ng tryptic HeLa digest with a 21 min gradient and analyzed with the reference proteome library. DIA, data-independent acquisition; IM, ion mobility; PASEF, parallel accumulation–serial fragmentation; py_diAID, Python package for DIA with an automated isolation design; SPD, samples per day.

py_diAID reached 99% and 94%, respectively. The original dia-PASEF method had already been extensively and manually optimized for the short gradient lengths and the tryptic HeLa digest employed here. This explains why the number of experimentally identified proteins is very similar between both methods (supplemental Table S3). However, even in this case, the optimal acquisition scheme of py_diAID increased the number of identified peptides by 6% in single-run injections (Fig. 2*B*) and across the entire RT (Fig. 2*C*), whereas the number of peptide identifications in replicate injections deviates only by 1%. Inspection of the data shows that the additional peptides originate both from the previously not covered regions and from the most dense elution times. More than 80% of all identified peptides were commonly identified by both methods (Fig. 2*D*). In other applications, such as phosphoproteomics, the gains by py_diAID were much larger (see below).

### *Deep Proteome Coverage in Short LC Gradients*

We next investigated if coupling our optimized dia-PASEF methods with project-specific, in-depth libraries yields higher peptide identification and improves quantification accuracy. To generate such an in-depth library, we separated 15 µg of the HeLa sample that we also used for single dia-PASEF acquisitions into 48 concatenated fractions using high-pH reverse-phase chromatography of the Spider fractionator (see the Experimental Procedures section) (32). These fractions were measured in DDA–PASEF mode and again

analyzed with FragPipe and its SpecLib workflow. We compared our "reference library" generated with limited sample amount (2.5 µg proteolytic digest) and 24 fractions to the new one with ample sample amount (15 µg) and twice as many fractions. As expected, the latter was substantially larger, containing 45% more peptides (counting all modifications) and 13% more proteins. Altogether, this deep library constructed from 21 min runs comprised 124,155 peptides and 8439 different protein groups (Fig. 3, *A* and *B*).

Next, we compared single dia-PASEF runs with reference *versus* deep library using DIA-NN and found a corresponding increase in the proteome depth (39% more peptides and 12% more proteins) (Fig. 3, *C* and *D*). Using the deep library identified 76,214 ± 1021 peptides and the reference library 51,711 ± 641 peptides (Fig. 3*C*). With the deep library, an astounding 7056 ± 8 proteins were identified with our optimized acquisition scheme in each of four replicate runs on average. Specifically, with the reference library, DIA-NN reported 14% significant protein identifications on the basis of one peptide, and this percentage decreased slightly to 11% with the deeper library (Fig. 3*D*).

Quantitative reproducibility between the quadruplicates was virtually identical when using the reference or deep library (4.5% *versus* 4.4% on protein level and 12.1% *versus* 13.45% on peptide level) (Fig. 3*E*). Taken together, we found that single-run identification benefited from a project-specific in-depth library while maintaining the accuracy of quantification. We therefore used the library of 48 fractions for all 21 min runs to generate equivalent libraries for evaluating a range of gradient lengths as described next (referred to as "project-specific deep libraries").

We next investigated the effect of even shorter gradients as well as somewhat longer gradients on proteome depths and quantitative accuracy. As before, each library was acquired with DDA–PASEF and 15 µg HeLa lysate separated into 48 fractions. Extending the gradient to 44 min (30 SPD method on the Evosep One system) identified an average of 7756 ± 6 proteins based on 100,900 ± 634 peptides (including all modifications). This represents an identification increase of 10% on protein level in comparison to the 21 min gradient. The median CV between the quadruplicates was 4% at the protein level for these technical replicates, and 7393 protein groups had CVs below 20% (Fig. 4, *A* and *C*).

We expected that the fast scan rate of the timsTOF, together with our optimized method, might still accurately measure a large part of the proteome even in very short gradients (6, 27). Indeed, the 100 SPD method (11 min gradient) still identified 6285 ± 18 proteins (59,811 ± 368 peptides). Quantitative accuracy reported by DIA-NN did not suffer and remained at a median CV of 4%. Taking only the proteins with CVs equal or below 20%, the 100 SPD method still resulted in 6121 proteins, covering 83% of proteins that could be accurately quantified with the 44 min gradient while substantially reducing the analysis time (Fig. 4*A*). Rank order reproducibility was also high for these technical replicates for all gradient lengths (supplemental Figs. S9 and S10, $r = 0.999$ for proteins and $r = 0.992$ for peptides). As expected, the number of peptides identified per minute decreased when increasing the gradient length, whereas the 11-min gradient reached the highest numbers (9330 peptides per minute translating to 155 peptide identifications per second at the apex, Fig. 4*D*).

In conclusion, our data show that our improved workflow constitutes a powerful technological platform capable of accurately quantifying a large part of the proteome at high throughput.

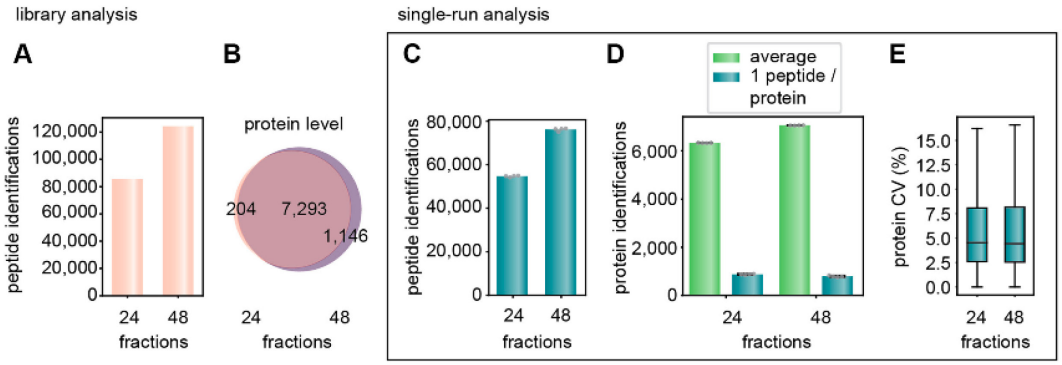



FIG. 3. **Workflow optimization for the 21 min gradient with project-specific deep libraries.** *A*, peptides identified of the reference *versus* the project-specific deep library for 21 min runs. *B*, shared proteins and depth on the protein level in the two libraries. *C*, average peptide identification of four single-run injections. These data and the one in (*D*) and (*E*) were generated from quadruplicate injections of 200 ng tryptic HeLa digest acquired with a 21 min gradient and searched with the reference (24 fractions) or project-specific library (48 fractions). *D*, average protein identifications and identifications with only one peptide in the single runs. *E*, CVs at the protein level based on the MaxLFQ algorithm of DIA-NN. Boxplots show the median (*center line*), 25th, and 75th percentiles (*lower* and *upper box limits*, respectively), and the 1.5× interquartile range (*whiskers*). n = 6384 (24 fractions) and 7121 (48 fractions) shown in *C*. DIA, data-independent acquisition.

**Optimal Design of dia-PASEF Windows**

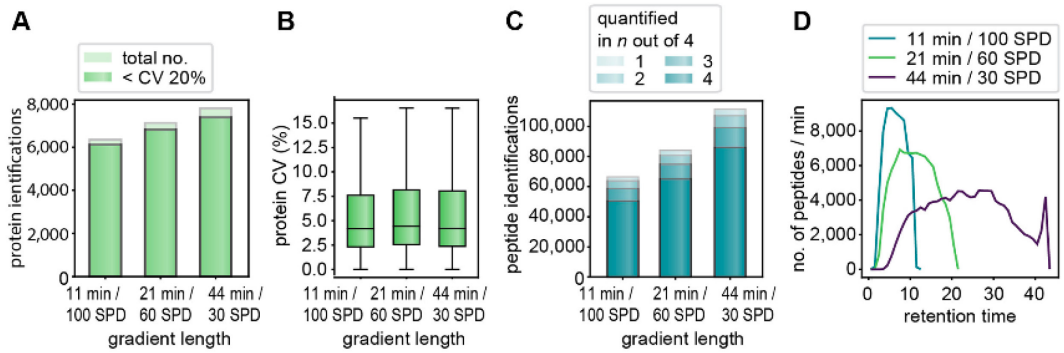



FIG. 4. **Comparison of different gradient lengths/throughput based on single-run analysis.** *A*, all single-run identifications and those with a CV <20% for the 11, 21, and 44 min gradients. *B*, CVs at the protein level based on the MaxLFQ algorithm of DIA-NN. Boxplots show the median (*center line*), 25th and 75th percentiles (*lower* and *upper box limits*, respectively), and the 1.5× interquartile range (*whiskers*). n = 6341 (11 min/100 SPD) and 7121 (21 min/60 SPD), and 7802 (44 min/30 SPD) shown in panel *A*. *C*, analysis of peptide quantification in n out of four technical replicates shows that the large majority is quantified consistently. *D*, the number of peptides per second over the retention time for the three gradient lengths. The data were acquired in quadruplicate injections of 200 ng HeLa digest and analyzed with 48 fraction, DDA–PASEF libraries each recorded with the corresponding gradient length. 11-min library: 8553 proteins and 122,105 peptides; 21-min library: 8439 proteins and 124,155 peptides; 44-min library: 9461 proteins and 175,839 peptides. DDA, data-dependent acquisition; DIA, data-independent acquisition; PASEF, parallel accumulation–serial fragmentation.

### *Comparison of Proteome Results Between DIA-NN and Spectronaut*

The aforementioned analyses were all performed with the DIA-NN package. To determine if our results depend on the software used, we employed Spectronaut (3), another widely used software package (11, 43). This revealed that both packages identified comparable numbers of proteins. For instance, in the 60 SPD method, Spectronaut reported 7285 significant protein groups, whereas DIA-NN reported 7056 significant protein groups (supplemental Fig. S11*A*). In the version tested (Spectronaut 16), this also held for even shorter gradients (6250 *versus* 6285).

Having established that the overall protein numbers are similar, we next investigated the overlap between the found proteins. As DIA-NN has a different protein grouping algorithm from Spectronaut, we performed this analysis on the level of genes and peptide precursors. Employing similar grouping schemes at the gene level showed a high level of concordance, with 548 genes unique to Spectronaut and 208 unique to DIA-NN out of a total of 7668 identified genes for both (supplemental Fig. S11*B*). For the total of 128,002 identified peptide precursors, the discrepancy was somewhat larger, with 28% unique identifications for Spectronaut and 5% for DIA-NN (supplemental Fig. S11*C*). Overall, based on these proteome results, we conclude that the gains achieved by py_diAID are independent of the DIA analysis software used.

### *Rapid Phosphoproteomics With Optimal Isolation Window Design*

Phosphorylation, one of the most prevalent and most studied PTM, refers to the addition of a phosphoryl group—usually on serine, threonine, or tyrosine amino acid residues. This introduces a mass and IM shift on the modified peptides, indicating that analysis of phosphopeptides can benefit from the additional IM dimension in PASEF (44, 45). To date, dia-PASEF has not been explored in a large-scale study of the phosphoproteome or any other post-translationally modified subproteome.

It is well known that the IM dimension separates peptides in clouds primarily reflecting their charge status. In the timsTOF case, Figure 5*A* depicts dense clouds containing doubly, triply, and quadruply charged peptide ions (46). In the case of phospho-enriched samples, projecting the distribution of phosphorylated peptides into the *m/z* and IM space revealed a substantial shift of ion cloud to higher *m/z* values and higher IM values because of the 80 Da increase in their mass, higher charge states, and conformational changes upon phosphorylation (Fig. 5*B*). These observations suggest that dia-PASEF methods need to be tailored for phosphoproteomics. To this end, we first generated an in-depth phospholibrary from EGF-stimulated HeLa cells that were separated into 76 fractions and then enriched for phosphorylated peptides. These enriched fractions were measured with the 60 SPD method, DDA-PASEF in little more than 1 day. We analyzed the results both by FragPipe combined with DIA-NN and by Spectronaut 16 (see the Experimental Procedures section). This generated an in-depth library of 187,730 modified or unmodified peptides, 123,133 phosphopeptides, and 107,154 phosphosites for DIA-NN. Spectronaut 16 obtained very similar results (194,309 modified or unmodified peptides, 132,270 phosphopeptides, and 114,158 phosphosites). The overlap between phosphopeptides was 83% based on the sequence

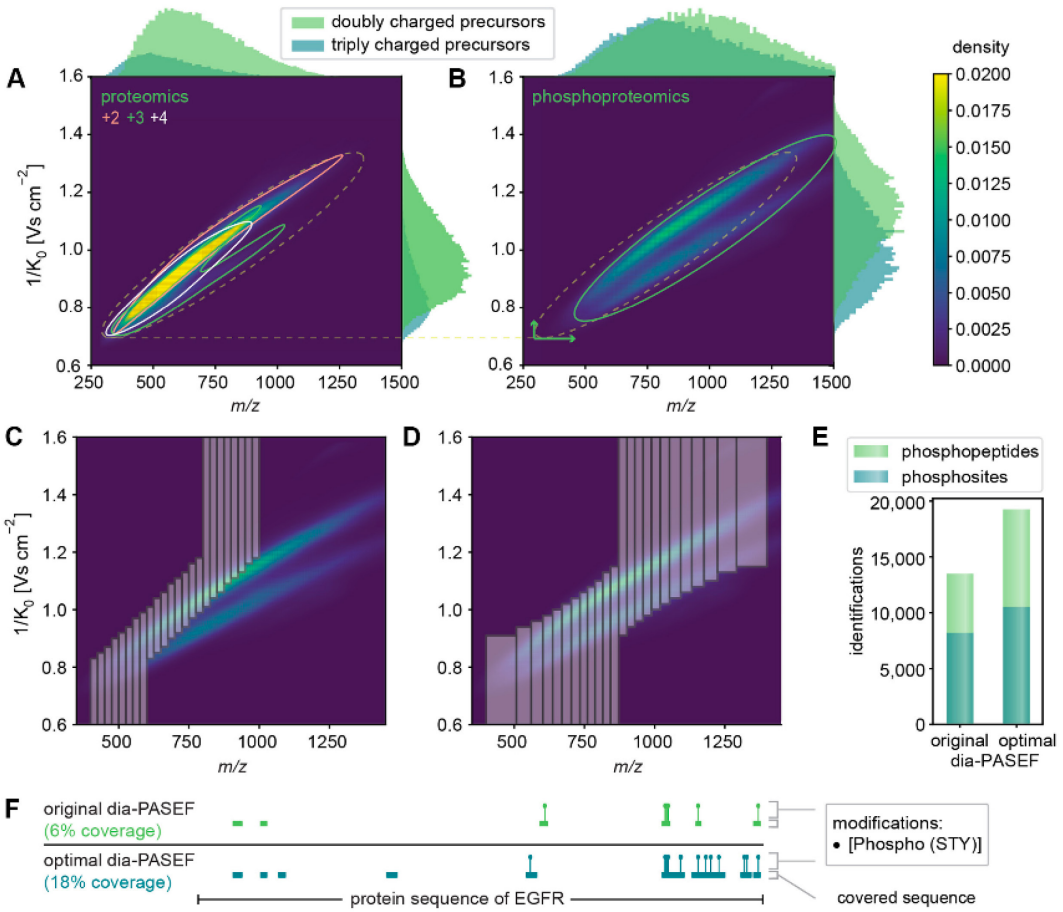



FIG. 5. **Method optimization specifically for phosphoproteomics.** *A*, peptide distribution of a proteomics digest displayed as kernel density estimation dependent on the charge and histograms of the abundance of differently charged precursors based on our deep proteomics library. *B*, peptide distribution of a phosphoproteomics digest displayed as kernel density estimation and histograms of the abundance of differently charged precursors based on our phosphopeptide library. *C*, original dia-PASEF method plotted on top of the phosphopeptide library. *D*, optimal dia-PASEF method tailored to the phospholibrary. *E*, identified phosphosites and phosphopeptides based on quadruplicates of 100 µg EGF-stimulated and enriched HeLa digest, separated within 21 min, and searched with DIA-NN against the phospholibrary. *F*, AlphaMap visualization (47): Protein sequence coverage of the EGF receptor (EGFR) depending on the acquisition method. DIA, data-independent acquisition; EGF, epidermal growth factor; PASEF, parallel accumulation–serial fragmentation.

without considering the modification localizations (supplemental Fig. S14*A*).

When we simulated the coverage of the original dia-PASEF method for the 21 min gradient (6), we found that it only reached a coverage of 34% of phosphopeptide ions in our deep phospholibrary, in contrast to the 81% achieved for unmodified peptides (Fig. 5*C*). Therefore, we used our phospholibrary as input for py_diAID to obtain a dia-PASEF method tailored for phosphoproteomics. This resulted in a theoretical coverage of 93% of all doubly charged and 92% of all triply charged phosphopeptide ions (Fig. 5*D*).

We next utilized this optimal dia-PASEF phosphomethod to measure the samples containing phosphorylated peptides enriched from 100 µg digest of EGF-stimulated HeLa cells. We first analyzed the resulting files with DIA-NN against our deep phospholibrary. In agreement with our simulations, the original dia-PASEF method identified 8199 phosphosites and 13,485 phosphopeptides, whereas the optimal method detected 28% more phosphosites (10,510) and 43% more phosphorylated peptides (19,258) (Fig. 5*E*). The STY ratios of the identified phosphopeptides were similar for both methods, and the optimal method quantified 15,817 peptides modified on a

serine, 3552 modified on a threonine, and 553 modified on a tyrosine (supplemental Fig. S12*A*). To illustrate this further, we mapped the experimentally acquired phosphopeptides to the EGFR sequence essential for transmitting the EGF signal using AlphaMap (47). This revealed that the optimal dia-PASEF phosphomethod doubled the number of detected phosphosites to a total of 14 (Fig. 5*F*).

The intensities of the phosphopeptides detected in our deep FragPipe phospholibrary in DDA–PASEF mode and 76 fractions span almost seven orders of magnitude (supplemental Fig. S13*A*). When searching single dia-PASEF phosphoruns against our phospholibrary using DIA-NN, we found that single short gradients covered 21% of the phosphopeptide sequences, ranging from 12% in the most abundant quintile to 0.3% in the least abundant one (supplemental Fig. S13*A*). Apart from the statistical analysis, the AlphaViz package (48), based on AlphaTims (49), allows visualization of any phosphopeptides of interest. This is shown for the phosphopeptide ELVEPLT[Phospho (STY)]PSGEAPNQALLR on EGFR, where the distinct precursor and fragment peaks are clearly visible in the RT dimension and even more important in the RT–IM plane, supporting the DIA-NN assignment (supplemental Fig. S13, *B* and *C*).

Next, we analyzed the same single-run phospho dataset with Spectronaut. To our surprise—especially given the comparable results at the proteome level—Spectronaut drastically increased the number of identified phosphosites to 28,980 (supplemental Fig. S14*B*). This was even more pronounced for identified phosphopeptides (72,216, supplemental Fig. S14*D*). Accordingly, the common overlap of phosphosites was only 26% (supplemental Fig. S14*B*) and 38% for the sequence of phosphopeptides without taking into account site localization (supplemental Fig. S14*C*).

We do not know the origin of this large discrepancy, but we encourage the providers of these software packages to resolve this, especially as the code is not available for inspection. In the context of our study, we decided to continue with the more extensive Spectronaut results, as they appeared to still correctly represent the regulation in the EGFR signaling experiment described later.

### *In-Depth Phosphoproteomics Analysis of the EGF-Signaling Pathway*

To benchmark our optimal dia-PASEF workflow, we chose the well-studied EGF signaling pathway in HeLa cells. The binding of EGF to the EGFR results in the activation of downstream kinases, which phosphorylate a repertoire of numerous substrates, regulating diverse cellular processes (50). We aimed to quantitatively and accurately measure the differential phosphorylation of proteins involved in this signaling pathway using our rapid and sensitive method. To this end, EGF-treated and control samples were collected in three biological replicates, digested into peptides, and enriched for phosphorylated peptides (see the Experimental Procedures section). Subsequently, we measured the enriched phosphopeptides with dia-PASEF in 21 min and searched the deep phosphopeptide library that we already employed for the method optimization described previously with Spectronaut 16.

With our workflow, we quantified 46,136 phosphorylation sites on 4300 proteins. Of these, 35,537 sites were identified with a high-site localization probability (75%, class I sites (51)) and 20,001 were quantified in all replicates of at least one experimental condition (Fig. 6*A*). Demonstrating the phosphoproteome depth, 62,057 phosphopeptides were reported by Spectronaut to have a modification on a serine, 19,513 on a threonine, and 2788 on a tyrosine (supplemental Fig. S12*B*). The dia-PASEF workflow allowed high reproducible quantification demonstrated by a median Pearson coefficient above 0.92 for replicates within conditions (Fig. 6*B*). Remarkably, a full 26% (5200, 5% FDR) and 10.5% (2117, 1% FDR) of phosphorylation sites were significantly modulated upon EGF treatment (Fig. 6*C*).

As expected, Gene Ontology enrichment analysis revealed strong overrepresentation of proteins involved in the EGFR signaling pathway (Gene Ontology Biological Process [GOBP]) and related pathways among the significantly EGF-upregulated phosphoproteins (Fig. 6*D*). Most are known to be critical for intact EGF signaling. For example, we detected phosphorylation of T693, Y1110, Y1172, and Y1197 on the receptor EGFR itself, Y427 on the adaptor protein Src homology 2 domain–containing-transforming protein C1 (SHC1), Y659 on growth factor receptor bound protein 2–associated protein 1 (GAB1), and on the downstream kinases mitogen-activated protein kinase 2 (MAP2K2) (T394) and mitogen-activated protein kinase 1 and 3 (MAPK1/3) (T185/Y187, T202/Y204) (52) (Fig. 6, *C* and *E*). These phosphosites are typically used to examine EGF signaling with classical methods such as immunoblotting or with targeted MS (9, 10, 53). These approaches, however, only allow relatively low-throughput analyses, which require dedicated assay development procedures or the generation of phosphospecific antibodies. In contrast, by combining the automated phosphoenrichment on the BRAVO platform with the robust Evosep and timsTOF setup, our approach achieves 60 SPD. This allows us to track and accurately quantify the induction of more than 60 phosphorylation events on proteins critical for EGF signaling (part of GOBP–EGFR signaling pathway) within a single 21-min run (supplemental Fig. S15). Importantly, besides the phosphorylations of the classical EGF signaling members, many other signaling events that, for example, result from signaling crosstalk downstream of the EGFR can also be detected, including S897 of the ephrin type-A receptor 2 (EPHA2), S339 of the C-X-C motif chemokine receptor 4 (CXCR4), and T701 of Erb-B2 receptor tyrosine kinase 2 (ERBB2) (supplemental Fig. S15).

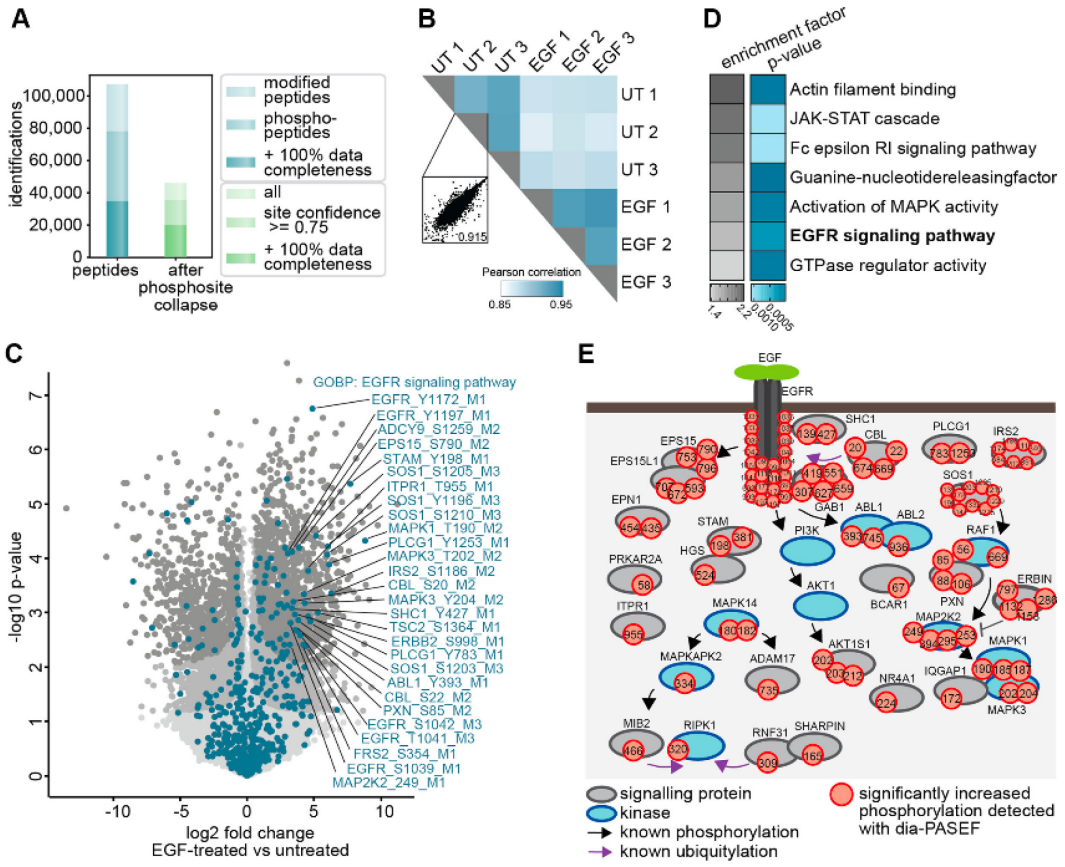



FIG. 6. **The dia-PASEF workflow allows the robust detection of characteristic EGF signaling events.** *A*, numbers of all identified phosphopeptides and phosphosites before and after filtering for localization probability and data completeness. *B*, phosphoproteome Pearson correlation matrix. Scatter plot shows the correlation of replicates within a condition. *C*, volcano plot of phosphosites regulated upon 15 min of EGF treatment in HeLa cells *versus* untreated cells. (Two-sided Student's *t* test, FDR <0.01 = *gray*, FDR <0.05 = *dark gray*). Protein's part of the GOBP term "EGFR signaling pathway" are highlighted in *turquoise*. *D*, Fisher's exact test of proteins with significantly increased phosphosites upon EGF treatment ($p < 0.002$). Enrichment annotations are GOBP, GOMF, and KEGG. *E*, scheme of significantly upregulated phosphosites that were detected in this study and are part of the GOBP term "EGFR signaling pathway" and/or changed significantly upon EGF stimulation (FDR < 0.05). DIA, data-independent acquisition; EGF, epidermal growth factor; EGFR, EGF receptor; FDR, false discovery rate; GOBP, Gene Ontology Biological Process; GOMF, Gene Ontology Molecular Function; KEGG, Kyoto Encyclopedia of Genes and Genomes; PASEF, parallel accumulation–serial fragmentation.

To identify functionally important phosphorylation events not directly linked to EGF signaling, we matched the functionality prediction score developed by Beltrao *et al.* (54) to the upregulated phosphorylation events. We identified 659 phosphosites with a high functional score of >0.5 to be significantly upregulated, which are not part of the GOBP term "EGFR signaling pathway" (FDR < 0.05) (supplemental Data 1). These include EGF-induced phosphorylation of E3 ligases like Mindbomb homolog 2 (MIB2) (S309) and members of the linear ubiquitin chain assembly complex ring finger protein 31 (RNF31) (S466) and sharpin (S165), which are most frequently studied in the context of tumor necrosis factor signaling (supplemental Fig. S15) (55–58). Similarly, phosphorylation of receptor interacting serine/threonine protein kinase 1 (RIPK1) on S320, which prevents tumor necrosis factor–induced cell death, was also increased upon EGF signaling (supplemental Fig. S15) (59, 60). This phosphorylation is mediated by MAP kinase–activated protein kinase 2 (MAPKAPK2), which is activated upon EGF stimulation demonstrated by its increased phosphorylation at T334. These are just some examples of functional candidates whose role in EGF signaling has still to be determined.

Together, this EGF study demonstrates the quantitative capabilities of the dia-PASEF-based phosphoproteomics workflow. We conclude that efficient analysis of ions separated in the IM and *m/z* space enables the investigation of signaling pathways with high sensitivity in a high-throughput manner.

## DISCUSSION

The optimal placement of dia-PASEF windows in the two-dimensional *m/z* and IM space is not trivial. We here developed py_diAID, which is available on GitHub at MannLabs and is installable as a Python module with a command line interface or as a graphical user interface on Windows, Mac, and Linux. It adjusts the isolation window width to the precursor density and optimally positions the isolation design in the *m/z*–IM space. This leads to near-complete theoretical precursor coverage for proteomics. Compared with the original dia-PASEF method (6), the gains for phosphorylated precursors are especially striking (34% *versus* 93%).

MS-based proteomics is a rapidly developing technology. For perspective, to cover 10 thousand proteins, we had to measure the samples for 12 days with 4-h gradients 10 years ago (61). Here, we coupled a robust high-throughput LC system to the TIMS-qTOF instrument employing the rapid sampling speed of a TOF analyzer. It offers short gradients and also low overhead time, enhancing the overall throughput capabilities (28). With this, we generated in-depth project-specific libraries of 9461 proteins in only 13% of the previous measurement time. Furthermore, once the libraries are ready, subsequent proteome characterization using py_diAID-generated methods happens in only 44 min to a depth of 7700 proteins (less than 1% of the measurement time necessary 10 years ago). Our workflow is also twice as fast as currently employed high-throughput screening strategies for cancer proteomics, while achieving greater proteome depth on cell lysate (62–64).

So far, there have been only a few reports of the timsTOF principle on phosphoproteomics (29). Here, we show that this instrument is capable of in-depth phosphoproteomics with very high sensitivity. Specifically, we identified 35,000 phosphosites in only 21 min in triplicates from 100 µg EGF-stimulated HeLa cell digests. Our workflow opens up the possibility to measure multiple pathways in a short time. We demonstrated that quantitatively analyzing the regulated phosphoproteome covers the well-studied EGF signaling pathway together with auxiliary pathways. Interestingly, our workflow is even faster than selected reaction monitoring employed as a targeted screening method for assessing the activation of signaling pathways (9). However, our method is generic to any pathway and applicable in principle to the entire phosphoproteome.

In the current implementation, the dia-PASEF windows are adjusted based on empirical data before the acquisition. These adjustments could also be implemented in real time based on the precursor density achieving an acquisition design optimized to the individual time points of an entire gradient. Furthermore, we employed in-depth libraries. While they can be generated quickly, current developments of *in silico*-generated DIA libraries or direct DIA methods may soon obviate the need for this step. Likewise, we expect that py_diAID will perform similarly for other PTMs.

## DATA AVAILABILITY

All dia-PASEF parameter files required for the acquisition, MS raw files corresponding to the spectral libraries and single-run experiments, and output information from DIA-NN, Spectronaut, and FragPipe have been deposited with the ProteomeXchange Consortium *via* the PRIDE partner (65) repository with the dataset identifier PXD034128. Supplemental data 2 is a roadmap linking the raw files. *Homo sapiens* (taxon identifier: 9606) proteome databases were downloaded from https://www.uniprot.org. py_diAID is a fully open-source package, and the code is freely available under the Apache 2.0 license at https://github.com/MannLabs/pydiAID.

*Supplemental data*—This article contains supplemental data (48, 49).

*Acknowledgments*—We thank Nagarjuna Nagaraj for introducing us to the acquisition software timsControl, which is the basis for py_diAID, and our colleagues in the Department of Proteomics and Signal Transduction at the Max Planck Institute of Biochemistry. We are particularly grateful for the help from Ankit Sinha, Igor Paron, Maria Wahle, Corazon Ericka Mae Itang, Isabell Bludau, Constantin Ammar, and Medini Steger. This study was supported by the Max-Planck Society for Advancement of Science, the Deutsche Forschungsgemeinschaft project "Chemical proteomics inside us" (grant no.: 412136960), and by the Bavarian State Ministry of Health and Care through the research project DigiMed Bayern (www.digimed-bayern.de).

*Author contributions*—P. S., M. T., M. C. T., F. M. H., O. K., A.-D. B., F. M., and M. M. conceptualization; P. S., M. T., M. C. T., F. M. H., O. K., A.-D. B., F. M., and M. M. methodology; P. S., E. V., and S. W. software; P. S., M. T., M. C. T., E. V., F. M. H., O. K., A.-D. B., F. M., and M. M. formal analysis; P. S., M. T., M. C. T., and F. M. H. investigation; P. S., M. T., S. W., O. K., and M. M. writing–original draft.

*Conflict of interest*—M. M. is an indirect investor in Evosep Biosystems. All other authors declare no competing interests.

*Abbreviations*—The abbreviations used are: ACN, acetonitrile; DDA, data-dependent acquisition; DIA, data-independent

acquisition; EGF, epidermal growth factor; EGFR, EGF receptor; FA, formic acid; FDR, false discovery rate; GOBP, Gene Ontology Biological Process; IM, ion mobility; LC, liquid chromatography; MS, mass spectrometry; PASEF, parallel accumulation–serial fragmentation; PTM, post-translational modification; py_diAID, Python package for DIA with an automated isolation design; RT, retention time; SPD, samples per day; TIMS, trapped ion mobility spectrometry.

Received May 31, 2022, and in revised form, July 31, 2022 Published, MCPRO Papers in Press, August 6, 2022, https://doi.org/10.1016/j.mcpro.2022.100279

REFERENCES


1. Aebersold, R., and Mann, M. (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355
2. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B. C., and Aebersold, R. (2018) Data-independent acquisition-based SWATH - MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **14**, 1–23
3. Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinović, S. M., Cheng, L. Y., Messner, S., *et al.* (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics* **14**, 1400–1410
4. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., *et al.* (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, 1–17
5. Chapman, J. D., Goodlett, D. R., and Masselon, C. D. (2014) Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom. Rev.* **33**, 452–470
6. Meier, F., Brunner, A. D., Frank, M., Ha, A., Bludau, I., Voytik, E., *et al.* (2020) diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nat. Methods* **17**, 1229–1236
7. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A., and Yates, J. R. (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1**, 39–45
8. Bekker-Jensen, D. B., Martínez-Val, A., Steigerwald, S., Rüther, P., Fort, K. L., Arrey, T. N., *et al.* (2020) A compact quadrupole-orbitrap mass spectrometer with FAIMS interface improves proteome coverage in short LC gradients. *Mol. Cell. Proteomics* **19**, 716–729
9. Picotti, P., and Aebersold, R. (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9**, 555–566
10. Keshishian, H., McDonald, E. R., Mundt, F., Melanson, R., Krug, K., Porter, D. A., *et al.* (2021) A highly multiplexed quantitative phosphosite assay for biology and preclinical studies. *Mol. Syst. Biol.* **17**, e10156
11. Bekker-Jensen, D. B., Bernhardt, O. M., Hogrebe, A., Martinez-Val, A., Verbeke, L., Gandhi, T., *et al.* (2020) Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat. Commun.* **11**, 1–12
12. Hansen, F. M., Tanzer, M. C., Brüning, F., Bludau, I., Stafford, C., Schulman, B. A., *et al.* (2021) Data-independent acquisition method for ubiquitinome analysis reveals regulation of circadian biology. *Nat. Commun.* **12**, 254
13. Steger, M., Demichev, V., Backman, M., Ohmayer, U., Ihmor, P., Müller, S., *et al.* (2021) Time-resolved *in vivo* ubiquitinome profiling by DIA-MS reveals USP7 targets on a proteome-wide scale. *Nat. Commun.* **12**, 5399
14. Mund, A., Coscia, F., Kriston, A., Hollandi, R., Kovács, F., Brunner, A.-D., *et al.* (2022) Deep visual proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* **40**, 1231–1240
15. Brunner, A., Thielert, M., Vasilopoulou, C., Ammar, C., Coscia, F., Mund, A., *et al.* (2022) Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **18**, e10798
16. Ridgeway, M. E., Lubeck, M., Jordens, J., Mann, M., and Park, M. A. (2018) Trapped ion mobility spectrometry: a short review. *Int. J. Mass Spectrom.* **425**, 22–35
17. Fernandez-Lima, F., Kaplan, D. A., Suetering, J., and Park, M. A. (2011) Gas-phase separation using a trapped ion mobility spectrometer. *Int. J. Ion Mobil. Spectrom.* **14**, 93–98
18. Fernandez-Lima, F. A., Kaplan, D. A., and Park, M. A. (2011) Note: integration of trapped ion mobility spectrometry with mass spectrometry. *Rev. Sci. Instrum.* **82**, 126106
19. McLean, J. A., Ruotolo, B. T., Gillig, K. J., and Russell, D. H. (2005) Ion mobility-mass spectrometry: a new paradigm for proteomics. *Int. J. Mass Spectrom.* **240**, 301–315
20. Meier, F., Brunner, A. D., Koch, S., Koch, H., Lubeck, M., Krause, M., *et al.* (2018) Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol. Cell. Proteomics* **17**, 2534–2545
21. Meier, F., Beck, S., Grassl, N., Lubeck, M., Park, M. A., Raether, O., *et al.* (2015) Parallel accumulation-serial fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J. Proteome Res.* **14**, 5378–5387
22. Beck, S., Michalski, A., Raether, O., Lubeck, M., Kaspar, S., Goedecke, N., *et al.* (2015) The impact II, a very high-resolution quadrupole time-of-flight instrument (QTOF) for deep shotgun proteomics. *Mol. Cell. Proteomics* **14**, 2014–2029
23. Meier, F., Park, M. A., and Mann, M. (2021) Trapped ion mobility spectrometry and parallel accumulation–serial fragmentation in proteomics. *Mol. Cell. Proteomics* **20**, 100138
24. Silveira, J. A., Ridgeway, M. E., Laukien, F. H., Mann, M., and Park, M. A. (2017) Parallel accumulation for 100% duty cycle trapped ion mobility-mass spectrometry. *Int. J. Mass Spectrom.* **413**, 168–175
25. Skowronek, P., and Meier, F. (2022) High-throughput mass spectrometry-based proteomics with dia-PASEF. *Methods Mol. Biol.* **2456**, 15–27
26. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., and Ralser, M. (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44
27. Demichev, V., Szyrwiel, L., Yu, F., Teo, G. C., Rosenberger, G., Niewienda, A., *et al.* (2022) dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nat. Commun.* **13**, 3944
28. Bache, N., Geyer, P. E., Bekker-Jensen, D. B., Hoerning, O., Falkenby, L., Treit, P. V., *et al.* (2018) A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Mol. Cell. Proteomics* **17**, 2284–2296
29. Ogata, K., Chang, C. H., and Ishihama, Y. (2021) Effect of phosphorylation on the collision cross sections of peptide ions in ion mobility spectrometry. *Mass Spectrom.* **10**, 1–8
30. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324
31. Stukalov, A., Girault, V., Grass, V., Karayel, O., Bergant, V., Urban, C., *et al.* (2021) Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **594**, 246–252
32. Kulak, N. A., Geyer, P. E., and Mann, M. (2017) Loss-less nano-fractionator for high sensitivity, high coverage proteomics. *Mol. Cell. Proteomics* **16**, 694–705
33. Yu, F., Teo, G. C., Kong, A. T., Haynes, S. E., Avtonomov, D. M., Geiszler, D. J., *et al.* (2020) Identification of modified peptides using localization-aware open search. *Nat. Commun.* **11**, 4065
34. Yu, F., Haynes, S. E., Teo, G. C., Avtonomov, D. M., Polasky, D. A., and Nesvizhskii, A. I. (2020) Fast quantitative analysis of timsTOF PASEF data with MSFragger and IonQuant. *Mol. Cell. Proteomics* **19**, 1575–1585
35. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., and Nesvizhskii, A. I. (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513–520
36. da Veiga Leprevost, F., Haynes, S. E., Avtonomov, D. M., Chang, H. Y., Shanmugam, A. K., Mellacheruvu, D., *et al.* (2020) Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat. Methods* **17**, 869–870
37. Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J. Proteome Res.* **7**, 29–34
38. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **4**, 923–925

39. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658
40. Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526
41. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., *et al.* (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740
42. Bruderer, R., Bernhardt, O. M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., *et al.* (2017) Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. *Mol. Cell. Proteomics* **16**, 2296–2309
43. Zhang, F., Ge, W., Ruan, G., Cai, X., and Guo, T. (2020) Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. *Proteomics* **20**, 1900276
44. Olsen, J. V., and Mann, M. (2013) Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol. Cell. Proteomics* **12**, 3444–3452
45. Doll, S., and Burlingame, A. L. (2015) Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem. Biol.* **10**, 63–71
46. Meier, F., Köhler, N. D., Brunner, A. D., Wanka, J. M. H., Voytik, E., Strauss, M. T., *et al.* (2021) Deep learning the collisional cross sections of the peptide universe from a million experimental values. *Nat. Commun.* **12**, 1–24
47. Voytik, E., Bludau, I., Willems, S., Hansen, F. M., Brunner, A.-D., Strauss, M. T., *et al.* (2022) AlphaMap: an open-source Python package for the visual annotation of proteomics data with sequence-specific knowledge. *Bioinformatics* **38**, 849–852
48. [preprint] Voytik, E., Skowronek, P., Zeng, W.-F., Tanzer, M. C., Brunner, A.-D., Thielert, M., *et al.* (2022) AlphaViz: visualization and validation of critical proteomics data directly at the raw data level. *bioRxiv*. https://doi.org/10.1101/2022.07.12.499676
49. Willems, S., Voytik, E., Skowronek, P., Strauss, M. T., and Mann, M. (2021) AlphaTims: indexing trapped ion mobility spectrometry-TOF data for fast and easy accession and visualization. *Mol. Cell. Proteomics* **20**, 100149
50. Wee, P., and Wang, Z. (2017) Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers (Basel).* **9**, 52
51. Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., *et al.* (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648
52. Martinez-Val, A., Bekker-Jensen, D. B., Steigerwald, S., Koenig, C., Østergaard, O., Mehta, A., *et al.* (2021) Spatial-proteomics reveals phospho-signaling dynamics at subcellular resolution. *Nat. Commun.* **12**, 7113
53. Mahmood, T., and Yang, P. C. (2012) Western blot: technique, theory, and trouble shooting. *N. Am. J. Med. Sci.* **4**, 429–434
54. Ochoa, D., Jarnuczak, A. F., Viéitez, C., Gehre, M., Soucheray, M., Mateus, A., *et al.* (2020) The functional landscape of the human phosphoproteome. *Nat. Biotechnol.* **38**, 365–373
55. Feltham, R., Jamal, K., Tenev, T., Liccardi, G., Jaco, I., Domingues, C. M., *et al.* (2018) Mind bomb regulates cell death during TNF signaling by suppressing RIPK1's cytotoxic potential. *Cell Rep.* **23**, 470–484
56. Tanzer, M. C., Bludau, I., Stafford, C. A., Hornung, V., and Mann, M. (2021) Phosphoproteome profiling uncovers a key role for CDKs in TNF signaling. *Nat. Commun.* **12**, 6053
57. Goffeau, A., Barrell, G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., *et al.* (1996) Life with 6000 genes. *Science* **274**, 546–567
58. Thys, A., Trillet, K., Rosińska, S., Gayraud, A., Douanne, T., Danger, Y., *et al.* (2021) Serine 165 phosphorylation of SHARPIN regulates the activation of NF-κB. *iScience* **24**, 101939
59. Jaco, I., Annibaldi, A., Lalaoui, N., Wilson, R., Tenev, T., Laurien, L., *et al.* (2017) MK2 phosphorylates RIPK1 to prevent TNF-induced cell death. *Mol. Cell* **66**, 698–710.e5
60. Mohideen, F., Paulo, J. A., Ordureau, A., Gygi, S. P., and Harper, J. W. (2017) Quantitative phospho-proteomic analysis of TNFα/NFκB signaling reveals a role for RIPK1 phosphorylation in suppressing necrotic cell death. *Mol. Cell. Proteomics* **16**, 1200–1216
61. Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., *et al.* (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548
62. Poulos, R. C., Hains, P. G., Shah, R., Lucas, N., Xavier, D., Manda, S. S., *et al.* (2020) Strategies to enable large-scale proteomics for reproducible research. *Nat. Commun.* **11**, 3793
63. Muazzam, A., Chiasserini, D., Kelsall, J., Geifman, N., Whetton, A. D., and Townsend, P. A. (2021) A prostate cancer proteomics database for SWATH-MS based protein quantification. *Cancers (Basel)* **13**, 5580
64. Tully, B., Balleine, R. L., Hains, P. G., Zhong, Q., Reddel, R. R., and Robinson, P. J. (2019) Addressing the challenges of high-throughput cancer tissue proteomics for clinical application: ProCan. *Proteomics* **19**, 1900109
65. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., *et al.* (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552

## Article 2. Synchro-PASEF allows precursor-specific fragment ion extraction and interference removal in data-independent acquisition.

*Mol Cell Proteomics*, 2023 (22, 2, 100489); DOI: 10.1016/j.mcpro.2022.100489

Patricia Skowronek[1], Florian Krohs[2], Markus Lubeck[2], Georg Wallmann[1], Ericka C. M. Itang[1], Polina Koval[1,3], Maria Wahle[1], Marvin Thielert[1], Florian Meier[4], Sander Willems[1,5,*], Oliver Raether[2,*], Matthias Mann[1,6,*]

*[1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [2]Research and Development, Bruker Daltonics GmbH & Co KG, Bremen, Germany. [3]Biomedicine and Neuroscience, Kyiv Academic University, Kyiv, Ukraine. [4]Functional Proteomics, Jena University Hospital, Jena, Germany. [5]Research and Development, Bruker Belgium nv., Kontich, Belgium. [6]NNF Center for Protein Research, Copenhagen, Denmark.*

**These authors contributed equally to this work as senior authors.*

Dia-PASEF with the py_diAID algorithm established variable vertical quadrupole isolation windows and an easily accessible optimal dia-PASEF method design based on precursor densities in the m/z-ion mobility plane. Although it parallelized TIMS and quadrupole, it is limited in the number of quadrupole steps. The quadrupole 'jumped' between positions so that no signal was recorded during the transition. Subsequently, hardware and software advances enabled the fast and precise movement of the quadrupole synchronously with the release of ions from the TIMS tunnel; thus, seamlessly and continuously following the natural shape of the precursor cloud in m/z and ion mobility dimension. The new scan mode, termed synchro-PASEF, increased the sampling rate and acquired fragment intensity, by enabling faster cycle times and efficiently sampling the precursor cloud in diagonal windows.

Moreover, synchro-PASEF slices the precursor in the ion mobility plane because of its continuous movement and scanning. Precursor slicing leaves a specific fragment ion pattern in adjacent synchro-PASEF scans. This lock-and-key mechanism allowed the definition of precursor-fragment and fragment-fragment relationships in highly convoluted DIA spectra and thus enables the reduction of spectral complexity to 'pure fragment spectra'. Together, this prevents misassignment and eliminates interferences used for quantification, resulting in reliable identification and quantification.

Synchro-PASEF adds another dimension to the data cube in timsTOF, but currently the missing or incomplete computational analysis limits the application of this proof-of-principle study of synchro-PASEF. To overcome this, our group is in the process of developing alphaDIA (Wallmann *et al.*, in preparation) a modular software search algorithm which efficiently takes the precursor slicing into account.

Together, synchro-PASEF enables fast cycle times of less than a second and a corresponding increase in fragment intensity. With this it also enhances the duty cycle and increases the ion sampling efficiency. This directly benefits sc-proteomics and high-sensitive applications.

***Contribution:***

Co-authorship. The project has been initiated by Matthias Mann, Patricia Skowronek, Sander Willems and our collaboration partners from Bruker Daltonics. As the lead on the collaboration with Bruker Daltonics, I was involved in the installation of the required hardware components and the biweekly discussions of the study. Further, I analyzed data for the synchronized experiments and discussed scan modes. Additionally, I assisted in the figure design. Finally, I revised and edited the manuscript together with the co-authors.

MCP TECHNOLOGICAL INNOVATION AND RESOURCES

# Synchro-PASEF Allows Precursor-Specific Fragment Ion Extraction and Interference Removal in Data-Independent Acquisition

**Authors**

Patricia Skowronek, Florian Krohs, Markus Lubeck, Georg Wallmann, Ericka C. M. Itang, Polina Koval, Maria Wahle, Marvin Thielert, Florian Meier, Sander Willems, Oliver Raether, and Matthias Mann

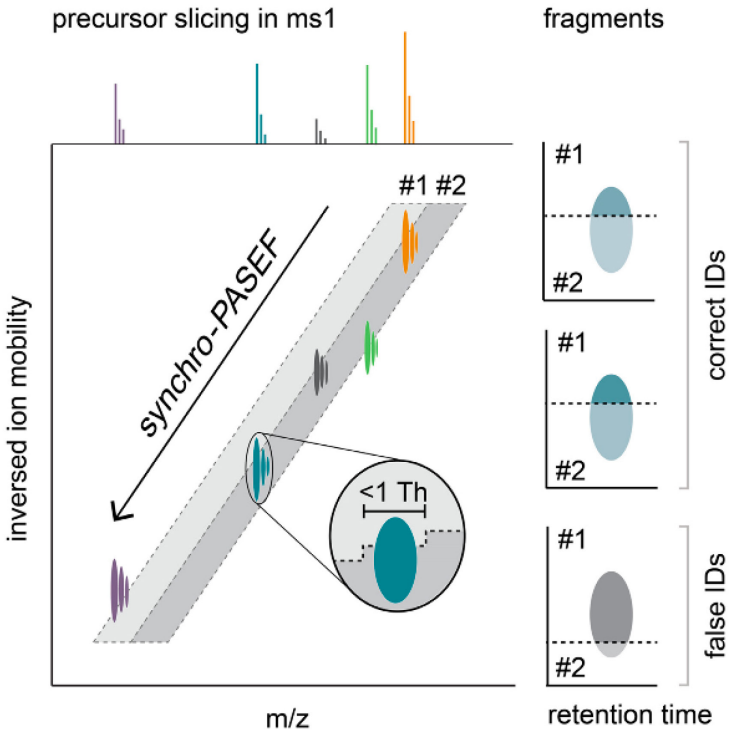
**Correspondence**

sander.willems@bruker.com;
oliver.raether@bruker.com;
mmann@biochem.mpg.de

**In Brief**

The novel scan mode synchro-PASEF efficiently follows the natural shape of the precursor cloud in the m/z and the trapped ion mobility space. This manifests in short cycle times and high sampling frequency of the eluting peptides and fragment signal amplification. Additionally, the seamlessly movement of the quadrupole nearly universally slices the precursor in the ion mobility dimension. The slicing position adds tremendous specificity, allowing deconvolution of the fragment space to 'pure fragmentation spectra' reminiscent of data-dependent acquisition spectra.

**Graphical Abstract**




**Highlights**

- Synchro-PASEF is a highly efficient scan mode for MS-based proteomics.
- It allows very fast cycle times, amplifying the fragment signal three-fold.
- Fragments are directly linked to precursors *via* precursor slicing.
- Precursor slicing enables deconvolution to pure fragmentation spectra.
- Synchro-PASEF unites the benefits of data-dependent and data-independent acquisition.

2023, Mol Cell Proteomics *22*(2), 100489
© 2022 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
https://doi.org/10.1016/j.mcpro.2022.100489

MCP TECHNOLOGICAL INNOVATION AND RESOURCES

# Synchro-PASEF Allows Precursor-Specific Fragment Ion Extraction and Interference Removal in Data-Independent Acquisition

Patricia Skowronek[1], Florian Krohs[2], Markus Lubeck[2], Georg Wallmann[1], Ericka C. M. Itang[1], Polina Koval[1,3], Maria Wahle[1], Marvin Thielert[1], Florian Meier[4], Sander Willems[1,5,*], Oliver Raether[2,*], and Matthias Mann[1,6,*]

**Data-independent acquisition (DIA) methods have become increasingly popular in mass spectrometry–based proteomics because they enable continuous acquisition of fragment spectra for all precursors simultaneously. However, these advantages come with the challenge of correctly reconstructing the precursor–fragment relationships in these highly convoluted spectra for reliable identification and quantification. Here, we introduce a scan mode for the combination of trapped ion mobility spectrometry with parallel accumulation—serial fragmentation (PASEF) that seamlessly and continuously follows the natural shape of the ion cloud in ion mobility and peptide precursor mass dimensions. Termed synchro-PASEF, it increases the detected fragment ion current several-fold at sub-second cycle times. Consecutive quadrupole selection windows move synchronously through the mass and ion mobility range. In this process, the quadrupole slices through the peptide precursors, which separates fragment ion signals of each precursor into adjacent synchro-PASEF scans. This precisely defines precursor–fragment relationships in ion mobility and mass dimensions and effectively deconvolutes the DIA fragment space. Importantly, the partitioned parts of the fragment ion transitions provide a further dimension of specificity *via* a lock-and-key mechanism. This is also advantageous for quantification, where signals from interfering precursors in the DIA selection window do not affect all partitions of the fragment ion, allowing to retain only the specific parts for quantification. Overall, we establish the defining features of synchro-PASEF and explore its potential for proteomic analyses.**

Data-independent acquisition (DIA) has shown enormous promise in recent years because of its signature advantage that every precursor is guaranteed to be fragmented once within every DIA acquisition cycle (1). Although this does not guarantee that every peptide is identified and quantified in every run, it goes a long way to make datasets more reproducible. Importantly, the fragment ions in DIA have elution profiles just like precursors, which dramatically improves to the quality of the fragment peaks. Frequently used software tools such as DIA-NN (2, 3) or Spectronaut (4) match peptides from a library into each of the DIA runs. These libraries were traditionally acquired experimentally with deep data-dependent acquisition (DDA)-based measurements of the proteome of interest but are now often generated directly from the DIA data (5) or *in silico* from the entire proteome using deep learning (6–11). Given these advantages, very deep and quantitatively accurate datasets can now routinely be generated by DIA.

Despite these rapid developments, many researchers still prefer DDA—often because only a single precursor is targeted for fragmentation, leading to simpler fragmentation spectra and more straightforward peptide identifications that may appear more trustworthy.

Following the original implementation of the SWATH acquisition scheme on the Sciex platform by Aebersold *et al.* (1, 12) and then on the Orbitrap instruments, where DIA first demonstrated in-depth proteome coverage (4), we have recently combined DIA with the Bruker timsTOF (13). In this instrument, peptides are confined in a trapped ion mobility spectrometry (TIMS) device by the opposing forces of a gas flow and an electric field and then released from the trap by lowering the electric field strength. Collisional cross-sections and molecular weights of charged peptides are highly correlated, which is the basis of the 'parallel accumulation—serial fragmentation' (PASEF) principle, which boosts fragmentation rates in DDA more than 10-fold at no loss of sensitivity

From the [1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany; [2]Research and Development, Bruker Daltonics GmbH & Co KG, Bremen, Germany; [3]Biomedicine and Neuroscience, Kyiv Academic University, Kyiv, Ukraine; [4]Functional Proteomics, Jena University Hospital, Jena, Germany; [5]Research and Development, Bruker Belgium nv., Kontich, Belgium; [6]Proteomics, NNF Center for Protein Research, Copenhagen, Denmark

*For correspondence: Matthias Mann, mmann@biochem.mpg.de; Oliver Raether, oliver.raether@bruker.com; Sander Willems, sander.willems@bruker.com.



© 2022 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

https://doi.org/10.1016/j.mcpro.2022.100489

(14, 15). For the DIA mode, we very recently described a refined acquisition scheme that simultaneously optimizes the ion mobility and mass ranges for best precursor coverage (16). However, in these schemes, the quadrupole selection window remains fixed for each segment of the acquisition cycle.

We reasoned that a synchronous movement of the quadrupole selection window with the ion mobility scan would better utilize the trapped ion cloud. Furthermore, the rapidly advancing quadrupole window across the mass and ion mobility ranges could lead to much more defined precursor–fragment relationships. Although this scan mode is reminiscent of the one-dimensional movement of the quadrupole in techniques such as SONAR (17, 18) and Scanning SWATH (19), there are important conceptional differences. Simply scanning the m/z range in a timsTOF would imply very wide selection windows to cover the extent of the ion cloud in the mass dimension and would fail to capture most precursors if it was not synchronized with the release of the ions from the TIMS device.

We therefore set out to develop a synchronized PASEF scan mode with very high speed, so as to cover the entire precursor mass range of interest ideally in a single TIMS ramp. We first investigate the analytical properties of 'synchro-PASEF' on simple protein mixtures. Our results demonstrate that it makes efficient use of the ion population for fragmentation, even with very short acquisition cycles. We also find that peptide precursors are nearly universally 'sliced' in the ion mobility dimension by adjacent synchro-PASEF scans and describe how this enables the extraction of pure fragmentation spectra and mitigates interference for quantification.

## EXPERIMENTAL PROCEDURES

### *Development of an Extremely Fast and Synchronized Scan Capability*

Previously in dda- and dia-PASEF, the analytical quadrupole was operated statically, *i.e.*, stepped through a defined number of isolation windows per TIMS ramp. While the quadrupole changed its position, the signal was not recorded, limiting the number of quadrupole steps. In synchro-PASEF, the quadrupole selection window changes incrementally and is constantly synchronized to the TIMS elution. To acquire the complete signal, we routed the digital analog converters (DACs) of the quadrupole radio frequency-amplitude, bias-direct current (DC), and resolution-DC to the same field programmable gate array (FPGA) of the instrument controller that controls the TIMS device. We also implemented the mass-to-set-value calculations based on the quadrupole calibration in the FPGA. Translation of the setpoint isolation masses and widths to the DAC values is implemented on the firmware level.

Together, this results in a more accurate correlation of the quadrupole with the TIMS ramp and when the quadrupole changes the mass position, the MS data are recorded simultaneously. However, these simultaneous events in synchro-PASEF bring their own challenges since there are several delays between triggering the programmed isolation position and recording the fragment spectra: First, the time constant of the electronics and thus the rise time of the actual values that are applied to the quadrupole rods need to be adapted to each other; second, the ions' mass and mobility-dependent transition time from the quadrupole to the time of flight (TOF) detector need to be accounted for. Both result in an apparent isolation mass shift of the quadrupole. Moreover, the transition time of ions through the quadrupole reduces the effective isolation width, given the very high scanning speed. To correct for this, we quantified the combined effects in terms of isolation mass and width offsets for scan rates up to 36,000 Th/s (see below) and determined linear correction functions to calibrate the quadrupole isolation mass and width in alignment with the synchronous movement.

### *Experimental Design and Statistical Rationale*

We performed the experiments with tryptic HeLa cell lysate digest obtained from HeLa S3 cells (American Tissue Culture Collection, ATCC) to illustrate the behavior of synchro-PASEF in a complex sample type and with a simple tryptic protein digest mixture consisting of bovine serum albumin (BSA, P02769), yeast alcohol dehydrogenase (ADH, P00330), yeast enolase (P00924), and rabbit phosphorylase b (P00489) for detailed analyses. This simple protein mixture was obtained from Waters GmbH.

The complete dataset consisted of 13 raw files of single runs for this proof-of-principle study. For experiments involving a complex proteome mixture, 200 ng HeLa digest was used; for experiments involving the simple protein mixture, 200 fmol was injected for LC gradients of 3.2 min [300 samples per day (SPD) on the Evosep system] and 5.6 min (200 SPD) and 400 fmol for LC gradients of 11 min (100 SPD) and 21 min (60 SPD). To analyze simple protein digests in a complex background, we spiked 200 fmol protein mixture into 100 ng HeLa. The figure legends contain additional experimental design and statistical rationale for the respective experiments.

### *Sample Preparation*

The complex proteome mixture used in our experiments was composed of HeLa S3 cells (ATCC). Cells were cultured in Dulbecco's modified Eagle's medium (Life Technologies Ltd) supplemented with 20 mM glutamine, 10% fetal bovine serum, and 1% penicillin-streptomycin. The sample preparation was performed following the in-StageTip protocol (20). After washing the cells in PBS and cell lysis, the proteins were reduced, alkylated, and digested by trypsin (Sigma-Aldrich) and LysC (WAKO) (1:100, enzyme/protein, w/w) in one step. The peptides were dried, resuspended in 0.1% TFA/2% acetonitrile (ACN), and 200 ng digest was loaded onto Evotips (Evosep). The Evotips were prepared by activation with 1-propanol, washed with 0.1% formic acid (FA)/99.9% ACN, and equilibrated with 0.1% FA. After loading the samples, tips were washed once with 0.1% FA.

The simple proteome mixture was composed in equal amounts of the purified and predigested tryptic standard proteins enolase, phosphorylase b, ADH, and BSA. The peptides were reconstituted in 0.1% FA, and either 200 or 400 fmol (see above) were loaded onto Evotips.

### *LC Analysis*

The Evosep One liquid chromatography system separated peptide mixtures at various throughputs using standardized gradients with 0.1% FA and 0.1% FA/99.9% ACN as the mobile phases. The 60 and 100 SPD runs were measured with an 8 cm × 150 μm reverse-phase column packed with 1.5 μm $C_{18}$-beads (Bruker Daltonics) connected to a fused silica 10 μm ID emitter (Bruker Daltonics). For measuring 200 and 300 SPD, we connected the same emitter to a 3 cm × 150 μm reverse-phase column packed with 1.9 μm $C_{18}$-beads (Bruker Daltonics). In both cases, the emitter was operated inside a nano-electrospray ion source (Captive spray source, Bruker Daltonics).

### MS Analysis

The Evosep LC system was coupled with a modified timsTOF Pro 2 mass spectrometer (Bruker Daltonics) to acquire data in synchro-PASEF mode and for comparison in dda- and dia-PASEF mode.

We acquired dda-PASEF with four PASEF scans per top-N acquisition cycle, an intensity threshold of 2500 arbitrary units (a.u.) as indicated by the Bruker acquisition software and a target value of 20,000 a.u. for precursor selection. Precursors that reached this target value were excluded for 0.4 min, which still permitted resequencing in simple mixtures, as shown in section two. Singly charged precursors were filtered out based on their m/z-ion mobility position, and precursors with a mass below 700 Da were isolated with a quadrupole selection window of 2 Th and otherwise with 3 Th.

For dia-PASEF, we used a method generated with our python tool py_diAID (16). That method covered an m/z-range from 350 to 1200 in 12 dia-PASEF scans with two isolation windows per dia-PASEF scan resulting in a cycle time of 1.4 s.

Additionally, we optimized three synchro-PASEF methods based on the tryptic HeLa reference library introduced in our optimal dia-PASEF study (16). The first method consists of four 25 Th isolation windows with a cycle time of 0.56 s and a second method of seven 15 Th windows (0.9 s cycle time). The third method had one 35 Th, two 25 Th, and one 50 Th isolation window (cycle time 0.56 s). The calculated precursor coverages and kernel density plots depicting a precursor ion cloud in the figures was based on this reference library, too.

If not stated otherwise, we used 100 ms to accumulate and elute ions in the TIMS tunnel. The m/z-range was acquired from 100 to 1700, and the ion mobility range from 1.3 to 0.7 V $cm^{-2}$ calibrated with Agilent ESI Tuning Mix ions (m/z, $1/K_0$: 622.02, 0.98 V $cm^{-2}$, 922.01, 1.19 V $cm^{-2}$, 1221.99, 1.38 V $cm^{-2}$). Experiments described to be 'with minimal collision energy' were measured at a constant collision energy of 10 eV (because the software does not allow setting this value to zero), which is not sufficient to cause peptide fragmentation. For experiments with collision energy, the energy was linearly decreased from 59 eV at ion mobility of 1.6 Vs $cm^{-2}$ to 20 eV at 0.6 V $cm^{-2}$.

### Raw Data Analysis

Data were analyzed by manual and visual inspection with the help of our publicly available Python tools AlphaTims (21) and AlphaViz (22). The MaxQuant evidence output table (Version 1.6.17.0) (23, 24) from a single HeLa dda-PASEF run was the basis for calculating the percentage of precursor slicing. We used the reviewed human proteome (UniProt, Nov 2021, 20,360 entries without isoforms) and default settings for the analyses. This comprised a false discovery rate of 1%, two missed cleavages, the cleavage pattern for trypsin, cysteine carbamidomethylation as fixed modification, and methionine oxidation and protein N-terminal acetylation as variable modifications. The mass tolerance was set to 10 ppm for the main search, and known contaminants were excluded from the dataset.

### Statistical Analysis

The raw data were analyzed and visualized with Python (3.8, Jupyter notebook) and the packages pandas (1.2.4), AlphaTims (1.0.4) and numpy (1.20.3, (25)) for data accession and py_diAID (0.0.16), AlphaViz (1.1.15), matplotlib (3.4.2), and astropy (5.1) for visualization.

## RESULTS

### Synchronizing Quadrupole Movement With Trapped Ion Mobility Separation

The timsTOF principle adds a combined trapping and separation device to the regular liquid chromatography quadrupole TOF set up (26–28) (Fig. 1*A*). The TIMS tunnel has two segments, the first of which accumulates the ions for typically 100 ms, while the second releases the ions from the tunnel in high to low order, with the ions that have the largest cross-sections (and typically mass) first. Although the TIMS device is capable of an ion mobility resolution of greater than 100, in complex mixtures, the ion mobility peaks are typically several ms wide.

The downstream quadrupole sets the molecular mass per charge (m/z or Th) isolation range of interest, and the selected precursors are subjected to fragmentation, with all fragment ions appearing at the same ion mobility value as their precursors. The PASEF principle utilizes the speed of the TOF analyzer and the correlation between ion mobility and molecular weight to subject more than one precursor per accumulation cycle for fragmentation. This implies a selection in two dimensions since all peptide precursors require isolation based on molecular weight and ion mobility (13, 15). Note that the selection in m/z dimension is performed explicitly by the quadrupole whereas the selection in ion mobility dimension happens implicitly by the timing of the quadrupole selection in relation to the release of ions from the TIMS device.

In DDA mode, PASEF allows the selection of more than ten precursors per TIMS ramp, which involves rapid placement of the selection window to the precursor positions in a zig-zag pattern (Fig. 1*B*, left panel). In our previous implementation of DIA, the quadrupole is instead stationary with an appropriate selection window for a part of the TIMS ramp and then jumps the next mass range (Fig. 1*B*, middle panel). Although highly efficient (16), neither scan mode naturally follows the 'banana-shaped' ion cloud in the m/z *versus* ion mobility plane.

We reasoned that a continuous and synchronized movement of the quadrupole could slice the ion cloud into several, equally complex TIMS ramps (Fig. 1*B*, right panel). To enable such a synchro-PASEF scan, we built upon the FPGA already present in the instrument that previously controlled among others the TIMS scan. In addition, the DACs of the quadrupole radio frequency-amplitude, the bias-DC, and the resolution-DC are all routed through this FPGA (Experimental Procedures). This allows the quadrupole to move continuously and extremely fast, covering about 8000 Th/s while being aligned with the TIMS release (Fig. 1*C*).

In this particular implementation, the mass spectrometer executes 927 TOF pulses during the 100 ms TIMS ramp, where each TOF acquisition takes 110 µs. The leading edge (lower m/z) of the quadrupole selection window of 25 Th starts at m/z 1170 and moves down to 335 in our method (Fig. 1*C*). Thus, the quadrupole control changes its position every 0.11 ms with a step size of 0.9 Th. In comparison, Scanning SWATH (19) moves the selection window every 2 ms with a step size of 2 Th. Given a width of the ion cloud of about 100 Th of the doubly charged precursors (Fig. 1*B*, top right panel), we found that four TIMS ramps with our 25 Th isolation width

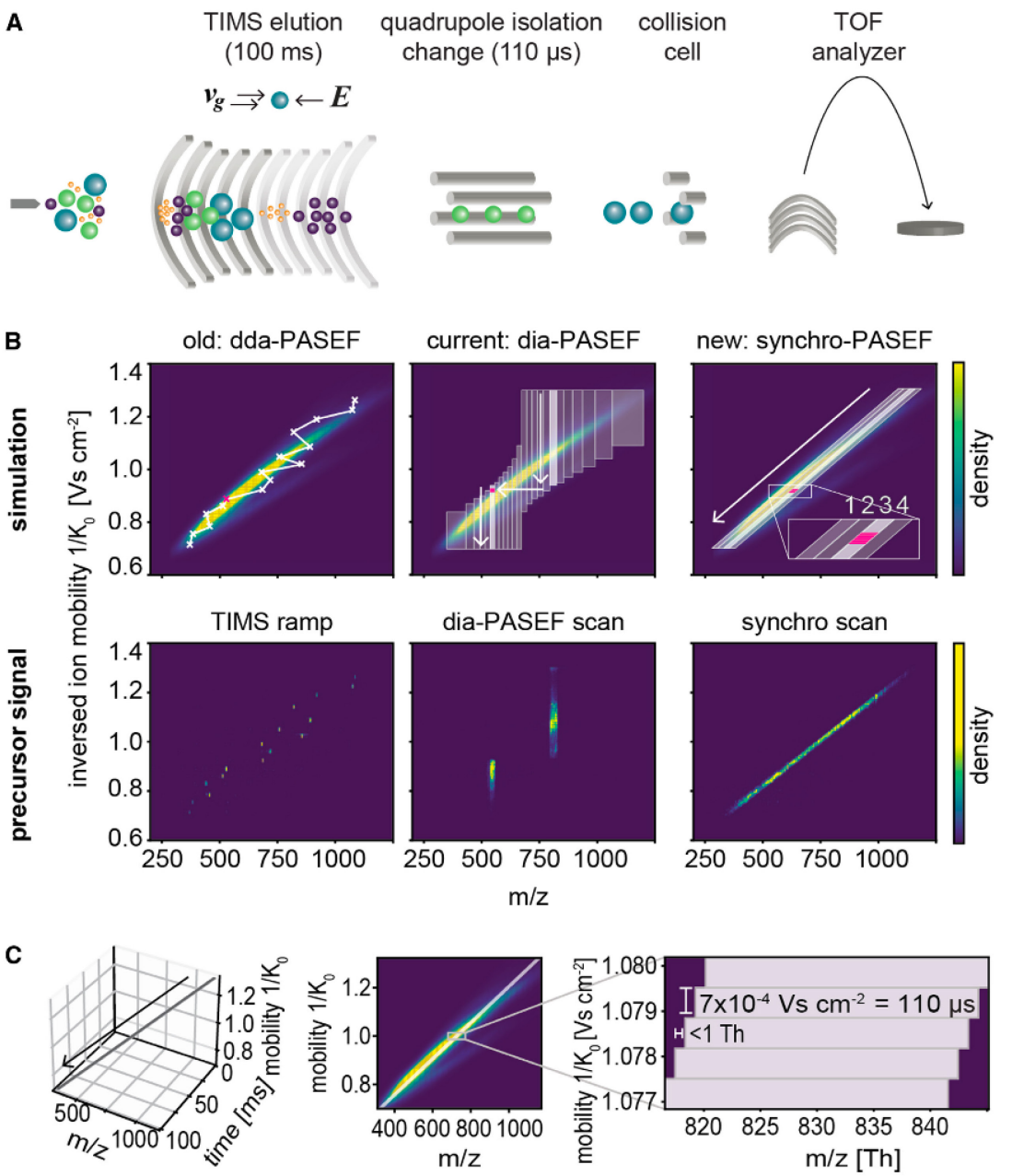



FIG. 1. **The principle of synchro-PASEF precursor isolation.** *A*, the timsTOF ion path: Ions enter the mass spectrometer after electrospray, accumulate and separate in the TIMS tunnel, and are then isolated by a mass-selective quadrupole. A collision cell fragments these precursor ions and a reflector TOF analyzer determines the mass spectrum every 110 us. *B*, *top panel*, simulations of dda-PASEF, dia-PASEF, and synchro-PASEF. The background ion cloud in m/z and ion mobility dimensions shows a kernel density distribution of a reference library from tryptic HeLa (see Experimental Procedures). *Bottom panel*, precursor signal of each scan mode from 200 ng tryptic HeLa digest acquired with minimal collision energy for a 21 min gradient (60 samples per day). dda-PASEF selects the top-N most intense precursors with an isolation

adequately cover these precursors. Beyond this, the window size can be adjusted to any desired value. Based on the reference library, this acquisition scheme leads to a precursor coverage of 76% (see Experimental Procedures). Adding one MS1 level TIMS ramp to measure precursors and transfer time, the total cycle time is only 0.56 s (supplemental Table S1), much faster than our previously optimized dia-PASEF method for short gradients (1.4 s, Experimental Procedures). Those methods were designed for comprehensive coverage, but only contain 8% of precursors per isolation window, in contrast to 19% for synchro-PASEF. The underlying reason for this advantage is that slicing the banana-shaped ion cloud vertically as we had done before requires many more steps than slicing it diagonally along its length while keeping the same width. Thus synchro-PASEF is better adapted to the ion cloud, thereby increasing precursor sampling and making it very fast and efficient.

### *Synchro-PASEF Combines Continuous Fragmentation With Short Cycle Times*

Having established the feasibility and parameters of synchro-PASEF scans on unfragmented precursor selection, we next investigated their benefits for the fragment spectra compared to dda- and dia-PASEF.

As mentioned above, dda-PASEF selects single precursors leading to a striped pattern of fragment ions in the ion mobility (Fig. 2*A*, left panel). Only a small proportion of precursors is covered in each TIMS ramp, but for these, the precursor-fragment and the fragment-fragment relationships are clearly defined. These fragmentation spectra can still be composed of more than one co-eluting precursor in the same precursor isolation window, but the ion mobility selection makes this less likely.

High data completeness is inherent to DIA modes since a broad mass range is recorded deterministically. Furthermore, DIA samples the precursor and fragments multiple times over an elution peak in the retention time and—in the case of dia-PASEF—in the ion mobility dimension. This sampling over the elution peaks in principle enables the reconstruction of the relationship between the precursor and its fragments based on the alignment of the elution profiles. However, the broad isolation windows of DIA lead to highly chimeric fragment spectra since multiple precursors are selected at once, which in practice limits a precise reconstruction of these relationships even in dia-PASEF (Fig. 2*A*, middle panel).

Synchro-PASEF maintains the principle of DIA—fragmenting multiple precursors at once—resulting in complex spectra. However, in contrast to our previous dia-PASEF methods, the fragmentation spectrum is continuous in the m/z and ion mobility dimension, filling nearly the entire m/z *versus* ion mobility plane. Furthermore, there are no artifactual borders due to the quadrupole jumping to different mass ranges (Fig. 2*A*, right panel).

To directly measure the increase in fragmentation yield in synchro-PASEF, we analyzed a low complexity mixture (Experimental Procedures). Taking the co-eluting phosphorylase b peptide GLAGVENVTELK and ADH peptide SISIVGSYVGNR measured with 21 min gradients (60 SPD on the Evosep system) as an example, we only observed a few fragmentation events across the elution peak in DDA (caused by 'resequencing events' in this low complexity mixture) (Fig. 2*B*, left panel). In contrast, our recently reported optimal dia-PASEF method (16) had five fragmentation data points per peak at a cycle time of 1.4 s (Fig. 2*B*, middle panel). As outlined above, the diagonal quadrupole movement of synchro-PASEF leads to much more rapid sampling of the eluting peaks, in this case multiplying the number of fragmentation data points compared to dia-PASEF (Fig. 2*B*, right panel). Importantly, the increased sampling also amplifies the integrated signal for each fragment. For these peptides, synchro-PASEF acquired three-fold more fragment signal than our previous dia-PASEF method, as expected from the nearly three-fold faster sampling (Fig. 2*C*).

### *Extensions of Synchro-PASEF*

So far, we have elucidated the features of synchro-PASEF with an acquisition method that had a constant isolation width of 25 Th, four synchro scans, a ramp time of 100 ms, and a quadrupole step size of 0.9 Th. This is, however, just a reference method and can be adjusted depending on the use case, for example to increase proteome coverage. Next, we explored three acquisition schemes that achieve different balances between quadrupole isolation width, cycle time, precursor coverage, and now also quadrupole step size (Fig. 3*A*).

The first one is the method predominantly used in this paper, the second one has seven 15 Th synchro scans, and the third method has four synchro scans with different window sizes adapted to the precursor cloud density but kept constant throughout each individual synchro scan. Since the

width of 2 to 3 Th. Each TIMS ramp isolates different precursors, and the 5656th TIMS ramp is visualized above (*left panel*). dia-PASEF isolates the precursor cloud in vertical stripes with variable isolation widths. A 20 Th isolation in the sixth dia-PASEF scan at frame 5649 is shown (*middle panel*). Synchro-PASEF follows the precursor cloud naturally and selects many more precursors per isolation window as the previous dia-PASEF schemes. The third synchro scan, frame 5654, is shown (*right panel*). *C*, visualization of the synchro-PASEF quadrupole movement in m/z in Th, inverse ion mobility ($1/K_0$), and TIMS ramp time dimensions in ms. The quadrupole control steps 0.9 Th (later referred to as quadrupole step size) to lower m/z values for every 110 μs TOF trigger, corresponding to a 0.0007 V $cm^{-2}$ downwards step in the inverted ion mobility direction. DDA, data-dependent acquisition; DIA, data-independent acquisition; PASEF, parallel accumulation—serial fragmentation; TIMS, trapped ion mobility spectrometry; TOF, time of flight.

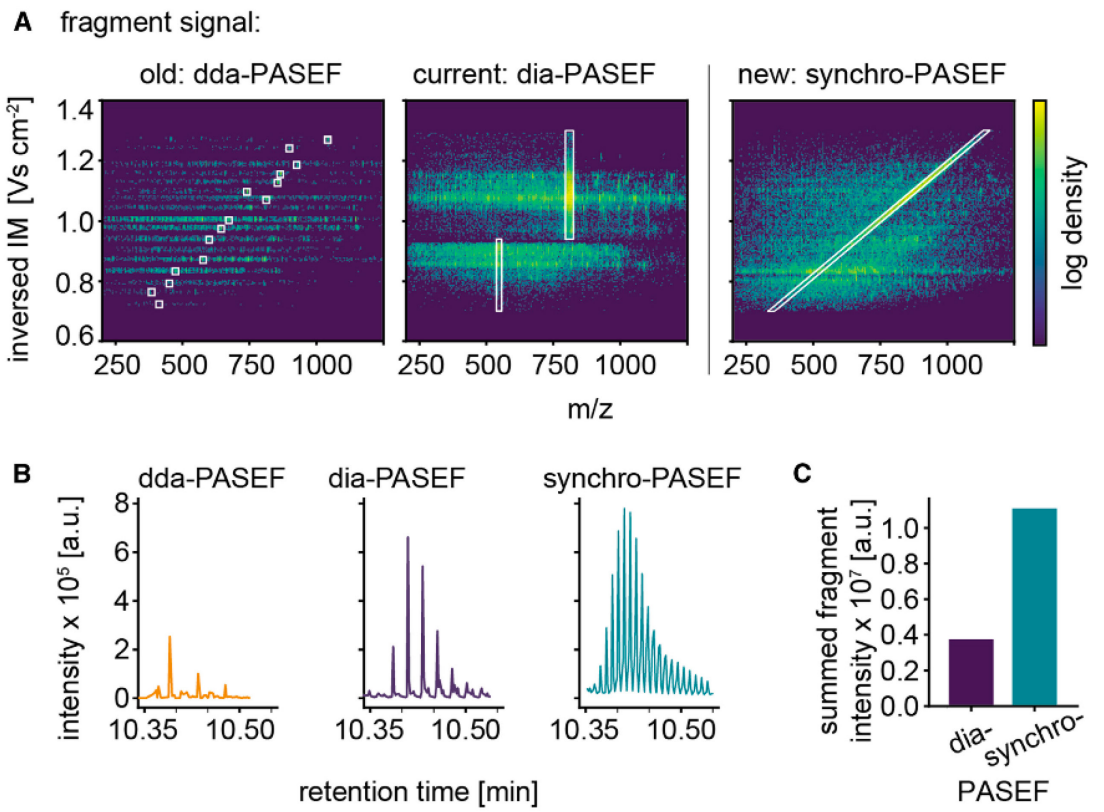



FIG. 2. **Fragmentation behavior for dda-, dia- and synchro-PASEF.** *A*, comparing fragment signals of dda-, dia- and synchro-PASEF. A tryptic digest of HeLa cells was measured in a 21 min gradient. The same frame numbers as in Figure 1*B* are shown but collision energy was applied. *Left panel*, fragmentation spectra of 15 precursors marked with their isolation windows in dda-PASEF. *Middle panel*, isolation windows and continuous fragment acquisition in dia-PASEF. There are no fragments at the transition point and at the precursor low mobility regions. *Right panel*, the isolation window of the third synchro-PASEF scan is shown. Note the uniform fragment distribution in the entire m/z – ion mobility plane. *B*, conversion efficiency to fragments for the three scan modes illustrated with the fragment intensities of GLAGVENVTELK (phosphorylase b) and SISIVGSYVGNR (ADH) co-eluting at a retention time of 10.35 to 10.55 min and an ion mobility of 0.93 to 0.955 V $cm^{-2}$. In DDA mode, the eluting peptides were fragmented three times in this low complexity mixture. Synchro-PASEF sampled the peptide many more times than dia-PASEF due to shorter cycle times. *C*, summed intensities for the fragment signals of the two peptides in (*B*). ADH, alcohol dehydrogenase; DDA, data-dependent acquisition; DIA, data-independent acquisition; PASEF, parallel accumulation—serial fragmentation.

second method has seven synchro scans, it also has the longest cycle time (899 ms instead of 561 ms, Fig. 3*B* and supplemental Table S1). Method 3 has the widest isolation windows and covers 86% of the precursor cloud, which is 10% more than the coverage of method 1 and 2, while having a short cycle time of 561 ms (Fig. 3*C*).

A TOF trigger, which is the duration of one TOF acquisition, takes 107.87 µs (for simplicity rounded to 110 µs above). At a TIMS ramp time of 100 ms, the TOF executes 927 triggers in the standard methods (100 ms TIMS ramp time/107.87 µs = 927 TOF trigger) (Fig. 1*C*), whereas doubling the ramp time to 200 ms also doubles the number of TOF triggers to 1854. In the case of a quadrupole selection window of 25 Th moving from 1170 to 335 Th, a 200 ms ramp time leads to a step size of 0.45 Th ((1170 Th − 335 Th)/1854 TOF trigger = 0.45 Th quadrupole step size) (Fig. 3*D*). Thus, the quadrupole step size can be made even smaller for longer TIMS ramps.

A TIMS ramp of 50 ms implies a step size of 1.8 Th (Fig. 3*D*) and a quadrupole scanning speed of about 16,000 Th/s. We characterized the isolation performance at these very fast scan rates to evaluate the implications for method optimization. Note that the electronics trigger the programmed mass position earlier than the actual recording of the fragment spectrum (Experimental Procedures). We quantified the isolation mass and width effects for scan rates up to 36,000 Th/s and determined linear calibration functions to correct for

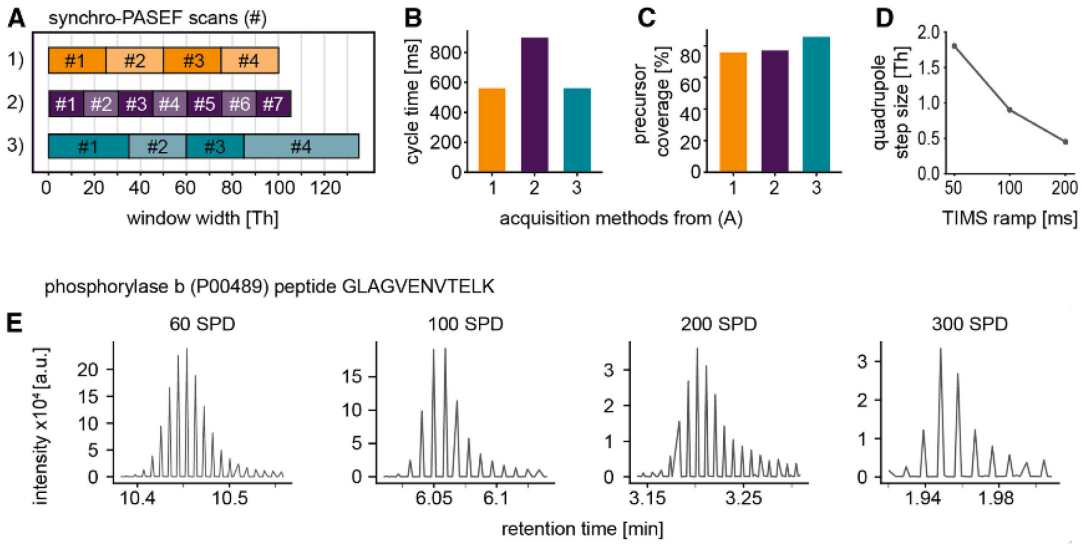



FIG. 3. **Parametrization of synchro-PASEF scan modes.** *A*, acquisition schemes for synchro scan methods with different quadrupole selection window widths. *B*, cycle times of the methods in (*A*). *C*, precursor coverage of methods in (*A*) calculated with the HeLa reference library. *D*, quadrupole step size in dependence on the TIMS ramp time. *E*, precursor elution profiles and sampled data points in the retention time dimension at a throughput of 60 SPD (21 min), 100 (11 min), 200 (5.6 min), and 300 (3.2 min) for the peptide GLAGVENVTELK. PASEF, parallel accumulation—serial fragmentation; SPD, samples per day; TIMS, trapped ion mobility spectrometry.

the synchronously moving quadrupole (supplemental Fig. S1*A*). After calibration, the mass deviation of the center of the 25 Th isolation window at a scan speed of 16,000 Th/s was less than ±0.4 Th for the relevant mass-to-charge range. Deviation of total isolation width was less than ±0.5 Th (supplemental Fig. S1*B*). Note that this is not the accuracy of the leading and trailing edges of the isolation window (see below).

Motivated by this quadrupole precision even at very high scan rates, we wished to investigate how much LC gradients could be shortened while retaining many data points per peak. We acquired our simple protein mixture with the standard method (four times 25 Th synchro scans) with progressively higher throughput (Fig. 3*E*). Taking the same peptide GLAGVENVTELK for illustration, we achieve high peak sampling for all gradients, including the ones with FWHM peak widths around 1 s. Remarkably, even at a throughput of 300 SPD corresponding to an LC gradient of 3.2 min, we still obtained eight data points per peak with a cycle time of 0.56 ms. This suggests that synchro-PASEF is suitable for high throughput, quantitative DIA applications.

### *Precursor Slicing by Synchro-PASEF Deconvolutes DIA Spectra*

Our results up to this point establish superior precursor sampling efficiency and speed for synchro-PASEF compared to previous scanning modes. However, when analyzing the data, we noticed that the diagonal synchro scans frequently intersected or 'sliced' precursors in the m/z and ion mobility plane. In fact, this was the rule rather than the exception and is caused by the fact that scan lines are closer together than the ion mobility peak widths (see below). We also noticed that precursor slicing divides the fragment signals into adjacent synchro scans.

This principle is depicted schematically in Figure 4*A*. When the quadrupole moves from high to low m/z values and from high to low inversed ion mobility values, the trailing edge of the quadrupole selection window (synchro scan #1) moves through the orange precursor and the top parts of all fragments for this precursor are recorded. The leading edge of the next synchro scan slices the precursor at the same position, this time recording the bottom parts of all fragments. Importantly, adding the top and bottom parts reconstitutes the total fragment signal with the same two-dimensional peak shape as the precursor signal.

To analyze how frequent such slicing events occur, we note that at each quadrupole position, the isolation window is 25 Th wide, and the scan diagonal of one synchro-PASEF scan is 3.2 ms high in the ion mobility dimension (supplemental Fig. S2*A*). With our dda-PASEF HeLa run from Figure 2*A*, we determined that the average ion mobility peak width was 4.3 ms (supplemental Fig. S2*B*). A total of 79% of the precursor peaks have an ion mobility profile wider than the synchro scan height. Altogether, 97.4% of the precursors that are

**Precursor-Specific DIA With Synchro-PASEF**

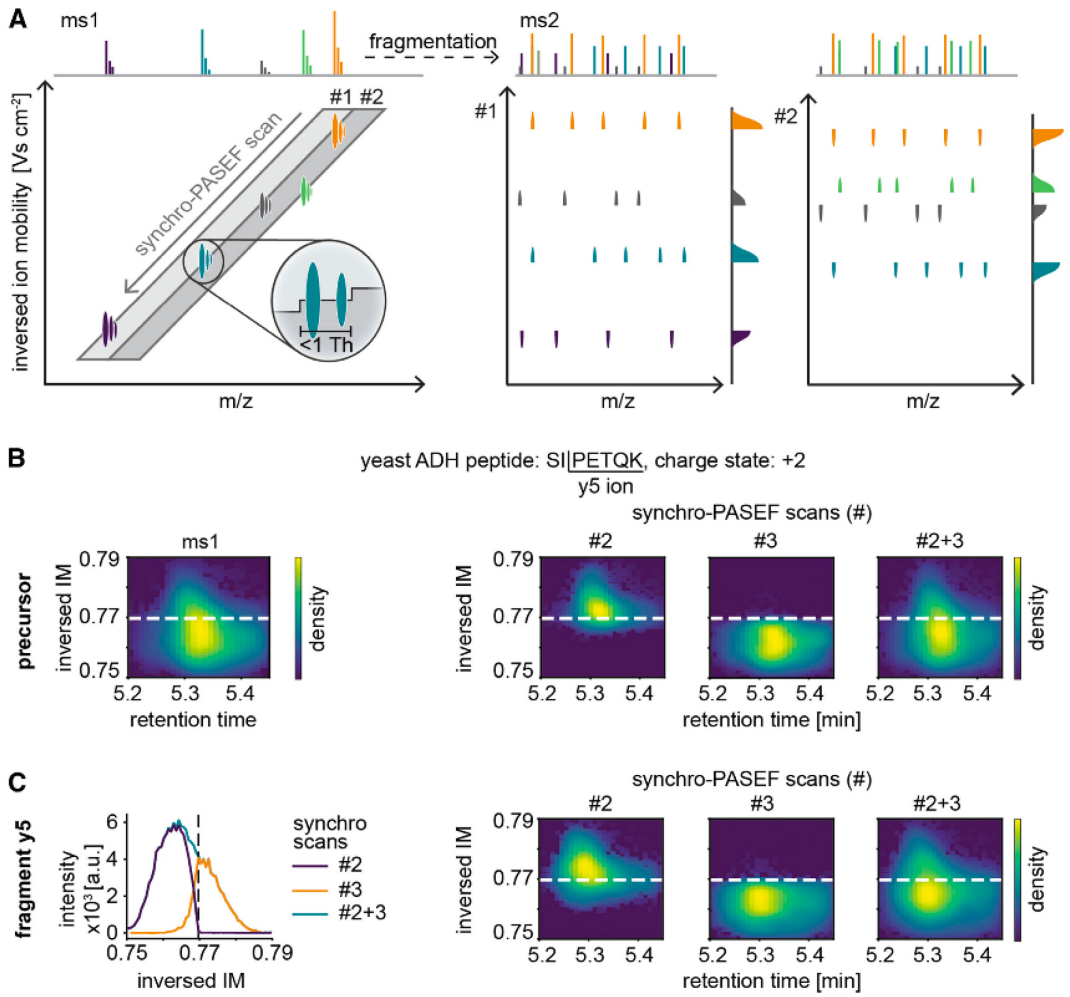



FIG. 4. **Principle of precursor slicing in synchro-PASEF.** *A*, two synchro scans and five precursors are schematically depicted (*left panel*). As the trailing edge of synchro scan #1 moves through the orange precursor, the top parts of all fragments appear at the same time, followed by the bottom part in scan #2 (*middle* and *right panels*, respectively). The slicing position is known for each precursor, and all pieces with the same slicing position belong together. This specifies the precursor–fragment and fragment–fragment relationships. *B*, slicing of the quadrupole of the yeast ADH peptide SIPETQK. For illustration, minimal collision energy was applied. The precursor peak can be quantitatively reconstructed from the sliced parts in two subsequent synchro scans. *C*, same as (*B*) but applying collision energy and focusing on the y5 ion. The traces in adjacent synchro scans add up to the full trace for this transition (*left panel*). The fragment is sliced in exactly the same way as the precursor in (*B*) (*right three panels*). ADH, alcohol dehydrogenase; PASEF, parallel accumulation—serial fragmentation.

within the area covered by the synchro scans are sliced at least once, making precursor slicing a nearly universal phenomenon (supplemental Fig. S2*C*). The percentage for each synchro-PASEF scan varies from 65% to 27% depending on the position of the acquisition scheme and the precursor density in the m/z and ion mobility plane (supplemental Fig. S2*D*). The percentage of precursors that are sliced depends on the height of the synchro scans in ion mobility direction, which in turn varies with the quadrupole selection window width and the TIMS ramp time. Referring back to the three synchro-PASEF methods in Figure 3*A*, the wider isolation windows of method 3 reduce the percentage of sliced precursors to 90.1%. Method 2 with the smallest isolation width (15 Th) also has the smallest height of the diagonal,

resulting in 99.7% of precursors being sliced (supplemental Fig. S2*C*).

To experimentally demonstrate precursor slicing, we measured our simple protein mixture, focusing in detail on the doubly charged yeast alcohol dehydrogenase peptide SIPETQK (modified sequence: acetylated protein N-term-SIPETQK, m/z 422.7). Acquiring this peptide with minimal collision energy in MS1 mode defines the peak shape in retention time and ion mobility space (Fig. 4*B*, leftmost panel). Given the known precursor m/z and ion mobility, it is straightforward to calculate the intersection point of the synchro scan that slices this precursor (dashed line in the figure). Applying synchro scans with minimal collision energy, the top part of the precursor appears in one synchro scan with the edge at the expected slicing position. The next synchro scan contains the remainder of the peak and both parts together reconstitute the full precursor peak (Fig. 4*B*, rightmost panel). This peak reconstitution can also be seen when signal intensities of the two synchro scans are separately plotted as a function of m/z, retention time or ion mobility (supplemental Fig. S3*A*).

Next, we examined how precisely the quadrupole slices. Inspection of the ion cloud after synchro scan isolation but with minimal fragmentation showed good overall alignment between the programmed isolation window and actually transmitted ions, with the leading edge of the quadrupole ejecting lower mass ions somewhat sharper than the trailing edge ejecting higher mass ions (supplemental Fig. S3, *B* and *C*). This is the expected behavior because the stability to instability transition of a quadrupole is steeper for the low mass ejection than the one for high mass ejection (29). When applying collision energy, we observe the slicing behavior for all fragments (shown for y5 in Fig. 4*C*).

Synchro-PASEF inherits the PASEF dimensions retention time, ion mobility, m/z, and intensity from dia-PASEF. However, precursor-fragment or fragment-fragment relationships are defined by the phenomenon of precursor slicing. As illustrated in our simple mixture and in more detail below, two-dimensional precursor slicing conceptually 'deconvolutes' the unprocessed DIA spectra because it establishes precursor–fragment relationships similar to DDA spectra. In the next section, we refine this concept and apply it to qualitatively and quantitatively disentangle individual peptides from the complex background of a cellular proteome.

### *Precise Identification and Quantification* Via *a Lock-and-Key Mechanism*

As described above, synchro-PASEF slices precursors and forms complementary fragment parts in adjacent synchro scans. When put back together at the known slicing position of the precursor, these parts should reconstitute its shape. We found these criteria to be exquisitely specific (see below), reminiscent of a 'lock-and-key mechanism'.

To illustrate, we first depict a precursor and the slicing plane in the m/z, retention time, and ion mobility dimensions (Fig. 5*A*). A number of fragment ions have roughly similar retention times and ion mobilities and could therefore potentially belong to the precursor (Fig. 5*B*). From the calculated slicing position of the precursor, the potential fragments can be verified based on their experimental slicing pattern (Fig. 5*C*). Investigating these peaks more closely, fragments 1, 3, and 4 have the same shape and slicing position as the precursor and are correct identifications. Fragment 2 co-elutes in retention time and ion mobility but is sliced at a different position then calculated for the precursor mass. Fragment 5 has a different shape in the ion mobility dimension and therefore cannot be fragments for this precursor.

Precursor slicing has also implications for quantification based on the requirement that fragment parts added together should reconstitute the shape of the precursor peak. The fraction of the precursor intensity in each synchro scan has fixed ratios given by the slicing position (Fig. 5*D*). If the top and bottom parts of a sliced fragment do not add up to the intensity distribution of the precursor, this typically indicates an interference of another precursor at this fragment mass.

We demonstrate the specificity of this principle on a precursor, which we calculate to be sliced at an ion mobility position of 0.895 V $cm^{-2}$ based on a precursor m/z of 563.786 (Fig. 5*E*). The fragmentation spectrum for this precursor is complex even after filtering for concordance in retention time and ion mobility (Fig. 5*F*) but this can be disentangled by inspecting the individual peaks in detail (Fig. 5*G*). The prominent peaks labeled signals 1 and 2 have a similar shape and slicing position as the precursor. Additionally, their intensity ratios (35% and 65%) correspond to the calculated precursor slicing (Fig. 5*H*). Fragment 3 has the correct slicing position, but the fragment part from synchro scan two is overlayed with another signal, changing the intensity ratios. Fragment 4 is sliced at a different position and has a different shape. As a result, only fragments 1 and 2 are valid. Note that only one part of fragment 3 has an interference; therefore, the second part could still be used for identification and quantification.

Going through all fragment signals leads to an extracted spectrum as simple or even simpler than a DDA spectrum (Fig. 5*G*) (see explanation in supplementary Figs. S4 and S5). The precursor and fragment masses extracted in this way excellently fit the doubly charged peptide LHVDPENFR (P02081) of hemoglobin fetal subunit beta that is present in the HeLa growth medium. Fragment 1 is the y5 ion, fragment 2 the y6, and fragment 3 the y7 ion.

To demonstrate the extraction of DDA-like spectra more objectively, we assigned precursors of the simple protein mixture spiked into a complex HeLa mixture. As before, we filtered spectra for co-elution in retention time and ion mobility and, in a second step, extracted peaks that conform to the constraints imposed by the additional precursor slicing dimension. Because virtually all fragments in the extracted

**Precursor-Specific DIA With Synchro-PASEF**

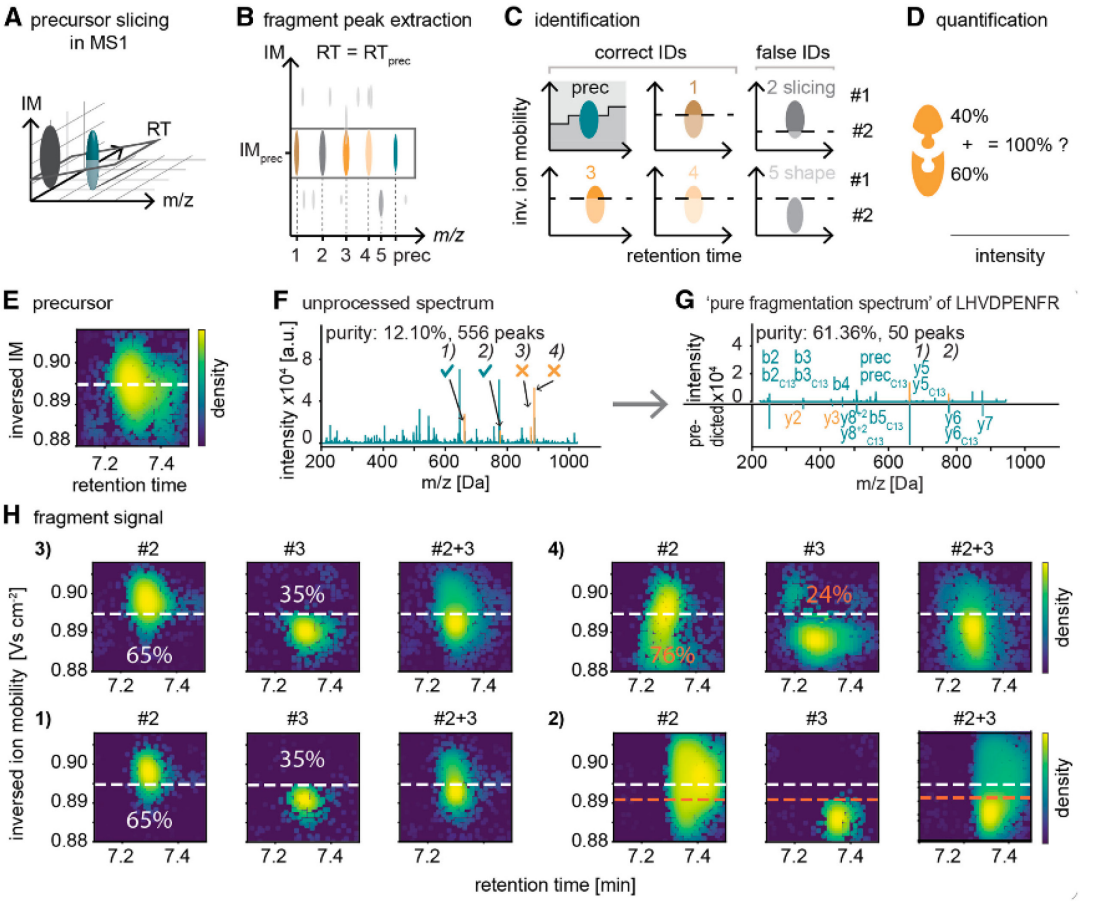



FIG. 5. **Precursor slicing makes identification and quantification more precise.** *A*, the quadrupole slices precursors into at least two parts. *B*, all features with the same retention time and ion mobility distribution as the precursor are potential fragments of the precursor. *C*, precise identification enabled by slicing: fragment 2 is sliced at the wrong position, and fragment 5 does not align in the retention time and ion mobility dimension; hence these fragments do not belong to this precursor. *D*, slicing improves the confidence for correct quantification since both sliced parts should add up to 100% normalized intensity. *E*, a particular precursor in the HeLa digest and its calculated slicing position indicated by a dashed line. *F*, fragment m/z spectrum extracted at the precursor's ion mobility and retention time position. *G*, pure fragmentation spectrum after inspecting the single fragments for correct precursor slicing and the mirrored spectrum predicted by AlphaPeptDeep (6). *H*, fragment signal from (*F*) in detail. m/z positions: (1) fragment y5: 662.33, (2) fragment y6: 777.35, (3) fragment y7: 876.42, (4) 887.46. IM, ion mobility; RT, retention time.

spectra belong to only one precursor (unlike in DDA spectra), we term them 'pure fragmentation spectra'. To computationally handle the slicing of the precursors, we calculated a correlation score between the *in-silico* sliced precursor and the measured fragment partitions. We found that a correlation threshold of at least 0.2 efficiently rejected unrelated peaks. Additionally, we required the intensity ratios for the summed intensities between the upper and lower part of the *in-silico* sliced precursor and the fragment partitions to deviate on an absolute log2 scale less than 3. We found these values to be an optimal balance for generating pure fragmentation spectra while retaining annotated fragments (supplemental Fig. S4). Although we used fixed cutoffs, we believe that future algorithms would benefit most from handling these scores as dynamic features. As a result of our analysis, the ion current in the 29 pure spectra easily explainable by the expected fragmentation spectrum is 58.2% on average, whereas this was only 11.4% without considering precursor slicing

(supplemental Fig. S5). To further validate the quality of the peptide-spectrum matches, we calculated the cross-correlation between the theoretical fragmentation pattern and the acquired spectrum. We used the XCorr score that is a standard score in DDA search algorithms (30, 31). Reassuringly, spectra filtered for precursor slicing showed a higher score compared to spectra without precursor slicing (supplemental Fig. S6), as expected by reducing the impact of background signal on the score. Inspecting these spectra manually, we found that eight of the 29 peptides showed interference in one of the fragment partitions (similar to fragment 3 in Fig. 5*H*), demonstrating that the lock-and-key mechanism can assist in preventing misassignment. For 21 peptides, we found at least one slicing position that did not agree with the expected one for this precursor (similar to fragment 4 in Fig. 5*H*); thus, the slicing position could assist in sorting out wrong precursor–fragment associations (supplemental Table S2).

## DISCUSSION

In this paper, we have developed and characterized a novel scanning method termed synchro-PASEF. A defining feature is that the quadrupole selection window moves diagonally through the m/z - ion mobility plane and synchronously to the ion mobility elution. In this sense, it is the two-dimensional extension of methods such as SONAR or Scanning SWATH. Compared to our previous DDA- or DIA-based PASEF methods, this movement samples the precursor cloud more efficiently, covering up to 19% of all precursors per isolation window. Additionally, it allows short cycle times which is reflected in up to three-fold higher fragment intensities and much higher peak sampling rates compared to our previous standard method. This implies adequate sampling even for very short LC gradients, in turn opening up for high throughput. The fragment signal amplification and higher sampling rate should also be beneficial for quantitative precision.

Another key and unanticipated feature of synchro-PASEF is the fact that precursors are sliced in m/z and ion mobility space by the fast-moving quadrupole selection window. Nearly all precursors are sliced, and their fragments are thereby separated into adjacent synchro scans with high precision and reproducibility. The slicing position deterministically depends on the precursor mass and establishes precursor–fragment and fragment–fragment relationships. We found that the separated fragment parts can be recombined quantitatively with very high specificity. This lock-and-key mechanism can serve as a highly specific criterium for identification and quantification, in effect deconvoluting the unprocessed DIA spectra to the simplicity of DDA spectra or perhaps even better. We plan to use this additional information to eliminate false identifications, pinpoint signals with interference and even correct them by retrieving only the interference-free part of the transition.

Taken together, synchro-PASEF adds another dimension to the already five-dimensional data cube in timsTOF—consisting of retention time, ion mobility, quadrupole selection, m/z, and intensity (28). It offers all the advantages of DIA, including a comprehensive picture of precursor and fragments over the elution profiles, as well as high data completeness resulting from a high ion beam usage with short cycle times. This is combined with the specificity of DDA owing to a defined precursor–fragment and fragment–fragment relationship established by two-dimensional precursor slicing. Clearly, both library-free and library-based analysis could benefit from this principle. In this sense, synchro-PASEF promises to combine the advantages of DDA and DIA.

Here, we only established these defining features and potential of synchro-PASEF but did not actually evaluate proteome coverage or quantitative accuracy in complex proteome mixtures at a global scale. This is due to the current absence of tools that can handle the data structures generated by synchro scans. Once the appropriate algorithms and software are available, it will be interesting to see how the principal advantages of synchro-PASEF translate to increases in proteome depth, coverage, and quantitation. This will also be a good point to optimize and extend the method parameters, similar to what we reported for dia-PASEF to design optimal methods (16). For instance, synchro scan borders could be chosen somewhat wider (as in method 3 in Fig. 3*A*) to cover more of the precursors. Likewise, the borders of the quadrupole selection windows could be offset between adjacent scans to ensure that all precursors are sliced.

Once the downstream informatic infrastructure is in place, synchro-PASEF may be particularly promising for proteome types that are traditionally challenging for DIA. We speculate that the additional confidence in fragment identity would be useful for studying posttranslational modifications, where each additional fragment can be crucial in pinpointing the modification site. Likewise, synchro-PASEF may help disentangle the very complex spectra generated by multiplex DIA approaches (32–34). Since precursor slicing is directly dependent on the precursor mass, it could link fragments to labeled precursors that are co-eluting and are co-isolated for fragmentation.

One of the most significant challenges of DIA is the data complexity, especially with increasing isolation window width and decreasing gradient length. Synchro-PASEF has several unique advantages that address these issues. The diagonal movement of the quadrupole allows for very short cycle times (sub-second) even when using standard isolation widths such as 25 Th. This sufficiently samples peptides and their transitions even for very short gradients, in turn increasing throughput. While the gradient time strongly influences the peak capacity in retention time, the separation in the m/z and ion mobility dimension are independent of this (35). Synchro-PASEF retains all the advantages of traditional DIA window schemes, but additionally disentangles data complexity by

introducing precursor slicing, which establishes a clear precursor–fragment relationship even if there would only be a single timepoint sampled across the elution profile. Since this feature only depends on the precursor mass and the position in the ion mobility dimension, we hypothesize that synchro-PASEF is much less affected by interference in extremely short gradients or in multiplexed DIA.

## DATA AVAILABILITY

All method parameter files required for the acquisition and mass spectrometry raw files have been deposited with the ProteomeXchange Consortium *via* the PRIDE partner (36) repository with the dataset identifier PXD037766 and with MassIVE with the identifier MSV000090925. *Homo sapiens* (taxon identifier: 9606) proteome database and the sequence of BSA (P02769), ADH (P00330), enolase (P00924), and phosphorylase b (P00489) were downloaded from https://www.uniprot.org.

*Supplemental data*—This article contains supplemental data (6, 23).

*Acknowledgments*—We thank our colleagues at the department of proteomics and signal transduction as well as at Bruker Daltonics for discussions and help. We are grateful for the input from Igor Paron, Bianca Splettstoesser and Fabian Opitz. Maximilian Strauss suggested using the cross correlation for retrieving fragments.

*Funding and additional information*—This study was supported by the Max-Planck Society for Advancement of Science, the Deutsche Forschungsgemeinschaft project "Chemical proteomics inside us" (grant no.: 412136960), European Union's Horizon 2020 research and innovation program under grant agreement No 874839 ISLET and by the Bavarian State Ministry of Health and Care through the research project DigiMed Bayern (www.digimed-bayern.de).

*Author contributions*—P. S., S. W., O. R., and M. M. conceptualization; P. S., S. W., M. M., F. K., M. L., and O. R. methodology; P. S., G. W., M. W., E. C. M. I., P. K., and F. M. investigation; P. S., F. K., G. W., P. K., M. T., F. M., S. W., and M. M. formal analysis; P. S. and M. M. witing-original draft.

*Conflict of interest*—M. M. is an indirect investor in Evosep Biosystems. F. K., M. L., S. W. and O. R. are employees of Bruker Daltonics. All other authors declare that they have no conflict of interest with the contents of this article.

*Abbreviations*—The abbreviations used are: ADH, alcohol dehydrogenase; BSA, bovine serum albumin; DAC, digital analog converter; DC, direct current; DDA, data-dependent acquisition; DIA, data-independent acquisition; FPGA, field programmable gate array; PASEF, parallel accumulation – serial fragmentation; SPD, samples per day; TIMS, trapped ion mobility spectrometry.

Received November 14, 2022, and in revised form, December 12, 2022 Published, MCPRO Papers in Press, December 22, 2022, https://doi.org/10.1016/j.mcpro.2022.100489

## REFERENCES

1. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B. C., and Aebersold, R. (2018) Data-independent acquisition-based SWATH - MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **14**, 1–23
2. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., and Ralser, M. (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44
3. Demichev, V., Szyrwiel, L., Yu, F., Teo, G. C., Rosenberger, G., Niewienda, A., *et al*. (2022) dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nat. Commun.* **13**, 3944
4. Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinović, S. M., Cheng, L. Y., Messner, S., *et al*. (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics* **14**, 1400–1410
5. Tsou, C. C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A. C., *et al*. (2015) DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* **12**, 258–264
6. Zeng, W.-F., Zhou, X.-X., Willems, S., Ammar, C., Wahle, M., Bludau, I., *et al*. (2022) AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics. *Nat. Commun* **13**, 7238
7. Mann, M., Kumar, C., Zeng, W.-F., and Strauss, M. T. (2021) Artificial intelligence for proteomics and biomarker discovery. *Cell Syst.* **12**, 759–770
8. Gessulat, S., Schmidt, T., Zolg, D. P., Samaras, P., Schnatbaum, K., Zerweck, J., *et al*. (2019) Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **16**, 509–518
9. Tiwary, S., Levy, R., Gutenbrunner, P., Salinas Soto, F., Palaniappan, K. K., Deming, L., *et al*. (2019) High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nat. Methods* **16**, 519–525
10. Ma, C., Ren, Y., Yang, J., Ren, Z., Yang, H., and Liu, S. (2018) Improved peptide retention time prediction in liquid chromatography through deep learning. *Anal. Chem.* **90**, 10881–10888
11. Zeng, W.-F., Zhou, X.-X., Zhou, W.-J., Chi, H., Zhan, J., and He, S.-M. (2019) MS/MS spectrum prediction for modified peptides using pDeep2 trained by transfer learning. *Anal. Chem.* **91**, 9724–9731
12. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., *et al*. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**. O111. 016717
13. Meier, F., Brunner, A. D., Frank, M., Ha, A., Bludau, I., Voytik, E., *et al*. (2020) diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nat. Methods* **17**, 1229–1236
14. Meier, F., Beck, S., Grassl, N., Lubeck, M., Park, M. A., Raether, O., *et al*. (2015) Parallel accumulation-serial fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J. Proteome Res.* **14**, 5378–5387
15. Meier, F., Brunner, A. D., Koch, S., Koch, H., Lubeck, M., Krause, M., *et al*. (2018) Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol. Cell. Proteomics* **17**, 2534–2545
16. Skowronek, P., Thielert, M., Voytik, E., Tanzer, M. C., Hansen, F. M., Willems, S., *et al*. (2022) Rapid and in-depth coverage of the (phospho-) proteome with deep libraries and optimal window design for dia-PASEF. *Mol. Cell. Proteomics* **21**, 100279
17. Juvvadi, P. R., Moseley, M. A., Hughes, C. J., Soderblom, E. J., Lennon, S., Perkins, S. R., *et al*. (2018) Scanning quadrupole data-independent acquisition, part B: application to the analysis of the calcineurin-interacting proteins during treatment of Aspergillus

fumigatus with Azole and Echinocandin antifungal drugs. *J. Proteome Res.* **17**, 780–793
18. Moseley, M. A., Hughes, C. J., Juvvadi, P. R., Soderblom, E. J., Lennon, S., Perkins, S. R., *et al.* (2018) Scanning quadrupole data-independent acquisition, part A: qualitative and quantitative characterization. *J. Proteome Res.* **17**, 770–779
19. Messner, C. B., Demichev, V., Bloomfield, N., Yu, J. S. L., White, M., Kreidl, M., *et al.* (2021) Ultra-fast proteomics with scanning SWATH. *Nat. Biotechnol.* **39**, 846–854
20. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324
21. Willems, S., Voytik, E., Skowronek, P., Strauss, M. T., and Mann, M. (2021) AlphaTims: indexing trapped ion mobility spectrometry-TOF data for fast and easy accession and visualization. *Mol. Cell. Proteomics* **20**, 100149
22. [preprint] Voytik, E., Skowronek, P., Zeng, W.-F., Tanzer, M. C., Brunner, A.-D., Thielert, M., *et al.* (2022) AlphaViz: visualization and validation of critical proteomics data directly at the raw data level. *bioRxiv*. https://doi.org/10.1101/2022.07.12.499676
23. Prianichnikov, N., Koch, H., Koch, S., Lubeck, M., Heilig, R., Brehmer, S., *et al.* (2020) Maxquant software for ion mobility enhanced shotgun proteomics. *Mol. Cell. Proteomics* **19**, 1058–1069
24. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
25. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., *et al.* (2020) Array programming with NumPy. *Nature* **585**, 357–362
26. Ridgeway, M. E., Lubeck, M., Jordens, J., Mann, M., and Park, M. A. (2018) Trapped ion mobility spectrometry: a short review. *Int. J. Mass Spectrom.* **425**, 22–35
27. Fernandez-Lima, F., Kaplan, D. A., Suetering, J., and Park, M. A. (2011) Gas-phase separation using a trapped ion mobility spectrometer. *Int. J. Ion Mobil. Spectrom.* **14**, 93–98
28. Meier, F., Park, M. A., and Mann, M. (2021) Trapped ion mobility spectrometry and parallel accumulation–serial fragmentation in proteomics. *Mol. Cell. Proteomics* **20**, 100138
29. Paul, W., and Steinwedel, H. (1953) Notizen: Ein neues Massenspektrometer ohne Magnetfeld. *Z. Naturforsch.* **8**, 448–450
30. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989
31. Eng, J. K., Hoopmann, M. R., Jahan, T. A., Egertson, J. D., Noble, W. S., and MacCoss, M. J. (2015) A deeper look into comet—implementation and features. *J. Am. Soc. Mass Spectrom.* **26**, 1865–1874
32. Derks, J., Leduc, A., Wallmann, G., Huffman, R. G., Willetts, M., Khan, S., *et al.* (2022) Increasing the throughput of sensitive proteomics by plexDIA. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01389-w
33. Pino, L. K., Baeza, J., Lauman, R., Schilling, B., and Garcia, B. A. (2021) Improved SILAC quantification with data-independent acquisition to investigate bortezomib-induced protein degradation. *J. Proteome Res.* **20**, 1918–1927
34. Griffiths, J. R., Chicooree, N., Connolly, Y., Neffling, M., Lane, C. S., Knapman, T., *et al.* (2014) Mass spectral enhanced detection of Ubls using SWATH acquisition: MEDUSA—simultaneous quantification of SUMO and ubiquitin-derived isopeptides. *J. Am. Soc. Mass Spectrom.* **25**, 767–777
35. Oliinyk, D., and Meier, F. (2022) Ion mobility-resolved phosphoproteomics with dia-PASEF and short gradients. *Proteomics*, e2200032
36. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., *et al.* (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552

## 3.2. Ultra-high sensitivity mass spectrometry and single-cell proteomics

### Article 3. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation.

*Molecular Systems Biology*, 2022 (18, e10798); DOI: 10.15252/msb.202110798

Andreas-David Brunner[1], Marvin Thielert[1], Catherine Vasilopoulou[1], Constantin Ammar[1], Fabian Coscia[2], Andreas Mund[2], Ole B Hoerning[3], Nicolai Bache[3], Amalia Apalategui[4], Markus Lubeck[4], Sabrina Richter[5,6], David S Fischer[5,6], Oliver Raether[4], Melvin A Park[7], Florian Meier[1,8], Fabian J Theis[5,6], Matthias Mann[1,2]

*[1]Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Martinsried, Germany. [2]NNF Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [3]EvoSep Biosystems, Odense, Denmark. [4]Bruker Daltonik GmbH, Bremen, Germany. [5]Helmholtz Zentrum München - German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany. [6]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany. [7]Bruker Daltonics Inc., Billerica, MA, USA. [8]Functional Proteomics, Jena University Hospital, Jena, Germany.*

After establishing sensitive dia-PASEF methods for the timsTOF platform, we addressed fundamental technological limitations with regards to ultimate sensitivity. Initially, in close collaboration with Bruker Daltonics, we contributed to the construction of an instrument featuring a brighter ion source, various ion optic components and fine-tuned parameters including detector voltage. This modified instrument, subsequently introduced as the timsTOF SCP (for sc-proteomics), demonstrated a 4.4-fold higher ion current. Moreover, we developed low flow rate gradients using the Evosep One LC system. We experimented with rates from 100 nL/min down to 25 nL/min, ultimately standardizing on 100 nL/min for stable operation over hundreds of runs. When compared to 1 µL/min flow rates, we observed a ten-fold signal intensity increase as electrospray theory would predict. Lastly, we incorporated the dia-PASEF scan mode for label-free single-cell quantification, which showed high data completeness and low coefficients of variation (CV).

To study biological responses to drug perturbation at the single-cell level, we trapped HeLa cells in specific cell cycle stages by thymidine and nocodazole (G1, G1/S, G2, G2/M) and measured in total 421 single cells. We utilized those single-cell proteome measurements to classify different cell cycle stages, similar to prior work done with single-cell RNA sequencing. By using cell cycle stage-specific marker proteins, we effectively differentiated arrested cell cycle transitions of G2/M and G1/S stages.

Moreover, we identified several regulated proteins previously not reported in cell cycle transitions, exemplified by the regulation of NACA.

In comparison to scRNA-seq, the single-cell proteome showed higher correlation between the cells on average. When investigating the variability of gene expression as a function of abundance, we found that for proteomics, CVs were very small compared to the ones of scRNA-seq. This resulted in our conclusion that the proteome is more stable than the transcriptome. With these findings, we also defined a single cell ‘core proteome’.

In summary, we co-developed a highly sensitive MS instrument and implemented a very efficient workflow for label-free sc-proteomics. We quantified cellular heterogeneity upon targeted perturbation. This work enabled for the first time in-depth analysis of single cells in a label-free approach, measuring a single cell one by one.

***Contribution:***

Co-authorship, second-author. The project was initiated by Andreas-David Brunner and Matthias Mann. I was involved in the co-development of the MS instrument together with Andreas-David Brunner and Bruker Daltonics. Additionally, I benchmarked and developed the low flow Whisper gradients with Ole B. Hoerning and Nicolai Bache from EvoSep Biosystems. Further, I designed and performed the single-cell experiments and analyzed the proteomic data together with Andreas-David Brunner. I curated the data and revised the text and figures of the manuscript.

Article

TRANSPARENT PROCESS OPEN ACCESS

molecular systems biology

# Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation

Andreas-David Brunner[1], Marvin Thielert[1], Catherine Vasilopoulou[1], Constantin Ammar[1], Fabian Coscia[2], Andreas Mund[2], Ole B Hoerning[3], Nicolai Bache[3], Amalia Apalategui[4], Markus Lubeck[4], Sabrina Richter[5,6], David S Fischer[5,6], Oliver Raether[4], Melvin A Park[7], Florian Meier[1,8], Fabian J Theis[5,6] & Matthias Mann[1,2,*]

## Abstract

**Single-cell technologies are revolutionizing biology but are today mainly limited to imaging and deep sequencing. However, proteins are the main drivers of cellular function and in-depth characterization of individual cells by mass spectrometry (MS)-based proteomics would thus be highly valuable and complementary. Here, we develop a robust workflow combining miniaturized sample preparation, very low flow-rate chromatography, and a novel trapped ion mobility mass spectrometer, resulting in a more than 10-fold improved sensitivity. We precisely and robustly quantify proteomes and their changes in single, FACS-isolated cells. Arresting cells at defined stages of the cell cycle by drug treatment retrieves expected key regulators. Furthermore, it highlights potential novel ones and allows cell phase prediction. Comparing the variability in more than 430 single-cell proteomes to transcriptome data revealed a stable-core proteome despite perturbation, while the transcriptome appears stochastic. Our technology can readily be applied to ultra-high sensitivity analyses of tissue material, posttranslational modifications, and small molecule studies from small cell counts to gain unprecedented insights into cellular heterogeneity in health and disease.**

**Keywords** drug perturbation; low-flow LC–MS; proteomics at single-cell resolution; single-cell heterogeneity; systems biology
**Subject Categories** Cell Cycle; Proteomics
**DOI** 10.15252/msb.202110798 | Received 8 November 2021 | Revised 6 February 2022 | Accepted 8 February 2022
**Mol Syst Biol. (2022) 18: e10798**

## Introduction

In single-cell analysis, biological variability can directly be attributed to individual cells instead of being averaged over an ensemble or complex tissue (Regev *et al*, 2017). While microscopy has always been single-cell based, specialized deep sequencing technologies have achieved this for systems biological approaches (Smith *et al*, 2010; Ramsköld *et al*, 2012; Jaitin *et al*, 2014; Schnitzbauer *et al*, 2017; Schaum *et al*, 2018; Lundberg & Borner, 2019). At the level of proteins, the functional actors of cells, single cells are currently studied by antibody-based technologies, which are by necessity directed against previously chosen targets (Uhlén *et al*, 2015; Stoeckius *et al*, 2017; Jackson *et al*, 2020). In contrast, mass spectrometry (MS)-based proteomics is unbiased in the sense that it measures all proteins within its range of detection (Larance & Lamond, 2015; Aebersold & Mann, 2016). Thus, it would be highly desirable to apply this technology to single cells if the required sensitivity and robustness could be achieved. Previous approaches that employed chemical multiplexing of peptides have labeled a small number of single cells but combined them with a dominant booster channel for MS analysis (Budnik *et al*, 2018; Tsai *et al*, 2020; Schoof *et al*, 2021), which can hamper signal deconvolution (Brenes *et al*, 2019; Cheung *et al*, 2021). Alternatively, proof of principle has been demonstrated for unlabeled approaches using sophisticated sample preparation methods in pico-liter devices (Li *et al*, 2018; Liang *et al*, 2020; Williams *et al*, 2020).

Here, we set out to develop an ultrasensitive MS-based workflow that would allow quantitatively precise and accurate MS proteomics data by injecting single cells one by one into the MS—which we call true single-cell–derived proteomics (T-SCP). To achieve scale, robustness, and community adoption, we aimed to combine technologies that could readily be made commercially available. We

1 Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Martinsried, Germany
2 NNF Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
3 EvoSep Biosystems, Odense, Denmark
4 Bruker Daltonik GmbH, Bremen, Germany
5 Helmholtz Zentrum München – German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany
6 TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany
7 Bruker Daltonics Inc., Billerica, MA, USA
8 Functional Proteomics, Jena University Hospital, Jena, Germany
*Corresponding author. E-mail: mmann@biochem.mpg.de

© 2022 The Authors Published under the terms of the CC BY 4.0 license

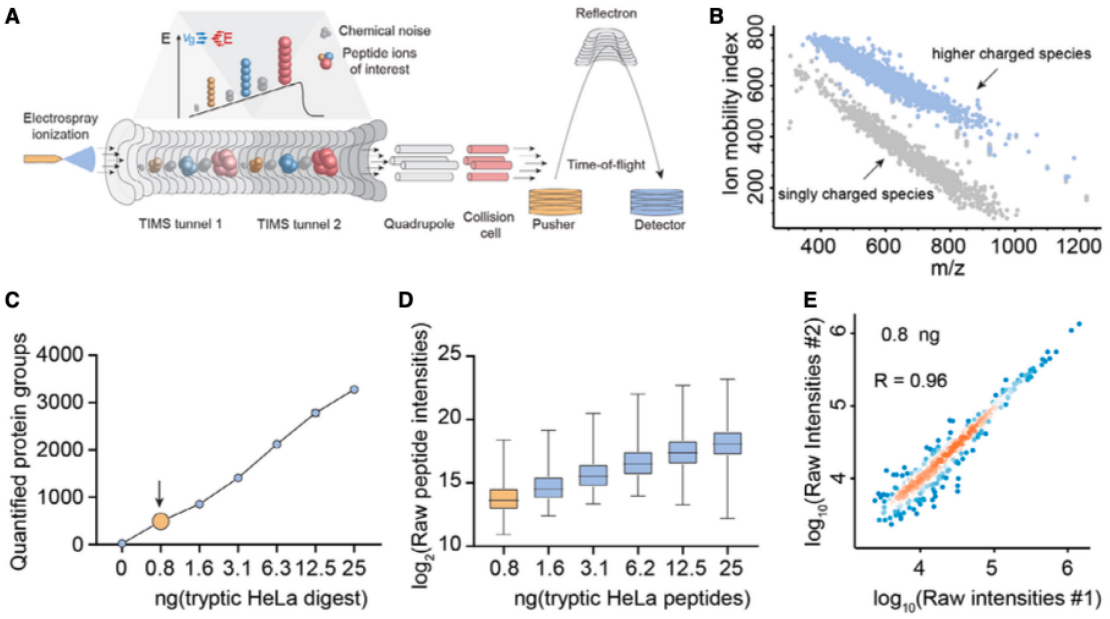



**Figure 1. TIMS enables virtually noise-free spectra and ultra-high sensitivity proteomics.**

A, B The TIMS-qTOF principle separating singly charged background peaks from multiply charged peptide precursor ions, making precursor ions visible at extremely low signal levels (0.8 ng HeLa digest).

C Quantified proteins from a HeLa digest dilution series from 25 ng peptide material down to 0.8 ng (arrow), roughly corresponding to the protein amount contained in three HeLa cells on our initial LC–MS setup (See Material and Methods).

D Linear quantitative response curve of the HeLa digest experiment in C (Box and Whiskers; The middle represents the median, the top and the bottom of the box represent the upper and lower quartile values of the data, and the whiskers represent the maximum and minimum value of the data).

E Quantitative reproducibility of two successive HeLa digest experiments at the lowest dilution (technical LC–MS/MS replicates).

apply our T-SCP technology to a drug perturbation experiment, capturing functional, dynamic responses on a single-cell population.

## Results

### Noise-reduced quantitative mass spectra

We recently introduced parallel accumulation–serial fragmentation (PASEF), a mass spectrometric acquisition scheme in which peptide ions are released from a trapped ion mobility (TIMS) device into the vacuum system in concentrated packages (Meier *et al*, 2015, 2018). Chemical noise is widely distributed as a result of its heterogeneous nature and the 10-fold increased peak capacity due to TIMS (Fig 1A and B; Meier *et al*, 2020b). These precursors can be fragmented in a highly sensitive manner, either in data-dependent (ddaPASEF) or data-independent (diaPASEF) mode, resulting in very high ion utilization and data completeness (Meier *et al*, 2020a). To explore sensitivity limits of our initial LC-MS setup (See Material and Methods), we measured a dilution series of HeLa cell lysate from 25 ng down to the equivalent of a few single cells on a quadrupole time-of-flight instrument (TIMS-qTOF). This identified more than 550 proteins from 0.8 ng HeLa lysate with the DDA acquisition mode and a conservative MaxQuant analysis (Fig 1C; Cox & Mann, 2008). Proteins were quantified with the linear signal response expected from the dilution factors (Fig 1D). Furthermore, quantitative reproducibility in replicates at the lowest level was still excellent (R = 0.96, Fig 1E). Given that the protein amount of a single HeLa cell is as low as 150 pg (Volpe & Eremenko-Volpe, 1970), and accounting for inevitable losses in sample preparation including protein digestion, we estimated that we would need to increase sensitivity by at least an order of magnitude to enable true single-cell proteomics.

### True single-cell proteome analysis

Three main factors govern MS sensitivity: ionization efficiency, transfer efficiency into the vacuum system, and ion utilization by the instrument (Wilm & Mann, 1996). We first constructed an instrument with a brighter ion source, introduced different ion optic elements and optimized parameters such as detector voltage. Together, this led to a more than fourfold higher ion current (Fig 2A). Next, we FACS sorted zero, one, and up to six single HeLa cells in quadruplicate into individual 384 wells, processed them separately, and analyzed them on this modified mass spectrometer. This resulted on average in 843, 1,279, and 1,890 identified proteins for one, two, and six cells, respectively. Note that this analysis

 © 2022 The Authors

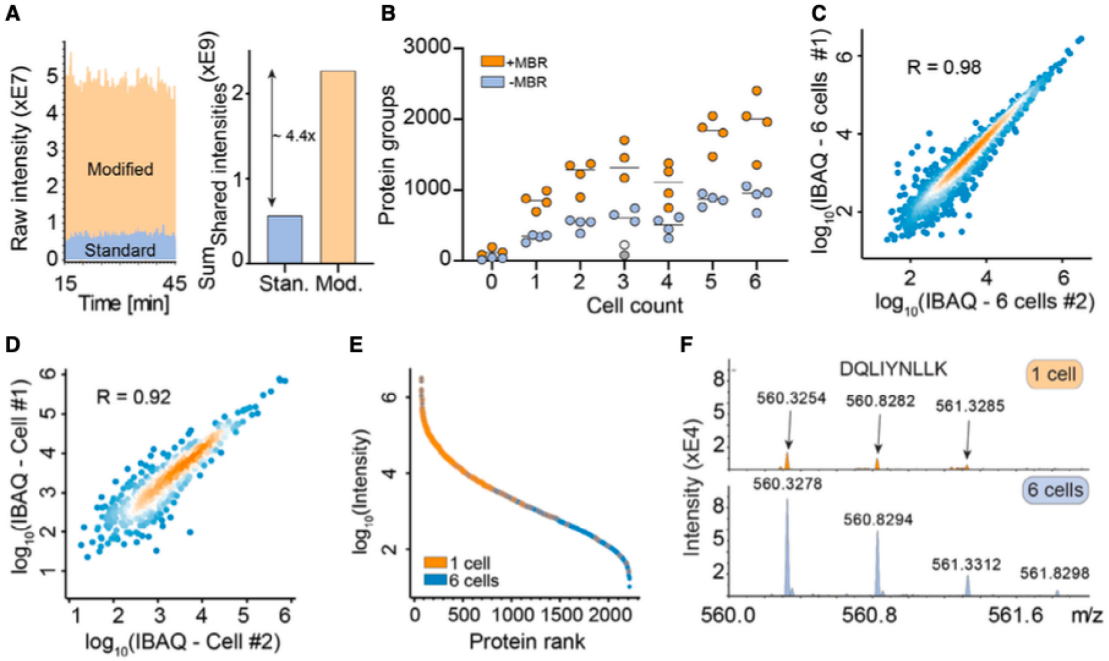



**Figure 2. A novel mass spectrometer allows the analysis of true single-cell proteomes.**

A Raw signal increase from standard versus modified TIMS-qTOF instrument (left) and at the evidence level (quantified peptide features in MaxQuant) (right).

B Proteins quantified from one to six single HeLa cells, either with "matching between runs" (MBR) in MaxQuant (orange) or without matching between runs (blue). The outlier in the three-cell measurement in grey (no MBR) or white (with MBR) is likely due to failure of FACS sorting as it identified a similar number of proteins as blank runs (Horizontal lines within each respective cell count indicate median values).

C Quantitative reproducibility in a rank order plot of a six-cell replicate experiment.

D Same as C for two independent single cells.

E Rank order of protein signals in the six-cell experiment (blue) with proteins quantified in a single cell colored in orange.

F Raw MS1-level spectrum of one precursor isotope pattern of the indicated sequence and shared between the single-cell (top) and six-cell experiments (bottom).

benefited from transferring peptide identifications on the MS1 level, as expected from extremely low sample amounts (Fig 2B). Protein identifications at zero cells are most likely a result of minimal contribution from previous runs since they map to the most abundant proteins of the six-cell measurements in a rank plot (Fig EV1A and B). Quantitative precision and accuracy were high when comparing single cells, not much reduced from comparing six cells (Fig 2C and D). A rank order abundance plot revealed that the measured single-cell proteome preferentially mapped to the higher abundant part of the six-cell proteome, indicating that proteome coverage depended deterministically on overall LC-MS sensitivity (Fig 2E). Inspecting shared peptides between the single-cell and six-cell experiment showed that clearly interpretable precursor isotope patterns were still present at high signal-to-noise levels even at single-cell level following the cell count intensity ratio trend (Fig 2F).

### Ten-fold sensitivity increase

As electrospray (ES) is concentration dependent, sensitivity increases with decreasing flow rate; however, very low flow systems are challenging to operate robustly and are consequently not widely available (Emmett & Caprioli, 1994; Wilm & Mann, 1996; Greguš *et al*, 2020). We recently described a chromatography system that decouples sample loading and gradient formation from the LC-MS run and operates at a standardized flow rate of 1 µl/min for high reproducibility (Bache *et al*, 2018). This flow is fully controlled by a single pump instead of the binary gradients produced by other systems. We found that it worked robustly at flow rates down to 25 nl/min but standardized on 100 nl/min, which enabled stable operation for the entire project with the same column-emitter setup (Fig EV1B and C). ES sprayer diameter and gradient length were optimized for turnover, minimizing carryover and stability.

MS-based T-SCP requires loss-less sample preparation by protein isolation and solubilization, followed by tryptic protein digestion and peptide purification ready for MS analysis (Budnik *et al*, 2018; Li *et al*, 2018; Zhu *et al*, 2018; Greguš *et al*, 2020; Williams *et al*, 2020). We found that small volumes of weak organic solvents in conical 384-well plates provided a versatile and automatable environment for efficient cell lysis and protein digestion in minimal volumes (Fig 3A). Briefly, single cells were sorted into wells

© 2022 The Authors 

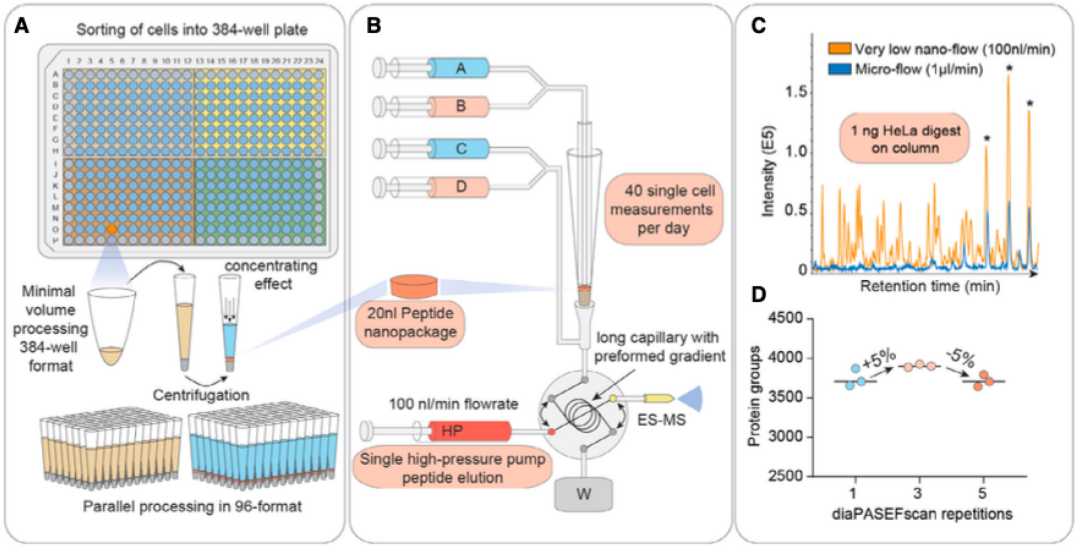



**Figure 3. Miniaturized sample preparation coupled to very low-flow chromatography and diaPASEF.**

A Single cells are sorted in a 384-well format into 1 µl lysis buffer by FACS with outer wells serving as qualitative and quantitative controls. Single cells are lysed and proteins are solubilized at 72°C in 20% acetonitrile, and digested at 37°C. Peptides are concentrated into 20 nl nanopackages in StageTips in a 96-well format.

B These tips are automatically picked and peptide nanopackages are eluted in a sub-100-nl volume. After valve switching, the peptide nanopackage is pushed on the analytical column and separated, fully controlled by the single high-pressure pump at 100 nl/min.

C Base–peak chromatogram of the standardized nanoflow (100 nl/min, orange) and microflow (1 µl/min, blue) gradients with 1 ng of HeLa digest on the StageTip. Asterices indicate polyethylene glycole contaminants in both runs.

D Nanoflow (100 nl/min) and short-gradient diaPASEF method combined. Summation of one to five diaPASEF scan repetitions was used to find the optimum for high-sensitivity measurements at 1 ng of HeLa digest.

containing 1 µl lysis buffer, followed by a heating step and further addition of buffer containing digestion enzymes to a total of 2 µl, all in an enclosed space. Peptides were concentrated in a standard EvoTip device, which resembles the functionality of a StageTip (Rappsilber *et al*, 2007)· into 20 nl nanopackages, from which they were eluted in minimal volumes (Fig 3B). To benchmark the effect of reduced flow rate and the concentrated peptide nanopackage elution, we directly compared signal traces of the normal 1 µl/min to the 100 nl/min setup. For 1 ng peptide material, this resulted in a 10-fold increase in signal (Fig 3C). To achieve high data completeness between hundreds of single-cell measurements, we next replaced ddaPASEF by diaPASEF, in which fragment-level matching is further supported by ion mobility data (Meier *et al*, 2020a). We found that combining subsequent diaPASEF scan repetitions further improved protein identification numbers. Together, the very low flow chromatography and this diaPASEF acquisition mode resulted in the highly reproducible identification and quantification of more than 3,900 HeLa proteins from only 1 ng (Fig 3D), a drastic increase from the 550 identified in our initial setup from a similar amount. Data completeness was at 92% and coefficient of variation (CV) < 10% for the selected scan repetition mode (Fig EV1D). This demonstrates that diaPASEF provides its advantages also at extremely low sample amounts, prompting us to adopt this acquisition mode for the single-cell workflow in the remainder of this work.

### T-SCP dissects arrested cell cycle states

The cell cycle is an important and well-studied biological process that has frequently been used as a test case in single-cell studies (Aviner *et al*, 2015; Ly *et al*, 2017). To investigate if our proteomics workflow could detect biological responses to drug perturbation at the single-cell level, we treated HeLa cells with thymidine and nocodazole to produce four cell populations enriched in specific cell cycle stages (231 cells; Fig 4A). We quantified up to 2,083 proteins per single cell and 2,501 overall using a HeLa dia spectral library with about 4,000 protein groups. This number ranged from a median of 1,018 in G1 to 1,932 in G1/S, 1,572 in G2, and 1,705 in G2/M (Fig 4B). The full data set, even though biologically heterogeneous, showed a median coefficient of variation (CV) of 0.3 across all genes and a clear dependence of the CV on protein intensity levels (Fig EV2A and B). To estimate the total protein amount per cell, we summed all protein signals based on their identifying peptides. Judged by protein amount, G2 cells were approximately 1.8-fold larger than G1 cells; thus, T-SCP correctly reflected the proliferation state, while highlighting a substantial heterogeneity within each cell cycle stage that would have been hidden in bulk sample analysis (Fig 4C). To be able to directly compare single-cell proteomes and cancel out protein abundance differences attributed to varying total protein amounts and

 © 2022 The Authors

identifications of each cell, we normalized our data using the retention time-dependent normalization module offered by the search engine DIA-NN across cells (Demichev *et al*, 2020; Fig EV2C). Furthermore, we stringently filtered our data set for at least 600 protein identifications per cell and more than 15% observations for each protein across remaining single cells (See Material and Methods). The proteomes of the different cell cycle states grouped together in a principal component analysis (PCA)

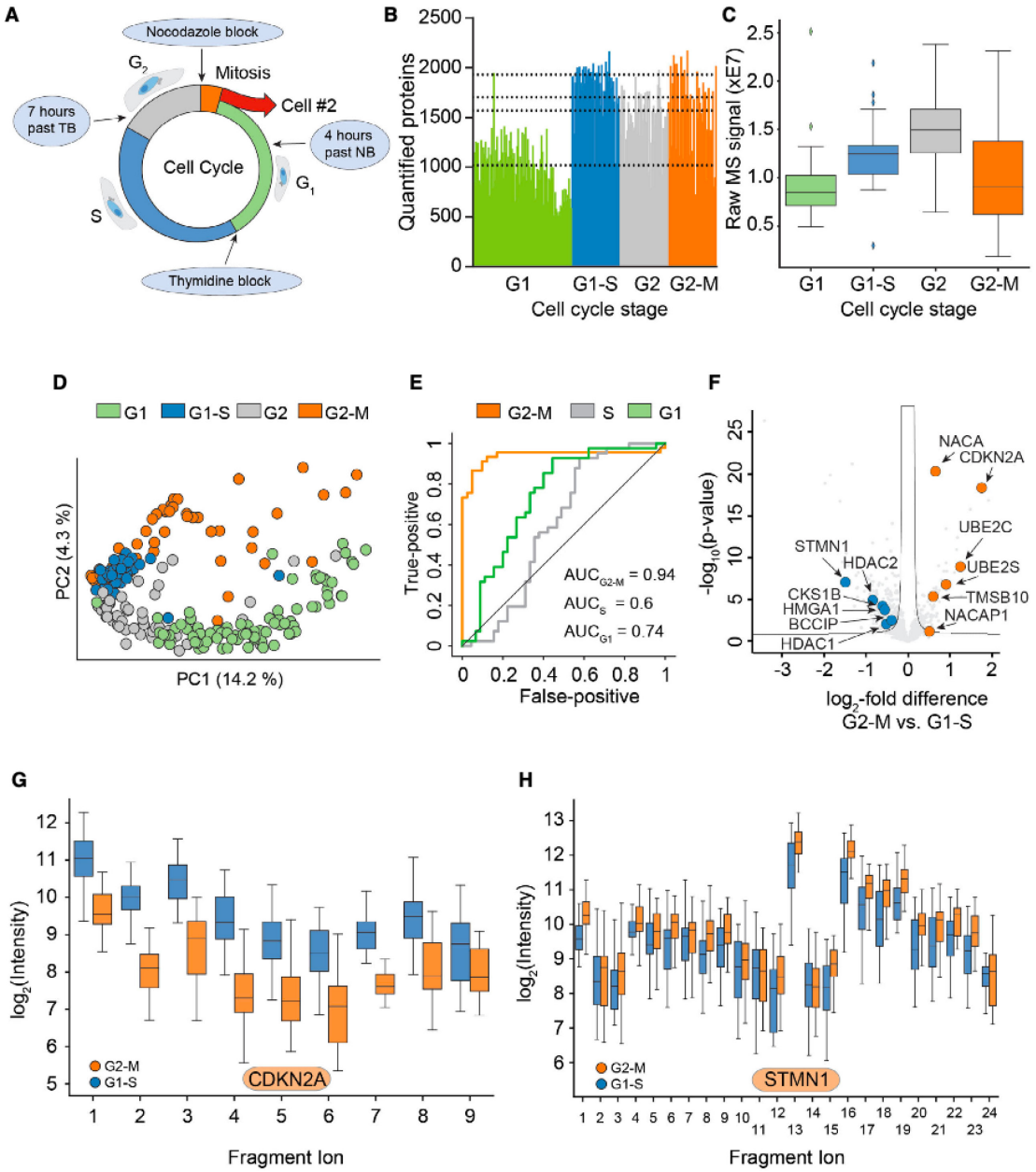



**Figure 4.**

© 2022 The Authors

**Figure 4. T-SCP correctly quantifies cell cycle states.**

A Arresting single cells by drug perturbation.

B Numbers of protein identifications across 231 cells in the indicated cell cycle stages as enriched by the drug treatments in A (Dashed lines indicate the median number of identifications for each respective cell cycle stage).

C Boxplot of total protein signals of the single cells in B after filtering for at least 600 protein identifications per cell and 15% data completeness per protein across cells (G1: $n = 84$; G1-S: $n = 41$; G2: $n = 52$; and G2-M: $n = 45$); (Box and Whiskers; The middle represents the median, the top and the bottom of the box represent the upper and lower quartile values of the data, and the whiskers represent the 1.5× IQR).

D PCA of single-cell proteomes of B.

E Receiver operator curves (ROC) for the distinction between G2-M cells and G1-S cells based on sets of marker proteins for G1, S, and G2-M phase, respectively, with the indicated area under the curve (AUC) scores. G1-S cells were used as positive targets for the G1 and S score, G2-M for the G2-M score.

F Volcano plot of quantitative protein differences in the two drug-arrested states. Arrows point toward colored significantly regulated key proteins of interest (Benjamini–Hochberg corrected multiple-sample *t*-test; FDR = 0.05; S = 0.2).

G Quantitative fragment ion-level data of CDKN2A-associated peptides (FDR < $10^{-15}$; Benjamini–Hochberg corrected multiple-sample *t*-test (Box and Whiskers; The middle represents the median, the top and the bottom of the box represent the upper and lower quartile values of the data, and the whiskers represent the 1.5× IQR).

H Quantitative fragment ion-level data of STMN1-associated peptides (FDR < $10^{-15}$; Benjamini–Hochberg corrected multiple-sample *t*-test (Box and Whiskers; The middle represents the median, the top and the bottom of the box represent the upper and lower quartile values of the data, and the whiskers represent the 1.5× IQR).

plot (Fig 4D). In addition to these drug-perturbed cells, we measured more than 200 untreated ones from two independent cell culture batches. The proteomes of these asynchronous cells distributed well across the cell cycle states, while different passage batches were enriched in the G1 and G2 phase (Fig EV2D). This highlights that biological variation dominates remaining technical variation. The T-SCP data set covered proteins assigned to many cellular compartments, membranes, and biological processes involved in biological regulation, metabolism, transport, and signal transduction at high quantitative precision despite severe systematic perturbation introducing stark biological variation and proteome remodeling (Fig EV2E, Dataset EV1).

Next, we asked whether single-cell proteome measurements can be used to assign cellular states, similar to how single-cell RNA sequencing (scRNA-seq) measurements have frequently been applied to cell type and state discovery, highlighted by cellular atlas projects (Regev *et al*, 2017). In previous proteomics studies, cell populations had been enriched for cell cycle states and sets of regulated proteins had been extracted (Aviner *et al*, 2015; Ly *et al*, 2017). We here selected cell cycle stage marker proteins as the top 60 most differentially expressed in the G2/M-, G1-, or S-phase protein set from Geiger and coworkers (Aviner *et al*, 2015), as it used similar drug treatment on bulk populations and investigated how likely cells from different cell cycle stages could be distinguished (Dataset EV2). We used these marker proteins to set up cell cycle stage-specific scores indicating the likelihood to belong to the respective phase previously used for scRNA-seq cell cycle stage predictions. This model clearly distinguished cells from G2/M and G1/S and also other comparisons (Figs 4E and EV2F; Wolf *et al*, 2018).

Next, we investigated the differentially expressed proteins between the drug-arrested cell cycle stage transition G2/M and G1/S. Among the significantly regulated proteins was a large number of known cell cycle regulators, some of which are highlighted (Dataset EV3; Fig 4F). Quantitative MS data at the fragment ion level were highly significant for these as illustrated by the cell cycle regulator CDKN2A, STMN1, and further examples (FDR < $10^{-15}$, Figs 4G and H, and EV3). Our single-cell data set also highlighted proteins not previously associated with the cell cycle and the G2/M transition. For instance, NACA was clearly identified and regulated (FDR < $10^{-15}$, Fig EV3).

## SC proteomes compared to transcriptomes

Given our set of more than 430 single-cell proteomes, we compared the T-SCP measurements after filtering with similar single-cell RNA sequencing data (scRNA-seq) (Hu *et al*, 2019; Schwabe *et al*, 2020). To achieve technology-independent insights, we selected assays from two widespread scRNA-seq technologies, Drop-seq (Macosko *et al*, 2015) and the lower-throughput SMART-Seq2 (Picelli *et al*, 2014), on the same cellular system. The Drop-seq assay is based on unique molecular identifiers (UMIs) to control for amplification biases in library preparation, whereas the SMART-Seq2 assay is not UMI controlled. Note that MS-based proteomics inherently does not involve any amplification and is not subject to associated artifacts.

Despite subtle differences, HeLa cell culture should reflect a characteristic global distribution of gene and protein expression states (Liu *et al*, 2019). This assumption would allow us to assess self-consistency of the measurement technologies. First, we computed the distribution over all pairwise correlation coefficients of cells within a technology (Svensson, 2020). We found that in the proteome measurement, cells have higher correlation on average than in the droplet-based and the SMART-Seq2 method (Fig EV4A). This is true when analyzing all available genes within each particular data set and also when analyzing all shared genes (Fig EV4A). Comparing all three data sets for gene or protein expression completeness (on average in 49% of the 2,480 proteins observed by MS-based proteomics), protein expression completeness per cell followed a normal distribution (Fig 5A). For SMART-Seq2, this was only 27 and 8% in the droplet-based protocol. Furthermore, when comparing the gene or protein expression completeness on shared gene level, the saturation sequencing effect in the SMARTseq2 data set becomes pronounced. In the Drop-seq data set, this is controlled for using UMI-based protocols (Islam *et al*, 2014; Svensson, 2020; Fig EV4B). Both single-cell RNA-sequencing technology data sets followed a bimodal gene completeness frequency distribution, while single-cell proteomes do not (Fig EV4C).

Next, we investigated whether there were systemic limitations of the detection in the protein measurements. Such effects are discussed for scRNA-seq measurements as "drop-out events" or "zero-inflation," although they are now much reduced in UMI-based protocols (Islam *et al*, 2014; Svensson, 2020). We identified signs of such detection limits as bimodality in the lower abundance range of

 © 2022 The Authors

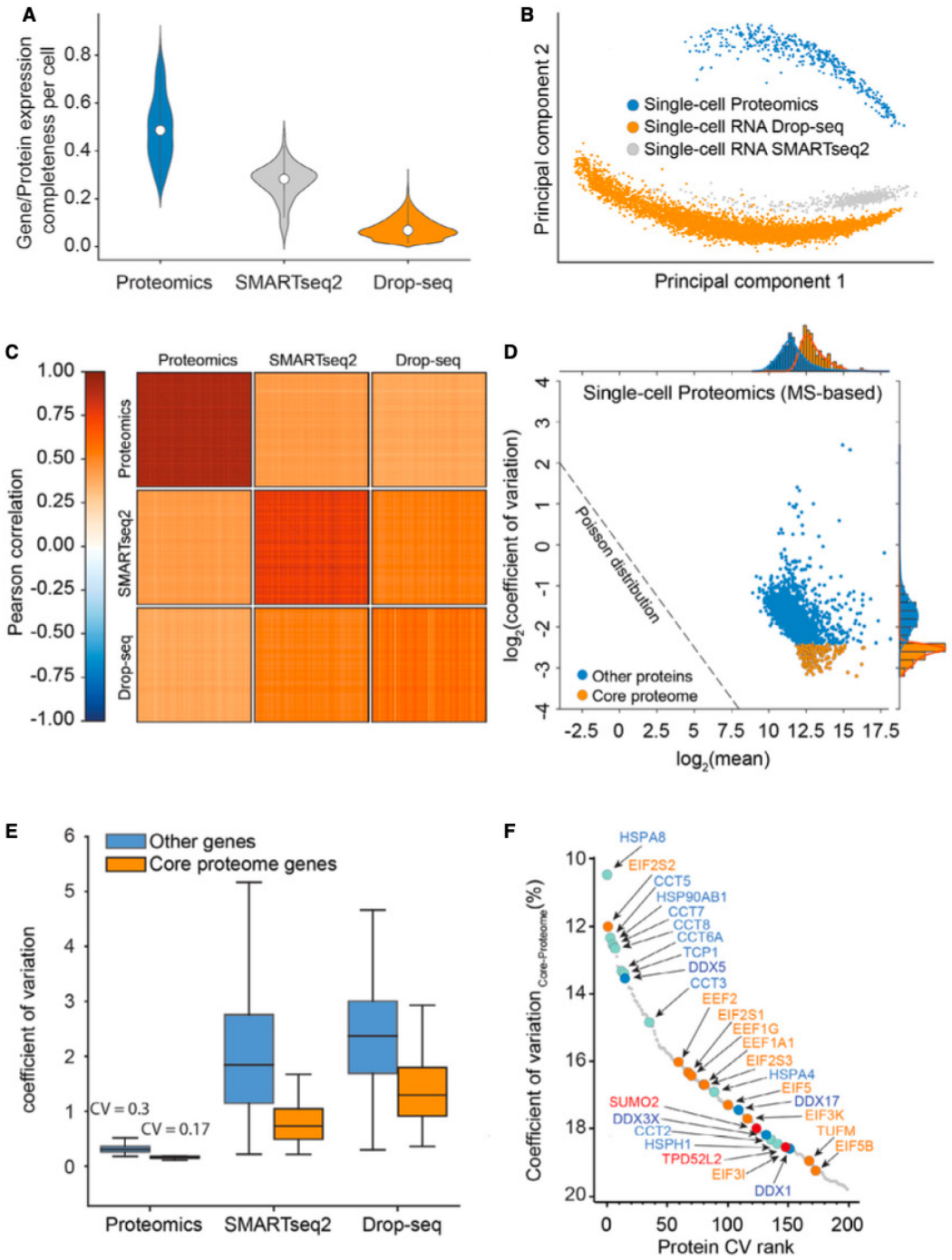



**Figure 5.**

© 2022 The Authors

**Figure 5. Single cells have a stable-core proteome but not transcriptome.**

A Gene or protein expression completeness per cell for T-SCP (Cells × Proteins: 424 × 2,480), SMARTseq2 (Cells × Genes: 720 × 24,990), or Drop-seq (Cells × Genes: 5,022 × 41,161) shown as violin plot; middle points represent the data set median.

B Principal component analysis of single-cell gene and protein expression measurements (1,672 shared genes).

C Heat map of cell–cell correlations across individual cells measured by proteomics and by both transcriptome technologies (1,672 shared genes).

D Coefficient of variation of single-cell protein expression levels in LC-MS based proteomics as a function of mean expression levels with the "core proteome" colored in orange.

E Boxplot of coefficient of variation of protein and transcript expression levels in LC-MS based proteomics, SMARTseq2, and Drop-seq technologies with a separate "core proteome" colored in orange (Box and Whiskers; The middle represents the median, the top and the bottom of the box represent the upper and lower quartile values of the data, and the whiskers represent the 1.5× IQR).

F Rank order abundance plot for the core proteome with color-coded protein classes (Red: SUMO2 and TDP52L2 proteins; Turquoise: Chaperonin and folding machinery-associated proteins. Orange: Translation initiation and elongation; Yellow: Structural proteins; Blue: DEAD box helicase family members).

the protein measurements (Fig EV4C–E). This suggests that—apart from increased sensitivity—our single-cell protein analysis could benefit from imputation or tailored likelihood-based parameter estimation methods (Risso *et al*, 2018; Lähnemann *et al*, 2020).

For bulk measurements, transcript levels generally correlate moderately with the corresponding protein levels, however, this correlation strongly depends on the biological situation (Buccitelli & Selbach, 2020). At the single-cell level, this effect is further convoluted by dissimilar measurement technologies and possibly by fundamental biological differences between the transcriptome and proteome. We asked to what degree scRNA-seq measurements could be used as a proxy for protein measurements in our data but found that protein measurements separate strongly from RNA in a principal component analysis (Fig 5B). While single-cell transcript expression levels correlate well between the scRNA-seq technologies, they diverge from single-cell protein measurements (Fig 5C).

As the transcriptome is measured as count data and the proteome as signal intensity levels, we further investigated the correlation between the transcriptome and proteome by correlating the coefficients of variation for shared genes between all data sets (Fig EV5A). This analysis reflects the quantitative variation of each gene at the single-cell transcriptome and single-cell proteome level across the cell cycle. Indeed, we found that the quantitative variation for the single-cell transcriptomes was high and correlation was consistent between both data sets (Fig EV5B).

In stark contrast, gene-level variation of both technologies did not correlate well with the single-cell proteome, highlighting again that both biological information levels are regulated fundamentally different at the single-cell level (Fig EV5C). This suggests that single-cell protein and RNA levels are very different, re-emphasizing that protein measurements yield complementary information to RNA measurements and do not simply re-iterate similar gene expression states. This implies distinct RNA and protein abundance regulation mechanisms on both modalities, dissection of which would not be possible with RNA measurements alone.

### T-SCP reveals a stable core proteome

Prompted by the divergent correlation values between the proteome and transcript levels, we next investigated the variability of gene expression as a function of abundance. For protein expression measurements, coefficients of variation were very small across covered abundances (Figs 5E and EV5A). This is consistent with a model in which the covered proteome is stable and probed deterministically across its full dynamic range. In contrast, the same analysis for UMI-controlled and not UMI-controlled scRNAseq data revealed a much higher overall transcriptome variability, as measured by the coefficient of variation of single-cell RNA-seq compared to protein measurements (Figs 5D and EV6A). Remarkably, this difference is already very apparent with the current sensitivity of MS-based proteomics, which will surely increase in the future. Comparing single-cell proteome measurements with six-cell proteomes (Fig 2C) suggests that a moderate increase in MS sensitivity would reveal a large part of the proteome to be quantitatively stably expressed.

We also observed that the single-cell transcriptome is dominated by shot noise, which has a Poisson distribution, because many of the transcripts are expressed at lower than one copy per cell on average (Fig EV6A). This means that a given single cell can have zero, one, or two transcripts of many of its expressed genes, leading to a Poisson distribution when summing up many single-cell measurements. For genes with higher expression values, there will always be transcripts present in each single cell and then the expression distribution of those genes is not shot noise dominated. For proteins, in contrast, the CVs depend only on the measurement sensitivity, as there are always sufficient copies in each cell to ensure that their expression levels are not biologically shot noise limited (Fig EV6A and 2A).

Based on these observations, we defined a "core-proteome" subset in the MS-based proteomics data by selecting the top 200 proteins with the lowest CVs of the proteins shared between at least 70% of the more than 430 single cells, including the drug perturbations (Dataset EV4). Interestingly, these proteins were distributed well across the covered dynamic range of the proteome (Fig 5D). Strikingly, we found the corresponding transcripts of the core proteome to be distributed across the full range of CVs in single-cell transcriptome data (Figs 5E and EV6B and C). The core proteome highlighted proteins frequently used for normalization such as HSP90, providing a positive control (Fig 5F). The CV rank plot of the core proteome also reveals a diverse set of proteins, including representatives of translation initiation and elongation, folding machineries, and nucleic acid helicases. Interestingly, we also identify TPD52L2 as one of the most stable proteins, which in turn is described as one of the most abundant proteins in HeLa cells (Hein *et al*, 2015) and SUMO2, which is known for its involvement in a plethora of essential regulatory cellular processes, suggesting a stable cellular SUMO2 pool even during stark proteome remodeling (Gareau & Lima, 2010).

 © 2022 The Authors

## Discussion

The T-SCP pipeline combines miniaturized sample preparation coupled to very low-flow liquid chromatography and a novel mass spectrometer resulting in at least one order of magnitude sensitivity gain at highest robustness for the analysis of single cells. We quantify cellular heterogeneity following targeted perturbation, which enables the direct analyses of drug responses in single-cell hierarchies on the proteome level. Furthermore, the comparison of single-cell RNA and proteome level revealed that the proteome is stable while the transcriptome is more stochastic, highlighting substantial regulation of translation and setting the stage for its elucidation at the single-cell level.

Although mainly demonstrated here for single-cell total proteome measurements, the sensitivity gain achieved in our workflow will be advantageous in any situation that is sample limited. This includes investigation of other compound classes such as metabolites or drugs, post-translational modifications from small numbers of cells or from *in vivo* material, and measurements directly from paraffin-embedded formalin-fixed (FFPE) pathology specimens, which we are already pursuing (preprint: Bhatia *et al*, 2021; preprint: Mund *et al*, 2021).

Our ion mobility-enhanced workflow is also compatible with chemical multiplexing with the advantage that the booster channel causing reporter ion distortions could be omitted or reduced (Ogata & Ishihama, 2020) and also benefit alternative multiplexing strategies like complementary TMT or EASI-Tag (Wühr *et al*, 2012; Winter *et al*, 2018). Furthermore, there are many opportunities for increasing overall sensitivity, including even brighter ion sources, improved chromatography, and better data analysis and modeling tools, similar to the rapid recent advances in the scRNAseq field.

## Material and Methods

**Reagents and Tools table**

| | | |
|---|---|---|
| **Chemicals, enzymes and other reagents** | | |
| Formic acid | Sigma Aldrich/Merck | Cat # 64-18-6 |
| Acetonitrile | Sigma Aldrich/Merck | Cat # 75-05-8 |
| Trifluoroacetic acid | Sigma Aldrich/Merck | Cat # 76-05-1 |
| Water, Optima™ LC/MS Grade | Fisher Chemical | Cat # W64 |
| Lysyl-Endopeptidase | Wako Chemicals | Cat # 129-02541 |
| Trypsin | Sigma Aldrich/Merck | Cat # T6576 |
| **Software** | | |
| MaxQuant (1.6.7.0) | https://maxquant.org/ | N/A |
| Perseus (1.6.7.0) | https://maxquant.org/perseus/ | N/A |
| Jupyter Notebook | https://jupyter.org/ | N/A |
| **Other** | | |
| 384-Well Plates | Eppendorf | Cat # 0030129547 |
| Adhesive PCR Sealing Foil Sheets | Thermo Scientific | Cat # AB-0626 |
| Empore SPE SDB-RPS disk | Sigma Aldrich/Merck | Cat # 66886-U |
| iST' sample preparation kit | PreOmics GmbH | Cat # P.O. 00001 |
| ThermoMixer® | Eppendorf | Cat # 460-0223 |
| NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer | Thermo Fisher | Cat # ND-ONEC-W |
| Concentrator plus | Eppendorf | Cat # F-45-48-11 |
| Mastercycler X50h | Eppendorf | Cat # 63160000 |
| EASY-nLC™ 1200 System | Thermo Fisher | Cat # LC140 |
| EvoSep One | EvoSep | Cat # EV-1000 |
| EvoTip | EvoSep | Cat # EV-2001 |
| 15 cm, 75 µm ID, 1.9 µm ID beads with 120A surface | EvoSep | Cat # EV-1112 |
| ZDV Emitter Sprayer 10 µm ID | Bruker Daltonik GmbH | Cat # 1865691 |
| timsTOF Pro | Bruker Daltonik GmbH | N/A |
| timsTOF SCP | Bruker Daltonik GmbH | N/A |
| Column oven | Sonation lab solutions | Cat # PRSO-V2 |

© 2022 The Authors

**Methods and Protocols**

***Sample preparation for bulk dilution experiments***

For all benchmark experiments, purified peptides from bulk HeLa cells were used. HeLa was cultured in Dulbecco's modified Eagle's medium at 10% fetal bovine serum, 20 mM glutamine, and 1% penicillin–streptomycin. Cells were collected by centrifugation, washed with phosphate-buffered saline (PBS), flash frozen in liquid nitrogen, and stored at −80°C. Cells were resuspended in PreOmics lysis buffer (PreOmics GmbH) and boiled for 20 min at 95°C, 1,500 rpm to denature, reduce, and alkylate cysteins, followed by sonication in a Branson, cooled down to room temperature, and diluted 1:1 with 100 mM Tris–HCl pH 8.5. Protein concentration was estimated by nanodrop measurement and 500 µg were further processed for overnight digestion by adding lysC and trypsin in a 1:50 ratio (µg of enzyme to µg of protein) at 37°C and 1,500 rpm. Peptides were acidified by adding 1% trifluoroacetic acid (TFA) and 99% isopropanol (IprOH) in a 1:1 ratio, vortexed, and subjected to StageTip (Rappsilber *et al*, 2007) clean-up via styrenedivinylbenzene reverse-phase sulfonate (SDB-RPS). Twenty microgram of peptides were loaded on two 14-gauge StageTip plugs. Peptides were washed two times with 200 µl 1% TFA and 99% IprOH followed by 200 µl 1% TFA and 99% IprOH in an in-house-made StageTip centrifuge at 2,000 g and elution with 100 µl of 1% Ammonia, 80% acetonitrile (ACN), 19% $ddH_2O$ into PCR tubes, and finally dried at 60°C in a SpeedVac centrifuge (Eppendorf, Concentrator plus). Peptides were resuspended in 0.1% TFA, 2% ACN, and 97.9% $ddH_2O$.

***Sample preparation for single-cell experiments (Protocol style)***

1 HeLa cells were cultured following a standard protocol as described above.
2 Supernatant was removed, cells were detached with trypsin treatment, followed by strong pipetting for cell aggregate dissociation.
3 Cells were washed three times with 500 µl ice-cold phosphate-buffered saline (PBS), pelleted by centrifugation, and the supernatant was removed.
4 For fluorescent-activated cell sorting (FACS), 5 µl DAPI was added to the 5 ml single-cell solution and sorting performed on the DAPI-negative live cell population.
5 Single cells were sorted into 384-well TwinTec Eppendorf plates containing 1 µl of 20% acetonitrile (ACN), 100 mM Tris–HCl pH 8.5, centrifuged briefly, sealed with aluminum foil and frozen at −80°C until further use (we cannot exclude that FACS sorting could lead to subtle changes in the proteome).
6 Single-cell containing 384-well plates were incubated for 30 min at 72°C in a PCR cycler, followed by 5 min sonication (Elmasonic P) at 37 kHz and room temperature.
7 Protein digestion was performed overnight at 37°C in a PCR cycler after adding 1 µl of 20% ACN, 100 mM Tris–HCl, pH 8.5, and 1 ng trypsin/lysC mix. (For the peptide bulk and cell count dilution experiments, peptides were resuspended in 4 µl of 2% ACN, 0.1% TFA, and 97.9% $ddH_2O$, and injected directly via NanoLC.)
8 Samples were dried in a SpeedVac at 30°C for 45 min
9 Single-cell-derived peptides were resuspended in 5 µl pure formic acid and incubated for 10 min on a thermo shaker at 25°C and 800 rpm.
10 EvoTips were activated following the standard EvoSep protocol (Sample loading protocol for Evotips). Then, 50 µl buffer A was added to each EvoTip followed by centrifugation at 200 g for 1 min (This leaves approximately 30 µl of buffer A on top of the SPE material.)
11 15 µl of buffer A (99.9% $ddH_2O$, 0.1% FA) were added to each single-cell well containing the dissolved single-cell peptides in 5 µl FA, followed by a 5-min shaking phase on a thermoshaker at 800 rpm, RT.
12 The single-cell peptides (20 µl total now) were transferred into the activated EvoTip, followed by centrifugation at 600 g for 1 min and two centrifugation steps after adding 50 µl buffer A. Last, 150 µl buffer A was added to each EvoTip and spun for 30 s at 300 g.

***Cell cycle experiments***

The drug-perturbed cell cycle arrest experiment was designed to enrich cells in four cell cycle stages—G1, the G1/S transition, G2, and the G2/M transition. HeLa cells were grown to approximately 30% confluence as described above, washed and treated for 24 h with 5 mM thymidine, released for 4.5 h, and treated again with 5 mM thymidine or 0.1 µg/ml nocodazole for 13 h. Cells of the G1/S phase (thymidine block) or G2/M phase (nocodazole block) were washed in PBS, trypsinated, subjected to strong pipetting to dissociate cell aggregates, and ice-cold PBS washes before DAPI-negative single live cell FACS sorting. A second set of G1/S phase and G2/M phase blocked cells was washed and cultured for 7 h or 2.5 h to enrich early G2 and G1 phase HeLa cells. These were washed with PBS, trypsinated, and subjected to DAPI-negative single live cell FACS sorting into 384-well plates pre-loaded with 1 µl 20% acetonitrile, 100 mM Tris–HCl, pH 8.5 lysis buffer. Furthermore, we prepared presumable unsynchronized cells sets from two independent cell cultures and subjected them to sample preparation as described below.

***High-pH reversed-phase fractionation***

To generate a deep library of HeLa precursors for all data-dependent benchmark experiments, peptides were fractionated at pH 10 with the spider fractionator (Kulak *et al*, 2017). Fifty micrograms of purified peptides were separated on a 30 cm $C_{18}$ column in 96 min and concatenated into 24 fractions with 2 min exit valve switches. Peptide fractions were dried in a SpeedVac and reconstituted in 2% ACN, 0.1% TFA, and 97.9% $ddH_2O$ for LC–MS analysis.

***Liquid chromatography***

For the initial benchmark experiments with HeLa bulk dilution and the cell count dilution, liquid chromatography analysis was performed with an EASY nanoLC 1200 (Thermo Fisher Scientific). Peptides were loaded on a 45 cm in-house packed HPLC column (75 µm inner diameter packed with 1.9 µm ReproSil-Pur C18-AQ silica beads, Dr. Maisch GmbH, Germany). Sample analytes were separated using a linear 60 min gradient from 5 to 30% B in 47.5 min followed by an increase to 60% for 2.5 min, by a 5 min wash at 95% buffer B at 300 nl/min, and re-equilibration for 5 min at 5% buffer B (buffer A: 0.1% formic acid (FA) and 99.9% $ddH_2O$; buffer B: 0.1% FA, 80% ACN, and 19.9% $ddH_2O$). The column temperature was kept at 60°C by an in-house manufactured oven.

 © 2022 The Authors

For all other proteome analyses, we used an EvoSep One liquid chromatography system (Bache *et al*, 2018) and analyzed the single-cell proteomes with a novel 35 min stepped pre-formed beta gradient eluting the peptides at 100 nl/min flow rate. We used a 15 cm × 75 µm ID column with 1.9 µm C18 beads (EvoSep) and a 10 µm ID zero dead volume electrospray emitter (Bruker Daltonik). Mobile phases A and B were 0.1% FA in water and 0.1% FA in ACN, respectively.

Both LC systems were coupled online to a modified trapped ion mobility spectrometry quadrupole time-of-flight mass spectrometer (timsTOF Pro, Bruker Daltonik GmbH, Germany) via a nanoelectrospray ion source (Captive spray, Bruker Daltonik GmbH).

***Construction of a novel mass spectrometer with higher sensitivity***

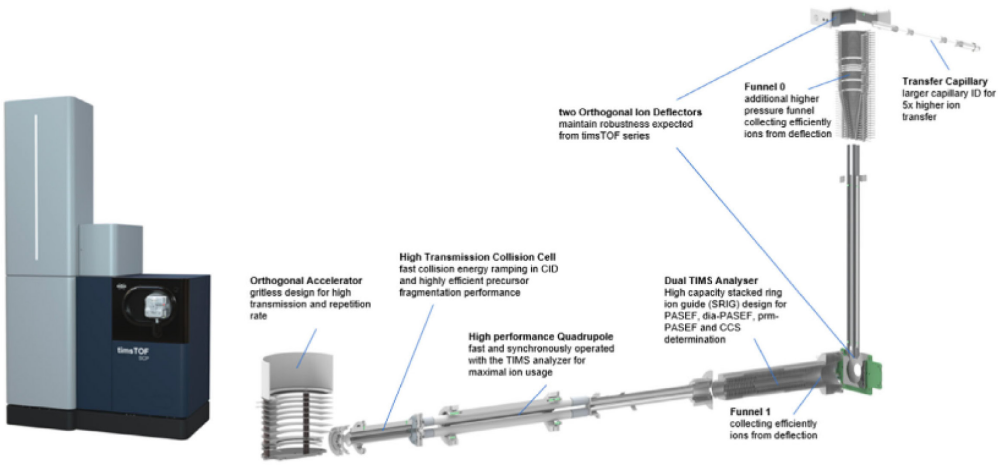
We updated our ion source to draw more ions into the vacuum system of the instrument. This is accomplished by modifying the glass capillary that conducts gas and ions between the ionization region at atmospheric pressure and the first pumping region. The added gas is eliminated by an additional pumping stage and associated prototype ion optics. These ion optics—a high pressure ion funnel and a RF multipole—confine the ions while the added gas is removed and moves them to the next vacuum region where TIMS analysis occurs. Importantly, the glass capillary is oriented orthogonal to the high-pressure funnel (as in prior designs) so that neutral contaminants and solvent droplets are directed by the gas flow away from the funnel. Furthermore, the high-pressure funnel and RF multipole are oriented orthogonal to the TIMS. This has the dual advantage of maintaining the gas dynamics of our original design, which is crucial for TIMS performance, and also that all remaining neutral contaminants are moved away from the TIMS entrance. This dual orthogonal design provides robustness against contamination in that neutrals, particles, and droplets are, in two places, driven past the ion optics, into the pumping ports. In studies of this prototype source, we estimate an improvement of a factor of 4.7 in ion transmission, and therefore overall increased signal intensity.

To accommodate the increased ion current, the TIMS analyzer has been updated to a new stacked ring (SRIG) design. This design uses a higher-order RF field in the ion accumulation region to create a larger effective ion storage volume than the low-order fields of previous designs. However, a low-order, quadrupolar field is maintained in the analyzer region to compress the ions toward the analyzer axis during elution to maintain high mobility resolution. In addition, the transition between the high-order and low-order portions of the device has been optimized relative to prior designs to further improve performance such as peak shape and resolution under practical conditions. This results in about a factor of 3 gain in ion (charge) capacity and therefore about a factor of 3 in the instrument's dynamic range.

Schematic representation of Bruker timsTOF SCP, which is a high-performance, ultra-high sensitivity benchtop mass spectrometer that combines a dual-TIMS analyzer coupled to a quadrupole, a collision cell that features fast collision energy ramping and a TOF mass analyzer.

***Mass spectrometry***

Mass spectrometric analysis was performed either in a data-dependent (dda) or data-independent (dia) PASEF mode. For ddaPASEF, 1 MS1 survey TIMS-MS and 10 PASEF MS/MS scans were acquired as per acquisition cycle. Ion accumulation and ramp time in the dual TIMS analyzer was set to 50/100/200 ms each and we analyzed the ion mobility range from $1/K_0 = 1.6$ Vs cm$^{-2}$ to 0.6 Vs cm$^{-2}$. Precursor ions for MS/MS analysis were isolated with a 2 Th window for $m/z < 700$ and 3 Th for $m/z > 700$ in a total m/z range 100–1,700 by synchronizing quadrupole switching events with the precursor elution profile from the TIMS device. The collision energy was lowered linearly as a function of increasing mobility starting from 59 eV at $1/K_0 = 1.6$ VS cm$^{-2}$ to 20 eV at $1/K_0 = 0.6$ Vs cm$^{-2}$. Singly charged precursor ions were excluded with a polygon filter (otof control, Bruker Daltonik GmbH). Precursors for MS/MS were picked at an intensity threshold of 1,500 arbitrary units (a.u.) and re-sequenced until reaching a "target value" of 20,000 a.u. considering a dynamic exclusion of 40 s elution. For DIA analysis, we made use of the correlation of ion mobility (IM) with m/z and synchronized the elution of precursors from each IM scan with the quadrupole isolation window. We used the described




© 2022 The Authors 

100 ms ddaPASEF method for the acquisition of a HeLa bulk single-shot library for the single-cell experiments and the short-gradient diaPASEF method as described in Meier *et al* (2020a), but performed up to five consecutive diaPASEF cycles before the next MS1 scan (see main text). The collision energy was ramped linearly as a function of the IM from 59 eV at 1/K0 = 1.6 Vs $cm^{-2}$ to 20 eV at 1/K0 = 0.6 Vs $cm^{-2}$.

***Raw data analysis***

ddaPASEF data for tryptic HeLa digest dilution series and the cell count experiment were analyzed in the MaxQuant environment (version 1.6.7) and searched against the human Uniprot databases (UP000005640_9606.fa, UP000005640_9606_additional.fa), which extracts features from four-dimensional isotope patterns and associated MS/MS spectra (Cox & Mann, 2008; Prianichnikov *et al*, 2020). False-discovery rates were controlled at 1% both on peptide spectral match (PSM) and protein levels. Peptides with a minimum length of seven amino acids were considered for the search, including N-terminal acetylation and methionine oxidation as variable modifications and cysteine carbamidomethylation as fixed modification, while limiting the maximum peptide mass to 4,600 Da. Enzyme specificity was set to trypsin cleaving C terminal to arginine and lysine. A maximum of two missed cleavages were allowed. Maximum precursor and fragment ion mass tolerance were searched as default for TIMS-DDA data. Peptide identifications by MS/MS were transferred by matching four-dimensional isotope patterns between the runs (MBR) with a 0.7-min retention time match window and a 0.05 $1/K_0$ ion mobility window in case of the single-cell count dilution experiment into a deep ddaPASEF library consisting of 24 fractionations of tryptic HeLa digest. These data were also searched without matching between runs to access the MBR-mediated identification increase. Either intensity-based absolute quantification (IBAQ) or label-free quantification was performed with the MaxLFQ algorithm and a minimum ratio count of 1 (Cox *et al*, 2014).

For all other single-cell experiments, we used a small library consisting of 25,376 peptides and 4,144 protein groups, which was acquired with the 100 ms ddaPASEF method described above and generated with the MSFRAGGER version 16 using default settings with the exception that cysteine carbamidomethylation was removed from fixed modification (Kong *et al*, 2017; preprint: Demichev *et al*, 2021). All single-cell measurements were searched against the human UniProt reference proteome (UP000005640_9606.fa and UP000005640_9606_additional.fa) of canonical and isoform sequences.

Due to recent software improvement driven by the implementation of the next generation of Spectronaut 15, followed by DIA-NN 1.8 (Bruderer *et al*, 2015; preprint: Demichev *et al*, 2021) for the analysis of diaPASEF raw files, which utilize the complex diaPASEF TIMS-TOF data much better by improved machine learning algorithms, we initially evaluated both software solutions for the analysis of our single-cell data set. It turned out that DIA-NN 1.8 using spectral libraries generated with MSFRAGGER, at that time, outperformed the library-based and directDIA analysis pipeline in Spectronaut 15 in our hands. This is consistent with the reports of DIA-NN being used for sample-limited analyses and very short gradients (Messner *et al*, 2020, 2021; preprint: Demichev *et al*, 2021; Fig EV7).

According to the DIA-NN authors, this benefit results from the advanced use of deep learning algorithms for signal deconvolution and for diaPASEF data specifically from the additional ion mobility resolution that makes full use of the correlation between ion mobility and m/z on MS2 level for fragment signal pattern identification and its matching to the precursor level. That said, our main focus in the current manuscript was not to present computational advances and a benchmarking of DIA-NN versus Spectronaut, but to highlight our scalable and robust MS-based single-cell proteomics workflow. Furthermore, because of its somewhat better performance and to present a unified analysis, we consistently analyzed all data with DIA-NN.

Raw files were analyzed with DIA-NN (version 1.8; Demichev *et al*, 2020) using default settings (e.g., 1% precursor and protein FDR), except changing the covered peptide length range to 7–50, precursor charge range to 2–4, enabling MBR, turning of protein inference to use the inference from the library generated by MSFRAGGER, quantification strategy to robust LC (high precision), and library generation to IDs, RT, and IM profiling.

***Visualization and FDR estimates of fragment ion intensities***

Quantitative fragment ion profiles were generated from the DIA-NN output table. Only fragment ions used for quantification in DIA-NN were included. To cancel out cell size-dependent abundance changes, one normalization factor was estimated per cell, using fold-change-based normalization of the whole data set, as described in the MS-EmpiRe method, which we also used for FDR control (Ammar *et al*, 2019). The intensities were $\log_2$ transformed and subsequently visualized.

***Proteomics downstream data analysis***

Proteomics data analysis was performed in the Perseus environment (version 1.6.7, 1.5.5) (Tyanova *et al*, 2016), GraphpadPrism (version 8.2.1), and Python (version 3.8.2). MaxQuant output tables were filtered for "Reverse," "Only identified by site modification," and "Potential contaminants" before further processing. Ontologies for the biological process and cellular compartment assignment for proteins were performed with the mainAnnot.homo_sapiens.txt.gz, followed by categorical counting across all proteins for each of the ontologies, and counts were exemplary visualized as frequency plot. For single-cell analysis, if not otherwise specified, the DIA-NN protein group data output was filtered first for at least 600 protein observations per cell and at least 15% quantification events across rows and log (x + 1) transformed resulting in the following cell numbers and protein quantifications (Fig EV8).

For correlation analysis of two protein expression vectors, transformed gene or protein quantification events of two cells were plotted against each other replacing missing values by zeros. For principal component analysis (PCA), missing values were imputed from a normal distribution with a width of 0.3 standard deviations that was downshifted by 1.8 standard deviations. Differential expression analysis by two-sided unpaired *t*-test was performed on two groups filtered for at least 50% row-wise quantification events within one group. False discovery rate control due to multiple hypothesis testing was performed by a permutation-based model and SAM statistic with an $S_0$ parameter of 0.3. For cell size estimation based on raw MS signal, intensity outputs within cell cycle resolved single-cell proteomics results were summed up and visualized as boxplots. The core proteome was calculated by filtering the

 © 2022 The Authors

whole single-cell proteomics data set for at least 70% quantification events for each protein followed by selection of the top 200 proteins with the smallest coefficient of variation across the data set.

***Single-cell protein and RNA comparison and dropout statistics***

The SMART-Seq2 (Hu *et al*, 2019) data set measured 720 HeLa cells in three different batches with a total of 24,990 expressed genes. The Drop-seq (Schwabe *et al*, 2020) data set contained three batches with a total of 5,665 cells and 41,161 expressed genes. We performed the single-cell analysis with scanpy v1.6.0 (Wolf *et al*, 2018). If not stated otherwise, we used standardized filtering across all data sets, removed cells with less than 600 genes expressed, and removed genes detected in < 15% of the remaining cells, resulting in 10,557 transcripts in 720 cells in the SMART-Seq2 data set and 5,022 transcripts and 6,701 cells measured with Drop-seq technology. Ratios of non-zero entries in the scRNAseq data sets and the number of identified proteins in our data are summarized as violin plots. To investigate data completeness across covered dynamic range, we computed the data completeness as a function of the mean log(x + 1)-transformed protein abundance of all non-zero/-NaN entries. We included the expected data completeness based on the assumption that missing values are purely due to shot-(Poisson)-noise as 1-exp(-x). For correlation analysis, the RNA abundance entries were linearly scaled to sum to the mean cell size of the respective data set per cell (231,281.56 for SMART-Seq2 and 7,808.12 for Drop-Seq) followed by log(x + 1) transformation of all abundance entries. Correlation values between the expressions of two cells were computed as the Pearson correlation on the 1,672 genes that were shared in all three data sets. Entries of missing protein abundance values were excluded from the specific computation. For the PCA plot of technological comparisons, the gene coverage intersection of all technologies (1,672) was isolated, NaNs were replaced by zeros, and expression values were linearly scaled to 1E6 followed by log (1 + x) transformation. In coefficient of variation (CV) versus CV plots comparing different technologies as well as the mean versus CV analysis (including the core proteome analysis) and the CV distribution boxplots, RNA expression vectors were scaled to the mean cell size of that measurement technology and mean and CV values were computed per gene under the assumption that single-cell RNA-sequencing data are not zero inflated (Svensson, 2020) while NaNs were excluded for the proteomics data. CV (Proteomics) versus CV (RNA-seq) plots show the comparison of CV values of proteins/genes that were shared between all data sets.

***Cell cycle state prediction***

Cell cycle predictions were performed using the scanpy method score_genes (Wolf *et al*, 2018) based on three sets of proteins that are specifically expressed in the G1 (MARCKS, KRT1, HIST1H1E, KRT18, HNRNPA1, CHCHD3, CD44, NASP, TARDBP, PODXL, SUMO2, STMN1, TRIM28, and SPTAN1), S (NOLC1, ATP2A2, CANX, TMX1, CKB, SLC25A3, SLC16A1, MT-CO2, SRPRB, CYB5R3, LETM1, and ANP32B), or G2/M phase (TOP2A, HMGB1, EIF5B, TMSB10, EIF3D, ANP32A, RCC2, FASN, LUC7L2, AARS, KPNA2, and CKAP5), respectively. The cell phase-specific protein sets were selected based on the z-scored fold-change ratios provided in Geiger and coworkers (Aviner *et al*, 2015). The top 60 highest differentially expressed genes were selected and filtered for quantification in at least 70% of our cells. This scoring method yields the average expression on the provided set of genes minus the average expression on a reference set of genes for each cell. The reference set is chosen to mirror the average expression of the target gene set. For this analysis, cells were filtered, log(x + 1) transformed, and missing values replaced by zeros. Plotted are the ROC curves for the three scores corresponding to the three sets of characteristic proteins (G1, S, and G2 M) used individually to discriminate between the cells of two cell cycle stages.

## Data availability

All mass spectrometry raw data, libraries, and outputs from each particular search engine analyzed in this study have been deposited to the ProteomeXchange Consortium via the PRIDEpartner repository. Project accession: PXD024043 (www.ebi.ac.uk/pride/archive?keyword = PXD024043).

The code used for data analysis can be found as a Jupyter Notebook at: https://github.com/theislab/singlecell_proteomics.

**Expanded View** for this article is available online.

### Acknowledgements

S. R. is supported by the Helmholtz Association under the joint research school "Munich School for Data Science"—MUDS. D.S.F. acknowledges support from a German Research Foundation (DFG) fellowship through the Graduate School of Quantitative Biosciences Munich (QBM). F.C. acknowledges the European Union's Horizon 2020 research and innovation program (Marie Skłodowska-Curie individual fellowship under grant agreement 846795). We thank our colleagues in the Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, at the Center for Protein Research at Copenhagen University, and at Bruker Daltonik for discussions and help. In particular, we thank I. Paron, A. Piras, and C. Deiml for technical support and J.B. Müller for column production. We also thank Mark Ridgeway, Craig Whitehouse, Andreas Brekenfeld, Niels Goedecke, and Christoph Gebhardt of Bruker for their contribution to the development, installation, and troubleshooting of the modified mass spectrometer. We are grateful for the FACS sorting support by the Imaging Core Facility at the Max Planck Institute of Biochemistry, in particular Martin Spitaler and Markus Oster. Open Access funding enabled and organized by ProjektDEAL.

## Author contributions

**Andreas-David Brunner:** Conceptualization; Data curation; Formal analysis; Validation; Investigation; Visualization; Methodology; Writing—original draft; Project administration; Writing—review and editing. **Marvin Thielert:** Data curation; Formal analysis; Investigation; Methodology. **Catherine Vasilopoulou:** Data curation; Formal analysis; Investigation; Methodology. **Constantin Ammar:** Data curation; Software; Formal analysis; Investigation; Methodology. **Fabian Coscia:** Resources; Investigation; Methodology. **Andreas Mund:** Resources; Investigation; Methodology. **Ole B Hoerning:** Resources; Software; Methodology. **Nicolai Bache:** Resources; Software; Investigation; Methodology. **Amalia Apalategui:** Resources; Software; Methodology. **Markus Lubeck:** Resources; Software; Investigation; Methodology. **Sabrina Richter:** Data curation; Software; Formal analysis; Investigation; Visualization; Methodology. **David S Fischer:** Conceptualization; Software; Formal analysis; Validation; Investigation; Visualization; Methodology. **Oliver Raether:**

© 2022 The Authors

Conceptualization; Resources; Software; Formal analysis; Supervision; Investigation; Methodology; Project administration. **Melvin A Park:** Resources; Methodology. **Florian Meier:** Conceptualization; Resources; Supervision; Investigation; Methodology. **Fabian J Theis:** Conceptualization; Resources; Software; Formal analysis; Supervision; Funding acquisition; Investigation; Methodology. **Matthias Mann:** Conceptualization; Resources; Supervision; Funding acquisition; Investigation; Methodology; Writing—original draft; Project administration; Writing—review and editing.

In addition to the CRediT author contributions listed above, the contributions in detail are:

AD-B and MM conceptualized and designed the study. AD-B, MT, FC, and AM performed experiments. MAP and OR designed the new mass spectrometer. AD-B, OBH, and NB conceived the new EvoSep gradient. OBH, NB, AD-B, and MT designed the new EvoSep gradient and optimized it for proteomics performance. DSF and FJT conceptualized the single-cell modeling. AD-B, SR, DSF, FJT, CA, MT, OBH, NB, CV, AA, ML, FM, and MM analyzed the data. AD-B and MM wrote the manuscript.

### Disclosure statement and competing interests

ML, OR, AA, and MAP are employees of Bruker Daltonik. OBH and NB are employees of EvoSep Biosystems. MM is an indirect shareholder in EvoSep Biosystems. FJT reports receiving consulting fees from Roche Diagnostics GmbH and Cellarity Inc., and ownership interest in Cellarity Inc. and Dermagnostix. All other authors have no competing interests.

## References


Aebersold R, Mann M (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* 537: 347–355

Ammar C, Gruber M, Csaba G, Zimmer R (2019) MS-EmpiRe utilizes peptide-level noise distributions for ultra-sensitive detection of differentially expressed proteins. *Mol Cell Proteomics* 18: 1880–1892

Aviner R, Shenoy A, Elroy-Stein O, Geiger T (2015) Uncovering hidden layers of cell cycle regulation through integrative multi-omic analysis. *PLoS Genet* 11: 1005554

Bache N, Geyer PE, Bekker-Jensen DB, Hoerning O, Falkenby L, Treit PV, Doll S, Paron I, Müller JB, Meier F *et al* (2018) A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Mol Cell Proteomics* 17: 2284–2296

Bhatia HS, Brunner A, Rong Z, Mai H, Todorov I, Ali M, Molbay M, Kolabas ZI (2021) DISCO-MS: proteomics of spatially identified tissues in whole organs. *BioRxiv* https://doi.org/10.1101/2021.11.02.466753 1–23 [PREPRINT]

Brenes A, Hukelmann J, Bensaddek D, Lamond AI (2019) Multibatch TMT reveals false positives, batch effects and missing values. *Mol Cell Proteomics* 18: 1967–1980

Bruderer R, Bernhardt OM, Gandhi T, Miladinović SM, Cheng L-Y, Messner S, Ehrenberger T, Zanotelli V, Butscheid Y, Escher C *et al* (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics* 14: 1400–1410

Buccitelli C, Selbach M (2020) mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet* 21: 630–644

Budnik B, Levy E, Harmange G, Slavov N (2018) SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* 19: 161

Cheung TK, Lee CY, Bayer FP, McCoy A, Kuster B, Rose CM (2021) Defining the carrier proteome limit for single-cell proteomics. *Nat Methods* 18: 76–83

Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 13: 2513–2526

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367–1372

Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* 17: 41–44

Demichev V, Yu F, Teo GC, Szyrwiel L, Rosenberger GA, Decker J, Kaspar-Schoenefeld S, Lilley KS, Mülleder M, Nesvizhskii AI *et al* (2021) High sensitivity dia-PASEF proteomics with DIA-NN and FragPipe. *bioRxiv* https://doi.org/10.1101/2021.03.08.434385 [PREPRINT]

Emmett MR, Caprioli RM (1994) Micro-electrospray mass spectrometry: ultra-high-sensitivity analysis of peptides and proteins. *J Am Soc Mass Spectrom* 5: 605–613

Gareau JR, Lima CD (2010) The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol* 11: 861–871

Greguš M, Kostas JC, Ray S, Abbatiello SE, Ivanov AR (2020) Improved sensitivity of ultralow flow LC-MS-based proteomic profiling of limited samples using monolithic capillary columns and FAIMS technology. *Anal Chem* 92: 14702–14712

Hein M, Hubner N, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak I, Weisswange I, Mansfeld J, Buchholz F *et al* (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* 163: 712–723

Hu W, Zhang X, Guo Q, Yang J, Yang Y, Wei S, Su X (2019) HeLa-CCL2 cell heterogeneity studied by single-cell DNA and RNA sequencing. *PLoS One* 14: e0225466

Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lönnerberg P, Linnarsson S (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* 11: 163–166

Jackson HW, Fischer JR, Zanotelli VRT, Ali HR, Mechera R, Soysal SD, Moch H, Muenst S, Varga Z, Weber WP *et al* (2020) The single-cell pathology landscape of breast cancer. *Nature* 578: 615–620

Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A *et al* (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343: 776–779

Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods* 14: 513–520

Kulak NA, Geyer PE, Mann M (2017) Loss-less nano-fractionator for high sensitivity, high coverage proteomics. *Mol Cell Proteomics* 16: 694–705

Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, Vallejos CA, Campbell KR, Beerenwinkel N, Mahfouz A *et al* (2020) Eleven grand challenges in single-cell data science. *Genome Biol* 21: 53

Larance M, Lamond AI (2015) Multidimensional proteomics for cell biology. *Nat Rev Mol Cell Biol* 16: 269–280

Li ZY, Huang M, Wang XK, Zhu Y, Li JS, Wong CCL, Fang Q (2018) Nanoliter-scale oil-air-droplet chip-based single cell proteomic analysis. *Anal Chem* 90: 5430–5438

Liang Y, Acor H, McCown MA, Nwosu AJ, Boekweg H, Axtell NB, Truong T, Cong Y, Payne SH, Kelly RT (2020) Fully automated sample processing and analysis workflow for low-input proteome profiling. *Anal Chem* 93: 1658–1666.


 © 2022 The Authors

Liu Y, Mi Y, Mueller T, Kreibich S, Williams EG, Van Drogen A, Borel C, Frank M, Germain P-L, Bludau I *et al* (2019) Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat Biotechnol* 37: 314 – 322

Lundberg E, Borner GHH (2019) Spatial proteomics: a powerful discovery tool for cell biology. *Nat Rev Mol Cell Biol* 20: 285 – 302

Ly T, Whigham A, Clarke R, Brenes-Murillo AJ, Estes B, Madhessian D, Lundberg E, Wadsworth P, Lamond AI (2017) Proteomic analysis of cell cycle progression in asynchronous cultures, including mitotic subphases, using PRIMMUS. *Elife* 6: e27574

Macosko E, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas A, Kamitaki N, Martersteck E *et al* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161: 1202 – 1214

Meier F, Beck S, Grassl N, Lubeck M, Park MA, Raether O, Mann M (2015) Parallel accumulation-serial fragmentation (PASEF): Multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J Proteome Res* 14: 5378 – 5387

Meier F, Brunner A-D, Koch S, Koch H, Lubeck M, Krause M, Goedecke N, Decker J, Kosinski T, Park MA *et al* (2018) Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol Cell Proteomics* 17: 2534 – 2545

Meier F, Brunner A-D, Frank M, Ha A, Bludau I, Voytik E, Kaspar-Schoenefeld S, Lubeck M, Raether O, Bache N *et al* (2020a) diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nat Methods* 17: 1229 – 1236

Meier F, Köhler N, Brunner A-D, Wanka J-M, Voytik E, Strauss M, Theis F, Mann M (2020b) Deep learning the collisional cross sections of the peptide universe from a million training samples. *Nat Commun* 12: 1185

Messner CB, Demichev V, Wendisch D, Michalick L, White M, Freiwald A, Textoris-Taube K, Vernardis SI, Egger A-S, Kreidl M *et al* (2020) Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. *Cell Syst* 11: 11 – 24

Messner CB, Demichev V, Bloomfield N, Yu JSL, White M, Kreidl M, Egger A-S, Freiwald A, Ivosev G, Wasim F *et al* (2021) Ultra-fast proteomics with Scanning SWATH. *Nat Biotechnol* 39: 846 – 854

Mund A, Coscia F, Hollandi R, Kovács F, Kriston A, Brunner A-D, Bzorek M, Naimy S, Mette Rahbek Gjerdrum L, Dyring-Andersen B *et al* (2021) AI-driven Deep Visual Proteomics defines cell identity and heterogeneity Proteomics Program. *bioRxiv* https://doi.org/10.1101/2021.01.25.427969 [PREPRINT]

Ogata K, Ishihama Y (2020) Extending the separation space with trapped ion mobility spectrometry improves the accuracy of isobaric tag-based quantitation in proteomic LC/MS/MS. *Anal Chem* 92: 8037 – 8040

Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9: 171 – 181

Prianichnikov N, Koch H, Koch S, Lubeck M, Heilig R, Brehmer S, Fischer R, Cox J (2020) Maxquant software for ion mobility enhanced shotgun proteomics. *Mol Cell Proteomics* 19: 1058 – 1069

Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC *et al* (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* 30: 777 – 782

Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2: 1896 – 1906

Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M *et al* (2017) The human cell atlas. *Elife* 6: e27041

Risso D, Perraudeau F, Gribkova S, Dudoit S (2018) A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat Commun* 9: 284

Schaum N, Karkanias J, Neff NF, May AP, Quake SR, Wyss-Coray T, Darmanis S, Batson J, Botvinnik O, Chen MB *et al* (2018) Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* 562: 367 – 372

Schnitzbauer J, Strauss MT, Schlichthaerle T, Schueder F, Jungmann R (2017) Super-resolution microscopy with DNA-PAINT. *Nat Protoc* 12: 1198 – 1228

Schoof EM, Furtwängler B, Üresin N, Rapin N, Savickas S, Gentil C, Lechman E, Keller UAD, Dick JE, Porse BT (2021) Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nat Commun* 12: 1 – 15

Schwabe D, Formichetti S, Junker JP, Falcke M, Rajewsky N (2020) The transcriptome dynamics of single cells during the cell cycle. *Mol Syst Biol* 16: e9946

Smith ZD, Nachman I, Regev A, Meissner A (2010) Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nat Biotechnol* 28: 521 – 526

Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, Smibert P (2017) Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 14: 865 – 868

Svensson V (2020) Droplet scRNA-seq is not zero-inflated. *Nat Biotechnol* 38: 147 – 150

Tsai C-F, Zhao R, Williams SM, Moore RJ, Schultz K, Chrisler WB, Pasa-Tolic L, Rodland KD, Smith RD, Shi T *et al* (2020) An Improved Boosting to Amplify Signal with Isobaric Labeling (iBASIL) Strategy for Precise Quantitative Single-cell Proteomics. *Mol Cell Proteomics* 19: 828 – 838

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13: 731 – 740

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A *et al* (2015) Tissue-based map of the human proteome. *Science* 347: 1260419

Volpe P, Eremenko-Volpe T (1970) Quantitative studies on cell proteins in suspension cultures. *Eur J Biochem* 12: 195 – 200

Williams SM, Liyu AV, Tsai C-F, Moore RJ, Orton DJ, Chrisler WB, Gaffrey MJ, Liu T, Smith RD, Kelly RT *et al* (2020) Automated coupling of nanodroplet sample preparation with liquid chromatography-mass spectrometry for high-throughput single-cell proteomics. *Anal Chem* 92: 10588 – 10596

Wilm M, Mann M (1996) Analytical properties of the nanoelectrospray ion source. *Anal Chem* 68: 1 – 8

Winter SV, Meier F, Wichmann C, Cox J, Mann M, Meissner F (2018) EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. *Nat Methods* 15: 527 – 530

Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 19: 15

Wühr M, Haas W, McAlister GC, Peshkin L, Rad R, Kirschner MW, Gygi SP (2012) Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster. *Anal Chem* 84: 9214 – 9221

Zhu Y, Piehowski PD, Zhao R, Chen J, Shen Y, Moore RJ, Shukla AK, Petyuk VA, Campbell-Thompson M, Mathews CE *et al* (2018) Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat Commun* 9: 1 – 10

**License:** This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Article 4. Robust dimethyl-based multiplex-DIA workflow doubles single-cell proteome depth via a reference channel

*Molecular Systems Biology*, 2023 (19, e11503); DOI: 10.15252/msb.202211503

Marvin Thielert[1,†], Ericka CM Itang[1,†], Constantin Ammar[1], Florian A Rosenberger[1], Isabell Bludau[1], Lisa Schweizer[1], Thierry M Nordmann[1], Patricia Skowronek[1], Maria Wahle[1], Wen-Feng Zeng[1], Xie-Xuan Zhou[1], Andreas-David Brunner[1,2], Sabrina Richter[3,4], Mitchell P Levesque[5], Fabian J Theis[3,4], Martin Steger[1,6,‡], Matthias Mann[1,6,‡]

*[1]Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [2]Boehringer Ingelheim Pharma GmbH &Co. KG, Drug Discovery Sciences, Biberach an der Riss, Germany. [3]Helmholtz Zentrum Muenchen–German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany. [4]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany. [5]Department of Dermatology, University of Zurich, University of Zurich Hospital, Zurich, Switzerland. [6]New address: NEOsphere Biotechnologies GmbH, Planegg, Germany. [7]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.*

*[†]These authors contributed equally to this work.*

*[‡]These authors contributed equally to this work as senior authors.*

In an effort to further enhance sc-proteomics, I implemented dimethyl labeling, a reliable and cost-efficient method for multiplexing, and combined it with DIA. Since non-isobaric channels are isolated from each other, we sought to introduce a ‘reference channel’, a higher input spike-in proteome in one of the channels. We termed the technology multiplex-DIA (mDIA).

In benchmarking experiments, we consistently found a near complete labeling efficiency. Multispecies experiment showed a high quantification accuracy, even at high ratios between the channels. Identification rate and quantification reproducibility was comparable to label-free analysis. We demonstrated that mDIA increased the throughput to five-plex, when using LysN as the protease instead of the commonly used trypsin and LysC. Because of the increased complexity, the five-plex analysis decreased protein identifications by 15% compared to one channel. However, mDIA on five-plex level will improve with search algorithms efficiently handling multiplex DIA data.

To optimize our single-cell workflow, we adapted the lysis buffer to facilitate the amine-directed chemistry of dimethyl labeling. Furthermore, we automated sample preparation to increase reproducibility and throughput. The ‘reference channel’ doubled the proteomic depth in single cells. Moreover, we implemented an algorithm called Reference Quantification (RefQuant), which utilizes the reference channel for quantification on the basis of ratios to the target channels. Testing this RefQuant

approach on varied datasets showed good reproducibility and quantitative accuracy for both MS1 and MS2 quantification, successfully retrieving four-fold differences in target channels of single-cell equivalents. Next, we analyzed 476 cells and identified a median 2,377 protein groups, and almost 4,000 protein groups in a few cells. Additionally, we confirmed that the concept of a stable proteome remains true at higher proteomic depth.

Finally, we demonstrated the modularity of mDIA by connecting it seamlessly with the DVP pipeline to increase spatial resolution and enhance the throughput at the MS and laser-capture microdissection levels. We profiled *in-situ* cancer cells in primary cutaneous melanoma. Analyzing epidermal and dermal melanoma, we found distinct tumor microenvironment patterns such as remodeling of the extracellular matrix.

In summary, mDIA is a powerful technology that can utilize a reference channel for higher identification and accurate quantification in sc-proteomics with a throughput of 80 single cells per day. We enhanced the proteomic depth and confirmed the concept of a stable proteome at this higher depth. In combination with DVP, mDIA enhanced spatial resolution and throughput and we showed the potential for precision oncology on primary cutaneous melanoma. Together, mDIA provides avenues for enhanced single-cell and clinical proteomics.

***Contribution:***

Shared first-authorship. The project was initiated by Matthias Mann, Martin Steger and myself. I supervised the benchmarking experiment on bulk digests. I designed and developed the automated sample preparation workflow and performed all single-cell experiments. I analyzed the single-cell data, wrote the manuscript together with Matthias Mann and designed the figures. Additionally, I designed the cover for that issue of the journal.

Method

SOURCE DATA | TRANSPARENT PROCESS | OPEN ACCESS

molecular systems biology

# Robust dimethyl-based multiplex-DIA doubles single-cell proteome depth via a reference channel

Marvin Thielert[1,†], Ericka CM Itang[1,†], Constantin Ammar[1], Florian A Rosenberger[1], Isabell Bludau[1], Lisa Schweizer[1], Thierry M Nordmann[1], Patricia Skowronek[1], Maria Wahle[1], Wen-Feng Zeng[1], Xie-Xuan Zhou[1], Andreas-David Brunner[1,2], Sabrina Richter[3,4], Mitchell P Levesque[5], Fabian J Theis[3,4], Martin Steger[1,6,‡] & Matthias Mann[1,7,*,‡]

## Abstract

**Single-cell proteomics aims to characterize biological function and heterogeneity at the level of proteins in an unbiased manner. It is currently limited in proteomic depth, throughput, and robustness, which we address here by a streamlined multiplexed workflow using data-independent acquisition (mDIA). We demonstrate automated and complete dimethyl labeling of bulk or single-cell samples, without losing proteomic depth. Lys-N digestion enables fivefold quantification at MS1 and MS2 level. Because the multiplexed channels are quantitatively isolated from each other, mDIA accommodates a reference channel that does not interfere with the target channels. Our algorithm RefQuant takes advantage of this and confidently quantifies twice as many proteins per single cell compared to our previous work (Brunner *et al*, PMID 35226415), while our workflow currently allows routine analysis of 80 single cells per day. Finally, we combined mDIA with spatial proteomics to increase the throughput of Deep Visual Proteomics seven-fold for microdissection and four-fold for MS analysis. Applying this to primary cutaneous melanoma, we discovered proteomic signatures of cells within distinct tumor microenvironments, showcasing its potential for precision oncology.**

**Keywords** DIA; dimethyl labeling; multiplexing; single cells; spatial proteomics

**Subject Categories** Methods & Resources; Proteomics

**DOI** 10.15252/msb.202211503 | Received 9 December 2022 | Revised 17 July 2023 | Accepted 25 July 2023 | Published online 21 August 2023

**Mol Syst Biol. (2023) 19: e11503**

## Introduction

Characterizing biology directly at the single-cell level is greatly advancing our knowledge of different cell types and cellular heterogeneity. Single-cell RNA sequencing (scRNAseq) is now routine and large atlases of human cell populations of different organs are being generated (Eraslan *et al*, 2022; Suo *et al*, 2022; Tabula Sapiens Consortium *et al*, 2022). To clearly define single cell types or sub-types, such measurements often encompass tens or hundreds of thousands of single-cell transcriptomes.

Single-cell mass spectrometry (MS)-based proteomics (scProteomics) is also generating much interest, because the dynamic proteome is thought to be a very informative reflection of the biological function of cells. Furthermore, scProteomics could complement other omics modalities in a multi-level description of cellular systems. However, this requires overcoming key technological challenges in four areas: lossless sample preparation, high-performance chromatography, high sensitivity MS measurements and optimal analysis of the low-level signals for quantification. A pioneering approach for reducing sample losses is called nanoPOTS (nanodroplet processing in one pot for trace samples) coupled sophisticated protein extraction to dedicated, narrow column chromatography (Zhu *et al*, 2018). SCoPE-MS employs the tandem mass tag (TMT) isobaric labeling strategy to differentially mass-label the single cells, to which a 'carrier channel' – originally consisting of hundreds of single-cell equivalents – is added (Budnik *et al*, 2018). Quantification is less straightforward in SCoPE-MS because their carrier channel also contributes to the reporter ion readout in TMT, and this has been shown to lead to ratio distortions (Cheung *et al*, 2020; Ye

1 Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany
2 Boehringer Ingelheim Pharma GmbH & Co. KG, Drug Discovery Sciences, Biberach an der Riss, Germany
3 Helmholtz Zentrum München – German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany
4 TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany
5 Department of Dermatology, University of Zurich, University of Zurich Hospital, Zurich, Switzerland
6 New address: NEOsphere Biotechnologies GmbH, Planegg, Germany
7 Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
*Corresponding author. Tel: +49 89 8578-2557; E-mail: mmann@biochem.mpg.de
†These authors contributed equally to this work
‡These authors contributed equally to this work as senior authors

© 2023 The Authors. Published under the terms of the CC BY 4.0 license

et al, 2022; Ctortecka et al, 2022b). Recently, label-free approaches have also shown promising results (Saha-Shah et al, 2019; Zhu et al, 2019; Cong et al, 2021; Matzinger et al, 2023).

Our group recently described an ultra-high sensitivity workflow for single-cell applications that builds entirely upon readily available components (Brunner et al, 2022). Samples are deposited into low adsorption 384-well plates, allowing parallel preparation with minimal losses, as also used by others (Liang et al, 2020; Schoof et al, 2021; Specht et al, 2021). The resulting peptides are deposited on standard Evotips, from which they are eluted in 'nano-packages' of 20 nl into a preformed gradient. This stored gradient is then separated at a very low flow 'Whisper' gradient (100 nl/min) on an analytical column attached to the Evosep system (Bache et al, 2018) and electrosprayed into a trapped ion mobility time of flight mass spectrometer (timsTOF) (Meier et al, 2018). We performed data acquisition by dia-PASEF (Meier et al, 2020) and achieved median protein identifications of about 1,000 in interphase HeLa cells and up to 2,000 in drug arrested mitotic cells.

Based on these previous experiences with scProteomics, we aimed to develop an improved workflow building on recent developments in multiplex-DIA (mDIA). Although DIA is overwhelmingly performed in a label-free and single-run manner, researchers have long investigated if some of the advantages of multiplexing could be transferred to the DIA setting as well. The principal challenge in mDIA is that multiplexing further multiplies the complexity of already highly complex DIA spectra. To our knowledge, the MEDUSA method first reported this concept (Griffiths et al, 2014). Here, ubiquitin and SUMO peptides were differentially labeled with mTRAQ reagents (Δ0, Δ4, and Δ8). Garcia and colleagues combined DIA with SILAC labeling achieving similar identification depth with an order of magnitude better quantitative accuracy than SILAC alone (Pino et al, 2021). Very recently, a powerful triplex mTRAQ workflow has been described, which approached similar proteome depth as label-free DIA data (Derks et al, 2022) and exceeded its quantitative precision within runs. Applied to single cells, their 'plexDIA' approach reached a depth of 1,000 proteins per cell in an active gradient of 30 min. Apart from these non-isobaric multiplexing methods, low mass reporter based isobaric methods or precursor coupled reporter tags have also been described (Tian et al, 2020; Ctortecka et al, 2022a).

Here we present an mDIA workflow that uses dimethyl labeling for multiplexing. This derivatization has been well established in proteomics from the original report (Hsu et al, 2003) and extensive subsequent work by the Heck group (Boersema et al, 2008, 2009; Taouatas et al, 2010). We introduce the principles of dimethyl mDIA, characterize its performance at the bulk proteomics level on modern MS instrumentation and extend to five-plex at the MS1 and MS2 level by using Lys-N enzyme. Next, we develop an automated workflow using robotic derivatization combined with Evotips for single-cell proteomics. We then evaluate the concept of a reference channel in mDIA – in which one of the channels is used as a spike-in proteome, making measurements universally comparable to each other. In the context of single-cell proteomics it doubles identifications and throughput. Further, we devise Reference Quantification (RefQuant) an algorithmic strategy to make optimal use of the reference channel for quantification and apply it to single-cell proteomics. Finally, we combine mDIA with our recently developed spatial technology termed Deep Visual Proteomics (DVP) (Mund et al, 2022) to illustrate the potential of mDIA in precision oncology using melanoma as an example.

## Results

### Exploration of dimethyl labeling-based multiplexed data-independent acquisition (mDIA)

Previous approaches to multiplexed data-independent acquisition (mDIA) mass spectrometry have used SILAC or amino-reactive labels, specifically the non-isobaric analog to the iTRAQ reagent, termed mTRAQ (Griffiths et al, 2014; Derks et al, 2022). To extend the repertoire of available tools for multiplexed analysis of complex proteomes by DIA, we investigated whether a complementary chemical labeling approach based on peptide dimethylation was also suited for this. We decided to explore dimethylation as the derivatization of peptides with dimethyl labels, which is quick, reliable, cost-efficient, and can easily be automated (Raaijmakers et al, 2008). Primary amines occurring on peptide N-termini or epsilon-amino groups of lysine residues are thereby derivatized through formaldehyde, to form an intermediate Schiff base that is subsequently reduced by sodium cyanoborohydride, to form a dimethylamino group (Hsu et al, 2003). Depending on the combination of stable isotope labeling reagents, this adds a 28.0313 Da, 32.0564 Da, or 36.0757 Da mass tag to each amino group in a given tryptic peptide (referred to as light (Δ0), intermediate (Δ4), and heavy labels (Δ8)), enabling three-plexed mDIA of proteomes (Fig 1A).

Following a previously established dimethyl labeling protocol (Boersema et al, 2009), we first evaluated the feasibility of combining it with mDIA. As a first step, we labeled tryptic peptides derived from bovine serum albumin (BSA) with dimethyl-Δ0/Δ4/Δ8, which resulted in a labeling efficiency of more than 99% (Fig 1B). To determine quantification accuracy in this setup, we prepared three samples combining the three channels in different ratios (Δ0/Δ4/Δ8 = 17:2:1; 7:2:1 and 5:3:2) and acquired the data in DIA mode (See Materials and Methods), followed by raw data processing with DIA-NN, a neural network-based software that recently has been benchmarked for multiplexed DIA-MS (Demichev et al, 2019, 2022; Derks et al, 2022). This revealed a high degree of quantification accuracy, even for the highest tested ratio (17:1) (Fig 1C) and as expected deviations from the true ratios were mainly observed for low-abundant peptides (Appendix Fig S1A and B). We therefore concluded that dimethyl labeling might also be suited for multiplexed acquisitions of complex proteomes in DIA mode.

### mDIA applied to in-depth quantification of complex proteomes

To determine the suitability of dimethyl labeling for mDIA of whole cell protein extracts, we derivatized tryptic peptides from HeLa cells with three dimethyl mass tags (Δ0, Δ4, and Δ8). Like the labeling of just one protein (Fig 1B), this resulted in an almost complete labeling (> 99%) of all detected peptides for all three channels (Appendix Fig S2A).

Next, we assessed whether identification rates and quantification precision might be compromised by dimethyl labels compared to label-free analysis. First, we measured triplicates of unlabeled and dimethyl labeled peptides (light channel only, Δ0) and processed the raw data with DIA-NN using spectral libraries predicted by AlphaPeptDeep (Zeng et al, 2022). Those predicted libraries account for potential differences in peptide fragmentation, collisional cross

 © 2023 The Authors

sections, and retention times between unlabeled and dimethyl labeled peptides.

We used a 75 min LC gradient and data acquisition either by regular DIA on an Orbitrap platform, or using a timsTOF HT instrument, with our recently reported optimal dia-PASEF acquisition schemes (Appendix Fig S3A, Materials and Methods) (Skowronek *et al*, 2022b). Whereas the timsTOF platform generally yielded higher identification numbers and greater quantification precision, the number of quantified precursors and protein groups between unlabeled and one-channel dimethyl labeled samples was similar on both instruments. This demonstrates that derivatization of peptides with dimethyl groups does not negatively impact peptide

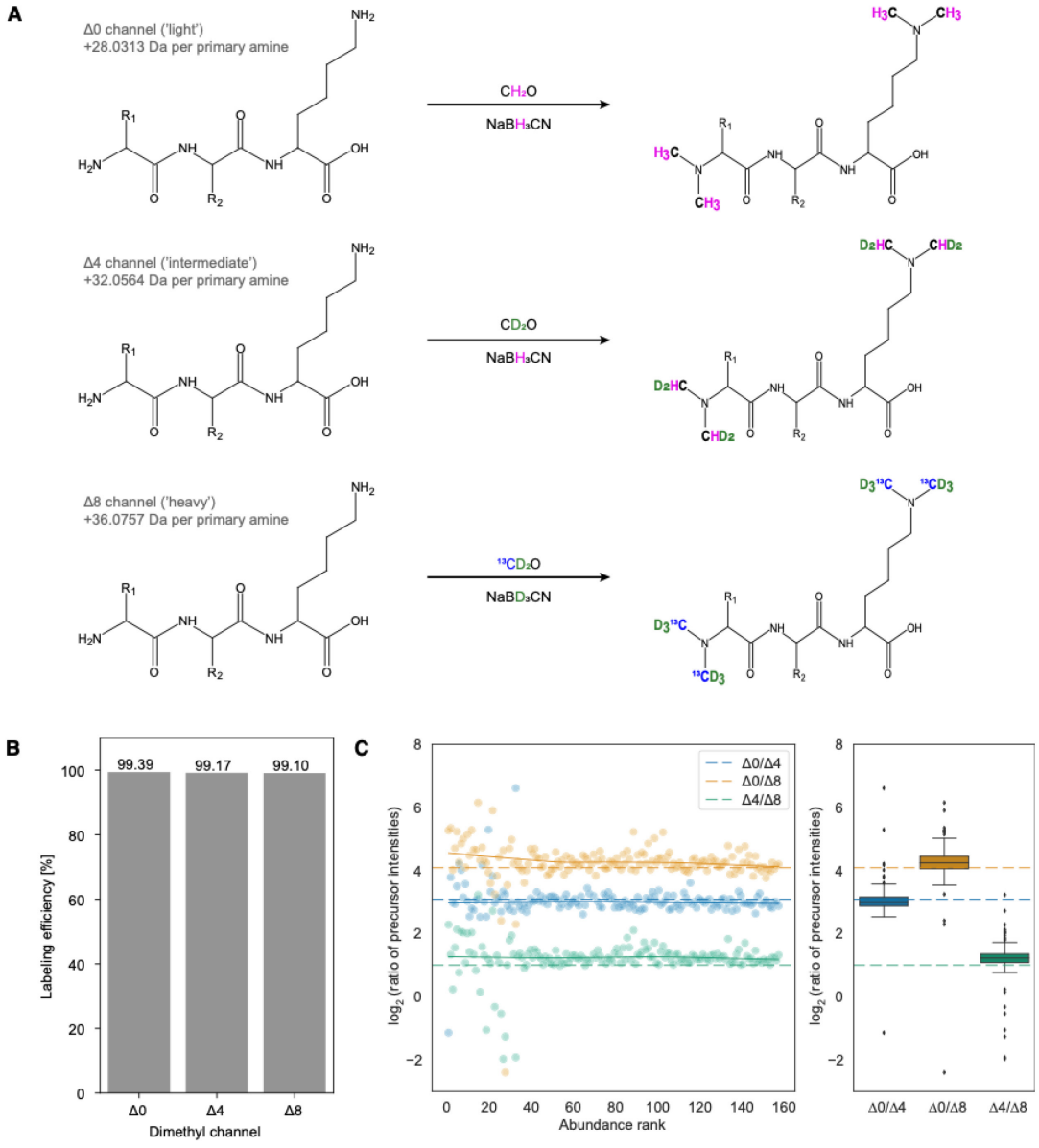



**Figure 1.**

© 2023 The Authors

**Figure 1. Dimethyl labeling of bovine serum albumin (BSA) combined with multiplexed data-independent acquisition (mDIA).**

A Stable isotope dimethyl labeling scheme for a three-plex mDIA setup. Six hydrogens can be replaced by deuterium and the two carbons by $^{13}C$ per dimethyl, of which there are one in tryptic peptides ending in arginine (N-terminus) and two in those ending in lysine (N-terminus and N-ε-Lys). Depending on the combination of stable isotope labeling reagents, mass tags of 28.0313 Da, 32.0564 Da, or 36.0757 Da are added to each primary amine group of a peptide. A tryptic peptide harboring a C-terminal lysine residue is depicted.

B Dimethyl labeling efficiency of peptides derived from intensity ratios of labeled peptides relative to all detected peptides in DDA mode ($n$ = 1).

C Quantification accuracy of dimethyl labeled peptides in DIA mode. Differentially labeled tryptic BSA peptides were mixed in a 17:2:1 ratio of Δ0/Δ4/Δ8 and the data was acquired in DIA mode. Scatterplots (left) illustrate the $\log_2$ intensity ratios as a function of the peptide abundance rank, which is summarized as a boxplot (right). The box depicts the interquartile range with the central band representing the median value of the dataset. The whiskers represent the furthest datapoint that is within 1.5 times the interquartile range (IQR). In both panels, the expected ratios are marked by colored dashed lines. 50 fmol BSA was injected per replicate (technical replicates, $n$ = 3).

identification rates (Fig 2A–C). Using the Orbitrap instrument, we quantified about 7,000 protein groups (85,000 precursors) from 125 ng of injected peptides in both unlabeled and dimethyl-Δ0-labeled samples, with a median CV of 4.8%. In contrast, the timsTOF yielded about 20% more protein groups and 50% more precursors in unlabeled samples (about 8,400 protein groups and 128,000 precursors) and 11% more protein groups and 40% more precursors (7,700 protein groups and 118,000 precursors) in labeled samples with the same injection amount (Fig 2A and B).

When increasing proteome complexity by adding a second channel (Δ4), identifications remained almost constant at both the precursor and the protein level on the timsTOF platform (about 115,000 precursors and 8,200 protein groups), whereas we observed a slight decrease on the Orbitrap instrument (about 20% for precursors and 4% for proteins groups). In a three-plex mDIA setup, protein group identifications further decreased on the Orbitrap and slightly decreased with the timsTOF (Fig 2B), indicating that the deconvolution of multiplexed spectra is challenging and that the ion mobility dimension of the timsTOF platform helps to resolve this complexity. Quantitative reproducibility measured by coefficient of variation (CV) decreased only slightly from single to triple labeling (median 4.1% at a single channel and 7.8% at three channels) (Fig 2C). We empirically determined the false discovery rate (FDR) in a two-species experiment (Fig EV1A). For bulk and equally abundant channels, the FDR was 1.5% at a 'Channel.Q.Value' cutoff of 0.01 (see Fig EV1A for comparison of 'Channel.Q.Value' and 'Translated.Q.Value'). These results indicate absent or minimal cross-talk between the dimethyl channels.

To investigate quantification accuracy in a three-plex setup we combined tryptic peptides from human (HeLa), *S. cerevisiae*, and *E. coli* at defined ratios for three labels (Fig 2D). This mixing scheme creates a benchmark of known protein ratios, which can be compared with the ratios of measured intensity values. In addition to our initially designed dia-PASEF method, we generated an alternative method, consisting of multiple MS1 scans in between dia-PASEF scans of each duty cycle (Appendix Fig S3B) similar to what was performed previously on Orbitrap instruments (Xuan *et al*, 2020; Derks *et al*, 2022). We used this MS1-centric method for direct comparison with our standard MS2-centric method, by quantifying protein ratios across channels in the three-plex, mixed species experiment (See Materials and Methods). This revealed that with both methods the measured protein group ratios largely agreed with the expected ones and that the MS2-centric method slightly outperformed the MS1-centric method in terms of accuracy, with similar identification rates and quantification precision (Figs 2E and EV1B and C).

### Extending the scope of multiplexing by dimethyl labels

Although trypsin is the most frequently used protease for MS-based proteomics applications, in the context of dimethyl labeling it limits multiplexing to only three channels if a mass difference between channels of 4 Da is desired at MS1 and MS2 level. At the MS1 level alone, Lys-C in combination with different isotopes of formaldehyde and cyanoborohydride enables five-plex dimethyl labeling (Wu *et al*, 2014).

Lys-C derived peptides harbor two primary amino groups for labeling – one of them is located on the N-terminus and the second one on the side chain of the C-terminal lysine residue. Digesting proteins with Lys-N instead generates a N-terminal lysine on each peptide, and thus two primary N-terminal amines for labeling (Raijmakers *et al*, 2010). Five-plex labeling of Lys-C or Lys-N-derived HeLa peptides thus results in isotopologues that are separated by 4 Da from each other, which would be sufficient for accurate quantification (Fig 3A).

Lys-N-derived and dimethylated peptides form more b-ions during MS2 fragmentation than tryptic peptides (Taouatas *et al*, 2008) and the two dimethyl labels are carried by b-ions. We reasoned that this should enable MS2-based quantification in five-plex experiments. In contrast, fragmented five-plex labeled peptides derived from Lys-C carry one dimethyl label on each b- and y-ion. Each of them separates from its isotopologues by only 2 Da on MS2 level, which leads to major overlaps of isotope clusters.

We therefore explored whether Lys-N-mediated digestion combined with five-plex dimethyl labeling could be a straightforward way of further increasing multiplexing for mDIA. We used HeLa lysates and digested them with Lys-N, followed by derivatization with five-plex dimethyl mass tags, achieving more than 97% labeling efficiency (Appendix Fig S2B). After combining the labeled peptides in equal ratios, we quantified them using the timsTOF instrument, with dia-PASEF as scan mode and employing methods optimized for Lys-N-digestion with py_diAID (Appendix Fig S3C) (Skowronek *et al*, 2022b). This led to the quantification of about 7,000 protein groups in both unlabeled and labeled samples with just one-channel (Δ0) (Fig 3B and C). In contrast to dimethyl labeled tryptic peptides, combining up to three labels did not significantly decrease protein group identifications and even in the five-plex setup, protein identifications decreased by only about 15% compared to one-channel dimethyl measurements. Similar to the three-plex mDIA result described above, CV values did not change until 3-plex, while they increased slightly in 5-plex (Fig 3D). The median fold changes of a mixing experiment largely agreed with their expected values in 5-plex (Fig EV1D). This demonstrates the feasibility of multiplexing up to five samples in mDIA experiments using dimethyl labeling.

 © 2023 The Authors

**Robust and automated ultra-high sensitivity mDIA workflow with a reference channel**

Having established the dimethyl mDIA derivatization for bulk proteomics samples, we set out to combine it with our recently published ultra-high sensitivity single-cell proteomics workflow (Brunner *et al*, 2022). We used a 384-well plate format for all ultra-high sensitivity experiments to reduce reaction volumes. Furthermore, we adapted the lysis and digestion buffer to an amine-free buffer using triethylammonium bicarbonate (TEAB) to enable the amine directed chemistry of dimethyl labeling (Fig 4A). The dimethylation reaction itself is performed by simple addition of the chemicals to the peptide

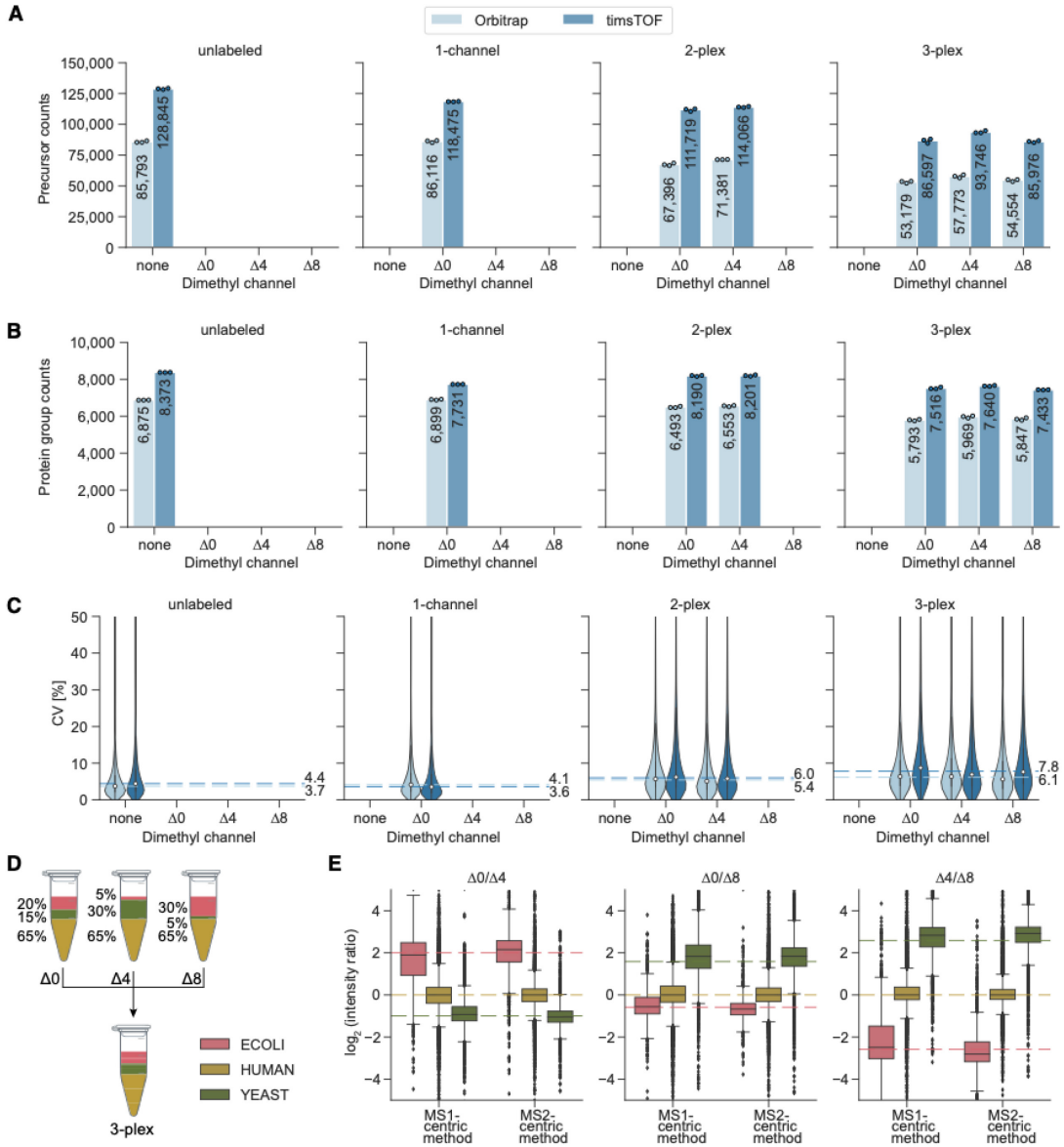



**Figure 2.**

© 2023 The Authors

**Figure 2. Identification rates, quantification precision and accuracy of dimethyl labeled peptides.**

A, B Number of quantified HeLa peptide precursors (A) and protein groups (B) for unlabeled, one-channel (Δ0), two-plex (Δ0 and Δ4), and three-plex (Δ0, Δ4, and Δ8) labeled samples. 125 ng of peptides were injected three technical replicates ($n$ = 3) per channel.

C Coefficients of variation (CV, %) of all protein groups identified per condition for Orbitrap and timsTOF instruments. Protein group intensities were calculated using MaxLFQ-based protein quantification from 'Precursor.Normalised' quantities. Median CVs for three technical replicates ($n$ = 3) are shown as dashed lines.

D Mixing scheme for tryptic peptides of HeLa, *S. cerevisiae*, and *E. coli* at different ratios prior to dimethyl labeling. The three channels were multiplexed in a 1:1:1 ratio and 450 ng total amount (150 ng per channel) was measured in three technical replicates ($n$ = 3) on the timsTOF platform using MS1- and MS2-centric methods (Appendix Fig S3A and B).

E Side-by-side comparison of quantification accuracies between MS1-centric and MS2-centric acquisition methods in a mixed species experiment (technical replicates, $n$ = 3). Protein group ratios are plotted as boxplots with expected ratios (dashed lines). The box depicts the interquartile range with the central band representing the median value of the dataset. The whiskers represent the furthest datapoint that is within 1.5 times the interquartile range (IQR).

mixtures in small volumes achieving a labeling efficiency greater than 99.5% for all channels at 1 ng tryptic HeLa peptides (Fig 4B). Subsequent cleanup of the derivatization reaction has previously been done in an in-column format (Boersema *et al*, 2009), but such a step is generally omitted in single-cell applications to avoid sample loss. In our format, however, the cleanup and combination of the separately labeled single cells can easily be performed in the Evotips. Each of the three in-solution labeled digests in a 384-well plate are sequentially deposited onto the same Evotip, thus, the derivatization reaction does not complicate the workflow and is completely transparent to the user. The entire sample preparation, including the combination of the labeled channel while loading of the Evotips, was automated on a standard Bravo robot (Agilent) to streamline the workflow, enhance reproducibility and automate for increasing throughput (Fig 4A, Materials and Methods). This also adds traceability and enables compliance and quality control.

In the Evosep instrument, peptides are eluted from the tips in nano-packages and analyzed at a very low 'Whisper' flow (100 nl/min) with a gradient of 31 min and overhead time to the next injection of 7 min ('40 samples per day method'). With such a low flow rate, it is important to avoid any post-column dead volumes. We therefore sought to eliminate the post-column connectors that we had employed previously by using pulled columns packed into the electrospray tip (IonOpticks) (preprint: Sandow *et al*, 2021). This setup makes the workflow very robust and reproducible, and also improved chromatographic resolution, with peaks eluting in full width half maximum (FWHM) with a median of about 2.7 s, corresponding to an elution volume of only 8 nl at peak width.

Finally, we analyzed the multiplexed single-cell samples using an optimal dia-PASEF method (Skowronek *et al*, 2022b) with eight dia-PASEF scans and a mass range of 300 to 1,200 and ion mobility of 0.7 to 1.3 on the timsTOF SCP (Appendix Fig S3D, See Materials and Methods).

In the bulk experiments above, we had established that the different non-isobaric channels in the mDIA workflow are decoupled from each other in terms of quantification. More importantly, different precursors that are fragmented together still do not contribute to the same 'reporter ions' as they may do in TMT labeling. Thus, the only interference in terms of quantification has to come from chemical noise of different precursors that happen to share a fragment within the MS2 resolution of the mass spectrometer. We therefore reasoned that we could use one of the mDIA channels as a 'reference channel' comprised of the same input material across all samples. Apart from providing a common standard for identification and quantification, such a reference channel conceptually decouples identification from quantification. This is because the reference channel proteome is easily identified due to its higher signal intensity and its uniformity across samples. Subsequently, the software needs to transfer the quantification boundaries from the reference channel to the target channels containing the single-cell proteomes.

To investigate the reference channel concept in mDIA, we first systematically increased its loading in the Δ0 channel (The specific datasets used for evaluation are summarized in Fig EV2). With our current chromatographic set-up, the number of identifications leveled off at roughly 10 ng, which we therefore chose as the reference channel amount for all subsequent experiments (Fig EV3A).

Next, we kept the reference amount constant, but varied the amount of peptide in the target channels to explore how much the reference channels would support identifications of weak signals. The DIA-NN software that we used for analysis does not directly have a notion of a reference channel, but considers the boundaries of the channel with the highest scoring precursor as internal reference for transfer. To assess identification confidence, DIA-NN reports a 'Translated.Q.Value' and a 'Channel.Q.Value' parameter (Fig EV3B). As noted above for bulk proteomics experiments, the former is unsuitable for FDR determination. Therefore, we experimentally determined the value of the 'Channel.Q.Value' at which features from the reference channel could safely be transferred into the target channels. To this end, we measured mDIA samples in which we left one or both of the target channels empty (scDecoy). This revealed that a 'Channel.Q.Value' of 0.45 led to a count-based FDR of 1% for precursors and 0.17 for protein groups (Fig EV3C and D, Materials and Methods). At the latter value, we also obtained accurate quantitative ratios between the target and reference channel as judged by defined mixing experiments (Fig EV3G). To ensure high confidence identifications in the target channels with single-cell equivalents and good quantification, we henceforth filtered the data with a 'Channel.Q.Value' of 0.15 as well as the recommended parameters of DIA-NN. Note that this same empirically determined cutoff was also supported by experiments on mouse liver tissue (Fig EV3E).

With these parameters, we investigated whether increasing amounts in the reference channel might influence the number of proteins found in single-cell equivalents in the target channels (scReference dataset, amounts from 0.25 ng to 10 ng in the reference channel and single-cell equivalents in the target channels). Importantly, while increasing the reference channel input amount beyond 5 ng, the target channel identifications remained the same (Fig 4C). We conclude that these channels are isolated from each other as expected from the mDIA concept. Additionally, to show that this identification increase is due to the reference channel and

 © 2023 The Authors

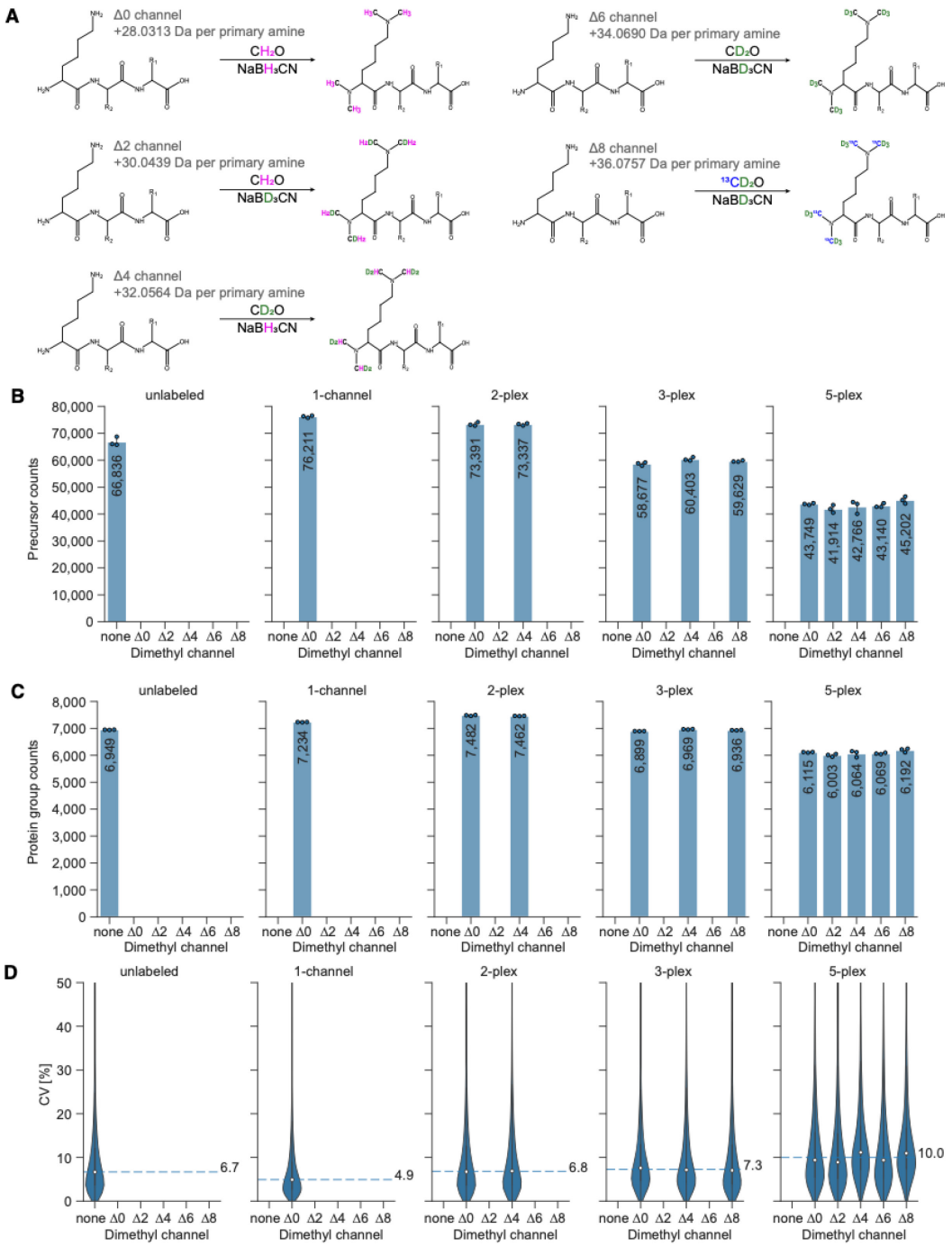



**Figure 3.**

© 2023 The Authors

**Figure 3. Five-plex dimethyl labeling using Lys-N protease and mDIA.**

A Stable isotope dimethyl labeling scheme for a five-plex mDIA experimental setup with Lys-N digested peptides. Depending on the combination of stable isotope labeling reagents, mass tags of 28.0313 Da, 30.0439 Da, 32.0564 Da, 34.0690 Da, or 36.0757 Da are added to each primary amine group on a peptide. Since Lys-N hydrolyses the N-terminal side of lysine residues, two labels are clustered on the N-termini of peptides (N-terminus and N-ε-Lys).

B, C Number of quantified HeLa peptide precursors (B) and protein groups (C) for unlabeled (none), one-channel (Δ0), two-plex (Δ0 and Δ4), three-plex (Δ0, Δ4, and Δ8), and five-plex (Δ0, Δ2, Δ4, Δ6, and Δ8) labeled samples. 100 ng of peptides were injected per channel (technical replicates, $n = 3$).

D Coefficients of variation (CV, %) of all protein groups identified per condition in the dimethyl five-plex setup, using Lys-N as protease. Protein group intensities are calculated using MaxLFQ-based protein quantification from 'Precursor.Normalised' quantities. Median CVs for three technical replicates ($n = 3$) are shown as dashed lines.

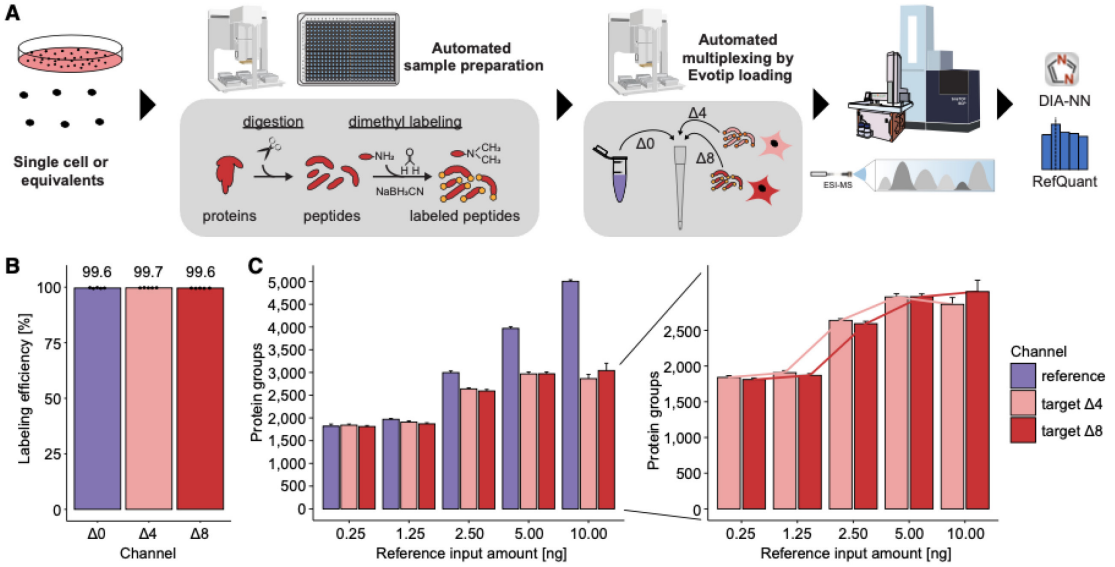



**Figure 4. Streamlined and automated ultra-high sensitivity mDIA workflow.**

A Single HeLa cells or single-cell equivalents are processed in a standard 384-well plate on a Bravo robot (Agilent) by lysis in TEAB and ACN, tryptic digestion, dimethyl labeling, and multiplexing by loading the different mDIA channels onto the same Evotip. LC–MS analysis is done by an Evosep One with low-flow chromatography coupled to a timsTOF SCP instrument. The data is analyzed by DIA-NN, followed by our algorithm RefQuant (see main text).

B 1 ng tryptic peptides from HeLa cells were labeled with dimethyl mass tags Δ0, Δ4, and Δ8 and acquired individually in DDA mode to determine labeling efficiency. Efficiency is calculated based on intensity ratios of labeled peptides relative to all detected peptides. The labeling efficiencies were consistently higher than 99.5% for all channels in quintuplicate measurements (technical replicates, $n = 5$).

C Effect of varying the protein input in the reference channel for protein identification across all channels. Increasing the total protein amount in the reference channel linearly increases protein identifications in the reference channel (left), but importantly protein identifications reach a plateau in the target channels (Δ4 and Δ8) with single-cell equivalents upon 5–10 ng in the reference channel (scReference dataset). Connected lines between increasing reference input amounts show a sigmoidal relation (right). Error bars represent the standard deviation of quintuplicate measurements (technical replicates, $n = 5$).

not simply because of adsorptive losses during sample analysis, we compared the identifications in the target channels with reference channel and without, while spiking 10 ng of unlabeled HeLa instead. Identifications from single-cell equivalents adding an unrelated 10 ng proteome were 1,247 proteins, while using a 10 ng reference channel were 2,018 protein groups (mean of five replicates in each target channel). Thus, the increase in identifications is overwhelmingly due to the reference channel, rather than simply an effect of adsorptive losses (Fig EV3F).

### RefQuant improves quantification by sampling ratios relative to the reference channel

Having shown the advantages of a reference channel to substantially increase protein identifications in the target channels (Fig 4C) and the 'Channel.Q.Value' cutoff of 0.15 to filter out false identifications (Fig EV3B, right panel), we investigated if its benefits also extend to quantification. We reasoned that the standardized reference proteome present at higher amounts might directly improve

 © 2023 The Authors

quantification in the target channels by effectively reducing technical variation across runs – an idea already implemented in the super-SILAC strategy (Geiger *et al*, 2010). To enable this analysis, we implemented an approach that we term **Ref**erence **Quant**ification (RefQuant), which is based on sampling all available ratios relative to the reference channel (Fig 5, Materials and Methods). In short, the ratios of individual fragment ions as well as MS1 peaks between reference and target channels are extracted, which results in a distribution of ratios (Figs 5A and EV4A). Subsequently, a best ratio R is estimated from the ratio distribution in a robust manner by taking the mean from the 40% lower quantile of ratios. We determined this threshold empirically using the scBenchmark dataset (Figs EV2 and EV4B). The intensity of the target channel is then calculated by multiplying the ratio R with a scaling factor representing the intensity of the reference channel. This gives a re-scaled precursor intensity value. The scaling factor is derived per precursors as the median reference intensity over all runs, in order to further stabilize the estimate. An interesting property of RefQuant is that it utilizes the high relative ratio to the reference channel in two ways: (i) the prior knowledge about the high expected ratio is used as an effective 'noise filter' (Fig EV4A–D) and (ii) the higher abundance of the reference channel stabilizes quantification, which should conceptually reduce overall noise (Appendix Text S1).

To test RefQuant, we initially applied it to a single-cell equivalent benchmark dataset (scBenchmark, target channels with 250 pg HeLa peptides, reference channel with 10 ng) and one in which we varied the amount of HeLa peptides in each of the target channels covering four-fold expression differences in both directions from a single-cell equivalent (scQuant, target channels from 62.5 to 2,000 pg HeLa peptides, reference channel with 10 ng; Fig EV2).

In the scBenchmark dataset, the ratio between the reference and target channel is well reconstructed by RefQuant, as compared to the standard method of only extracting the target ion intensities from DIA-NN without using the reference channel (Fig 5B). This indicates good reproducibility and also good quantitative accuracy. This was the case both for MS1 ('Ms1.Area') and MS2 ('Precursor.Translated' and 'Precursor.Normalised') quantification. In particular, quantification without RefQuant had more scatter and RefQuant best reconstructed the true ratios. DIA-NN reports a difference between quantification of tryptic peptides with C-terminal lysine ('lysine precursors') and arginine ('arginine precursors'). RefQuant also accurately quantifies arginine precursors, which are particularly challenging to quantify, presumably because they have shared y-ions between mDIA channels. Next, we specifically investigated the effects of varying the peptide amount in the reference channel with or without RefQuant. We analyzed the ratios to reference over a range from single-cell equivalents (0.25 ng) to 10 ng in the reference channel (Appendix Fig S4A–E). Without RefQuant, we observed a clear dependence of the distribution spread (measured via the standard deviation) on the amount of reference channel at the MS1 level. This spread was substantially reduced on the MS2 level (assessed via 'Precursor.Translated' and the 'Precursor.Normalised' quantities from DIA-NN). However, when applying RefQuant to the same data virtually no dependence of the standard deviation on reference channel amount was observed. Additionally, only RefQuant eliminated systematic ratio compression as long as the reference channel was dominant. This demonstrates that RefQuant specifically utilizes the information of the reference channel and thereby removes potential artifacts.

We next turned to assess the accuracy of different fold changes in single-cell equivalents, the most important outcome of single-cell proteomics. The scQuant dataset shows that RefQuant on mDIA accurately retrieves four-fold differences in the target channel with either decreasing or increasing amount in the target channel. There were no differences in the two target channels (Δ4 and Δ8) after median correction (Figs 5C and EV4E). The quantification of lysine peptides, which have no shared fragments between channels, has less variation compared to arginine precursors although the results of the arginine peptides may be skewed by their small number at 62.5 pg (Fig 5C). Finally, we investigated the quantitative reproducibility of single-cell equivalents in dependence on the reference channel amount (scReference). The Pearson correlations were always greater than 0.88 using RefQuant or 0.8 for MS1 quantities and did not differ between target channels Δ4 and Δ8 (Appendix Fig S4B and C).

### mDIA doubles protein identifications per single cells

Having shown the benefits of the reference channel for identification and quantification, we next applied our mDIA workflow to single cells. We designated one of the channels as the reference (Δ0), leaving two target channels (Δ4 and Δ8) for single-cell analysis. With this setup, we routinely analyzed 80 single cells per day.

The only difference to the ultra-high sensitive mDIA workflow described above is that we upfront sorted single HeLa cells from culture by flow cytometry into 384-well plates. We FACS-isolated the cells in an unbiased manner; thus we expected a range of cell sizes depending on the cell cycle stage. This was reflected in a wide range of protein identifications across analyzed cells (Fig 6A). To investigate the cause of this spread, we considered the summed MS signal of all peptides as a proxy for total protein amount in our single cells. Plotting this signal against protein identifications revealed a strong dependency of input amount on proteome depth (Fig 6B), but is not caused by doublets (Appendix Fig S5A). This suggests that further gains in ion signals will allow even deeper coverage in the reference mDIA workflow in the future. We also note that FACS-sorting may lead to some drop outs which might be overcome with more sensitive sorting strategies as well as the possibility to take pictures of each isolated single cell to validate an intact cell for analysis as done in the cellenONE instrument (preprint: Hartlmayr *et al*, 2021). This may help to reduce drop outs and throughput.

However, even at this stage, we already identified 2,377 protein groups and 7,607 peptides per single cell and reached almost 4,000 protein groups in a few single cells (disregarding a single outlier at 4,600 identifications) (Fig 6A, Appendix Fig S5B and C). Across our 476 cells in the entire dataset, we obtained 5,997 proteins, a substantial part of the proteome expressed by a single cell. Even compared to some of the deepest proteomes that were achieved in this cell type with hundreds of µg rather than pg input (Bekker-Jensen *et al*, 2017), our single-cell proteomes together account for half of the proteins.

This is more than twice the number of the median identified in interphase cells in our previous publication, almost entirely attributable to the improvement due to the reference channel (Fig 6A) (Brunner *et al*, 2022). The overall coefficient of variation due to technical and biological factors was 0.4 without bias between the target channels (Fig 6C). Overall data completeness was high. After we filtered out proteins that were only detected in 15% of the

© 2023 The Authors

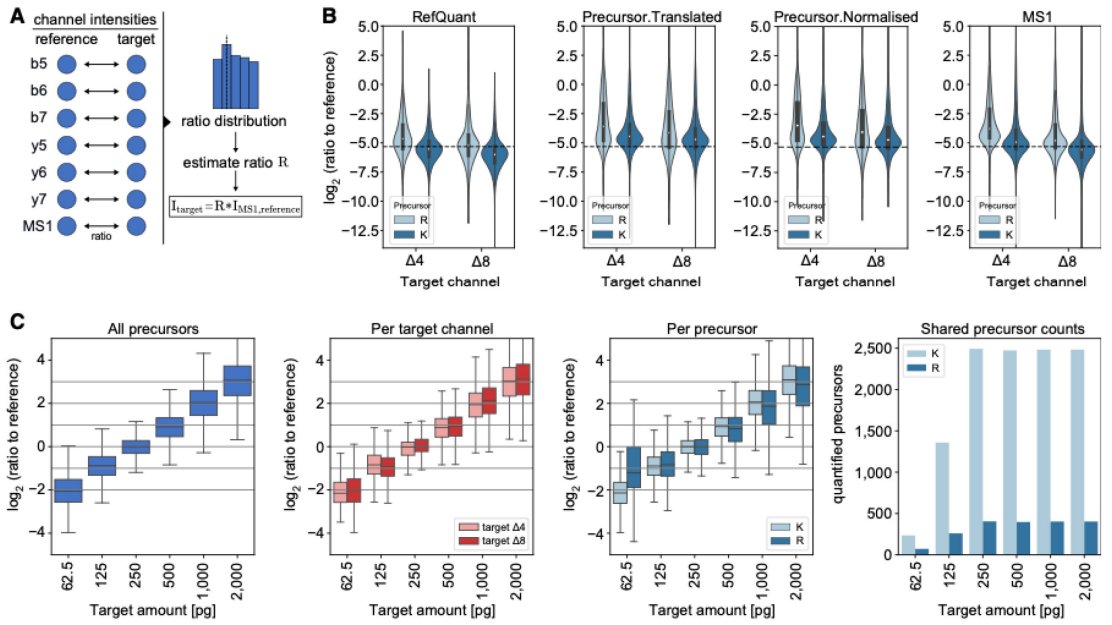



**Figure 5. RefQuant quantifies single-cell equivalents accurately based on sampling ratios relative to the reference channel.**

A Concept of RefQuant in calculating ratios between each fragment and MS1 peak for each precursor between reference and each target channel. The resulting ratio distribution is filtered for the first 40% of the quantile and the ratio factor R is calculated by the mean of the resulting distribution. This ratio factor R is then multiplied by the reference intensity to retrieve the RefQuant target intensity.

B Ratio of reference channel to each target channel by arginine (R) and lysine (K) precursor based on quantities from RefQuant (left), 'Precursor.Translated' and 'Precursor.Normalised' by DIA-NN (middle) and MS1 by DIA-NN (right) on the scBenchmark dataset (technical replicates, $n = 5$). RefQuant showed best the expected ratios in both target channels compared to DIA-NN quantities. The violin plot shows the distribution of the data while the box depicts the interquartile range with the central band representing the median value of the dataset. The whiskers represent the furthest datapoint within 1.5 times the interquartile range (IQR).

C RefQuant accurately quantified four-fold differences in protein amount between reference and target channels (scQuant dataset, 62.5 to 2000 pg in target channel, 10 ng in reference channel) and correctly extracted the expected ratios across channels for lysine (K) and arginine (R) precursors (technical replicates, $n = 5$). The ratio of reference to target channel was calculated using RefQuant. All $\log_2$ ratios were normalized to the single-cell equivalent sample in the target channel (250 pg) and only shared precursors are used (right panel). The box depicts the interquartile range with the central band representing the median value of the dataset. The whiskers represent the furthest datapoint within 1.5 times the interquartile range (IQR).

Source data are available online for this figure.

cells – possibly due to biological variation – data completeness of the remaining 3,696 proteins was 59.3% across the entire dataset. As expected, this depended on the abundance of the proteins, with completeness of the top rank order quartile at 94.1% and the lowest quartile at 36.2% (Fig 6D). The median overall Pearson correlation of the single cells was 0.79 (Fig 6E).

Overall, our protein signals cover more than four orders of magnitude of dynamic range, in which we identify many transcription factors, like the Hox family, STAT1-3 and SOX6, which range over the whole abundance range, but are mainly expressed at low amounts (Fig 6F). This constitutes to 20% of all human transcription factors (Lambert *et al*, 2018).

### Stable proteome at deeper proteome coverage

In our recent single-cell publication, we discovered evidence for a stable proteome while the transcriptome had a much higher overall variability as measured by CV. We attributed this difference to the fact that single cells need a complete proteome to function, whereas transcripts for genes expressed in these same cells are only needed rarely (Brunner *et al*, 2022). Thus, numbers of transcripts for many expressed genes are often below one on average across many single cells. A limitation of our previous study was the coverage of the single-cell proteome – about 1,000 proteins in interphase cells and up to 2,000 proteins in drug arrested mitotic cells.

Here, we used the same strategy as in our label-free single-cell dataset (Brunner *et al*, 2022). In brief, we investigated the variability of protein expression by the CV as a function of abundance by MS intensity (See Materials and Methods). Remarkably, we still see relatively small CVs across covered abundances for the whole measured dynamic range (Appendix Fig S5D). Previously, we had defined a 'core proteome' of the top 200 proteins with at least 70% data completeness. We also observe lower CVs and higher mean MS intensity for these compared to noncore proteins (Appendix Fig S5E and F).

 © 2023 The Authors

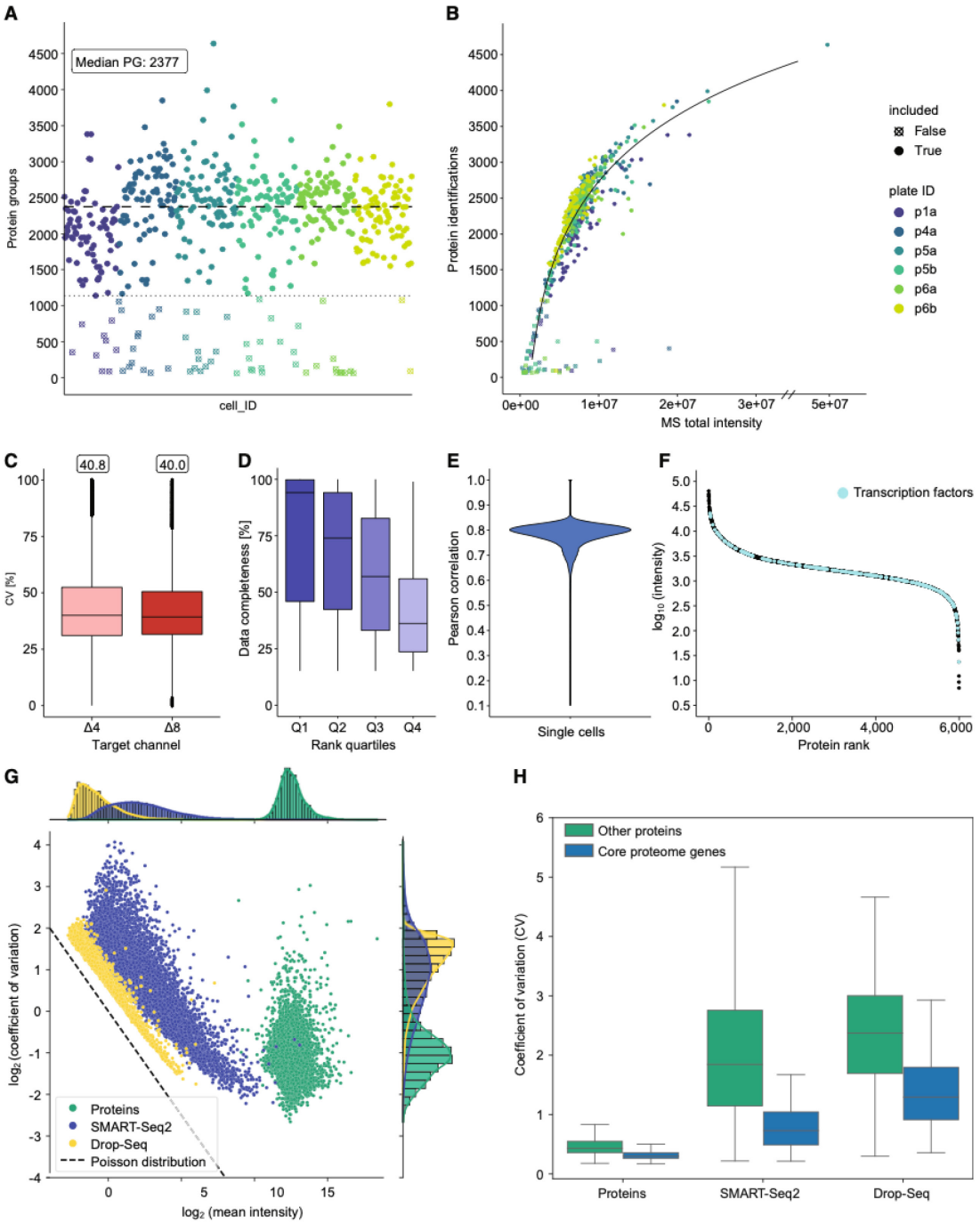



**Figure 6.**

© 2023 The Authors

**Figure 6. Reference channel doubles protein identifications per single cell and the concept of a stable proteome is still valid.**

A Protein identifications of single-cell measurements. Four hundred and seventy-six single HeLa cells reveal a median identification of 2,377 with few single cells of identifications up to 4,000 protein groups over six 384-well plates (disregarding a single outlier of 4,600 protein groups).

B Protein identifications versus the total sum of MS signal intensity per cell measurement shows a logarithmic dependency.

C CV distribution of single cells in each target channel (biological replicates, $n = 476$). The box depicts the interquartile range with the central band representing the median value of the dataset. The whiskers represent the furthest datapoint within 1.5 times the interquartile range (IQR).

D Completeness of ranked quartiles (Q1: 1–924; Q2: 925–1,848; Q3: 1,849–2,772; Q4: 2,773–3,696). Proteins with high rank show higher data completeness compared to low ranked proteins (biological replicates, $n = 476$). The box depicts the interquartile range with the central band representing the median value of the dataset. The whiskers represent the furthest datapoint within 1.5 times the interquartile range (IQR).

E Pearson correlation of single cells of the mDIA workflow (median 0.79).

F Rank plot of close to 6,000 proteins. Transcription factors are highlighted across abundance range (cyan).

G Coefficients of variation of single-cell mDIA protein expression plotted against the mean intensity of each protein for single-cell proteins (green) and scRNAseq of SMART-Seq2 (blue) and Drop-Seq (yellow).

H Comparison of coefficients of variation between single-cell proteomics and scRNAseq with regards to the 'core proteome' (blue) and other genes (green) (biological replicates, scProteomics $n = 476$, SMART-Seq2 $n = 720$, Drop-seq $n = 5{,}665$). The box depicts the interquartile range with the central band representing the median value of the dataset. The whiskers represent the furthest datapoint within 1.5 times the interquartile range (IQR).

Source data are available online for this figure.

The somewhat increased CVs of the proteins that were only quantified here, is readily explained by their lower signal in the MS.

As before, we next compared our mDIA single-cell dataset to scRNAseq data based on Drop-Seq (Macosko *et al*, 2015) and SMART-Seq2 (Picelli *et al*, 2014) on the same cell type. In the count distribution plot, scRNAseq and mDIA scProteomics data again separate from each other, indicating the stable proteome at higher proteomic coverage (Fig 6G). CV values for the proteins are tight and low while scRNAseq has higher CV values (Fig 6H). Together these data show that the concept of a stable proteome still holds true at the higher proteomic depth achieved here.

### mDIA advances single cell type resolved spatial proteomics in oncology

We reasoned that the multiplexing and quantitative attributes of mDIA should be of great advantage in tissue proteomics, especially in the context of diseases. In particular, we wanted to integrate it with our recent technology termed Deep Visual Proteomics (Mund *et al*, 2022). DVP combines artificial intelligence-driven image analysis of cellular phenotypes with automated single-cell laser microdissection and ultra-high-sensitivity mass spectrometry, effectively linking protein abundance to complex cellular phenotypes while preserving the spatial context. To explore the advantages of mDIA in this context, we profiled cancer cells in-situ in primary cutaneous melanoma. Using routine histopathological markers, we segmented single melanoma cells and further stratified them according to their spatial location to the epidermal or dermal compartment, thereby taking the tumor microenvironment into account (Fig 7A). In recent work on superficial spreading melanoma, we had cut 700 shapes from 2.5 µm-thin formalin-fixed paraffin-embedded (FFPE) tissue sections (Mund *et al*, 2022). We found that the mDIA workflow seamlessly integrated into the DVP pipeline. For the design of the reference channel, we used bulk digest from a consecutive tissue slide, containing mainly but not exclusively tumor material. We expected mDIA to be much more sensitive in this context and therefore only excised 100 single cell shapes (corresponding to 20 cell equivalents at 2.5 µm thickness). We added the reference channel to the epidermal and dermal target channels in a roughly estimated ten-fold excess. With a Whisper20 SPD gradient (58 min active gradient), we quantified 4,000 protein groups in melanoma cells within the epidermis and 2,740 in the dermis, similar to our previous report on the same cell type. However, our mDIA-DVP pipeline used seven times less input amount, a shorter gradient (1 h vs. 2 h) and measured two samples per run. This represents an increased throughput of four-fold for MS analysis without losing proteomic depth and a seven-fold decreased cutting time on the laser microdissection instrument, a large gain for the overall DVP pipeline.

To test the cell type-specific and biological validity, we looked for typical melanocytic markers (SOX10, MITF, DCT, MLANA, PMEL, TYR, TYRP1) (Belote *et al*, 2021) in our data. We detected all but TYRP1, including SOX10, the most important (transcription) factor of melanocytic lineage used in routine clinical diagnostics (Fig 7B). Interestingly, expression of these identity markers was lower in dermal melanoma cells, highlighting the role of the tumor microenvironment in cancer-cell identity and dedifferentiation (Fig 7C). Next, we sought to assess proteomic differences between epidermal and dermal melanoma cells. In principal component analysis (PCA), component 1 clearly separated dermal from epidermal melanoma cells (Fig 7D). Importantly, this was irrespective of the target channel used for labeling individual replicates, which we ascertained by label swapping. Compared to melanoma cells in the epidermis, melanoma cells located in the dermal compartment had significantly higher expression of proteins involved in remodeling of the extracellular matrix (e.g., DPT, COL1A2, COL1A1, COL3A1, COL6A1, DCN, TGFBI, PRELP, DCN, FGA, FN1, and LUM) (Fig 7E). In contrast, melanoma cells of the dermal compartment had significantly lower expression of TACSTD2 and LGALS7, amongst others (Fig 7E). While downregulation of TACSTD2 is part of a gene expression profile signature that predicts increased risk of cancer progression, LGALS7 is involved in promoting cellular apoptosis and could therefore lead to increased tumor cell survival (Biron-Pain *et al*, 2013; Gerami *et al*, 2015). The protein Keratin 15 (KRT15), which is normally expressed in basal keratinocytes of the epidermis, has also been found to be associated with tumor stage and prognosis in metastatic melanoma patients, with higher expression in primary tumors and loss in metastases (Han *et al*, 2021). We also observed a reduced expression of KRT15 in cells located in the dermal compared to the epidermal melanoma cells according to the depth of invasion (Fig 7E). Pathway enrichment revealed multiple and highly relevant signaling cascades enriched in dermal melanoma cells, such as senescence, PI3K-AKT signaling, and IL-18

 © 2023 The Authors

signaling (Fig 7F). Taken together, the combination of mDIA and DVP enables high-throughput, in-depth and biologically relevant insights of cancer cells from single patients, including the spatial component.

## Discussion

Here, we described a robust and economical mDIA workflow that we explored using standard proteomics samples and single cells.

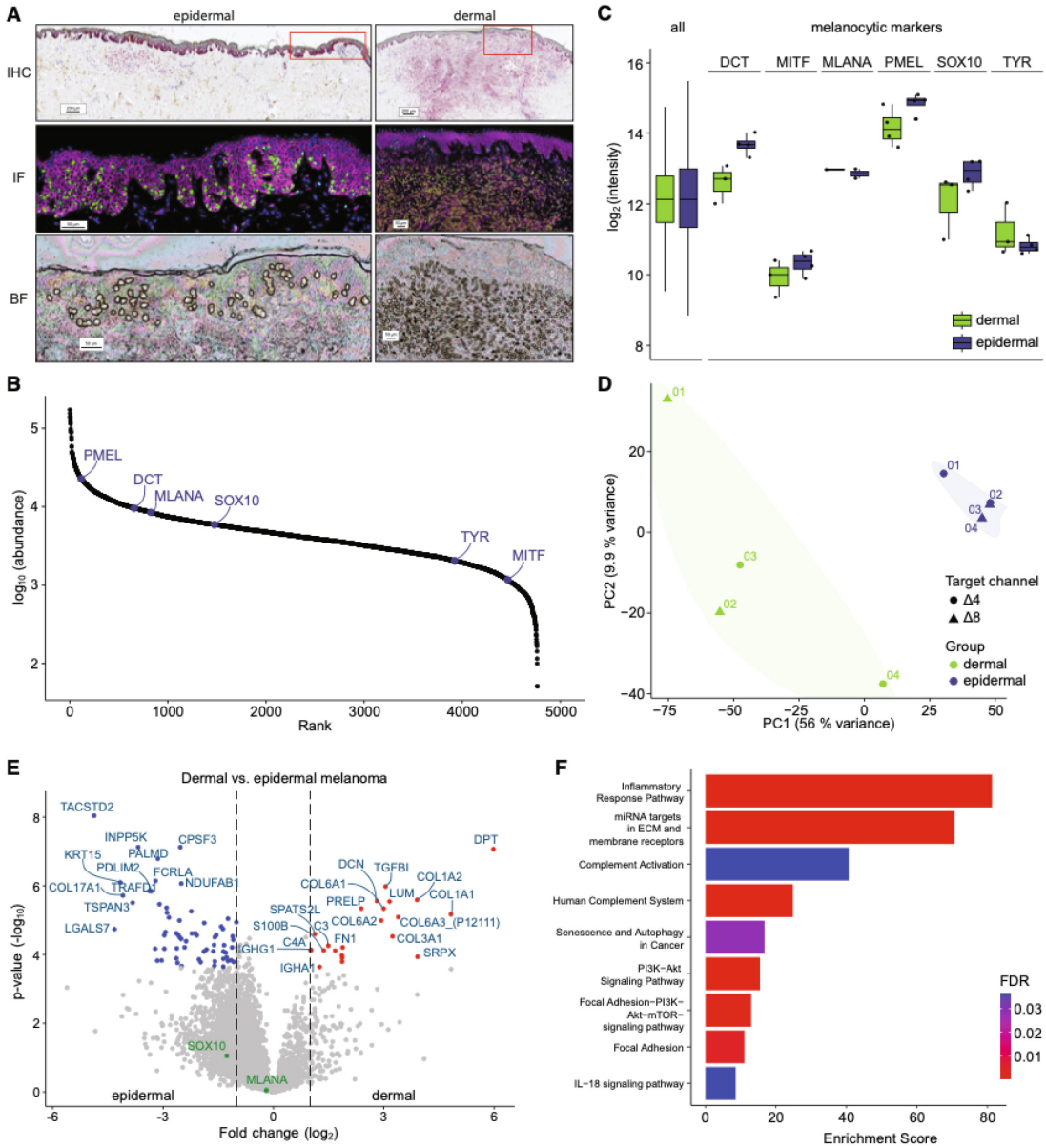



**Figure 7.**

© 2023 The Authors

**Figure 7. mDIA-DVP applied to primary cutaneous melanoma showcases its potential for precision oncology.**

A Macroscopic overview of primary cutaneous melanoma (type SSM, Breslow 3.5 mm) including Melan-A immunohistochemistry (IHC, pink) and CD44/Sox10 immunofluorescence (IF, pink/green), as well as a brightfield (BF) image after laser microdissection. Segmented melanoma cells are highlighted (yellow outlines). Quadruplicates of 100 shapes (20 cell equivalents) were laser-microdissected from the epidermal (left) and dermal (right) compartment using Deep Visual Proteomics (DVP).

B Rank plot of all proteins identified in melanoma cells. Six out of seven melanocyte identity markers were identified (blue).

C Boxplot of $\log_2$ intensities of all proteins compared to six out of seven identified melanocytic markers (Belote *et al*, 2021) (biological replicate, $n = 4$ in each group). The box depicts the interquartile range with the central band representing the median value of the dataset. The whiskers represent the furthest datapoint within 1.5 times the interquartile range (IQR).

D Principal component analysis of dermal and epidermal melanoma cells which differentiate on PC1 irrelevant of the target channel used.

E Differential protein expression between epidermal (left) and dermal (right) melanoma cells. Significant proteins (Benjamini–Hochberg corrected multiple-sample *t*-test, $q$-val/FDR $< 0.01$) with a $\log_2$ fold change $> 1$ (red) and $< 1$ (blue) are highlighted, respectively.

F Overrepresentation analysis of significantly enriched proteins in dermal melanoma using the Wikipathway database. Color represents FDR (threshold $< 0.05$).

One of its main advantages is that it uses dimethyl labeling, a well-established, very simple and robust technology (Hsu *et al*, 2003; Boersema *et al*, 2009). We consistently achieved labeling efficiencies of more than 99% for peptides generated by trypsin and Lys-C – proteases typically employed in proteomics. The labeling protocol consists of only three pipetting steps, hence it was easily automatable on a standard liquid handling robot. The differentially labeled peptides are combined into one stage-tip for purification and storage, thus avoiding sample loss. Another practical advantage is that the reagents themselves are very stable, whereas broadly used NHS (N-Hydroxysuccinimide)-chemistry-based reagents such as mTRAQ or TMT are more labile.

Dimethyl labeling enables quantification at both the MS1 and MS2 level, but is limited to three-plex when used in combination with trypsin. Here we show that Lys-N digestion followed by dimethyl labeling accommodates a five-plex format at both levels as well, with the interesting property that b-ions are split up into five quantifiable states separated by 4 Da. The y-ions are less intense due to the absence of a C-terminal lysine, however, this is compensated for identification purposes by the fact that y-ions from all channels stack on top of each other. This will be particularly interesting in the context of fragmentation with Electron Transfer Dissociation (ETD), as this has been shown to generate nearly pure c-ion series in many cases on triply charged precursors (Hennrich *et al*, 2010). It might also be interesting if synchro-PASEF (Skowronek *et al*, 2022a) can distinguish y-ions from different channels by different slicing positions of the precursor-fragment pair, since the different channels slightly differ in their ion mobility versus m/z position. More generally, mDIA should fully benefit from new scan modes and different mass spectrometry platforms.

Current software tools do not yet make optimal use of dimethyl mDIA data. Unlike $^{13}$C-based labels, dimethyl labels can slightly influence retention times on LC columns. Our results show that this is not a major impediment to quantification, but we believe that results will improve once the retention times are directly modeled and corrected for. We have begun to do this with AlphaPeptDeep (Zeng *et al*, 2022), which we also used to derive the predicted DIA libraries.

While proteome depth is not compromised by dimethyl labeling per se, we did notice a drop in identifications for multiplexed as opposed to single proteomes (although overall identifications and quantifications are still vastly higher). As this correlates with the number of labeled channels, we attribute it to the increased complexity of the DIA data which will eventually lead to an overlap in DIA transitions. At least some of this drop in identifications should be retrievable by further optimization of the DIA software as the extra complexity follows a clearly defined pattern in the channel spacing. In tri-plex format and combined with the relatively short gradients on the Evosep instruments, this achieves a throughput of 3 × 60 samples per day for proteome measurements, which would make it quite practical to incorporate an up-front fractionation step in the future.

We also explored the idea of a reference channel in single-cell mDIA. Note that this is conceptually different from the carrier channel employed in the SCoPE-MS method (Budnik *et al*, 2018) because the fragments of the mDIA channels are offset from each other and do not contribute to a common low reporter mass (Derks *et al*, 2022). Existing DIA software such as Spectronaut or DIA-NN, which we used here, do not explicitly make use of the reference channel yet, and the concept of a separate FDR for individual target channels is not fully developed yet. Nevertheless, we found that identifications roughly doubled, from just above thousand proteins for interphase cells in our previous single-cell publication (Brunner *et al*, 2022) to a median of 2,370 now. Encouragingly, almost 4,000 proteins were identifiable for some of the cells. When combined with further ongoing instrumental and sample preparation advances, this may enable characterization of more than half of the expressed proteome in single cells in the near future. For our single-cell analysis, we use the low flow, high sensitivity Whisper gradients on the Evosep, leading to a throughput of 80 single cells per day when using a reference channel, which would increase to 160 per day or 1,000 per week with the Lys-N five-plex method (subtracting one channel for the reference).

We are currently exploring the applicability of mDIA for a wide variety of biological hypotheses and projects that can now be addressed with proteomics. In regular analyses, we have found that the method retains most of the advantages of DIA, while tripling the throughput at little if any extra cost, as others have found as well (Derks *et al*, 2022). When transferred to our DVP technology (Mund *et al*, 2022), mDIA is especially attractive because different cell states that are spatially separated can be directly compared to each other. Furthermore, we found that the exquisite sensitivity of mDIA with a reference channel allows us to drastically reduce the number of shapes. Rare cell types can now rapidly be profiled at great proteomic depth and increased spatial resolution. Importantly, this also

 © 2023 The Authors

enables the analysis of true single cells in a spatial context (preprint: Rosenberger *et al*, 2022). Altogether, we believe that mDIA-DVP will become a key enabling technology for precision oncology.

We look forward to further exploration of the reference channel concept. Combined with the inherent simplicity of the DIA acquisition method, it may lead to much more complete proteomes. Furthermore, the reference channel will make proteomic results much more comparable across experiments and between laboratories, something that should greatly benefit the community. This is applicable particularly for clinical cohorts like for example plasma proteomics (Bader *et al*, 2023). Together, the mDIA workflow and the reference channel opens up new opportunities for single-cell proteomics and clinical proteomics.

## Materials and Methods

### Reagents and Tools table

| Reagent/Resource | Source | Identifier |
|---|---|---|
| **Chemicals, Enzymes, and other reagents** | | |
| Bovine Serum Albumin (BSA) | New England BioLabs Inc. | P8108S |
| HeLa tryptic digest | Pierce | 1862824 |
| Yeast (*S. cerevisiae*) tryptic digest | Promega | V746A |
| *E. coli* tryptic digest | Waters | 186003196 |
| Triethylammonium bicarbonate (TEAB) | Sigma Aldrich | T7408 |
| Ammonium bicarbonate (ABC) | Merck | 1.01131 |
| Urea | Sigma Aldrich | U1250 |
| Dithiothreitol (DTT) | Sigma Aldrich | D0632 |
| Chloroacetamide (CAA) | Sigma Aldrich | C0267 |
| Formaldehyde: 'light' ($CH_2O$) | Sigma Aldrich | F8775 |
| Formaldehyde-d2-solution: 'intermediate' ($CD_2O$) | Sigma Aldrich | 492620 |
| Formaldehyde-$^{13}C$, d2-solution: 'heavy' ($^{13}CD_2O$) | Sigma Aldrich | 596388 |
| Sodium cyanoborohydride: 'light' ($NaBH_3CN$) | Sigma Aldrich | 156159 |
| Sodium cyanoborodeuteride: 'heavy' ($NaBD_3CN$) | Sigma Aldrich | 190020 |
| Lichropur® ammonia solution, 25% | Merck | 5.33003 |
| LC–MS-grade water | Fisher Chemical | W6-4 |
| Trifluoroacetic acid (TFA) | Sigma Aldrich | 8.08260 |
| Formic acid (FA) | Sigma Aldrich/Merck | #64-18-6 |
| Acetonitrile | Sigma Aldrich/Merck | #75-05-8 |
| CDS Empore C18 Extraction Disks | Empore | 13-110-018 |
| Trypsin | Sigma-Aldrich | T6567 |
| Lys-C | FUJIFILM Wako | 125-05061 |
| Lys-N | ImmunoPrecise Antibodies (Europe) BV | L101 |

**Reagents and Tools table** (continued)

| Reagent/Resource | Source | Identifier |
|---|---|---|
| **Software** | | |
| MaxQuant (2.0.1) | https://maxquant.org/ | N/A |
| DIA-NN (1.8.1) | https://github.com/vdemichev/DiaNN | N/A |
| RStudio (2022.07.2) and R (4.2.2) | https://posit.co/download/rstudio-desktop/ | N/A |
| Jupyter Notebook | https://jupyter.org/ | N/A |
| RefQuant | https://github.com/MannLabs/refquant | N/A |
| **Other** | | |
| 384-well plates | Eppendorf | #0030129547 |
| Adhesive PCR Sealing Foil Sheets | Thermo Scientific | #AB-0626 |
| Bravo robot | Agilent | |
| ThermoMixer | Eppendorf | #460-0223 |
| Mastercycler X50h | Eppendorf | #63160000 |
| NanoDrop One/OneC Microvolume UV–Vis Spectrophotometer | Thermo Fisher | @ND-ONEC-W |
| EvoSep One | EvoSep | #EV-1000 |
| Evotip Pure | EvoSep | #EV-2011 |
| timsTOF HT | Bruker Daltonik GmbH | N/A |
| timsTOF SCP | Bruker Daltonik GmbH | N/A |
| timsTOF Ultra | Bruker Daltonik GmbH | N/A |
| Orbitrap Exploris 480 | Thermo Scientific | N/A |
| Column oven | Sonation lab solutions | #PRSO-V2 |
| Aurora Elite CSI | IonOpticks | AUR3-15075C18-CSI |

### Methods and Protocols

#### *Sample preparation of bulk samples*

*Bovine serum albumin*

CAM-modified tryptic digest of bovine serum albumin (New England BioLabs Inc.) was labeled with dimethyl using the in-solution labeling protocol (Boersema *et al*, 2009). Briefly, 500 pmol of the BSA digest was reconstituted in 300 µl of 100 mM triethylammonium bicarbonate (TEAB, pH 7) buffer and aliquoted into three separate tubes. To each of these, 4 µl of 4% (vol/vol) formaldehyde ($CH_2O$, $CD_2O$ or $^{13}CD_2O$) and 4 µl of 600 mM sodium cyanoborohydride ($NaBH_3CN$ or $NaBD_3CN$) were added sequentially. The samples were incubated at room temperature for 1 h on a bench-top mixer. To quench the reaction, 16 µl of 1% (vol/vol) ammonia ($NH_4OH$) was added to each tube. The solutions were then dried at room temperature using a speed vacuum concentrator, reconstituted in 200 µl of buffer A (0.1% formic acid), and desalted using in-house prepared, C18 StageTips (Rappsilber *et al*, 2003). The solutions were then adjusted to have a concentration of 50 fmol/µl. Three different mixtures were prepared by combining the three channels (Δ0/Δ4/Δ8) in varying mixing ratios: Mix 1 (8.5 µl/1.0 µl/0.5 µl), Mix 2 (7 µl/2 µl/1 µl), and Mix 3 (5 µl/3 µl/2 µl). A 1 µl

© 2023 The Authors

aliquot of each mix was injected into the Thermo Orbitrap Exploris 480™ in triplicates, corresponding to an injection amount of 50 fmol per replicate.

*Tryptic HeLa*

A similar in-solution labeling protocol was followed for tryptic HeLa digests (Pierce #1862824) using 10 µg of starting material per channel. After C18 desalting, four different mixes were prepared: 62.5 ng/µl of label-free solution, 62.5 ng/µl of 1-channel (Δ0), 125 ng/µl of 2-plex (1:1 Δ0/Δ4), and 187.5 ng/µl of 3-plex (1:1:1 Δ0/Δ4/Δ8). A 2 µl aliquot of each mix was injected into the Thermo Orbitrap Exploris 480™ in triplicates and another 2 µl aliquot of each mix was injected into the Bruker timsTOF HT in triplicates. These correspond to injection amounts of 125 ng for unlabeled and one-channel solutions, 250 ng for two-plex solutions, and 375 ng for three-plex solutions, per replicate.

*Mixed species (human/yeast/E. coli)*

For the mixed species experiment, three different mixtures with varying mixing ratios of HeLa tryptic digest (Pierce #1862824), *S. cerevisiae* tryptic digest (Promega V746A), and *E. coli* tryptic digest (Waters #186003196) were prepared prior to labeling: Mix 1 (5.20 µg/1.20 µg/1.60 µg Human(H)/Yeast(Y)/*E. coli*(E)), mix 2 (5.20 µg/2.40 µg/0.40 µg H/Y/E), and mix 3 (5.20 µg/0.40 µg/2.40 µg H/Y/E). Mix 1 was then labeled with Δ0, mix 2 with Δ4, and mix 3 with Δ8, following the in-solution labeling protocol. After C18 desalting, each mixture was adjusted to 150 ng/µl using buffer A (0.1% formic acid) and the three labeled mixtures were combined with each other in a 1:1:1 ratio. A 3 µl aliquot of each mixture was injected into the Bruker timsTOF HT in triplicates for each of the two acquisition methods (20S and 16S4MS1, S = dia-PASEF scan) tested in this study. This corresponds to an injection amount of 450 ng total per replicate for each acquisition method.

*Sample preparation of mixed species for FDR estimation*

To empirically determine the false discovery-rate (FDR), 4 µg S. cerevisiae tryptic digest (Promega V746A) and HeLa tryptic digest (Pierce #1862824) each was mixed together and labeled with Δ0. Additionally, 4 µg yeast tryptic peptides were labeled with Δ4 and Δ8, respectively. After C18 desalting, each solution was adjusted to 100 ng/µl using buffer A (0.1% formic acid) and the three labeled solutions were then mixed together in a 2:1:1 (Δ0/Δ4/Δ8) ratio to have equally distributed amount of yeast in each channel. A 4 µl aliquot of each mixture was injected into the Bruker timsTOF HT in triplicates, corresponding to an injection amount of 400 ng per replicate.

*Lys-N-derived HeLa*

The Lys-N digest was prepared in-house. A HeLa cell pellet was lysed in 8.0 M urea, 50 mM ABC pH 8.5 and incubated in a Bioruptor for 10 min. Reduction of disulfide bonds was performed by adding dithiothreitol (DTT) to a final concentration of 10 mM DTT and incubating at room temperature for 45 min. Carbamidomethylation of the free sulfhydryl groups was achieved by adding chloroacetamide (CAA) to a final concentration of 40 mM CAA and incubating in the dark at room temperature for 15 min. The reaction was then quenched with DTT to a final concentration of 10 mM DTT. The resulting lysate was diluted to ensure that the concentration of urea in the solution is less than 2.0 M (diluted to 1.0 M). Lys-N (ImmunoPrecise Antibodies) was added to the lysate in a 1:100 (enzyme/protein) ratio and the digestion was left to digest at 37°C overnight. The mixture was acidified to a final concentration of 1% TFA to quench the digestion and desalted using in-house prepared, C18 stage tips. The same in-solution labeling protocol was used for labeling 25 µg of HeLa Lys-N peptides per channel. Five different mixes were prepared from the labeled peptides: 50 ng/µl of unlabeled solution, 50 ng/µl of one-channel (Δ0), 100 ng/µl of two-plex (1:1 Δ0/Δ4), 150 ng/µl of three-plex (1:1:1 Δ0/Δ4/Δ8), and 250 ng/µl of five-plex (1:1:1:1:1 Δ0/Δ2/Δ4/Δ6/Δ8). A 2 µl aliquot of each mix was injected into the Bruker timsTOF HT in triplicates. These correspond to injection amounts of 100 ng for unlabeled and one-channel solutions, 200 ng for two-plex solutions, 300 ng for three-plex solutions, and 500 ng for 5-plex solutions, per replicate. For the 5-plex accuracy test, purified labeled HeLa samples were mixed together in a ratio of 1:2:4:2:1 (Δ0, Δ2, Δ4, Δ6, Δ8) and a total of 500 ng was used per injection.

***Sample preparation for reference channel and single-cell equivalents***

HeLa cells were cultured in Dulbecco's modified Eagle's medium at 10% fetal bovine serum, 20 mM glutamine and 1% penicillin–streptomycin. The supernatant was carefully removed, cells were detached by accutase treatment for 5 min at 37°C followed by pipetting for cell aggregate dissociation. Cells were washed three times with ice-cold PBS, lysed by 60 mM TEAB and 20% acetonitrile (ACN) at 72°C for 30 min. Proteins were digested by trypsin and Lys-C in a 1:50 enzyme:protein ratio. Peptides were cleaned up by C18 desalting as described above. Cleaned up peptides were labeled with light dimethyl (Δ0) as described above in 100 µl 60 mM TEAB (pH 8.5) to be comparable to the single-cell workflow. Evotips were loaded as described in the single-cell experiments.

***Sample preparation for single-cell experiments***

HeLa cells were cultured and harvested following standard protocol as above. Cells were washed three times with ice-cold PBS, pelleted by centrifugation and resuspended in ice-cold PBS to a solution of $2 \times 10^6$ cells per ml. Two microliters propidium iodide (PI, BioRad, 1351101) was added to the single-cell suspension and sorting was performed on the PI-negative live cell population using fluorescent-activated cell sorting (FACS). Single cells were sorted into 384-well plates containing 1 µl of 20% ACN, 0.01% n-Dodecyl β-D-maltopyranoside (DDM), 60 mM TEAB pH 8.5, centrifuged briefly, sealed with aluminum seal sheets and frozen at −80°C until further processing. Single cells were incubated in the 384-well plate for 30 min at 72°C in a PCR cycler with a lid temperature of 110°C, digested overnight at 37°C (lid temperature 55°C) after addition of 1 µl of 20% ACN, 0.01% DDM, 60 mM TEAB and 0.5 ng trypsin and Lys-C each. Digested peptides were derivatized with dimethyl by adding 1 µl 0.6% formaldehyde (final concentration 0.15%) and 1 µl 92 mM cyanoborohydride (final concentration 23 mM). Single cells in a consecutive row pattern (B, D, F, ...) were labeled with intermediate dimethyl (Δ4) using intermediate formaldehyde ($CD_2O$) and light cyanoborohydride, while other wells (C, E, G, ...) were labeled with heavy dimethyl (Δ8) using heavy formaldehyde ($^{13}CD_2O$) and heavy cyanoborohydride. After 1 h incubation at room temperature, the dimethyl labeling reaction was stopped by

 © 2023 The Authors

adding 1 µl 0.65% ammonia solution (final concentration 0.13%). Before Evotip loading, samples were acidified by 1 µl 6% TFA (final concentration 1%). All pipetting steps were done with a Bravo robot (Agilent). Evotips Pure were loaded with the Bravo robot (Agilent), by activation with 1-propanol, washing two times with 50 µl buffer B (99.9% ACN, 0.1% FA), activation with 1-propanol and two wash steps with 50 µl buffer A (99.9% $H_2O$, 0.1% FA). In between Evotips were spun at 700 *g* for 1 min. For sample loading, Evotips were prepared with 70 µl buffer A and a short spin at 700 *g*. Ten microliters of 1 ng/µl prepared and aliquoted reference channel proteome (labeled with light dimethyl, Δ0) was pipetted into Evotips, followed by an intermediate and heavy dimethyl labeled single cell with the same tip of the Bravo robot to avoid plastic contacts. Single-cell wells were washed with 15 µl buffer A and also loaded onto the Evotips. Evotips were spun at 700 *g* to load the sample and washed ones with 50 µl buffer A and stored at 4°C with buffer A on top until measured.

#### Multiplexed-DIA Deep Visual Proteomics (mDIA-DVP)

##### *Specimen preparation and immunofluorescence staining*

Primary cutaneous melanoma (subtype superficial spreading melanoma, Breslow-depth 2.5 mm) was obtained as formalin-fixed paraffin-embedded (FFPE) tissue according to the institutional review board protocol (BASEC. Nr:2014-0425) at the University Hospital of Zurich, Switzerland. The tissue was collected in accordance with the Declaration of Helsinki following standard operating procedures. 2.5 µm thin tissue sections were mounted on Polyethylen-Naphthalat (PEN) membrane slides (MicroDissect GmbH, MDG3P40AK), coated with Vectabond (Biozol, VEC-SP-1800). Tissue specimens were processed as described previously (Mund *et al*, 2022; preprint: Nordmann *et al*, 2023). In brief, slides were deparaffinized (2 min xylene, 1 min 100% EtOH, 95% EtOH, 75% EtOH, 30% EtOH, and ddH20, repeated twice, respectively) and glycerol-supplemented heat induced epitope-retrieval (G-HIER) (Tris/EDTA pH 9 HIER buffer (DAKO, S2367), 10% glycerol (v/v; Sigma, G7757)) was performed for 20 min at 88°C. Thereafter, unspecific binding sites were blocked using 5% bovine serum albumin (BSA) in PBS for 30 min at RT, followed by the application of primary antibodies for CD44 (rabbit, Abcam ab51037; 1:200) and SOX10 (mouse, NordicBiosite bsh-7959-1; 1:50) for 90 min in a humidified staining chamber. Secondary antibodies against mouse-IgG (Alexa-647, Invitrogen A32728, 1:400) and rabbit-IgG (Alexa-555, Invitrogen A32732, 1:400) were applied for 1 h, followed by 7 min incubation with SYTOX green Nucleic Acid Stain (Invitrogen S7020; 1:700 in ddH20). Potential membrane distortions were released using a needle (30G) and sealed thereafter, as described previously (preprint: Nordmann *et al*, 2023). Cover glasses were mounted using Slowfade Diamond Antifade Mountant (Invitrogen, S36967) and removed after high-resolution imaging in ddH20. All staining steps were performed at room temperature unless otherwise specified.

##### *Artificial intelligence (AI)-based image analysis and cell extraction*

High-content images of mounted tissue specimens were acquired on a Zeiss Axioscan Z7 at 20× magnification with a 10% tile overlap. Resulting data were submitted to AI-based cell segmentation and classification using the Biology Image Analysis Software (BIAS, Cell Signaling). Cell outlines were defined based on CD44 staining and supervised machine learning (Mode: Deep Neural Network) to further stratify melanoma cells according to SOX10 positivity. Epidermal and dermal location was specified by manual region annotation. Contours of identified cells were transferred to a LMD7 laser microdissection instrument (Leica Microsystems) and cell registration was achieved using three reference points in the tissue regions of interest. Cells were extracted using the middle pulse function and an outline offset of 0.25 µm into a 384-well plate by gravity. Samples were collected in quadruplicates for the respective number of contours.

##### *Sample processing for mDIA-DVP*

DVP samples were processed on a liquid handling platform (Bravo, Agilent Technologies). Extracted cells were collected at the bottom of each well by acetonitrile-supported centrifugation and vacuum evaporation until dryness. Thereafter, lysis was accomplished in 6 µl 70 mM TEAB, 0.013% DDM and incubation at 95°C for 60 min, complemented by additional 60 min at 75°C in 12% (v/v) acetonitrile (by adding 1 µl 80% ACN). Proteins were digested using 4 ng LysC and 6 ng trypsin overnight at 37°C and labeled as described above. For the reference channel, a consecutive tissue slide – containing mainly but not exclusively tumor material – was collected from a PEN membrane slide into a 96-well plate (Covaris) using a scalpel and processed analogous to the DVP samples using 42 µl of lysis buffer and additional sonification (LE220-plus: peak Power: 450.0, duty factor: 50%, cycles: 200, average power: 225, Covaris). The following steps of digestion, purification and labeling were performed as described above in sample prepartion of the reference channel. Evotip loading was accomplished as described above using 25 ng of the bulk tissue as reference channel (Δ0) combined with the target channels. Two replicates of each melanoma type were measured as target Δ4 and Δ8, respectively.

#### Acquisition of bulk data

All samples were injected using the EASY-nLC 1200 (Thermo Scientific) and separated via in-house pulled columns (50 cm length, 75 µm ID) packed with 1.9 µm ID ReproSil-Pur 120 C18-AQ (Dr. Maisch GmbH). These samples were sprayed into either an Orbitrap Exploris 480™ (Thermo Scientific) or a timsTOF HT (Bruker Daltonics) for MS analysis. Buffer A is 0.1% formic acid in LC–MS-grade water. Buffer B is 80% acetonitrile in LC–MS-grade water with 0.1% formic acid.

##### *Bovine serum albumin*

For BSA samples, a 30-min active gradient was used as follows: 2–7% Buffer B (minutes 0–1), 7–30% Buffer B (minutes 1–15), 30–65% Buffer B (minutes 15–18), 65–95% Buffer B (minutes 18–21), 95% Buffer B (minutes 21–24), 95–5% Buffer B (minutes 24–27), and 5% Buffer B (minutes 27–30). The flow rate was kept constant at 300 nl/min. *DDA*. For determining the labeling efficiency of each dimethyl channel, shotgun DDA was employed on the Thermo Orbitrap Exploris 480™. The MS1 full scan range was 300–1,650 m/z with a resolution of 60,000, normalized automatic gain control (AGC) target of 300%, and a maximum injection time of 25 ms. This shotgun DDA approach was filtered for an intensity threshold of 5,000, with inclusion of charge states 2–5 and a dynamic exclusion duration of 30 s. There are five data dependent MS2 scans with a resolution of 15,000, normalized AGC target of 100%, maximum injection time of 28 ms, 1.4-Th isolation window width, and 27%

© 2023 The Authors

normalized HCD collision energy. *DIA*. The DIA mode consists of one MS1 full scan followed by 16 MS2 windows with variable widths (Appendix Table S1). Each MS1 full scan (300–1,650 m/z) was conducted at 120,000 resolving power, 300% normalized AGC target, and 100 ms maximum injection time. Each MS2 scan was conducted at 30,000 resolving power, 3,000% normalized AGC target, and 30% normalized HCD collision energy with a default charge of 2. The RF lens was set to 40%.

*Tryptic HeLa*

For the tryptic HeLa digests, a 75 min active gradient was used as follows: 2–7% Buffer B (minutes 0–1), 7–30% Buffer B (minutes 1–60), 30–50% Buffer B (minutes 60–66), 50–60% Buffer B (minutes 66–70), 60–90% Buffer B (minutes 70–71), and 90% Buffer B (minutes 71–75). The flow rate was kept constant at 300 nl/min. *Orbitrap DDA*. For determining the labeling efficiency of each dimethyl channel, shotgun DDA was employed on the Thermo Orbitrap Exploris 480™. The MS1 full scan range was 300–1,650 m/z with a resolution of 60,000, normalized automatic gain control (AGC) target of 300%, and a maximum injection time of 25 ms. This shotgun DDA approach was filtered for an intensity threshold of 100,000, with inclusion of charge states 2–5 and a dynamic exclusion duration of 30 s. There are 15 data dependent MS2 scans with a resolution of 15,000, normalized AGC target of 100%, maximum injection time of 28 ms, 1.4-Th isolation window length, and 30% normalized HCD collision energy. *Orbitrap DIA*. The DIA mode consists of one MS1 full scan followed by 44 MS2 windows with variable widths (Appendix Table S2) (Steger *et al*, 2021). Each MS1 full scan (300–1,650 m/z) was conducted at 120,000 resolving power, 300% normalized AGC target, and 100 ms maximum injection time. Each MS2 scan was conducted at 30,000 resolving power, 3,000% normalized AGC target, and normalized HCD collision energy in stepped mode of 25%, 27.5%, and 30%. The default charge was 3 and the RF lens was set to 40%. *timsTOF dia-PASEF*. For the dia-PASEF measurements on the Bruker timsTOF HT platform, one MS1 full scan was followed by 20 dia-PASEF scans with variable widths that were optimized for the precursor densities of tryptic HeLa digests using py_diAID (Skowronek *et al*, 2022b). The method covers an m/z range from 300 to 1,200 with two IM windows per dia-PASEF scan ranging from 0.3 to 1.7 Vs $cm^{-2}$ (Appendix Fig S3A, Appendix Table S3). Since the MS1 scan and each dia-PASEF scan measures 100 ms, the total cycle time for this method is 2.1 s. The collision energy is a linear ramp from 20 eV at 1/K0 = 0.6 Vs $cm^{-2}$ to 59 eV at 1/K0 = 1.6 Vs $cm^{-2}$.

*Mixed species experiment (human/yeast/E. coli)*

The mixed species experiment was performed using the same gradient as the one used for tryptic HeLa digests. They were mainly measured in dia-PASEF mode on the Bruker timsTOF HT platform using two methods, hereby referred to as MS2-centric method (20S) and MS1-centric method (16S4MS1). The MS2-centric method has the same settings as the one described for tryptic HeLa digests where one MS1 scan is followed by 20 dia-PASEF scans (Appendix Fig S3A). The MS1-centric method was also optimized using py_diAID to cover an m/z range from 300 to 1,200 with two IM windows per dia-PASEF scan ranging from 0.3 to 1.7 Vs $cm^{-2}$ (Appendix Fig S3B). In this method, there are 4 × (1 MS1 scan + 4 dia-PASEF scans) that also leads to a total cycle time of 2.0 s (Appendix Fig S3B, Appendix Tables S4 and S5). The collision energy is a linear ramp from 20 eV at 1/K0 = 0.6 Vs $cm^{-2}$ to 59 eV at 1/K0 = 1.6 Vs $cm^{-2}$.

*Lys-N-derived HeLa*

All Lys-N-derived HeLa measurements were acquired on the Bruker timsTOF HT using the 75-min gradient used for the tryptic HeLa digests. They were mainly measured in dia-PASEF mode using an MS2-centric method (20S) optimized for Lys-N-derived HeLa peptides (Appendix Fig S3C, Appendix Table S6). The 5-plex accuracy experiment was measured on the EvoSep One system using the 30SPD method (44 min active gradient) on a 15 cm × 150 µm column with 1.9 µm C18-beads (PepSep) at 40°C. The analytical column were connected to a fused silica ID emitter (10 µm ID; Bruker Daltonics).

***LC–MS/MS analysis of ultra-high sensitivity and single-cell data***

All samples were loaded onto Evotips Pure and measured with the Evosep One LC system (EvoSep) coupled to a timsTOF SCP mass spectrometer (Bruker). The Whisper40 SPD (samples per day) method was used with the Aurora Elite CSI third generation column with 15 and 75 µm ID (AUR3-15075C18-CSI, IonOpticks) at 50°C inside a nanoelectrospray ion source (Captive spray source, Bruker). The mobile phases comprised 0.1% FA in LC–MS-grade water as buffer A and 99.9% ACN/0.1% FA as buffer B. The timsTOF SCP was operated in dia-PASEF mode with variable window widths. Optimal dia-PASEF methods cover the precursor cloud highly efficient in the m/z – ion mobility (IM) plane while providing deep proteome coverage. For method generation with py_diAID, the precursor density distribution in m/z and IM was estimated based on a tryptic 48 high-pH fraction library (Skowronek *et al*, 2022b). We calculated the optimal cycle time based on the chromatographic peak width of 5 ng HeLa single runs. DIA-NN reported a base-to-base peak width of 7.6 s, translating into eight data points per peak at a cycle time of 0.96 s. The optimal dia-PASEF method consisted of one MS1 scan followed by eight dia-PASEF scans with two IM ramps per dia-PASEF scan, covering a m/z range from 300 to 1,200 and IM of 0.7 to 1.3 Vs $cm^{-2}$ (Appendix Fig S3D, Appendix Table S7). The mass spectrometer was operated in high sensitivity mode, the accumulation and ramp time was specified as 100 ms, capillary voltage was set to 1,400 V and the collision energy was a linear ramp from 20 eV at $1/K_0$ = 0.6 Vs $cm^{-2}$ to 59 eV at $1/K_0$ = 1.6 Vs $cm^{-2}$.

For the mDIA-DVP (mDVP) experiment, a timsTOF Ultra mass spectrometer (Bruker) was used in combination with the Whisper20 SPD method. The optimal dia-PASEF method consisted of one MS1 scan followed by twelve dia-PASEF scans with two IM ramps per dia-PASEF scan, covering a m/z range from 350 to 1,200 and IM of 0.7 to 1.3 Vs $cm^{-2}$ (Appendix Table S8). All other parameters were kept the same.

The label efficiency of 1 ng HeLa peptides was assessed on the timsTOF SCP operating in dda-PASEF mode with ten PASEF/MSMS scans per topN acquisition cycle. Singly charged precursors were excluded by their position in the m/z-IM plane using a polygon shape, and precursor signals over an intensity threshold of 1,000 arbitrary units were picked for fragmentation. Precursors were isolated with a 2 Th window below m/z 700 and 3 Th above, as well as actively excluded for 0.4 min when reaching a target intensity of

 © 2023 The Authors

20,000 arbitrary units. All spectra were acquired within a m/z range of 100–1,700. All other settings were as described above.

***Raw data analysis with MaxQuant for label efficiency assessment***

MaxQuant version 2.0.1 was used to process up to quintuplicate DDA data for calculating the labeling efficiency of each dimethyl channel. All dimethyl groups (Var DimethNter0, Var DimethLys0, Var DimethNter2, Var DimethLys2, Var DimethNter4, Var DimethLys4, Var DimethNter6, Var DimethLys6, Var DimethNter8, and Var DimethLys8) were first configured as variable modifications in MaxQuant. They were selected as variable modifications together with Oxidation (M) and Acetyl (Protein-N-term). For example, BSA samples labeled with dimethyl Δ0 were searched with Var DimethNter0, Var DimethLys0, Oxidation (M), and Acetyl (Protein-N-term) as variable modifications. Carbamidomethyl (C) was selected as a fixed modification for bulk but unselected for ultra-high sensitivity acquisitions, since peptides were not reduced and alkylated. The maximum number of modifications per peptide was set to 3. Depending on the sample, either trypsin/P or Lys-N was selected as the protease and the maximum number of missed cleavages was set to 2. Labeling efficiency was calculated based on intensity ratios of labeled peptides relative to all detected peptides.

***Raw data analysis with DIA-NN***

DIA-NN version 1.8.1 was used to search DIA raw files and diaPASEF files for precursor and fragment identifications based on 2D or 3D peak position (retention time, m/z precursor, and IM) using AlphaPeptDeep-predicted spectral libraries (*dimethyl labeled tryptic HeLa: 6331933 precursor entries*, *label-free tryptic HeLa: 6341464 precursor entries*, *dimethyl labeled tryptic Human/Yeast/E.Coli mix: 9716057 precursor entries*, *dimethyl labeled Lys-N-derived HeLa: 3387578 precursor entries*, *dimethyl labeled tryptic HeLa for ultra-high sensitivity samples: 6294089 precursor entries*).

The DIA-NN search included the following settings: Protein inference = 'Genes', Neural network classifier = 'Single-pass mode', Quantification strategy = 'Robust LC (high precision)', Cross-run normalization = 'RT-dependent', Library Generation = 'IDs, RT and IM Profiling', and Speed and RAM usage = 'Optimal results'. Mass accuracy and MS1 accuracy were set to 0 for automatic inference from the first run. The following settings were also enabled: 'Use isotopologues', 'MBR', 'Heuristic protein inference', and 'No shared spectra'. For dimethyl labeled BSA and tryptic HeLa samples, the following additional commands were entered into the DIA-NN command line GUI: *(1)* {--fixed-mod Dimethyl, 28.0313, nK}, *(2)* {--channels Dimethyl, 0, nK, 0:0; Dimethyl, 4, nK, 4.0251:4.0251; Dimethyl, 8, nK, 8.0444:8.0444}, *(3)* {--original-mods}, *(4)* {--peak-translation}, *(5)* {--ms1-isotope-quant}, *(6)* {--report-lib-info}, and *(7)* {-mass-acc-quant 10.0}. Note that *(7)* is removed from the command line for the MS1-centric method. For dimethyl labeled Lys-N-derived HeLa samples, additional channels were inserted into the DIA-NN command line GUI: {--channels Dimethyl, 0, nK, 0:0; Dimethyl, 2, nK, 2.0126:2.0126; Dimethyl, 4, nK, 4.0251:4.0251; Dimethyl, 6, nK, 6.0377:6.0377; Dimethyl, 8, nK, 8.0444:8.0444}. For label-free HeLa tryptic/Lys-N samples, only the following commands were used: *(1)* {--original-mods}, *(2)* {--ms1-isotope-quant}, *(3)* {--report-lib-info} and *(4)* {-mass-acc-quant 10.0}. In cases where only one channel or two channels were present, the actual number of channels were inserted into the DIA-NN command line GUI and not the full set of channels. We used reannotate with the fasta SwissProt database of reviewed sequences (April, 2022).

Ultra-high sensitivity measurements were searched in the same way as tryptic HeLa samples, except that the scan window was always set to a fixed value of 9, as well as MS1 and mass accuracy to 15 ppm, as this was identified as the optimal settings after several runs and kept constant for minimizing variance between the datasets. The optimal settings were evaluated by an independent search with 'unrelated runs' enabled. In cases of empty channels, the full set of channels were still added to the DIA-NN command line GUI. For the single-cell dataset and mDIA-DVP, a combined library approach was used to speed up the searching time. A subset of the single-cell dataset (50 raw files) or DIA runs (50 ng) of the reference channel sample were searched with same settings against the AlphaPeptDeep-predicted library. The generated report-lib library (file name: 'report-lib.tsv', from 'first-pass' search) was then used to analyze all raw files. The report-lib library was adjusted by renaming the 'Protein.Group' column to 'UniprotID' and searched without the 'reannotate' function in DIA-NN to keep the same protein grouping. The first-pass-result of this search was used for data analysis, since a MBR cycle was already performed in the report-lib library generation.

***AlphaPeptDeep library prediction***

Libraries were predicted by AlphaPeptDeep using the same fasta file as described above. Predefined standard models of MS2, RT and CCS were refined by transfer learning using either MaxQuant, AlphaPept or DIA-NN output files (refined models are uploaded to PRIDE and MassIVE, see Data availability). Peptide labeling was set to 'Dimethyl@K' and 'Dimethyl@Any N-term' with a mass of 28.0313 (Δ0). Please note that the library only needs to be predicted with one channel (in our case Δ0, as used for the reference channel), since DIA-NN creates the library for the other channels on its own by the added commands. All other settings were used as standard. In short, 'None' was used as decoy method and trypsin was selected as protease. A maximum missed cleavage of 1 was allowed. 'Carbamidomethyl@C' was set as fixed modification (where appropriate) and 'Oxidation@M' as variable modification, with a maximum of one variable modification. The peptide length was set from 7 to 30, the precursor charge from 1 to 4 and the precursor m/z from 200 to 1800. b- and y-ions were used as fragment types with a maximum fragment charge of 2. The top 12 fragments per precursor were kept from 200 to 2000 m/z. Predicted RT was converted to iRT and a tsv file was exported for DIA-NN search.

***Protein quantification for bulk data***

The DIA-NN R package was used to calculate the MaxLFQ abundance for protein groups (Demichev *et al*, 2019). The MaxLFQ abundance (Cox *et al*, 2014) was calculated based on the 'Precursor.Normalised' column output by DIA-NN for MS2-centric methods whereas it is based on the 'Ms1.Area' column output by DIA-NN for the MS1-centric method. For all datasets, the output results from the R package were then filtered for 'Global.PG.Q.Value' < 0.01 and 'PG.Q.Value' < 0.05. An additional filtering for 'Channel.Q.Value' < 0.01 was applied for labeled datasets and 'Q.Value' < 0.01 for label-free datasets. For the mixed species three-plex sample, the calculated protein ratios were normalized using the human protein ratios which are present in equal amounts in the different channels.

© 2023 The Authors

To control the comparison, we used intersected protein groups between the two methods ($n$ = 8,707).

#### RefQuant and its protein quantification

The filtered DIA-NN output report table was processed with the 'RefQuant' package implemented in Python. RefQuant (https://github.com/MannLabs/refquant) imports the DIA-NN table and extracts the relevant information used for quantification (quantities used: 'Fragment.Quant.Raw', 'Ms1.Area', 'Precursor.Translated', 'Precursor.Normalised'). Subsequently, on each precursor, the following operations are performed: The target quantity is divided by the reference quantity for each of the available extracted quantities. This results in a list of ratios. The ratios are sorted in ascending order and the first 40% of ratios are retained. The overall ratio R between target and reference is then obtained by taking the mean of the remaining ratios. The ratio R is then multiplied by a precursor-specific scaling factor, representing the intensity in the reference channel, which results in an overall intensity estimate of the precursor in the target channel. The scaling factor for each precursor is derived by taking the median of the reference intensity over all available runs for this particular precursor. The default value for the reference intensity is derived from the 'Ms1.Area'. In case there are remaining precursors with less than two not-null 'Ms1.Area' values, the scaling factor is derived from 'Precursor.Translated' or 'Precursor.Normalised'. Otherwise, if there are still remaining precursors with less than two not-null reference intensity values, the scaling factor is derived from the summed available ion intensities. Processing with RefQuant resulted in a quantification matrix (protein group, precursors, experiment and channel RefQuant intensity).

The RefQuant quantification matrix was then further processed with the iq package (Pham *et al*, 2020) in order to derive MaxLFQ protein quantities.

#### Proteomics downstream data analysis

Proteomics data analysis was performed in RStudio 2022.07.2 with R 4.2.2 and Python (version 3.8.2). MaxQuant output tables were filtered for 'Reverse', 'Only identified by site modification' and 'Potential contaminants' before further processing. For ultra-high sensitivity measurements, the RefQuant output was filtered for 'Lib.PG.Q.Value' < 0.01, 'Q.value' < 0.01 and 'Channel.Q.Value' < 0.15. Seven hundred and sixty-four cells were measured and 534 cells were not empty (above 1% of identified proteins). Cells in a range of 1.5 standard deviations around the median number of identified proteins were used for further analysis, except three outliers which showed a highly increased MS signal intensity compared to their proteomic depth (SM03p5b_30_5329_label4, SM03p4a_68_5210_label8, SM03p6a_65_5473_label8).

#### Comparison of single-cell protein and RNA

The SMART-Seq2 (Hu *et al*, 2019) dataset contained 720 HeLa cells with a total of 24,990 expressed genes and the Drop-seq (Schwabe *et al*, 2020) dataset measured a total of 5,665 cells and 41,161 expressed genes, both measured as three batches. Single-cell analysis was performed with scanpy v1.6.0 (Wolf *et al*, 2018) similar as before, while proteins were normalized based on the scepter 'normalize' method that iteratively adjusts the median protein expression between the different batches and different targets (Schoof *et al*, 2021). If not stated otherwise, standardized filtering across all datasets, removed cells with less than 600 genes expressed, and removed genes detected in < 15% of the remaining cells. This resulted in 10,557 transcripts in 720 cells in the SMART-Seq2 dataset and 5,022 transcripts and 6,701 cells measured with Drop-seq technology. The data completeness across covered dynamic range was computed as a function of the mean log(x + 1)-transformed protein abundance of all non-zero/-NaN entries. The expected data completeness was included based on the assumption that missing values are purely due to shot-(Poisson)-noise as 1-exp(−x). For correlation analysis, the RNA abundance entries were linearly scaled to sum to the mean cell size of the respective dataset per cell (231,281.56 for SMART-Seq2 and 7,808.12 for Drop-Seq). This was followed by log(x + 1) transformation of all abundance entries. Entries of missing protein abundance values were excluded from the specific computation. RNA expression vectors were scaled to the mean cell size of that measurement technology in the CV versus mean intensity plots comparing different technologies as well as the mean versus CV analysis (including the core proteome analysis) and the CV distribution boxplots. Mean and CV values were computed per gene under the assumption that scRNAseq data are not zero inflated (Svensson, 2020) while NaNs were excluded for the proteomics data. CV (RNA-seq) versus CV (Proteomics) plots compared CV values of proteins/genes that were shared between all datasets.

## Data availability

All mass spectrometry proteomics raw data, libraries, refined AlphaPeptDeep models for library prediction and outputs from each particular search engine analyzed in this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol *et al*, 2019), Project accession: PXD038632 (https://www.ebi.ac.uk/pride/archive?keyword=PXD038632), and additionally, via the MassIVE partner repository, Project accession: MSV000090956 (https://massive.ucsd.edu/).

The source code of RefQuant is openly available on GitHub: https://github.com/MannLabs/refquant. It will be made available through PyPI with "pip install RefQuant" as well. The code used for data analysis will be made publicly available upon publication on GitHub.

**Expanded View** for this article is available online.

### Acknowledgements

This study was supported by the Max Planck Society for Advancement of Science, European Union's Horizon 2020 research and innovation program under grant agreement No. 874839 (ISLET), the Deutsche Forschungsgemeinschaft project "Chemical proteomics inside us" (grant No.: 412136960) and by the Bavarian State Ministry of Health and Care through the research project DigiMed Bayern (www.digimed-bayern.de). MT and MW acknowledge support from the International Max Planck Research School for Life Sciences – IMPRS-LS. SR is supported by the Helmholtz Association under the joint research school "Munich School for Data Science" – MUDS. FAR is an EMBO postdoctoral fellow (ALTF 399-2021). We are grateful for the FACS support by the Imaging Core Facility at the Max Planck Institute of Biochemistry, in particular Martin Spitaler and Markus Oster. We thank our colleagues in the Department of Proteomics and Signal Transduction, Max

 © 2023 The Authors

Planck Institute of Biochemistry and at the Center for Protein Research at Copenhagen University, for discussions and support. In particular, we thank Jakob Bader for DIA-NN library discussions, Igor Paron and Tim Heymann for technical support and column production and Medini Steger for scientific administration support. Open Access funding enabled and organized by Projekt DEAL.

### Author contributions

**Matthias Mann:** Conceptualization; formal analysis; supervision; funding acquisition; writing – original draft; project administration; writing – review and editing. **Marvin Thielert:** Conceptualization; data curation; software; formal analysis; supervision; validation; investigation; visualization; methodology; writing – original draft; project administration; writing – review and editing. **Ericka CM Itang:** Formal analysis; validation; investigation; writing – original draft; writing – review and editing. **Constantin Ammar:** Software; formal analysis; visualization. **Florian A Rosenberger:** Formal analysis; investigation. **Isabell Bludau:** Formal analysis; validation. **Lisa Schweizer:** Validation; investigation. **Thierry M Nordmann:** Formal analysis; investigation. **Patricia Skowronek:** Data curation; formal analysis. **Maria Wahle:** Formal analysis; validation; investigation. **Wen-Feng Zeng:** Software. **Xie-Xuan Zhou:** Software. **Andreas-David Brunner:** Formal analysis. **Sabrina Richter:** Formal analysis; visualization; methodology. **Mitchell P Levesque:** Resources; formal analysis. **Fabian J Theis:** Formal analysis; supervision; visualization; methodology. **Martin Steger:** Conceptualization; formal analysis; supervision; investigation; writing – original draft; writing – review and editing.

### Disclosure and competing interests statement

MM is an indirect shareholder in EvoSep Biosystems. All other authors have no relevant competing interests. MM and FJT are editorial advisory board members. This has no bearing on the editorial consideration of this article for publication.

## References

Bache N, Geyer PE, Bekker-Jensen DB, Hoerning O, Falkenby L, Treit PV, Doll S, Paron I, Müller JB, Meier F *et al* (2018) A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Mol Cell Proteomics* 17: 2284 – 2296

Bader JM, Albrecht V, Mann M (2023) MS-based proteomics of body fluids: the end of the beginning. *Mol Cell Proteomics* 22: 100577

Bekker-Jensen DB, Kelstrup CD, Batth TS, Larsen SC, Haldrup C, Bramsen JB, Sørensen KD, Høyer S, Ørntoft TF, Andersen CL *et al* (2017) An optimized shotgun strategy for the rapid generation of comprehensive human proteomes. *Cell Syst* 4: 587 – 599

Belote RL, Le D, Maynard A, Lang UE, Sinclair A, Lohman BK, Planells-Palop V, Baskin L, Tward AD, Darmanis S *et al* (2021) Human melanocyte development and melanoma dedifferentiation at single-cell resolution. *Nat Cell Biol* 239: 1035 – 1047

Biron-Pain K, Grosset AA, Poirier F, Gaboury L, St-Pierre Y (2013) Expression and functions of galectin-7 in human and murine melanomas. *PloS One* 8: e63307

Boersema PJ, Aye TT, Van Veen TAB, Heck AJR, Mohammed S (2008) Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. *Proteomics* 8: 4624 – 4632

Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJR (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* 4: 484 – 494

Brunner A-D, Thielert M, Vasilopoulou C, Ammar C, Coscia F, Mund A, Hoerning OB, Bache N, Apalategui A, Lubeck M *et al* (2022) Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol Syst Biol* 18: e10798

Budnik B, Levy E, Harmange G, Slavov N (2018) SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* 19: 161

Cheung TK, Lee CY, Bayer FP, McCoy A, Kuster B, Rose CM (2020) Defining the carrier proteome limit for single-cell proteomics. *Nat Methods* 181: 76 – 83

Cong Y, Motamedchaboki K, Misal SA, Liang Y, Guise AJ, Truong T, Huguet R, Plowey ED, Zhu Y, Lopez-Ferrer D *et al* (2021) Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. *Chem Sci* 12: 1001 – 1006

Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 13: 2513 – 2526

Ctortecka C, Krššáková G, Stejskal K, Penninger JM, Mendjan S, Mechtler K, Stadlmann J (2022a) Comparative proteome signatures of trace samples by multiplexed data-independent acquisition. *Mol Cell Proteomics* 21: 100177

Ctortecka C, Stejskal K, Krššáková G, Mendjan S, Mechtler K (2022b) Quantitative accuracy and precision in multiplexed single-cell proteomics. *Anal Chem* 94: 2434 – 2443

Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M (2019) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* 17: 41 – 44

Demichev V, Szyrwiel L, Yu F, Teo GC, Rosenberger G, Niewienda A, Ludwig D, Decker J, Kaspar-Schoenefeld S, Lilley KS *et al* (2022) dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nat Commun* 131: 1 – 8

Derks J, Leduc A, Wallmann G, Huffman RG, Willetts M, Khan S, Specht H, Ralser M, Demichev V, Slavov N (2022) Increasing the throughput of sensitive proteomics by plexDIA. *Nat Biotechnol* 41: 50 – 59

Eraslan G, Drokhlyansky E, Anand S, Fiskin E, Subramanian A, Slyper M, Wang J, Van Wittenberghe N, Rouhana JM, Waldman J *et al* (2022) Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science* 376: eabl4290

Geiger T, Cox J, Ostasiewicz P, Wisniewski JR, Mann M (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Methods* 75: 383 – 385

Gerami P, Cook RW, Wilkinson J, Russell MC, Dhillon N, Amaria RN, Gonzalez R, Lyle S, Johnson CE, Oelschlager KM *et al* (2015) Development of a prognostic genetic signature to predict the metastatic risk associated with cutaneous melanoma. *Clin Cancer Res* 21: 175 – 183

Griffiths JR, Chicooree N, Connolly Y, Neffling M, Lane CS, Knapman T, Smith DL (2014) Mass spectral enhanced detection of ubls using SWATH acquisition: MEDUSA-simultaneous quantification of SUMO and Ubiquitin-derived isopeptides. *J Am Soc Mass Spectrom* 25: 767 – 777

Han W, Hu C, Fan ZJ, Shen GL (2021) Transcript levels of keratin 1/5/6/14/15/16/17 as potential prognostic indicators in melanoma patients. *Sci Rep* 11: 1023

© 2023 The Authors

Hartlmayr D, Ctortecka C, Seth A, Mendjan S, Tourniaire G, Mechtler K, Biocenter V (2021) An automated workflow for label-free and multiplexed single cell proteomics sample preparation at unprecedented sensitivity. *bioRxiv* https://doi.org/10.1101/2021.04.14.439828 [PREPRINT]

Hennrich ML, Mohammed S, Altelaar AFM, Heck AJR (2010) Dimethyl isotope labeling assisted *de novo* peptide sequencing. *J Am Soc Mass Spectrom* 21: 1957 – 1965

Hsu JL, Huang SY, Chow NH, Chen SH (2003) Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* 75: 6843 – 6852

Hu W, Zhang X, Guo Q, Yang J, Yang Y, Wei S, Su X (2019) HeLa-CCL2 cell heterogeneity studied by single-cell DNA and RNA sequencing. *PloS One* 14: e0225466

Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT (2018) The human transcription factors. *Cell* 172: 650 – 665

Liang Y, Acor H, Mccown MA, Nwosu AJ, Boekweg H, Axtell NB, Truong T, Cong Y, Payne SH, Kelly RT (2020) Fully automated sample processing and analysis workflow for low-input proteome profiling. *Anal Chem* 93: 1658 – 1666

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM *et al* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161: 1202 – 1214

Matzinger M, Müller E, Dürnberger G, Pichler P, Mechtler K (2023) Robust and easy-to-use one-pot workflow for label-free single-cell proteomics. *Anal Chem* 95: 4435 – 4445

Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, Goedecke N, Decker J, Kosinski T, Park MA *et al* (2018) Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol Cell Proteomics* 17: 2534 – 2545

Meier F, Brunner AD, Frank M, Ha A, Bludau I, Voytik E, Kaspar-Schoenefeld S, Lubeck M, Raether O, Bache N *et al* (2020) diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nat Methods* 17: 1229 – 1236

Mund A, Coscia F, Kriston A, Hollandi R, Kovács F, Brunner AD, Migh E, Schweizer L, Santos A, Bzorek M *et al* (2022) Deep visual proteomics defines single-cell identity and heterogeneity. *Nat Biotechnol* 408: 1231 – 1240

Nordmann TM, Schweizer L, Metousis A, Thielert M, Rodriguez E, Mette Rahbek-Gjerdrum L, Stadler P-C, Bzorek M, Mund A, Rosenberger FA *et al* (2023) A standardized and reproducible workflow for membrane glass slides in routine histology and spatial proteomics. *bioRxiv* https://doi.org/10.1101/2023.02.20.529255 [PREPRINT]

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M *et al* (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47: D442 – D450

Pham TV, Henneman AA, Jimenez CR (2020) iq: an R package to estimate relative protein abundances from ion quantification in DIA-MS-based proteomics. *Bioinformatics* 36: 2611 – 2613

Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9: 171 – 181

Pino LK, Baeza J, Lauman R, Schilling B, Garcia BA (2021) Improved SILAC quantification with data independent acquisition to investigate bortezomib-induced protein degradation. *J Proteome Res* 20: 1918 – 1927

Raaijmakers MAJ, Smidts DP, Sergeant JA, Maassen GH, Posthumus JA, Van Engeland H, Matthys W (2008) Executive functions in preschool children with aggressive behavior: impairments in inhibitory control. *J Abnorm Child Psychol* 36: 1097 – 1107

Raijmakers R, Neerincx P, Mohammed S, Heck AJR (2010) Cleavage specificities of the brother and sister proteases Lys-C and Lys-N. *Chem Commun* 46: 8827 – 8829

Rappsilber J, Ishihama Y, Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* 75: 663 – 670

Rosenberger FA, Thielert M, Strauss MT, Ammar C, Mädler SC, Schweizer L, Metousis A, Skowronek P, Wahle M, Gote-Schniering J *et al* (2022) Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome. *bioRxiv* https://doi.org/10.1101/2022.12.03.518957 [PREPRINT]

Saha-Shah A, Esmaeili M, Sidoli S, Hwang H, Yang J, Klein PS, Garcia BA (2019) Single cell proteomics by data-independent acquisition to study embryonic asymmetry in *Xenopus laevis*. *Anal Chem* 91: 8891 – 8899

Sandow JJ, Infusini G, Dagley LF, Larsen R, Webb AI, Hall E (2021) Simplified high-throughput methods for deep proteome analysis on the timsTOF Pro. *bioRxiv* https://doi.org/10.1101/657908 [PREPRINT]

Schoof EM, Furtwängler B, Üresin N, Rapin N, Savickas S, Gentil C, Lechman E, Keller UAD, Dick JE, Porse BT (2021) Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nat Commun* 12: 3341

Schwabe D, Formichetti S, Junker JP, Falcke M, Rajewsky N (2020) The transcriptome dynamics of single cells during the cell cycle. *Mol Syst Biol* 16: e9946

Skowronek P, Krohs F, Lubeck M, Wallmann G, Itang ECM, Koval P, Wahle M, Thielert M, Meier F, Willems S *et al* (2022a) Synchro-PASEF allows precursor-specific fragment ion extraction and interference removal in data-independent acquisition. *Mol Cell Proteomics* 22: 100489

Skowronek P, Thielert M, Voytik E, Tanzer MC, Hansen FM, Willems S, Karayel Ö, Brunner A-D, Meier F, Mann M (2022b) Rapid and in-depth coverage of the (phospho-)proteome with deep libraries and optimal window design for dia-PASEF. *Mol Cell Proteomics* 21: 100279

Specht H, Emmott E, Petelski AA, Huffman RG, Perlman DH, Serra M, Kharchenko P, Koller A, Slavov N (2021) Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol* 22: 50

Steger M, Demichev V, Backman M, Ohmayer U, Ihmor P, Müller S, Ralser M, Daub H (2021) Time-resolved *in vivo* ubiquitinome profiling by DIA-MS reveals USP7 targets on a proteome-wide scale. *Nat Commun* 121: 1 – 13

Suo C, Dann E, Goh I, Jardine L, Kleshchevnikov V, Park JE, Botting RA, Stephenson E, Engelbert J, Tuong ZK *et al* (2022) Mapping the developing human immune system across organs. *Science* 376: eabo0510

Svensson V (2020) Droplet scRNA-seq is not zero-inflated. *Nat Biotechnol* 38: 147 – 150

Tabula Sapiens Consortium, Jones RC, Karkanias J, Krasnow MA, Pisco AO, Quake SR, Salzman J, Yosef N, Bulthaup B, Brown P *et al* (2022) The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science* 376: eabl4896

Taouatas N, Drugan MM, Heck AJR, Mohammed S (2008) Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat Methods* 55: 405 – 407

Taouatas N, Heck AJR, Mohammed S (2010) Evaluation of metalloendopeptidase Lys-N protease performance under different sample handling conditions. *J Proteome Res* 9: 4282 – 4288

Tian X, de Vries MP, Permentier HP, Bischoff R (2020) A versatile isobaric tag enables proteome quantification in data-dependent and data-independent acquisition modes. *Anal Chem* 92: 16149 – 16157

Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 19: 15

 © 2023 The Authors

Wu Y, Wang F, Liu Z, Qin H, Song C, Huang J, Bian Y, Wei X, Donga J, Zou H (2014) Five-plex isotope dimethyl labeling for quantitative proteomics. *Chem Commun* 50: 1708 – 1710

Xuan Y, Bateman NW, Gallien S, Goetze S, Zhou Y, Navarro P, Hu M, Parikh N, Hood BL, Conrads KA *et al* (2020) Standardization and harmonization of distributed multi-center proteotype analysis supporting precision medicine studies. *Nat Commun* 11: 5248

Ye Z, Batth TS, Rüther P, Olsen JV (2022) A deeper look at carrier proteome effects for single-cell proteomics. *Commun Biol* 5: 150

Zeng W-F, Zhou X-X, Willems S, Ammar C, Wahle M, Bludau I, Voytik E, Strauss MT, Mann M (2022) AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics. *Nat Commun* 131: 1 – 14

Zhu Y, Piehowski PD, Zhao R, Chen J, Shen Y, Moore RJ, Shukla AK, Petyuk VA, Campbell-Thompson M, Mathews CE *et al* (2018) Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat Commun* 91: 1 – 10

Zhu Y, Scheibinger M, Ellwanger DC, Krey JF, Choi D, Kelly RT, Heller S, Barr-Gillespie PG (2019) Single-cell proteomics reveals changes in expression during hair-cell development. *Elife* 8: e50777

**License:** This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

**Article 5. Making single-cell proteomics biologically relevant.**

*Nature Methods*, 2023 (20, 320-323); DOI: 10.1038/s41592-023-01771-9

Marvin Thielert[1,#], Florian A. Rosenberger[1,#], Matthias Mann[1,3]

*[1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [3]Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.*
*[#]These authors contributed equally: Florian A. Rosenberger, Marvin Thielert.*

Following the advances of sc-proteomics, I shared my experience and perspective in a brief outline of advances and future directions of sc-proteomics to biology.

An important aspect in sc-proteomics is the size of the cell, since the protein content highly correlated with the volume of a cell. Typically, the average cell contains about 250 ng of proteins but there is a large range from very tiny cells like T cells with 20 pg to cardiomyocytes or huge oocytes (factor 200,000). In the paper we make the case that with increasing proteomic depth and throughput, biological applications will follow. In our experience, 5,000 protein groups cover a multitude of biological pathways and protein classes that allow characterization of functional cell states. Additionally, we note that techniques for analyzing just the cells of interest without other cells enhances the observed biological effects. In evaluating the throughput, we propose that proteomics will be able to draw biological conclusions from fewer cell numbers compared to scRNA-seq, because of its nature of a complete and stable proteome leading to lower variances in proteome measurement compared to the stochastic transcriptome.

Several techniques exist for preparing single-cell samples. Single cells can be disaggregated from any source such as cell culture or tissue and sorted into individual reaction wells. Cells may also be selected based on fluorescent features. Further, LCM is a technique to preserve the spatial information. Building on this approach, we developed a workflow called scDVP (**Article 7)**, which enables the analysis of single cells in the spatial context and is particularly suited to the strengths of proteomics.

We expect that spatial sc-proteomics will greatly aid in studying proteomic changes of individual cells, for example, along various metabolic microenvironment gradients such as oxygen and nutrients. This technique could further offer understanding into the space directly adjacent to tumor cells or between metabolically cooperating cells.

***Contribution:***

Shared first-authorship. Matthias Mann, Florian Rosenberger and I wrote the paper. I contributed with my ideas and perspective to the manuscript. I designed the figures together with Florian Rosenberger and the help of a graphical illustrator.

Comment

https://doi.org/10.1038/s41592-023-01771-9

# Making single-cell proteomics biologically relevant

Florian A. Rosenberger, Marvin Thielert & Matthias Mann

Check for updates

Recent technological advances in mass spectrometry promise to add single-cell proteomics to the biologist's toolbox. Here we discuss the current status and what is needed for this exciting technology to lead to biological insight – alone or as a complement to other omics technologies.

As biological researchers, we want to identify rules that characterize a given system. In the past, much of this work has taken place either in vitro in cell culture or in vivo in whole animals or tissues. However, surprisingly often those discoveries hold true only in the chosen experimental conditions. The nascent field of single-cell omics captures the wide range of potential behaviors with various metabolic microenvironments, cell types, and (epi)genetic states. Tremendous developments in RNA sequencing (RNA-seq) technologies have enabled robust measurement of cellular transcriptomes of thousands of single cells[1]. This has recently been extended into the spatial dimension and, increasingly, to other genomic readouts[2]. As proteins are in most cases the biological actors in cellular life, an unbiased proteomic dimension would be a logical next addition to the omics arsenal of single-cell biology. However, apart from antibody-based technologies, which typically only cover a small number of predefined protein targets, mass spectrometry (MS)-based proteomics has so far been hampered by enormous technological challenges, mainly in sensitivity and throughput.

## Technological challenges in single-cell proteomics

The task of single-cell MS-based proteomics (sc-proteomics) is to identify and accurately quantify as many different proteins as possible across the vast dynamic range of the proteome, where proteins are expressed at copy numbers from dozens to a few millions per cell. Total protein material is in the few hundred picograms range for typical cells but can even be less than 20 pg for resting T cells (Fig. 1a)[3]. As there is no equivalent to oligonucleotide amplification, the quest is to develop the most sensitive sample preparation and analytical pipeline to obtain quantitative and reproducible data that reveal the relevant biology.

The amount of input material required for a biological proteomics experiment has exponentially decreased since the dawn of MS-based proteomics, from the milligrams range that was common until a few years ago to the nanogram range. The development of techniques that are starting to allow quantification of single-cell proteomes is a natural extension of this Moore's law-like trajectory. In fact, the number of proteins identified in one single cell in culture has itself increased from less than a hundred to more than a thousand in just a few years (Fig. 1b). This was made possible by efforts of various groups in four key technological areas: (1) sample preparation, (2) chromatography, (3) mass spectrometry, and (4) bioinformatic data acquisition and processing.

Inspired by single-cell RNA-seq (scRNA-seq) developments, various routes have been explored for lossless sample preparation from the intact single cell to enzymatically digested peptides ready for analysis by liquid chromatography–mass spectrometry (LC–MS), with the common theme of miniaturization to optimize reagent-to-protein ratios. A pioneering example is 'nanodroplet processing in one pot for trace samples' (nanoPOTS), in which nanoliter reactions are carried out on a glass chip in a humidified chamber[4]. In our group, we use 384-well plastic plates coated to reduce peptide adhesion without compromising peptide identification downstream[5].This format makes sc-proteomics sample preparation versatile and accessible in typical laboratory settings, an important feature for widespread use of sc-proteomics technologies upstream of specialized core facility instrumentation in the future.

Peptide mixtures are then injected onto a liquid chromatographic column for online LC–MS analysis. Flow rates correlate inversely with electrospray efficiency and sensitivity, making ultra-low flow rates and extremely narrow column diameters attractive[6]. However, this often comes with the disadvantage of broader elution peaks and diminished robustness of operation, which can potentially be avoided by photolithographically patterned columns termed µPAC[7,8] or commercial pulled emitter columns[9]. We recently described an automated system in which single-cell samples are injected in 'nano-packages' into an LC system with preformed gradients that are eluted at very low flow rates, increasing sensitivity tenfold while maintaining routine operation[5].

The most spectacular gains in sensitivity have come from the mass spectrometers themselves – for instance, by admitting a much larger proportion of ions into the instruments than before. Furthermore, new signal processing algorithms have increased mass spectrometric resolution during the past years on the widely used Orbitrap analyzers. In parallel, trapped ion mobility spectrometry (TIMS) coupled to time of flight (TOF) analyzers has become a viable contender for ultra-high sensitivity analysis[5]. TIMS introduces an extra dimension by separating peptides in an electrical field opposing the drag force of the nitrogen gas flow carrying the ions into the instrument. It allows the controlled release of selected ion mobility packages in a millisecond time scale perfectly placed between the second-wide chromatography peaks and the microsecond TOF scans. While most single-cell work has been done on the Orbitrap systems, in our laboratory the timsTOF currently has a sensitivity and scan speed advantage.

A milestone in proteomics in general and sc-proteomics in particular has been the transition from data-dependent acquisition (DDA) schemes, in which the top *n* mass precursors at a given elution time are fragmented, to data-independent acquisition (DIA), in which the entire mass range is divided into segments that the mass spectrometer cycles through for fragmentation[10]. This inherently retains fragment

# Comment

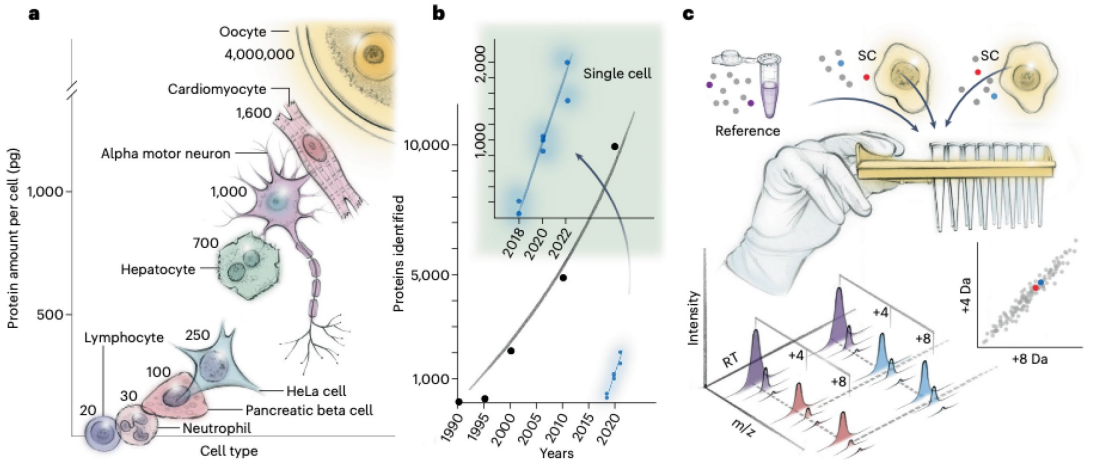



**Fig. 1 | The dawn of single-cell proteomics. a**, Cell types and their expected protein content. Numbers are estimates from cellular volume[3]. **b**, Historic perspective of protein depth in non-input-limited samples (estimates) versus single cells (green inset). **c**, Multiplexed data-independent acquisition (mDIA) allows use of a standard, high-input reference proteome to decouple identification and quantification, improving proteome depth and quantitative accuracy. The reference proteome and two single cells (SCs) are dimethyl labeled with a mass-to-charge ($m/z$) shift between channels of +0, +4 and +8 Da. RT, retention time.

information on all precursor ions over the LC elution peak – basically the entire proteome within the dynamic range of the instrument, which needs to be extracted from quite convoluted fragmentation spectra[11]. This data analysis is at the core of future computational proteomics developments that are needed to deal with the complexity of raw data in a time-efficient manner. Open source and open data policies are especially important here, as a dataset recorded in DIA mode needs to be matched to a library with consistency and transparency across different software packages.

## Biological requirements and emerging technological solutions

While most of the building blocks for sc-proteomics are in place, applications to date have mostly been technological proofs of principle. In our opinion, this is mainly due to limited depth of quantitative coverage achieved so far and also due to the limited throughput. Nearly complete proteome coverage in a given human cell type requires the identification of at least 12,000 to 13,000 different proteins (ignoring most protein isoforms for now)[12]. Such deep proteome coverage is difficult to achieve by LC–MS even with extensive fractionation of bulk samples, whereas a reasonably comprehensive human proteome of 10,000 proteins can be reached in a single run today in specialized laboratories[13] (Fig. 1b). To capture the biological state of a single cell, the currently reported numbers of around 1,000 proteins are clearly on the low side, and we believe that about 5,000 to 6,000 proteins would be a good number to strive for. In our experience such a proteome depth covers a multitude of pathways and protein classes that allow reconstruction of much of the functional state of the cell when measured with quantitative accuracy. Furthermore, by its nature sc-proteomics avoids the dilution of cell types of interest in a large background of unrelated cells that is intrinsic to bulk proteomics, allowing the biological phenotype to emerge more clearly ('biological enrichment').

Continuous developments across all the four key areas mentioned above will continue to improve sc-proteomics, perhaps on a similar trajectory as before. An immediate and promising avenue for greater depth is the decoupling of identification and quantification. This idea has already been introduced in the form of SCoPE-MS, an isobaric isotope-based method in which a 'booster channel' with the equivalent of hundreds of cells is used to mark the peptide location for DDA fragmentation[14]. However, the isobaric channels are intrinsically linked, leading to unreliable quantification at such high ratios, and the use of DDA limits comparability across many samples[15,16]. In contrast, multiplexed DIA (mDIA or plexDIA) features non-isobaric channels that are isolated from each other in terms of quantification[17–19] Building on this, we have doubled proteins identified and quantified in single cells by implementing the concept of a reference channel in mDIA[20]. The contours of each peptide feature in the reference channel are transferred onto single-cell channels (target channels) by matching retention time, ion mobility and mass shift, yielding the precise area in which signals need to be integrated for quantification (Fig. 1c). This method decouples identification and quantification, increasing the number of quantified peptides and proteins and greatly enhancing reproducibility and robustness of DIA-based sc-proteomics. Across hundreds of single cells, we currently quantify up to 4,000 proteins, with robust and reproducible quantification of 2,400 of them.

Another key issue is the throughput achieved, which will inevitably be much lower than in scRNA-seq due to the absence of barcoding.

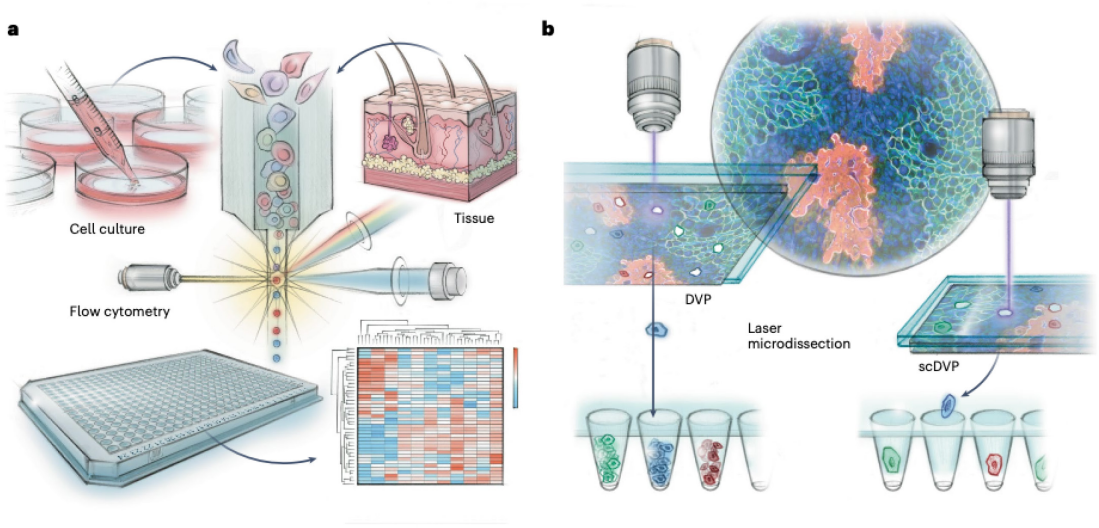



**Fig. 2 | Approaching biology with sc-proteomics. a**, Single-cell proteomics is currently performed on cells flow sorted from cell culture and could be extended to disaggregated tissue. **b**, Deep Visual Proteomics (DVP) retains spatial context of phenotypically identical microdissected single cells that are pooled from tissue at high proteome depth and throughput, while single-cell DVP (scDVP) resolves differences between individual cells.

Currently, up to five biologically independent samples can be acquired with DIA-compatible non-isobaric labels[20]. Thirty-minute gradients and fivefold multiplexing allow the measurement of about two hundred cells per day per instrument. Even shorter gradients[21] or higher multiplexing could further increase the throughput, although perhaps at the cost of proteomic depth. We recently highlighted that single cells have a complete proteome, in contrast to a stochastic transcriptome, which is due to the very low number of transcripts for most genes in single cells[5]. This leads to much lower variance of the proteome measurements, which in turn allows us to draw biological conclusions from many fewer cells — theoretically, a factor of a hundred lower for a tenfold lower variance. Thus, despite the throughput limitation, sc-proteomics can still address a large range of interesting applications.

## The spatial context in sc-proteomics

Sc-proteomics continues to generate much excitement in the proteomics community, and the field has already produced stunning technological advances with single-cell suspensions. This could be extended to disaggregated, fluorescently marked cells from tissues that are flow sorted into individual reaction wells (Fig. 2a). We believe that sc-proteomics — or close variants of it — will be most useful for biological applications when also conserving the spatial context of the cells of interest. We recently described an approach called Deep Visual Proteomics (DVP) that offers the key advantages of sc-proteomics. In DVP artificial intelligence identifies cell types after (immuno)staining of formalin-fixed, paraffin-embedded (FFPE) or frozen tissue sections, and a few hundred cells identical in appearance are automatically laser dissected for ultra-low-input proteomics of regions of interest (Fig. 2b, left panel)[22]. The equivalent of a hundred cells can readily be analyzed, which in our hands currently yields a depth of around 5,000 proteins while the throughput requirements are likewise reduced a hundred-fold compared to individual single-cell analyses. The coordinates and neighborhood information of the cells are automatically recorded, adding a rich spatial context to the proteomics results. We have found that this approach captures biological variability while averaging out the random variability in populations with the same functional attributes, offering biological enrichment.

Conversely, DVP requires cell markers and high-content imaging to accurately excise cells in the desired cell states, and it does not capture residual variability in the population of cells that are excised and analyzed together. In cases where markers of cell type and state are not initially known or there is reason to believe that there is more than one functional class in the selected population, cells can be excised individually by laser microdissection and analyzed one by one. This presents a formidable analytical challenge, first, because analysis from a tissue is more difficult than from suspension and, second, because single-cell shapes from a 2.5 to 10 µm tissue section contain only a fraction of a single cell. With this in mind, we have recently investigated the concept of single-cell DVP (scDVP; Fig. 2b, right panel) in the context of functionally and spatially distinct classes of hepatocytes in the liver context[23]. Hundreds of single-shape measurements to a median of 1,700 quantified proteins revealed that half of this proteome is under spatial control ('liver zonation').

Spatial sc-proteomics could be especially powerful when the metabolic microenvironment shapes the proteomes of individual cells. Examples include the proximity to blood vessels that generates gradients of oxygen and nutrients, the niche between metabolically

# Comment

cooperating cells, or the space directly adjacent to tumor cells. Events inside cells will be the key focus of sc-proteomics, yet many diseases are inevitably linked to the composition and accumulation of proteinaceous extracellular matrix, for instance in cancer or fibrosis. Ultra-high-sensitivity MS workflows are equipped to address this in their spatial context, which might identify new driver cell states in proximity to diseased extracellular matrix. By extension, subcellular structures or enigmatic structures connecting cells are fair game for sc-proteomics as well.

## Conclusion and outlook

Although we are only at the beginning of the sc-proteomics era, a multitude of biological questions can already be addressed from this new experimental angle. At the same time, the technological challenges in depth of proteome coverage and throughput remain substantial if biological questions other than those in very specialized applications are to be addressed. We believe that these challenges can mostly be met by a combination of improvements in MS instrumentation, measurement strategies and algorithmic developments. A depth of more than 5,000 proteins combined with a throughput of several hundred single cells per day would open up many mainstream areas. Proteomics researchers should be mindful that scRNA-seq is now well established and has high throughput, so they should seek out areas where sc-proteomics can naturally complement those capabilities. For example, the detection and quantification of novel protein isoforms is currently not realistic by most library-dependent search strategies of MS data, but could become more accessible by supplementing these with RNA-seq data.

We see immediate and exciting applications for DVP, a variant of sc-proteomics, as it provides the proteomic depth and throughput needed while adding the important spatial component to sc-proteomics. This technology seems tailor-made to study biologically meaningful differences that are hidden in bulk data and unfold only when we enrich for functionally near-identical cells. This most prominently applies to many types of cancers, in which DVP can uncover cellular heterogeneity and functional roles of stem cell niches. This comes with the requirement to predefine the cell states of interest (or have an AI do this), but recent developments such as scDVP can even bypass this. Complex tissues can be studied either with sc-proteomics alone or in concert with other omics methods. Among its unique features, sc-proteomics extends to protein complexes that are inaccurately quantified by scRNA-seq data, such as the oxidative phosphorylation system[24], unfolded protein aggregates in aging or deposited solid extracellular matrix in fibrosis.

Once the next technological steps have been taken — maybe in the course of the next few years — there will be new and unique vistas for sc-proteomics. As each single cell has a complete proteome, it will be fascinating for the future to study protein isoforms and post-translational modifications in this context. Beyond covering the steady states of cells, sc-proteomics can then be expanded to the dynamic responses to extracellular cues that propagate through signaling cascades. While this is realistic at the moment only for proteins with high copy numbers or enormous cells, the prospect of combining these aspects of the proteome with functional single-cell heterogeneity would be truly revolutionary.

**Florian A. Rosenberger** [1,3], **Marvin Thielert** [1,3] & **Matthias Mann** [1,2] ✉

[1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [2]Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [3]These authors contributed equally: Florian A. Rosenberger, Marvin Thielert.

✉e-mail: mmann@biochem.mpg.de

Published online: 10 March 2023

## References

1. Elmentaite, R., Dominguez Conde, C., Yang, L. & Teichmann, S. A. *Nat. Rev. Genet.* **23**, 395–410 (2022).
2. Deng, Y. et al. *Nature* **609**, 375–383, https://doi.org/10.1038/s41586-022-05094-1 (2022).
3. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. *Nucleic Acids Res.* **38**, D750–D753 (2010).
4. Zhu, Y. et al. *Nat. Commun.* **9**, 882 (2018).
5. Brunner, A.-D. et al. *Mol. Syst. Biol.* **18**, e10798 (2022).
6. Kelly, R. T. *Mol. Cell. Proteomics* **19**, 1739–1748 (2020).
7. Stejskal, K., Op de Beeck, J., Dürnberger, G., Jacobs, P. & Mechtler, K. *Anal. Chem.* **93**, 8704–8710 (2021).
8. Müller, J. B. et al. *Nature* **582**, 592–596 (2020).
9. Sandow, J. J., Infusini, G., Dagley, L. F., Larsen, R. & Webb, A. I. Preprint at *bioRxiv* https://doi.org/10.1101/657908 (2021).
10. Ludwig, C. et al. *Mol. Syst. Biol.* **14**, e8126 (2018).
11. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. *Nat. Methods* **17**, 41–44 (2020).
12. Bekker-Jensen, D. B. et al. *Cell Syst.* **4**, 587–599.e4 (2017).
13. Meier, F., Geyer, P. E., Virreira Winter, S., Cox, J. & Mann, M. *Nat. Methods* **15**, 440–448 (2018).
14. Budnik, B., Levy, E., Harmange, G. & Slavov, N. *Genome Biol.* **19**, 161 (2018).
15. Cheung, T. K. et al. *Nat. Methods* **18**, 76–83 (2021).
16. Ye, Z., Batth, T. S., Rüther, P. & Olsen, J. V. *Commun. Biol.* **5**, 150 (2022).
17. Minogue, C. E. et al. *Anal. Chem.* **87**, 2570–2575 (2015).
18. Derks, J. et al. *Nat. Biotechnol.* **41**, 50–59 (2023).
19. Griffiths, J. R. et al. *J. Am. Soc. Mass Spectrom.* **25**, 767–777 (2014).
20. Thielert, M. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2022.12.02.518917 (2022).
21. Messner, C. B. et al. *Nat. Biotechnol.* **39**, 846–854 (2021).
22. Mund, A. et al. *Nat. Biotechnol.* **40**, 1231–1240 (2022).
23. Rosenberger, F. A. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2022.12.03.518957 (2022).
24. Kühl, I. et al. *eLife* **6**, e30952 (2017).

## Acknowledgements

We thank the members of our single-cell and DVP teams in Munich and Copenhagen for discussions. The authors are supported by grants from the Max Planck Society for the Advancement of Science, Germany, and the European Union's Horizon 2020 research and innovation program ISLET (No. 874839). F.A.R. is an EMBO postdoctoral fellowship holder (ALTF 399-2021). We thank Juliet Percival (Percival Press, Durham, UK) for the scientific illustrations.

## Author contributions

M.M., M.T. and F.A.R. conceived the original idea. M.M., F.A.R. and M.T. wrote and edited the original draft. F.A.R., M.T. and M.M. conceptualized and guided the illustration process. All authors wrote, edited and gave final approval to the manuscript.

## Competing interests

M.M. is an indirect shareholder in Evosep Biosystems. The remaining authors declare no competing interests.

## 3.3. Spatial proteomics

Although the achievements of a sensitive workflow of sc-proteomics as label-free and multiplexed-DIA approach are significant in themselves, they only mark the beginning of a promising future. This sensitive pipeline has also demonstrated great potential for spatial proteomics and should be applicable to other intriguing fields with limitations in sensitivity such as phospho proteomics and PTMs in general as well as immunopeptidomics (discussed in **Article 11**) and peptidome approaches.

Our laboratory in Copenhagen recently developed 'Deep Visual Proteomics' (for further details, refer to 1.4.1) to enable the analysis of single-cell states and cell types in the spatial context. DVP connects high-content imaging, AI-guided laser-capture microdissection and ultra-sensitive MS. I was instrumental in the transfer of this technology to Munich, together with Lisa Schweizer. We evaluated the efficiency of DVP at our Munich location and found that the workflow needed certain optimization for reproducible staining of PEN membranes (**Article 6**). After establishing this, we combined our ultra-sensitive pipeline including mDIA with DVP to achieve single-cell resolution in spatial tissue, in a technology we termed single-cell DVP (scDVP) (**Article 7**). DVP is able to resolve two-dimensional structures of tissue slices. However, our organism and organs are three-dimensional structures. Therefore, I contributed to combining our technology with tissue clearing to achieve three-dimensional spatial resolution and the analysis in intact organs, termed DISCO-MS (**Article 8**). Further we applied DVP to interesting biological questions to show the biological relevance of the technology in ovarian cancer and cutaneous inflammatory disease (**Article 9** and **Article 10**).

## Article 6. A standardized and reproducible workflow for membrane glass slides in routine histology and spatial proteomics.

*Mol Cell Proteomics*, 2023 (22, 10, 100643); DOI: 10.1016/j.mcpro.2023.100643

Thierry M Nordmann[1,‡], Lisa Schweizer[1,‡], Andreas Metousis[1], Marvin Thielert[1], Edwin Rodriguez[1], Lise Mette Rahbek-Gjerdrum[2], Pia-Charlotte Stadler[3], Michael Bzorek[2], Andreas Mund[4], Florian A Rosenberger[1,*], Matthias Mann[1,4,*]

*[1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [2]Department of Pathology, Zealand University Hospital, Roskilde, Denmark. [3]Department of Dermatology and Allergy, University Hospital, LMU Munich, Germany. [4]Proteomics Program, Faculty of Health and Medical Sciences, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark.*

*[‡]These authors contributed equally to this work.*

**These authors contributed equally to this work as senior authors.*

In the staining process on membrane slides for DVP, we noticed a distortion of the PEN membrane resulting in poor reproducibility. This caused issues in the integrity of the tissue, including its detachment, leading to the need of long image analysis of many focal planes (z-level). Furthermore, liquid was able to pass the membrane which made LCM single-cell extraction impossible. This was because the membrane became more permeable with long treatments at high temperatures, which eventually led to bursting of the membrane. However, high temperatures are necessary for heat-induced antigen retrieval (HIER) to retrieve epitopes from proteins of FFPE tissue to produce the highest quality of images for single-cell segmentation and classification. We reasoned that enabling a higher viscosity might prevent the permeabilization of the liquid between the membrane and the glass. The addition of glycerol into the HIER (G-HIER) buffer solved the reproducibility issue. The integrity of the membrane was retained and only a small bubble was reproducibly formed at the top of the membrane. When we mounted tissue at the lower level of the membrane, we were able to remove the bubble, enabling high-content imaging on only one focal plane. Subsequently, we showed that G-HIER does not impact proteomic and spatial analysis. We only found differences between HIER treatment with different pH buffer independent of glycerol which was due to the solubility of proteins on isoelectric point and hydrophobicity of the measured peptides. Furthermore, we showed that G-HIER, also in a multiplexed imaging setup, can be used on automatic staining platforms to scale the workflow to routine histology and make it more reproducible.

***Contribution:***

Co-authorship. The concept and idea to include glycerol for optimized staining reproducibility of PEN membrane glass slides in the antigen-retrieval was initiated by Florian A. Rosenberger and spear-headed by Lisa Schweizer and Thierry Nordmann. Together with them, I designed and performed the experiments. I analyzed the effect of glycerol on LCM-cut tissue sections. Additionally, I developed the MS workflow and applied it to the high-sensitive DVP samples. I designed figures and wrote the MS part of the manuscript. Finally, I revised the manuscript.

MCP TECHNOLOGICAL INNOVATION AND RESOURCES

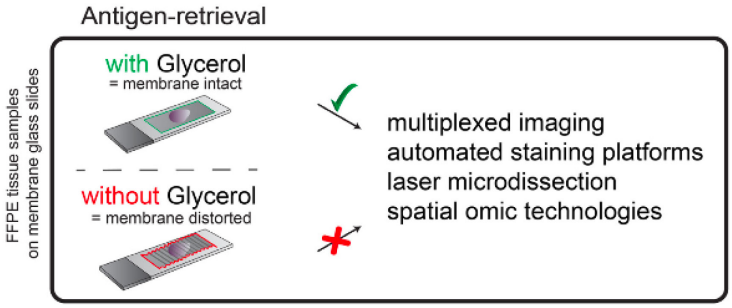
# A Standardized and Reproducible Workflow for Membrane Glass Slides in Routine Histology and Spatial Proteomics

**Authors**
Thierry M. Nordmann, Lisa Schweizer, Andreas Metousis, Marvin Thielert, Edwin Rodriguez, Lise Mette Rahbek-Gjerdrum, Pia-Charlotte Stadler, Michael Bzorek, Andreas Mund, Florian A. Rosenberger, and Matthias Mann

**Correspondence**
rosenberger@biochem.mpg.de;
mmann@biochem.mpg.de

**In Brief**
Integrating spatial omics with laser microdissection has been challenging due to incompatibility between routine histology and glass membrane slides. We introduce glycerol-modified antigen retrieval. This standardizes membrane slide use in routine histology and aligns with automated staining platforms. Importantly, it enables integration of multiplexed imaging, laser microdissection, and spatial proteomic workflows, advancing spatial omic technologies towards clinical use.

**Graphical Abstract**




**Highlights**

- Glycerol-enhanced antigen retrieval enables membrane slide use in routine histology.
- G-HIER is compatible with automated staining platforms and multiplexed imaging.
- G-HIER integrates with laser microdissection and spatial omic technologies.

2023, Mol Cell Proteomics *22*(*10*), 100643
© 2023 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
https://doi.org/10.1016/j.mcpro.2023.100643

MCP TECHNOLOGICAL INNOVATION AND RESOURCES

# A Standardized and Reproducible Workflow for Membrane Glass Slides in Routine Histology and Spatial Proteomics

Thierry M. Nordmann[1,‡], Lisa Schweizer[1,‡], Andreas Metousis[1], Marvin Thielert[1], Edwin Rodriguez[1], Lise Mette Rahbek-Gjerdrum[2], Pia-Charlotte Stadler[3], Michael Bzorek[2], Andreas Mund[4], Florian A. Rosenberger[1,*], and Matthias Mann[1,4,*]

**Defining the molecular phenotype of single cells *in situ* is key for understanding tissue architecture in health and disease. Advanced imaging platforms have recently been joined by spatial omics technologies, promising unparalleled insights into the molecular landscape of biological samples. Furthermore, high-precision laser microdissection (LMD) of tissue on membrane glass slides is a powerful method for spatial omics technologies and single-cell type spatial proteomics in particular. However, current histology protocols have not been compatible with glass membrane slides and LMD for automated staining platforms and routine histology procedures. This has prevented the combination of advanced staining procedures with LMD. In this study, we describe a novel method for handling glass membrane slides that enables automated eight-color multiplexed immunofluorescence staining and high-quality imaging followed by precise laser-guided extraction of single cells. The key advance is the glycerol-based modification of heat-induced epitope retrieval protocols, termed "G-HIER." We find that this altered antigen-retrieval solution prevents membrane distortion. Importantly, G-HIER is fully compatible with current antigen retrieval workflows and mass spectrometry–based proteomics and does not affect proteome depth or quality. To demonstrate the versatility of G-HIER for spatial proteomics, we apply the recently introduced deep visual proteomics technology to perform single-cell type analysis of adjacent suprabasal and basal keratinocytes of human skin. G-HIER overcomes previous incompatibility of standard and advanced staining protocols with membrane glass slides and enables robust integration with routine histology procedures, high-throughput multiplexed imaging, and sophisticated downstream spatial omics technologies.**

Spatial omics technologies, such as genomics, transcriptomics, and proteomics (1, 2), have tremendously advanced our understanding of tissue architecture by resolving the complex space-dependent interplay of heterogenous cell types. Laser microdissection (LMD) enables accurate extraction of tissue and single cells from defined anatomical regions, thereby preserving the all-important spatial information (3). Furthermore, LMD is compatible with many downstream omics technologies, providing the opportunity for multiomics analyses from the same sample (4, 5). Glass or metal frame slides coated with a plastic membrane are essential for precise LMD tissue extraction, in particular when attempting to retrieve anatomical structures at cellular levels or even subcellular levels. There are different types of membranes including polyethylene naphthalate (PEN), polyphenylene sulfide, polyethylene terephthalate, polyester, or fluorocarbon, each with unique physical and imaging properties (www.leica-microsystems.com/science-lab/application-specific-consumables-for-laser-microdissection/#gallery-11). Despite this variety, the incompatibility of membrane slides with routine histology strategies such as standard heat-induced epitope retrieval (HIER) has been a long-standing problem in spatial omics technologies (6). Specifically, multiplex immunofluorescence (mIF) can visualize numerous cell types, adding an important layer for omics analyses in a single tissue, but is hampered by the absence of robust multiplexed staining procedures on membrane slides (7, 8). In addition, the majority of studies that use membrane slides employ manual staining techniques, forgoing the advantages of reproducible and high-throughput staining systems. We propose a solution to these unsolved challenges by introducing a novel approach for handling of

From the [1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany; [2]Department of Pathology, Zealand University Hospital, Roskilde, Denmark; [3]Department of Dermatology and Allergy, University Hospital, LMU Munich, Germany; [4]Proteomics Program, Faculty of Health and Medical Sciences, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

‡These authors contributed equally to this work.

*For correspondence: Florian A. Rosenberger, rosenberger@biochem.mpg.de; Matthias Mann, mmann@biochem.mpg.de.



© 2023 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

https://doi.org/10.1016/j.mcpro.2023.100643

glass membrane slides for routine histological procedures and mIF. Our protocol processes membrane glass slides in high throughput for LMD, without altering or affecting downstream proteomic analysis. We show that this enables highly precise microdissection at the single-cell level while preserving all spatial information, allowing us to profile cell types using our recently published Deep Visual Proteomics (DVP) (9).

## EXPERIMENTAL PROCEDURES

### *Tissue Samples*

Formalin fixed and paraffin embedded (FFPE) human tissue samples were collected according to standard operating procedures. Briefly, skin specimens were stored in 5% formalin for 24 to 48 h at room temperature (RT), trimmed, and placed in embedding cassettes prior to automated processing (Tissue-TEK VIP; Sakura). Tissue specimens for the microarray block were fixed in formalin for 24 to 72 h at RT, trimmed, and manually embedded in paraffin.

### *Tissue Sectioning*

Two micrometer PEN membrane slides (MicroDissect GmbH; MDG3P40AK) were pretreated with VECTABOND (Biozol; VEC-SP-1800) according to the manufacturer's protocol and dried overnight at RT. FFPE tissue blocks were cooled to −17 °C at least 1 h prior to sectioning. Then, 2.5 µm sections were cut using a rotary microtome and placed onto a water bath at 37 °C. Sections were transferred to pretreated PEN membrane slides and dried overnight at 37 °C.

### *Visual Evaluation of Membrane Stability*

PEN membrane slides were treated identically to the initial experimental procedure of the DVP staining protocol described later. After antigen retrieval, slides were washed twice with double distilled water ($ddH_2O$) and imaged using the UV-scan mode of a gel documentation system (Axygen; GD-100) and Canon EOS 5D.

### *mIF Staining*

Human tonsil tissue sections were mounted on regular glass slides (SuperFrost Plus) or PEN membrane glass slides. Mounted slides were heated at 56 °C for 20 min and deparaffinized (2 × 2 min xylene, 2 × 1 min 100% EtOH, 95% EtOH, 75% EtOH, 30% EtOH, and ddH20, respectively). Glycerol-supplemented heat-induced epitope-retrieval (termed "G-HIER") was performed using preheated 1× target retrieval solution (TRS) pH 9 HIER buffer (DAKO; S2367)/10% glycerol (v/v; Sigma; G7757) in 50 ml conical tube placed in a water bath at 88 °C for 20 min. All further steps were performed according to the manufacturer's instructions, using the commercially available 8-color immuno-oncology panel (Immuno8 FixVUE; Ultivue). This kit contained antibodies against CD3, CD4, CD8, CD68, FoxP3, PD1, PDL1, and cytokeratin (CK). Nuclear counterstain was performed with 4′,6-diamidino-2-phenylindole and used for image registration using the UltiStacker software (Ultivue). Images were acquired on a Zeiss Axioscan Z7 at 20× magnification with a 10% tile overlap.

### *DAKO Staining Platform*

A small tissue microarray block composed of FFPE skin, appendix, cerebellum, adrenal gland, melanoma, and tonsillar tissue was sectioned and mounted on PEN slides as described previously. Next, sections were deparaffinized, rehydrated, and loaded wet on the fully automated instrument Omnis (Dako) based on dynamic gap staining technology and capillary forces. Sections were subjected to antigen retrieval using citrate buffer pH 6 (trisodium citrate dihydrate), TRS pH 9 (Dako; S2367), and TRS pH 6 (Dako; S2369) with or without 10% glycerol (Sigma–Aldrich/Merck; G7757) and heated for 60 min at 90 °C. Slides were subsequently incubated with anti-CK5 (1:200 dilution; Leica Biosystems, clone XM26, NCL-L-CK5) for 30 min at 32 °C. After washing and blocking of endogenous peroxidase activity, the reactions were detected/visualized using Envision FLEX+ High pH kit (Dako; GV800 + GV821) and Envision DAB+ Substrate Chromogen System (Dako; GV825) according to the manufacturer's instructions. Finally, slides were rinsed in water, counterstained with Mayer's hematoxylin, and air-dried prior to mounting.

### *Sample Preparation for Proteomics of Bulk Tissue Sections*

Tonsil tissue that was stained as described previously was collected from corresponding regions on consecutive slides using LMD. Subsequently, samples were lysed in 300 mM Tris–HCl (pH 8.0), 12.5% acetonitrile (ACN) including 5 mM Tris(2-carboxyethyl)phosphine and 20 mM chloroacetamide for 10 min at 90 °C, followed by focused ultrasonication (Covaris; Adaptive Focused Acoustic technology) and repeated heating for 80 min at 90 °C. Samples were then further processed as described previously (10).

### *Deep Visual Proteomics*

*Immunofluorescence Staining*—Mounted slides were heated at 56 °C for 20 min and deparaffinized (2 × 2 min xylene, 2× 1 min 100% EtOH, 95% EtOH, 75% EtOH, 30% EtOH, and $ddH_2O$, respectively). G-HIER was performed using preheated 1× TRS (pH 9) HIER buffer (DAKO; S2367)/10% glycerol (v/v; Sigma; G7757) in 50 ml conical tube placed in a water bath at 88 °C for 20 min. Subsequent to sample blocking using 5% bovine serum albumin in PBS for 30 min at RT, primary antibodies targeting pan-CK (rabbit; DAKO, Z0622, 1:100 dilution) and KRT10 (mouse; Abcam, ab76318, 1:800 dilution) were incubated for 1.5 h at 37 °C in a wet staining chamber. Following two wash steps in PBS, an incubation with secondary antibodies against mouse immunoglobulin G (A647; Invitrogen, A32728, 1:400 dilution) and rabbit immunoglobulin G (A555; Invitrogen, A32732; 1:400 dilution) was performed for 1 h at 37 °C in a wet staining chamber, followed by 7 min incubation with SYTOX green Nucleic Acid Stain (Invitrogen, S7020; 1:700 dilution in $ddH_2O$) at RT. Slides were then washed twice in $ddH_2O$ and allowed to dry briefly, before puncturing the membrane at the proximal end of the slide using a needle (30G) to eliminate the localized membrane elevation followed by subsequent adhesive sealing. Then, tissue sections were mounted with a cover glass using Slowfade Diamond Antifade Mountant (Invitrogen; S36967). After slide scanning, cover glasses were removed from membrane slides upon imaging by an incubation period of 5 to 10 min in $ddH_2O$ and air dried at RT.

*Image Analysis and LMD*—Artificial intelligence–guided cell recognition, classification, and extraction followed by mass spectrometry (MS)–based profiling was performed as described recently (9). In brief, fluorescence images were acquired on a Zeiss Axioscan Z7 at 20× magnification with a 10% tile overlap and analyzed using the Biology Image Analysis Software (Cell Signaling). Keratinocytes were identified with a deep neural network on the basis of pan-CK. After removal of duplicates at tile-overlapping regions, we used a supervised machine learning approach to differentially classify $KRT10^{pos}$ (*e.g.*, suprabasal) from $KRT10^{neg}$ (*e.g.*, basal) keratinocytes. Contour outlines were then exported along with reference points for image registration. Finally, 700 contours of each group were excised in quadruplicates on an LMD7 (Leica Microsystems). Each sample (consisting of 700 contours) was collected into a separate well of the underlying 384-well plate.

*Sample Processing and MS*—Sample replicates were processed following our recently published workflow (9). We used a liquid handling platform (Agilent Technologies; Bravo) to ensure reproducibility and high-throughput processing of samples. In brief, single-cell

shapes of the same type were collected at the bottom of each well using centrifugation and vacuum evaporation upon the addition of ACN. Subsequently, cells were lysed in 4 µl 60 mM triethylamine bicarbonate for 60 min at 95 °C, followed by further 60 min at 75 °C in 12% (v/v) ACN. Protein digest was performed overnight using 4 ng LysC and 6 ng trypsin in a total sample volume of 7.5 µl, respectively. The enzymatic reaction was quenched in a final concentration of 1% (v/v) TFA. Subsequently, samples were loaded on Evotips Pure tips following the manufacturer's instructions.

### *LC–MS/MS Analysis of Bulk and Ultra-High Sensitivity Data*

Bulk samples were reconstituted in buffer A* (2% ACN/0.1% formic acid [FA] in LC–MS grade water). MS data were acquired by an EASY nanoLC 1200 (Thermo Fisher Scientific) coupled to a timsTOF Pro2 mass spectrometer (Bruker Daltonics) with a nano-electrospray ion source (CaptiveSpray; Bruker Daltonics). Peptide amounts were determined using a Nanodrop VIS-spectrophotometer (Thermo Fisher Scientific), and 200 ng of peptides were loaded on a 50 cm in-house packed HPLC column (75 µm inner diameter packed with 1.9 µm ReproSil-Pur C18-AQ silica beads; Dr Maisch GmbH). The column temperature was kept at 60 °C by an in-house manufactured oven. Sample analytes were separated using a linear 120 min gradient from 3 to 30% buffer B in 95 min, followed by an increase to 60% for 5 min and to 95% buffer B for 5 min, as well as a 5 min wash at 95% buffer B and re-equilibration for 5 min at 5% buffer B (buffer A: 0.1% FA/dd$H_2O$; buffer B: 0.1% FA, 80% ACN, and 19.9% dd$H_2O$). The flow rate was kept constant at 300 nl/min. The mass spectrometer was operated in data-dependent acquisition-parallel accumulation serial fragmentation (PASEF) mode as previously described (11). Briefly, one MS1 scan was followed by ten PASEF MS/MS scans per acquisition cycle. The ion accumulation and ramp time in the dual TIMS analyzer was 100 ms, and ion mobility (IM) range was set from 1/K0 = 1.6 V $cm^{-2}$ to 0.6 V $cm^{-2}$. Single charged precursor ions were excluded with a polygon filter (timsControl, version 3.0.20.0; Bruker Daltonics), and precursors for MS/MS were picked at an intensity threshold of 2500 arbitrary units and resequenced until reaching a target value of 20,000 arbitrary units considering a dynamic exclusion of 40 s elution.

For the high-sensitivity DVP samples, samples were loaded onto Evotips Pure and measured with the Evosep One LC system (Evosep) coupled to a timsTOF SCP mass spectrometer (Bruker Daltonics) employing a nano-electrospray ion source (Bruker Daltonics). The Whisper 20 samples per day method was used with the Aurora Elite CSI third generation column with 15 cm and 75 µm ID (AUR3-15075C18-CSI, IonOpticks) at 50 °C. The mobile phases comprised 0.1% FA in LC–MS grade water as buffer A and 99.9% ACN/0.1% FA as buffer B. The timsTOF was operated in data-independent acquisition (DIA)-PASEF mode with variable window widths. Optimal DIA-PASEF methods cover the precursor cloud highly efficient in the *m/z*–IM plane while providing deep proteome coverage. For method generation with py_diAID, the precursor density distribution in *m/z* and IM was estimated based on a tryptic 48 high-pH fraction library (12). We calculated the optimal cycle time based on the chromatographic peak width of 5 ng HeLa single runs. The optimal DIA-PASEF method consisted of one MS1 scan followed by 12 DIA-PASEF scans with two IM ramps per DIA-PASEF scan, covering an *m/z* range from 300 to 1200 and IM of 0.7 to 1.3 V $cm^{-2}$. All other settings were as described previously.

### *Data Processing*

Data acquired in data-dependent acquisition-PASEF mode were analyzed with MaxQuant (version 2.0.1.0) using standard settings in reference to UniProt human databases (UP000005640_9606.fasta containing 21,010 proteins/20,243 genes and UP000005640_9606_additional.fasta containing 58,856 proteins/15,630 genes) (13). Oxidation (M) and acetyl (protein-N-term) was set as variable and carbamidomethyl (C) as fixed modification. The maximum number of modifications per peptide was set to 3. Trypsin/P was selected as protease, and the maximum number of missed cleavages was set to 2. MS data acquired in DIA-PASEF mode were processed in the DIA-NN software (version 1.8.1, (14)) using the same fasta files and standard settings. Search parameters deviated from the standard settings as follows: peptide length range is 7 to 55, precursor *m/z* range is 100 to 1700, fragment ion *m/z* range is 100 to 1700, quantification strategy "any LC (high accuracy)," cross-run normalization "global," with and without MBR ("match between runs") enabled. Mass tolerance for precursor ions was determined by DIA-NN (15.049 ppm, MBR) and 15.0349 ppm (no MBR). Mass tolerance for fragment ions was determined by DIA-NN (20.1046 ppm, MBR and 20.1046 ppm, no MBR). Precursor false discovery rate (FDR) was set to 1%. The report.pg_matrix output file of DIA-NN was used for further data analysis.

### *Bioinformatics Data Analysis*

Data analysis was performed in R, version 4.2.2. Statistical analysis of bulk data was performed with limma, version 3.52.4. For the comparisons shown, the number of significant hits (FDR <5%) was corrected for multiple testing using a Benjamini–Hochberg correction. The theoretical isoelectric point of a protein was calculated with the computePI function of seqinr v4.2-23 for every protein within one protein group. Hydrophobicity was estimated with the hydrophobicity function (scale = "KyteDoolittle") in the package peptides, version 2.4.4. Amino acid sequences for UniProt identifiers were retrieved with the UniProt.ws package v2.36.5. For protein groups (sequences that cannot be distinguished by the underlying peptide identifications), the mean value of individual proteins is presented.

For the analysis of DVP data, differential protein expression was determined as described previously. Biological pathway enrichments were performed using the WebGestaltR package, v.0.4.4 for an overrepresentation analysis in reference to the "Reactome" database and an FDR threshold of 5% in the background of all identified proteins in the dataset (organism: *Homo sapiens*) (15). Spatial data from xml files were plotted with the package sf v1.0-9 as described previously (16). For the principal component analysis, data were filtered for 80% valid values in each group and imputed from left-shifted normal distribution (shift = 1.8, scale = 0.3). The FactoMineR v2.6 package was used to perform the principal component analyses. The Venn diagram and UpsetR package were used to display the number of intersections and unique proteins per condition. Visual evaluation of membrane stability was performed in Python 3.9.12 using the NumPy, Matplotlib, and Scikit-image packages. Intensity values of the acquired images were normalized to the median intensity of each image and then plotted as heatmaps using the "viridis" colormap, indicating intensity values.

### *Ethics Approval and Patient Consent Statement*

Skin tissue sections were collected following informed consent and ethical approval (EK 22-0343). All experiments were performed in accordance with the Declaration of Helsinki. Regarding the tissue microarrays, according to Danish Data Protection Act (2018), they are exempted from patient consent and permission from legal authorities for fully anonymized material.

## RESULTS

Combining digital pathology with proteomics using LMD is an area of intense interest. However, there are many practical challenges to overcome in achieving this. In our initial work with

PEN membranes, we regularly observed substantial disintegration of the membrane and formation of air pockets at random locations underneath during standard HIER protocols. This renders the majority of slides incompatible with routine histological staining workflows, scanning, and potential downstream LMD procedures (Fig. 1*A* and supplemental Fig. S1*A*). To address this issue, we aimed at developing a robust membrane-compatible HIER protocol. Building on the observation that glycerol in HIER buffers enhances antigen retrieval (17), we tested the ability of glycerol in membrane stabilization. Indeed, we found that membrane integrity was preserved when using glycerol-adapted HIER, while fully eliminating the formation of random membrane distortions (Fig. 1*B* and supplemental Fig. S1*B*). Following G-HIER, tissue sections could easily be processed for staining and high-quality imaging, even in a single focal plane. For very large tissue sections, we additionally punctured the membrane at its proximal end with a syringe to remove any potentially remaining irregularities and subsequently resealed it (supplemental Fig. S1*C*). With this in place, G-HIER allows high-resolution

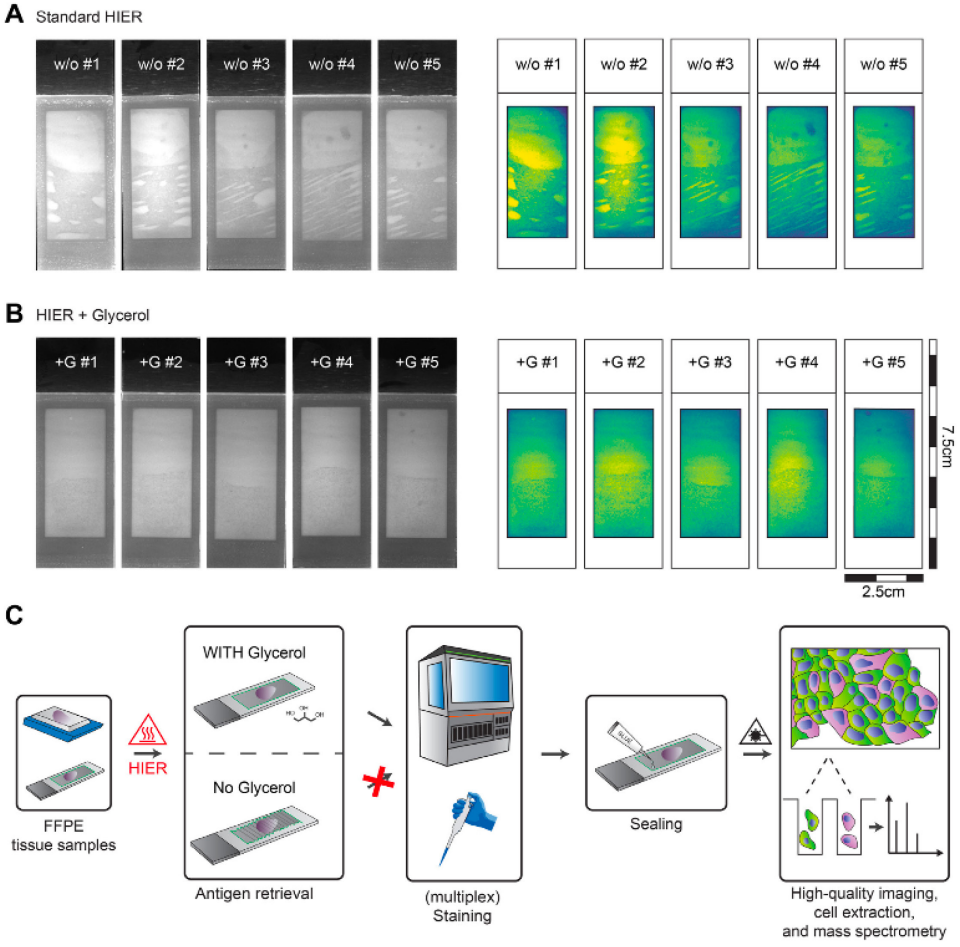



FIG. 1. **A robust workflow for heat-induced epitope retrieval (HIER) on membrane microscopy slides.** *A*, membrane distortion resulting from a standard HIER protocol shown as UV scans (*left*) and heatmap of normalized median UV signal intensities across slides. *B*, intact membrane stability using glycerol-supplemented antigen retrieval in the same arrangement. *C*, workflow schematic for the preparation of LMD-compatible membrane slides. FFPE archival tissue was mounted on 2 µm PEN membrane slides as in routine pathology followed by standard HIER without and with 10% glycerol. While the slides without glycerol were unusable for further staining procedures upon heating in solution, the presence of glycerol stabilized the membrane. Minor accumulation of gas below the membrane was removed using minimal invasive methods followed by glue-based sealing. FFPE, formalin fixed and paraffin embedded; +G, HIER buffer supplemented with 10% glycerol; LMD, laser microdissection; PEN, polyethylene naphthalate; w/o, without glycerol.

and rapid whole-slide imaging in a single focal plane, minimizing the amount of data storage necessary and reducing stitching errors that are detrimental for subsequent LMD. Having optimized HIER on membrane slides for routine histology workflows, we tested its compatibility with multiplexed imaging, automated staining device compatibility (Dako Omnis; Agilent), and downstream MS-based proteomics (Fig. 1*C*).

Multiplexed imaging enables the spatial characterization of diverse cellular population within the tissue microenvironment and has become increasingly popular in the scientific community. However, to our knowledge, it has not been performed on glass membrane slides and accordingly has not been combined with downstream LMD-based spatial omics technologies. To test the compatibility of our novel antigen retrieval method with membrane slides in combination with multiplexed imaging, we stained tonsil tissue with a commercially available 8-color immuno-oncology panel (Immuno8 FixVUE; Ultivue). Without further optimization, G-HIER resulted in high-quality images on PEN membrane slides and showed equal performance with all antibodies compared with regular glass slides (Fig. 2). Importantly, our staining unambiguously visualized the different cell types of the panel while also clearly highlighting cell boundaries required for automated cell segmentation. These results show that G-HIER enables multiplexed imaging on glass membrane slides, and both are seamlessly compatible with glass membrane slides.

Next, we evaluated the impact of G-HIER on protein identification using MS-based proteomics. We compared the proteome depth and information content of FFPE tonsil tissue samples after antigen retrieval with six different buffer compositions: 10 mM citrate buffer at pH 6, and commercially available DAKO buffer at pH 6 and pH 9, with or without addition of 10% glycerol, respectively. To ensure a robust and efficient experimental process, we directly integrated G-HIER and CK5 staining of FFPE tissue sections mounted on PEN-membrane slides on an automated staining platform (DAKO Omnis), followed by LMD and MS. We found that membrane glass slides were fully compatible with an automated staining device when using G-HIER. Proteomic analysis revealed identical numbers of identified proteins across all buffer conditions, at 5468 ± 12 proteins (mean ± SD, Fig. 3*A*). In our experience, such proteome depth provides substantial coverage of most biological pathways. Likewise, distribution and median of MS intensity per protein and the coefficient of variation as a measure of variability between replicates were comparable, demonstrating that glycerol has no negative impact on proteomic data acquisition (Fig. 3, *B–D*). The degree of common identifications and the distinct separation between treatment groups was not affected by the "MBR"

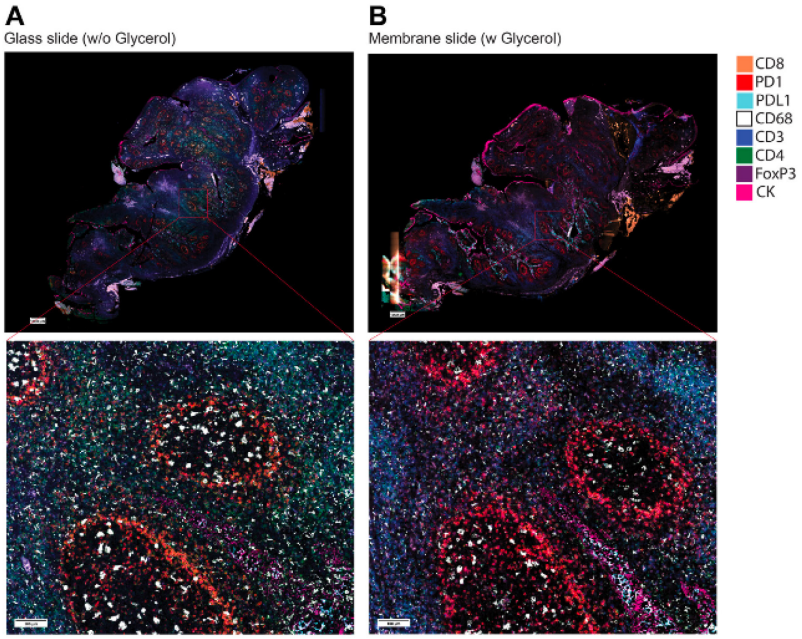



Fig. 2. **Comparison of multiplex immunofluorescence imaging in human tonsil tissue.** Eight-color multiplex immunofluorescence imaging (Immuno8 FixVUE; Ultivue) was performed side by side, in tissue mounted on regular (*A*) or PEN-membrane glass slides (*B*). Apart from G-HIER at 88 °C, all other steps were performed as recommended by the manufacturer. The following markers were used: CD8 (*orange*), PD1 (*red*), PDL1 (*turquoise*), CD68 (*white*), CD3 (*blue*), CD4 (*green*), FoxP3 (*purple*), and CK (*pink*). G-HIER, glycerol-supplemented heat-induced epitope retrieval; PEN, polyethylene naphthalate.

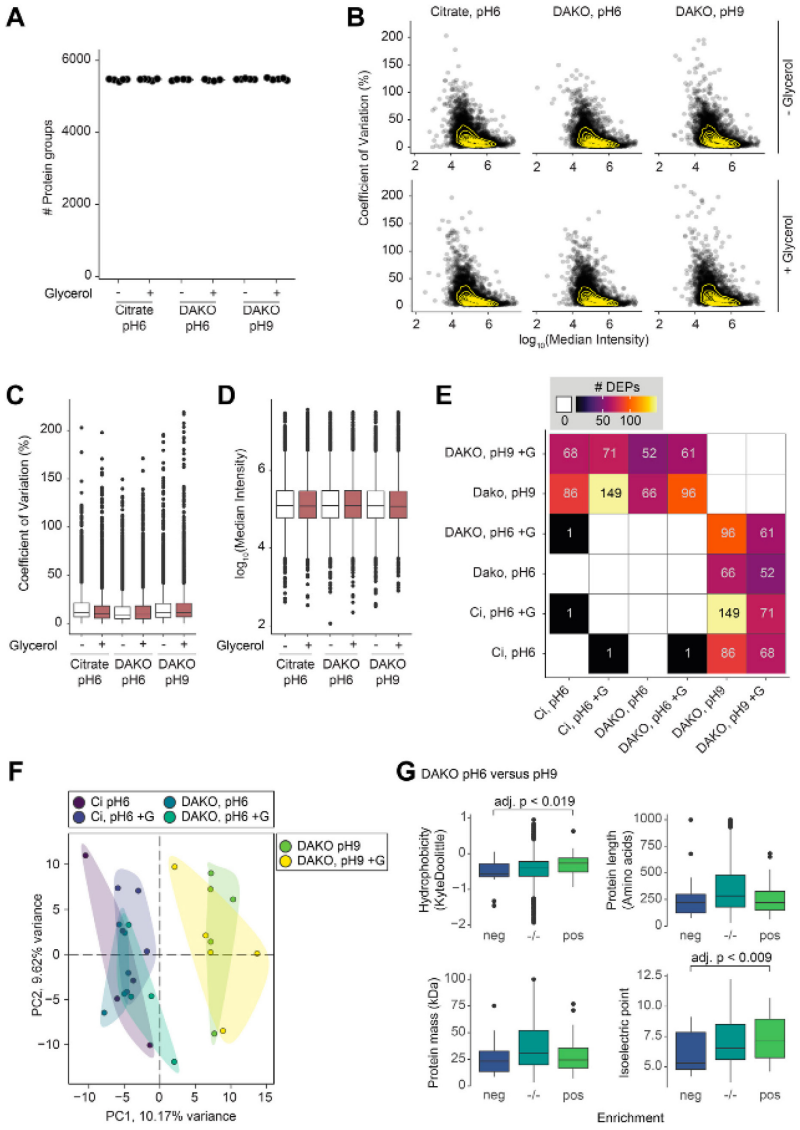



FIG. 3. **Effect of glycerol in various antigen retrieval solutions on the bulk proteome of tissue slices.** *A*, number of proteins detected in laser-dissected tonsil tissue after LC–MS/MS. *Points* are individual samples, the *line* represents the median within one group (n > 4). *B*, log-transformed MS signal *versus* coefficient of variation (%) per solution type. *Points* are median values of individual proteins per group, density is estimated by *yellow lines*. *C*, coefficient of variation and (*D*) $\log_{10}$-transformed MS signal as in (*B*) per group (n > 4). *E*, number of differentially expressed proteins (multiple-testing adjusted *p* value <0.05) between indicated comparisons, additionally illustrated with a color gradient. *White* and *empty* indicates that no significant hits were found. *F*, principal component analysis of all included samples (n = 28). *Colors* denote groups with different retrieval solutions. The amount of variation explained per principal component is given with the axis legend. *G*, comparison of proteome composition after heat-induced antigen retrieval with two solutions at pH 6 or pH 9 (without glycerol). Hydrophobicity and isoelectric point are calculated values. For protein groups, the median value of each contributing protein is shown (n > 4). *Boxplots* represent median and the 25th and 75th percentiles, and *whiskers* span the 1.5-fold interquartile range. Outliers are shown, if not indicated differently. adj. *p*., adjusted *p* value; Ci, citrate; DEP, differentially expressed proteins; +G, respective buffer supplemented with 10% glycerol; MS, mass spectrometry.

algorithm of the data processing software, despite slightly lower protein numbers (supplemental Fig. S2, *A–D*). Notably, in our comparison of different antigen retrieval conditions, we only found a difference because of the pH of the HIER buffer, while addition of glycerol was not a main driver of sample separation along principal component 1 and 2 (Fig. 3, *E* and *F*). We speculated that the pH could affect solubility of proteins depending on their physiochemical properties. Indeed, we found that the theoretical isoelectric point and hydrophobicity, rather than protein length or mass, affected which proteins were retained upon HIER at pH 6 compared with pH 9 (Fig. 3*G*). Although this does not affect G-HIER, it can be taken into consideration when establishing the staining of particular challenging protein targets. Thus, our optimized workflow efficiently combines HIER on membrane slides with downstream proteomics and automated staining techniques, without compromising reproducibility and MS data depth.

With a reliable protocol in hand for multiplexed imaging on glass membrane slides that preserves the potential of downstream proteomic analysis, we next aimed to demonstrate its applicability to the field of spatial proteomics. For this, we integrated G-HIER into the recently developed DVP workflow. Using G-HIER, we performed immunofluorescence for pan-CK and CK10 on a 3 μm thin FFPE skin section (10). Despite the inherent challenges of handling skin tissue sections during harsh histological conditions, we were able to successfully perform the entire staining procedure (including antibody titration experiments) on glass membrane slides without any signs of tissue detachment. This allowed us to profile adjacent epidermal cell types of healthy human skin, yielding spatially resolved proteomic profiles. Artificial intelligence–based algorithms segmented each cell on the basis of pan-CK and further classified it into basal and suprabasal cells by the presence or the absence of CK10 (Fig. 4*A*). Laser-assisted cell extraction and subsequent MS enabled clear differentiation between the cell types based on the identification and quantification of more than 3500 proteins across all samples (Fig. 4, *B* and *C*). Again, the degree of common identifications and the distinct separation between biological groups was not affected by the "MBR" algorithm (supplemental Fig. S3, *A–D*). Between the two cell types, 17% of the proteome was profoundly differentially expressed (minimum absolute fold change: 1.5, adjusted $p$ value <0.05, Fig. 4*D*). CK10 (KRT10), the initial staining target of our study, and KRT1 were upregulated in suprabasal cells, providing positive controls. In the basal cell layer, common markers of basal keratinocytes such as CK15 (KRT15) and basal membrane anchorage fibrils COL4A1 and COL7A1 were highly enriched. Summarizing the proteomics results by a biological pathway enrichment analysis, processes such as "keratinization" and "metabolism of lipids" reflected the formation of the epidermal barrier in the suprabasal layer, whereas "type I hemidesmosome assembly" and other cell–cell interactions indicated structural anchoring of cells in the basal layer (Fig. 4*E*). We next took advantage of the integration of spatial and proteomic information to reconstruct the cellular architecture of the skin specimen, mapping the mean protein intensities of KRT1 and KRT15 into the cellular environment (Fig. 4, *F* and *G*). These results demonstrate the robust nature of G-HIER and DVP and how this combination enables the cell type–resolved characterization of human skin by MS-based proteomics while preserving spatial information.

## DISCUSSION

HIER workflows have not traditionally been compatible with glass membrane slides for LMD. This long-standing issue has limited the integration of LMD-based approaches with current and future technologies, such as automated staining platforms, mIF imaging, and visual omics approaches.

To address these challenges, we here establish G-HIER. This optimized technique effectively stabilizes the membrane during antigen retrieval and enables subsequent high-throughput staining and LMD. Our protocol drew inspiration from the previous finding that the addition of 10% glycerol to the antigen retrieval process greatly enhances the quality of immunostainings in FFPE tissue sections (17). This earlier study highlighted the potential benefits of glycerol in providing a universal antigen retrieval buffer that enables excellent staining properties for a large variety of epitopes. Our own findings align with these results. Beyond this advantage, we here expanded the benefits of glycerol to the favorable impact on PEN membrane slides during the antigen retrieval process. We found that our optimized workflow drastically enhances the quality of antibody-based staining results on PEN glass slides, enabling their standard use in routine pathology and even multiplexed staining procedures. In addition, fully automated staining platforms such as the Dako Omnis can now be integrated seamlessly to further increase throughput and reproducibility, while still matching the requirements for high-throughput tissue extraction *via* laser-capture microdissection. Furthermore, our workflow is fully compatible with downstream MS-based proteomics analysis and provides reproducible and in-depth acquisition of the proteome.

These advancements are a crucial step toward robust single-cell or cell type–specific proteomics workflows. This is especially important, when working with extremely sparse material (16). Indeed, clinical material is often precious and scarce. In this context, G-HIER provides a standardized framework for efficient and lossless use of the few sections that may be available for research purposes. We demonstrate this powerful combination by differentiating between basal and suprabasal cells of the human skin, but this can be transferred to any cell type of interest. We have now standardized on G-HIER for all applications of our novel DVP technology and are evaluating the routine combination of automated staining with G-HIER.

In summary, G-HIER now enables routine use of glass membrane slides in standard staining procedures and more

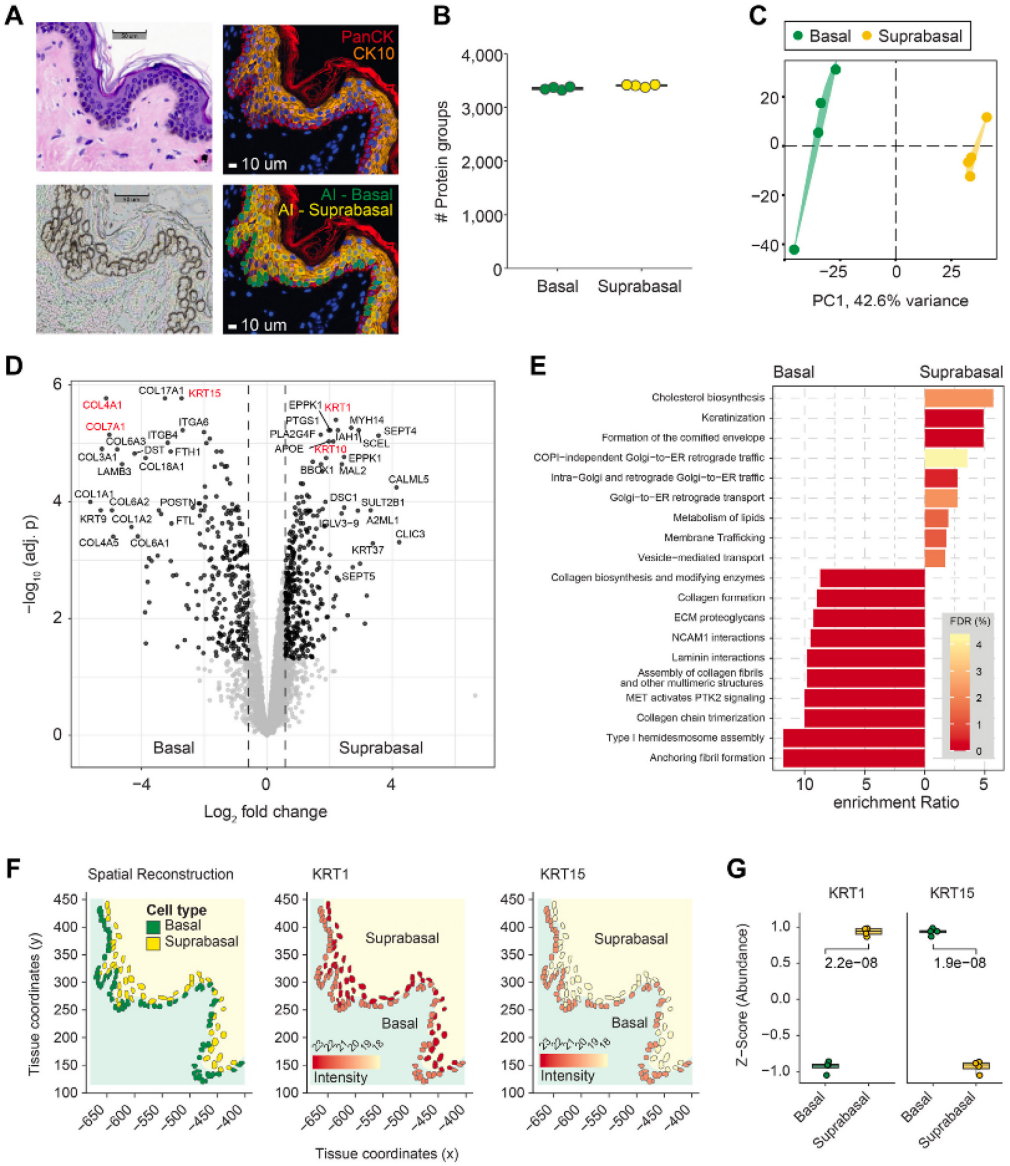



FIG. 4. **Cell type–resolved spatial proteomics of the human skin.** *A*, exemplary region of skin specimen shown in H&E (*upper left*), IF for pan-CK and CK10 (*upper right*), cell segmentation and classification by artificial intelligence (*lower right*), and subsequent to cell extraction by laser capture microdissection (*lower left*). *B*, number of identified protein groups for basal and suprabasal cells of the skin. *C*, principal component analysis separating skin cell types in the first two dimensions. *D*, differential protein expression of cell types. Significant proteins (minimum fold change: 1.5, adjusted *p* value <0.05) are highlighted in *black*, proof-of-principle proteins are shown in *red*. *E*, biological pathway enrichment analysis based on the "pathway Reactome" database. *F*, *in silico* spatial reconstruction of the tissue architecture showing the localization of basal and suprabasal cells in an exemplary tissue region (*left panel*). Enabled by DVP, mean protein intensities for the representative markers KRT1 (*middle panel*) and KRT15 (*right panel*) were mapped onto the tissue architecture. *G*, normalized abundances for the previously shown markers in each cell type. An unpaired Student's *t* test was used to annotate the significance for each comparison. adj. *p*.: adjusted *p* value; CK, cytokeratin; DVP, Deep Visual Proteomics; IF, immunofluorescence.

complex mIF imaging, which in turn can readily be coupled to LMD-assisted tissue extraction. G-HIER broadens applications of glass membrane slides in pathology and enables their streamlined use in the growing field of spatial omics technologies.

## DATA AVAILABILITY

MS data of this study have been deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository with the dataset identifier PXD040281 (preliminary username: reviewer_pxd040281@ebi.ac.uk, password: NpBxbVb9).

*Supplemental data*—This article contains supplemental data.

*Acknowledgments*—We thank members of the department of Proteomics and Signal Transduction for help and fruitful discussions and especially Katharina Zettl for superbly performing the immunofluorescence staining and Dirk Wischnewski for meticulous sample preparations. Susanne Vondenbusch-Teetz is acknowledged for the scientific implementation of her photography skills. We thank Peter Horvath and Single Cell Technologies Ltd for their technical support and Fabian Coscia for valuable feedback. This study has further been supported by the Max-Planck Society for Advancement of Science, the Federal Ministry of Education and Research (BMBF) through project CLINSPECT-M (16LW0243K), the Chan Zuckerberg Initiative (CZF2019-002448), and by grants from the Novo Nordisk Foundation, Denmark (grant agreements: NNF14CC0001 and NNF15CC0001).

*Funding and additional information*—T. M. N. is supported by a Swiss National Science Foundation Early Postdoc Mobility fellowship (P 2ZHP3-199648), and F. A. R. is an EMBO postdoctoral fellow (ALTF 399-2021).

*Author contributions*—F. A. R. original idea; T. M. N., L. S., and F. A. R. conceptualization; T. M. N., L. S., E. R., and F. A. R. formal analysis; T. M. N., L. S., Andreas Metousis., M. T., E. R., L. M. R.-G., P.-C. S., M. B., and F. A. R. investigation; T. M. N., L. S., F. A. R., and M. M. writing–original draft; Andreas Mund and M. M. supervision.

*Conflict of interest*—M. M. is an indirect investor in Evosep Biosciences. All other authors declare no competing interests.

*Abbreviations*—The abbreviations used are: ACN, acetonitrile; CK, cytokeratin; ddH2O, double distilled water; DIA, data-independent acquisition; DVP, Deep Visual Proteomics; FA, formic acid; FDR, false discovery rate; FFPE, formalin fixed and paraffin embedded; G-HIER, glycerol-supplemented HIER; HIER, heat-induced epitope retrieval; IM, ion mobility; LMD, laser microdissection; MBR, match between runs; mIF, multiplex immunofluorescence; MS, mass spectrometry; PASEF, parallel accumulation serial fragmentation; PEN, polyethylene naphthalate; RT, room temperature; TRS, target retrieval solution.

Received July 6, 2023, and in revised form, August 28, 2023 Published, MCPRO Papers in Press, September 7, 2023, https://doi.org/10.1016/j.mcpro.2023.100643

## REFERENCES

1. Zhao, T., Chiang, Z. D., Morriss, J. W., LaFave, L. M., Murray, E. M., Del Priore, I., *et al*. (2022) Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. *Nature* **601**, 85–91
2. Nirmal, A. J., Maliga, Z., Vallius, T., Quattrochi, B., Chen, A. A., Jacobson, C. A., *et al*. (2022) The spatial landscape of progression and immunoediting in primary melanoma at single-cell resolution. *Cancer Discov.* **12**, 1518–1541
3. Espina, V., Wulfkuhle, J. D., Calvert, V. S., VanMeter, A., Zhou, W., Coukos, G., *et al*. (2006) Laser-capture microdissection. *Nat. Protoc.* **1**, 586–603
4. Shen, S., Li, J., Huo, S., Ma, M., Zhu, X., Rasam, S., *et al*. (2021) Parallel, high-quality proteomic and targeted metabolomic quantification using laser capture microdissected tissues. *Anal. Chem.* **93**, 8711–8718
5. Nichterwitz, S., Chen, G., Aguila Benitez, J., Yilmaz, M., Storvall, H., Cao, M., *et al*. (2016) Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nat. Commun.* **7**, 12139
6. Gjerdrum, L. M., Lielpetere, I., Rasmussen, L. M., Bendix, K., and Hamilton-Dutoit, S. (2001) Laser-assisted microdissection of membrane-mounted paraffin sections for polymerase chain reaction analysis: identification of cell populations using immunohistochemistry and *in situ* hybridization. *J. Mol. Diagn.* **3**, 105–110
7. Hickey, J. W., Neumann, E. K., Radtke, A. J., Camarillo, J. M., Beuschel, R. T., Albanese, A., *et al*. (2022) Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging. *Nat. Methods* **19**, 284–295
8. Tan, W. C. C., Nerurkar, S. N., Cai, H. Y., Ng, H. H. M., Wu, D., Wee, Y. T. F., *et al*. (2020) Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. *Cancer Commun. (Lond.)* **40**, 135–153
9. Mund, A., Coscia, F., Kriston, A., Hollandi, R., Kovacs, F., Brunner, A. D., *et al*. (2022) Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* **40**, 1231–1240
10. Coscia, F., Doll, S., Bech, J. M., Schweizer, L., Mund, A., Lengyel, E., *et al*. (2020) A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *J. Pathol.* **251**, 100–112
11. Brunner, A. D., Thielert, M., Vasilopoulou, C., Ammar, C., Coscia, F., Mund, A., *et al*. (2022) Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **18**, e10798
12. Skowronek, P., Thielert, M., Voytik, E., Tanzer, M. C., Hansen, F. M., Willems, S., *et al*. (2022) Rapid and in-depth coverage of the (phospho-) proteome with deep libraries and optimal window design for dia-PASEF. *Mol. Cell. Proteomics* **21**, 100279
13. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
14. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., and Ralser, M. (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44
15. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., and Zhang, B. (2019) WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205
16. [preprint] Rosenberger, F. A., Thielert, M., Strauss, M. T., Ammar, C., Mädler, S. C., Schweizer, L., *et al*. (2022) Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome. *bioRxiv*. https://doi.org/10.1101/2022.12.03.518957
17. Sorrelle, N., Ganguly, D., Dominguez, A. T. A., Zhang, Y., Huang, H., Dahal, L. N., *et al*. (2019) Improved multiplex immunohistochemistry for immune microenvironment evaluation of mouse formalin-fixed, paraffin-embedded tissues. *J. Immunol.* **202**, 292–299

## Article 7. Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome.

*Nature Methods*, 2023 (20, 1530-1536); DOI: 10.1038/s41592-023-02007-6

Florian A. Rosenberger[1], Marvin Thielert[1], Maximilian T. Strauss[2], Lisa Schweizer[1], Constantin Ammar[1], Sophia C. Mädler[1], Andreas Metousis[1], Patricia Skowronek[1], Maria Wahle[1], Katherine Madden[1], Janine Gote-Schniering[3], Anna Semenova[3], Herbert B. Schiller[3], Edwin Rodriguez[1], Thierry M. Nordmann[1], Andreas Mund[2], Matthias Mann[1,2]

*[1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [2]Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [3]Comprehensive Pneumology Center (CPC) / Institute of Lung Health and Immunity (LHI), Helmholtz Munich; Member of the German Center for Lung Research (DZL), Munich, Germany.*

As discussed above, DVP overcomes depth and throughput limitations of spatial proteomics by combining a required number of cells with similar staining patterns and morphological features and to identify statistically and analytically robust cellular phenotypes (‘biological fractionation’). However, with the advances in MS sensitivity, we wanted to resolve spatial proteomics on the single-cell level, by developing single-cell DVP (scDVP).

The major structures in the liver are the ‘liver lobules’. In these repetitive hexagonal-shaped units, blood rich in oxygen enters at the portal vein (PV) at the periphery and drains at the central vein (CV) in the middle of the lobule. Hepatocytes are the most prominent cell type in the liver and they are known to have distinct functions dependent on their position along this metabolic spatial gradient, called liver zonation. We established a ‘liver paint’ staining marking PV with E-cadherin and CV with Glul (glutamate-ammonia ligase), thus visualizing the sinusoidal architecture. Phalloidin outlined the plasma membrane, and wheat germ agglutinin (WGA) the nucleus for cell segmentation.

We adapted our most advanced single-cell workflow using mDIA to DVP. This enhanced sensitivity enabled the analysis of one single contour of hepatocytes at a time, 100-fold less cell surface area than in DVP. We measured 455 single hepatocyte shapes in three biological replicates and reached a median depth of over 1,700 protein groups. The first principal component corresponded to the distance of the measured hepatocyte to PV and CV. Our spatial sc-proteomics data revealed zonation-dependent expression of almost half of the detected proteins, including known zonation markers like Ass1 and Glul. Overall, portal regions were more similar in their expression signatures than central

ones. Importantly, we found complete metabolic pathways differentially abundant along the PV-CV gradient. As biological positive control, proteins participating in ammonia fixation of the urea cycle and oxidative phosphorylation (OXPHOS) were highly expressed at the portal vein, while proteins involved in ammonia capture on glutamate were strong close to the pericentral vein. We also found new pathways correlated with zonation like peroxisome proliferator-activated receptor (PPAR) signaling, which was mainly expressed in central parts.

Contributing to the central question of the degree of correlation between transcriptome and proteome, we found that the overall correlation was very low (Pearson's R = 0.12 for transcript versus protein changes along the zonation axis). Intriguingly, however, this correlation was very high for a set of prominent zonation markers (R = 0.97), presumably reflecting tightly controlled regulation of this process at both levels of gene expression. In addition, having the two-dimension data of imaging and proteomics available, we reasoned that the image data may predict the proteome to some degree. Indeed, we successfully predicted the proteome intensities of five proteome classes with a correlation of R = 0.78.

In summary, scDVP is the first technology which directly links sc-proteomics and spatial proteomics. We demonstrated that single-cell data reveals subtle biological differences not visible when excising larger areas. Additionally, our work opens the application of DVP for tissue systems in which no established markers are known, to study the full heterogeneity at the spatial level.

***Contribution:***

Co-authorship, second-author. I originated the concept of scDVP together with Florian Rosenberger and Matthias Mann. I previously developed the mDIA workflow, essential for scDVP. Together with Florian Rosenberger, I designed, performed and analyzed the single-cell experiments, designed figures and the manuscript structure. Finally, I revised and edited the manuscript.

nature methods

Article

https://doi.org/10.1038/s41592-023-02007-6

# Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome

Received: 9 December 2022

Accepted: 15 August 2023

Published online: 2 October 2023

Check for updates

Florian A. Rosenberger[1], Marvin Thielert[1], Maximilian T. Strauss[2], Lisa Schweizer[1], Constantin Ammar[1], Sophia C. Mädler[1], Andreas Metousis[1], Patricia Skowronek[1], Maria Wahle[1], Katherine Madden[1], Janine Gote-Schniering[3], Anna Semenova[3], Herbert B. Schiller[3], Edwin Rodriguez[1], Thierry M. Nordmann[1], Andreas Mund[2] & Matthias Mann[1,2]

Single-cell proteomics by mass spectrometry is emerging as a powerful and unbiased method for the characterization of biological heterogeneity. So far, it has been limited to cultured cells, whereas an expansion of the method to complex tissues would greatly enhance biological insights. Here we describe single-cell Deep Visual Proteomics (scDVP), a technology that integrates high-content imaging, laser microdissection and multiplexed mass spectrometry. scDVP resolves the context-dependent, spatial proteome of murine hepatocytes at a current depth of 1,700 proteins from a cell slice. Half of the proteome was differentially regulated in a spatial manner, with protein levels changing dramatically in proximity to the central vein. We applied machine learning to proteome classes and images, which subsequently inferred the spatial proteome from imaging data alone. scDVP is applicable to healthy and diseased tissues and complements other spatial proteomics and spatial omics technologies.

Mass spectrometry (MS)-based single-cell proteomics (scProteomics) has made tremendous progress within just a few years, and can now quantify more than 1,000 proteins in cultured cells[1–3]. While this trajectory is promising, proteome depth, throughput and lack of spatial context limit biological use. We have recently introduced deep visual proteomics (DVP), a spatial technology that combines imaging, cell segmentation, laser microdissection and MS into a single workflow to investigate complex tissues with various cell types and metabolic niches[4]. DVP overcomes depth and throughput limitations with pooling the required number of cells with similar morphological features and staining patterns to identify statistically and analytically robust cellular phenotypes ('biological fractionation'). By its nature, it depends on prior knowledge of adequate markers of the cells of interest that resolve their heterogeneity. These markers might not be available for all subtypes of cells or those tissues that have rapidly changing proteome types such as heterogeneous tumors. To address this, we here developed single-cell DVP (scDVP), a complementary approach that extends scProteomics technologies into the intact tissue context.

In this Article, we use scDVP to explore spatial characteristics of hepatocyte subsets in mammalian liver—a highly organized and functionally repetitive tissue, in which the proteome of hepatocytes is determined by paracrine signaling, as well as oxygen and nutrient gradients[5]. These metabolic gradients require distinct functional cell states along the portal vein (PV) to central vein (CV) axis. This phenomenon of liver zonation has been described by single-cell RNA sequencing (scRNAseq) for hepatocytes[6,7], fluorescence-activated cell sorting (FACS) and MS-based proteomics[8], and multiplexed imaging[9]. Despite this long and varied background, the extent of spatial heterogeneity and proteome variation in hepatocyte remains an open question.

[1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [2]Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [3]Comprehensive Pneumology Center (CPC) / Institute of Lung Health and Immunity (LHI), Helmholtz Munich; Member of the German Center for Lung Research (DZL), Munich, Germany.
e-mail: mmann@biochem.mpg.de

Article https://doi.org/10.1038/s41592-023-02007-6

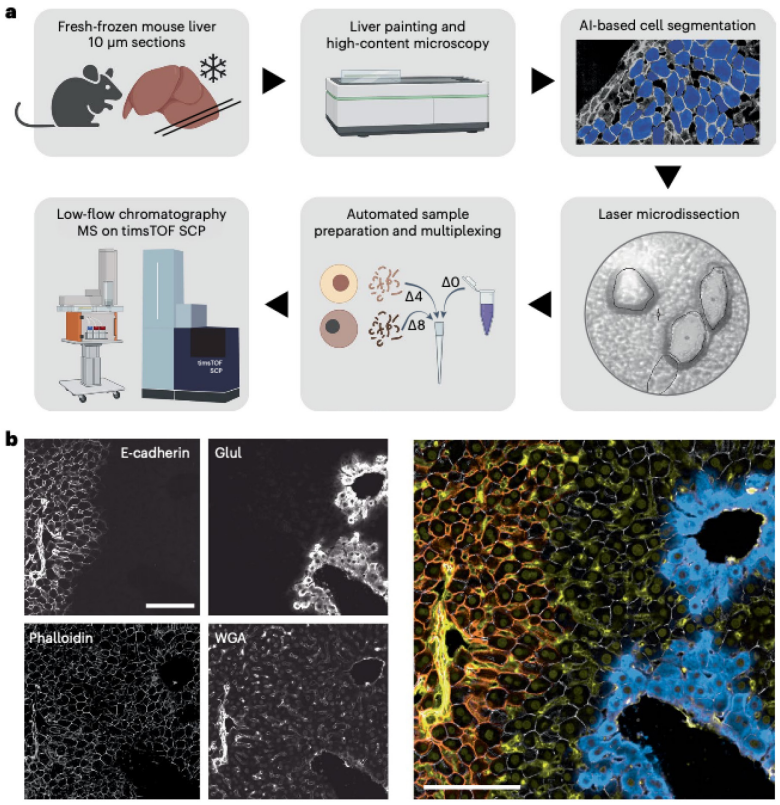



**Fig. 1 | Isolation and characterization of individual hepatocyte shapes in situ. a**, The scDVP workflow comprised embedding of fresh mouse liver tissue, staining and high-content microscopy, AI-guided hepatocyte segmentation, cutting and sorting of cells on a laser microdissection microscope, and peptide preparation with or without dimethyl labeling. The Δ0 channel contains the reference proteome and Δ4 and Δ8 contain two individual samples, which are all analyzed by ultra-high-sensitivity mass spectrometry. Created with BioRender.com. **b**, Liver painting with four stains. Left: E-cadherin marks PV regions, glutamate-ammonia ligase (Glul) surrounds the CV, the cell segmentation marker phalloidin, and the sinusoidal and nuclear counterstain WGA. Right: false color overlay of all channels: orange, E-cadherin; yellow, WGA, gray, phalloidin; turquoise, Glul. Scale bars, 100 µm.

## Results

### Robust isolation and characterization of hepatocyte shapes

To map the proteome of mouse hepatocytes at single-cell resolution, we established a modular and automated workflow aimed at loss-less sample preparation of the initial input cell for injection into the mass spectrometer (Fig. 1a). Mice livers were embedded and immediately frozen after cardiac arrest. We fixed 10 µm sections and stained them with a one-step protocol marking PVs and CVs, the sinusoidal architecture, nuclei and cell membranes (Fig. 1b and Methods). Individual cells were segmented by deep learning as before[4], and the resulting masks transferred to a laser microdissection microscope that automatically excised and collected individual shapes in 384-well plates. Given hepatocyte sizes of 20–30 µm, one shape cut from a 10 µm section corresponds to a third or half of a hepatocyte, or approximately 250 pg of protein input, equivalent to the protein content of one HeLa cell. We automated protein extraction and digestion by reagent addition into the same plate, omitting extra transfer steps, followed by peptide separation on the Evosep system[10] and injection into a trapped ion mobility time-of-flight single-cell proteomics (timsTOF SCP) mass spectrometer (Fig. 1a).

To establish an efficient workflow, we applied our scProteomics protocol[2] and titrated the number of cells required to obtain a robust signal (Extended Data Fig. 1a). To confirm biological ground truth, we performed initial experiments on five adjacent shapes per well (corresponding to about two complete hepatocyte cell masses), cut from randomly chosen locations. With these five shapes, we reached a median depth of 1,235 proteins across 230 samples (Extended Data Fig. 1b). The results confirmed expected liver biology, for instance, by differential expression of the PV marker argininosuccinate lyase (Asl) and central Cytochrom P450 2E1 (Cyp2e1, Extended Data Fig. 1c). Using zonation anchor proteins to arrange all the samples in pseudo-space (Extended Data Fig. 1d), we characterized spatially enriched gene sets along the zonation axis. While the protein sets for electron transport chain and oxidative phosphorylation (OXPHOS) were among processes upregulated in proximity to the PV, biotransformation and oxidations by cytochrome P450 were increased proximal to the CV, providing positive controls for low-input proteomics (Extended Data Fig. 2 and Supplementary Tables S1 and S2).

### Multiplex-DIA drastically increases proteome depth

Encouraged by these spatial results, we next asked whether single shapes alone could produce deep and interpretable proteomic results. To improve sensitivity further, we adopted and optimized elements of

Article https://doi.org/10.1038/s41592-023-02007-6

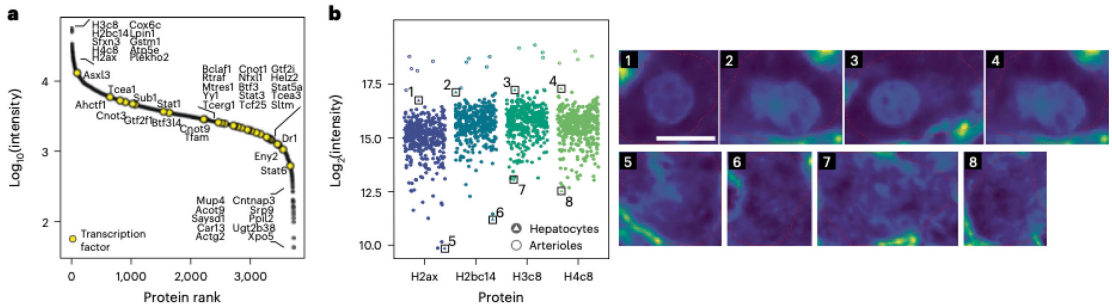



**Fig. 2 | Depth of single-shape proteomes and estimation of the nuclear compartment. a**, Unique proteins quantified in our scDVP workflow with mDIA ranked by signal intensity (two single-shape and a reference proteome channels, 31 min Evosep gradient, 15 cm column at 100 nl min$^{-1}$, dia-PASEF with optimized window design, library-dependent search in DIA-NN). Names of highest- and lowest-ranking proteins, as well as transcription factors, are indicated. **b**, Left: intensity of the top four histone proteins across all samples, including hepatocytes and quality control arteriole structures. Colors are specific for the indicated histone subunit. Right: WGA stain of cells corresponding to marked data points in the scatter plot. The color scale is signal intensity. Scale bar, 10 µm. Data from three mice were pooled.

our scProteomics workflow[11]. These include addition of the surfactant *n*-dodecyl-β-D-maltoside (DDM) to maximize peptide recovery[12], lowering the chromatographic flow rate to 100 nl min$^{-1}$ for increased ionization efficiency[3] ('Whisper gradients' on the Evosep system) and achieving higher chromatographic resolution with zero dead volume columns (IonOpticks)[13]. Most importantly, we added a labeled reference channel for multiplexed data-independent acquisition (mDIA) that decouples identification and quantification[11] (Fig. 1a).

For scDVP, we constructed a dimethyl-labeled bulk liver reference. Our robotic sample preparation setup achieved about 99% labeling efficiency in all three channels (Extended Data Fig. 3a). We co-injected 10 ng of the reference proteome together with the labeled proteomes of two single shapes at a mean size of 600 µm$^2$ (Extended Data Fig. 3b). This resulted in a doubling of identified proteins with a median number of 1,712 proteins across three biological replicates and 455 single shapes, at twice the previous throughput (Extended Data Fig. 3c). A maximum of 2,769 proteins were identified in one shape, and 3,738 unique proteins were found across all samples (Fig. 2a and Extended Data Fig. 3c). Four histone components ranked in the top ten, but we also found many transcription factors (Fig. 2a). In contrast, plasma proteins produced in hepatocyte were of medium abundance and hemoglobin subunits were not detected. This suggests little to no contamination from surrounding blood, a common issue in bulk proteomics (Extended Data Fig. 3d). The number of detected proteins correlated logarithmically with the microdissected area (Extended Data Fig. 3e), indicating that scDVP requires the highest possible MS sensitivity. Data completeness across all samples increased with median intensity per protein. Coefficients of variation were less than 50% and strongly depended on cell size and position along the zonation axis, reflecting biological heterogeneity in the data (Extended Data Fig. 3f,g). We hypothesized that the nuclear proportion in the cell slice would correlate with the intensity of these histones. Indeed, shapes with lowest histone intensities did not have any evident nuclear signal, while top intensities were in shapes with large or two nuclei. In addition to this, the intensity of the top four abundant histone proteins was highest in arterioles that we cut as technical control structures, and which are composed of more than one cell and nucleus (Fig. 2b and Extended Data Fig. 3h).

### Single-shape proteomes accurately reflect hepatocyte zonation

To test the biological validity of our proteomics data, we first reduced dimensionality in a principal component analysis (PCA), which revealed that PC1 represented the measured distance of a hepatocyte to PV and CV (Fig. 3a,b). Overlays of known liver zonation markers including Cyp2e1 and argininosuccinate lyase (Asl) showed opposite visual enrichment along PC1 (Extended Data Fig. 4a,b). In contrast, PC2 did not correlate with measured distance or hepatocyte zonation markers but rather with cytoskeletal components (Extended Data Fig. 4c,d). PC2 was also the dimension in which portal arterioles, which we excised as technical controls, separated from hepatocytes (Extended Data Fig. 4e).

We asked whether single-cell resolution provides any benefit over combining adjacent shapes. To this end, we iteratively combined shape information and averaged protein levels of cells with the same relative location along the zonation axis ('pseudo-neighbors'). Starting from the combination of as little as two shapes, PC1 continuously gained importance as measured by interquartile range and variance explained, whereas PC2 and all subsequent components dropped in explanatory value (Extended Data Fig. 5). This demonstrates that single-cell data retains subtle biological differences compared to the excision of larger areas.

On the basis of the distance ratio of PV and CV, we grouped the data into 20 spatial bins—approximately the maximum number of cells along the zonation axis. Analysis of variance (ANOVA) testing across all bins revealed that 49% of all proteins detected in at least half of the samples were significantly different between zones (false discovery rate (FDR) <0.05; Extended Data Fig. 6a and Supplementary Table S2). Zonation was also apparent after spatial sorting at the total proteome level (Fig. 3c and Supplementary Table S3) and for known hepatocyte zonation markers (Fig. 3d). Only 5.8% of these proteins were expressed equally in all zones (multiple testing-adjusted Shapiro–Wilk test, $P > 0.05$), including electron transfer flavoprotein β (Etfb), the electron acceptor in mitochondrial fatty acid β-oxidation (Extended Data Fig. 6b).

The first principal component along the zonation axis indicated that portal and periportal regions were more similar to one another than central and pericentral zones (Fig. 3b). Indeed, the spatial expression of the top ten significant zonation markers for each portal and central regions followed a hockey-stick curve from portal to central (Fig.3e), similar to Wnt-controlled transcripts in a scRNAseq dataset[6] and in line with a CV origin of Wnt signaling[14]. In contrast, this pattern was absent for the hits with the highest *P* values (least zonated hits; Extended Data Fig. 6d).

A cross-omics comparison with scRNAseq data[6] confirmed the directionality of the most prominent zonation markers (Pearson's $R = 0.97$, Extended Data Fig. 7a,b), while correlation was low across all proteins and transcripts (Pearson's $R = 0.12$). Notably, a number of proteins were regulated only in the RNA or protein dimension, or

Article https://doi.org/10.1038/s41592-023-02007-6

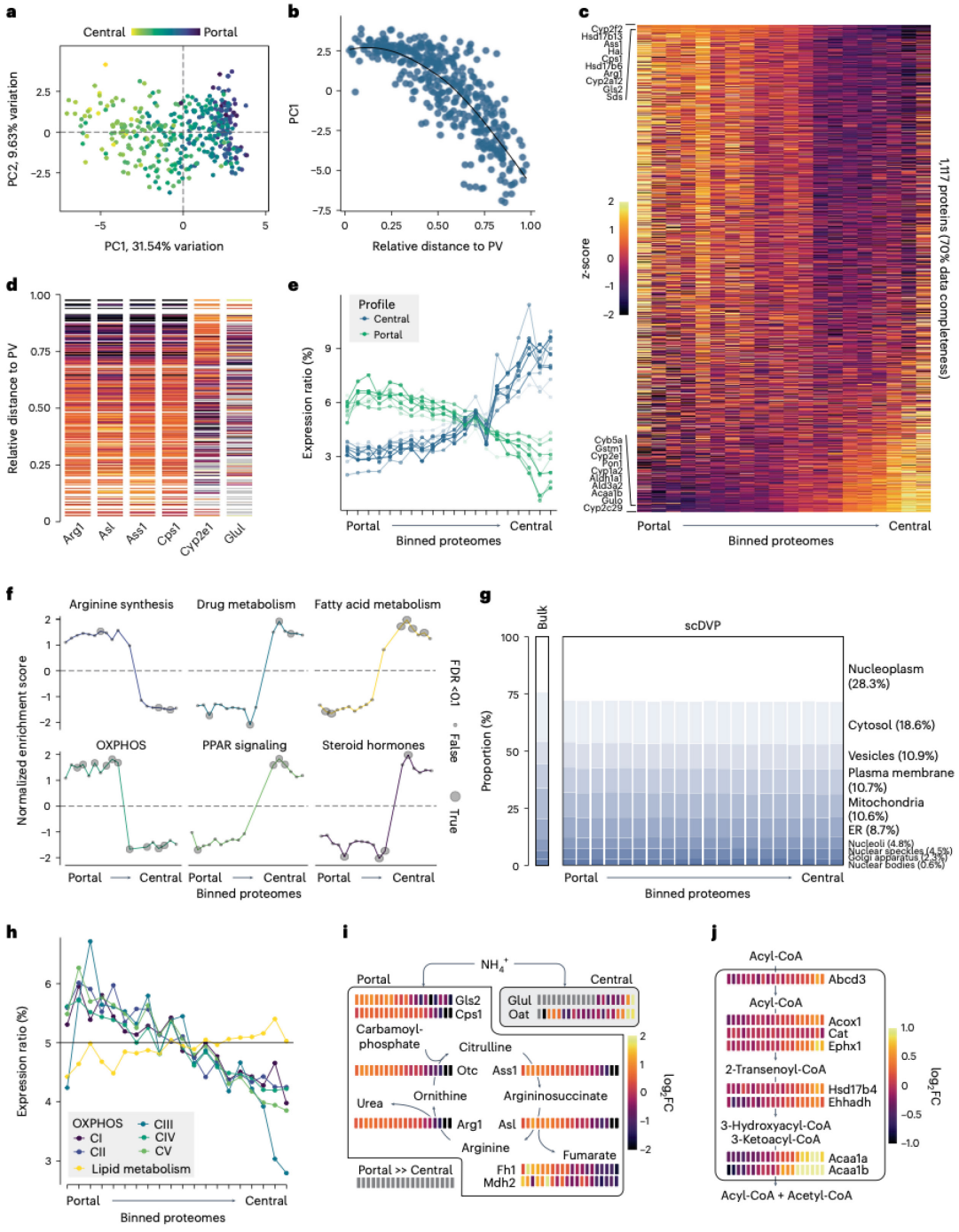


**Fig. 3 | Single-shape proteomes are accurate descriptors of zonated hepatocytes. a**, PCA of all hepatocytes. The color overlay corresponds to the ratio of measured distance PV over CV in the microscopy image. **b**, Measured distance ratio versus PC1. Relative distance of 0 is at the PV and of 1 is at the CV. Black: smoothing curve. **c**, Heat map of protein expression as *z*-score per protein across all samples. Proteins are ordered according to ANOVA fold change (FC) across 20 spatial equidistant bins, summarizing samples with a similar distance ratio to PV and CV. The ten top and bottom proteins are given. Only proteins that were detected in 70% of all samples are included. **d**, Protein expression as z-score of selected marker proteins, ordered by relative distance to PV and CV. One line is one shape measurement. Gray: protein not detected. **e**, Expression of the top 20 significant proteins in 20 spatial bins, relative to total expression from portal to central. Zonation peak at PV: positive ANOVA fold change ($n = 10$), and vice versa ($n = 10$). **f**, Selected gene sets in individual spatial bins versus all others bins, depicting normalized enrichment score after gene set enrichment analysis. Dot size: significance after multiple testing adjustment. **g**, The proportion of protein signal stratified by subcellular compartment in a bulk mouse liver proteome and the scDVP dataset. Percentages refer to mean across spatial bins in the scDVP data. **h**, Relative expression in 20 bins from PV to CV of proteins constituting mitochondrial OXPHOS components (C) I–V, and mitochondrial lipid metabolism. **i**, Levels of urea cycle and connected enzymes from portal (left) to central (right) bins as $\log_2$FC relative to median expression in the two center bins. Portal box: active in portal regions. Central box: active in central region. **j**, Levels of peroxisomal enzymes related to very-long chain fatty acid degradation, spatially resolved as in **g**. Data from three mice were pooled.

even inversely correlated (Extended Data Fig. 7c), such as dimethylglycine dehydrogenase in the choline catabolic pathway. Similarly, when we compared our data with a FACS-based hepatocyte proteome[8], we found slightly lower correlation of markers and better overlap overall (Pearson's $R$ of 0.16 versus 0.12, Extended Data Fig. 7d–f). Members of glutathione metabolism had similar spatial distribution in both datasets (Extended Data Fig. 7g). This underlines that the scDVP dataset provides orthogonal insight into liver physiology instead of merely complementing existing datasets.

Enrichment of functional protein sets across the spatial bins confirmed that arginine biosynthesis and OXPHOS were highly enriched toward the PV (Fig. 3f). When we added subcellular annotations to our dataset, we found negligible differences to a bulk mouse liver proteome for many compartments including the plasma membrane (summed intensity of 10.6% in the library versus 10.7% in this scDVP data), highlighting that laser microdissection is suitable to excise the entire shape (Fig. 3g and Supplementary Table S6). On a biological level, we found only small changes of summed organellar intensities across spatial bins (Extended Data Fig. 8 and Supplementary Table S6), namely decreasing mitochondrial and endoplasmic reticulum mass and increasing Golgi apparatus and nucleoplasm from PV to CV. When cross-mapping the scDVP data with the mitochondrial protein library Mitocarta 3.0 (ref. 15), the five complexes of OXPHOS decreased collectively by more than 25%, yet mitochondrial proteins related to fatty acid metabolism mildly increased conspatially, suggesting differential regulation within the same cellular compartment (Fig. 3h). Remarkably, these protein sets reach their spatial expression at the midpoint between PV and CV in contrast to the hockey-stick distribution of the top ten differentially expressed proteins (Fig. 3e,h), suggesting that the mitochondrial compartment is not dependent on the Wnt-signaling gradient.

The scDVP data correctly confirmed that proteins participating in ammonia fixation of the urea cycle were highly expressed in portal regions, while those involved in ammonia capture on glutamate were strongly pericentral (Fig. 3i). To our surprise, several other signaling-related pathways were also zonated including peroxisome proliferator-activated receptor (PPAR) signaling (Fig. 3f and Supplementary Table S5). This was corroborated by prominent central expression of enzymes required for peroxisomal degradation of very-long-chain fatty acids, and ω-oxidation of dicarboxylic C12 fatty acids, enriched in, for instance, coconut oil (Fig. 3j). We conclude that the spatial proteome data from single hepatocyte shapes is biologically accurate and informative, and furthermore, contains rich biological information to be mined.

### Spatial context regulates single-cell proteomes

Combining the single-shape proteomes with their inherent spatial information and staining intensities, scDVP revealed clear dependence of fluorescent intensities with the eight proteome classes established above (Fig. 4a,b).

Encouraged by the evident complementarity between extensive proteomics and spatial data, we reasoned that the microscopic image could contain sufficient information to predict the proteome. To this end, we trained a machine-learned (ML) model on 17 features to predict the proteome classes from imaging data. We grouped the training set into five proteome classes by $k$-means clustering (Extended Data Fig. 9a), and used the information in all imaging channels as predictors (Extended Data Fig. 9b). This model reached an average precision of 0.94 (Extended Data Fig. 9c,d), correctly assigning the proteome class of almost all cells. Errors occurred exclusively between spatially neighboring classes (Fig. 4c).

We tested the model performance on a new section (not used in training), from which we measured 60 single-shape proteomes. Visual inspection indicated that the predicted classes were correctly located in proximity to CV or PV, even in the presence of cutting artifacts (Fig. 4d). We used the class probabilities as weights to predict the spatial proteome, which accurately approximated overall protein intensities ($R = 0.78$ between prediction and measurement, Fig. 4e). When predicting the proteome of a larger section for all quantified proteins, the ML model correctly assigned the spatial directionality of zonation markers, as well as their expected extension into the intermediate zone (Fig. 4f). Thus, the model confirms the accuracy of measured single-shape proteomes, and is furthermore a potent predictor of spatial proteomes across any imaged areas.

## Discussion

Here, we present a single-cell spatial map of the murine liver acquired by MS-based proteomics. Our approach successfully combines microscopic imaging data with ultra-high-sensitivity proteomics, building on four major technological advances: (1) artificial intelligence (AI)-assisted segmentation and laser microdissection, (2) multiplex-DIA (mDIA), (3) low-flow gradients and (4) the ultra-high sensitivity of a timsTOF SCP mass spectrometer.

To date, MS-based scProteomics has been exclusively reported for cell suspensions. State-of-the-art workflows currently reach a proteomic depth of up to 2,000 proteins in cultured cells, with about 250 pg of cellular protein mass. This is similar to the protein material in our sliced hepatocytes, taking the section thickness of 10 µm and hepatocyte size of 20–30 µm into account. With our scDVP workflow, we achieved more than 1,700 proteins per single shape (and up to 2,700) despite working from sections that were fixed, stained, imaged and laser dissected. Laser microdissection successfully separated hepatocytes from surrounding material including blood remnants, which holds promise for smaller cell types in more complex tissue environments. The size of our shapes correlated strongly with the number of identified proteins, suggesting that scDVP is currently limited by MS sensitivity and will thus profit from continuous technical developments. We established the scDVP protocol to combine one reference channel with two single shapes (effective two-plex) and used a 40 samples per day chromatography method. This can be

Article https://doi.org/10.1038/s41592-023-02007-6

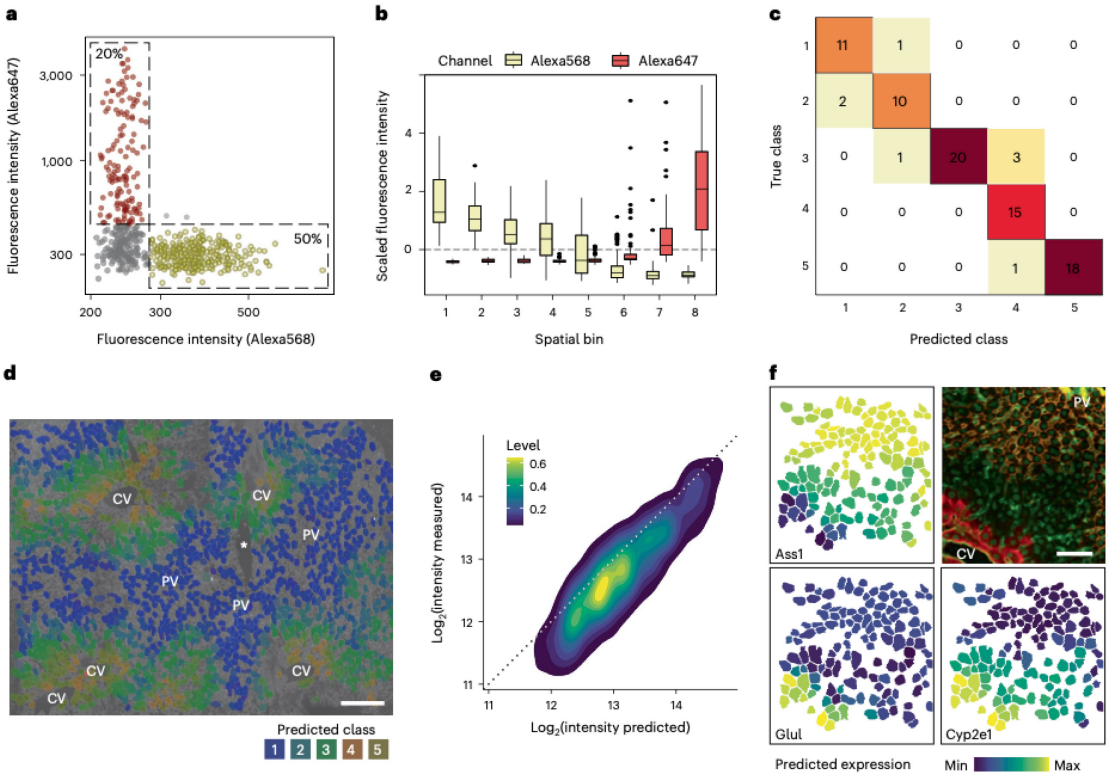



**Fig. 4 | Combining imaging and proteome data for a ML model. a**, Fluorescence intensities of Alexa568 (PV marker E-cadherin) and Alexa647 (CV marker Glul), with percentages in indicated bins. Each dot represents one shape. **b**, Intensities of the spatial markers as in **a** across eight spatial bins. The boxes are first and third quartiles, the thick line is the median, whiskers are ±1.5 interquartile range and outliers are indicated as individual points. **c**, Confusion matrix of a ML model with five classes, informed by microscopy and proteomics data. Colors scale with counts in each box. **d**, Predicted classes of segmented hepatocytes. The hue is the maximum class probability. **e**, Density plot of predicted versus measured intensities of a section excluded from machine learning ($R = 0.78$). **f**, Spatial depiction of one biological replicate with microscopy ground truth on top right, and three predictions. Orange, E-cadherin; red: Glul; green: WGA. *Sectioning artifact. Scale bar, 150 µm.

further scaled to five-plex (effective four-plex) and 80 samples per day, scaling to 320 shapes per day[11]. Given the more stable core proteome compared with single-cell transcriptomes[2] and the resulting lower required sample number, scDVP experiments encompassing a few hundred single shapes could be done in just a few days. Furthermore, because of the very low quantities and absence of proprietary reagents, marginal costs are extremely low.

Our proteomics data from single shapes correctly and accurately recapitulates hepatocyte physiology by direction, extent and spatial organization of zonation. More than half of quantified proteins were significantly different between portal and central zones, in line with scRNAseq data[6,16] and FACS-based proteomics data[8]. The fact that we detected all of the previously used markers of liver zonation[6] suggests that our proteomic depth is sufficient to integrate into other omics datasets. This became further apparent on the level of functional pathways, including signaling and disease pathways. Interestingly, peroxisomal degradation of very-long-chain fatty acids, as well as dicarboxylic C12 fatty acids, was enriched in proximity to the CV. Biochemical evidence by radiolabeling experiments support the notion that nonmitochondrial fatty acid oxidation localizes to pericentral regions[17]. We report an almost linear decrease of mitochondrial mass and OXPHOS subunits along the zonation axis. This is in line with intravital microscopy data showing decreasing mitochondrial membrane potential[18]. A rhythmic expression pattern has been previously shown for a large number of liver transcripts and proteins[16,19]. While we have not covered the temporal aspect here, the scDVP approach could contribute to such studies by adding a spatial dimension.

In the previously described DVP workflow, we used pools of cells combined on the basis of common features, such as the expression intensity of already known markers, or morphology[4]. This approach allows a deep, rapid and robust proteome characterization that accurately represents the underlying biology. By analyzing single cellular shapes without prior assumptions, scDVP now removes the dependency on established markers or features. This makes it a promising approach in heterogeneous tissues with partially or not defined subtypes of cells, such as in many tumor tissues. Moreover, scDVP can be a method of choice to map proteomic disturbances along gradients of, for instance, signaling factors, nutrients or gases, and in physiological settings that may create impediments for other omics methods, for instance, in extracellular fibrotic scars.

Our results demonstrate that the central challenge of scDVP is the sensitivity of the overall workflow. Although we have here reduced the area

required for laser microdissection by 100-fold compared with our initial DVP report[4], we note that one excised hepatocyte shape contains approximately ten times more protein than the smallest cells of interest, such as typical resting lymphocytes[20]. While the required sensitivity is being developed, the original DVP approach using pooled cells of the same type is a powerful tool for this kind of problems. We also note recent success in drastically scaling down DVP for formalin-fixed and paraffin-embedded samples, which are readily available in many clinical settings[21].

There have been advances in the quantification of posttranslational modifications from ultra-low-input material, such as from 1 µg down to the material corresponding to single cell, for instance in the enrichment protocol µPhos[22]. In combination with scDVP, this holds promise for single cells, although the biological information in single-cell phosphoproteomics data would currently be limited to a few high-abundance proteins with high modification stoichiometries. Subtle signaling events, such as the liver-dominant Wnt signaling, will require additional technological developments for in-depth biological description of signaling in single cells by MS.

We have shown that single-cell data can be used to train an accurate ML model that predicts the proteome class from visual information only. Evidence suggests that morphological features such as nuclear vacuolation and texture associate with zonation, and can even serve as a progression and stratification marker of nonalcoholic fatty liver disease[23]. Combining such easily available features and extensive proteomic sampling can clearly lead to higher precision of the predictive models. Transfer learning might then extend the approach to many new areas, as already shown for single-cell transcriptomics data[24]. The modular nature of scDVP, especially the open format from laser microdissection to 384-well plates for sample preparation, makes it widely applicable and also compatible with other spatial omics technologies such as spatial transcriptomics, epigenomics[25] or multiplexed imaging. In conclusion, scDVP is a powerful tool for basic discovery science, working in concert with DVP and other omics methods to enrich spatial workflows.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-023-02007-6.

## References


1. Kelly, R. T. Single-cell proteomics: progress and prospects. *Mol. Cell Proteomics* **19**, 1739–1748 (2020).
2. Brunner, A.-D. et al. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **18**, e10798 (2022).
3. Derks, J. et al. Increasing the throughput of sensitive proteomics by plexDIA. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01389-w (2022).
4. Mund, A. et al. Deep visual proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* **40**, 1231–1240 (2022).
5. Cunningham, R. P. & Porat-Shliom, N. Liver zonation—revisiting old questions with new technologies. *Front. Physiol.* **12**, 732929 (2021).
6. Halpern, K. B. et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 352–356 (2017).
7. Aizarani, N. et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* **572**, 199–204 (2019).
8. Ben-Moshe, S. et al. Spatial sorting enables comprehensive characterization of liver zonation. *Nat. Metab.* **1**, 899–911 (2019).
9. Guilliams, M. et al. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* **185**, 379–396.e38 (2022).
10. Bache, N. et al. A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Mol. Cell Proteomics* **17**, 2284–2296 (2018).
11. Thielert, M. et al. Robust dimethyl-based multiplex-DIA workflow doubles single-cell proteome depth via a reference channel. *Mol. Syst. Biol.* https://doi.org/10.15252/msb.202211503 (2023).
12. Tsai, C.-F. et al. Surfactant-assisted one-pot sample preparation for label-free single-cell proteomics. *Commun. Biol.* **4**, 1–12 (2021).
13. Sandow, J. J., Infusini, G., Dagley, L. F., Larsen, R. & Webb, A. I. Simplified high-throughput methods for deep proteome analysis on the timsTOF Pro. Preprint at *bioRxiv* https://doi.org/10.1101/657908 (2021).
14. Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. Self-renewing diploid Axin2+ cells fuel homeostatic renewal of the liver. *Nature* **524**, 180–185 (2015).
15. Rath, S. et al. MitoCarta3.0: an updated mitochondrial proteome now with suborganelle localization and pathway annotations. *Nucleic Acids Res.* **49**, D1541–D1547 (2021).
16. Droin, C. et al. Space–time logic of liver gene expression at sublobular scale. *Nat. Metab.* **3**, 43–58 (2021).
17. Guzmán, M., Bijleveld, C. & Geelen, M. J. H. Flexibility of zonation of fatty acid oxidation in rat liver. *Biochem. J.* **311**, 853–860 (1995).
18. Kang, S. W. S. et al. A spatial map of hepatic mitochondria uncovers functional heterogeneity shaped by nutrient-sensing signaling. Preprint at *bioRxiv* https://doi.org/10.1101/2023.04.13.536717 (2023).
19. Robles, M. S., Humphrey, S. J. & Mann, M. Phosphorylation Is a central mechanism for circadian control of metabolism and physiology. *Cell Metab.* **25**, 118–127 (2017).
20. Rosenberger, F. A., Thielert, M. & Mann, M. Making single-cell proteomics biologically relevant. *Nat. Methods* **20**, 320–323 (2023).
21. Makhmut, A. et al. A framework for ultra-low input spatial tissue proteomics. Preprint *bioRxiv* at https://doi.org/10.1101/2023.05.13.540426 (2023).
22. Oliinyk, D., Will, A., Schneidmadel, F. R., Humphrey, S. J. & Meier, F. µPhos: a scalable sensitive platform for functional phosphoproteomics. Preprint at *bioRxiv* https://doi.org/10.1101/2023.04.04.535617 (2023).
23. Segovia-Miranda, F. et al. Three-dimensional spatially resolved geometrical and functional models of human liver tissue reveal new aspects of NAFLD progression. *Nat. Med.* **25**, 1885–1893 (2019).
24. Lotfollahi, M., Wolf, F. A. & Theis, F. J. scGen predicts single-cell perturbation responses. *Nat. Methods* **16**, 715–721 (2019).
25. Deng, Y. et al. Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* **609**, 375–383 (2022).


**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023

Article https://doi.org/10.1038/s41592-023-02007-6

## Methods

### Mouse experiments and organ collection

Pathogen-free male and female 10-week-old C57BL/6J-rj mice were purchased from Janvier and maintained at the appropriate biosafety level under constant temperature and humidity conditions with a 12 h light cycle. Animals were allowed food and water ad libitum. All experiments were performed on 12- or 13-week-old wild-type mice. These were killed by cervical dislocation, and the liver was rapidly excised through a ventral opening of the peritoneum. The organ was rinsed in cold phosphate-buffered saline (PBS), and the left lateral lobe was divided into three pieces. For this study, the distal-caudal quarter was embedded in optimal cutting temperature medium (Sakura Finetek) in 15 mm disposable cryomolds (Sakura Finetek) and frozen in isopentane that was precooled to dew point in liquid nitrogen. Fully solidified blocks were transferred to dry ice, and then to a −80 °C freezer until further processing. Animal handling and organ withdrawal were performed in accordance with the governmental and international animal welfare guidelines and ethical oversight by the local government for the administrative region of Upper Bavaria (Germany), registered under ROB-55.2-2532.Vet_02-16-208.

### Immunofluorescence staining

Two micrometer polyethylene naphthalate membrane slides were pretreated by ultraviolet ionization for 1 h at 254 nm. Without delay, slides were consecutively washed for 5 min each in 350 ml acetone and 7 ml VECTABOND reagent to 350 ml with acetone, and then washed in ultrapure water for 30 s before drying in a gentle nitrogen air flow. For sectioning, tissue blocks were transferred to a cryostat (Leica CM3050) at −18 °C chamber and −15 °C object temperature, and left to equilibrate for 30 min. Blocks were then trimmed, and final sections were cut at 10 µm thickness with a disposable high-profile blade (Leica 818). Frozen sections were transferred to pretreated, cold polyethylene naphthalate membrane slides, and melted for less than 5 s on a room temperature surface. The sections were then fixed in prewarmed 4% paraformaldehyde in PBS at 37 °C, then in 95% ethanol at room temperature and finally again in 4% paraformaldehyde in PBS at 37 °C. Slides were rinsed in PBS and left in 5% BSA–PBS blocking solution for 1 h until staining. Sections were stained for 1 h at 37 °C in a humid and dark chamber with 200 µl of a one-step liver painting in 1% BSA: 1:300 phalloidin coupled to Atto-425 (Sigma 66939), 1:200 wheat germ agglutinin (WGA) coupled to Alexa Fluor 488 (Invitrogen, W11261), 1:100 anti-E-cadherin coupled to Alexa Fluor 555 (BD 560064), anti-glutamine synthase (Abcam, ab176562) and 1:500 anti-rabbit nanobody coupled to Alexa Fluor 647 (Chromotek srbAF647-1-100). Slides were washed three times for 2 min in PBS in the dark, and mounted with 21 FL ProLong Diamond mounting medium (Invitrogen, P36961) and a 22 × 22 mm #1.5 coverslip. Slides were stored until imaging in 50 ml tubes with desiccating material at 4 °C.

### High-content imaging

Sections were imaged on an OperaPhenix high-content microscope, controlled with Harmony v4.9 software, at 40× magnification, with binning of two and a per tile overlap of 10%. At excitation wavelengths of 425 nm, 555 nm and 647 nm, an 80% laser intensity were used at an illumination time of 100 ms, while in the 488 nm (CFP) channel, 20% and 20 ms were used. E-cadherin and glutamine synthetase were imaged simultaneously, while phalloidin and WGA were imaged consecutively.

### Image postprocessing

Acquired images were flat-field corrected using the Harmony software. Stitching of image tiles was performed using the ashlar Python API (application programing interface)[26] with a max shift value of 30. Stitched images were exported as .tif files and imported into the Biological Image Analysis Software (BIAS, Single-Cell Technologies Ltd.)[4] with the packaged import tool. In BIAS, large tif images were first retiled to 1,024 × 1,024 px at an overlap of 5%. Hepatocytes were identified with a deep neural network for histological cytoplasm segmentation on the basis of phalloidin staining at 1.2 input spatial scaling, 40% detection confidence and 30% contour confidence. Only contours between 135 µm$^2$ and 1,350 µm$^2$ were taken into consideration, and no further exclusion criteria were applied. After removal of duplicates and false identifications by supervised machine learning, contours were exported without additional shape offset together with three calibration points that were chosen at characteristic tissue positions. Contour outlines were simplified by removing 99% of data points. For five-shape proteomes, directly adjacent shapes forming a pentagon-like structure were manually picked. Single shapes were randomly picked and every 15–25th shape was assigned to adjacent wells in a 384-well plate. Arterioles were manually assigned based on WGA signal, ellipticity and proximity to the E-cadherin positive PV.

### Laser microdissection

Contour outlines were imported after reference point alignment, and shapes were cut by laser microdissection with the LMD7 (Leica) in a semi-automated manner at the following settings: power 59, aperture 1, speed 60, middle pulse count 1, final pulse −1, head current 48–52%, pulse frequency 3,282 and offset 100. For the five-shape experiment, the microscope was controlled with LMD v8.2, with which five directly adjacent shapes were sorted into a low-binding 384-well plate (Eppendorf 0030129547) with one empty well between samples. Single shapes were cut and sorted with the software LMD beta 10 after calibration of the gravitational stage shift into 384-well plates into all wells, leaving the outermost rows and columns empty. A 'wind shield' plate was used on top of the sample stage to avoid erroneous sorting. Plates were sealed, centrifuged at 1,000*g* for 5 min and then frozen at −20 °C until further processing.

### Reference peptide preparation for five-shape and single-shape proteomes

The proximal part of two biologically independent lobes of the same mice as in the scDVP experiments was used to construct a library. The tissue embedded in optimal cutting temperature medium was removed from −80 °C and directly disintegrated in a plastic bag with a manual tissue homogenizer (rubber hammer). Pieces of approximately 1 mm$^3$ were transferred into a low-binding 96-well plate with magnets (BeatBox Tissue Kit, Preomics), covered with 50 µl of 60 mM triethylammonium bicarbonate buffer with 10% acetonitrile (ACN; lysis buffer), and lysed in a BeatBox (Preomics) at standard settings for 10 min. Samples were then boiled at 96 °C for 20 min, transferred to 1.5 ml low-binding tubes, filled up to 500 µl with lysis buffer and sonicated for five times 30 s on/off cycles. After centrifugation at 2,000*g* for 1 min, the protein concentration in the supernatant was estimated on a NanoDrop, and LysC and trypsin were added at a protein-to-enzyme ratio of 1:100. After digest for 20 h, samples were acidified to 1% trifluoroacetic acid (TFA), centrifuged at 3,000*g* for 10 min at room temperature, and dried in a SpeedVac for 30 min. Digest was filled to 1 ml with buffer A (0.1% formic acid (FA)), and desalted on C-18 columns (Waters WAT036820). They were activated and equilibrated with 2 ml of methanol, 2 ml of buffer B (100% ACN, 0.1% FA) and 2 ml of buffer A, before sample loading. Peptides were washed with buffer A two times, eluted in 80% ACN with 0.2% FA and dried down.

### Library fractionation for five-shape proteomes

Peptides were reconstituted in 18 µl buffer A* (0.1% FA, 2% ACN) fractionated on a 30-cm-long 1.9 µm ReproSil C-18 column (PepSep) using a 100 min high-pH gradient. The concentration of buffer B was increased from 3% to 30% in 45 min, to 40% in 12 min, to 60% in 5 min, to 95% in 10 min, kept constant for 10 min, reduced to 5% in 10 min and kept constant for 8 min. The eluted peptides were automatically collected into 48 fractions with a concatenation time of 90 s per fraction. The

fractions were dried in a SpeedVac, reconstituted in 0.1% FA and directly loaded onto Evotips.

### Labeling of single-shape reference proteome

Peptides were reconstituted to 0.125 µg µl$^{-1}$ in 60 mM triethylammonium bicarbonate buffer with 10% ACN, pH 8.5. The peptides were then dimethyl labeled with 0.15% light formaldehyde ($CH_2O$) and 0.023 M light sodium cyanoborohydrate ($NaBH_3CN$) for 1 h at room temperature, quenched with 0.13% ammonia and acidified to 1% TFA. After drying in a SpeedVac, pellets were reconstituted in 100 µl buffer A, and desalted via 5 µg C-18 columns on an AssayMap (Agilent) following the standard protocol. The resulting reference peptides were dried, and reconstituted to 1 ng µl$^{-1}$ in buffer A.

### Peptide preparation of single shapes and dimethyl labeling for multiplexing

Peptides were prepared semi-automated on a Bravo pipetting robot (Agilent), similarly to as described previously[11]. During each incubation step, plates were tightly sealed with two stacked aluminum lids to avoid evaporation (Thermo Fisher Scientific, AB0626). For this, plates were removed from the freezer and centrifuged. The wells were then washed on the robot with 28 µl of 100% ACN and dried in a SpeedVac (Eppendorf) at 45 °C for 20 min. Shapes were then resuspended in 6 µl of 60 mM triethylammonium bicarbonate buffer (pH 8.5, Sigma) with 0.013% DDM (Sigma), and cooked for 30 min at 95 °C in a PCR cycler at a lid temperature of 110 °C. After addition of 1 µl of 80% ACN (final concentration 10%), samples were incubated for another 30 min at 75 °C, cooled briefly, and 1 µl with 4 ng LysC and 6 ng trypsin was added. The samples were digested for 18 h, and then 1 µl of either intermediate ($CD_2O$) or heavy formaldehyde ($^{13}CD_2O$) was added to a final concentration of 0.15%. Without delay, either light ($NaBH_3CN$) or heavy ($NaBD_3CN$) sodium cyanoborohydrate were added to 0.023 M to retrieve Δ4 and Δ8 dimethyl-labeled single-shape samples. The sealed plate was then incubated at room temperature for 1 h, and the reaction was quenched to 0.13% ammonia and acidified to 1% TFA.

### Peptide loading onto C-18 tips

C-18 tips (Evotip Pure, EvoSep) were activated for 5 min in 1-propoanl, washed twice with 50 µl of buffer B (99.9% ACN, 0.1% FA), activated for 5 min in 1-propanol, and washed twice with 50 µl buffer A (0.1% formic acid). Single-shape samples were then loaded automatically with the Agilent Bravo robot into 30 µl buffer in the tip that was spun through the C-18 disk for a few seconds only. For loading, 10 µl of 1 ng µl$^{-1}$ reference peptides (Δ0) were pipetted first, followed by Δ4, and Δ8 samples with the same tip. Wells were rinsed with 15 µl buffer A that was also loaded onto the tip. After peptide binding, the disk was further washed once with 50 µl buffer A and then overlayed with 150 µl buffer A. All centrifugation steps were performed at 700*g* for 1 min, except sample loading for 2 min.

For five-shape proteomes, plates were treated as above, with the exception of lysis in 4.5 µl 60 mM triethylammonium bicarbonate buffer, pH 8.5 without DDM, and consecutive addition of 1 µl LysC and 1.5 µl trypsin to achieve the same digestion volume as above. Five-shape samples were not dimethyl labeled and multiplexed, but acidified directly after digest, and loaded manually onto Evotips following the protocol described above.

### LC–MS/MS analysis of five shapes

Samples were measured with the Evosep One LC system (EvoSep) coupled to a timsTOF SCP mass spectrometer (Bruker Daltonics). The 30 samples per day method was used with the Evosep Performance column 15 cm, 150 µm ID (EV1137 EvoSep) at 40 °C inside a nanoelectrospray ion source (Bruker Daltonics) with a 10 µm emitter (ZDV Sprayer 10, Bruker Daltonics). The mobile phases were 0.1% FA in liquid chromatography (LC)–MS-grade water (buffer A) and 99.9% ACN/0.1% FA (buffer B). We used a dia-PASEF method with 16 dia-PASEF scans separated into four ion mobility windows per scan covering an $m/z$ range from 400 to 1,200 by 25 Th windows and an ion mobility range from 0.6 to 1.6 V s cm$^{-2}$ ('standard scheme'[27]). The mass spectrometer was operated in high sensitivity mode, with an accumulation and ramp time at 100 ms, capillary voltage set to 1,750 V and the collision energy as a linear ramp from 20 eV at $1/K_0$ = 0.6 V s cm$^{-2}$ to 59 eV at $1/K_0$ = 1.6 V s cm$^{-2}$.

### LC–MS/MS analysis of single shapes

Samples were measured with the Evosep One LC system (EvoSep) coupled to a timsTOF SCP mass spectrometer (Bruker Daltonics). The Whisper40 samples per day method was used with the Aurora Elite CSI third generation 15 cm and 75 µm ID (AUR3-15075C18-CS IonOpticks, Australia) at 50 °C inside a nanoelectrospray ion source (Bruker Daltonics). The mobile phases were 0.1% formic acid in LC–MS-grade water (buffer A) and 99.9% ACN/0.1% FA (buffer B). The timsTOF SCP was operated with an optimal dia-PASEF method generated with our Python tool py_diAID[28]. The method contained eight dia-PASEF scans with variable width and two ion mobility windows per dia-PASEF scan, covering an $m/z$ from 300 to 1,200 and an ion mobility range from 0.7 to 1.3 V s cm$^{-2}$, as previously used on the same gradient and similar input material amount[11]. The mass spectrometer was operated in high sensitivity mode, with an accumulation and ramp time at 100 ms, capillary voltage set to 1,400 V and the collision energy as a linear ramp from 20 eV at $1/K_0$ = 0.6 V s cm$^{-2}$ to 59 eV at $1/K_0$ = 1.6 V s cm$^{-2}$.

The labeling efficiency was accessed on the same LC–MS/MS in data-dependent acquisition (dda)-PASEF mode with ten PASEF scans per topN acquisition cycle. Singly charged precursors were excluded by their position in the $m/z$-ion mobility plane using a polygon shape, and precursor signals over an intensity threshold of 1,000 arbitrary units were picked for fragmentation. Precursors were isolated with a 2 Th window below $m/z$ 700 and 3 Th above, as well as actively excluded for 0.4 min when reaching a target intensity of 20,000 arbitrary units. All spectra were acquired within a $m/z$ range of 100–1,700. All other settings were kept as described before.

### Spectral library generation

The spectral library was generated on five dda-PASEF single shots from 50 ng mouse reference peptide, using the same chromatography method as above. Spectra were search with FragPipe v18.0 (ref. 29) using MSFragger v3.5, Philosopher v4.4.0 and EasyPQP v0.1.32 against a mouse FASTA reference file with 55,319 entries used throughout this study, excluding 50% decoys. Standard settings of the DIA_SpecLib_Quant workflow were used with the following exceptions: N-terminal and lysine mass shift of 28.0313 Da were set as fixed modifications, and methionine oxidation as variable modification. Carbamidomethylation was unselected as samples were not reduced and alkylated. One missed cleavage was accepted. The precursor charge ranged from 2 to 4. The peptide mass range was set to 300–1,800, and peptide length from 7 to 30. For DIA-NN compatibility, the column 'FragmentLossType' was removed in the output library file.

### Spectral search

All 263 files were search together in DIA-NN (version 1.8.1) (ref. 30) against the above-generated library, using a mass and MS1 mass accuracy of 15.0, scan windows of 9, and activated isotopologues, Match-between-Runs (MBR), heuristic protein inference and no shared spectra, in single-pass mode. Proteins were inferred from genes. Library generation was set as 'IDs, RT & IM profiling', and 'Robust LC (high precision)' as the quantification strategy. Dimethyl labeling at N-termini and lysins was set as fixed modification at 28.0313 Da, and Δ4 or Δ8 were spaced 4.0251 Da or 8.0444 Da from the reference Δ0 ({–fixed-mod Dimethyl, 28.0313, nK} and {–channels Dimethyl, 0, nK, 0:0; Dimethyl, 4, nK, 4.0251:4.0251; Dimethyl, 8, nK, 8.0444:8.0444}). Additional settings were {–original-mods}, {–peak-translation}, {–ms1-isotope-quant}, {–report-lib-info}.

### Data analysis

(1) RefQuant: to determine the quantities of the precursors in the DIA-NN report.tsv file, we utilized the Python-based RefQuant algorithm[11]. In brief, RefQuant determines the ratio between target and reference channels for each individual fragment ion and MS1 peak that is available. This gives a collection of ratios from which RefQuant estimates a likely overall ratio between target and reference. The ratio between target and reference was rescaled by the median reference intensity over all runs for the given precursor, thereby giving a meaningful intensity value for the target channel. The RefQuant quantification matrix was filtered for 'Lib.PG.Q.Value' <0.01, 'Q.value' <0.01 and 'Channel.Q.Value' <0.15 and was then collapsed to protein groups using the MaxLFQ algorithm[31] as implemented in the R package iq (version 1.9.6) (ref. 32) with median normalization turned off.

(2) Sample filtering and normalization: protein group data were then further analyzed in R v4.2.1 operating in RStudio v2022.07.2. Samples were excluded if the number of detected proteins was below 1.5 or above 3 s.d. from the sample identification median, or within (806, 3,362) identified proteins, resulting into a dropout of 8.9% (41 of 459 samples). Four samples were removed due to their outlier position on the PCA, see Supplementary Table S3. Eight samples were removed due to their cell sizes larger than the BIAS cutoff of 1,350 µm$^2$. This resulted in 406 included samples, of which 400 were hepatocytes and 6 endothelial structures for validation. After sample filtering, data was median normalized to a set of proteins that were quantified across all samples (175 proteins quantified in 100% of included samples; Supplementary Table S3), thus correcting for the dependence of protein numbers on shape size. For hepatocyte specific analysis, the arteriole proteomes were removed before normalization.

(3) Figure generation: we chose 20 classes for all comparative spatial analyses as this matches the approximate number of cells from PV to CV, and five classes for machine learning as a compromise between meaningful separation and having enough samples per class. Proteome bins were based on an equidistant split of PC1, distance classes accordingly on a split of PV over CV distance ratios, and applied as indicated. PCA were performed with the PCAtools v2.8.0 package. Limma v3.52.4 was used for statistical testing across proteome bins on a 50%-complete protein data matrix. 'FDR' was applied for multiple testing correction, and an FDR cutoff of 5% was considered significant. Heat mapping was performed with pheatmap 1.0.12, the completeness of the data matrix is indicated in the figure legends. Proteomic gene set enrichment analyses were done with WebGestalt 2019 (ref. 33) in an R environment using Kyoto Encyclopedia of Genes and Genomes metabolic pathways or Wikipathway as functional library and an FDR threshold for reporting of 1. Significance was defined as FDR <10%, and normalized enrichment scores are reported here. Subcellular localization and mitochondrial functional protein sets were retrieved from mouse Mitocarta 3.0 (ref. 15). Urea cycle and peroxisomal fatty acid degradation proteins were manually curated. Normality was assessed with Shapiro–Wilk's test, and *P* values were corrected for multiple testing and expressed as FDR. Spatial data from xml files was plotted with the package sf v1.0-9. For comparisons to scRNAseq data, the dataset of Halpern et al.[6] was used, for which we binned the proteome data into nine equidistant spatial bins as described above. We used the dataset by Ben-Moshe et al.[8] to compare scDVP data with FACS-based proteomics data, binning our samples into eight equidistant spatial bins.

### Image processing

Image data analysis was done in Python (3.8.11). Image shapes were extracted from the stitched tiles using Pillow (9.0.0). For each shape, the bounding box was calculated by taking the floor and ceiling of each shape coordinate and taking the maximum and minimum in *x* and *y*. The bounding rectangle was used to crop out the respective region of interest of the image. For image with offset extraction, the center of each bounding rectangle was calculated and rounded to the next integer. An offset of 1,000 was added to each direction to additionally capture the surrounding environment, and the bounding box was highlighted. For composite images, each image was exported per channel with matplotlib (3.5.1), reloaded, merged with NumPy (1.4.2) and saved again. ImageJ was used to manually measure the distance of a shape to its proximal PV and CV.

### Machine learning

For each shape and in all four channels (cyan fluorescent protein, Alexa488, Alexa568 and Alexa647), the mean, median, minimum and maximum intensity of each bounding box were calculated, as well as the shape area. This feature list was saved with pandas (1.22.3). Proteomics data were clustered with a *k*-means algorithm into five clusters. Next, we used a supervised learning approach to classify the proteomic clusters based on the feature list. The training was performed using the scikit-learn package (1.0.2). Data ($n = 408$) were randomly split in train and test datasets (split of 0.2). For classification, we used a RandomForestClassifier (n_estimators=200) and achieved a testing accuracy of 0.90. To export probabilities, we used the predict_proba functionality of RandomForest. Diagnostic plots were generated using the Yellowbrick package (1.5). The random state was set to 23 for train/test-split and RandomForestClassifier.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[34] partner repository with the ID PXD038699. Imaging data has been deposited to BioImages[35] with the accession number S-BIAD596.

## Code availability

The R and Python code used to produce the figures can be downloaded from the Mann lab Github repository via https://github.com/MannLabs/single-cell-DVP.

## References


26. Muhlich, J. L. et al. Stitching and registering highly multiplexed whole-slide images of tissues and tumors using ASHLAR. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btac544 (2022).
27. Meier, F. et al. diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nat. Methods* **17**, 1229–1236 (2020).
28. Skowronek, P. et al. Rapid and in-depth coverage of the (phospho-)proteome with deep libraries and optimal window design for dia-PASEF. *Mol. Cell. Proteomics* **21**, 100279 (2022).
29. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513–520 (2017).
30. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).

31. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ*. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
32. Pham, T. V., Henneman, A. A. & Jimenez, C. R. iq: an R package to estimate relative protein abundances from ion quantification in DIA-MS-based proteomics. *Bioinformatics* **36**, 2611–2613 (2020).
33. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
34. Perez-Riverol, Y. et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).
35. Hartley, M. et al. The BioImage Archive—building a home for life-sciences microscopy data. *J. Mol. Biol.* **434**, 167505 (2022).


## Acknowledgements

We thank our colleagues at the Department of Proteomics and Signal Transduction at the Max Planck Institute of Biochemistry as well as our colleagues at the Center for Proteome Research in Copenhagen for their input and support. We are particularly grateful for input and help from I. Bludau, A. Brunner and L. Zeitler. We thank Peter H. and Single-Cell Technologies Ltd. for their technical support. F.A.R. is an EMBO postdoctoral fellow (ALTF 399-2021). S.C.M. is a PhD fellow of the Boehringer Ingelheim Fonds. J.G.-S. has received funding from the European Respiratory Society and the European Union's H2020 research and innovation program under the Marie Sklodowska-Curie RESPIRE4 grant agreement no. 847462. This study has been supported by the Horizon-2020 under the MICROB-PREDICT program (M.M., no. 825694) and ISLET (M.M., no. 874839), by the Max Planck Society for Advancement of Science (M.M.), by the Chan Zuckerberg Initiative (M.M., CZF2019-002448) and by grants from the Novo Nordisk Foundation, Denmark (M.M., grant agreements NNF14CC0001 and NNF15CC0001). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Author contributions

F.A.R., M.T. and M.M. conceptualized the scDVP workflow. M.T., F.A.R., P.S. and M.W. acquired MS data. F.A.R., M.T.S., L.S., A. Metousis and K.M. performed data analysis. M.T.S. trained the ML model. C.A. developed quantification software. S.C.M. optimized the high-content microscopy pipeline. F.A.R., M.T., L.S., A. Metousis, S.C.M., E.R., T.M.N. and A. Mund developed and optimized the experimental scDVP workflow. J.G.-S., A.S. and H.B.S. provided mouse samples. F.A.R., M.T.S., P.S. and T.M.N. curated data. M.M. and F.A.R. supervised the project. F.A.R. and M.M. wrote the original manuscript draft. All authors read, revised and approved the manuscript.

## Funding

Open access funding provided by Max Planck Society

## Competing interests

M.M. is an indirect investor in Evosep. All other authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at https://doi.org/10.1038/s41592-023-02007-6.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41592-023-02007-6.

**Correspondence and requests for materials** should be addressed to Matthias Mann.

**Peer review information** *Nature Methods* thanks Albert Heck, Shalev Itzkovitz, and Sinem Saka for their contribution to the peer review of this work. Primary Handling Editor: Arunima Singh, in collaboration with the *Nature Methods* team. Peer reviewer reports are available.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Article 8. Spatial proteomics in three-dimensional intact specimens.**

*Cell*, 2022 (185, 26, 5040-5058); DOI: 10.1016/j.cell.2022.11.021

Harsharan Singh Bhatia[1,2,‡], Andreas-David Brunner[3,4,‡], Furkan Öztürk[1,‡], Saketh Kapoor[1], Zhouyi Rong[1,2,13], Hongcheng Mai[1,2,13], Marvin Thielert[3], Mayar Ali[1,5], Rami Al-Maskari[2,7,8], Johannes Christian Paetzold[1,7,8,10], Florian Kofler[7,8,9,12], Mihail Ivilinov Todorov[1,2], Muge Molbay[1,2,13], Zeynep Ilgin Kolabas[1,2,5], Moritz Negwer[1], Luciano Hoeher[1], Hanno Steinke[19], Alina Dima[7,8], Basavdatta Gupta[1], Doris Kaltenecker[2,14], Özüm Sehnaz Caliskan[15,16], Daniel Brandt[15,16], Natalie Krahmer[15,16], Stephan Müller[17,18], Stefan Frieder Lichtenthaler[5,6,17,18], Farida Hellal[1,2], Ingo Bechmann[19], Bjoern Menze[7,8,11], Fabian Theis[20,21,22], Matthias Mann[3,23,*], Ali Ertürk[1,2,5,6,*]

*[1]Insititute for Tissue Engineering and Regenerative Medicine (iTERM), Helmholtz Zentrum München, 85764 Neuherberg, Germany. [2]Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians University Munich, 81377 Munich, Germany. [3]Department for Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. [4]Boehringer Ingelheim Pharma GmbH & Co. KG, Drug Discovery Sciences, Birkendorfer Str. 65, D-88400 Biberach Riss, Germany. [5]Graduate School of Neuroscience (GSN), 82152 Munich, Germany. [6]Munich Cluster for Systems Neurology (SyNergy), 81377 Munich, Germany. [7]Center for Translational Cancer Research (TranslaTUM) of the TUM, 81675 Munich, Germany. [8]Image-Based Biomedical Modeling, Department of Informatics, Technical University of Munich, 85748 Garching, Germany. [9]Helmholtz AI, Helmholtz Zentrum München, 85764 Neuherberg, Germany. [10]Biomedical Image Analysis Group, Department of Computing, Imperial College London, London SW7 2AZ, UK. [11]Department for Quantitative Biomedicine, University of Zurich, 8006 Zurich, Switzerland. [12]Department of Neuroradiology, Klinikum rechts der Isar, 81675 Munich, Germany. [13]Munich Medical Research School (MMRS), 80336 Munich, Germany. [14]Institute for Diabetes and Cancer, Helmholz Zentrum München, 85764 Neuherberg, Germany. [15]Institute for Diabetes and Obesity, Helmholz Zentrum München, 85764 Neuherberg, Germany. [16]German Center for Diabetes Research, Helmholz Zentrum München, 85764 Neuherberg, Germany. [17]German Center for Neurodegenerative Diseases (DZNE), 81377 Munich, Germany. [18]Neuroproteomics, School of Medicine, Klinikum Rechts der Isar, Technical University of Munich, 81675 Munich, Germany. [19]Institute of Anatomy, University of Leipzig, 04109 Leipzig, Germany. [20]Institute of Computational Biology, Helmholz Zentrum München, 85764 Neuherberg, Germany. [21]TUM School of Life Sciences Weihenstephan, Technical University of Munich, 85354 Freising, Germany. [22]Department of Mathematics, Technical University of Munich, 85748 Garching, Germany. [23]NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark.*

*‡These authors contributed equally.*

**These authors contributed equally to this work as senior authors.*

With the achievements of two-dimensional (2D) spatial proteomics, we reasoned – since our body is a three-dimensional complex organism – that a three-dimensional (3D) spatial map will add further valuable information. To this end, we collaborated with the Ertürk group to combine whole-organ or whole-tissue clearing and imaging with our

sensitive MS technology, termed 3D imaging of solvent-cleared organs profiled by MS (DISCO-MS).

To enable this, we first established a proteomic sample preparation which is capable of dealing with the challenging nature of cleared tissue and which is universal across the clearing technique. We evaluated various combinations of solubilizing agents and several conditions in order to achieve high protein extraction efficiency, peptide recovery and qualitative and quantitative reproducibility. The combination of SDS-SDC buffer with multiple heating steps and acetone precipitation enabled the application of MS-based proteome analysis on cleared specimen at high depth and quantitative accuracy. Cleared tissue only differed in ‘blood microparticle’ proteins compared to fresh frozen or FFPE tissue, which is the part that gets lost during tissue clearing.

The potential of DISCO-MS was evaluated by studying mild traumatic brain injury (mTBI) in mice, Alzheimer’s disease in FAD model mice and coronary artery disease (CAD) in an intact human heart. The analyzed area was 500,000 $\mu m^3$ corresponding to about 60 cells.

mTBI causes discrete local inflammation and showed significant differences in axonal damage and repair like stathmin1 and neurocan between regions of interest in disease and control mice. In FAD mice, we characterized early- and late-stage amyloid beta (Aβ) plaques from whole brains in 3D and found involvement of vesicle fusion and transport, myelin sheath function and complement system related family members in early-stage plaque development. In CAD, we discovered proteomic changes associated with systematic gene ontology terms such as platelet aggregation, hypertrophic cardiomyopathy and blood coagulation.

Additionally, our collaborators developed a robotic extraction system (DISCO-bot) which enabled automatic non-destructive tissue extraction, since sectioning and imaging for LCM is impractical for large-volume samples at scale.

In summary, DISCO-MS is a powerful tool to obtain unbiased proteomic information on heterogeneous tissue regions with known spatial locations in 3D and intact organs. It offers the ability to benchmark mice models against human samples on molecular and anatomical levels, which can be applied to clinical samples and various biomedical research questions.

***Contribution:***

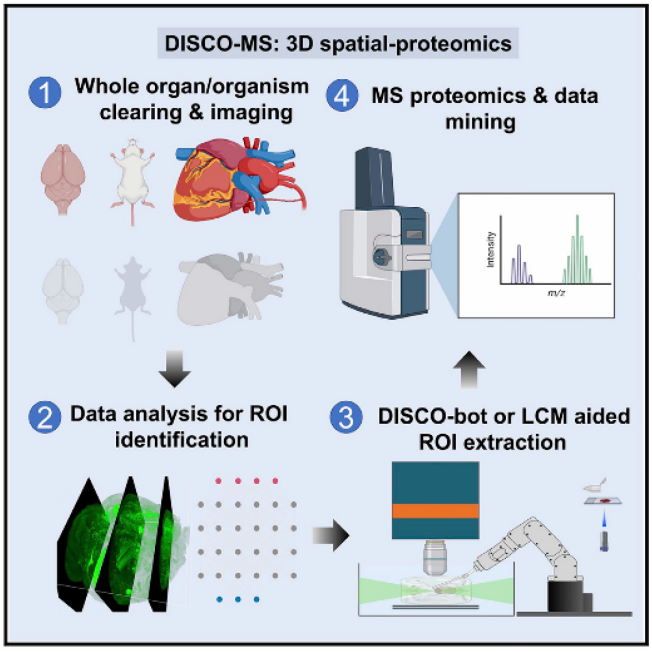
Co-authorship. Ali Ertürk, Matthias Mann, Harsharan Singh Bhatia and Andreas-David Brunner conceptualized and designed the study. The study was a great collaborative effort with the Ertürk laboratory, known for tissue engineering and regenerative medicine. Andreas-David Brunner performed most of the proteomics experiments. I did MS-based proteomics analysis for the human data.

Cell

Resource

# Spatial proteomics in three-dimensional intact specimens

## Graphical abstract




## Authors

Harsharan Singh Bhatia, Andreas-David Brunner, Furkan Öztürk, ..., Fabian Theis, Matthias Mann, Ali Ertürk

## Correspondence

mmann@biochem.mpg.de (M.M.), erturk@helmholtz-muenchen.de (A.E.)

## In brief

DISCO-MS and DISCO-bot allows for unbiased spatial proteome analysis of small tissue regions identified by panoptic imaging of large cleared samples, *in silico* reconstructions, and automated minimally invasive robotic tissue extraction.

## Highlights

- DISCO-MS is a spatial proteomics technology in optically cleared whole specimens
- DISCO-MS is aided by AI and robotics and yields proteome similar to fresh samples
- DISCO-bot-aided DISCO-MS reveals spatial immune cell heterogeneity in mouse bones
- DISCO-bot-aided DISCO-MS reveals plaque heterogeneity in human coronary artery

Bhatia et al., 2022, Cell *185*, 5040–5058
December 22, 2022 © 2022 Elsevier Inc.
https://doi.org/10.1016/j.cell.2022.11.021

CellPress

CellPress

Cell

Resource

# Spatial proteomics in three-dimensional intact specimens

Harsharan Singh Bhatia,[1,2,24] Andreas-David Brunner,[3,4,24] Furkan Öztürk,[1,24] Saketh Kapoor,[1] Zhouyi Rong,[1,2,13] Hongcheng Mai,[1,2,13] Marvin Thielert,[3] Mayar Ali,[1,5] Rami Al-Maskari,[2,7,8] Johannes Christian Paetzold,[1,7,8,10] Florian Kofler,[7,8,9,12] Mihail Ivilinov Todorov,[1,2] Muge Molbay,[1,2,13] Zeynep Ilgin Kolabas,[1,2,5] Moritz Negwer,[1] Luciano Hoeher,[1] Hanno Steinke,[19] Alina Dima,[7,8] Basavdatta Gupta,[1] Doris Kaltenecker,[2,14] Özüm Sehnaz Caliskan,[15,16] Daniel Brandt,[15,16] Natalie Krahmer,[15,16] Stephan Müller,[17,18] Stefan Frieder Lichtenthaler,[5,6,17,18] Farida Hellal,[1,2] Ingo Bechmann,[19] Bjoern Menze,[7,8,11] Fabian Theis,[20,21,22] Matthias Mann,[3,23,*] and Ali Ertürk[1,2,5,6,25,*]

[1]Insititute for Tissue Engineering and Regenerative Medicine (iTERM), Helmholtz Zentrum München, 85764 Neuherberg, Germany
[2]Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians University Munich, 81377 Munich, Germany
[3]Department for Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany
[4]Boehringer Ingelheim Pharma GmbH & Co. KG, Drug Discovery Sciences, Birkendorfer Str. 65, D-88400 Biberach Riss, Germany
[5]Graduate School of Neuroscience (GSN), 82152 Munich, Germany
[6]Munich Cluster for Systems Neurology (SyNergy), 81377 Munich, Germany
[7]Center for Translational Cancer Research (TranslaTUM) of the TUM, 81675 Munich, Germany
[8]Image-Based Biomedical Modeling, Department of Informatics, Technical University of Munich, 85748 Garching, Germany
[9]Helmholtz AI, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[10]Biomedical Image Analysis Group, Department of Computing, Imperial College London, London SW7 2AZ, UK
[11]Department for Quantitative Biomedicine, University of Zurich, 8006 Zurich, Switzerland
[12]Department of Neuroradiology, Klinikum rechts der Isar, 81675 Munich, Germany
[13]Munich Medical Research School (MMRS), 80336 Munich, Germany
[14]Institute for Diabetes and Cancer, Helmholz Zentrum München, 85764 Neuherberg, Germany
[15]Institute for Diabetes and Obesity, Helmholz Zentrum München, 85764 Neuherberg, Germany
[16]German Center for Diabetes Research, Helmholz Zentrum München, 85764 Neuherberg, Germany
[17]German Center for Neurodegenerative Diseases (DZNE), 81377 Munich, Germany
[18]Neuroproteomics, School of Medicine, Klinikum Rechts der Isar, Technical University of Munich, 81675 Munich, Germany
[19]Institute of Anatomy, University of Leipzig, 04109 Leipzig, Germany
[20]Institute of Computational Biology, Helmholz Zentrum München, 85764 Neuherberg, Germany
[21]TUM School of Life Sciences Weihenstephan, Technical University of Munich, 85354 Freising, Germany
[22]Department of Mathematics, Technical University of Munich, 85748 Garching, Germany
[23]NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
[24]These authors contributed equally
[25]Lead Contact
*Correspondence: mmann@biochem.mpg.de (M.M.), erturk@helmholtz-muenchen.de (A.E.)
https://doi.org/10.1016/j.cell.2022.11.021

SUMMARY

Spatial molecular profiling of complex tissues is essential to investigate cellular function in physiological and pathological states. However, methods for molecular analysis of large biological specimens imaged in 3D are lacking. Here, we present DISCO-MS, a technology that combines whole-organ/whole-organism clearing and imaging, deep-learning-based image analysis, robotic tissue extraction, and ultra-high-sensitivity mass spectrometry. DISCO-MS yielded proteome data indistinguishable from uncleared samples in both rodent and human tissues. We used DISCO-MS to investigate microglia activation along axonal tracts after brain injury and characterized early- and late-stage individual amyloid-beta plaques in a mouse model of Alzheimer's disease. DISCO-bot robotic sample extraction enabled us to study the regional heterogeneity of immune cells in intact mouse bodies and aortic plaques in a complete human heart. DISCO-MS enables unbiased proteome analysis of preclinical and clinical tissues after unbiased imaging of entire specimens in 3D, identifying diagnostic and therapeutic opportunities for complex diseases.

CellPress

## INTRODUCTION

At their early stages, many diseases have modest pathological changes in mostly unknown tissue regions, making them hard to identify and characterize. For example, early changes in dementia may include the activation of a few local inflammatory cells, changes in the microvasculature, and the appearance of just a few initial amyloid-beta plaques in uncharacterized brain regions.[1] Such small regional changes are extremely hard to identify using standard histology, limiting our ability to investigate initial stages of diseases for early diagnosis and therapy. Recent advances in tissue clearing technologies allow fluorescence imaging of complete biological tissues including mouse organs and whole bodies as well as intact human organs.[2–4] After tissues are rendered transparent, end-to-end laser scanning microscopy reveals their cellular and subcellular details. Leveraging artificial intelligence (AI)-guided image analysis, even tiny changes in cellular structures, which otherwise would be missed, can be identified readily and quickly.[5,6] However, visually pinpointing these regions alone does not answer mechanistic questions at the molecular level.

In parallel, enormous progress has been made to increase the throughput of single-cell transcriptomic technologies,[7–9] even though some studies show that RNA expression may weakly correlate with the protein expression.[10–13] More recently, single-cell analysis of proteomes has also become possible with the advent of ultra-high-sensitivity mass spectrometry (MS)-based approaches.[10,14–17]

At the RNA level, it is possible to preserve the spatial context in two-dimensional (2D) samples as spatially resolved transcriptome analysis is quickly becoming a mainstay of the molecular biology toolkit.[18–25] Methods for spatial proteome analysis are actively developed, but most methods are either limited to analyzing fewer than 100 proteins (e.g., highly multiplexed or multi-round immunostaining) or limited to 2D samples (e.g., MS-based spatial proteomics) or both.[26–30] In addition, it has been unclear to what extent these methods can be made compatible with advanced whole-organ imaging of optically cleared tissue.

Here, we combine whole-organ and whole-mouse tissue clearing and imaging with ultra-high-sensitivity MS-based proteome analysis and can characterize the proteome of samples isolated from cleared tissue comprising fewer than 100 cells. We term this method three-dimensional (3D) imaging of solvent-cleared organs profiled by mass spectrometry, or DISCO-MS (Figure S1A). Using DISCO-MS, we successfully analyzed fluorescently labeled small target regions isolated from whole mouse or human organs to discover spatial-molecular profiles of initial pathological events in various disease models (Video S1). To facilitate sample extraction, we developed a robotic tissue extraction system (called DISCO-bot) for more challenging specimens including whole adult mouse bodies and whole human organs (Video S2). Combining DISCO-MS and DISCO-bot, we studied the spatial proteome heterogeneity of immune cell-enriched tissues from bone marrow niches of intact mice and aortic plaques from the human heart.

## RESULTS

### MS-based proteomics of solvent-cleared tissue

Tissue clearing is a chemical process that relies on tissue permeabilization and subsequent extraction of different biomolecules including water (organic-solvent-based methods) and lipids (in most clearing methods).[31] Whether the proteome of tissues after these diverse extraction steps remains intact had been unclear. To investigate this, we employed MS-based proteomics, which can provide unbiased in-depth insights into the composition, structure, and function of the entirety of expressed proteins.[32]

To this end, we subjected solvent-cleared samples to proteomics workflows with respect to protein recovery and qualitative and quantitative reproducibility. Using fresh-frozen tissues as controls, we started with 3DISCO and uDISCO clearing methods (Figures S1B–S1D), two commonly used methods that are quick and known to provide the highest tissue and organ transparency.[3] We tested several protein-solubilization approaches and figured that the combination of SDS-based protein solubilization and tissue pulverization, followed by SDC resolubilization and protein digestion, yielded qualitatively and quantitatively very similar proteomes between fresh and solvent-based clearing methods (Figures S1B and S1C). We identified up to 5,500 proteins across conditions with Pearson correlation coefficients between 0.89 and 0.99 (Figure S1D).

While 3DISCO and uDISCO work well for fluorescent dye imaging, the signal of endogenous fluorescent proteins such as EGFP is rather unstable and decays with time.[33] To avoid this, we developed vDISCO, which uses fluorescent-dye-conjugated nanobodies to stabilize and enhance fluorescence signals.[34] As vDISCO includes several additional steps that might change proteome constitution, we also tested our proteomics workflow on these tissues. Identification of more than 6,000 proteins across all clearing conditions including vDISCO and replicates with Pearson correlations ranging from 0.85 to 0.94 confirmed the suitability of our workflow (Figure 1A). Next, we asked whether archived human brain tissues can be analyzed and cleared human brain tissue stored in formalin for more than 5 years with the DISCO and SHANEL methods.[4] DISCO-MS identified more than 5,000 proteins in all clearing conditions very similar to paraformaldehyde (PFA)-fixed controls at high quantitative reproducibility (R = 0.91–0.96) (Figure 1B). We conclude that our sample preparation workflow allows the MS-based proteome analysis of cleared mouse and human specimens at high depth and quantitative accuracy comparable to fresh and PFA-fixed control samples.

### High proteome yield in vDISCO cleared tissues

Next, we examined the in-depth proteomes of vDISCO-cleared tissues to investigate potential protein depletions introduced by the clearing process and compared with fresh-frozen samples. We identified close to 8,000 proteins, a very substantial proportion of the total proteome across conditions and biological replicates with a high quantitative reproducibility (R = 0.94; Figure 1C and S2A). Coefficients of variation (CVs) within fresh (non-perfused) and cleared (perfused and fixed) conditions were below 0.2, demonstrating that vDISCO clearing yields proteomes that are qualitatively and quantitatively similar to fresh tissue and highly reproducible across biological replicates

CellPress

Cell
Resource

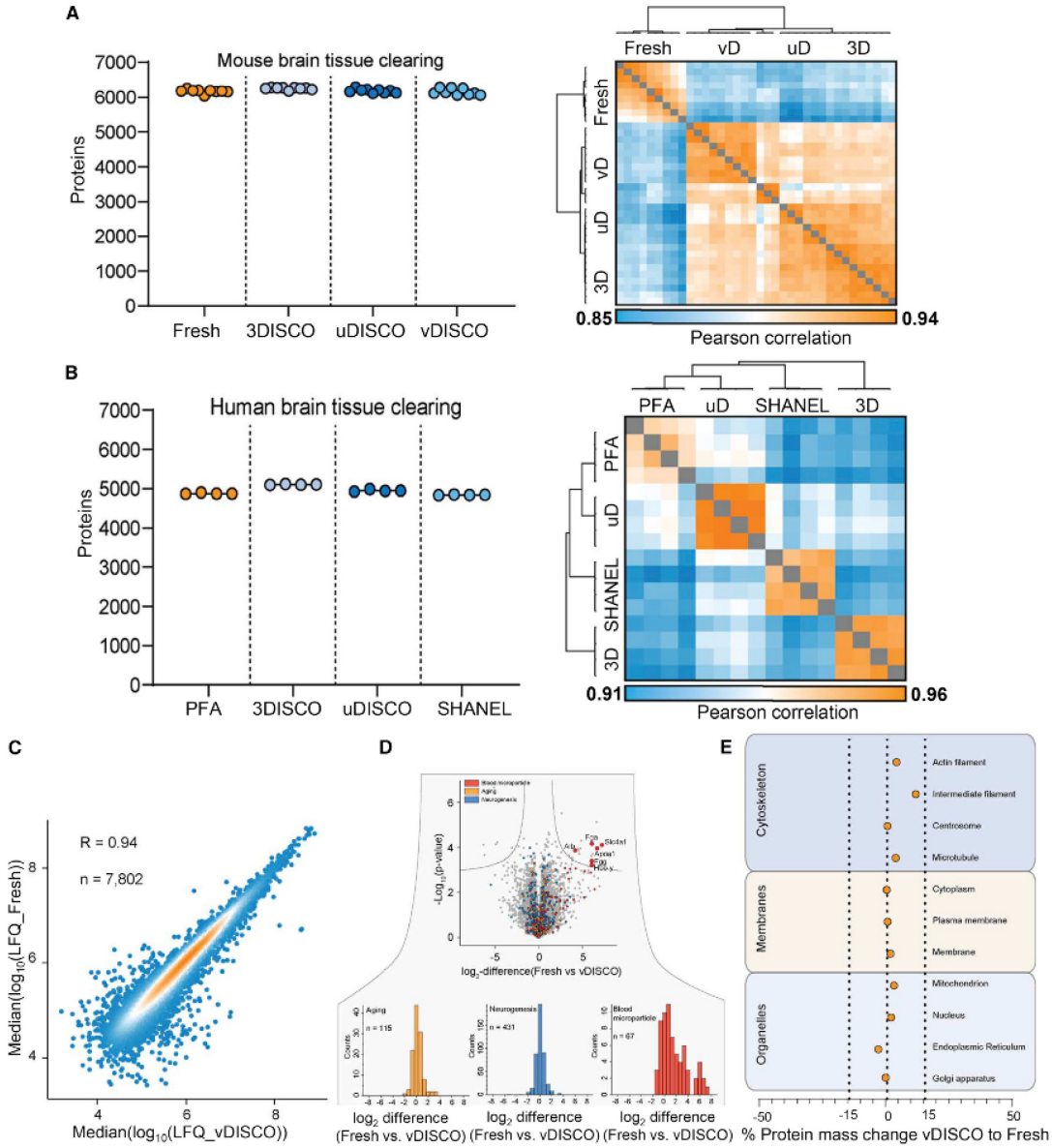



**Figure 1. Proteome of cleared rodent and human tissues**

(A) Proteome analysis from mouse brain tissues after different organic solvent-based tissue clearing methods vs. fresh controls. Protein identifications and proteome correlations across all clearing techniques and fresh tissue are shown. (N = 3 biological replicates, n = 9 total experimental replicates per condition).

(B) Archived human brain cortex blocks cleared and number of detected protein groups with proteome correlations across all clearing methods are compared with the numbers in PFA fixed blocks. n = 4 experimental replicates.

(C) Quantitative reproducibility of vDISCO-cleared vs. fresh.

(D) DE analysis of vDISCO-cleared vs. fresh sample proteomes highlighting the expected change in "blood microparticle" due to blood perfusion step for tissue clearing in contrast to fresh samples. Otherwise, proteins in other GO groups were unchanged.

*(legend continued on next page)*

(Figures S2B–S2E). The only altered gene ontology (GO) term was "blood microparticle" proteins, not surprising as fresh tissues were not perfused. Other GO keywords are quantitatively and qualitatively preserved (Figure 1D and S2F). Next, we checked the protein mass distributions between fresh and vDISCO mouse brains for membrane, organelle, and cytoskeleton terms. We found that all percentage protein mass differences were well below 15% across sub-terms and that the protein mass change associated with all membrane-related terms was below 3% (Figure 1E). Moreover, we also looked at the specific membrane-bound protein class, which are plasma membrane proteins. The percentage total protein mass change was below 0.5%. As an example, we highlighted epidermal growth factor receptor (EGFR), a transmembrane receptor, between fresh and vDISCO-cleared brain samples (Figure S2G). In summary, even the strongest organic solvent-based tissue clearing approach, vDISCO, yields qualitatively and quantitatively very similar proteomes compared to the fresh tissue.

### Proteomes of micro-dissected tissues imaged in 3D

After establishing a high-quality and reproducible MS-compatible sample preparation workflow for solvent-cleared tissues, we turned to the unbiased proteome analysis of smaller target tissue regions (~0.0005 $mm^3$) previously imaged and located in 3D. We successfully met three major challenges: (1) reliable dissection of small tissue regions identified by 3D imaging of cleared tissues, (2) analysis of deep proteomes from only a few nanograms of dissected and rigid solvent-cleared tissue, and (3) registration of small tissue regions back onto a complete 3D light-sheet image stack. To solve the first challenge, we developed a series of steps to render cleared rigid tissue soft for precise cryo-sectioning and laser capture microdissection (LCM) without deformation. In short, we reversed the clearing protocol, rehydrated the cleared tissue stepwise, and cryo-preserved it. This workflow avoided rupturing of the tissue during cryo-sectioning and allowed us to laser micro-dissect tissue regions as small as 0.0005 $mm^3$ corresponding by volume to approximately 60 cells. Next, we miniaturized our sample preparation workflow and then performed MS-based proteome analysis on a modified trapped ion mobility MS platform developed for highest sensitivity down to the level of single cells.[10] Lastly, we automated registration of the 2D LCM images back to the 3D light-sheet imaging stack (Figure 2A). Briefly, following light-sheet microscopy imaging, the cleared mouse brains were rehydrated, cryoprotected, sectioned, and imaged prior to cutting by LCM. These images were aligned (rigid registration) to produce a reconstructed 3D stack of LCM sections. These were then coordinated slice by slice to a subset of the light-sheet imaging stack (affine registration), and the highest-scoring section was subsequently used to project the regions of interest (ROIs) onto the 3D light-sheet imaging data (Figure 2B and S3Q; see Methods S1).

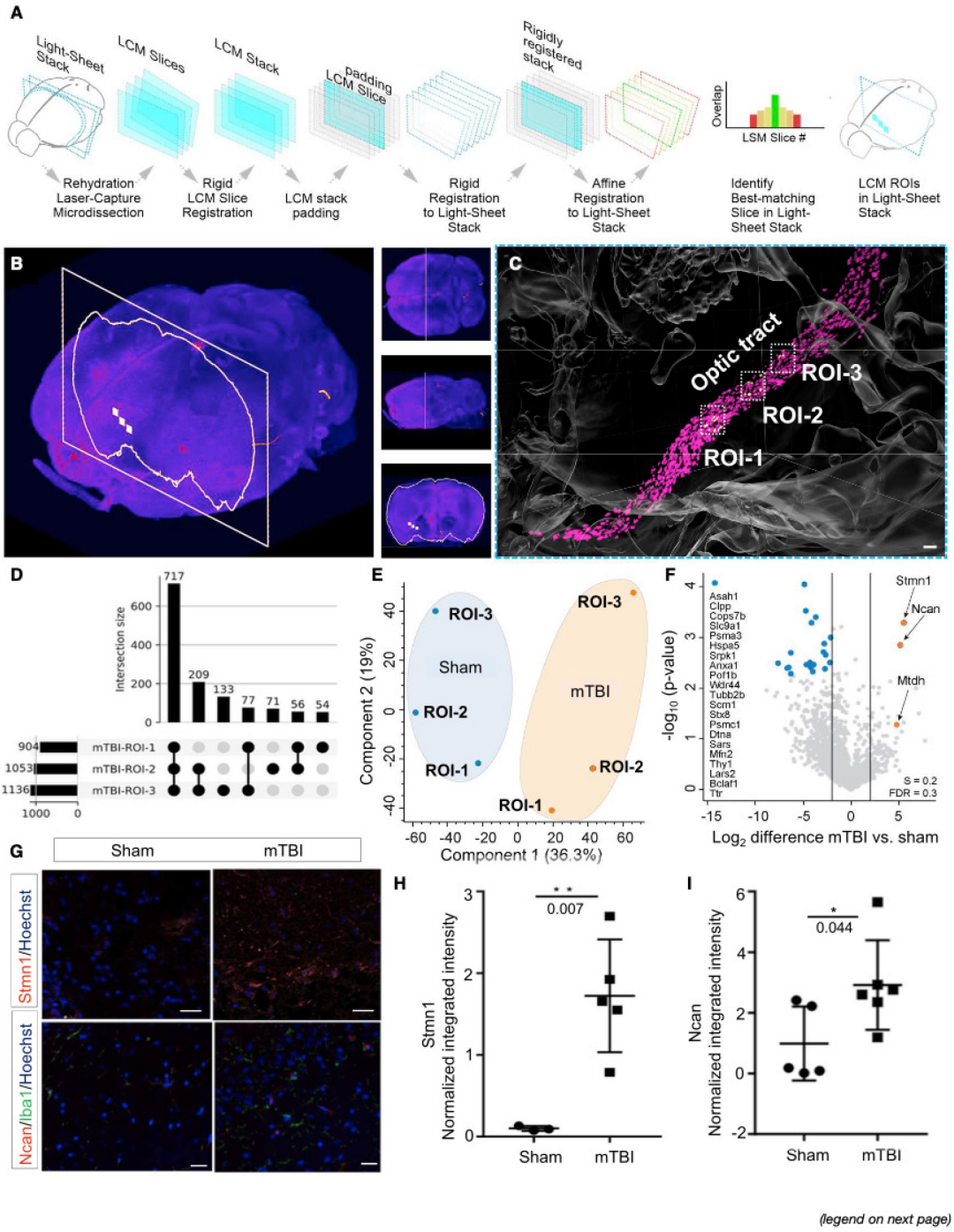
To explore the potential of our technology in biological applications, we first used a mild traumatic brain injury (mTBI) mouse model to identify and analyze proteomes of brain regions containing discrete local inflammation. mTBI and concussions are common injuries that can lead to long-term morbidities such as sleep disorders, neuropsychiatric disorders, and even early onset of dementia.[35] They are characterized by chronic inflammation, which can induce neurodegeneration in selected brain regions, particularly along the stretched axonal tract.[36] We used a repetitive mTBI injury model on CX3CR1-EGFP mice (Figures S3A and S3B), in which all microglia are labeled with an EGFP-fusion construct. The mild nature of this injury was confirmed by behavioral test that showed no significant alterations after 8 weeks post injury in Barnes maze and beam-walk tests (Figures S3C and S3D). Brains of CX3CR1-EGFP mice were processed for vDISCO labeling and clearing 8 weeks post injury, showing microglia activation in discrete regions of whole brain. ClearMap quantification[37] identified activated microglia with enlarged morphology in diverse brain regions, especially along the axonal tracts including the optic tract and the corpus callosum when compared with sham-operated animals (Figures S3E–S3L). The same mTBI injury model on Thy1-GFP-M reporter mice (expressing GFP only in neurons) confirmed the axonal abnormalities in the same brain regions (Figures S3M–S3P).

We then used DISCO-MS on isolated ROIs including locally activated microglia with known spatial information (Figures 2B and 2C; Video S3). Analyzing three ROIs from the optic tract as small as 0.0005 $mm^3$ compared to corresponding regions in sham control animals, we quantified up to 1,400 proteins per ROI. Overall, we found 602 common proteins in all ROIs of mTBI and sham. Comparing ROIs from mTBI among themselves, we found a shared proteome signature comprising 717 proteins, with each ROI having unique sets of proteins (Figure 2D). Principal-component analysis (PCA) separated the proteomes of ROIs between mTBI and controls (Figure 2E). Interestingly, several proteins related to axonal damage and repair were strongly differentially regulated between conditions. For example, stathmin1 (Stmn1; 32-fold increase) is a protein involved in the regulation of microtubule filament system. Its overexpression has been shown to promote disassembly of microtubules in blast-induced mTBI.[38] Neurocan (Ncan; 30-fold increase) is a chondroitin sulfate proteoglycan involved in the modulation of cell adhesion and migration, which is upregulated in injured brains.[39] Eight proteins were uniquely detected in mTBI, including metadherin (Mtdh) present in glial cells but not yet been described in mTBI. Furthermore, we found many previously described proteins in the context of brain injury, providing positive controls, and many yet unknown proteins to be downregulated in the mTBI model. For example, acid ceramidase (aCDase, Asah1), an enzyme implicated in sphingolipid metabolism and in multiple diseases, was downregulated[40–42] and has previously been associated with morphological defects in cultured neurons.[43] Our data now suggest its potential involvement in axonal deformity in mTBI. We also observed alterations in mitochondria proteins along the optical tract such as caseinolytic protease proteolytic subunit (Clpp), a mitochondrial matrix

(E) Percentage change of protein mass distribution between vDISCO-cleared vs. fresh samples. Percentage changes are shown as a median change within one group for organelles, membranes, and cytoskeleton GO terms (N = 3 biological replicates, n = 9 total experimental replicates per condition).
See also Figures S1 and S2.

CellPress

Cell
Resource




(legend on next page)

protease,[44] and mitofusin 2 (Mfn2), a mitochondrial membrane protein that participates in mitochondrial fusion and contributes to the maintenance and operation of mitochondrial network[45] (Figure 2F). We further validated the enrichment of Stmn1 and Ncan in mTBI brain tissues by immunofluorescence (Figures 2G–2I). Our data demonstrate that DISCO-MS is a powerful approach for obtaining unbiased proteomic information on heterogeneous tissue regions with known spatial locations.

## Scalable and robust pathology identification using deep learning

One of the early hallmarks of Alzheimer's disease (AD) pathology is the accumulation of amyloid-beta (Aβ) plaques in the brain parenchyma.[46] We anticipated that the unbiased detection of Aβ plaques, followed by their equally unbiased proteome analysis using DISCO-MS, would provide valuable insights into the initial stages of AD. To this end, we used the 5xFAD mouse model of AD to identify Congo red labeled Aβ plaques in young mouse brains.

As the locations of these initial plaques are unknown, we developed a deep learning (DL) approach to identify all Aβ plaques rapidly and reliably in whole mouse brain scans. In short, our network architecture is based on U-Net, a well-established approach for biomedical image analysis[47] (Figure 3A). To assess our segmentation quality, we calculated a wide range of voxel-wise and Aβ plaque segmentation metrics. Our DL architecture for automated Aβ plaque detection showed high performance in volumetric accuracy (0.99 ± 0.00), volumetric (0.71 ± 0.06), and surface (0.94 ± 0.03) as well as overall Dice scores (0.89 ± 0.09) per Aβ plaque. After segmenting all plaques in the entire brain using DL, we registered our data to the Allen brain atlas to obtain region-wise quantifications for over a thousand brain subregions (Figure 3B).[48,49] We then grouped them into the major brain regions as defined by the Allen mouse brain ontology for simplicity and visualized plaque volume per region in early and later stages of disease (Figures 3C and 3D) using brainrender-an open source Python package for interactive visualization.[50]

Our DL model identified few Aβ plaques in 6-week-old mouse brains and identified several thousand at a later stage of the disease (6-month-old 5xFAD), (Video S4) whereas plaques were absent at 5 weeks of age. We also confirmed our finding of Aβ plaques in 6-week-old mice in the same brain regions by immunohistochemistry with anti-Aβ monoclonal antibodies (Figure S4A). To compare the plaques between 6-week- and 6-month-old 5xFAD mice in 1,238 different brain regions, we grouped 72 major brain regions according to Allen brain atlas and plotted 27 disease-relevant regions. Some of the main brain regions with initial plaques were retrohippocampal region, medulla, molecular layer of cerebellar cortex, fiber tracts, subiculum areas, visual area, and hippocampal formation. The subiculum area and retrohippocampal region showed much larger number of plaques at later stage (6 months) of disease: 1,689 and 10,630, respectively (Figures 3E and 3F). We observed the largest plaques in the midbrain (motor related, 3,438 $\mu m^3$) followed by temporal association areas (2,880 $\mu m^3$), posterior amygdalar nucleus (2,450 $\mu m^3$), and auditory areas (2,366 $\mu m^3$), with an average volume ranging between 2,000 $\mu m^3$ and 3,500 $\mu m^3$ in early stage (Figure 3G). Interestingly, the plaque volume of these regions were significantly reduced at 6 months, indicating a temporal change in plaque morphology, either caused by adverse biological effects by their surrounding microenvironments or potentially causing changes to it during disease progression.

Following the DL-based identification of early Aβ plaques in the 5xFAD mouse model, we isolated four ROIs (volume: ~0.0005 $mm^3$) from the hippocampal region vs. corresponding brain regions from the control mice and subjected them to MS-based proteomics (Figures 4A–4D). We compared ~2,000 proteins across replicates and PCA plot separated the ROIs with Aβ plaques from the control brain regions (Figure 4E). Differential expression (DE) analysis revealed that many well-characterized AD-associated proteins were enriched in 5xFAD ROIs including the Aβ precursor protein,[46,51] (32-fold increase) and the thimet oligopeptidase 1 (8-fold increase) (Figure 4F). Apart from known and well-established AD-related proteins,[52,53] we also detected less-characterized proteins in early-stage Aβ plaques such as a member of the calcium-binding protein family S100a11.

Moreover, we asked how similar the proteomes of our ROIs with early Aβ plaques were compared to each other. Plaques with more than 1,200 protein identifications each shared 768 proteins, defining a core proteome of early-stage Aβ plaque formation (Figure 4G). An abundance rank plot of the shared early-stage Aβ plaques' core proteome revealed several members of the Ywhaz (14-3-3) and the S100a protein family. In the early-stage Aβ plaque ROIs, we also found many other proteins involved in AD such as two isoforms of Mapt, namely Mapt-4 and Mapt-5, illustrating the specificity of MS-based proteomics (Figure S5A). Our proteomics data also highlight early-stage Aβ plaque variability (Figure S5B) with respect to well-characterized AD proteins (Mapt, Tmed10, App), proteins of the S100a family,

**Figure 2. DISCO-MS reveals effects of mTBI in discrete regions of whole brain**

(A) Schematic representation of the alignment procedure.

(B) Location of the LCM slice (outline) inside the light-sheet stack, with the dissected 200 × 200 μm areas (white squares) inside the optic tract.

(C) 3D reconstruction of stiched images of an exemplary CX3CR1-GFP/+ mouse brain after mTBI, highlighting substantial increase in activated microglia (in magenta) along optic tract region. 3 neighboring ROIs along optic tract were identified, laser captured, and subjected to proteomic analyses. Scale bar, 100 μm.

(D) The number of shared and unique set of proteins in mTBI.

(E) PCA plot showing the distribution of individual ROIs from mTBI vs. ROIs from control (sham) with the same spatial location.

(F) Volcano plot showing the significant enrichment of proteins.

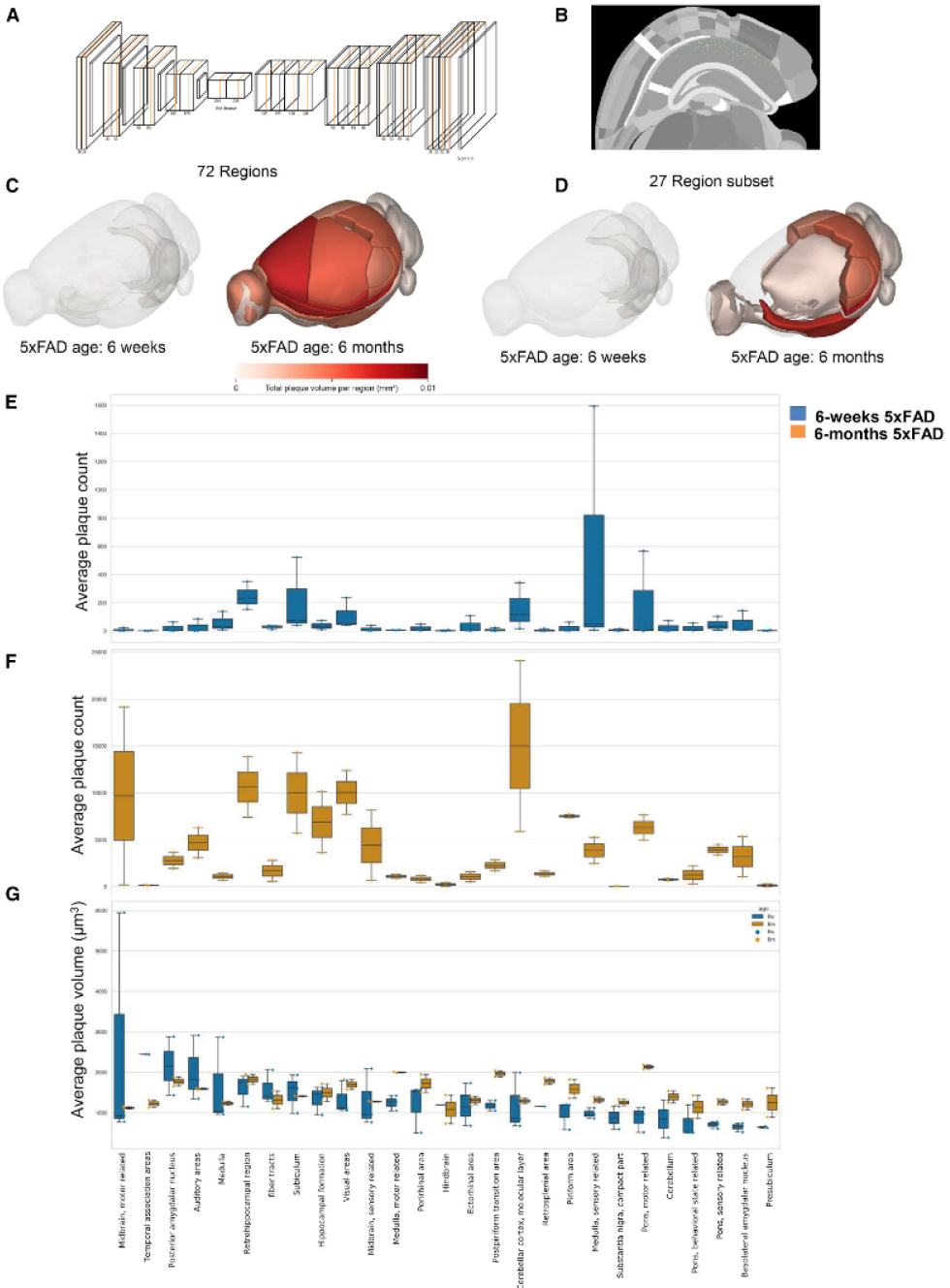
(G) Histological validation of the top 2 proteins in the optic tract: (1) stathmin (Stmn1) in red and nuclear marker Hoechst dye in blue, (2) neurocan (Ncan) shown in magenta along with microglia marker (Iba1) in green and Hoechst dye in blue. Scale bars, 20 μm.

(H and I) Intensity quantification of Stmn1 immunostaining signal ($p = 0.007$, $n = 3$ from total 8 sections) and Ncan immunostaining signal ($p = 0.044$, $n = 3$, animals from total 11 sections) in mTBI vs. sham, respectively (unpaired two-sided Student's *t* test, data presented as ± SD).

See also Figure S3, Methods S1, and Video S3

CellPress

Cell
Resource




(legend on next page)

peptidases (Thop1, Ppia), proteins of the Ywha family, and other structure-determining proteins including Nefm and Map2. Finally, we confirmed the presence of S100a11 and Thop1 in early-stage plaques of 5xFAD brain slices by immunofluorescence, whereas these proteins are absent from the respective regions in wild-type (WT) mice (Figures S4B–S4D).

We next compared plaque regions to neighboring non-plaque regions in the dorsal and ventral subiculum, where we observed initial plaque formations (Figures 4H and 4I). We found that 29 proteins were upregulated whereas 14 proteins were downregulated in these early plaque isolates. Among others, we found significant quantitative change of proteins related to vesicle fusion, vesicle-mediated transport, and secretory pathways (Bnip1, Cpd; >10-fold increase in both; $p < 0.0001$ for Bnip; $p < 0.05$ for Cpd), which are potential markers for an early AD onset (Figure 4K). Known plaque-associated proteins were regulated in a region-specific manner, which were previously identified only at the later stages of AD without precise spatial location (Figure 4L). For example, an increased expression of Manf had been reported in APP/PS1 transgenic mice only at later stages of the disease.[54] Actl6b is involved in transcriptional activation and repression of select genes by chromatin remodeling and has been investigated in in the context of later-stage AD (6–15 months).[55] Mutations in human Actl6b have been associated with intellectual disability,[56] suggesting a role of this protein in the early stage of neurological conditions, including AD. Mbp, Cnp, and Plp1 have been associated with the structural integrity of the myelin sheath and its function in age-related AD.[57] In addition to many significantly enriched proteins in early plaques vs. non-plaque areas of the same region, we also found substantial subregional heterogeneity (Figures S5C–S5E). Furthermore, many of these proteins were uniquely distributed across these two subregions. These could potentially drive disease progression in a region-specific manner, a concept that has been little explored thus far due to technical limitations.

Next, we profiled the later-stage plaque (6 months) microenvironment in a region-specific manner and compared this with early plaque protein enrichments (Figures S5F and S5G). Dorsal subiculum data showed a significant upregulation of 49 proteins, whereas 21 proteins were downregulated. Among the upregulated proteins, we found proteins related to S100a family (S100a13), complement activation and microglia phagocytosis pathway (C1qa, C1qb, Itgb), and tau protein binding (Apoe, S100b, Clu). We also observed a distinct proteomic signature in dorsal vs. ventral subiculum at later stages, indicating region- and time-specific changes in the proteomic landscape around these plaques (Figures S5H and S5I).

Taken together, DISCO-MS allowed us to pinpoint early- and late-stage Aβ plaques from the whole brains in 3D and analyze their spatial proteomic makeup. We recovered many known markers of Aβ plaques in AD as well as less-characterized proteins in AD pathogenesis. Our data also suggest the significant involvement of S100a, Ywhaz (14-3-3), vesicle fusion and transport, myelin sheath function, and the complement system related family members in early-stage Aβ plaque development.

### DISCO-MS from large-volume samples by DISCO-bot tissue extraction

Next, we aimed to develop DISCO-MS for even larger samples including whole mouse bodies and whole human organs. As sectioning/imaging for LCM is impractical for large-volume samples at scale, we developed a robotic extraction system (named DISCO-bot), using biopsy needles to isolate ROIs for subsequent proteomics analysis (Figure 5A). This required (1) stabilization of whole mouse body for robotic extraction while imaging, (2) minimizing biopsy needle deflections during extractions, and (3) biopsy needles that can penetrate into hard cleared tissues.

Firstly, to stabilize the sample, we investigated different resins and agarose concentrations to modulate bed stiffness, compatibility with the clearing solutions, and imaging and figured that 2% agarose embedding was well suited for our purpose. Next, we customized different 3D-printed mouse holder adaptors and needle holders, according to their strength and force-deflection criteria and chose the one with the least deflection (Figure 5B–5E and S6). Last, we tested various needle sizes and shapes with stylet inside and found that needles size 18 gauge (G) and/or 22 G provided a good penetration and sample extraction precision without contamination of undesired tissue (Figure 5F; see details in STAR Methods). We further optimized the DISCO-bot to work concomitantly with the light-sheet microscope (Figures 5G–5I). The resulting DISCO-bot system allowed the extraction of small tissue regions from cleared samples while imaging, thereby enabling non-destructive and repetitive tissue isolation at scale for DISCO-MS (Figures 5J–5M; Video S5; see DISCO-MS Handbook as Methods S1).

To demonstrate the utility of the DISCO-bot in our DISCO-MS pipeline, we cleared LysM-eGFP mice using vDISCO. We co-labeled nuclei with propidium iodide (PI) to identify the ROIs for DISCO-bot extraction. Imaging of the whole body allowed us to spatially identify locations of all LysM-eGFP$^+$ cells, which were mostly found to be in bone marrow niches (Figure 6A; Video S6). We then focused on the cranium and scapula, two bones with irregular 3D structures and thus hard to study with standard 2D sections. We chose three ROIs from the parietal cranium region and six ROIs from scapula (three from the lateral border (LysM-eGFP) and three ROIs from the medial border (LysM-eGFP$^+$) [Figures 6B–6E]). DISCO-bot-extracted ROIs were then

**Figure 3. Deep learning analysis of plaques in whole 5xFAD mouse brains**

(A) 3D U-Net architecture including layer information and feature sizes.
(B) Segmented plaques overlaid on the hierarchically and randomly color-coded atlas to reveal annotated regions available.
(C) Visualization of plaque volume per region in 72 regions.
(D) Visualization of plaque volume per region in a subset of 27 major regions.
(E) Quantification of the number of plaques in the major brain areas of 6-week-old mice (n = 3, ±SD).
(F) The number of plaques in the major brain areas of 6-month-old mice (n = 2, ±SD).
(G) Quantification of the plaque volume in 6-week-old mice (in blue, n = 3, ±SD) and 6-month-old mice (in orange, n = 2, ±SD).
See also Video S4

CellPress

Cell
Resource

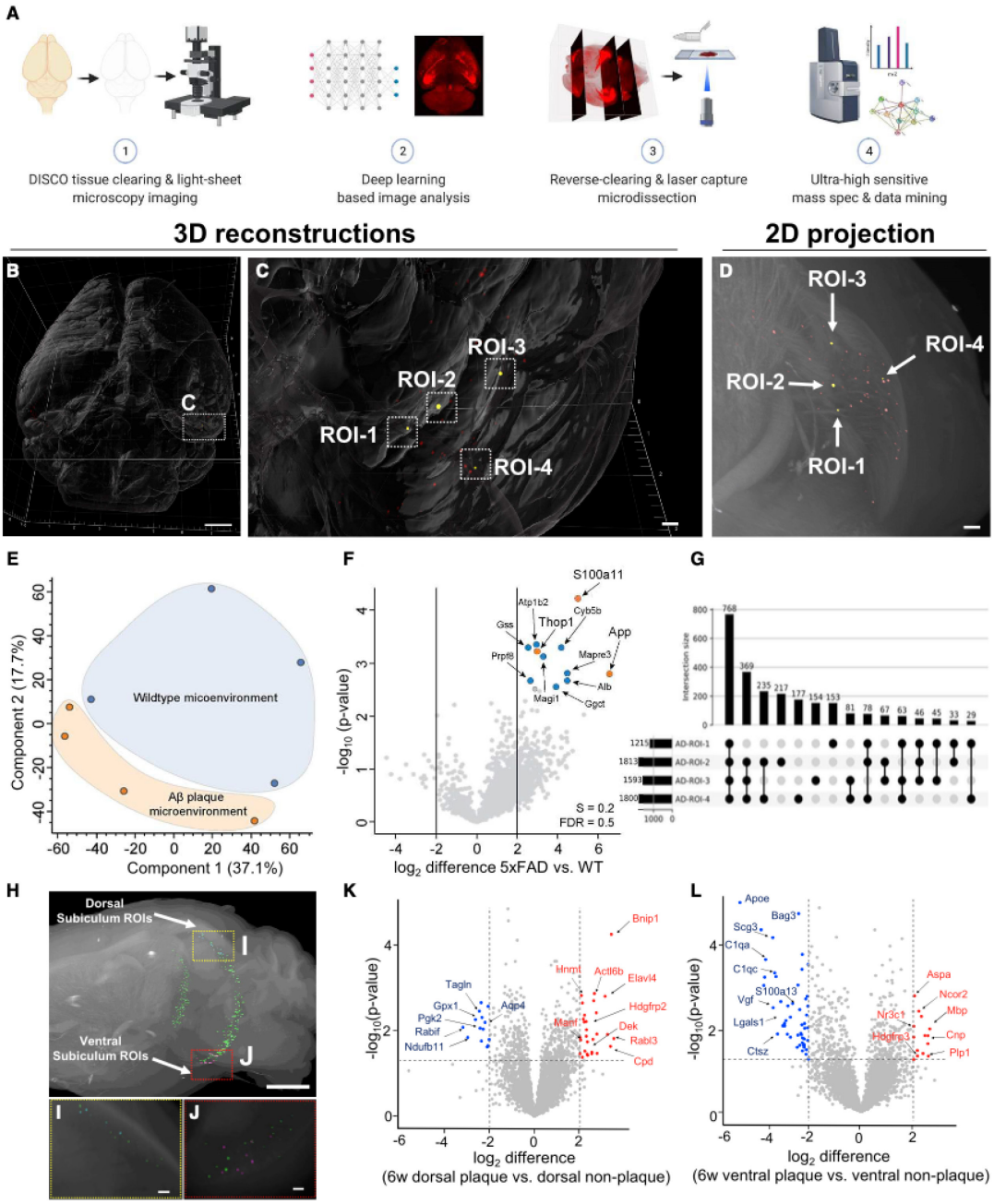



**Figure 4. DISCO-MS unravels the single-plaque proteome in AD mouse model**

(A) Major steps of DISCO-MS.

(B) 3D visualization of Aβ plaques (in red) in stiched images of 5xFAD mouse brains (n = 4 experimental replicates). Scale bar, 500 μm.

*(legend continued on next page)*

analyzed using the DISCO-MS pipeline as described above. In the cranium, we identified the shared signature of 1,984 proteins in all three ROIs with Pearson correlation between 0.72 and 0.76 (Figures 6F and 6G). To verify the precision of DISCO-bot extraction, we compared our results with proteomics of freshly isolated skull marrow cells.[58] We found ~2,200 shared proteins out of total 2,550 identified protein groups when compared with isolated skull-proteome, confirming the high precision of DISCO-bot extraction. Among others, seven protein groups identified here were earlier shown to be expressed both at transcript and proteome level in freshly isolated skull marrow, which further validates the extraction precision using robotic arm (Figure 6H). In the scapula, we observed distinct signal of LysM-eGFP in medial vs. lateral border bone. PCA clearly separated the two groups as well as the ROIs itself, particularly from medial border, indicating inter-regional and intra-regional heterogeneity among these extracts. We identified 1,250 proteins across conditions with high Pearson correlation coefficients (0.88–0.98) and found a common signature of 764 proteins between lateral and medial border bone, whereas 336 and 22 proteins were unique to the respective regions (Figures S7A–S7C). DE analyses showed upregulation of ten proteins, including those related to the innate and adaptive immune system such as antigen-presenting molecule H2-L (MHC class1b), B cell/T cell receptor pathway-related proteins, and cytokine signaling proteins such as signal transducer and activator of transcription 3 (Stat3), which are involved in biological processes of inflammatory response regulation to antigenic stimuli. Among 33 downregulated proteins were signalosome-related proteins (Cops7a) and proteins related to actin filament network formation (Fhod1) (Figure 6I–6L and S7D). These results demonstrate that DISCO-MS can be applied to whole adult mouse bodies after end-to-end imaging to investigate spatial-molecular heterogeneity and diverse biology.

### Spatial proteomics of coronary arteries in SHANEL-cleared human heart

Next, we tested our DISCO-bot-aided DISCO-MS approach in coronary artery disease (CAD). Acute myocardial infarction (MI) is the major contributor to cardiovascular mortality, the leading cause of deaths worldwide.[59,60] It occurs when atherosclerotic plaques (largely made up of lipids and calcified tissue) slowly build up in the inner lining of a coronary artery and then suddenly rupture, causing thrombus formation, occluding the artery and ultimately prevent blood flow to other parts of body. Recent advances in single-cell RNA sequencing (scRNA-seq) technologies have provided a deeper understanding of cardiac cells in both health and disease[61,62] but lack precise spatial information. Moreover, transcript levels may weakly correlate with proteome data, even at the single cells.[10]

We obtained PFA-fixed human heart, labeled its vessels with dextran, and cleared the whole heart using SHANEL human organ clearing protocol (Figures 7A and 7B). This readily located the calcified atherosclerotic plaques of any size along regions without plaques in the same coronary artery. We extracted six ROIs around the large plaque-associated regions and six ROIs from the little-to-no plaque-associated regions, which could represent early plaque formation along the coronary artery of the same heart for DISCO-MS analysis (Figures 7C–7F; Video S7). We quantified 1,300 proteins in each ROI and found 53 downregulated and 6 upregulated proteins (Figure 7G). Systematic GO term and pathway analysis of regulated proteins suggests the regulation of hypertrophic cardiomyopathy, cardiac muscle contraction (TPM1, TPM2, MYL2, MYL3, MYH6, MYBPC3, DES), focal adhesion, blood coagulation, plasminogen activation, platelet aggregation, fibrinogen complex, blood-clotting cascade (SERPING1, FGA, FGG, FGB, FN1) (Figure 7H, S7E, and S7F). Some of these proteins were regulated in a region-specific manner including MYH10, MYH11, FGA, FGB, and FGG, which have already been reported to be associated with plaques (Figure 7I). Furthermore, we found upregulation (2.2-fold, $p < 0.05$) of myosin heavy chain (MYH8), a less-characterized protein in context of atherosclerotic plaque. As its related forms MYH10 and MYH11 were recently discovered as biomarker for atherosclerotic plaque formation,[63,64] our findings of MYH8 in the context of CAD is encouraging and invites further investigation. We also observed a 2.5-fold ($p < 0.01$) increase in glyoxalase 1 (GLO1), a ubiquitous cellular enzyme that participates in the detoxification of methylglyoxal, a cytotoxic byproduct of glycolysis that induces protein modification (advanced glycation end products, AGEs), oxidative stress, and apoptosis. GLO1 and AGEs are implicated in the pathogenesis of aging and diabetes. A large-scale meta-analyses showed a gene network involved in antigen processing to be strongly associated with CAD.[65] Key drivers of this network included GLO1, which strengthens the potential role of this protein in reprogramming atherosclerotic plaque microenvironment. We also observed a significant enrichment (2.3-fold, $p < 0.01$) of carbonic anhydrase 6 (CA6), which might play a role in vascular calcification, hence plaque progression, and has not been associated with CAD before. CA are a group of isoenzymes that catalyze the reversible conversion of carbon dioxide into bicarbonate. CA isoenzymes were previously shown to be involved in the vascular calcification in humans.[66,67] We further found perturbations in members of recently discovered myosin heavy-chain family

(C) Enlarged view of region marked in (B) 4 different ROIs (in yellow) from hippocampus, each containing single-plaque, selected and isolated for mass spectrometric measurements. Scale bar, 100 μm.
(D) 2D projection of selected ROIs. Scale bar, 100 μm.
(E) PCA plot of ROIs' proteome from 5xFAD vs. the same regions from control.
(F) Volcano plot showing the significantly enriched proteins.
(G) The number of shared and unique sets of proteins in 5xFAD.
(H) Selection of early plaque and non-plaque ROIs from 2 subregions of subiculum. Scale bar, 1,000 μm.
(I and J) Enlarged view of regions marked in (H). Scale bar, 100 μm.
(K) Volcano plot showing significantly enriched proteins in plaque vs non-plaque areas in dorsal region.
(L) In ventral region (n = 14–20 ROIs per region).
See also Figures S4 and S5 and Video S1.

CellPress

Cell
Resource

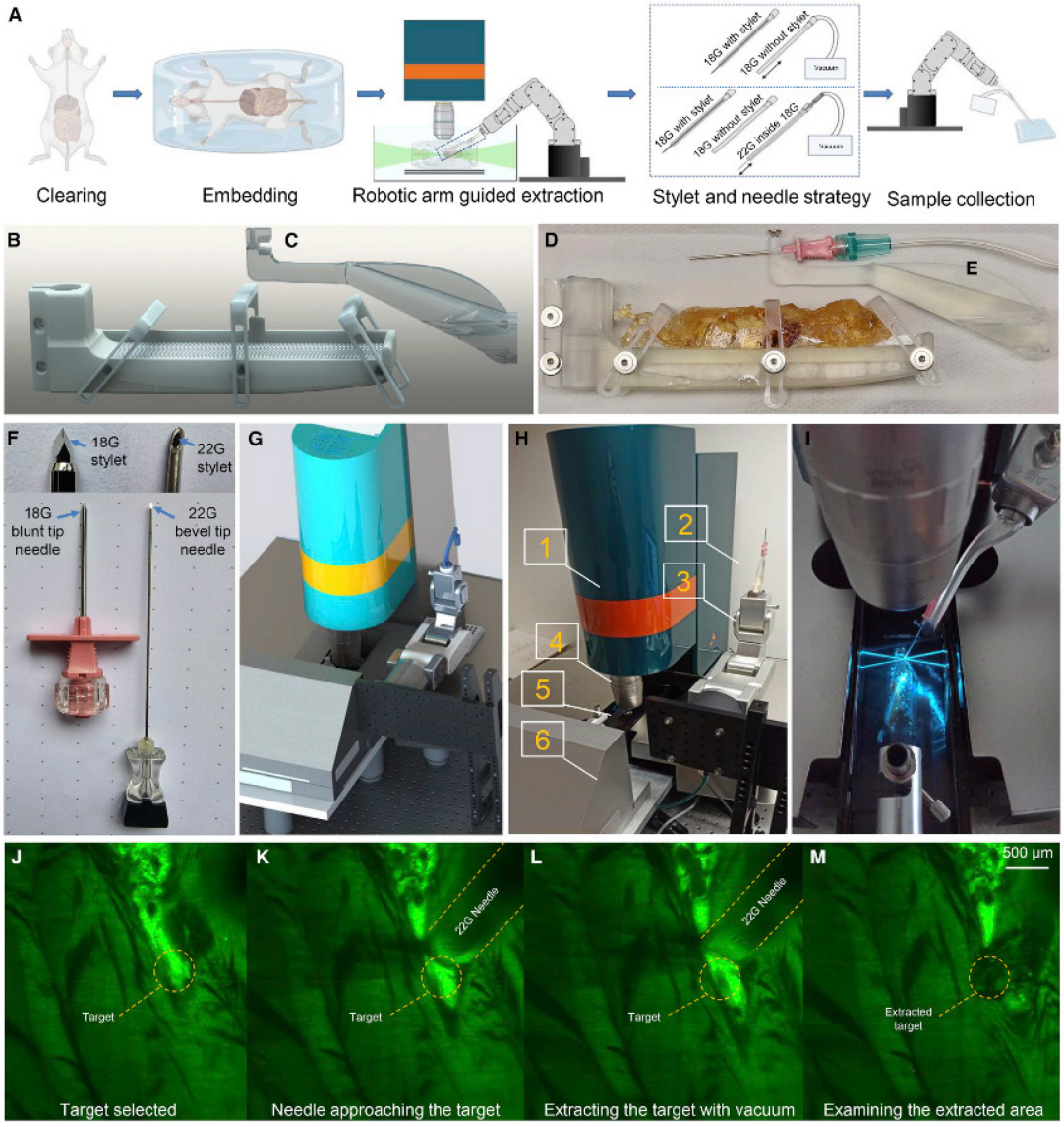



**Figure 5. DISCO-bot enables non-destructive tissue extraction from large, cleared samples while imaging**

(A) Schematic of robotic extraction.
(B) 3D model of the mouse holder.
(C) 3D model of needle holder.
(D) Cleared mouse on optimized mouse holder.
(E) Needle attached to the needle holder.
(F) The biopsy needles with sealing stylets.
(G) CAD model of microscope with robotic arm.
(H) Components of microscope and robotic setup annotated as (1) light-sheet microscope, (2) needle attached on the needle holder, (3) Meca500 6 DoF robotic arm with vacuum tube, (4) objective, (5) sample mounted on the mouse holder, (6) mouse holder stage.
(I) Meca500 robotic arm.

*(legend continued on next page)*

(MYH10, MYH11) biomarker and MYH6, a less-characterized form in the context of late-stage atherosclerotic plaque microenvironment. Moreover, we observed a significant reduction in the expression of MACROH2A.1 histone, which has been little studied in context of regulating autophagy and cholesterol efflux.[68]

## DISCUSSION

Deciphering tissue heterogeneity is essential to understand normal physiology and pathological processes. However, despite the enormous progress made in single-cell and spatially resolved -omics technologies,[7–10,14–17] spatial-omics analysis of tissue regions imaged in whole organs and organisms remained challenging.

We addressed this challenge by developing and applying DISCO-MS for proteome analysis of small tissue regions identified by panoptic imaging of large samples, *in silico* 3D reconstructions using AI, and automated minimally invasive tissue isolation for large-volume tissues using robotics. Employing DISCO-MS, we identified brain regions affected by the early stages of Aβ plaque genesis in the brains of a young AD mouse model. Most of the initial plaques appeared in the retrohippocampal regions, including the entorhinal area, and some in hindbrain regions, including pons and medulla.[69] The proteome data provided by our work can facilitate the discovery of diagnostic and therapeutic approaches tailored for an early-stage intervention of AD. Future studies are warranted as, e.g., time course and molecular details might differ between the 5xFAD mouse model we used and other widely employed models such as the 3xTg-AD mice.[70] Our method also allows the analysis of human post-mortem samples to explore the clinical relevance of our findings.

Earlier studies obtained the proteomic profiles either from the lysed whole brains of AD mice[71] or provided only average information on the plaque pathology from pre-selected brain regions without spatial details on the whereabouts of the plaques.[72,73] Moreover, some of these studies were performed only on certain isolated cell types or brain sections,[74,75] again missing the spatial context at the whole-organ level. More recently, attempts were made to characterize the plaques in whole brains of 5xFAD mice; however, these studies lack spatial proteome information.[76,77] In contrast, we obtain the proteomic data on isolated single-plaque microenvironments with the known spatial locations in whole brains. Thus, we provide both molecular information related to single plaques, and their spatial location, which would be critical for localized targeted treatments.

Our technology performs equally well on mouse and human tissues and yields qualitative and quantitative proteomics data nearly indistinguishable from uncleared samples, even for the harshest organic solvent-based tissue clearing approaches. The broad compatibility with different tissues and organisms and the relative ease of the DISCO-MS procedure will allow benchmarking different animal models against human samples on the molecular and anatomical levels.

Although tissue clearing presumably removes the lipidic cast of membrane proteins, we observed that the plasma membrane protein GO class was hardly affected, suggesting that DISCO-MS can be used to identify surface markers for drug targeting. Beyond neuroscience, this technology may transform spatial molecular investigations in many other biomedical research areas including clinical samples. We already applied DISCO-MS to human heart samples with atherosclerotic plaques and identified molecules involved in CAD such as GLO1, CA6, MYH6, and MACROH2A.1. After SHANEL clearing, we quickly navigated through the whole coronary artery in 3D and pinpointed the spatial distribution of plaque-associated regions. Given the precision and sensitivity of our method, we were able to isolate and analyze neighboring tissue regions with and without pathology for comparison.

A key advantage of our approach is that the robot-guided extraction of tissue isolates directly from the cleared tissue makes it scalable, as no prior sectioning and registration is needed. This allows the quick streamlined isolation of many samples from identical specimens while retaining complete spatial information and allowing for the resampling of other parts of the same specimen in future analyses.

In conclusion, we present a spatial unbiased proteome profiling technology comprising complete 3D-imaging data of whole organs and organisms, enabling unbiased identification and automated extraction of interesting tissue regions and including subsequent molecular characterization. Notably, the DISCO-MS technology presented here is versatile across labeling and solvent-cleared methods for whole organs. It is not only applicable to reporter mouse lines but can also be utilized where reporter lines are unavailable. In those cases, deep tissue labeling with dyes and potentially antibodies can be performed against the antigen of interest, imaged as a whole organ, and subjected to MS-based proteome profiling. DISCO-MS should be of great interest to researchers with archived solvent-cleared organs and imaging data, where molecular data are missing. This method will further enable to interrogate the molecular basis of a pathological milieu located in a seemingly random manner to advance research in complex diseases.

### Limitations of our study

Although we have demonstrated that DISCO-MS enabled us to perform unbiased proteome analysis after unbiased imaging of entire specimens in 3D, further developments would be needed to extract fewer cells for downstream proteome analyses. Our current method works well for down to about 60 cells and provides us with average molecular details of these cells and their microenvironments. While this is sufficient to study diverse biological questions such as plaque heterogeneity, it is still far from the single-cell level. Single-cell resolution is important to understand the cell compositions and cellular interactions in the biological processes under investigation. To reach that goal, the current protocol needs to be refined for smaller sample

(H and I) Coarse calibration with mouse in imaging chamber prior to extraction.
(J–M) Robotic extraction example: (J) target selection, (K) needle approaching the target, (L) extracting the target under vacuum, (M) evaluation of extracted region.
See also Figure S6, Methods S1, and Videos S2 and S5

CellPress

Cell
Resource

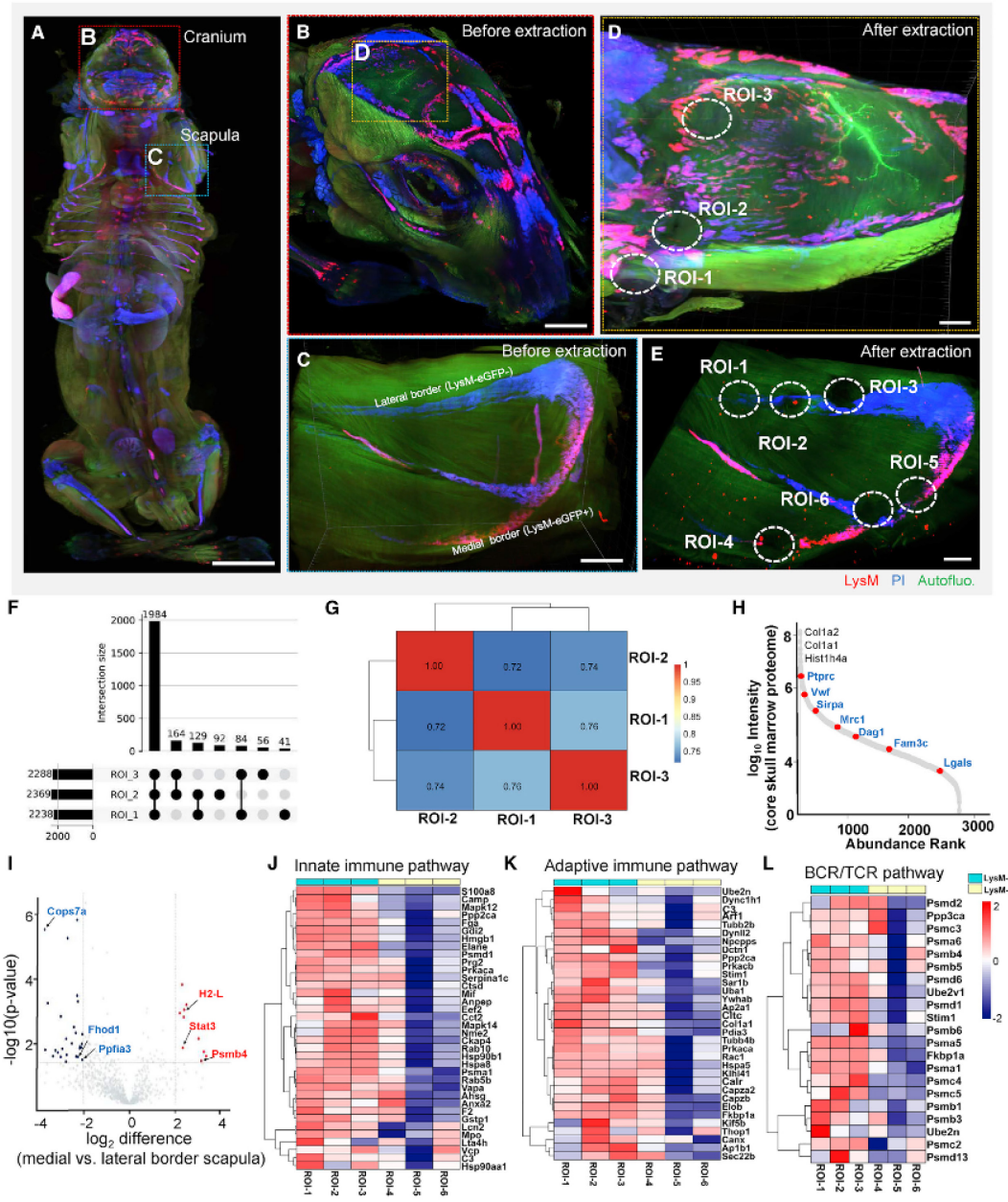



**Figure 6. DISCO-bot-aided DISCO-MS on bone marrow niches of whole mouse bodies**

(A) Stiched images of whole body of LysM-EGFP mouse after clearing and imaging (n = 4). Scale bar, 700 μm.

(B) 3D visualization of stiched images of whole head zooming into parietal surface before extraction.

(C) 3D visualization of stiched images of right scapula showing lateral and medial borders before extraction (PI in blue, LysM in red). Scale bar, 500 μm.

(D) Marked 3 ROIs in parietal surface after robotic extraction. Scale bar, 500 μm.

(E) Marked 6 ROIs, 3 in lateral border and 3 in medial border after extraction. Scale bar, 500 μm.

*(legend continued on next page)*

volumes to extract sub-nanogram tissues from the cleared samples. Furthermore, the proteomic sample preparation for cleared tissues could be miniaturized to the level of single cells or pooled cell states without losing proteomics depth.[17]

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
  - Human samples
- METHOD DETAILS
  - Repeated closed head mild traumatic brain injury (mTBI)
  - Perfusion and tissue preparation
  - Congo red labeling of whole brains of 5xFAD animals
  - Clearing of brains using DISCO methods
  - vDISCO whole-brain passive immunostaining, clearing and imaging
  - vDISCO whole body immuno labeling, clearing and imaging
  - SHANEL sample preparation and clearing
  - Behavioral assessment
  - Immunofluorescence and confocal microscopy
  - Light-sheet microscopy and image processing
  - Registration of light-sheet and LCM images to correlate probe selection
  - Optimization of cleared tissue for cryopreservation and sectioning
  - Laser-capture microdissection
  - Automated robotic proteomic probe extraction from whole body
  - Precision and accuracy of sampling
  - Needle gauge
  - Specimen holder and needle holder strategy
  - Embedding of the whole mouse body
  - Manipulation system
  - Optimization of DISCO cleared sample preparation for mass spectrometry analysis
  - High-pH reversed-phase fractionation
  - Liquid chromatography and mass spectrometry (LC-MS)
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ClearMap quantification
  - Deep learning analyses
  - Proteomics data processing
  - Proteomics downstream data analysis
- ADDITIONAL RESOURCES

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2022.11.021.

A video abstract is available at https://doi.org/10.1016/j.cell.2022.11.021#mmc9.

## ACKNOWLEDGMENTS

A.-D.B. acknowledges support from the International Max Planck Research School for Life Sciences (IMPRS-LS). This work was supported by the Vascular Dementia Research Foundation, Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the framework of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy, ID 390857198), the Bundesministerium für Bildung und Wissenschaft (BMBF) for Grant CLINSPECT-M (FKZ161L0214C, to S.F.L. and M.M.), ERC Consolidator Grant (AE, GA 865323), Nomis Heart Atlas Project Grant (Nomis Foundation), the Helmholtz AI program through grant Deeproad, and the Max Planck Society for the Advancement of Science. We thank our colleagues in the Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, for discussions and help. In particular, we thank I. Paron, A. Piras, and C. Deiml for technical support and J.B. Müller for column production. We thank Dr. Rupinder Kaur for her support. Some of the graphical illustrations used in the manuscript were prepared using BioRender.com. We thank Mr. Marin Bralo at iTERM for his technical support throughout this study.

## AUTHOR CONTRIBUTIONS

A.E., M. Mann, H.S.B., and A.-D.B. conceptualized and designed the study. H.S.B., Z.R., and H.M. performed mice experiments, solvent-based organ clearing, light-sheet imaging procedures, and stitching of data. H.S.B. developed the LCM-based isolation procedure for target tissues from solvent-cleared organs. A.-D.B. developed the sample preparation workflow for proteomics analysis. A.-D.B., S.K., H.S.B., M. Molbay, and M.T. performed mass spectrometry-based proteomics analysis. R.A.-M., J.C.P., F.K., M.A., B.M., and F.T. developed deep learning models. M.A. and F.T. performed data analysis. M.I.T. performed atlas registration of 5xFAD brains. H.S.B. and D.K. performed ClearMAP analyses. A.E. conceptualized, and F.O. designed the robotic extraction system. Z.I.K. labeled, cleared, and imaged whole body of mouse. F.O., H.S.B., and S.K. extracted tissue from whole body and whole organs. S.K. performed mass spectrometry on robotic extracted samples. H.S.B. and S.K. analyzed the data. M.N., J.C.P., and A.D. performed 2D LCM registration to 3D data. S.M. and S.F.L. helped with the prototyping experiment. H.S.B., A.-D.B., M. Mann, and A.E. wrote the manuscript.

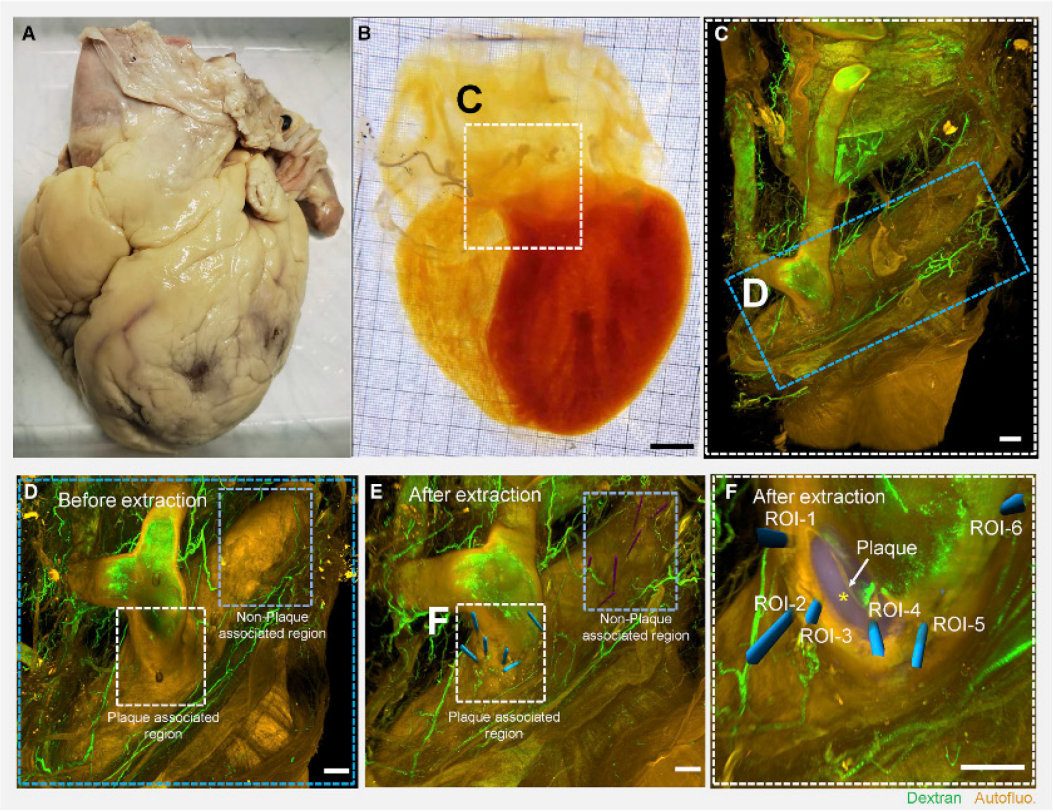
## DECLARATION OF INTERESTS

F.T. reports receiving consulting fees from Roche Diagnostics GmbH and Cellarity, Inc., and ownership interest in Cellarity, Inc., and Dermagnostix. A.E.,

(F) The number of shared and unique set of proteins in cranium.
(G) Pearson correlation among cranium samples (n = 3 experimental replicates).
(H) Rank order of core protein abundance signals in cranium.
(I) Volcano plot showing the significantly enriched proteins in medial vs. lateral border of scapula.
(J) Heatmap showing the proteins related to innate immune system in medial vs. lateral border of scapula.
(K) Proteins related to adaptive immune system.
(L) Related to B cell receptor/T cell receptor pathway in scapula bone marrow niches.
See also Figure S7 and Video S6.

CellPress

Cell
Resource



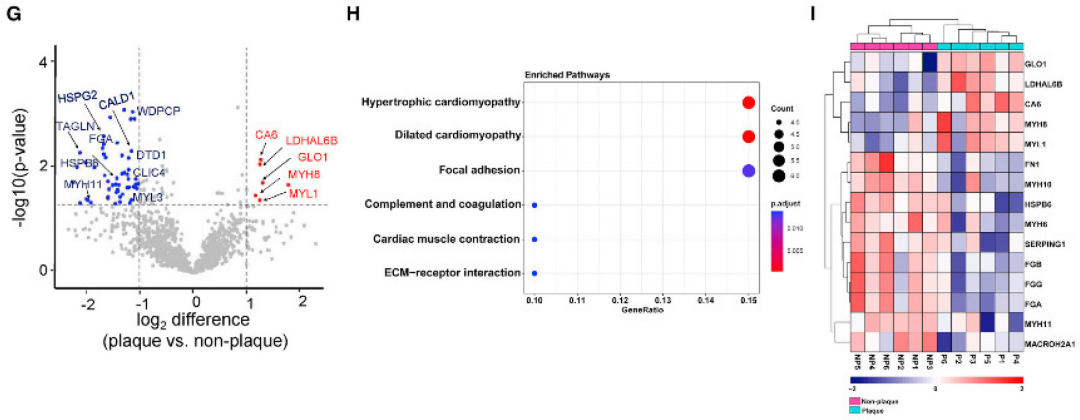




**Figure 7. Human heart clearing, imaging, and proteomics from atherosclerotic plaques in coronary artery**

(A) Image of whole human heart before clearing.

(B) Image of human heart after SHANEL clearing. Scale bar, 2 mm.

(C) 3D visualization of stiched images of portion of human heart showing dextran-labeled vessels (in green), plaques as autofluorescence (in gold). Scale bar, 1,000 μm.

(D) Stiched images of coronal artery with atherosclerotic plaque-associated regions and non-plaque regions before robotic guided extraction. Scale bar, 1,000 μm.

*(legend continued on next page)*

H.S.B., F.O., S.K., M.I.T., and M.N. have filed a patent related to robotics technologies presented in this work.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research. We worked to ensure sex balance in the selection of non-human subjects. We worked to ensure diversity in experimental samples through the selection of the genomic datasets. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. One or more of the authors of this paper self-identifies as a gender minority in their field of research. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community. We avoided "helicopter science" practices by including the participating local contributors from the region where we conducted the research as authors on the paper.

Received: May 3, 2021
Revised: June 13, 2022
Accepted: November 18, 2022
Published: December 22, 2022

## REFERENCES

1. Braak, H., Thal, D.R., Ghebremedhin, E., and Del Tredici, K. (2011). Stages of the Pathologic Process in Alzheimer Disease: Age Categories From 1 to 100 Years. J. Neuropathol. Exp. Neurol. *70*, 960–969. https://doi.org/10.1097/NEN.0b013e318232a379.
2. Belle, M., Godefroy, D., Couly, G., Malone, S.A., Collier, F., Giacobini, P., and Chédotal, A. (2017). Tridimensional Visualization and Analysis of Early Human Development. Cell *169*, 161–173.e12. https://doi.org/10.1016/j.cell.2017.03.008.
3. Ueda, H.R., Ertürk, A., Chung, K., Gradinaru, V., Chédotal, A., Tomancak, P., and Keller, P.J. (2020). Tissue clearing and its applications in neuroscience. Nat. Rev. Neurosci. *21*, 61–79. https://doi.org/10.1038/s41583-019-0250-1.
4. Zhao, S., Todorov, M.I., Cai, R., -Maskari, R.A., Steinke, H., Kemter, E., Mai, H., Rong, Z., Warmer, M., Stanic, K., Schoppe, O., Paetzold, J.C., Gesierich, B., Wong, M.N., Huber, T.B., Duering, M., Bruns, O.T., Menze, B., Lipfert, J., Puelles, V.G., Wolf, E., Bechmann, I., and Erturk, A. (2020). Cellular and Molecular Probing of Intact Human Organs. Cell *180*, 796–812.e19. https://doi.org/10.1016/j.cell.2020.01.030.
5. Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., and Van Valen, D. (2019). Deep learning for cellular image analysis. Nat. Methods *16*, 1233–1246. https://doi.org/10.1038/s41592-019-0403-1.
6. Sullivan, D.P., Winsnes, C.F., Åkesson, L., Hjelmare, M., Wiking, M., Schutten, R., Campbell, L., Leifsson, H., Rhodes, S., Nordgren, A., Smith, K., Revaz, B., Finnbogason, B., Szantner, A., and Lundberg, E. (2018). Deep learning is combined with massive-scale citizen science to improve large-scale image classification. Nat. Biotechnol. *36*, 820–828. https://doi.org/10.1038/nbt.4225.
7. Ji, A.L., Rubin, A.J., Thrane, K., Jiang, S., Reynolds, D.L., Meyers, R.M., Guo, M.G., George, B.M., Mollbrink, A., Bergenstråhle, J., Larsson, L., Bai, Y., Zhu, B., Bhaduri, A., Meyers, J.M., Rovira-Clave, X., Hollmig, S.T., Aasi, S.Z., Nolan, G.P., Lundeberg, J., and Khavari, P.A. (2020). Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. Cell *182*, 497–514.e22. https://doi.org/10.1016/j.cell.2020.05.039.
8. Stuart, T., and Satija, R. (2019). Integrative single-cell analysis. Nat. Rev. Genet. *20*, 257–272. https://doi.org/10.1038/s41576-019-0093-7.
9. VanHorn, S., and Morris, S.A. (2021). Next-Generation Lineage Tracing and Fate Mapping to Interrogate Development. Dev. Cell *56*, 7–21. https://doi.org/10.1016/j.devcel.2020.10.021.
10. Brunner, A.D., Thielert, M., Vasilopoulou, C., Ammar, C., Coscia, F., Mund, A., Hoerning, O.B., Bache, N., Apalategui, A., Lubeck, M., Richter, S., Fischer, D.S., Raether, O., Park, M.A., Meier, F., Theis, F.J., and Mann, M. (2022). Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. Mol. Syst. Biol. *18*, e10798. https://doi.org/10.15252/msb.202110798.
11. Reimegård, J., Tarbier, M., Danielsson, M., Schuster, J., Baskaran, S., Panagiotou, S., Dahl, N., Friedländer, M.R., and Gallant, C.J. (2021). A combined approach for single-cell mRNA and intracellular protein expression analysis. Commun Biol *4*, 624. https://doi.org/10.1038/s42003-021-02142-w.
12. Johnson, E.C.B., Carter, E.K., Dammer, E.B., Duong, D.M., Gerasimov, E.S., Liu, Y., Liu, J., Betarbet, R., Ping, L., Yin, L., Serrano, G.E., Beach, T.G., Peng, J., De Jager, P.L., Haroutunian, V., Zhang, B., Gaiteri, C., Bennett, D.A., Gearing, M., Wingo, T.S., Wingo, A.P., Lah, J.J., Levey, A.I., and Seyfried, N.T. (2022). Large-scale deep multi-layer analysis of Alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level. Nat. Neurosci. *25*, 213–225. https://doi.org/10.1038/s41593-021-00999-y.
13. Wegler, C., Ölander, M., Wiśniewski, J.R., Lundquist, P., Zettl, K., Åsberg, A., Hjelmesæth, J., Andersson, T.B., and Artursson, P. (2020). Global variability analysis of mRNA and protein concentrations across and within human tissues. NAR Genom Bioinform *2*, lqz010. https://doi.org/10.1093/nargab/lqz010.
14. Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018). SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. *19*, 161. https://doi.org/10.1186/s13059-018-1547-5.
15. Cheung, T.K., Lee, C.-Y., Bayer, F.P., McCoy, A., Kuster, B., and Rose, C.M. (2021). Defining the carrier proteome limit for single-cell proteomics. Nat. Methods *18*, 76–83. https://doi.org/10.1038/s41592-020-01002-5.
16. Cong, Y., Motamedchaboki, K., Misal, S.A., Liang, Y., Guise, A.J., Truong, T., Huguet, R., Plowey, E.D., Zhu, Y., Lopez-Ferrer, D., et al. (2021). Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. Chem. Sci. *12*, 1001–1006. https://doi.org/10.1039/D0SC03636F.
17. Mund, A., Coscia, F., Kriston, A., Hollandi, R., Kovács, F., Brunner, A.-D., Migh, E., Schweizer, L., Santos, A., Bzorek, M., Naimy, S., Rahbek-Gjerdrum, L.M., Dyring-Andersen, B., Bulkescher, J., Lukas, C., Eckert, M.A., Lengyel, E., Gnann, C., Lundberg, E., Horvath, P., and Mann, M. (2022). Deep Visual Proteomics defines single-cell identity and heterogeneity. Nat. Biotechnol. *40*, 1231–1240. https://doi.org/10.1038/s41587-022-01302-5.
18. Alon, S., Goodwin, D.R., Sinha, A., Wassie, A.T., Chen, F., Daugharthy, E.R., Bando, Y., Kajita, A., Xue, A.G., Marrett, K., Prior, R., Cui, Y., Payne, A.C., Yao, C.C., Suk, H.J., Wang, R., Yu, C.C.J., Tillberg, P., Reginato, P., Pak, N., Liu, S., Punthambaker, S., Iyer, E.P.R., Kohman, R.E., Miller, J.A., Lein, E.S., Lako, A., Cullen, N., Rodig, S., Helvie, K., Abravanel, D.L., Wagle, N., Johnson, B.E., Klughammer, J., Slyper, M., Waldman, J., Jane-Valbuena, J., Rozenblatt-Rosen, O., Regev, A., IMAXT Consortium, Church, G.M., Marblestone, A.H., Boyden, E.S., Al Sa'd, M., Alon, S.,

(E) Stiched images of coronary artery after extraction of ROIs, extraction points are shown in blue (plaque region) and in magenta (non-plaque region). Scale bar, 1,000 μm.
(F) Enlarged view of ROIs around the segmented plaque marked in (E). Scale bar, 1,000 μm.
(G) Volcano plot showing the significantly enriched proteins in plaque (ROIs 1–6) vs. non-plaque areas.
(H) Pathway enrichment of differentially regulated proteins.
(I) Heatmap of differentially regulated proteins in each ROIs.
See also Figure S7 and Video S7.

Aparicio, S., Battistoni, G., Balasubramanian, S., Becker, R., Bodenmiller, B., Boyden, E.S., Bressan, D., Bruna, A., Burger, M., Caldas, C., Callari, M., Cannell, I.G., Casbolt, H., Chornay, N., Cui, Y., Dariush, A., Dinh, K., Emenari, A., Eyal-Lubling, Y., Fan, J., Fatemi, A., Fisher, E., Gonzalez-Solares, E.A., Gonzalez-Fernandez, C., Goodwin, D., Greenwood, W., Grimaldi, F., Hannon, G.J., Harris, O., Harris, S., Jauset, C., Joyce, J.A., Karagiannis, E.D., Kovacevic, T., Kuett, L., Kunes, R., Kupcu Yoldas, A., Lai, D., Laks, E., Lee, H., Lee, M., Lerda, G., Li, Y., McPherson, A., Millar, N., Mulvey, C.M., Nugent, F., O'Flanagan, C.H., Paez-Ribes, M., Pearsall, I., Qosaj, F., Roth, A.J., Rueda, O.M., Ruiz, T., Sawicka, K., Sepulveda, L.A., Shah, S.P., Shea, A., Sinha, A., Smith, A., Tavare, S., Tietscher, S., Vazquez-Garcia, I., Vogl, S.L., Walton, N.A., Wassie, A.T., Watson, S.S., Weselak, J., Wild, S.A., Williams, E., Windhager, J., Whitmarsh, T., Xia, C., Zheng, P., and Zhuang, X. (2021). Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. Science *371*, eaax2656. https://doi.org/10.1126/science.aax2656.

19. van den Brink, S.C., Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivié, J., Baillie-Johnson, P., Nichols, J., Sonnen, K.F., Martinez Arias, A., and van Oudenaarden, A. (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. Nature *582*, 405–409. https://doi.org/10.1038/s41586-020-2024-3.

20. Chen, W.-T., Lu, A., Craessaerts, K., Pavie, B., Sala Frigerio, C., Corthout, N., Qian, X., Laláková, J., Kühnemund, M., Voytyuk, I., Wolfs, L., Mancuso, R., Salta, E., Balusu, S., Snellinx, A., Munck, S., Jurek, A., Fernandez Navarro, J., Saido, T.C., Huitinga, I., Lundeberg, J., Fiers, M., and De Strooper, B. (2020). Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. Cell *182*, 976–991.e19. https://doi.org/10.1016/j.cell.2020.06.038.

21. Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., Norris, E., Pan, A., Li, J., Xiao, Y., Halene, S., and Fan, R. (2020). High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. Cell *183*, 1665–1681.e18. https://doi.org/10.1016/j.cell.2020.10.026.

22. Mahdessian, D., Cesnik, A.J., Gnann, C., Danielsson, F., Stenström, L., Arif, M., Zhang, C., Le, T., Johansson, F., Schutten, R., Backstrom, A., Axelsson, U., Thul, P., Cho, N.H., Carja, O., Uhlen, M., Mardinoglu, A., Stadler, C., Lindskog, C., Ayoglu, B., Leonetti, M.D., Ponten, F., Sullivan, D.P., and Lundberg, E. (2021). Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. Nature *590*, 649–654. https://doi.org/10.1038/s41586-021-03232-9.

23. Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science *363*, 1463–1467. https://doi.org/10.1126/science.aaw1219.

24. Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, A., Ponten, F., Costea, P.I., Sahlen, P., Mulder, J., Bergmann, O., Lundeberg, J., and Frisen, J. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82. https://doi.org/10.1126/science.aaf2403.

25. Takei, Y., Yun, J., Zheng, S., Ollikainen, N., Pierson, N., White, J., Shah, S., Thomassie, J., Suo, S., Eng, C.-H.L., Guttman, M., Yuan, G.C., and Cai, L. (2021). Integrated spatial genomics reveals global architecture of single nuclei. Nature *590*, 344–350. https://doi.org/10.1038/s41586-020-03126-2.

26. Giesen, C., Wang, H.A.O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P.J., Grolimund, D., Buhmann, J.M., Brandt, S., Varga, Z., Wild, P.J., Gunther, D., and Bodenmiller, B. (2014). Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat. Methods *11*, 417–422. https://doi.org/10.1038/nmeth.2869.

27. Ko, J., Wilkovitsch, M., Oh, J., Kohler, R.H., Bolli, E., Pittet, M.J., Vinegoni, C., Sykes, D.B., Mikula, H., Weissleder, R., and Carlson, J.C.T. (2022). Spatiotemporal multiplexed immunofluorescence imaging of living cells and tissues with bioorthogonal cycling of fluorescent probes. Nat. Biotechnol. *40*, 1654–1662. https://doi.org/10.1038/s41587-022-01339-6.

28. Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S., and Nolan, G.P. (2018). Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. Cell *174*, 968–981.e15. https://doi.org/10.1016/j.cell.2018.07.010.

29. Saka, S.K., Wang, Y., Kishi, J.Y., Zhu, A., Zeng, Y., Xie, W., Kirli, K., Yapp, C., Cicconet, M., Beliveau, B.J., Lapan, S.W., Yin, S., Lin, M., Boyden, E.S., Kaeser, P.S., Pihan, G., Church, G.M., and Yin, P. (2019). Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. Nat. Biotechnol. *37*, 1080–1090. https://doi.org/10.1038/s41587-019-0207-y.

30. Radtke, A.J., Kandov, E., Lowekamp, B., Speranza, E., Chu, C.J., Gola, A., Thakur, N., Shih, R., Yao, L., Yaniv, Z.R., Beuschel, R.T., Kabat, J., Croteau, J., Davis, J., Hernandez, J.M., and Germain, R.N. (2020). IBEX: A versatile multiplex optical imaging approach for deep phenotyping and spatial analysis of cells in complex tissues. Proc. Natl. Acad. Sci. USA. *117*, 33455–33465. https://doi.org/10.1073/pnas.2018488117.

31. Ueda, H.R., Dodt, H.-U., Osten, P., Economo, M.N., Chandrashekar, J., and Keller, P.J. (2020). Whole-Brain Profiling of Cells and Circuits in Mammals by Tissue Clearing and Light-Sheet Microscopy. Neuron *106*, 369–387. https://doi.org/10.1016/j.neuron.2020.03.004.

32. Aebersold, R., and Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. Nature *537*, 347–355. https://doi.org/10.1038/nature19949.

33. Richardson, D.S., Guan, W., Matsumoto, K., Pan, C., Chung, K., Ertürk, A., Ueda, H.R., and Lichtman, J.W. (2021). TISSUE CLEARING. Nat Rev Methods Primers *1*, 84. https://doi.org/10.1038/s43586-021-00080-9.

34. Cai, R., Pan, C., Ghasemigharagoz, A., Todorov, M.I., Förstera, B., Zhao, S., Bhatia, H.S., Parra-Damas, A., Mrowka, L., Theodorou, D., Rempfler, M., Xavier, A.L.R., Kress, B.T., Benakis, C., Steinke, H., Liebscher, S., Bechmann, I., Liesz, A., Menze, B., Kerschensteiner, M., Nedergaard, M., and Erturk, A. (2019). Panoptic imaging of transparent mice reveals whole-body neuronal projections and skull-meninges connections. Nat. Neurosci. *22*, 317–327. https://doi.org/10.1038/s41593-018-0301-3.

35. Langlois, J.A., Rutland-Brown, W., and Thomas, K.E. (2005). The incidence of traumatic brain injury among children in the United States: differences by race. J. Head Trauma Rehabil. *20*, 229–238. https://doi.org/10.1097/00001199-200505000-00006.

36. Ertürk, A., Mentz, S., Stout, E.E., Hedehus, M., Dominguez, S.L., Neumaier, L., Krammer, F., Llovera, G., Srinivasan, K., Hansen, D.V., Liesz, A., Scearce-Levie, K.A., and Sheng, M. (2016). Interfering with the Chronic Immune Response Rescues Chronic Degeneration After Traumatic Brain Injury. J. Neurosci. *36*, 9962–9975. https://doi.org/10.1523/JNEUROSCI.1898-15.2016.

37. Renier, N., Adams, E.L., Kirst, C., Wu, Z., Azevedo, R., Kohl, J., Autry, A.E., Kadiri, L., Umadevi Venkataraju, K., Zhou, Y., Wang, V.X., Tang, C.Y., Olsen, O., Dulac, C., Osten, P., and Tessier-Lavigne, M. (2016). Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes. Cell *165*, 1789–1802. https://doi.org/10.1016/j.cell.2016.05.007.

38. Chen, M., Song, H., Cui, J., Johnson, C.E., Hubler, G.K., DePalma, R.G., Gu, Z., and Xia, W. (2018). Proteomic Profiling of Mouse Brains Exposed to Blast-Induced Mild Traumatic Brain Injury Reveals Changes in Axonal Proteins and Phosphorylated Tau. J Alzheimers Dis *66*, 751–773. https://doi.org/10.3233/JAD-180726.

39. Asher, R.A., Morgenstern, D.A., Fidler, P.S., Adcock, K.H., Oohira, A., Braistead, J.E., Levine, J.M., Margolis, R.U., Rogers, J.H., and Fawcett, J.W. (2000). Neurocan is upregulated in injured brain and in cytokine-treated astrocytes. J. Neurosci. *20*, 2427–2438. https://doi.org/10.1523/jneurosci.20-07-02427.2000.

40. Alaamery, M., Albesher, N., Aljawini, N., Alsuwailm, M., Massadeh, S., Wheeler, M.A., Chao, C., and Quintana, F.J. (2021). Role of sphingolipid metabolism in neurodegeneration. J. Neurochem. *158*, 25–35. https://doi.org/10.1111/jnc.15044.

41. Gebai, A., Gorelik, A., Li, Z., Illes, K., and Nagar, B. (2018). Structural basis for the activation of acid ceramidase. Nat. Commun. *9*, 1621. https://doi.org/10.1038/s41467-018-03844-2.

42. Parveen, F., Bender, D., Law, S.-H., Mishra, V.K., Chen, C.-C., and Ke, L.-Y. (2019). Role of Ceramidases in Sphingolipid Metabolism and Human Diseases. Cells *8*, E1573. https://doi.org/10.3390/cells8121573.

43. Kyriakou, K., W Lederer, C., Kleanthous, M., Drousiotou, A., and Malekkou, A. (2020). Acid Ceramidase Depletion Impairs Neuronal Survival and Induces Morphological Defects in Neurites Associated with Altered Gene Transcription and Sphingolipid Content. Int. J. Mol. Sci. *21*, E1607. https://doi.org/10.3390/ijms21051607.

44. Rumyantseva, A., Popovic, M., and Trifunovic, A. (2022). CLPP deficiency ameliorates neurodegeneration caused by impaired mitochondrial protein synthesis. Brain *145*, 92–104. https://doi.org/10.1093/brain/awab303.

45. Shi, Y., Yi, C., Li, X., Wang, J., Zhou, F., and Chen, X. (2017). Overexpression of Mitofusin2 decreased the reactive astrocytes proliferation in vitro induced by oxygen-glucose deprivation/reoxygenation. Neurosci. Lett. *639*, 68–73. https://doi.org/10.1016/j.neulet.2016.12.052.

46. Meyer-Luehmann, M., Spires-Jones, T.L., Prada, C., Garcia-Alloza, M., de Calignon, A., Rozkalne, A., Koenigsknecht-Talboo, J., Holtzman, D.M., Bacskai, B.J., and Hyman, B.T. (2008). Rapid appearance and local toxicity of amyloid-beta plaques in a mouse model of Alzheimer's disease. Nature *451*, 720–724. https://doi.org/10.1038/nature06616.

47. Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. https://doi.org/10.48550/arXiv.1505.04597.

48. Todorov, M.I., Paetzold, J.C., Schoppe, O., Tetteh, G., Shit, S., Efremov, V., Todorov-Völgyi, K., Düring, M., Dichgans, M., Piraud, M., Menze, B., and Erturk, A. (2020). Machine learning analysis of whole mouse brain vasculature. Nat. Methods *17*, 442–449. https://doi.org/10.1038/s41592-020-0792-1.

49. Wang, Q., Ding, S.-L., Li, Y., Royall, J., Feng, D., Lesnar, P., Graddis, N., Naeemi, M., Facer, B., Ho, A., Dolbeare, T., Blanchard, B., Dee, N., Wakeman, W., Hirokawa, K.E., Szafer, A., Sunkin, S.M., Oh, S.W., Bernard, A., Phillips, J.W., Hawrylycz, M., Koch, C., Zeng, H., Harris, J.A., and Ng, L. (2020). The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. Cell *181*, 936–953.e20. https://doi.org/10.1016/j.cell.2020.04.007.

50. Claudi, F., Tyson, A.L., Petrucco, L., Margrie, T.W., Portugues, R., and Branco, T. (2021). Visualizing anatomically registered data with brainrender. Elife *10*, e65751. https://doi.org/10.7554/eLife.65751.

51. Lichtenthaler, S.F. (2006). Ectodomain shedding of the amyloid precursor protein: cellular control mechanisms and novel modifiers. Neurodegener. Dis. *3*, 262–269. https://doi.org/10.1159/000095265.

52. Meckelein, B., de Silva, H., Roses, A.D., Rao, P., Pettenati, M.J., Xu, P.T., Hodge, R., Glucksman, M.J., and Abraham, C.R. (1996). Human endopeptidase (THOP1) is localized on chromosome 19 within the linkage region for the late-onset alzheimer disease AD2 locus. Genomics *31*, 246–249. https://doi.org/10.1006/geno.1996.0041.

53. Pollio, G., Hoozemans, J.J., Andersen, C.A., Roncarati, R., Rosi, M.C., van Haastert, E.S., Seredenina, T., Diamanti, D., Gotta, S., Fiorentini, A., Magnoni, L., Raggiaschi, R., Rozemuller, A.J., Casamenti, F., Caricasole, A., and Terstappen, G.C. (2008). Increased expression of the oligopeptidase THOP1 is a neuroprotective response to Aβ toxicity. Neurobiol. Dis. *31*, 145–158. https://doi.org/10.1016/j.nbd.2008.04.004.

54. Xu, S., Di, Z., He, Y., Wang, R., Ma, Y., Sun, R., Li, J., Wang, T., Shen, Y., Fang, S., Feng, L., and Shen, Y. (2019). Mesencephalic astrocyte-derived neurotrophic factor (MANF) protects against Aβ toxicity via attenuating Aβ-induced endoplasmic reticulum stress. J. Neuroinflammation *16*, 35. https://doi.org/10.1186/s12974-019-1429-0.

55. Arisi, I., D'Onofrio, M., Brandi, R., Felsani, A., Capsoni, S., Drovandi, G., Felici, G., Weitschek, E., Bertolazzi, P., and Cattaneo, A. (2011). Gene expression biomarkers in the brain of a mouse model for Alzheimer's disease: mining of microarray data by logic classification and feature selection. J Alzheimers Dis *24*, 721–738. https://doi.org/10.3233/JAD-2011-101881.

56. Marom, R., Jain, M., Burrage, L.C., Song, I.-W., Graham, B.H., Brown, C.W., Stevens, S.J., Stegmann, A.P., Gunter, A.T., Kaplan, J.D., Gavrilova, R.H., Shinawi, M., Rosenfeld, J.A., Bae, Y., Tran, A.A., Chen, Y., Lu, J.T., Gibbs, R.A., Eng, C., Yang, Y., Rousseau, J., de Vries, B.B., Campeau, P.M., and Lee, B. (2017). Heterozygous variants in ACTL6A, encoding a component of the BAF complex, are associated with intellectual disability. Hum. Mutat. *38*, 1365–1371. https://doi.org/10.1002/humu.23282.

57. Depp, C., Sun, T., Sasmita, A.O., Spieth, L., Berghoff, S.A., Steixner-Kumar, A.A., Subramanian, S., Möbius, W., Göbbels, S., Saher, G., Zampar, S., Wirths, O., Thalmann, M., Saito, T., Saido, T., Krueger-Burg, D., Kawaguchi, R., Willem, M., Haass, C., Geschwind, D., Ehrenreich, H., Stassart, R., and Nave, K.A. (2021). Ageing-associated myelin dysfunction drives amyloid deposition in mouse models of Alzheimer's disease. https://doi.org/10.1101/2021.07.31.454562.

58. Kolabas, Z.I., Kuemmerle, L.B., Perneczky, R., Förstera, B., Büttner, M., Caliskan, O.S., Ali, M., Rong, Z., Mai, H., Hummel, S., Bartos, L.M., Biechele, G., Zatcepin, A., Albert, N.L., Unterrainer, M., Gnorich, J., Zhao, S., Khalin, I., Rauchmann, B.S., Molbay, M., Sterr, M., Kunze, I., Stanic, K., Besson-Girard, S., Kopczak, A., Katzdobler, S., Palleis, C., Gokce, O., Lickert, H., Steinke, H., Bechmann, I., Buerger, K., Levin, J., Haass, C., Dichgans, M., Havla, J., Kumpfel, T., Kerschensteiner, M., Simons, M., Plesnila, N., Krahmer, N., Bhatia, H.S., Erener, S., Hellal, F., Brendel, M., Theis, F.J., and Erturk, A. (2021). Multi-omics and 3D-imaging reveal bone heterogeneity and unique calvaria cells in neuroinflammation. https://doi.org/10.1101/2021.12.24.473988.

59. Nabel, E.G., and Braunwald, E. (2012). A tale of coronary artery disease and myocardial infarction. N. Engl. J. Med. *366*, 54–63. https://doi.org/10.1056/NEJMra1112570.

60. Wong, N.D. (2014). Epidemiological studies of CHD and the evolution of preventive cardiology. Nat. Rev. Cardiol. *11*, 276–289. https://doi.org/10.1038/nrcardio.2014.26.

61. Fernandez, D.M., Rahman, A.H., Fernandez, N.F., Chudnovskiy, A., Amir, E.-A.D., Amadori, L., Khan, N.S., Wong, C.K., Shamailova, R., Hill, C.A., Wang, Z., Remark, R., Li, J.R., Pina, C., Faries, C., Awad, A.J., Moss, N., Bjorkegren, J.L.M., Kim-Schulze, S., Gnjatic, S., Ma'ayan, A., Mocco, J., Faries, P., Merad, M., and Giannarelli, C. (2019). Single-cell immune landscape of human atherosclerotic plaques. Nat Med *25*, 1576–1588. https://doi.org/10.1038/s41591-019-0590-4.

62. Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C.L., Lindberg, E.L., Kanda, M., Polanski, K., Heinig, M., Lee, M., Nadelmann, E.R., Roberts, K., Tuck, L., Fasouli, E.S., DeLaughter, D.M., McDonough, B., Wakimoto, H., Gorham, J.M., Samari, S., Mahbubani, K.T., Saeb-Parsy, K., Patone, G., Boyle, J.J., Zhang, H., Zhang, H., Viveiros, A., Oudit, G.Y., Bayraktar, O.A., Seidman, J.G., Seidman, C.E., Noseda, M., Hubner, N., and Teichmann, S.A. (2020). Cells of the adult human heart. Nature *588*, 466–472. https://doi.org/10.1038/s41586-020-2797-4.

63. Meng, L.B., Li, J.Y., Xu, H.X., Wu, D.S., Shan, M.J., Chen, Y.H., Xu, J.P., Liu, L.T., Chen, Z., Li, Y.J., Gong, T., and Liu, D. (2022). A potential biomarker for clinical atherosclerosis: A novel insight derived from myosin heavy chain 10 promoting transformation of vascular smooth muscle cells. Clin. Transl. Med. *12*, e672. https://doi.org/10.1002/ctm2.672.

64. Takahashi, L., Ishigami, T., Tomiyama, H., Kato, Y., Kikuchi, H., Tasaki, K., Yamashita, J., Inoue, S., Taguri, M., Nagao, T., Chikamori, T., Ishikawa, Y., and Yokoyama, U. (2021). Increased Plasma Levels of Myosin Heavy Chain 11 Is Associated with Atherosclerosis. J. Clin. Med. *10*, 3155. https://doi.org/10.3390/jcm10143155.

65. Mäkinen, V.-P., Civelek, M., Meng, Q., Zhang, B., Zhu, J., Levian, C., Huan, T., Segrè, A.V., Ghosh, S., Vivar, J., Nikpay, M., Stewart, A.F.R., Nelson, C.P., Willenborg, C., Erdmann, J., Blakenberg, S., O'Donnell, C.J., Marz, W., Laaksonen, R., Epstein, S.E., Kathiresan, S., Shah, S.H., Hazen, S.L., Reilly, M.P., Lusis, A.J., Samani, N.J., Schunkert, H., Quertermous, T., McPherson, R., Yang, X., and Assimes, T.L. (2014). Integrative genomics reveals novel molecular pathways and gene networks for coronary

artery disease. PLoS Genet. *10*, e1004502. https://doi.org/10.1371/journal.pgen.1004502.

66. Adeva-Andany, M.M., Fernández-Fernández, C., Sánchez-Bello, R., Donapetry-García, C., and Martínez-Rodríguez, J. (2015). The role of carbonic anhydrase in the pathogenesis of vascular calcification in humans. Atherosclerosis *241*, 183–191. https://doi.org/10.1016/j.atherosclerosis.2015.05.012.
67. Yuan, L., Wang, M., Liu, T., Lei, Y., Miao, Q., Li, Q., Wang, H., Zhang, G., Hou, Y., and Chang, X. (2019). Carbonic Anhydrase 1-Mediated Calcification Is Associated With Atherosclerosis, and Methazolamide Alleviates Its Pathogenesis. Front. Pharmacol. *10*, 766. https://doi.org/10.3389/fphar.2019.00766.
68. Robichaud, S., Fairman, G., Vijithakumar, V., Mak, E., Cook, D.P., Pelletier, A.R., Huard, S., Vanderhyden, B.C., Figeys, D., Lavallée-Adam, M., Baetz, K., and Ouimet, M. (2021). Identification of novel lipid droplet factors that regulate lipophagy and cholesterol efflux in macrophage foam cells. Autophagy *17*, 3671–3689. https://doi.org/10.1080/15548627.2021.1886839.
69. Whitesell, J.D., Buckley, A.R., Knox, J.E., Kuan, L., Graddis, N., Pelos, A., Mukora, A., Wakeman, W., Bohn, P., Ho, A., Hirokawa, K.E., and Harris, J.A. (2019). Whole brain imaging reveals distinct spatial patterns of amyloid beta deposition in three mouse models of Alzheimer's disease. J. Comp. Neurol. *527*, 2122–2145. https://doi.org/10.1002/cne.24555.
70. Oddo, S., Caccamo, A., Shepherd, J.D., Murphy, M., Golde, T.E., Kayed, R., Metherate, R., Mattson, M.P., Akbari, Y., and LaFerla, F.M. (2003). Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Abeta and synaptic dysfunction. Neuron *39*, 409–421. https://doi.org/10.1016/S0896-6273(03)00434-3.
71. Bai, B., Wang, X., Li, Y., Chen, P.-C., Yu, K., Dey, K.K., Yarbro, J.M., Han, X., Lutz, B.M., Rao, S., Jiao, Y., Sifford, J.M., Han, J., Wang, M., Tan, H., Shaw, T.I., Cho, J.H., Zhou, S., Wang, H., Niu, M., Mancieri, A., Messler, K.A., Sun, X., Wu, Z., Pagala, V., High, A.A., Bi, W., Zhang, H., Chi, H., Haroutunian, V., Zhang, B., Beach, T.G., Yu, G., and Peng, J. (2020). Deep Multilayer Brain Proteomics Identifies Molecular Networks in Alzheimer's Disease Progression. Neuron *105*, 975–991.e7. https://doi.org/10.1016/j.neuron.2019.12.015.
72. Kim, D.K., Han, D., Park, J., Choi, H., Park, J.-C., Cha, M.-Y., Woo, J., Byun, M.S., Lee, D.Y., Kim, Y., and Mook-Jung, I. (2019). Deep proteome profiling of the hippocampus in the 5XFAD mouse model reveals biological process alterations and a novel biomarker of Alzheimer's disease. Exp. Mol. Med. *51*, 1–17. https://doi.org/10.1038/s12276-019-0326-z.
73. Neuner, S.M., Wilmott, L.A., Hoffmann, B.R., Mozhui, K., and Kaczorowski, C.C. (2017). Hippocampal proteomics defines pathways associated with memory decline and resilience in normal aging and Alzheimer's disease mouse models. Behav. Brain Res. *322*, 288–298. https://doi.org/10.1016/j.bbr.2016.06.002.
74. Boza-Serrano, A., Yang, Y., Paulus, A., and Deierborg, T. (2018). Innate immune alterations are elicited in microglial cells before plaque deposition in the Alzheimer's disease mouse model 5xFAD. Sci. Rep. *8*, 1550. https://doi.org/10.1038/s41598-018-19699-y.
75. Rangaraju, S., Dammer, E.B., Raza, S.A., Rathakrishnan, P., Xiao, H., Gao, T., Duong, D.M., Pennington, M.W., Lah, J.J., Seyfried, N.T., and Levey, A.I. (2018). Identification and therapeutic modulation of a pro-inflammatory subset of disease-associated-microglia in Alzheimer's disease. Mol Neurodegeneration *13*, 24. https://doi.org/10.1186/s13024-018-0254-8.
76. Gail Canter, R., Huang, W.-C., Choi, H., Wang, J., Ashley Watson, L., Yao, C.G., Abdurrob, F., Bousleiman, S.M., Young, J.Z., Bennett, D.A., Delalle, I., Chung, K., and Tsai, L.H. (2019). 3D mapping reveals network-specific amyloid progression and subcortical susceptibility in mice. Commun Biol *2*, 360. https://doi.org/10.1038/s42003-019-0599-8.
77. Oblak, A.L., Lin, P.B., Kotredes, K.P., Pandey, R.S., Garceau, D., Williams, H.M., Uyar, A., O'Rourke, R., O'Rourke, S., Ingraham, C., Bednarczyk, D., Belanger, M., Cope, Z.A., Little, G.J., Williams, S.P.G., Ash, C., Bleckert, A., Ragan, T., Logsdon, B.A., Mangravite, L.M., Sukoff Rizzo, S.J., Territo, P.R., Carter, G.W., Howell, G.R., Sasner, M., and Lamb, B.T. (2021). Comprehensive Evaluation of the 5XFAD Mouse Model for Preclinical Testing Applications: A MODEL-AD Study. Front. Aging Neurosci. *13*, 713726. https://doi.org/10.3389/fnagi.2021.713726.
78. Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M.T., Baker, M., Browne, W.J., Clark, A., Cuthill, I.C., Dirnagl, U., Emerson, M., Garner, P., Holgate, S.T., Howells, D.W., Karp, N.A., Lazic, S.E., Lidster, K., MacCallum, C.J., Macleod, M., Pearl, E.J., Petersen, O.H., Rawle, F., Reynolds, P., Rooney, K., Sena, E.S., Silberberg, S.D., Steckler, T., and Wurbel, H. (2020). The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. PLoS Biol. *18*, e3000410. https://doi.org/10.1371/journal.pbio.3000410.
79. Kulak, N.A., Geyer, P.E., and Mann, M. (2017). Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics. Mol. Cell. Proteomics *16*, 694–705. https://doi.org/10.1074/mcp.O116.065136.
80. Meier, F., Brunner, A.-D., Frank, M., Ha, A., Bludau, I., Voytik, E., Kaspar-Schoenefeld, S., Lubeck, M., Raether, O., Bache, N., Aebersold, R., Collins, B.C., Rost, H.L., and Mann, M. (2020). diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. Nat. Methods *17*, 1229–1236. https://doi.org/10.1038/s41592-020-00998-0.
81. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.
82. Silversmith, W. (2021). cc3d: Connected components on multilabel 3D & 2D images. https://doi.org/10.5281/ZENODO.5535250.
83. Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. Mol. Cell. Proteomics *13*, 2513–2526. https://doi.org/10.1074/mcp.M113.031591.
84. Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinović, S.M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., and Reiter, L. (2015). Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. Mol. Cell. Proteomics *14*, 1400–1410. https://doi.org/10.1074/mcp.M114.044305.
85. Callister, S.J., Barry, R.C., Adkins, J.N., Johnson, E.T., Qian, W.j., Webb-Robertson, B.-J.M., Smith, R.D., and Lipton, M.S. (2006). Normalization Approaches for Removing Systematic Biases Associated with Mass Spectrometry and Label-Free Proteomics. J. Proteome Res. *5*, 277–286. https://doi.org/10.1021/pr050300l.
86. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods *17*, 41–44. https://doi.org/10.1038/s41592-019-0638-x.
87. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13*, 731–740. https://doi.org/10.1038/nmeth.3901.
88. Wiśniewski, J.R., Hein, M.Y., Cox, J., and Mann, M. (2014). A "Proteomic Ruler" for Protein Copy Number and Concentration Estimation without Spike-in Standards. Mol. Cell. Proteomics *13*, 3497–3506. https://doi.org/10.1074/mcp.M113.037309.

Cell
Resource

CellPress

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Antibodies** | | |
| Iba1 | Wako | Cat.# 019–19741; RRID: AB_839504 |
| Stathmin 1 | Novus Biologicals | Cat.# NBP1-76798; RRID:AB_11015725 |
| Neurocan | abcam | Cat.# ab31979; RRID:AB_2149711 |
| S100a11 | R&D | Cat.# MAB5167; RRID:AB_1207907 |
| Thop1 | Novus Biologicals | Cat.# NB400-146; RRID:AB_10128241 |
| MOAB2 | Novus Biologicals | Cat.# NBP2-13075; RRID:AB_2923428 |
| Goat anti-rabbit IgG Alexa Fluor 647 | Invitrogen | Cat.# A21245; RRID:AB_2535813 |
| Goat anti-Mouse IgG Alexa Fluor 488 | Invitrogen | Cat.# A11029; RRID:AB_2534088 |
| Goat anti-Mouse IgG Alexa Fluor 594 | Invitrogen | Cat.# A11032; RRID:AB_2534091 |
| Goat anti-Rat IgG Alexa Fluor 594 | Invitrogen | Cat.# A11007; RRID:AB_10561522 |
| Goat anti-Rat IgG Alexa Fluor 488 | Invitrogen | Cat.# A11006; RRID:AB_141373 |
| Atto647N conjugated anti GFP nanobooster | Chromotek | Cat.# gba647n-100; RRID:AB_2629215 |
| **Chemicals, peptides, and recombinant proteins** | | |
| Congo Red | Sigma-Aldrich | Cat.#C6277 |
| Propidium iodide (PI) | Sigma-Aldrich | Cat.#P4864 |
| Paraformaldehyde (PFA) | Sigma-Aldrich | Cat.#P6148 |
| Tetrahydrofuran (THF) | Sigma-Aldrich | Cat.# 186562 |
| *tert*-butanol | Sigma-Aldrich | Cat.# 360538 |
| Dichloromethane (DCM) | Sigma-Aldrich | Cat.# 270997 |
| Diphenyl ether (DPE) | Alfa Aesar | Cat.# A15791 |
| Vitamin E (DL-alpha-tocopherol) | Alfa Aesar | Cat.# A17039 |
| Benzyl benzoate | Sigma-Aldrich | Cat.#W213802 |
| Benzyl alcohol | Sigma-Aldrich | Cat.# 24122 |
| CHAPS | Carl Roth | Cat.# 1479.4 |
| N-Methyldiethanolamine | Sigma-Aldrich | Cat.# 471828 |
| Trypsin | Sigma-Aldrich | Cat.#T6567 |
| Formic acid | Sigma-Aldrich | Cat.# 64-18-6 |
| Acetonitrile | Sigma-Aldrich | Cat.# 75-05-8 |
| **Experimental models: Organisms/strains** | | |
| $CX_3CR\text{-}1^{GFP}$ mice: B6.129P2(Cg)-$Cx3cr1^{tm1Litt}$/J | Jackson Laboratory | Strain # 005582; RRID: IMSR_JAX:005,582 |
| Thy1-GFP-M mice: Tg(Thy1-EGFP)MJrs/J | Jackson Laboratory | Strain # 007788; RRID: IMSR_JAX:007,788 |
| 5xFAD: B6.Cg-Tg(APPSwFlLon, PSEN1*M146L*L286V)6799Vas/Mmjax | Jackson Laboratory | Strain #034848-JAX; RRID: MMRRC_034848-JAX |
| LysM-eGFP: B6.129(Cg)-$Lyz2^{tm1.1Graf}$/Mmmh | MMRC | Strain #012039-MU; RRID: MMRRC_012039-MU |
| C57BL/6J mouse line | Jackson Laboratory | Strain #:000,664; RRID: IMSR_JAX:000,664 |
| **Software and algorithms** | | |
| ImSpector | Aberrior/LaVision | https://www.lavisionbiotec.com |
| Imaris | Bitplane AG | https://imaris.oxinst.com/ |
| Vision4D | Arivis | https://www.arivis.com/de/imaging-science/arivis-vision4d |
| Fiji | Schindelin et al.[81] | https://ImageJ.net/software/fiji/ |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| syGlass VR | syGlass | https://www.syglass.io |
| Ansys Academic Research Mechanical | Ansys | https://www.ansys.com/academic |
| RoboDK 5.5 | RoboDK Software S.L. | https://robodk.com/ |
| MaxQuant (1.6.7.0) | Tyanova et al.[87] | https://maxquant.org/ |
| Perseus (1.6.7.0) | Tyanova et al.[87] | https://maxquant.org/perseus/ |
| GraphPad Prism(8.2.1) | GraphPad Software | https://www.graphpad.com |
| Code for 2D to 3D image registration | This Paper | Methods S1 |
| **Deposited Data** | | |
| Mass spectrometry raw data, libraries and outputs | This paper | https://www.ebi.ac.uk/pride/; ID# PXD034027 and PXD025316 |
| **Others** | | |
| 18G A-max paed Bone marrow biopsy needle 40,901,803 | HVM Medical Products GmbH | https://hvm-medical.com/produktkatalog-biopsie/ |
| 22G Chiba Fine Needle 07,742,208 | HVM Medical Products GmbH | https://hvm-medical.com/produktkatalog-biopsie/ |

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Ali Ertürk (erturk@helmholtz-muenchen.de).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- All mass spectrometry raw data, libraries and outputs from each particular search engine analyzed in this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier: Database: PXD034027 and Database: PXD025316.
- The source code for registration of 2D to 3D images is provided in supplemental information under Methods S1.
- Any additional information required to reanalyze the data reported in this work paper is available from the Lead Contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Animals

We used the following animals in the study: mixed gender for CX3CR1-eGFP (10 months old), males only for Thy-1-GFPM (10 months old), mixed gender for 5xFAD (6 weeks and 6-months old), mixed gender for LysM-eGFP (3 months old) and mixed gender for C57Bl6/J (3 months old) from Jackson Laboratory. The animals were housed under a 12/12 h light/dark cycle. The animal experiments were conducted according to institutional guidelines: Klinikum der Universität München/Ludwig Maximilian University of Munich and after approval of the Ethical Review Board of the Government of Upper Bavaria (Regierung von Oberbayern, Munich, Germany) and the Animal Experiments Council under the Danish Ministry of Environment and Food (2015-15-0201-00535) and following the European directive 2010/63/EU for animal research. All data are reported according to the ARRIVE.[78] Sample sizes were chosen based on prior experience with similar models.

### Human samples

Intact human brains from a 92 years-old female and intact heart from a 97 years-old female were taken from human body donors with no known neuropathological diseases. The donors gave their informed and written consent to explore their cadavers for research and educational purposes, when still alive and well. The signed consents are kept at the Anatomy Institute, University of Leipzig, Germany. Institutional approval was obtained in accordance to the Saxonian Death and Funeral Act of 1994. The signed body donor consents are available on request.

Cell
Resource

CellPress

## METHOD DETAILS

### Repeated closed head mild traumatic brain injury (mTBI)

Before mTBI, tin foil was taped and tightened to the U-shaped stage made of clear plastic container (38 × 27 × 27 $cm^3$) containing a sponge collection (38 × 25 × 15 $cm^3$). Then, mice were pre-treated with buprenorphine (1:15 saline, 50ul/20mg, ip) and anesthetized with 4% isoflurane using 1.0 L $min^{-1}$ air until non-responsive to a paw or tail pinch. To ensure the head acceleration and rotation following the head impact, the mice were placed under an impact tip on the tin foil, which contains holes according to the shape of a mouse and can support the body weight of the mice. The mice were kept under light anesthesia with continued 2% isoflurane. mTBI was produced using a stereotaxic impactor device with a 5-mm round tip coated with 1mm thick rubber, which can preserve an intact skull after impact. The impact tip was placed and covered the scalp's area from just behind the eyes to the midline of the ears, the center of the tip at approximately midway along the sagittal suture. The injury was produced without skin incision (velocity of 5 m/s, depth of 0.5mm, and dwell time of 0.1 s). The mouse was removed quickly from the collection sponge and transferred to the recovery box maintained at 32 °c. In total, mice received four hits with a 48 h interval in seven days. Sham mice receive identical handling and exposure to the same time length of anesthesia as the mTBI mice but receive no impact.

### Perfusion and tissue preparation

Mice were anesthetized using a combination of midazolam, medetomidine and fentanyl (MMF) (1mL/100g of body mass for mice; i.p.). As soon as the animals did not show any pedal reflex, they were intracardially perfused with cold heparinized 0.1 M PBS (10 U/mL of Heparin, Ratiopharm; 100–125 mmHg pressure using a Leica Perfusion One system) for 5–10 min at room temperature until the blood was washed out, followed by ice-cold 4% paraformaldehyde (PFA) in 0.1 M PBS (pH 7.4) (Sigma) for 10 min. Then, the brains were extracted and post-fixed in 4% PFA for 1 day at 4°C and later washed with 0.1 M PBS for 10 min 3 times at room temperature. The whole brain clearing or nanoboosting procedure was started immediately. For the collection of fresh frozen samples, animals were sacrificed by cervical dislocation and brains were quickly snapped frozen in liquid nitrogen and stored in −80°C until further processing.

### Congo red labeling of whole brains of 5xFAD animals

Whole brains were dehydrated with gradual addition of methanol in PBS (50% x1, 80% x1, 100% x2, each for 1 h). Overnight bleaching with 5% hydrogen peroxide in methanol was done at 4°C. Brains were then gradually rehydrated in 100%, 80%, 50% methanol in PBS (1 h for each step, followed by 2 additional washes in PBS). Detergent washing was then performed in PBS with 0.2% Triton X-100 for 2 h, brains were incubated overnight at 37°C in PBS with 0.2% Triton X-100 and 0.3 M glycine, followed by blocking in PBS with 0.2% Triton X-100 and 6% goat serum for 7 days. Following blocking, the tissue was washed for 1 h twice in PBS with 0.2% Tween 20 and 10 μg/mL heparin (PTwH). Next, brains were incubated with 10 μM Congo Red (Sigma, C6277) at 37°C in PTwH for 5 days. After that, brains were washed in PTwH for 2 days with periodic solution changes and gradually dehydrated using 3DISCO clearing as described next.

### Clearing of brains using DISCO methods

We followed the 3DISCO and uDISCO passive clearing protocol as described previously. In brief, dissected brains were placed in 5 mL tubes (Eppendorf, 0,030,119.401) and covered with 4.5 mL of clearing solution. All incubation steps were performed in a fume hood with gentle shaking or rotation, with the samples covered with aluminum foil to keep them in dark. To clear the samples using 3DISCO, gradient of tetrahydrofuran (THF) in distilled water (v/v %), 2 h for each step, was used as 50%, 70%, 90%, 100% and overnight 100% THF; after dehydration, the samples were incubated for 45 min in dichloromethane (DCM, Sigma, 270,997), and finally in BABB (benzyl alcohol + benzyl benzoate 1:2, Sigma, 24,122 and W213802) until transparency. Next for uDISCO a gradient of tert-butanol (Sigma, 360,538) in distilled water (v/v %) was used as 50, 70, 90, 100 twice at 32°C for 12 h each step, followed by immersion in DCM for 45 min at room temperature and finally incubated with the refractive index matching solution BABB-D15 containing 15 parts BABB, 1 part diphenyl ether (DPE) (Alfa Aesar, A15791) and 0.4% Vol vitamin E (DL-alpha-tocopherol, Alfa Aesar, A17039), for at least 6 h at room temperature until achieving transparency.

### vDISCO whole-brain passive immunostaining, clearing and imaging

Passive vDISCO was performed on dissected organs as performed by Cai R et al. First, the post-fixed brains were pre-treated with permeabilization solution containing 1.5% goat serum, 0.5% Triton X-100, 0.5 mM of Methyl-beta-cyclodextrin, 0.2% *trans*-1-Acetyl-4-hydroxy-L-proline and 0.05% Sodium Azide 0.1 M for 2 days at 37°C with gentle shaking. Subsequently, the brains were incubated in 4.5 mL of this same permeabilization solution plus the nanobooster Atto647N conjugated anti-GFP (1:600, which is ~5–8 μg of nanobooster in 4.5 mL) for CX3CR1-eGFP and Thy-1-GFPM brains for 12–14 days at 37°C with gentle shaking, then brains were washed for 2 h 3 times and once overnight with the washing solution (1.5% goat serum, 0.5% Triton X-100, 0.05% of sodium azide in 0.1 M PBS) at room temperature and in the end washed for 2 h 4 times with 0.1 M PBS at room temperature. The immunostained brains were cleared with 3DISCO clearing first they were put in the Eppendorf 5 mL tubes and then incubated at room temperature with gentle shaking in 4.5 mL of the following gradient of THF in distilled water (v/v %), 2 h for each step: 50, 70, 90, 100 THF and

overnight 100% THF; after dehydration, the samples were incubated for 45 min in DCM, and finally in BABB until transparency. During all the clearing steps, the tubes were wrapped with aluminum foil to keep them in dark.

### vDISCO whole body immuno labeling, clearing and imaging

The detailed protocol of vDISCO was described previously.[34] The mouse bodies were placed inside a 300 mL glass chamber (Omnilab, 5,163,279), to be filled with the appropriate solution regarding the protocol to cover the entire body of the animal (~250–300mL). A transcardial circulator system was established in order to allow peristaltic pumping of the solutions (ISMATEC, REGLO Digital MS-4/8 ISM 834; reference tubing, SC0266), with the pressure being set at 180–230 mmHg (50–60 rpm). The tubing was set to allow pumping of the solutions through the heart (attached to a perfusion needle (Leica, 39,471,024)) into the vasculature with the same entry point used for PBS and PFA perfusion steps described above. The other end of the tube was immersed into the chamber with a loose end to allow suction of the solution into the body. The samples were initially perfused with a decolorization solution (25% of CUBIC reagent 1 (55) which is composed of 25 wt % urea (Carl Roth, 3941.3), 25 wt % N,N,N′,N′-tetrakis (2-hydroxypropyl) ethylenediamine (Sigma, 122,262) and 15 wt % Triton X-100 (AppliChem, A4975,1000) in 0.1 M PBS)) for 2 days, refreshing the solutions every 12h. Samples were washed with PBS for 3 × 2h. Then, decalcification solution (10 w/v % EDTA in 0.01 PBS, pH~8–9, Carl Roth, 1,702,922,685) was perfused for 2 days followed by half a day with permeabilization solution composed of 0.5% Triton X-100, 1.5% goat serum (GIBCO, 16,210,072), 0.5 mM of Methyl-beta-cyclodextrin (Sigma, 332,615), 0.2% *trans*-1-Acetyl-4-hydroxy-L-proline (Sigma, 441,562), 0.05% sodium azide (Sigma, 71,290) in 0.01 M PBS. The setup was adjusted to initiate the PI labeling and boosting. The free end of the perfusion tube was connected to a 0.22 μm syringe filter (Sartorius, 16,532) and an infrared lamp (Beuer, IL21) was aimed at the chamber to enable the solution's temperature to be around 26–28°C. This setup was then left running for 6 days after the addition of 35 μL of nanobooster (stock concentration 0.5–1 mg/mL) and 290 μL of propidium iodide (stock concentration 1 mg/mL) which was added directly into the refreshed permeabilization solution. Next, the body was placed into a 50 mL tube (Falcon, 352,070), with the same permeabilization and labeling solution, and an extra 5 μL of nanobooster was added. The tube was then put on a shaker at RT for 2 additional days for labeling. Atto647N conjugated anti GFP nanobooster (Chromotek, gba647n-100) and propidium iodide (PI, Sigma, P4864) was used to boost the signal from the LysM-eGFP animals and stain cell nuclei respectively in the study. Then, the animals were placed back into the initial perfusion setup, where washing solution was perfused for 2 × 12h, which was composed of; 1.5% goat serum, 0.5% Triton X-100, 0.05% of sodium azide in 0.1 M PBS. 0.1 M PBS was used to wash the sample 3 × 2h. 3DISCO protocol was applied for whole body clearing. The animals were demounted from the perfusion system but kept in glass chambers and placed on top of shakers (IKA, 2D digital) at room temperature inside a fume hood. Glass chambers were sealed with parafilm and covered with aluminum foil along with the 3DISCO application. For dehydration, sequential immersion of tetrahydrofuran (THF) (Sigma,186,562) (50 vol % THF, 70 vol % THF, 80 vol % THF, 100 vol % THF and again 100 vol % THF) was applied every 12 h. Then 3 h of dichloromethane (DCM) (Sigma, 270,997) immersion for delipidation was followed by indefinite immersion in BABB (benzyl alcohol + benzyl benzoate 1:2, Sigma, 24,122 and W213802) solution for refractive index matching.

### SHANEL sample preparation and clearing

Archived human samples were obtained in PFA which were stored for a long period of time (>5 years) at 4°C and subjected to our previously published SHANEL clearing protocol with some modifications.[4] Briefly, samples were dehydrated with EtOH/$dH_2O$ series at RT: 50%, 70%, 100% for 1 h for each step, then incubated with 10 mL DCM/MetOH (2:1 v/v) (freshly prepared) for 6h at RT followed by rehydration with EtOH/$dH_2O$ series at RT: 100%, 70%, 50%, $dH_2O$ for 1 h each step then incubated with 0.5M acetic acid (30 mL/L) at RT for 2 h, then wash with $dH_2O$ twice for 15 min and then incubated with 4M guanidine hydrochloride (382.12 g/L), 0.05M sodium acetate (4.1 g/L), 2% v/v Triton X-100 in $dH_2O$, (measure pH: 6.0) at RT for 2 h, then wash with $dH_2O$ twice for 15 min each and wash with PBS twice for 15 min each. Afterward samples were incubated with 10% CHAPS, 25% N-Methyl diethanolamine in $dH_2O$ at 37°C for 4 h and then washed with $dH_2O$ twice for 15 min each. Since we did not perform any deep antibody labeling in these samples, we started clearing these samples without prior blocking or antibody labeling steps. Clearing was done with THF in water with dilutions (v/v %) of 50%, 70%, 90%, 100%, 1h each, 100% overnight, DCM 45 min and incubated in BABB until the samples were transparent.

### Behavioral assessment

#### *Barnes maze*

Briefly, a maze consisting of a surface bright circular platform with an escape black box can be recessed and located at the bottom of one of the 20 holes. Visual shapes were placed on 3 walls of the room as cues. For all trials, mice were placed in a cylinder black start chamber in the center of the maze for 10 s. After the chamber lifted and the test started, mice were given 3 min to locate and enter the target box during the spatial acquisition time. For a period of 4 days, 4 trials were given per day with an inter-trial of 15 min. The trial ended when the mouse entered the escape box or after 3 min had elapsed. Mice were allowed to remain in the escape box for 1 min. A system (Ethovision XT) was used to continually track and record the movement of the mice. Escape latency was measured as the time taken for the mouse to enter the box.

Cell
Resource

CellPress

### *Assessment of motor function*

The mice were given 3 trials training per day for 3 days to walk along a 1 cm diameter and 100 cm long wood beam with a goal box on the end of the beam before mild TBI. The beam was placed 1 m above ground. The latency that it takes walking to cross the beam after 8 weeks from mild TBI were recorded. Mice that were unable to cross the beam were removed in the training.

### Immunofluorescence and confocal microscopy

Briefly, mice were sacrificed after 8 weeks of injury or at six weeks of age following transcardial perfusion with PBS and with 4% cold PFA. Brains were post-fixed in 4% PFA at 4°C overnight. Either frozen sections or cleared-rehydrated frozen sections were treated with 0.2% Triton X-100 in PBS for 15 min, blocked for 1 h at room temperature with 10% serum in PBST. Then incubation with primary antibodies Iba1 (1:1000, Wako, 019–19741), Stathmin 1 (1:300, Novus, NBP1-76798), Neurocan (1:300, abcam, ab31979), S100a11 (1:300, R&D, MAB5167), Thop1 (1:300, Novus, NB400-146), MOAB2 (1:1000, Novus, NBP2-13075), at 4°C for overnight and Alexa conjugated secondary antibodies (1:1000, Goat anti-rabbit IgG Alexa Fluor 647, Invitrogen, A21245; Goat anti-Mouse IgG Alexa Fluor 488, A11029; Goat anti-Mouse IgG Alexa Fluor 594, A11032; Goat anti-Rat IgG Alexa Fluor 594, A11007; Goat anti-Rat IgG Alexa Fluor 488, A11006) were incubated for 1 h at room temperature. Slices were mounted after being stained with Hoechst 33,342 (Invitrogen). Images were acquired with 10x, 40x, and 63× objective of confocal microscope (ZEISS LSM880).

### Light-sheet microscopy and image processing

Single plane illuminated (light-sheet) image stacks were acquired using an Ultramicroscope II (LaVision BioTec) and UltraMicroscope Blaze (Miltenyi Biotec), featuring an axial resolution of 4 μm with following filter sets: ex 470/40 nm, em 535/50 nm; ex 545/25 nm, em 605/70 nm; ex 640/40 nm, em 690/50 nm. Whole brains were imaged individually using high magnification objectives: 4× objective (Olympus XLFLUOR 4× corrected/0.28 NA [WD = 10 mm]), LaVision BioTec MI PLAN 12× objective (0.53 NA [WD 10 = mm]) coupled to an Olympus revolving zoom body unit (U-TVCAC) kept at 1x. High magnification tile scans were acquired using 20-35% overlap and the light-sheet width was reduced to obtain maximum illumination in the field. Processing, data analysis, 3D rendering and video generation for the rest of the data were done on an HP workstation Z840, with 8 core Xeon processor, 196 GB RAM, and Nvidia Quadro k5000 graphics card and HP workstation Z840 dual Xeon 256 GB DDR4 RAM, nVidia Quadro M5000 8GB graphic card. We used Imaris (Bitplane), Fiji (ImageJ2), Vision 4D (Arivis) and syGlass (for 3D and 2D image visualization). Tile scans were stitched by Fiji's stitching plugin49.

### Registration of light-sheet and LCM images to correlate probe selection

In order to map the proteomic samples to the whole brain light-sheet imaging we developed a registration protocol. The LCM method is acquired in order to select the ROI to be cut from the sample of brain tissue. The registration of light-sheet and LCM images is very challenging because it is a multimodal registration problem and the imaged tissue undergoes physical changes when the re-clearing process is applied after the light-sheet imaging, introducing scale differences and sometimes even damaged tissue. Moreover, the tissue contrast and lighting distribution are different.

For the purpose of registering the images, we acquired consecutive 2D LCM slices (0.65 x 0.65 $\mu m^2$, slice thickness 12 um) which correspond to the 3D light-sheet volume (1.63 × 1.63 × 4 $\mu m^3$). We down sampled all images to the largest dimension to achieve an isotropic resolution of 12um. We experimented with a multitude of registration steps, such as: 3D-3D volume registration, 2D-2D slice registration, various transformations (rigid, affine, BSpline), edge extraction, binary thresholding, normalization strategies, and similarity metrics. We found the best performance for the following protocol using elastix:

1) Pre-processing: Both LCM and light-sheet images were clipped to the relevant contrast ranges and normalized to the [0, 255] range. Additionally, the LCM slices undergo histogram equalization to mitigate the severe lighting differences, 2) first, a single LCM slice is rigidly registered to each 3D light-sheet slice 3) next, we calculate different image similarity scores between the registered LCM slice and its corresponding LS slice. We choose the most similar slices as an initial match in z-Dimension 4) then, we create a 3D LCM volume by padding the single LCM slice and then do a full 3D affine registration to the 3D light-sheet volume to account for scaling differences. These four steps lead to our registration result, where the two red boxes indicate a potential cut region of interest in the LCM image and its corresponding spatial location in the light-sheet image. We find that the registration of the individual LCM slices is very successful, whereby the equalization plays a crucial role. The rigid registration is also able to find a good initialization for the affine registration in the next step. The third step poses harder challenges, as the differences between the LCM and light-sheet images outlined above persist. For example, we observe that the brightness of the light-sheet image increases radially outwards, whereas in the LCM the light is more region-dependent. Furthermore, in the LCM images various regions are so hypointense, that even the extraction of low-level features such as edges is not very informative for the registration. These challenges particularly impact the registration of the LCM volume in the z direction. We are convinced that this registration problem is a highly non-trivial task and believe that future research will be of high relevance for the community beyond our application.

### Optimization of cleared tissue for cryopreservation and sectioning

After acquiring the whole brain images from CX3CR1-eGFP, 5xFAD, C57BL/6J mice the brains were further optimized for cryopreservation and sectioning. The course of tissue clearing and imaging in BABB makes the tissue brittle and hard to process further. To solve this, we rehydrated the samples with the respective clearing solutions to be able to process samples for cryosectioning.

Thereafter, samples were washed with PBS twice for 15 min each and cryopreserved overnight with 30% sucrose solution in 4°C. To avoid any ice crystals formation, samples were further embedded in Optimal cutting temperature compound (OCT compound) under the chilled isopentane container placed on dry ice. Samples were stored in −80°C until cryosectioning.

### Laser-capture microdissection

For the microdissection of cells and plaques, we used both the PALM Micro-Beam system (Zeiss) and Leica LMD 7000 (Leica). PALM Micro-Beam uses a focused laser beam to cut out and isolate the selected specimen without contact. The laser catapult quickly isolates the region of interest and uncontaminated in the adhesive cap mounted in the RoboMover upwards. The Leica system projects the samples downwards. Briefly, in case of the PALM system, after cryosectioning, sections were mounted on the polyethylene naphthalate (PEN, Zeiss) slides and were either stored at −80°C in 50 mL falcon tubes filled with molecular sieves (Sigma-Aldrich) or processed further for serial dehydration with ethanol and air dried for 15 min under the hood. Cells in the optic tract from mTBI/Sham brains and amyloid-beta plaques from the 5xFAD and from respective WT brain regions were micro-dissected by laser pressure catapulting (LPC) UV LCM system (Palm Zeiss Microlaser Technologies, Munich, Germany) consisting of an inverted microscope with a motorized stage, an UV laser and an X-Cite 120 fluorescence illuminator (EXFO). The microdissection process was visualized with an AxioCam ICc camera coupled to a computer and was controlled by Palm RoboSoftware (Zeiss, Germany). An area of approximately 200 × 200 μm (corresponding to 40–60 cells) were cut by laser using a 20× objective (LD Plan-Neofluar 20×/0.4 corr M27) and catapulted against gravity into the adhesive cap. Tissues were quickly lysed in 20 μL of lysis buffer, spun down and kept in dry ice or stored in −80°C. To avoid any uncertainties in capturing ROI, each time after catapulting as well as after lysing and spun down, the cap was examined under the camera.

### Automated robotic proteomic probe extraction from whole body

The strategy applied in this study was image guided extraction of targeted samples in the bone marrow of scapula and cranium on whole mouse body with a precise robotic biopsy solution. The imaging takes place under the Ultramicroscope Blaze manufactured by Miltenyi Biotec. The robotic arm Meca500 R3 manufactured by Mecademic Robotics is used to control the motion of the needle sets to extract multiple targets. During the imaging, the embedded whole mouse body was placed on a bed designed for this study that holds the sample stable under the load applied by the needle. The Meca500 was mounted above the microscope to utilize the slim gap between the submerged lens and the walls of the sample bath for the operation on the mouse during live imaging. The extraction was conducted by using two sizes of biopsy needles with stylets sealing the shaft of the needle to prevent unwanted tissue contamination while approaching the target. The larger diameter needle was fixed on the robotic arm with a holder designed and manufactured to ensure the access of the finer needle to the target with precision.

### Precision and accuracy of sampling

The most important problem to overcome in the deep tissue sampling in a whole mouse body is to achieve the required precision and accuracy of the sampling needle tip. Since the operation happens in toughened, inhomogeneous multilayer tissue, the forces acting on the needle and the tissue during the penetration are highly fluctuating and axially asymmetric. The borders of two different types of tissues e.g., bone and muscle, tend to deflect the needle in the direction of the muscle. These forces result in bending of the needle which leads to inaccuracies when approaching the target. Since it is problematic to move the needle sideways or rotate it inside the tissue due to high traction, it is complicated to compensate for the inaccuracy after the needle has penetrated into the tissue.

Providing precision and accuracy to the operation requires every component on the load path to remain as rigid as possible. The translational stage of the microscope is fixed on the main structure and holds the mouse bed, the mouse bed supports the agarose embedding around the mouse and the agarose embedding supports the whole mouse body including the extremities and inner organs. On the other end of the setup the robotic arm was fixed on the main structure. The needle holder is screwed on the last actuator of the robot arm. The needle is fixed on the needle holder with a screw. The load transfer mechanism can be reduced to some basic components to model the bending behavior of the structure and calculate the deflections. The total deflection of all these components under load is optimized to provide an accurate positioning of the needle deep in the tissue.

### Needle gauge

The needle gauge plays a key role in the total deflection. The needle can be considered as a cylindrical cantilever beam and the maximum deflection on the needle caused by a concentrated load at the tip is defined by the formula:

$$\delta_{max} = \frac{Pl^3}{3EI}$$

In this equation the maximum deflection $\delta_{max}$ is proportional to the force applied on the tip $P$, proportional to the cube of the length of the needle $l$ and inversely proportional to the Young's modulus of the needle's material $E$ and the static moment of inertia $I$.

The needle's deflection in the deep tissue is not only caused by the concentrated force applied on the piercing tip, but also by the distributed load on the needle, for which case the deflection at the tip is defined by the formula: $\delta_{max} = \frac{\omega l^4}{8EI}$.

In this case the maximum deflection $\delta_{max}$ is proportional to the distributed load applied on the needle $\omega$, proportional to the 4th power of the length of the needle $l$ and inversely proportional to the Young's modulus of the needle's material $E$ and the static moment of inertia $I$. The static moment of inertia $I$ can be calculated as:

$$I_y = I_z = \frac{\pi(R^4 - r^4)}{4}$$

For both of these equations the static moment of inertia $I$ is proportional to the difference of the 4th power of the outer $R$ and inner $r$ radii of the needles. These calculations suggest a 22G needle would deflect around 9 times further away from target in comparison to an 18G needle.

Although a larger needle is more accurate meeting the target due to lower deflection, there are also disadvantages such as collecting a larger sample including unwanted neighboring tissue. Furthermore, since the penetration area is larger with a larger needle, the force required to achieve the penetration stress on the tissue is also larger. Occasionally, it becomes impossible to apply the necessary force for penetration, since the tissue cannot support such loads and the harder bone tissue gets pushed into the softer muscle tissue. We found that the optimum solution to this problem was using two needles in combination. The procedure applied in this study was to extract bone marrow with 18G blunt tip biopsy needle A-MAX PAED and 22G Chiba biopsy needle. Both of these needles include a stylet to block the hollow shaft of the needle until the biopsy target is reached. Initially the 18G needle was fixed on the needle holder on the Meca500 and the coordinates of the tip of the needle are calibrated under the microscope. The needle is advanced with the Meca500 axially to the coordinates where the stylet is about to touch the target. Then the 18G stylet is removed to create an open channel for the 22G needle to move until the target without any resistance or deflection. As soon as the 22G needle reaches the target the 22G stylet is removed from the needle and the target is sucked in to the needle's shaft with the slight vacuum supplied by a syringe and the chopping motion of the beveled needle tip around the target. When the targets acquisition is confirmed observing the light sheet image, the needle is retracted without stopping or increasing the vacuum applied by the syringe and the sample is deployed into a low binding 0.5 mL Eppendorf tube with some extra ECi (ethyl-cinnamate) to make sure any remaining pieces of the sample in the needle is flushed before the next target.

### Specimen holder and needle holder strategy

During the imaging process the sample and the lens are submerged into an imaging solution with a refractive index of 1.56. The common options for this medium are mostly ECi or benzyl benzoate and benzyl alcohol (BABB) solution with a 2:1 ratio. The mouse holder and needle holder are manufactured using the clear resin on a Form3 manufactured by Formlabs. This process provides agile prototype iterations but the manufactured parts lose their integrity and start falling apart upon contact with BABB. Therefore, these experiments are conducted under ECi. The mouse bed and the needle holder are modeled using finite elements method (FEM) to simulate their behavior under loading using Ansys Academic Research Mechanical 2022 R2. Using the stresses and deflections calculated from the static FEM analyses, an iterative method was employed to optimize the structures of the mouse bed and the needle holder. The mass of the mouse bed was reduced to around 80 g and the needle tip to around 20 g, while the simulated total deflection of both pieces combined was reduced to less than a millimeter from an initial total deflection around 7mm under extreme load conditions such as 40 N axial load applied on the needle and the same amplitude applied on the mouse bed in the direction of needle's approach.

### Embedding of the whole mouse body

Although the whole mouse body is cleared and fixed, it still deflects far enough for the needle to miss the target. Especially when the target is in a bone, the forces required for penetration go beyond 10N with the 18G needle. Embedding the body with a gel or resin is crucial to limit the deformation of the mouse body. The uneven structure of the internal organs, gaps in the chest cavity and curved outer surfaces create incalculable sources for tissue and needle deflection. To stabilize the cleared mouse bodies during robotic tissue extraction, they had to be embedded in a block of agarose. To do this, we prepared a 2% (w/v) solution of agarose by dissolving the agarose powder in deionized water. This mixture was heated in a microwave until the agarose had completely dissolved. Then, the mixture was allowed to cool until it had reached a temperature of about 50–60°C. This cooled agarose solution was then poured onto the sample, and it was left to cool until it had completely solidified. After it had solidified, excess parts of agarose around the samples were carefully trimmed away using a sharp scalpel.

To clear the agarose and prepare it for light sheet imaging, the sample was dehydrated in a serial dilution of tetrahydrofuran (ROTISOLV, Carl Roth). The dilutions used were as follows: (v/v) 20%, 30%, 50%, 70%, 90%, 100% -2x. After the sample had completely dehydrated, it was immersed in a solution of BABB which is a mixture of benzyl alcohol and benzyl benzoate to complete the clearing process. The sample was immersed in these solutions for a minimum of 24 h at every stage.

### Manipulation system

A common solution for micrometer accurate targeting is a micromanipulator but in this study the target was located deep in the hardened and tough tissue. Considering the force required in case of a bone penetration is in the range of 10 N and micromanipulators operate in the range of hundreds of mN, the standard solution would not suffice for this specific problem.

For the image guided sample extraction, a robot arm with harmonic drives was selected. The advantages in comparison to a micromanipulator or a piezo electric stage system is the level of applicable force without compromising accuracy. The Meca500 can apply a force of 110N and still remain within a precision tolerance of 5 μm. Being able to control such a force with high precision in a large working envelope with 260mm reach in 6 degrees of freedom, provided us the freedom to conduct experiments with various assembly positions, tool deigns, needle dimensions delivering repeatable results.

The system is controlled via the programming and simulation software for industrial robots RoboDK. The 3D models of the optical table, microscope and other structural parts are assembled corresponding to their real locations in a 3D CAD software and exported as.STEP file to the RoboDK environment. From that point on Meca500 is controlled over TCP/IP socket communication with ASCII commands. Having the whole environment in the control software, paths are calculated in segments to avoid any collision. Any misalignment with the needle tip in real world and computer model is also calibrated in the software before sample extraction.

### Optimization of DISCO cleared sample preparation for mass spectrometry analysis

Several conditions and combinations of solubilizing agents for the isolation of proteins from tissue cleared mouse brain, heart, and lung samples were initially evaluated for protein extraction efficiency, peptide recovery, and qualitative and quantitative reproducibility keeping fresh or PFA-fixed as reference. Our goal was to establish a workflow that recovers proteomes that are as similar as possible to non-cleared tissue and is universal for all tissue clearing techniques.

Cleared organs or cryosections were removed from the refractive index matching solution BABB and washed five times with 1x PBS solution. The organ was then flash-frozen and pulverized in a Covaris CP02. Afterward, the samples were resuspended in different protein solubilizing solutions (6% SDS500 mM TrisHCl, pH 8.5 (SDS buffer); 2% Sodium deoxycholate, 100 mM TrisHCl pH 8.5, 10 mM Tris-(2-carboxyethyl)-phosphine (TCEP), 40 mM Chloroacetamide (SDC buffer); 50% Trifluoroethanol, 100 mM TrisHCl, pH 8.5 (TFE buffer), followed by protein extraction at 95°C, 1.000rpm for 45 min. Then the samples were subjected to sonication (Branson) at maximum frequency for 30 cycles at 50% output, followed by another heating step at 95°C, 1.000 rpm for 45 min. From here on, processing steps diverged for each protocol.

Proteins solubilized in the SDS buffer were precipitated with ice-cold acetone at 80% v/v ratio overnight at −80°C, followed by centrifugation at max. g for 15 min at 4°C. The supernatant was removed, the pellet was washed with 5 mL ice-cold 80% v/v Acetone/ddH2O, followed by 30 min precipitation on dry ice. The acetone wash steps were repeated two times for a total of three washes. Proteins solubilized in the TFE buffer, were subjected to solvent evaporation in a SpeedVac at 45°C until dryness before further processing.

In case of SDS-SDC or TFE-SDC protocol, in which SDS or TFE protein extraction was coupled to an SDC-based protein digestion, SDS- or TFE-solubilized proteins were resuspended in 1mL of SDC buffer and heated to 95 °C at 1.000 rpm for 10 min to denature proteins, reduce cysteine bridges and alkylate free cysteine residues. Afterward, samples were sonicated for 15 cycles each 30 s at max power in a Bioruptor, followed by another heating step for 10 min at 95°C, 1.000 rpm in a Thermoshaker.

SDC-only, SDS-SDC, TFE-SDC solubilized protein solutions were cooled down to room temperature, diluted 1:1 with 100 mM TrisHCl, pH 8.5, followed by protein concentration estimation by Nanodrop. Extracted and solubilized proteins were digested overnight at 37°C and 1.000 rpm, with trypsin and LysC at a protein to enzyme w/w ratio of 1:50. Next day, trypsin and LysC were added again at a protein to enzyme w/w ratio of 1:50 and proteins were digested further for 4 h at 37°C, 1.000 rpm. Resulting peptides were acidified with 1% TFA 99% Isopropanol in a 1:1 ratio and vortexed, followed by centrifugation at 22.000 xg RT to pellet residual particles. The supernatant was transferred into a fresh tube and subjected to StageTip clean-up via SDB-RPS. 20 μg of peptides were loaded on two 14-gauge stage-tip plugs. Peptides were washed twice with 200 μL 1% TFA 99% ddH2O followed by 200 μL 1% TFA 99% isopropanol in an in-house-made StageTip centrifuge at 2,000 xg. Peptides were eluted with 100 μL of 5% Ammonia, 80% ACN into PCR tubes and dried at 45°C in a SpeedVac centrifuge (Eppendorf, Concentrator plus). Peptides were resuspended in 0.1% TFA, 2% ACN, 97.9% ddH2O.

After evaluation of protein extraction efficiency, all sample preparation for Fresh, PFA-fixed, uDISCO-, 3DISCO-, SHANEL-cleared tissue was performed following the SDS-SDC protocol. For LCM sample preparation, LCM samples were caught on PCR tubes with adhesive caps and successful isolation was verified by visual inspection. 20μL of SDS-buffer was added to each tube. The tube was closed and vortex for 30 s, followed by centrifugation for 5 min in a table-top centrifuge to 'catch' the LCM sample in the protein solubilization buffer, which was confirmed afterward by visual inspection. Sample preparation was performed as described for the SDS-SDC protocol, except for the following modifications: No shaking during cooking steps; Instead of a Branson sonicator, a Bioruptor was used for each sonication step; No Covaris CP02 was used for crushing the sample; Acetone precipitation was performed at 100 μL total volume; SDC resuspension and protein digestion was performed in a 20 μL volume.

### High-pH reversed-phase fractionation

To generate a deep library of experiment-specific precursors, peptides were fractionated at pH 10 with the spider-fractionator.[79] 50 μg of purified peptides were separated on a 30 cm $C_{18}$ column in 96 min and concatenated into 16 or 24 fractions with 2 min exit valve switches. Peptide fractions were dried in a SpeedVac and reconstituted in 2% ACN, 0.1% TFA, 97.9% $ddH_2O$ for LC-MS analysis.

Cell
Resource

CellPress

### Liquid chromatography and mass spectrometry (LC-MS)

LC-MS was performed on an EASY nanoLC 1200 (Thermo Fisher Scientific) coupled online either to a quadrupole Orbitrap mass spectrometer (Q Exactive HFX, Thermo Fisher Scientific), or a trapped ion mobility spectrometry quadrupole time-of-flight mass spectrometer (timsTOF Pro, Bruker Daltonik GmbH, Germany) via a nano-electrospray ion source (Captive spray, Bruker Daltonik GmbH). Peptides were loaded on a 50 cm in-house packed HPLC-column (75 μm inner diameter packed with 1.9 μm ReproSil-Pur C18-AQ silica beads, Dr. Maisch GmbH, Germany). Sample analytes were either separated using a linear 100min gradient from 5 to 30% B in 80 min followed by an increase to 60% for 4 min, and by a 4 min wash at 95%, a decrease to 5% B for 4 min, and a re-equilibration step at 5% B for 4 min, or separated on a linear 120 min gradient from 5 to 30% B in 90 min followed by an increase to 60% for 10 min, and by a 5 min wash at 95%, a decrease to 5% B for 5 min, and a re-equilibration step at 5% B for 5 min (Buffer A: 0.1% Formic Acid, 99.9% ddH2O; Buffer B: 0.1% Formic Acid, 80% CAN, 19.9% ddH2O). Peptides derived from LCM and matching libraries were separated using a linear 70 min gradient from 3 to 30% B in 45 min followed by an increase to 60% for 5 min, an increase to 95% in 5min, followed by 5 min at 95% B, a decrease to 5% B for 5 min, and an equilibration step at 5% B for 5 min. Flow-rates were constant at 300 nL/min. The column temperature was kept at 60°C by an in-house manufactured oven.

Mass spectrometry analysis for the evaluation of sample preparation on a Q Exactive HFX was performed in data dependent scan mode. For full proteome measurements, MS1 spectra were acquired at 60.000 resolution and an m/z range of 300–1.650 with an automatic gain control (AGC) target of 3E6 ions and a maximum injection time of 20 ms. The top 15 most intense ions with a charge of two to eight from each MS1 scan were isolated with a width of 1.4 Th, followed by higher-energy collisional dissociation (HCD) with a normalized collision energy of 27% and a scan range of 200–2,000 m/z. MS/MS spectra were acquired at 15,000 resolution with an AGC target of 1E5, a minimum AGC target of 2.9E3, and a maximum injection time of 28 ms. Dynamic exclusion of precursors was set to 30 s.

Deep proteomes and comparisons of clearing conditions with the SDS-SDC protocol were acquired on a standard timsTOF Pro in a data-dependent PASEF mode with 1 MS1 survey TIMS-MS and 10 PASEF MS/MS scans per acquisition cycle. Ion accumulation and ramp time in the dual TIMS analyzer was set to 100 ms each and we analyzed the ion mobility range from $1/K_0 = 1.6$ Vs $cm^{-2}$ to 0.6 Vs $cm^{-2}$. Precursor ions for MS/MS analysis were isolated with a 2 Th window for $m/z < 700$ and 3 Th for $m/z > 700$ in a total m/z range of 100–1.700 by synchronizing quadrupole switching events with the precursor elution profile from the TIMS device. The collision energy was lowered linearly as a function of increasing mobility starting from 59 eV at $1/K_0 = 1.6$ VS $cm^{-2}$ to 20 eV at $1/K_0 = 0.6$ Vs $cm^{-2}$. Singly charged precursor ions were excluded with a polygon filter (otof control, Bruker Daltonik GmbH). Precursors for MS/MS were picked at an intensity threshold of 2.500 a.u. and re-sequenced until reaching a 'target value' of 20.000 a.u taking into account a dynamic exclusion of 40 s elution.

Peptides derived from LCM samples were acquired on a timsTOF Pro modified for highest ion transmission and sensitivity, as described in Brunner et al., in a data-dependent or data independent acquisition PASEF mode.[10] In DDA, 1 MS1 survey TIMS-MS and 5 PASEF MS/MS scans represents one acquisition cycle. Ion accumulation and ramp time in the dual TIMS analyzer was set to 50 ms each and we analyzed the ion mobility range from $1/K_0 = 1.6$ Vs $cm^{-2}$ to 0.6 Vs $cm^{-2}$. Precursor ions for MS/MS analysis were isolated with a 2 Th window for $m/z < 700$ and 3 Th for $m/z > 700$ in a total m/z range of 100–1.700 by synchronizing quadrupole switching events with the precursor elution profile from the TIMS device. The collision energy was lowered linearly as a function of increasing mobility starting from 59 eV at $1/K_0 = 1.6$ VS $cm^{-2}$ to 20 eV at $1/K_0 = 0.6$ Vs $cm^{-2}$. Singly charged precursor ions were excluded with a polygon filter (otof control, Bruker Daltonik GmbH). Precursors for MS/MS were picked at an intensity threshold of 1.500 a.u. and re-sequenced until reaching a 'target value' of 20.000 a.u taking into account a dynamic exclusion of 40 s elution. For DIA analysis, we made use of the correlation of ion mobility (IM) with m/z and synchronized the elution of precursors from each IM scan with the quadrupole isolation window. We used the short-gradient diaPASEF method as described in Meier et al.[80] but performed five consecutive diaPASEF cycles before the next MS1 scan. The collision energy was ramped linearly as a function of the IM from 59 eV at 1/K0 = 1.6 Vs $cm^{-2}$ to 20 eV at 1/K0 = 0.6 Vs $cm^{-2}$.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### ClearMap quantification

To quantify microglia distribution in whole brains of mTBI and sham animals, we used ClearMap. As the script was originally developed for quantification of the $cFos^+$ cells, to comply with the offered method, we performed the following pre-processing steps on our microglia data using Fiji before ClearMap:

1. Background equalization to homogenize intensity distribution and appearance of the microglia cells over different regions of the brain, using pseudo-flat-field correction function from Bio-Voxxel toolbox.
2. Convoluted background removal, to remove all particles bigger than relevant cells. This was done with the median option in the Bio-Voxxel toolbox.
3. Two-dimensional median filter to remove remaining noise after background removal. The filter radius was chosen to ensure the removal of all particles smaller than microglia cells.

4. Unshapen mask to amplify the high-frequency components of a signal and increase overall accuracy of the cell detection algorithm of ClearMap.

After pre-processing, ClearMap was applied by following the original publication and considering the threshold levels that we obtained from the pre-processing steps. As soon as the quantification was completed, the data was exported as an Excel file for further analysis.

### Deep learning analyses

The segmentation of the stained Aβ plaques represents a key step toward a reliable quantification thereof. We develop a customized, three-dimensional deep learning approach to optimize segmentation of Aβ plaques in the whole brains of 5xFAD animals. Our network architecture is inspired by the well-established U-Net architecture. Our loss function is an equally weighted combination of Dice and binary cross entropy loss. We use the Ranger optimizer, which combines Rectified Adam, gradient centralization and LookAhead. Annotation of our dataset was performed in Fiji and double checked by multiple experts.[81] Our dataset consists of 98 image volumes (300 × 300 × 300 voxel) from one Alzheimer brain; where 34 vol include Aβ plaques, and 64 vol do not contain any plaques. An ensemble of experts including the scientist who imaged the brains labeled all images. We randomly sampled our training set of 85 vol (21 with AD plaques, 64 without AD plaques), our validation set of seven volumes and our separate test set of six volumes. During training and testing we applied suitable data augmentation protocols. In order to assess the quality of our segmentation we calculate a wide range of voxel-wise and Aβ plaque wise segmentation metrics. Based on the reliable segmentation of individual Aβ plaques we continued toward a statistical evaluation of the number and size of Aβ plaques per brain region. First, we registered all of our brains to the Allen brain Atlas, enabling a single voxel assignment to brain structures. For whole brain segmentation we extracted the single connected components using cc3d,[82] which represent our individual segmented Aβ plaques and calculate their total size in voxels as a biomarker. Using this registration and biomarker, we calculated per brain region statistics for the presence and size of Aβ plaques across the whole brain.

### Proteomics data processing

Raw files were either searched against the mouse Uniprot databases (UP00000589_10090.fa, UP00000589_10090_additional.fa) or human Uniprot databases (UP000005640_9606.fa, UP000005640_9606_additional.fa). For DDA raw file analysis, we used the MaxQuant version 1.6.7.0 which extracts features from four-dimensional isotope patterns and associated MS/MS spectra. False-discovery rates were controlled at 1% both on peptide spectral match (PSM) and protein level. Peptides with a minimum length of seven amino acids were considered for the search including N-terminal acetylation and methionine oxidation as variable modifications and cysteine carbamidomethylation as fixed modification, while limiting the maximum peptide mass to 4.600 Da. Enzyme specificity was set to trypsin cleaving c-terminal to arginine and lysine. A maximum of two missed cleavages were allowed. Maximum precursor tolerance in the first search and fragment ion mass tolerance were searched as default for TIMS-DDA data. Main search tolerance was set to 20 ppm. The median absolute mass deviation for the dataset was 1.57 ppm for precursors. Peptide identifications by MS/MS were transferred by matching four-dimensional isotope patterns between the runs with a 0.7 min retention-time match window and a 0.05 $1/K_0$ ion mobility window. Label-free quantification was performed with the MaxLFQ algorithm, and a minimum ratio count of 1.[83] For DIA data analysis of the 5xFAD samples isolated from different brain regions, a hybrid library containing all single-shot DIA files and region-specific libraries (16 fractions each) were created using the Spectronaut software suite (version 14.10.201222.47784; Biognosys AG, Schlieren, Switzerland).[84] All files were searched against the mouse Uniprot databases (UP00000589_10090.fa, UP00000589_10090_additional. fa) of canonical and isoform sequences. For DIA single-shot analysis, a minimum of three fragments per peptide, and a maximum of six fragments were included for data extraction. Searches used protein N-terminal acetylation and methionine oxidation as variable modifications. Protein intensities were normalized using the "Local Normalization" (Q-value = 0.2) algorithm based on a local regression model.[85] A protein and precursor FDR of 1% was used. Default settings were used for other parameters. Within our regional brain tissue FAD comparison, after filtering and before performing statistical analysis, our data were filtered for at least 3,500 protein identifications per samples and at least 50% data completeness per protein. This resulted in a total of 4,296 protein identifications across 60 samples at a data completeness of above 95%. For downstream analysis, the 5% missing values were imputed from a downshifted normal distribution. For DIA analysis of human heart samples, library free approach was used and the data was analyzed using DIA-NN (Version 1.).[86]

### Proteomics downstream data analysis

Proteomics data analysis was performed in the Perseus environment (version 1.6.7.0),[87] Prism (GraphPad Software, version 8.2.1). MaxQuant output tables were filtered for 'Reverse', 'Only identified by site modification', and 'Potential contaminants' before further processing. Protein and peptide identifications were reported after filtering as described above. Proteome correlations across technical/analytical/biological replicates were performed after $\log_{10}$-transformation. Coefficients of variation (CVs) were calculated across the full dataset or within experimental groups on raw intensity levels for shared observations of more than one. Hierarchical clustering was performed in Perseus with default parameters and Pearson correlation as distance parameters. Before DE analysis, data were filtered for at least two observations in one group to be compared, followed by $\log_2$-transformation and imputation from a normal distribution modeled as the dataset with a downshift of 1.8 standard deviations and a width of 0.3 standard deviations. Deep

proteomes of biological replicates from fresh or vDISCO cleared tissue were tested for differences by a two-sided t-test. False-discovery rate control due to multiple hypothesis testing was performed by a permutation-based model and SAM-statistic with an $S_0$-parameter of 0.2 and an FDR of 0.01. Ontologies for cellular compartment assignment and keywords was performed with the mainAnnot.Mus_musculus.txt.gz followed by $log_2$-fold difference frequency counts for the terms 'Extracellular space', 'Blood microparticle', 'Neurodegeneration', 'Aging', 'Neurogenesis', 'Receptor', 'Virus-Host', 'Immunity', Wound healing' and 'Cell migration'. 1D enrichment analysis was performed on the two-sided t-test difference and only enriched terms with a size of larger than ten were displayed in the comparison of fresh versus vDISCO deep proteomes. CVs rank plots were calculated within each of the deep proteome groups and plotted against the median abundance of each protein within each group after $log_{10}$-transformation.

For the calculation of systematic ontology-related protein mass shifts, total protein copy number estimations of the deep fresh and vDISCO cleared proteomes of biological replicates were calculated using the Perseus plugin 'Proteomic ruler'.[88] Protein copy numbers were calculated with the following settings: Averaging mode. 'All columns separately', Molecular masses: 'Molecular weight [kDa]', Scaling mode: 'Histone proteomic ruler', Ploidy: '2', Total cellular protein concentration: '200 g/L'. Proteins were annotated with regards to their cellular compartment by gene ontology from the mainAnnot.mus_musculus.txt.gz. For protein mass estimates, we multiplied the resulting protein copy number by its protein mass for each conditional replicate and summed up all protein masses to obtain the total protein mass for each representative proteome reflecting 100% of the protein mass. To calculate the subcellular protein mass contribution, we calculated the protein mass proportion for the GOCC terms related to the cytoskeleton: 'Actin filament', 'Intermediate filament', 'Centrosome', 'Microtubule'; Membranes: 'Cytoplasm', 'Plasma membrane', 'Membrane'; Organelles: 'Mitochondrion', 'Nucleus', 'ER', 'Golgi apparatus'. For calculating the organellar change between the respective Fresh and vDISCO sub-proteomes, individual protein mass contributions were normalized by its total proteome mass first, followed by ratio calculation to obtain the percentage shift of protein mass between Fresh and vDISCO brains.

For PCA of both LCM applications (mTBI and FAD), data were grouped according to their condition, filtered for at least 760 or 900 proteins for the FAD or mTBI experiment respectively and at least 2 observations within one of the two conditions, column-wise median normalized, and missing values were imputed from a normal distribution with a width of 0.3 standard deviations that was down-shifted by 1.8 standard deviations. DE analysis for the FAD and mTBI experiment was performed by two-sided Welch's t-test on LFQ or IBAQ data respectively. False-discovery rate control due to multiple hypothesis testing was performed by a permutation-based model and SAM-statistic with an $S_0$-parameter of 0 or 0.2 and an FDR of 0.3 or 0.5 for the mTBI and FAD comparison, respectively.

## ADDITIONAL RESOURCES

- Videos related to DISCO-MS work: http://discotechnologies.org/DISCO-MS/
- Further details on the vDISCO protocol: http://discotechnologies.org/vDISCO/
- Videos related to workshops: http://discotechnologies.org/workshop/

## Article 9. Spatial proteo-transcriptomic profiling reveals the molecular landscape of borderline ovarian tumors and their invasive progression.

Preprint, *medRxiv*, 2023 (Nov, 13); DOI: 10.1101/2023.11.13.23298409

Lisa Schweizer[1,2], Rahul Krishnan[2], Aasa Shimizu[2], Andreas Metousis[1], Hilary Kenny[2], Rachelle Mendoza[4], Thierry M. Nordmann[1], Sarah Rauch[2], Lucy Kelliher[2], Janna Heide[2], Florian A. Rosenberger[1], Agnes Bilecz[2], Sanaa Nakad Borrego[2], Maximillian T. Strauss[3], Marvin Thielert[1], Edwin Rodriguez[1], Johannes B. Müller-Reif[1], Mengjie Chen[5], S. Diane Yamada[2], Andreas Mund[3], Ricardo R. Lastra[4], Matthias Mann[1,3,#], and Ernst Lengyel[2,#]

*[1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [2]Department of Obstetrics and Gynecology/Section of Gynecologic Oncology, University of Chicago, Chicago, IL, USA. [3]Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [4]Department of Pathology, The University of Chicago, Chicago, IL, USA. [5]Medicine/Section of Genetic Medicine, The University of Chicago, Chicago, IL, USA.*

*[#]Correspondence: elengyel@uchicago.edu or mmann@biochem.mpg.de*

In an example of investigating clinically relevant samples with DVP, we explored the molecular mechanisms of ovarian cancer. Serous borderline tumors (SBT) typically have a good prognosis after surgical tumor removal, but they often recur as highly invasive low-grad serous cancer (LGSC). Mutational profiling positions SBT as precursor lesions of LGSC, but little is known about the transition and their molecular landscape.

We applied DVP to a clinical cohort of four different disease states (SBT, SBT-MP, LGSC-PT and LGSC-Met) to resolve the epithelium (Epcam positive) and stroma (Decorin positive). Our data suggests that the protein expression of micropapillary SBT (SBT-MP) aligns more closely with SBT than with LGSC, positing it as an intermediary step in the tumorigenesis to invasive cancer. We identified alterations in the MAPK-signaling pathway and a gradual upregulation of cancer markers such as Kras and Rras. SBT-MP and LGSC displayed an upregulation of ADAM17, which promotes epithelial to mesenchymal transition (EMT) – a crucial step in tumor metastasis. SBT expressed PAX8, a primary marker for secretory cells of the fallopian tube. However, early stages of SBT also still expressed many proteins characteristic for normal ciliated cells including CAPS, TPPP3 and NUDC, in contrast to SBT-MP and LGSC.

In the stroma compartment, we demonstrated changes as the tumor becomes invasive. Most notably, proteins involved in EMT and ECM remodeling are upregulated in stroma and disease progression, indicating a dense tumor microenvironment. Interestingly, we detected a high immune response in SBT-MP at its onset, which decrease as the tumors became more invasive. In parallel, we also integrated our spatial proteomics approach with spatial transcriptomics and found that those technologies to be complementary.

The functional significance of our findings was tested using representative cell lines, which positively correlated on proteome level with LGSC-PT epithelium. Cell lines revealed significant inhibition of migration and invasion by knockout of the alternative splicing regulator NOVA2, SNCG, ADAM15, CLIC3 and POSTN, which we had found to be differentially expressed between SBT and LGSC. Proliferation was strongly inhibited upon NOVA2 and SNCG knockdown. We detected NOVA2 exclusively in epithelial cells in invasive LGSC. Furthermore, FDA-approved drugs acting on these pathways decreased proliferation and inhibited migration in the cell line models.

In conclusion, we combined spatial proteomics with spatial transcriptomics approaches and showed that their information is complementary. Of note, only proteomics identified the potentially causal proteins such as NOVA2. Our technology successfully identified many possible pathway contributions and a molecular switch for metastasis in cancer biology. This study is exemplary for further opportunities to solve molecular mechanisms of cancer biology.

***Contribution:***

Co-authorship. Matthias Mann and Ernst Lengyel conceived the study. Lisa Schweizer conducted the project as first author. I contributed to the study by supporting the DVP application, workflow and interpretation. More specifically, I performed the MS data acquisition of DVP samples and the reference cell lines. Additionally, I helped to revise the manuscript.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

# *Spatial proteo-transcriptomic profiling reveals the molecular landscape of borderline ovarian tumors and their invasive progression*

Lisa Schweizer[1,2], Rahul Krishnan[2], Aasa Shimizu[2], Andreas Metousis[1], Hilary Kenny[2], Rachelle Mendoza[4], Thierry M. Nordmann[1], Sarah Rauch[2], Lucy Kelliher[2], Janna Heide[2], Florian A. Rosenberger[1], Agnes Bilecz[2], Sanaa Nakad Borrego[2], Maximillian T. Strauss[3], Marvin Thielert[1], Edwin Rodriguez[1], Johannes B. Müller-Reif[1], Mengjie Chen[5], S. Diane Yamada[2], Andreas Mund[3], Ricardo R. Lastra[4], Matthias Mann[1,3,#], and Ernst Lengyel[2,#]

[1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany
[2]Department of Obstetrics and Gynecology/Section of Gynecologic Oncology, University of Chicago, Chicago, IL, USA
[3]Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[4]Department of Pathology, The University of Chicago, Chicago, IL, USA
[5]Medicine/Section of Genetic Medicine, The University of Chicago, Chicago, IL, USA
[#]Correspondence: elengyel@uchicago.edu or mmann@biochem.mpg.de

## Abbreviations

| Abbreviation | Description |
|---|---|
| AI | Artificial Intelligence |
| BIAS | Biological Image Analysis Software |
| DVP | Deep Visual Proteomics |
| LGSC-Met | Low-Grade Serous Cancer - Metastasis |
| LGSC-PT | Low-Grade Serous Cancer - Primary Tumor |
| MS | Mass Spectrometry |
| PCA | Principal Component Analysis |
| ROIs | Regions of Interest |
| SBT | Serous Borderline Tumor |
| SBT-MP | Micropapillary Serous Borderline Tumor |

## Abstract

Serous borderline tumors (SBT) are epithelial neoplastic lesions of the ovaries that commonly have a good prognosis. In 10-15% of cases, however, SBT will recur as low-grade serous cancer (LGSC), which is deeply invasive and responds poorly to current standard chemotherapy[1,2,3]. While genetic alterations suggest a common origin, the transition from SBT to LGSC remains poorly understood[4]. Here, we integrate spatial proteomics[5] with spatial transcriptomics to elucidate the evolution from SBT to LGSC and its corresponding metastasis at the molecular level in both the stroma and the tumor. We show that the transition of SBT to LGSC occurs in the epithelial compartment through an intermediary stage with micropapillary features (SBT-MP), which involves a gradual increase in MAPK signaling. A distinct subset of proteins and transcripts was associated with the transition to invasive tumor growth, including the neuronal splicing factor NOVA2, which was limited to expression in LGSC and its corresponding metastasis. An integrative pathway analysis exposed aberrant molecular signaling of tumor cells supported by alterations in angiogenesis and inflammation in the tumor microenvironment. Integration of spatial transcriptomics and proteomics followed by knockdown of the most altered genes or pharmaceutical inhibition of the most relevant targets confirmed their functional significance in regulating key features of invasiveness. Combining cell-type resolved spatial proteomics and transcriptomics allowed us to elucidate the sequence of tumorigenesis from SBT to LGSC. The approach presented here is a blueprint to systematically elucidate mechanisms of tumorigenesis and find novel treatment strategies.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

## Introduction

Serous ovarian tumors are the most common ovarian cancers. They are characterized as either high-grade (HGSC) or low-grade (LGSC), the latter displaying only mild atypia and few mitotic figures (Fig. 1a). Patients with LGSC tend to be younger than patients with HGSC and have slower growing, widely invasive tumors surrounded by dense fibrotic stroma, making complete surgical removal challenging. If the tumor cannot be entirely removed, metastatic LGSC (LGSC-Met) has an indolent clinical course with late recurrences and a low chance of cure. Although it is a different disease entity, LGSC is treated with the same standard platinum/taxane chemotherapy as HGSC. In LGSC, however, this treatment generally results in minimal clinical response and continued slow progression[1].

Mutational profiling has suggested that serous borderline tumors (SBT) are precursor lesions of LGSC[4]. SBT are neoplasms of epithelial origin, which mostly are unilateral without any stromal invasion (Fig. 1a, b). Patients with SBT usually have an excellent prognosis after operative tumor removal, but 10-15% will recur with an LGSC, which has very low response rates to current treatment strategies[2,3]. In contrast to HGSC, SBT and LGSC have wild-type p53, few DNA copy-number changes, and a low rate of somatic mutations. Both SBT and LGSC have mutually exclusive mutations in *BRAF*, *KRAS*, or *ERBB2*[6-8], all upstream regulators of the mitogen-activated protein kinase (MAPK), known to drive cell proliferation. A role for the MAPK pathway and estrogen receptor signaling has been identified in SBT and LGSC, but otherwise, little is known about their molecular landscape. This knowledge is a prerequisite to the identification of novel treatment options.

Here we set out to study the malignant transformation of SBT through putative intermediary steps (micropapillary SBT) to invasiveness in LGSC with spatial omics technologies. Deep Visual Proteomics (DVP) integrates high-content imaging, artificial intelligence (AI)-based cell recognition, laser microdissection, and mass spectrometry to preserve spatial information while quantifying the proteome in an unbiased manner[5]. We combined DVP with probe-based spatial transcriptomics to discover novel pathway alterations in LGSC and to find more effective treatments from the changes in protein and transcript expression during the transition from SBT to LGSC. The transcripts and proteins discovered with this approach included NOVA2, which is not normally expressed in the reproductive system but appears during the LGSC transition. We confirmed the relevance of the candidates identified by spatial omics by targeted functional screens in which a knock-down of the respective genes reduced invasion and migration.

## Results

### Spatial multi-omics of the transition of borderline tumor to low-grade serous cancer

Mutational profiling and histologic observations suggest a progression from serous borderline tumors (SBT) to micropapillary SBT (SBT-MP) and ultimately to low-grade serous carcinoma (LGSC), which may metastasize if not diagnosed and treated early[9] (Fig. 1a). Histologically, this hypothetical transition is characterized by irregularly contoured papillae with a hierarchical branching pattern in SBT that may progress to a micropapillary tissue architecture[10], with mild cytologic atypia, signifying a higher risk of progression to LGSC[11]. In contrast to these non-invasive phenotypes, once cells evolve to LGSC, they destructively invade into the stroma and frequently metastasize to the omentum[12].

We employed spatial proteomics (DVP), and transcriptomics (GeoMX) on a cohort of patients with SBT, SBT-MP, primary LGSC-PT, and corresponding omental metastasis (LGSC-Met) (Supplementary Table 1) for compartment-resolved characterization of different cell types (Fig. 1b). Using specific markers for epithelium (EPCAM) and stroma (Decorin), we stained all tissues with immunofluorescence and recorded high-content images. For DVP, cell populations were segmented on the images using the nucleAIzer AI algorithm[13] integrated in Biology Image Analysis Software (BIAS), and laser microdissected for ultra-high sensitivity mass spectrometry (MS) data acquisition

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

(Fig. 1c)[14]. DVP detected and quantified a median of 5,456 different proteins in the epithelium and 3,919 in the stroma (Supplementary Table 2), with little inter- and intra-specimen variability and excellent data completeness in all anatomic regions (Extended Data 1 a-f). A unique advantage of DVP is that it records and preserves the morphological characteristics of each cell within the tissue architecture. This is illustrated by more uniform cell sizes (Extended Data 1 g) and a longer distance to the closest cellular neighbor (Extended Data 1h) in LGSC-PT as compared to SBT, the latter being indicative of stromal invasion. All proteins and transcripts highlighted in the following sections are described in Supplementary Table 3.

## Progression of borderline tumors to metastatic low-grade serous carcinoma revealed by DVP

Principal Component Analysis (PCA) of our spatial proteomics results in the epithelial compartment clustered and ordered all four histologies (SBT, SBT-MP, LGSC-PT, LGSC-Met) sequentially. The analysis placed micropapillary growth as an intermediary stage between non-invasive serous borderline and invasive low-grade cancer (Fig. 2a). The LGSC and subsequent metastasis clustered closely, indicating that cells of the primary tumor and the metastasis were very similar after the transition to invasiveness. Linear regression analysis of the four histological stages showed an increase or decrease in 195 proteins in the invasive phenotypes; this panel of proteins represents an interesting subset representing the entire transition from SBT to LGSC and corresponding metastasis (adj. p-value <= 0.05, logFC > 1.5, Extended Data 2a, b, Supplementary Table 4).

Serous borderline tumors with micropapillary features in the background of an SBT or uniform SBT-MP have a higher risk of recurrence as LGSC than conventional SBT[11]. However, whether SBT-MP (Fig. 1a) is a precursor to LGSC is controversial. In our proteomic data, the most significant changes between SBT and SBT-MP included metabolic changes, such as abundance changes in the argininosuccinate synthase (ASS1), a key enzyme in the arginine biosynthetic pathway with ALDH6A1 and ALDH2 (Extended Data 3a). Several proteins showed major expression changes between SBT and SBT-MP, the first step in the hypothesized transition, and maintained their abundance in LGSC and its corresponding metastasis (Extended Data 2a, b, Extended Data 4). We found an increased abundance of the transcription factors AHDC1 and ERF, as well as proteins associated with cancer stemness (AQP5 and ASB6), which were higher expressed in SBT-MP when compared to SBT (Extended Data 4a,b). The tumor suppressor CDKN2A[15] was already downregulated in SBT-MP and this low protein expression persisted in LGSC and LGSC-Met (Extended Data 4c). This data suggests that SBT-MP protein expression is more similar to SBT than to LGSC, consistent with the PCA (Fig. 2a); however, SBT-MP protein expression reflects several hallmarks of malignant transformation. One patient had both SBT-MP and LGSC adjacent on the same ovarian section, presenting a unique opportunity to study the putative transformation of SBT-MP to LGSC (Extended Data 5a). The gradual progression from SBT to invasive cancer identified in the overall cohort, was confirmed in the PCA of this specific case (Extended Data 5b, c). The differential protein expression mirrored findings obtained by comparing SBT-MP and LGSC in the main cohort, including the downregulation of the ubiquitin ligase, TRIM25, the second most significant protein in LGSC tissue of the main cohort (Extended Data 5d).

A direct comparison of SBT to LGSC-PT in the epithelial compartment identified 963 significantly differentially expressed proteins (Supplementary Table 5). These included several upregulated proteins previously identified in high-grade serous ovarian cancer (SHMT1, TAGLN, Fig. 2b). Two tumor suppressors involved in epigenetic regulation were progressively lost in LGSC and its corresponding metastasis (ZMYND10, OSCP1/NOR1), as was Anterior Gradient Protein 2/3 (AGR2/3) previously implicated in the progression of SBT to LGSC[16] (Extended Data 2b). Comparing SBT and LGSC-PT, 14.5% of all proteins showed a significantly different abundance, in contrast to only 0.3% between LGSC-PT and LGST-Met, indicating the high degree of similarity between primary tumors and corresponding metastasis at the single cell type level. Neudesin (NENF),

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

involved in the proliferation of neural progenitor cells[17], was one of the few upregulated proteins in the metastasis (Extended Data 3a). We found a gradual increase in the adipogenesis regulatory factor ADIRF, which is associated with cisplatin resistance[18], and the folate receptor alpha (FOLR1), which is a target for Mirvetuximab Soravtansine, a treatment of platinum resistant high-grade serous cancers[19] that is not currently used for LGSC (Extended Data 2a).

The MAPK-signaling pathway has been found to be altered in more than 50% of LGSC and 60% of serous borderline tumors[20]. We found that oncogenic MAPK, ERK1/2 and Met signaling were enriched in LGSC when compared to the borderline tumors (Fig. 2c, Extended Data 3b, c, Supplementary Table 6). Specifically, KRAS and RRAS gradually increased from SBT to LGSC (Fig. 2c), as did downstream effectors of the c-Met receptor tyrosine kinase (Extended Data 3c). Likewise, several proteins downstream of TGF-β signaling were gradually upregulated either in SBT-MP or in LGSC, including oncogenic ADAM17, which promotes epithelial to mesenchymal transition (EMT)[21] (Extended Data 3d). Most SBT and LGSC express hormone receptors and are at least temporarily responsive to anti-hormonal treatment[22]. Accordingly, there was also an increase from SBT to SBT-MP in progesterone binding protein, PGRMC2, indicating a role for progesterone during the early transition to LGSC (Fig. 2d)[23].

We also confirmed the previous finding that serous borderline tumors express PAX8, a key marker for secretory cells of the fallopian tube[15]. However, SBT expressed a large number of proteins reflecting normal ciliated cells of the fallopian tube including CAPS, TPPP3 and NUDC[24], which, interestingly, were lost upon progression to LGSC (Extended Data 3b,e,f, Supplementary Table 6).

Immunofluorescence staining confirmed the proteomics results, detecting only sporadic ciliated cells in LGSC (Extended Data 3f). This was accompanied by a decrease in proteins of the apoptotic machinery, including FADD, CASP-3/8, and TRADD (Fig. 2e). Among proteins identified by MS in LGSC and metastasis but absent in SBT and SBT-MP, we found exclusive expression of the splicing regulator NOVA2, which is physiologically expressed in the brain, but not in the healthy ovary[25] (Fig 2f, g, Supplementary Table 7). Of note, NOVA2 was detected in the medium abundance range of MS intensities rather than indicating incomplete MS data acquisition by values in the low abundance range, thereby validating our previous proteomic analysis (Extended Data 3g). NOVA2 expression was confirmed using IHC (Fig 2h).

## Spatial proteomic analysis of the tumor microenvironment

Deep Visual Proteomics allows protein detection in different tissue compartments, taking advantage of AI-based recognition of individual cell types (Fig. 3a). When analyzing stromal cells (as opposed to epithelial cells before) which we identified by decorin IF staining, a total of 178 proteins significantly changed during the transition from SBT to LGSC (Extended Data 6, Supplementary Table 4). SBT and SBT-MP global protein expression was clearly distinct from that of LGSC and its corresponding metastasis in a PCA (Fig. 3b), demonstrating that major stromal changes occur as the tumor becomes invasive (Extended Data 7a, Supplementary Table 5). In contrast to the epithelial compartment (Fig. 2a), the stroma of SBT and SBT-MP was very similar, showing no intermediate progression towards LGSC.

Stromal protein expression changed primarily during the transition from SBT to LGSC (629 differentially expressed proteins), including the upregulation of ANXA2 and its regulator S100A10, which play a central role in cancer cell proliferation (Fig. 3c). Proteins involved in EMT such as AHNAK, LMCD1 and Periostin (POSTN) were among the most significantly differentially expressed stromal proteins of the SBT to LGSC transition. These changes in expression were accompanied by significant alterations in ECM remodeling proteins (PLEC, FBLN1/2, MXRA5, FN1 and its receptor ITGB5), indicating a dense tumor microenvironment.

Biological processes upregulated in LGSC-PT were primarily associated with cell dynamics; we found an upregulation of actin (ACTG1) and its central regulator ROCK2, as well as other components involved in cytoskeletal organization (ARPC2-4, CFL1, ITSN1, PAK1/2), all known to be associated with ephrin signaling (Fig. 3d, Supplementary Table 6). Elastic fiber formation

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

gradually increased towards invasive phenotypes (Extended Data 6b). Similarly, proteins involved in cell-cell adhesion were more abundant in the invasive LGSC and included the suppressors of apoptosis, COMP and GAS6, confirming our findings of decreased apoptosis in the epithelium (Extended Data 7c, Fig. 2e). In addition, we found an acute inflammatory response in micropapillary SBT, which was less evident in the invasive tumors (Fig. 3e, Extended Data 7d). Thus, during early transformation (SBT-MP), our data suggest that there is a strong host immune reaction; however, once an invasive tumor (LGSC) is established, the expression of proteins associated with an immune response is downregulated in the tumor organ.

Our data also suggest that the stromal transition from non-invasive to invasive tumor stages is accompanied by increased glucose metabolism, including the upregulation of FBP1 and PCK2/PEPCK, which are rate-limiting enzymes in gluconeogenesis[26] (Fig. 3f). The master metabolic fibroblast regulator, NNMT, previously shown by us to regulate the transition from normal fibroblasts to cancer-associated fibroblasts (CAFs)[27], was among the most highly upregulated proteins in invasive LGSC as confirmed by IHC (Fig. 3g). In the IHC analysis NNMT expression was also high in the epithelium.

## The progression of serous borderline tumor to invasive low-grade serous carcinoma characterized by spatial transcriptomics

Protein expression best reflects the functional phenotype of a cell; however, we reasoned that correlating protein expression to gene expression changes would help us better understand the biology of LGSC. Combining spatial proteomics with spatial transcriptomics could provide information about bidirectional communication between tumor cells and the surrounding stroma and identify treatment targets missed by either method alone[28].

Using the GeoMx technology, which allowed us to define specific regions of interest (ROIs), we performed spatial transcriptomics of the serial sections of all four histologies that had been used for H&E and spatial proteomics (Fig. 1a, b). Following hybridization with over 18,000 RNA-probes, focused UV light allowed us to selectively release the probe barcodes that hybridized to the cell types of interest for NGS analysis (Supplementary Table 8). Sequencing counts of transcript probes generally indicated high sequencing saturation (9,872 targets detected after QC, Extended Data 8). To integrate the spatial transcriptome with the proteome, we matched ROIs of transcript expression by immunofluorescence staining using compartment specific antibodies (Fig. 4a).

A total of 1,386 transcripts in the epithelium were differentially expressed between SBT and LGSC-PT (Supplementary Table 9). In SBT, the mucin-regulating AGR2 and the gel-forming mucin MUC5B were upregulated when compared to LGSC (Fig 4b). More abundant transcripts in LGSC-PT included the CA-125 binding partner mesothelin (MSLN) and the serine protease KLK6, which predicts the recurrence of borderline and low-grade ovarian tumors and is regulated by MAPK[29]. Interestingly, several gene transcripts highly expressed in the nervous system, such as SPOCK2 and gamma-synuclein (SNCG) were upregulated in LGSC-PT. In the proteomics data, gamma synuclein was found to have been upregulated during the early transition of SBT to SBT-MP (Extended Data 2a, 4).

During the step-wise transition from SBT to LGSC, the transcriptional regulator MAZ, which is regulated by MAPK, and the transcription factor DUX4[30], which is involved in oocyte development, were upregulated early in SBT-MP (Extended Data 9a). The expression of JUNB, FOS, and the Fos-related antigen 2 (FOSL2), in conjunction with their upstream regulators EGFR and SHC1, gradually increased as SBT transitioned through SBT-MP to LGSC (Fig. 4c, Extended Data 9b, c, and Extended Data 10a, b). In contrast, the transcriptional repressor HIC1, which is involved in the stabilization of p53, was downregulated in SBT-MP when compared to SBT (Supplementary Table 9). GSEA pathway analysis highlighted changes in mucins such as MUC3A, MUC5B, and MUC16 (CA-125) as well as tissue migration, semaphorin signaling, and response to EGF (Fig. 4d and Extended Data 9d-f, Supplementary Table 6). Another receptor tyrosine kinase (RTK) receptor, c-Met, and its downstream effectors showed a very significant increase in both LGSC and LGSC-Met

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

(Fig. 4e), which was confirmed by IHC (Fig. 4f) and is consistent with the spatial proteomics results (Extended Data 3c).

PCA analysis of the Nanostring data in the epithelium showed differences between SBT and SBT-MP, while stromal gene expression distinctly separated SBT/SBT-MP from LGSC-PT/LGSC-Met (PCA, Fig. 4g, Extended Data 11b). Several known stromal transcripts upregulated between SBT and LGSC, such as S100A10, C3, and NNMT (Fig. 4h, Supplementary Table 9) had also been identified by spatial proteomics (Fig. 3c, Supplementary Table 5). Others, only identified as transcripts, such as the metalloproteinase, ADAM15, and transcripts involved in oxygen-hemostasis (e.g. VEGFA, HIF1α), increased during the transition from SBT to SBT-MP (Fig. 4i). In the early differences between SBT and SBT-MP, we found an enrichment of the stem cell markers LGR5 and FOXL2 in SBT (Extended Figure 10c,d, Extended Data 11a). The neural axon guidance factor SLIT2 was highly expressed in LGSC-Met, while the cell-adhesion protein CDHR1 was downregulated (Extended Data 11a). Upregulated pathways in CAFs[31] included genes involved in angiogenesis and hypoxia (Fig. 4j, Extended Data 11c).

## Integration of spatial proteomics and transcriptomics technologies

We reasoned that the multiple spatial data layers – single cell-type proteomics, transcriptomics and H&E staining – might complement each other when trying to understand the tumor organ. In particular, integrating protein and transcript expression could uncover the complex interplay of transcription and translation. Among the proteins and transcripts that did not change between SBT and LGSC-PT (Supplementary Tables 5, 9), 75% of the proteins and 38% of the transcripts overlapped between DVP and Nanostring in the stroma, while 43% of the proteins and 29% of the transcripts overlapped in the epithelium. In contrast, there was less overlap of the differentially expressed proteins and transcripts – 4% of proteins and 2% of transcripts in the stroma as well as 3% of proteins and 2 of transcripts in the epithelium with many Nanostring values not present in the DVP data and *vice versa* (Extended Data 12a, b).

These values reflect differences in the gene products detectable by the two technologies and potential differences in biological regulation[32]. We therefore focused on genes for which both the transcript and the protein were quantified, integrating the two datasets into a list of proteins/transcripts acquired through spatial proteomics and transcriptomics.

To determine markers with a higher probability of impacting tumorigenesis, we selected the most relevant proteins/transcripts acquired with both technologies for epithelium and stroma, respectively (selection scheme in Extended Data 12c). Downregulated targets were included if present in a tumor suppressor database[33]. The resulting list of 70 potential targets was integrated with a comprehensive literature review to annotate for biological relevance (Extended Data 12c-g, Supplementary Table 10). Notably, multiple proteins in both the stroma and the epithelium were targetable with clinically approved or pre-clinical inhibitors. These included CCN2 and ANXA2 in the stromal compartment, FOLR1, CRYAB and HTRA1 in the epithelium, and ADAM15 in both cell types.

Using serial sections for H&E, DVP, and Nanostring allowed perfect alignment of regions of interest from the H&E to transcriptomics and proteomics (Fig. 5a). Correlating the proteome and transcriptome data sets across the four histologies, a total of 1,142 of 4,992 genes present in the epithelium showed a positive correlation across the transition from SBT to LGSC-Met (Fig. 5b). In the stroma, 84 of 3,746 significantly correlated between transcripts and proteins (Fig. 5c). Only very few genes were anti-correlated between the two technologies in the epithelial compartment (total of 32, including ABT1, ZMIZ1, CIR1), and only a single gene in the stroma (CHST14). The majority of the 70 genes of potential biological relevance in the transition from SBT to LGSC selected above showed a similar trend between technologies (Extended Data 12c-g, Supplementary Table 10).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

## From molecular signatures to functional validation: Characterizing the transition of serous borderline tumor to low-grade serous cancer

Having characterized the proteomic and transcriptomic landscape through the progression to metastatic LGSC, we set out to understand the functional significance of the putative drivers of the transition between SBT and LGSC. We used the most significantly altered protein/transcript panel (Extended Data 12c-e, Supplementary Table 10) to evaluate specific alterations from SBT to LGSC. To this end, we first characterized several LGSC cell lines using proteomics and correlated these results with the epithelial compartment acquired with spatial proteomics in human tissue (Fig. 2). All five low-grade cell lines correlated positively with the proteomic findings in the epithelium of SBT and LGSC-PT, with the VOA4627 cell line showing the highest correlation with LGSC-PT (Fig. 6a). Based on these results, a siRNA screen was performed using the VOA4627 LGSC cell line to study migration. Knockdown of ADAM15, CLIC3, NOVA2, POSTN and SNCG, significantly inhibited VOA4627 migration (Fig. 6b). Individual knockdown of all five genes inhibited proliferation of VOA4627 and invasion of VOA4627 and VOA6406 cells, while knockdown of NOVA2 and SNCG significantly inhibited VOA6406 proliferation. (Fig. 6c, Extended Data 13a-c).

We explored the role of prominent pathways and targets indicated by our integrative omics approach (Extended Data 12) testing FDA-approved drugs on cancer cell proliferation and the invasion of LGSC cell lines. The c-MET inhibitors, Savolitinib, Icotinib and Cabozantinib, the EGFR inhibitors, Lapatinib and Erlotinib, and the CDK inhibitors, Abemaciclib and Milciclib, all inhibited proliferation and invasion, while a FOLR1-antibody drug conjugate, Mirvetuximab Soravtansine only inhibited the proliferation of VOA4627 and VOA6406 cells, consistent to its mechanisms of action on tubulin which regulates cell proliferation (Fig. 6d, Extended Data 13d). In summary, all pathways projected to be of biological significance from the spatial proteomics or transcriptomics results, turned out to be of functional significance when blocked using FDA approved inhibitors.

NNMT was the most upregulated putative stromal driver in LGSC. Knockdown of NNMT in immortalized human CAF cells[27] inhibited the CAF-conditioned media-driven proliferation and invasion of VOA4627 and VOA6406 cells (Fig. 6e, Extended Data 13e). A robust and reliable *in vivo* mouse model of LGSC was generated by *in vivo* passaging of VOA6406 cells. Intra-tumoral treatment with an NNMT inhibitor (NNMTi) reduced subcutaneous tumor growth of VOA6406ip1 after 4 weeks of treatment (Fig. 6f).

## Discussion

We performed a detailed analysis of the molecular transition from serous borderline to micropapillary serous borderline tumors and low-grade serous carcinoma and its corresponding metastasis, using cell-type resolved spatial proteomics and transcriptomics. This was enabled by the wider availability of spatial transcriptomics and ongoing improvements of the DVP technology[5], which now utilizes an enhanced, glycerol-based slide staining protocol[34], robotic sample preparation, and an increased cutting throughput. Employing this strategy, we identified up to 6,000 proteins in an equivalent of 200 cells in over 100 surgical tissue samples. The tissue-based nature of the DVP and the Nanostring GeoMx spatial transcriptomics platforms permits the use of serial tissue sections for the acquisition of complementary results for the two layers of gene expression and the correlation of the results with H&E staining. Overall, both transcriptomics and proteomics adequately covered previously known biological changes in cancer progression. Currently, transcriptomics provides greater depth, but proteomics was better able to cover changes in the stromal compartment. Proteomics accurately quantified protein level changes during progression, and it is expected that the depth of proteome coverage will increase even further.

Compared to conventional SBTs, serous borderline tumors with micropapillary features (SBT-MP) often involve both ovaries and have a higher risk of recurrence as LGSC[35]. Our proteomic data revealed that SBT-MP is an intermediate stage in the transition between SBT and LGSC, as had been postulated based on histology. In the transition of SBT to SBT-MP, we noted the upregulation of genes involved in tumor progression as well as stem cell markers. At the pathway

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

level, this included MAPK signaling, amino acid metabolism changes, and various transcription factors (AHDC1, MAZ, ERF). While transcript and protein expression in the stromal compartment generally did not change between SBT and SBT-MP, proteomics revealed a sudden increase in proteins associated with an inflammatory response, although this subsided once SBT-MP progressed to an invasive cancer (LGSC, LGSC-Met). We speculate that, during the early phase of transformation, the host immune response is activated but then is abrogated when the tumor becomes invasive. Indeed, DUX4, one of the transcription factors we found upregulated in SBT-MP and maintained in LGSC, promotes cancer immune evasion[36].

The previous molecular characterization of SBT and LGSC focused mainly on genomic profiling and hormone receptor expression. Those studies consistently highlighted changes in RAS genes, *BRAF*, ERBB2, and NF1, all regulators of the MAPK pathway, which is altered in more than 50% of all LGSC and borderline tumors[20,37]. By extending such studies to the proteomics and transcriptomics levels, our results unraveled changes in other targetable components of the MAPK pathways, from receptor tyrosine kinases to serine/threonine kinase and transcription factors, revealing a gradual activation of the entire MAPK network[38] as borderline tumors undergo a transformation from SBT to SBT-MP to LGSC, and metastasize. By systematically analyzing the transition from SBT to LGSC-Met and separately analyzing changes in the stroma and epithelium, we expanded these insights into other molecular changes in LGSC beyond MAPK (Fig 6 g). A strength of our analysis was the excision of epithelial tumor cells directly from their microenvironment allowing for cell-type specific resolution. Additionally, our spatial compartment-resolved analysis enabled us to study the stromal compartment in closest proximity to the tumor.

Our data suggest that EMT is an essential mechanism in the transition of SBT to LGSC. Indeed, the transcript and protein signatures in the stroma surrounding LGSC suggest marked ECM remodeling and support the observation that stiffness and a denser microenvironment is often associated with progression[39]. The transition of ductal carcinoma in situ to invasive breast cancer is similarly marked by changes in the EMT and a dense tumor stroma[40]. Beyond recapitulating EMT as tumors dedifferentiate, the spatial omics technologies identified specific, actionable targets in the stroma (e.g. ADAM15, CCN2, ANXA2). The most prominent protein and transcript change was NNMT, which reprograms normal fibroblasts into CAFs[27]. We found that a knock-down of NNMT reduced the proliferation in in vitro experiments and in vivo tumor growth, opening up the possibility of stromal-directed treatment approaches as part of LGSC treatment.

One of our most striking findings in the transition to invasiveness was the appearance of proteins associated with the brain. Specifically, proteomics identified NOVA2, a brain-specific splicing regulator, as highly expressed in LGSC and its corresponding metastasis, but absent in SBT and SBT-MP. Integrating spatial proteomics and transcriptomics, followed by validation and functional studies, identified previously unknown mechanisms for LGSC: CLIC3, POSTN, ADAM15 and SNCG represent targets previously unknown in LGSC. The highly significant proteins identified by both spatial techniques and selected for further studies were further narrowed by an siRNA screen and confirmed by functional assays. Based on our combined in vivo and in vitro results, our data supports further development of c-Met, EGFR, NNMT, and treatment using an FOLR1-specific antibody-drug conjugate in LGSC.

In summary, this first multi-omic spatial atlas of low-grade serous ovarian cancer progression molecularly dissected a previously hypothetical transition pathway, identifying a plethora of functionally important proteins and pathways, and suggested novel treatment targets. By characterizing the molecular landscape of serous borderline tumors and low-grade serous cancer on two omics levels, we provide an in-depth resource for further downstream analyses of the diseases, while demonstrating the translational relevance of our results in-vitro and in-vivo in this study. The strategy exemplified here could serve as a blueprint for research into other cancers or diseases that would benefit from multi-modal, cellularly, and spatially resolved elucidation of the transition from benign to malignant phenotypes.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

## Acknowledgments

We would like to thank Dr. Annette Feuchtinger and her team from the Core Facility Pathology & Tissue Analytics at the Helmholtz Center, Munich, Germany for their support during image data acquisition; Dirk Wischnewski (MPI Munich), Destiny Brown (U of C) and Paul Miller (U of C) for technical support; members of the Department of Proteomics and Signal Transduction at the Max Planck of Biochemistry in Munich, Germany, the ovarian cancer research laboratory in Chicago, USA, as well as members from the Center for Protein Research at Copenhagen University, Denmark, for fruitful discussions. We thank Pieter Farber, Ph.D. from the U of C Genomics facility for help with the spatial transcriptomics. We thank Dr. Mark Carey (The University of British Columbia, Division of Gynecologic Oncology, USA), Dr. Ie-Ming Shih (The Johns Hopkins Hospital, Department of Pathology, USA) and Dr. Fernando Guimarães (University of Campinas, Experimental Pathology Laboratory, Brazil) for low-grade cancer cell lines. We thank Gail Isenberg (U of C) for editing the manuscript.

This research has been supported by generous philanthropic gifts from Linda Usher and family (SD Yamada and E Lengyel), Bears Care - the charitable beneficiary of the Chicago Bears Football Club (H. Kenny and E. Lengyel), and a NIH/NCI R35 CA264619 (E Lengyel). The Cellular Screening Center and Human Tissue Resource Center Cores at the University of Chicago are funded by the Cancer Center Support Grant (P30CA014599). Furthermore, this study was supported by the Max Planck Society for Advancement of Science (M.M.). L.S. acknowledges funding from the German Federal Ministry of Education and Research under the funding code 16LW0243K. L.S. and M.T. were supported by the International Max Planck Research School for Life Sciences – IMPRS-LS. F.A.R. is an EMBO postdoctoral fellow (ALTF 399-2021) and T.M.N. is supported by a Swiss National Science Foundation (SNSF) Early Postdoc Mobility (P2ZHP3-199648) and Postdoc Mobility Fellowship (P500PM_210917). M.T. was supported by European Union's Horizon 2020 research and innovation program under grant agreement No. 874839 (ISLET).

## Contributions

The study was conceived by E. L. and M.M. Deep Visual Proteomics (DVP) application, data acquisition and interpretation were performed by L.S., R.K. and A.M. with the support of M.T., T.M.N., F.A.R., E.R. and A.B. Spatial transcriptomics data were acquired by R.K., L.S., A.B. and S.N.B. Omics data were statistically evaluated and integrated by L.S., M.S., M.C., and J.B.M-R. The patient cohort was identified by E.L., R.M., S.D.Y, R.K, and R.R.L. Pathological expertise was provided by R.M., R.R.L. and A.B. Functional experiments were designed by A.S., H.K., L.S., J.H., and E.L., and performed by A.S., H.K., S.R., L.K. and J.H. Figures were prepared by L.S. and H.K.
The paper was written by L.S. and E.L. The manuscript was edited by H.K., A.M., and M.M. All authors reviewed, provided feedback, and approved the final version of the manuscript.

## Data availability

All data will be publicly available upon publication in a journal.

## Competing interests

E.L. receives research funding to study ovarian cancer from Arsenal Bioscience and AbbVie through the University of Chicago unrelated to this work. The authors declare no competing interests in the context of this manuscript.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

# Methods

## Patient cohort and ethics

Patients who underwent primary surgery for a newly diagnosed borderline ovarian cancer or low-grade serous ovarian cancer at the University of Chicago were retrieved from the University of Chicago ovarian cancer database[42]. All patients gave written informed consent in compliance with the University of Chicago Institutional Review Board-approved protocol and in accordance with the Declaration of Helsinki. All surgeries were performed by board certified gynecologic oncologists (S.D.Y, E.L). The tumor pathology was reviewed and confirmed by two gynecologic pathologists (R.R.L., R.M.) prior to inclusion in the study.

## Deep Visual Proteomics

### Immunofluorescence

Formalin-fixed paraffin embedded (FFPE) tissue was sectioned (2.5 µm) from paraffin blocks and mounted on 1.0 PEN membrane slides (MicroDissect, MDG3P40AK). To enhance tissue adhesion, membrane slides were incubated for 1 hour under UV light and treated with VECTABOND reagent (Biozol, VEC-SP-1800) according to the manufacturer's instructions. Mounted slides were incubated at 55°C for 30 min and deparaffinized through xylene (2 min) and 100% ethanol, 95% ethanol, 75% ethanol, 30% ethanol, and distilled water (1 min) two times. Antigen retrieval occurred in 1x DAKO pH9 HIER buffer (Agilent Dako)/ 10% glycerol (v/v) (Sigma) in a preheated water bath at 90°C for 60 min, followed by the blocking of nonspecific binding sites using 5% BSA in PBS for 60 min at room temperature.

Directly conjugated antibodies targeting EPCAM (Abcam, recombinant Alexa Fluor® 555, ab275122, 1:100) and Decorin (Abcam, recombinant Alexa Fluor® 647, ab28132, 1:100) were diluted in antibody diluent solution (Agilent Dako, S080983-2) and incubated at 4°C overnight in a wet staining chamber. Nuclei were stained with SYTOX™ green (1:400, Thermo Fisher Scientific) for 10 min at room temperature and mounted with aqueous mounting medium (SlowFade™ Diamond Antifade Mountant, Thermo Fisher Scientific).

### Image acquisition and analysis

High-resolution fluorescence images were acquired on an Axio Scan.Z1 (Zeiss) microscope, coupled to a 20x/0.8 M27 dry objective and the scanned slides saved as '.czi' files. Images were recorded implementing a 10% tile overlap, five z-stacks (offset -5um) and a bin mode of 1x.1, using optimized exposure times for each fluorescent channel (AF647: 40ms, AF555: 300ms, AF488: 10ms). For image post-processing, z-planes were collapsed into a single plane based on the variance of pixel values ('Extended Depth of Focus - EDF') and stitched to achieve precise matching of tiles. Then, images were imported into the Biology Image Analysis Software ('BIAS') and analyzed[14]. Briefly, segmentation was performed for epithelial cells (Algorithm: 'Generic cytoplasm segmentation v1.0', Settings: input spatial scaling: 1.0, detection confidence: 50%, contour confidence: 50%, region properties: 10-500 µm$^2$) and stromal cell equivalents (Algorithm: Generic nucleus segmentation v1.0. Settings: input spatial scaling: 2.4, detection confidence: 50%, contour confidence: 50%, region properties: max. 100 µm$^2$, dilated by 9 µm$^2$).

Once regions of interest were selected using matching regions in spatial transcriptomics, shapes of single cells were exported while defining three reference points for coordinate system transfer. To improve the efficiency of laser-guided shape extraction, polygon reduction was accomplished by implementing the Ramer-Douglas-Peucker algorithm. To facilitate this, an interactive web interface was developed using Python (version 3.8.5) in conjunction with the Streamlit library (version 1.19.0). Internally, data manipulation tasks were carried out using the numpy (version 1.22.2) and pandas (version 1.4.0) libraries. Visualization of both original and reduced shapes was performed using the plotly library (version 5.5.0). Epsilon values were chosen interactively to find an optimum to preserve shapes and reduce points. Upon image acquisition, cover

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

glasses were removed and the tissue dried thoroughly at room temperature to enable precise laser cutting thereafter.

### Laser microdissection

Cells were cut from the tissue using the laser microscope (Leica Microsystems) and collected into a dry 384-well plate (Eppendorf) while maintaining a stable temperature of 31.9°C. AI-defined shapes of cells were imported using the reference points defined in the BIAS software and minimal correction of shape alignment was performed. Laser extraction was performed directing a diode-pumped solid-state laser (349 nm) *via* a HC PL FLUOTAR L 63x/0.70 objective (power: 59, aperture: 1, speed: 20-25, head current: 42-49%, pulse frequency: 2450-2600, offset: 214-219) conducting a final middle pulse to collect the shapes vertically into the well. Considering the surface area of the cell as indicator of final protein amounts injected into the mass spectrometer, a total of 700 epithelial cell shapes and 150 stroma equivalents were collected to compensate for differences in area sizes.

### Sample preparation

All laser dissected samples in the 384-well plate were processed in parallel by the implementation of an automated liquid handling platform (Agilent Bravo). Extracted cells were concentrated at the bottom of each well by addition of 28 µL of 100% acetonitrile, centrifugation at 2,000 g for 10 min and vacuum evaporation for 15 min (60°C). Cells were lysed in 4 µl of 60 mM triethylammonium bicarbonate (TEAB) in $H_2O$ at 95°C for 60 min. After the addition of 1µL of 60% acetonitrile (final concentration of 12% (v/v)), the samples were incubated at 75°C for 60 min. Proteins were digested sequentially, adding 1 µL of 4ng/µL LysC (Wako, 129-02541; in 60 mM TEAB, 12% I) for 3 hours and 1.5 µL of 4 ng/µL trypsin (Sigma-Aldrich, T6567; in 60 mM TEAB, 12% I) overnight at 37°C. The enzymatic digest was quenched using a final concentration of 1% (v/v) trifluoroacetic acid (TFA), centrifuged for 5 min at 1,000g and vacuum dried at 60°C. Samples were resuspended in 2% acetonitrile (v/v), 0.1% trifluoroacetic acid (v/v) and the entire volume injected for MS data acquisition.

### Mass spectrometry measurements and data processing

The LC system of choice was an EASY nanoLC 1200 (Thermo Fisher Scientific). Peptides were separated on a 50 cm in-house packed HPLC column[43] (75 um inner diameter packed with 1.9 um ReproSil-Pur C18-AQ silica beads (Dr. Maisch GmbH)) with a linear gradient of 120 min from 3 to 30% buffer B in 95 min, followed by an increase to 60% for 5 min, washed at 95% buffer B for 10 min and re-equilibration for 10 min at 5% buffer B (buffer A: 0.1% formic acid (FA) and 99.9% ddH2O; buffer B: 0.1% FA, 80% ACN, and 19.9% ddH2O). The flow rate was kept constant at 300 nl/min, and the column was heated to 60°C by an in-house manufactured oven. The EASY LC system was coupled to a timsTOF SCP mass spectrometer (Bruker) via a nanoelectrospray ion source (Captive spray source, Bruker). The mass spectrometer was operated in dia-PASEF mode using the 16 diaPASEF scan acquisition scheme (standard scheme)[44]. The method covered a m/z range from 400 to 1200 and ion mobility of 0.6 to 1.6 Vs $cm^{-2}$. All other settings were described previously[45]. For cell line experiments, the LC system was coupled to a timsTOF Pro2 mass spectrometer (Bruker) with settings as described above.

Raw data was searched using the DIA-NN software[46,47] (version 1.8.0). Searches were performed separately for the epithelial and stromal compartments using the library-free search and the human Uniprot databases (UP000005640_9606 with isoforms, February 28, 2022). In short, a deep-learning module, match-between-runs (MBR) and heuristic protein inference ('--relaxed-prot-inf') was enabled. N-terminal methionine excision and carbamidomethylation were set as fixed modifications, 'IDs, RT & IM profiling' was used for library generation, 'robust LC (high accuracy)' for quantification and 'Global' for cross-run normalization. The 'pg_matrix.tsv' output file was used for further data analysis.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

## Bulk proteomics of cell lines

The proteome of five cell lines was measured. Cells were harvested and washed in PBS. Pellets were lysed for 10 min at 90°C in 50 uL lysis buffer (12.5% acetonitrile, 300 mM Tris/HCl pH 8.0, 5 mM TCEP, 25 mM CAA), sonicated in a Bioruptor Plus (Diagenode, sonication cycles of 30 sec for 15 min) and heated at 90°C for 30 min. Protein concentration was measured using a Nanodrop instrument (ThermoFisher Scientific), proteins were digested using LysC and trypsin in a ratio of 1:100 over night at 37°C. Digests were quenched by adding trifluoroacetic acid (TFA) to a final concentration of 1% and peptides purified using two layers of SDB-RPS as active matrix in stage tips (Wash 1: isopropanol, 1% TFA; Wash 2: 0.2% TFA; elution buffer (80% I, 1% $NH_4$+)). Purified peptides were vacuum dried for 40 min at 60°C and resuspended in 2% acetonitrile (v/v), 0.1% trifluoroacetic acid (v/v). A total of 200ng were injected into the mass spectrometer and measured as described above.

## GeoMx Spatial Transcriptomics

FFPE tissue sections (5 µm) sliced consecutively to Deep Viusal Proteomics samples were processed following the Nanostring GeoMx user manual (MAN-10132-04). In brief, tissue slides were deparaffinized and rehydrated, followed by protein target retrieval using antigen retrieval buffer (Tris-EDTA) for 20 min in a pressure cooker. RNA target retrieval was accomplished by a digest with proteinase K for 15 min at 37°C. Tissue was post-fixed in 10% NBF, followed by an overnight hybridization at 37°C with the RNA probes (Human Whole Transcriptome Atlas, 18,000 protein-coding genes coupled to UV-cleavable oligonucleotide barcodes). Excess probes were removed the next day by washing the samples twice in stringent washes at 37°C, followed by an incubation period in blocking buffer W (Nanostring LOT#:2-23020032). Thereafter, tissue was stained using conjugated primary antibodies targeting pan-cytokeratin (mouse monoclonal antibody, Novus NBP2-33200AF488) and vimentin (mouse monoclonal IgG1κ, Santa Cruz sc-373717) as well as the SYTO13 nuclear stain (ThermoFisher Scientific, S7575) for one hour at room temperature.
For transcriptomics data acquisition, slides were placed on the GeoMx Digital Spatial Profiler (DSP) and scanned in 20x magnification. Based on immunofluorescence images, regions of interest (ROI) were collected under supervision of an experienced pathologist (AB) and matched to the regions selected in DVP. Upon ROI selection, oligonucleotide barcodes were collected into a 96-well plate by UV-ablation while precisely separating compartments in each region. Sequencing libraries were prepared using the Illumina TruSeq technology for the ligation of sequencing adaptors and amplification. Amplified libraries were purified using two rounds of Ampure XP magnetic bead cleanup (ratio 1.2:1 (beads: library)) (HighPrep PCR from MAGBIO (Cat#: AC-60500). Purified libraries were sequenced in an Illumina NovaSeq flowcell PE50 at a depth of 100x /µm2.

## Data analysis

Bioinformatics analyses of Deep Visual Proteomics and spatial transcriptomics data was performed in the R statistical environment (version 4.2.2). Protein intensities and GeoMx counts were log2 transformed and evaluated alongside as described below. For each dataset, samples belonging to the main cohort were selected, while addition samples ('case studies') were processed separately. For proteomic quality control, data completeness was determined by calculating the number of regions (0-3) in which a protein was consistently identified across a specimen. Coefficients of variations (CVs) were determined on non-logarithmic data and complete MS intensities (no 'NA' values) (i) within a specimen (intra-specimen) by calculating the variation between different regions and the mean CV values for each group and (ii) between specimen (inter-specimen) by calculating mean MS intensities of regions for each specimen and the CVs of mean region values within each cohort group.

GeoMx raw count data was processed and normalized using the R GeoMxTools package [1]. Segments with fewer than 1000 raw reads, below ~75% for % Aligned, ~80% for %Trimmed and Stitched sequencing reads, were removed. We also removed segments with sequencing saturation <50%, negative count < 1 and No Template Control count > 9000. A probe is removed globally if the

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

geometric mean of that probe's counts from all segments divided by the geometric mean of all probe counts representing the target from all segments is less than 0.1 or the probe is an outlier according to the Grubb's test in at least 20% of the segments. Segments with less than 5% of the genes detected were removed and genes detected in at least 5% of the segments were kept. Counts were then normalized by Q3 normalization.

Differential protein/transcript abundances of both datasets were determined using the 'Limma' package (v3.54.2) using the false discovery rate (FDR) for multiple testing correction and fixed parameters (adjusted p-value <= 0.05 and logarithmic foldchange > 1.5) for the assessment of significance. Likewise, 'Limma' was used for regression analyses across groups using a linear model fit followed by empirical Bayes statistics, while the same significance cutoff was used as described above. An overlap of MS protein identification was determined counting 'NA' and valid values per group and a combination matrix for upset plots, as well as all heatmaps, were generated using the ComplexHeatmap package (v2.14.0). For all Principal Component Analyses (PCA) and omics data integration, data were filtered per protein for valid values in 2 out of 3 regions in all specimens of at least one group. Imputed sample-from normal distribution (*width of 0.3, downshift of 1.8*). PCA were performed using the FactoMineR package (v2.8). Biological pathway enrichments were accomplished by a gene set enrichment analysis (GSEA) on significantly differentially regulated proteins/transcripts using the WebGestalt gene set analysis toolkit (v0.4.5) in reference to the 'Reactome Pathway' and 'non-redundant geneontology Biological Process' databases and a Benjamini-Hochberg FDR correction (cutoff: 0.05). Ciliated and secretory cell markers were extracted from the top 100 abundant transcripts of the ciliated and secretory cells in the post-menopausal fallopian tube as presented in recently published work[24]. Remaining plots were generated using the ggplot2 package (v3.4.2) and mean comparisons of p-values were added using paired Student's t-tests and equal variances. For the integration of omics data, regions of interest were filtered for dataset counterparts and protein/transcript names were matched by first prioritizing non-isoform proteins and then prioritizing the percentage of valid (not 'NA') values when selecting between isoforms. If two ROIs of the spatial transcriptomics matched one DVP region, mean GeoMx counts were calculated. Previously imputed data were used for the proteomic dataset. Correlation between datasets was tested by Pearson correlation using the 'stats' package (v4.2.2) and Benjamini-Hochberg (FDR) correction. Stacked image creation of our multi-omics approach was done with a custom Python script (version 3.10.11) and Adobe Illustrator (27.5) using the following libraries: pandas (1.5.3), untangle (1.2.1), numpy (1.24.3), PIL (9.4.0) and matplotlib (3.7.1). Initially, the shape data was superimposed on raw slide images, followed by manual adjustments as necessary. Subsequently, each color layer was exported separately as an overlay and integrated into Adobe Illustrator. Schematics were created with BioRender.com (Figure 1) and Adobe Illustrator (Figure 6).

## Antibody-based validation staining

### Immunohistochemistry

FFPE tissue was sectioned (5 µm) on SuperfrostTM Plus Microscope Slides (Fisher Scientific, 22-037-246). Then, the immunohistochemistry was performed on Leica Bond RX automated stainer. After the standard procedures for deparaffinization and rehydration, tissue sections were treated with Proteinase K (Agilent, S3004) for 5 minutes pre-treatment at room temperature. Anti- NOVA2 [(Novus Biologicals, # NBP1-92196, 1:150], c-MET [Invitrogen, #18-7366, 1:150], NNMT [Santa Cruz Biotechnology, #sc-376048, 1:100], PGRMC2 ThermoScientific, #PA5-59465,1:200], FOLR1 [ThermoScientific, #PA5-24186, 1:200], JUN [Cell Signaling technology, #9165, 1:200], EGFR [ProScience, #33-350, 1:25] was applied on tissue sections for 60 minutes incubation at room temperature. The antigen-antibody binding was detected by Bond Polymer Refine Detection (DS9800, Leica Microsystem). Tissue sections were briefly immersed in hematoxylin for counterstaining and were covered with cover glasses. These slides were imaged using Olympus VS200 Slideview. Images for publication were exported from QuPath version 0.3.2.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

### Immunofluorescence

FFPE tissue was sectioned (5 µm) on SuperfrostTM Plus Microscope Slides (Fisher Scientific, 22-037-246) and deparaffinized as described above (see section 'Deep Visual Proteomics'). For antigen-retrieval, samples were heated in a water bath at 90°C for 30 min in 1x DAKO pH9 HIER buffer (Agilent Dako) and blocked using 5% BSA in PBS for 45 min at room temperature. Primary antibodies for PAX8 (Mouse mAb 28556, Cell Signaling Technology, 1:400) and CAPS (Rabbit polyclonal PA5-60401, Thermo Fisher Scientific, 1:50) were diluted in antibody diluent solution (Agilent Dako, S080983-2) and incubated at 4°C overnight in a wet staining chamber. Secondary antibodies (anti-Mouse IgG1 AF647, A-21240, Thermo Fisher Scientific, 1:200; anti-rabbit IgG AF750, ab175735, Abcam, 1:200) and ConcavalinA conjugated to tetramethylrhodamine (C860, Thermo Fisher Scientific, 1:150) were incubated for 60 min at room temperature. Nuclear staining was performed using SYTOX$^{TM}$ green (1:400, Thermo Fisher Scientific) for 10 min at room temperature and slides were mounted (Vectashield Vibrance Antifade Mounting Medium, Vector Laboratories, H-1700). The resulting staining was visualized on a Axio Scan.Z1 (Zeiss) microscope as described above (see section 'Deep Visual Proteomics').

### Cell lines and culture conditions

VOA3448, VOA4627, and VOA6406 (M. Carey, University of British Columbia) cells were cultured in a 1:1 mixture of MCDB 105 (Cell Application Inc.) and Medium 199 (Gibco), supplemented with sodium bicarbonate (Corning), L-glutamine (Corning), and 10% fetal bovine serum (FBS). The cells were maintained at 37°C in a humidified incubator at 5% CO2. The cell lines were banked in liquid nitrogen and one vial of each passage were confirmed *Mycoplasma* negative using the STAT-Myco kit. The cell lines were validated using short tandem repeat DNA fingerprinting with the AmpFℓSTR Identifier kit and compared with known fingerprints by IDEXX BioAnalytics Laboratories (Columbia, MO). Cells were passaged 2–10 times after thawing before commencing with experiments.

### siRNA transfection

For transient transfections, LGSC were transfected with 25 nM small interfering siRNA pools using DharmaFECT 1 (Horizon Discovery) in M105:M199 media without FBS. Each pool contained four siRNA sequences for *ADAM15* (L-004505-00), *ADIRF* (L-012306-01), *ALDH2* (L-009766-00), *CLIC3* (L-011805-00), *CRYAB* (L-009743-00), *FOLR1* (L-010403-00), *FOSL2* (L-004110-00), *HTRA1* (L-006009-00), *LGALS1* (L-011718-00), *MFGE8* (L-021466-00), *MSLN* (L-006346-00), *NDRG1* (L-010563-00), *NNMT* (L-010351-00), *NOVA2* (L-012590-00), *POSTN* (L-020118-00), *RFTN1* (L-023452-02), *SHMT1* (L-004617-00), *SNCG* (L-011396-00), *SPCARC* (L-003710-00), *TAGLN* (L-003714-00 ), *THBS1* (L-015337-00), *TNS1* (L-009976-00), ON-TARGET plus Non-targeting siRNA #1 (D-001810-01), or siGLO RISC-Free Control regent (Horizon Discovery; Catalog number; D-001600-01-05).

### Inhibitor testing

The MET inhibitors, Savolitinib (HY-15959), Icotinib (HY-15164A) and Cabozantinib (YT-13016) and the EGFR inhibitors, Lapatinib (HY-50898) and Erlotinib (HY-50896) were purchased from MedChem Express. The FOLR1-targeted antibody drug target, Mirvetuximab was purchased from ImmunoGen, Inc. The CDK inhibitors, Abemaciclib (LY2835219) and Milciclib (PHA-848125) were purchased from SelleckChem. The inhibitors were tested in proliferation and invasion assays using IC50 doses.

### Quantitative PCR with reverse transcription (RT-qPCR)

RT-qPCR was conducted by StepOnePlus Real-Time PCR System (Applied Biosystems). Total RNA was extracted using TRIzol (Invitrogen) and transcribed into cDNA using TaqMan RNA Reverse Transcription Kit (Applied Biosystems). A TaqMan endogenous control was used to normalize mRNA expression. Each PCR assay was performed in triplicate, and relative levels of *NOVA2* expression were calculated using the $2^{-\Delta\Delta}C_t$ method.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

## Functional assays

### Migration/siRNA screen

Ovarian cancer cells were plated in 96-well plates (17,000 cells/well) and reverse transfected with target or control siRNA. The wells were washed thoroughly with phosphate buffered saline (PBS) to remove detached cells and growth media added to each well before scratching with a wound making machine. The cells were monitored on the IncuCyte® Live Cell Analysis System (Sartorius), with the wells imaged every hour to evaluate migration.

### Proliferation

Ovarian cancer cells were plated onto 96-well plates (10,000 cells/well) and incubated for 24 hours to allow cells to adhere before media changed. For siRNA testing, the cells were pre-transfected with target or control siRNA 24 hours prior to assay and treated with growth media. For inhibitor testing, the cells were treated with IC50 concentration of inhibitors in growth media. For conditioned media testing, 24h conditioned media was collected from primary human CAFs that stably expressednon-targeting shCtrl (5′-GCAGTTATCTGGAAGATCAGG-3′) or shNNMT (5′-GCTACACAATCGAATGGTT-3′) as previously described27. The LGSC cells were treated, incubated at 37°C with 5% CO2, and monitored on the IncuCyte® Live Cell Analysis System (Sartorius), with the wells imaged every 4-12 hours to evaluate proliferation.

### Invasion

Ovarian cancer cells were plated in the top well of the QCMTM 96-Well Cell Invasion Assay plate (20,000 cells/well). For siRNA testing, the cells were pre-transfected with target or control siRNA 24 hours prior to assay and plated in serum-free media. For inhibitor testing, the cells were treated with IC50 concentration of inhibitors in serum-free media. Growth media was placed in the bottom chamber for the siRNA and inhibitor testing. For conditioned media testing, conditioned media was collected from primary human CAFs that stably expressed shCtrl or shNNMT as described above. The LGSC cells were plated in CAF growth media (control media), and in the bottom chamber CAF control media, shCtrl-expressing CAF conditioned media, or shNNMT-expressing CAF conditioned media, was added. The cells were incubated for 48 hours to allow the cells to invade through the chamber. The invaded cells were detached, lysed, and stained with CyQuant GR Dye according to manufacturer instructions. Total fluorescence (480/520 nm) was acquired on a fluorescent plate reader (SpectraMax iD5).

## Western blot analyses

LGSC cells were plated at a concentration of $4 \times 10^5$ cells per well to a 6-well plate and cultured for 24-48 hours. The cells were then lysed with an SDS lysis buffer containing 4% SDS and 10 mM HEPES, at a pH of 8.5. Next, 15-25 µg of each cell lysate was separated by SDS-PAGE using 5–20% gel (BioRad) and transferred to a nitrocellulose membrane (Thermo Fisher Scientific). The membrane was blocked with 5% non-fat dry milk (NFDM) (Lab Scientific) in Tris-buffered saline with Tween 20 (TBST) for 1 hour at room temperature The membranes were incubated with primary antibodies against NOVA2 (1:1000) and GAPDH (1:1000) (Proteintech) in 2% bovine serum albumin (Sigma-Aldrich) in TBST at 4 °C overnight. After washing with TBST, the membrane was incubated with secondary antibodies (Cell Signaling) conjugated to horseradish peroxidase at 1:2000 dilution in 5 % NFDM/TBST for 1 hour at room temperature. The proteins were visualized with Clarity Western ECL Substrate (Bio-Rad) or SuperSignal West Femto Substrate (Thermo Fisher Scientific) under ChemiDoc XRS+ System (BIO-RAD). Full-length immunoblots are available in Supplementary Data.

## In vivo xenograft model

*Development of LGSC xenograft model*. Female HSD:Athymic Nude-Foxn1$^{nu}$ (athymic nude; #069(nu)/070(nu/+) mice at age 5-6 weeks and approximately 20 grams were purchased from Charles River. All procedures involving animal care were approved by the Committee on Animal Care at the University of Chicago. Mice were irradiated and injected intra-peritoneally (i.p.) with 5

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

million VOA6406 cells. The mice were sacrificed 12 weeks post cancer cell injection. The omental tumors were collected, minced into 1mm$^3$ pieces, and digested with 2.5U/ml dispase II (17105041, Gibco) at 37 °C and 5% CO2 for 30 minutes. Single cells were collected by 70 µm filtration, washed twice in PBS, and plated in growth media. After a 30-minute incubation, the unadhered cells were collected and plated in a new flask. These mouse tumor derived cells, VOA6406ip1, were passaged 10 times and validated using short tandem repeat DNA fingerprinting with the AmpFℓSTR Identifier kit and compared with known fingerprints by IDEXX BioAnalytics Laboratories (Columbia, MO).

*NNMTi intervention study experiment.* Five million VOA6406ip1 cells were subcutaneously injected with 50 µl of Matrigel (354262 Corning) into 6-week-old athymic nude mice. The mice that developed subcutaneous tumors were randomized. For treatments, 100 µg NNMTi, 100 µg of the less active enantiomer of the NNMTi, or no drug (vehicle control) in 50 µl PBS was mixed with 50 µl of alginate (final concentration: 10mg/kg alginic acid in PBS containing 10% DMSO). These three agents were administrated via intratumor injection biweekly starting 11 days post cancer cell injection. Tumor volume was measured once a week. All mice were sacrificed after 4 weeks of treatment.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

## References

1 Gershenson, D. M. *et al.* Recurrent low-grade serous ovarian carcinoma is relatively chemoresistant. *Gynecol Oncol* **114**, 48-52 (2009).
2 Gershenson, D. M. *et al.* Hormonal Maintenance Therapy for Women With Low-Grade Serous Cancer of the Ovary or Peritoneum. *J. Clin. Oncol.*, JCO2016710632 (2017).
3 Tang, M. *et al.* PARAGON: A Phase II study of anastrozole in patients with estrogen receptor-positive recurrent/metastatic low-grade ovarian cancers and serous borderline ovarian tumors. *Gynecol. Oncol.* **154**, 531-538 (2019).
4 Singer, G. *et al.* Patterns of p53 mutations separate ovarian serous borderline tumors and low- and high-grade carcinomas and provide support for a new model of ovarian carcinogenesis: A mutational analysis with immunohistochemical correlation. *American Journal of Surgical Pathology* **29**, 218-224 (2005).
5 Mund, A. *et al.* Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* (2022).
6 Ho, C. L., Kurman, R. J., Dehari, R., Wang, T. L. & Shih Ie, M. Mutations of BRAF and KRAS precede the development of ovarian serous borderline tumors. *Cancer Res* **64**, 6915-6918 (2004).
7 Anglesio, M. S. *et al.* Mutation of ERBB2 provides a novel alternative mechanism for the ubiquitous activation of RAS-MAPK in ovarian serous low malignant potential tumors. *Mol Cancer Res* **6**, 1678-1690 (2008).
8 Manning-Geist, B. *et al.* MAPK Pathway Genetic Alterations Are Associated with Prolonged Overall Survival in Low-Grade Serous Ovarian Carcinoma. *Clin Cancer Res* **28**, 4456-4465 (2022).
9 Shih, I. M., Wang, Y. & Wang, T. L. The Origin of Ovarian Cancer Species and Precancerous Landscape. *Am J Pathol* **191**, 26-39 (2021).
10 Silva, E. G. *et al.* Precursors in the ovarian stroma: another pathway to explain the origin of ovarian serous neoplasms. *Hum Pathol* **127**, 136-145 (2022).
11 Vang, R. *et al.* Long-term Behavior of Serous Borderline Tumors Subdivided Into Atypical Proliferative Tumors and Noninvasive Low-grade Carcinomas: A Population-based Clinicopathologic Study of 942 Cases. *Am J Surg Pathol* **41**, 725-737 (2017).
12 Lengyel, E. Ovarian cancer development and metastasis. *Am J Pathol* **177**, 1053-1064 (2010).
13 Hollandi, R. *et al.* nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer. *Cell Syst* **10**, 453-458 e456 (2020).
14 Mund, A. *et al.* Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat Biotechnol* **40**, 1231-1240 (2022).
15 Hunter, S. M. *et al.* Molecular profiling of low grade serous ovarian tumours identifies novel candidate driver genes. *Oncotarget* **6**, 37663-37677 (2015).
16 King, E. R. *et al.* The anterior gradient homolog 3 (AGR3) gene is associated with differentiation and survival in ovarian cancer. *Am J Surg Pathol* **35**, 904-912 (2011).
17 Kimura, I., Konishi, M., Miyake, A., Fujimoto, M. & Itoh, N. Neudesin, a secreted factor, promotes neural cell proliferation and neuronal differentiation in mouse neural precursor cells. *J Neurosci Res* **83**, 1415-1424 (2006).
18 Scott, B. J., Qutob, S., Liu, Q. Y. & Ng, C. E. APM2 is a novel mediator of cisplatin resistance in a variety of cancer cell types regardless of p53 or MMR status. *Int J Cancer* **125**, 1193-1204 (2009).
19 Matulonis, U. A. *et al.* Efficacy and Safety of Mirvetuximab Soravtansine in Patients With Platinum-Resistant Ovarian Cancer With High Folate Receptor Alpha Expression: Results From the SORAYA Study. *J Clin Oncol* **41**, 2436-2445 (2023).
20 Gershenson, D. M. *et al.* The genomic landscape of low-grade serous ovarian/peritoneal carcinoma and its impact on clinical outcomes. *Gynecol Oncol* **165**, 560-567 (2022).
21 Xu, M. *et al.* ADAM17 promotes epithelial-mesenchymal transition via TGF-beta/Smad pathway in gastric carcinoma cells. *Int J Oncol* **49**, 2520-2528 (2016).
22 Cheasley, D. *et al.* Genomic analysis of low-grade serous ovarian carcinoma to identify key drivers and therapeutic vulnerabilities. *J Pathol* **253**, 41-54 (2021).
23 Llaurado Fernandez, M. *et al.* Hormone receptor expression and outcomes in low-grade serous ovarian carcinoma. *Gynecol Oncol* **157**, 12-20 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

24 Lengyel, E. *et al.* A molecular atlas of the human postmenopausal fallopian tube and ovary from single-cell RNA and ATAC sequencing. *Cell Rep* **41**, 111838 (2022).
25 Saito, Y. *et al.* Differential NOVA2-Mediated Splicing in Excitatory and Inhibitory Neurons Regulates Cortical Development and Cerebellar Function. *Neuron* **101**, 707-720 e705 (2019).
26 Wang, Z. & Dong, C. Gluconeogenesis in Cancer: Function and Regulation of PEPCK, FBPase, and G6Pase. *Trends Cancer* **5**, 30-45 (2019).
27 Eckert, M. A. *et al.* Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts. *Nature* **569**, 723-728 (2019).
28 Akhoundova, D. & Rubin, M. A. Clinical application of advanced multi-omics tumor profiling: Shaping precision oncology of the future. *Cancer Cell* **40**, 920-938 (2022).
29 White, N. M. *et al.* KLK6 and KLK13 predict tumor recurrence in epithelial ovarian carcinoma. *Br J Cancer* **101**, 1107-1113 (2009).
30 Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* **49**, 925-934 (2017).
31 de Visser, K. E. & Joyce, J. A. The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. *Cancer Cell* **41**, 374-403 (2023).
32 Shi, H., Chai, P., Jia, R. & Fan, X. Novel insight into the regulatory roles of diverse RNA modifications: Re-defining the bridge between transcription and translation. *Mol Cancer* **19**, 78 (2020).
33 Zhao, M., Kim, P., Mitra, R., Zhao, J. & Zhao, Z. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res* **44**, D1023-1031 (2016).
34 Nordmann, T. M. *et al.* A standardized and reproducible workflow for membrane glass slides in routine histology and spatial proteomics. *Mol Cell Proteomics*, 100643 (2023).
35 Longacre, T. A., McKenney, J. K., Tazelaar, H. D., Kempson, R. L. & Hendrickson, M. R. Ovarian serous tumors of low malignant potential (borderline tumors): outcome-based study of 276 patients with long-term (> or =5-year) follow-up. *Am J Surg Pathol* **29**, 707-723 (2005).
36 Chew, G. L. *et al.* DUX4 Suppresses MHC Class I to Promote Cancer Immune Evasion and Resistance to Checkpoint Blockade. *Dev. Cell* **50**, 658-671 e657 (2019).
37 Etemadmoghadam, D. *et al.* EIF1AX and NRAS Mutations Co-occur and Cooperate in Low-Grade Serous Ovarian Carcinomas. *Cancer Res* **77**, 4268-4278 (2017).
38 Dhillon, A. S., Hagan, S., Rath, O. & Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **26**, 3279-3290 (2007).
39 Naba, A. *et al.* Characterization of the extracellular matrix of normal and diseased tissues using proteomics. *J. Proteome Res.* **16**, 3083-3091 (2017).
40 Risom, T. *et al.* Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. *Cell* **185**, 299-310 e218 (2022).
41 Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* **50**, D543-D552 (2022).
42 Sawada, K. *et al.* c-Met overexpression is a prognostic factor in ovarian cancer and an effective target for inhibition of peritoneal dissemination and invasion. *Cancer Res* **67**, 1670-1679 (2007).
43 Muller-Reif, J. B. *et al.* A New Parallel High-Pressure Packing System Enables Rapid Multiplexed Production of Capillary Columns. *Mol Cell Proteomics* **20**, 100082 (2021).
44 Meier, F. *et al.* diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat Methods* **17**, 1229-1236 (2020).
45 Brunner, A. D. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol Syst Biol* **18**, e10798 (2022).
46 Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* **17**, 41-44 (2020).
47 Demichev, V. *et al.* dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nature Communications* **13**, 3944 (2022).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

## Main Manuscript Figures

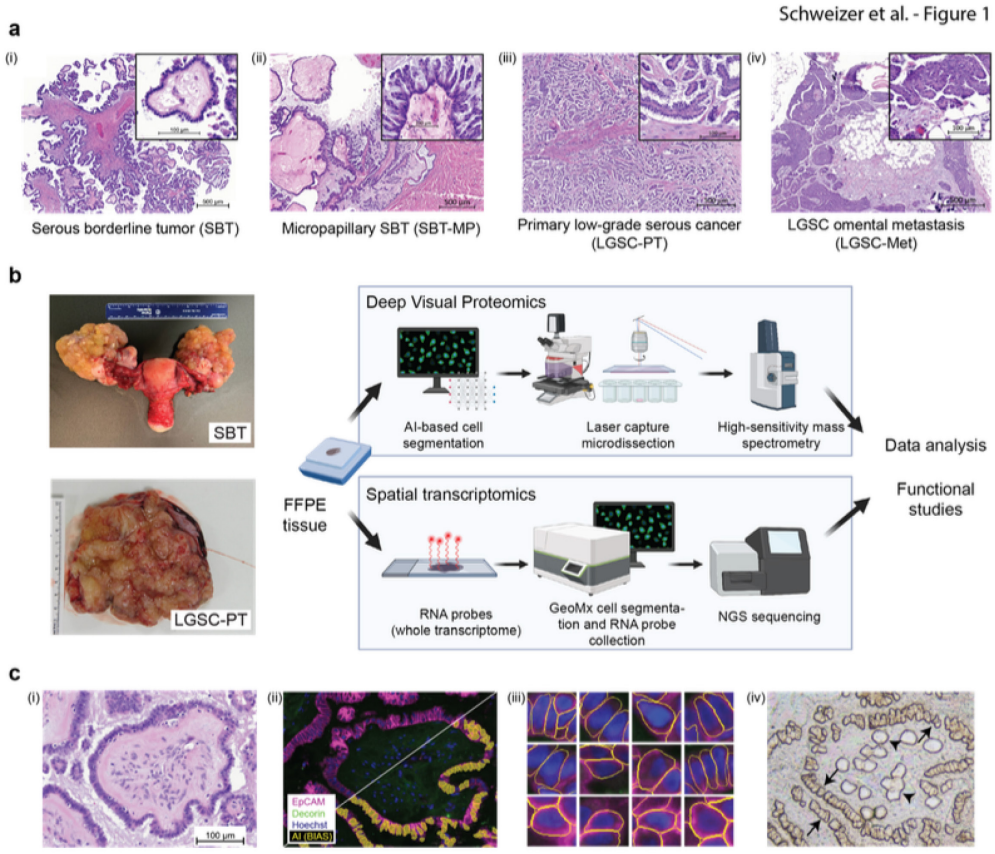



**Fig. 1 | Characterization of serous borderline and low-grade ovarian cancer**
**a)** Representative H&E images of the putative transformation sequence from (i) serous borderline tumor (SBT, not invasive) via (ii) micropapillary lesions (SBT-MP) to (iii) low-grade serous cancer-primary tumor (LGSC-PT), and (iv) metastasis (LGSC-Met). The papillary architecture with hierarchical branching pattern is characteristic of SBT. **b)** Gross pathology of a bilateral borderline tumor and low-grade serous cancer. SBT often present as bilateral adnexal tumors (left). Experimental strategy, bioinformatics and functional studies (right).
**c)** AI-based cell recognition and laser-based dissection in DVP. (i) H&E of an SBT. (ii, iii) Immunofluorescence. Staining for malignant epithelial cells (EpCAM, purple) and stromal (decorin, green) as well as artificial intelligence (AI)-based recognition of tumor cells (yellow – below diagonal line) using the BIAS software. (iii) Morphology of single epithelial cells as recognized by artificial intelligence. (iv) Brightfield image of the same sample in i and ii showing the tissue after single cell extraction. Microdissected epithelial cells (arrows) and stroma (arrowheads).

Serous borderline tumor (SBT), micropapillary SBT (MP-SBT), low-grade serous cancer primary tumor (LGSC-PT) and corresponding metastases (LGSC-Met).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

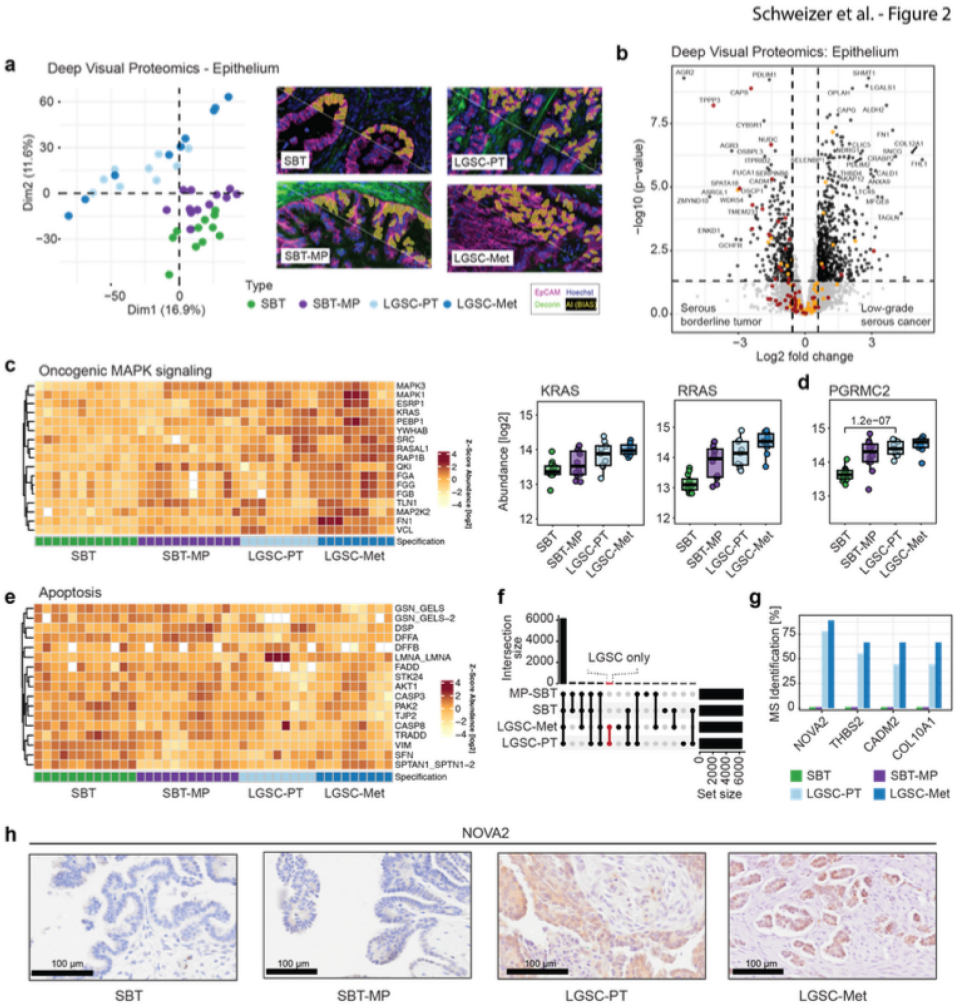



**Fig. 2 | Deep Visual Proteomics on the epithelial tumor compartment confirms known and identifies novel pathways in transition of SBT to LGSC**

**a)** Principal Component Analysis (PCA) for the epithelium clearly separates serous borderline tumors, serous borderline tumors with a micropapillary pattern, invasive low grade serous cancer, and corresponding metastasis. This transition is evident in the diagonal of PC1 and PC2 from lower right to upper left. AI-based recognition of epithelial cells using immunofluorescence (EpCAM-purple, decorin-green) below the white diagonal line, followed by AI segmentation (yellow).

**b)** Volcano plot of the differential epithelial protein expression between SBT and LGSC-PT in the epithelial compartment. A fold change cutoff of 1.5 and a q-value cutoff of 0.05 are indicated by vertical and horizontal lines, respectively. Proteins matching the significance for differential regulation (DR) criteria are highlighted in black, markers of secretory cells in orange and ciliated cells in red.

**c)** Proteins of the mitogen-activated protein kinase (MAPK)-signaling pathway show a gradual increase towards LGSC and corresponding metastasis (Heatmap). Commonly altered Ras and Ras-regulating proteins (box plots).

**d)** Boxplots of significantly changed membrane-associated progesterone receptor component 2 (PGRMC2) between the four groups (Student's T-test).

**e)** Heatmap. Proteins involved in apoptosis show reduced abundance from SBT to LGSC- Met.

**f)** Upset plot. Comparison of MS-detected peptides/proteins detected in specific groups but completely absent in others and therefore not included in Fig. 1e and 2b (methods). The set size is the number of identified proteins, while the intersection size shows the number of overlapping proteins.

**g)** Bar plot of the protein subset highlighted in (f). Bars present the percentage of samples in which the four most frequent proteins were identified per group using mass spectrometry. NOVA2 was solely detectable in more than 75% of LGSC-PT and LGSC-Met, but not in SBT and SBT-MP.

**h)** Immunohistochemistry for NOVA2.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

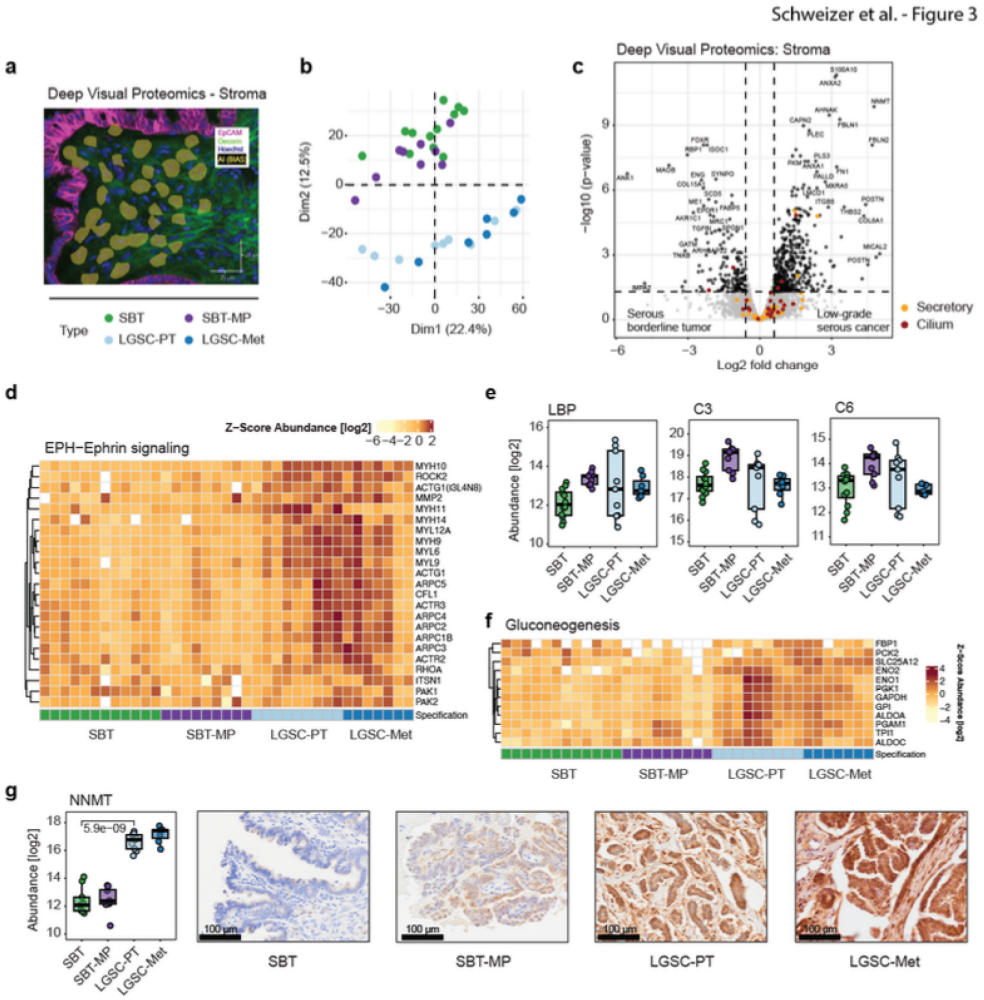



**Fig. 3 | Deep Visual Proteomics of the stromal compartment uncovers a bimodality in the transition of SBT to LGST**

**a)** Immunofluorescence outlining the extraction of cell equivalents from the stroma (EpCAM-purple, decorin-green, AI-segmentation - yellow).
**b)** Principal Component Analysis comparing stromal protein expression shows the separation of serous borderline and micropapillary tumors from invasive low grade serous cancer and corresponding metastases. The variability was most evident in dimension 2 (12.5%).
**c)** Volcano plot of differential stroma protein expression between SBT and LGSC-PT in the epithelial compartment. A fold change cutoff of 1.5 and a q-value cutoff of 0.05 are indicated by vertical and horizontal lines, respectively. Proteins matching the significance for differential regulation (DR) criteria are highlighted in black, markers of secretory in orange and ciliated cells in red.
**d)** Proteins that are involved in actin-rearrangement suggested by ephrin (EPH) signaling show an increase towards LGSC and corresponding metastasis in the stromal compartment.
**e)** Boxplots of proteins involved in acute inflammation and the complement system across the transition.
**f)** Proteins involved in gluconeogenesis show an increase towards LGSC and corresponding metastasis.
**g)** NNMT protein abundance (left) and immunohistochemistry (right).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

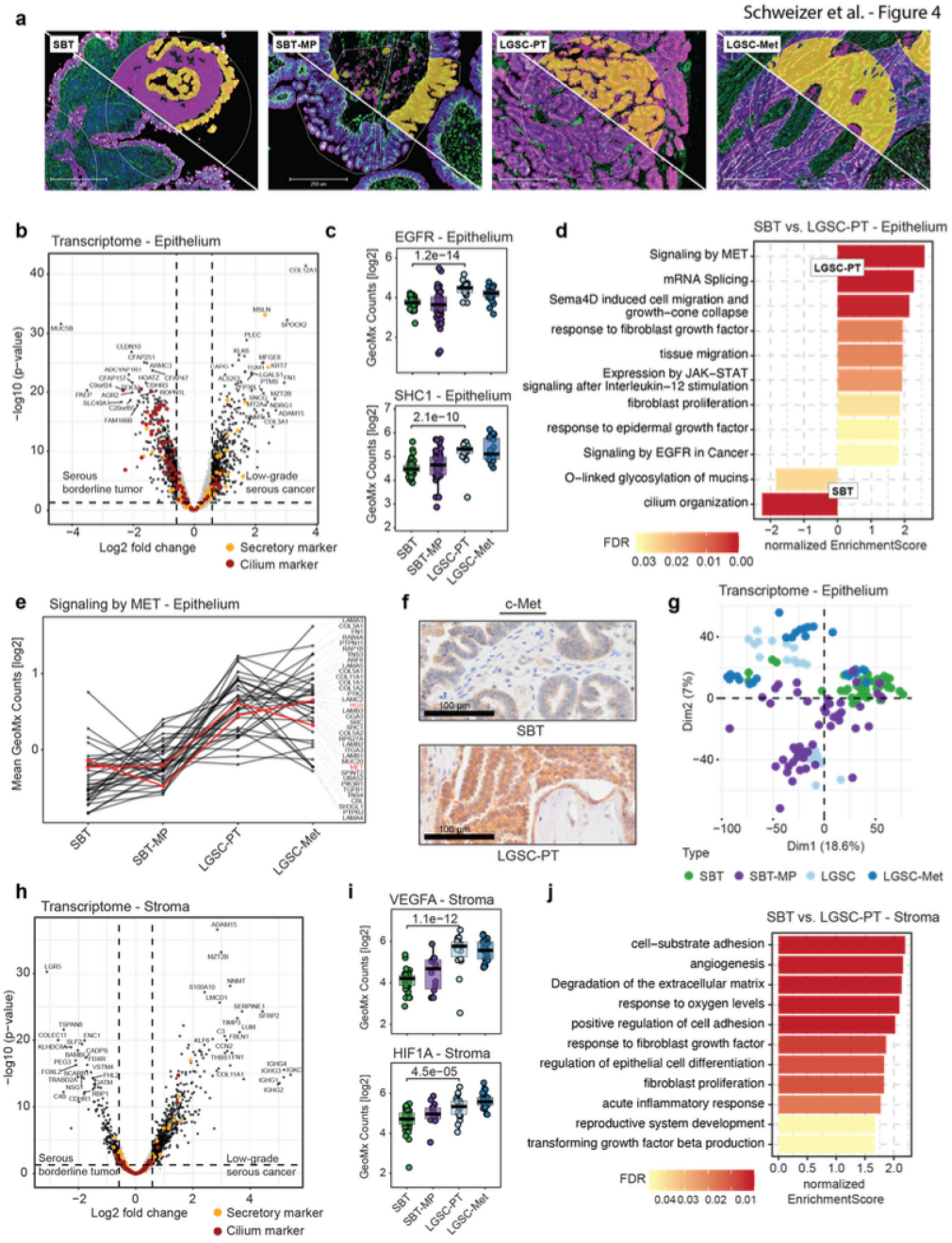



**Fig. 4 | Spatial transcriptomics of SBT and LGSC**
**a)** Immunofluorescence for pan-cytokeratin (purple), decorin (green), and nuclei (blue) for exemplary regions. Tumor and stroma compartments for subsequent UV illumination are shown above the white diagonal line in yellow and magenta, respectively. Regions of interest are outlined with fine white lines.
**b)** Epithelium. Differential transcript abundance of borderline versus low-grade serous cancer (Volcano plot). Protein markers for ciliated and secretory cells are highlighted in orange and red, respectively.
**c)** GeoMx counts for EGFR and SHC1 transcripts across the progression series (Student's t-test).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

**d)** GSEA biological pathway enrichment analysis based on the spatial transcript results comparing SBT and LGSC-PT in the epithelial compartment (Pathway REACTOME, Gene Ontology Biological Processes) on the comparison in (b).
**e)** Profile plot of pathway-associated proteins determined in (d) for 'Signaling by MET'. Proteins with critical roles in the pathway (MET, HGS) are annotated in red. All other proteins are summarized in Supplementary Table 8.
**f)** IHC of c-MET in SBT and LGSC-PT.
**g)** Nanostring Principal Component Analysis using transcripts in the epithelium for the indicated histologies.
**h)** Stroma. Differential transcript abundance of SBT versus LGSC-PT. Protein markers for ciliated and secretory cells are highlighted in orange and red, respectively.
**i)** GeoMx counts for VEGFA and HIF1α across the progression series (Student's t-test).
**j)** GSEA biological pathway enrichment analysis using the spatial transcript results in SBT and LGSC-PT in the stromal compartment (Pathway REACTOME, Gene Ontology Biological Processes) on the comparison in (h).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Schweizer et al. - Figure 5

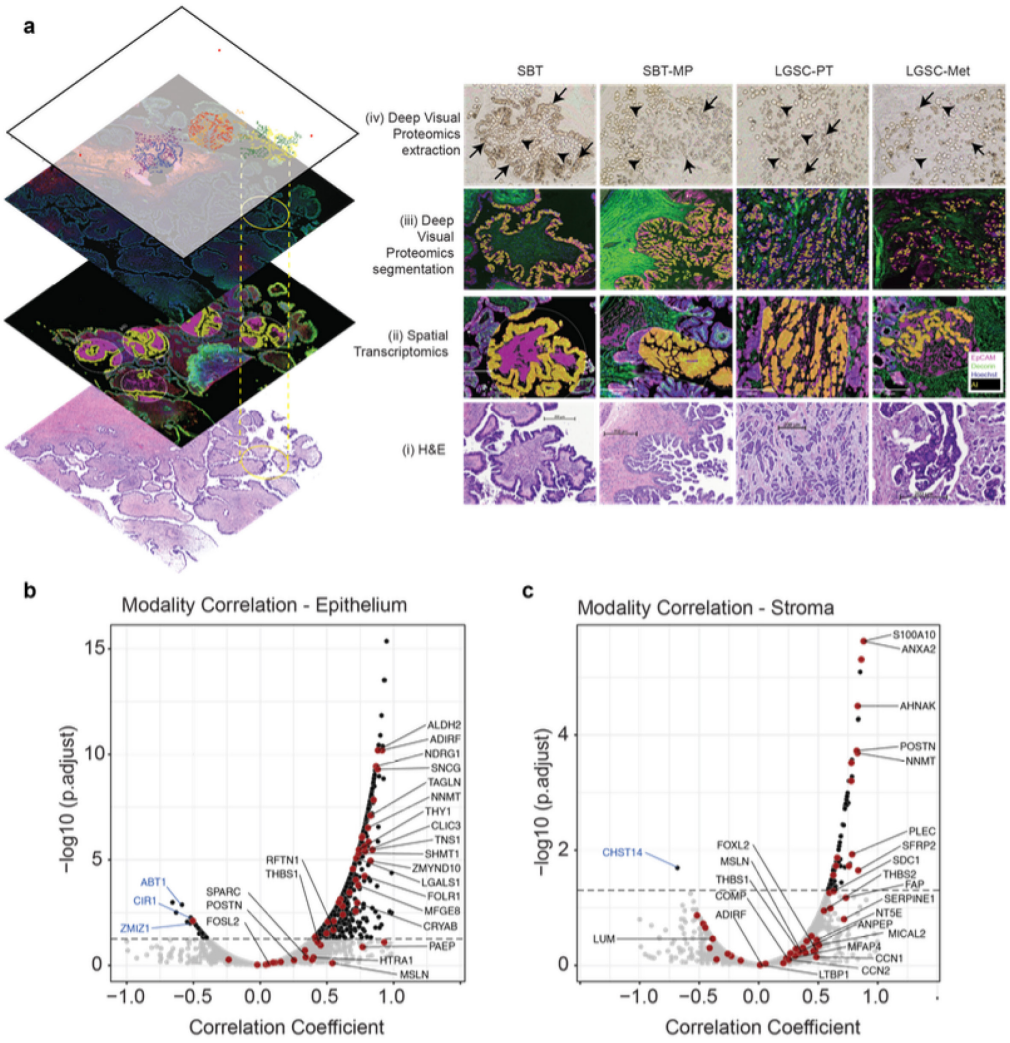



**Fig. 5 | Integration of spatial transcriptomics and proteomics**
**a)** Multi-layer integration of Deep Visual Proteomics and spatial transcriptomics for the four histologic subtypes. (i) H&E, (ii) spatial transcriptomics regions of interest, (iii) Deep Visual Proteomics including AI based cell recognition/segmentation, and (iv) brightfield image after DVP laser microdissection. Immunofluorescence showing malignant epithelial cells (EpCAM, purple) and stroma (decorin, green) in both spatial proteomics and transcriptomics (ii, iii). AI-based recognition in the DVP or Nanostring technology is shown in yellow (tumor compartment) and magenta (stromal compartment in spatial transcriptomics), respectively. Microdissected epithelial cells (black arrows) and stroma (arrowheads) (iv). ROIs of spatial transcriptomics matched to the previously selected regions. The top layer of the visual integration (left panel) shows the laser microdissected cells in different regions for the epithelium (blue/yellow/red) and in the stroma (purple/orange/green) used in DVP.
**b, c)** Correlation of protein and transcript expression comparing spatial proteomics and transcriptomics. Genes that are significantly correlated are in black and are above the dashed line (Pearson coefficient ≥ 0.05). A negative correlation coefficient indicates opposite trends of protein and transcript expression. Anticorrelating genes discussed in the text are annotated in blue. Significant genes among the set of 70 of biological interest are highlighted in red and annotated in black (see also Extended Data 12).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

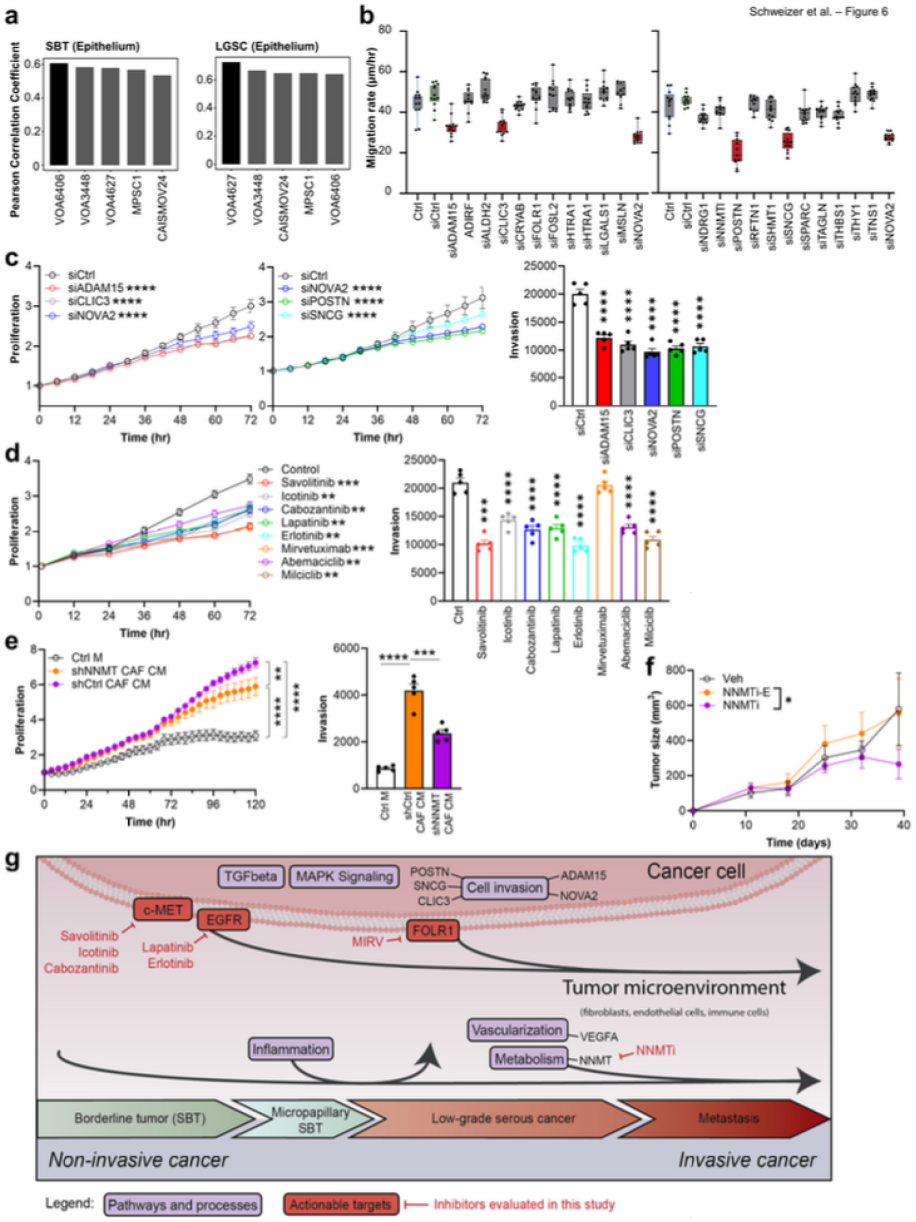



**Fig. 6 | Functional studies characterizing the significant 'omics' genes and pathways in the transitions of SBT to LGSC**

**a)** Pearson coefficient. Correlation analysis of protein expression (DVP) from SBT (left) and LGSC (right) human epithelial tissue compared to epithelial LGSC cell lines. Cell lines of highest similarity are highlighted in black.
**b)** siRNA screening of 23 priority genes (Fig. 5) using the LGSC cell line VOA4627 measuring cell migration. Box plots +/- standard deviation.
**c)** Validation of the most promising siRNA hits in the LGSC cell line VOA4627 measuring proliferation (left) and invasion (right).
**d)** Inhibitor testing in VOA4627 proliferation (left) and invasion (right).
**e)** VOA4627 cells were treated with either condition media from immortalized human CAFs where NNMT was inhibited using shRNA (shNNMT CAF CM), shRNA control transfected CAF condition media (shCtrl CAF CM) or control media (Ctrl M).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.13.23298409; this version posted November 13, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

**f)** Nude mice were injected intraperitoneal with the VOA6406ip1 cell line at day 0. Treatment of mice started at day 11 and was continued daily with either the NNMTi, the inactive enantiomer or vehicle control biweekly and tumor volume measured from 11-40 days after cancer cell injection.
All growth curves and bar graphs show mean +/- SEM. The data presented in c-e was repeated in 3 independent experiments. Significance levels by p-value: * 0.05, ** 0.01, ***, 0.001, *** 0.0001.
**g)** Schematic summary on pathways and distinct proteins/transcripts involved in the transition from SBT to LGSC.

## Article 10. Single-cell spatial proteomics identifies the JAK/STAT pathway as an actionable therapeutic target in lethal cutaneous drug reactions.

Preprint, *medRxiv*, 2023 (Nov, 12); DOI: 10.1101/2023.11.11.23295492

Thierry M. Nordmann[1,2,*], Holly Anderton[3,4], Akito Hasegawa[5], Lisa Schweizer[1], Peng Zhang[6], Pia-Charlotte Stadler[1,7], Ankit Sinha[1], Andreas Metousis[1], Florian A. Rosenberger[1], Maximillian Zwiebel[1], Takashi K. Satoh[7], Florian Anzengruber[2,8], Maria C. Tanzer[3,4,9], Yuki Saito[5], Ting Gong[6], Marvin Thielert[1], Haruna Kimura[5], Natasha Silke[3,4], Edwin H. Rodriguez[1], Restivo Gaetana[2], Hong Ha Nguyen[5], Annette Gross[10], Mitchell P. Levesque[2], Peter J. Murray[10], Saskia Ingen-Housz-Oro[11], Andreas Mund[12], Riichiro Abe[5], John Silke[3,4], Chao Ji[6,13,*], Lars E. French[7,14,*], Matthias Mann[1,*]

*[1]Department of Proteomics and Signal Transduction; Max Planck Institute of Biochemistry; Martinsried, Bavaria, 82152; Germany. [2]Department of Dermatology, University Hospital Zurich, University of Zurich, Zurich; Switzerland. [3]Inflammation division, Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia. [4]Department of Medical Biology, University of Melbourne, Parkville, VIC 3052, Australia. [5]Division of Dermatology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan. [6]Department of Dermatology, The First Affiliated Hospital of Fujian Medical University, Fuzhou, Fujian, China. [7]Department of Dermatology and Allergy, University Hospital, Ludwig Maximilian University (LMU) Munich, Germany. [8]Department of Internal Medicine, Division of Dermatology, Cantonal Hospital Graubuenden, Chur, Switzerland. [9]Advanced Technology and Biology division, Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia. [10]Immunoregulation Research Group, Max Planck Institute of Biochemistry, Martinsried, Germany. [11]Dermatology Department, AP-HP, Henri Mondor Hospital, Créteil, France. [12]Proteomics Program, The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Faculty of Health and Medical Sciences, Blegdamsvej 3B, 2200 Copenhagen, Denmark. [13]Key Laboratory of Skin Cancer of Fujian Higher Education Institutions, Fujian Medical University, Fuzhou, Fujian, China. [14]Dr. Philip Frost, Department of Dermatology and Cutaneous Surgery, University of Miami Miller School of Medicine, Miami, FL.*
**Corresponding authors.*

Following the example of cancer biology, we next investigated fatal drug-induced skin reactions. These diseases range from self-resolving conditions to severe states, specifically drug reaction with eosinophilia and systemic symptoms (DRESS) and toxic epidermal necrolysis (TEN). These states can be triggered by routine medication and patients may experience sudden epidermal detachment. As no effective therapy currently exists, supportive care remained the consensus treatment.

We employed DVP to examine the molecular mechanisms occurring within keratinocytes and immune cells (CD45 positive) at both spatial and cell type levels. As anticipated, DVP successfully differentiated between keratinocytes and immune cells upon proteomic analysis. Keratinocytes separated by the different disease states with

STAT3 emerging as one of the main drivers along with a strong interferon signature. DRESS exhibited upregulated MHC class-I proteins, suggesting a participation of keratinocytes in antigen presentation. In the case of immune cells, TEN again revealed a dominant interferon signature with STAT1 as a principal driver. However, DRESS was the most distinguishable disease state, with its drivers of signals from viral pathways, in contrast to TEN.

Together, the interferon signature was the most prominent feature in TEN in both cell types, and thus we reasoned that activation of the JAK/STAT pathway is one of the main drivers. Given the strength of this signature and the existence of FDA-approved drugs modulating the JAK/STAT pathway, we focused our efforts on finding a therapeutic strategy on this particular pathway. Note, however, that our data suggests alternative treatment strategies such as the inhibition of EZH2 or IDO1, which are associated with reducing cytotoxic reactions in the tumor microenvironment. Both proteins have been subjects of FDR-approved cancer treatments, though their involvement in inflammatory or allergic diseases has not been shown yet.

To establish clinical relevance, we pioneered a novel *in-vitro* model for cutaneous drug reactions by co-culturing primary keratinocytes taken from TEN survivors with activated patient-matched peripheral blood mononuclear cells (PBMCs). Keratinocytes were killed by activated PBMCs within three days. However, treatment with the FDA and EMA approved pan-JAK inhibitor tofacitinib resolved keratinocyte cytotoxicity *in vitro*. These positive results from our *in-vitro* model were also supported by the treatment of two independent and established mouse models of TEN, with normalization of dermal thickness upon treatment.

Due to these promising results and the urgent clinical need, our collaborators treated four human patients with JAK1 inhibitor rescue therapy with abrocitinib for ten days. In all patients, disease progression stopped and their skin re-epithelialized resulting in full recovery from TEN. All treated patients displayed significant decrease in pSTAT1 levels after treatment. The fact that these patients not only survived but also had no side effects show the promise for efficacy of this treatment.

In summary, this study exemplifies the profound utility of spatial and clinical proteomics. Given the urgent clinical need and the availability of FDA-approved drugs, we navigated the journey from a biological question all the way to the treatment of the disease in humans.

***Contribution:***

Co-authorship. Thierry Nordmann, Lars E. French and Matthias Mann conceived the study. Thierry Nordmann conducted the project as first author. I contributed to the proteomic part of the study. I performed the MS measurements, provided feedback in the data interpretation and figure design. Additionally, I edited and revised the manuscript.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

# Single-cell spatial proteomics identifies the JAK/STAT pathway as an actionable therapeutic target in lethal cutaneous drug reactions

Thierry M. Nordmann[1,2]*, Holly Anderton[3,4], Akito Hasegawa[5], Lisa Schweizer[1], Peng Zhang[6], Pia-Charlotte Stadler[1,7], Ankit Sinha[1], Andreas Metousis[1], Florian A. Rosenberger[1], Maximillian Zwiebel[1], Takashi K. Satoh[7], Florian Anzengruber[2,8], Maria C. Tanzer[3,4,9], Yuki Saito[5], Ting Gong[6], Marvin Thielert[1], Haruna Kimura[5], Natasha Silke[3,4], Edwin H. Rodriguez[1], Restivo Gaetana[2], Hong Ha Nguyen[5], Annette Gross[10], Mitchell P. Levesque[2], Peter J. Murray[10], Saskia Ingen-Housz-Oro[11], Andreas Mund[12], Riichiro Abe[5], John Silke[3,4], Chao Ji[6,13]*, Lars E. French[7,14]*, Matthias Mann[1]*

1 Department of Proteomics and Signal Transduction; Max Planck Institute of Biochemistry; Martinsried, Bavaria, 82152; Germany
2 Department of Dermatology, University Hospital Zurich, University of Zurich, Zurich; Switzerland
3 Inflammation division, Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia.
4 Department of Medical Biology, University of Melbourne, Parkville, VIC 3052, Australia.
5 Division of Dermatology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan.
6 Department of Dermatology, The First Affiliated Hospital of Fujian Medical University, Fuzhou, Fujian, China.
7 Department of Dermatology and Allergy, University Hospital, Ludwig Maximilian University (LMU) Munich, Germany.
8 Department of Internal Medicine, Division of Dermatology, Cantonal Hospital Graubuenden, Chur, Switzerland.
9 Advanced Technology and Biology division, Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia.
10 Immunoregulation Research Group, Max Planck Institute of Biochemistry, Martinsried, Germany.
11 Dermatology Department, AP-HP, Henri Mondor Hospital, Créteil, France.
12 Proteomics Program, The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Faculty of Health and Medical Sciences, Blegdamsvej 3B, 2200 Copenhagen, Denmark.
13 Key Laboratory of Skin Cancer of Fujian Higher Education Institutions, Fujian Medical University, Fuzhou, Fujian, China.
14 Dr. Philip Frost, Department of Dermatology and Cutaneous Surgery, University of Miami Miller School of Medicine, Miami, FL.

*Corresponding authors. mmann@biochem.mpg.de; Lars.French@med.uni-muenchen.de; jichaofy@fjmu.edu.cn.; nordmann@biochem.mpg.de

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

## Abstract

**Toxic epidermal necrolysis (TEN) is a fatal drug-induced skin reaction and an emerging public health issue. Triggered by common medications, TEN patients undergo severe and sudden epidermal detachment caused by keratinocyte cell death. Although molecular mechanisms driving keratinocyte cell death have been proposed, the main drivers remain unknown and no effective therapy exists. To systematically map molecular changes that are associated with TEN and identify potential druggable targets, we employed the single-cell spatial proteomics technique Deep Visual Proteomics. We analyzed formalin-fixed paraffin-embedded archived skin-tissue biopsies of three types of cutaneous drug reactions with varying severity and quantified over 5,000 proteins in keratinocytes and skin-infiltrating immune cells. Most strikingly, this revealed a robust enrichment of Type-I and -II interferon signature in the immune cell and keratinocyte compartment of TEN patients, along with a drastic activation of pSTAT1. Targeted inhibition with pan-JAK inhibitor (JAKi) tofacitinib reduced keratinocyte-directed cytotoxicity in a novel live-cell imaging assay, using patient-derived keratinocytes and peripheral blood mononuclear cells (PBMCs). Furthermore, oral administration of pan-JAKi tofacitinib or baricitinib ameliorated clinical and histological disease severity in two distinct mouse models of TEN. Lastly, JAKi treatment was safe and associated with rapid cutaneous re-epithelialization and recovery in four patients with TEN. This study uncovers the JAK-STAT and interferon signaling pathways as key pathogenic drivers of TEN and demonstrates the potential of targeted JAK inhibition as a curative therapy.**

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

## Introduction

Cutaneous adverse drug reactions (CADRs) are the most common immune-mediated adverse drug reactions, with up to 2% of CADRs classified as severe and life-threatening[1]. The spectrum of CADRs ranges from self-resolving maculopapular rash (MPR) to rare and life-threatening conditions, such as drug reaction with eosinophilia and systemic symptoms (DRESS) and toxic epidermal necrolysis (TEN)[2]. DRESS is a severe multiorgan hypersensitivity reaction triggered by culprit drugs and sustained by the reactivation of latent viruses, leading to aberrant T-cell and eosinophilic response[3]. In comparison, TEN is characterized by fulminant epidermal detachment of over 30 % of the body surface area and is fatal in a third of all cases. TEN represents a considerable socio-economic burden due to specialized in-hospital care and lost productivity among survivors[4-6]. Despite diverse proposed pathogenic mechanisms[7-21], the main drivers of cytotoxicity remain elusive. Accordingly, consensus therapy of TEN is still supportive care[22].

Over the last 5 years, spatial omics has emerged as a powerful strategy to obtain a global, molecular view of intact tissue samples[23-25]. When applied at single-cell resolution, these techniques provide an opportunity to profile different cells without removing them from their native environment and thus provide a more physiologically relevant view of cellular function and dysfunction. However, although a broad variety of such technologies exist, translation into direct clinical benefit has remained largely aspirational. For example, application of spatial omics to TEN has been limited by the rarity of severe cutaneous drug reactions, and inability of existing techniques to explore different skin-residing cellular subsets especially in the context of clinically relevant formalin-fixed paraffin-embedded (FFPE) skin biopsies. To address these technical limitations, we have recently introduced Deep Visual Proteomics (DVP) as a novel spatial single-cell proteomics technology that combines the use of high content imaging, AI guided cell segmentation and classification, and laser micro dissection of individual target cells to achieve ultra-sensitive proteomics[26]. By using FFPE-archived biopsies, DVP enables high-yield, in-depth analysis of cells of interest, using easily accessible tissue specimens, even in rare disorders. Here, we employed DVP to study mechanisms of cutaneous drug reactions, investigate the predominant molecular features in the severe types of CADRs, and functionally validate targeted small molecule inhibitors for rapid therapeutic intervention in the most severe CADR, TEN.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

## Cohort assembly and spatial proteomics workflow

To study molecular mechanisms of cutaneous drug reactions, we assembled an age and sex matched, retrospective cohort of lesional FFPE skin biopsies from patients with mild (MPR) or severe (TEN, DRESS) CADRs alongside healthy controls (Fig. 1a and Extended Data Figure 1, N = 21). After clinical and histological re-validation of all cases, we stained 3 µm FFPE tissue sections for CD45 to visualize immune cells and with pan-cytokeratin for keratinocytes (Fig. 1b). A machine learning (ML) algorithm ensured proper segmentation and prevented downstream analysis of overlapping cellular compartments (Extended Data Fig. 1b). Thereafter, contours of immune cells and keratinocytes from each patient were laser micro-dissected and each separately pooled into multi-well plates for subsequent peptide extraction and ultra-high sensitivity mass spectrometry (MS)-data acquisition (Fig. 1b). This DVP approach achieved a proteomic depth of 4992 unique proteins in immune cells and 5009 in keratinocytes across all patients (Fig. 1c). The biological separation of keratinocytes from immune cells greatly exceeded interpatient heterogeneity (Fig. 1c and Extended Data Fig. 1c). A third of all identified proteins were differentially expressed, including the upregulation of LGALS7B and KRT5 in keratinocytes and of ANXA6 and CORO1A in immune cells, confirming the cell-type resolution of the DVP data (Extended Data Fig. 1d). Taken together, our cohort, which included 5-6 individuals per CADR indication, was amenable to the DVP workflow, resulting in robust and reproducible quantification of ~5,000 proteins and clear dissection of keratinocytes and immune cells.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

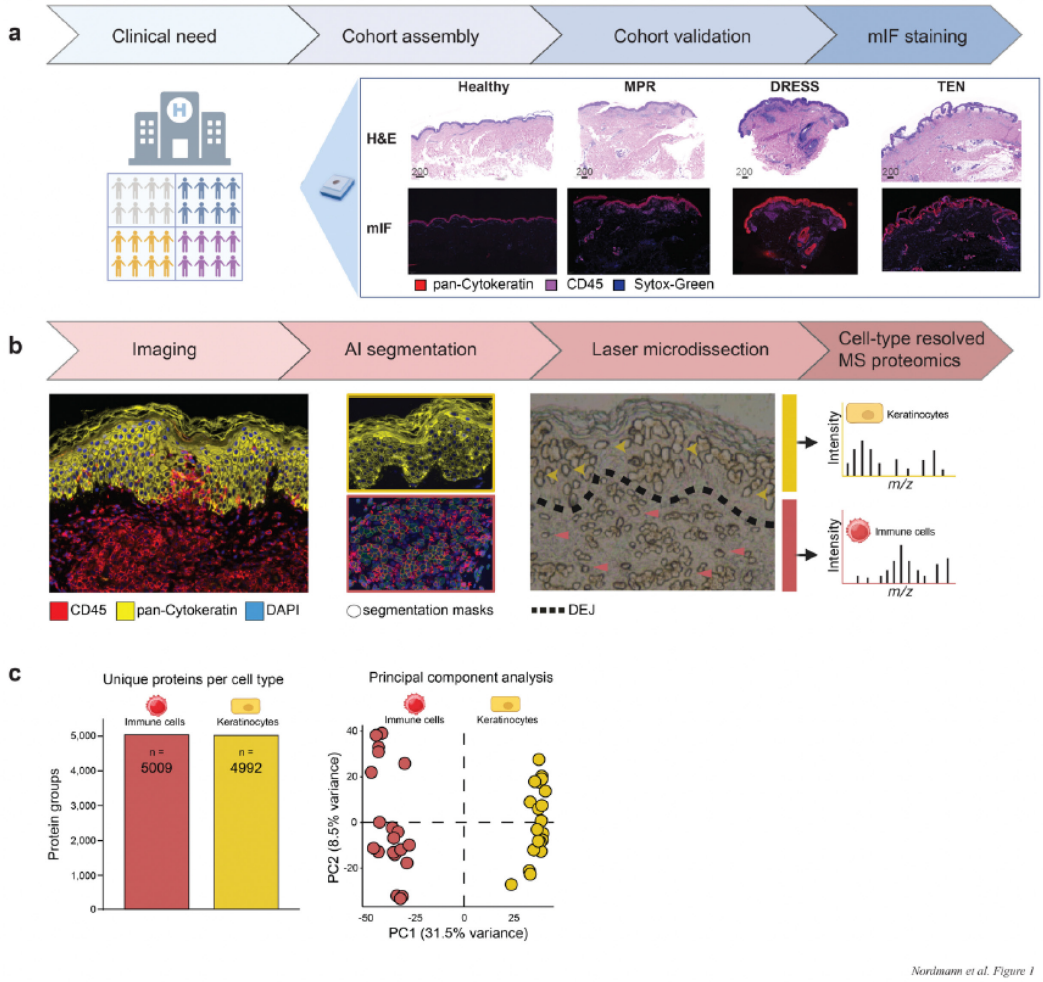



**Fig. 1 | Deep Visual Proteomic workflow for in-depth proteomic data with cellular resolution in cutaneous adverse drug reactions. a.** Cohort assembly and validation (N=21). Hematoxylin & Eosin (H&E) and multicolor immunofluorescence staining of a representative 3µm FFPE tissue section from the indicated conditions (red: pan-cytokeratin$^{+}$ keratinocytes, purple: CD45$^{+}$ immune cells, blue: sytox-green stained nuclei). Maculopapular rash (MPR), drug rash with eosinophilia and systemic symptoms (DRESS), toxic epidermal necrolysis (TEN) and healthy individuals. N = 5 individuals per cohort (TEN, DRESS, MPR and healthy). **b.** The Deep Visual Proteomics pipeline, including high-content imaging, machine learning-based cell segmentation and classification, followed by laser microdissection and ultra-sensitive mass spectrometry-based proteomics at cellular resolution. Representative area of excised keratinocyte (yellow arrow) and immune cell (red arrow) from FFPE tissue sections. **c.** Identification of unique protein groups and principal component analysis across all patients in the indicated cell type. Each dot is a patient (N = 21), color represents the cell type. This figure was partially created with Biorender.com.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

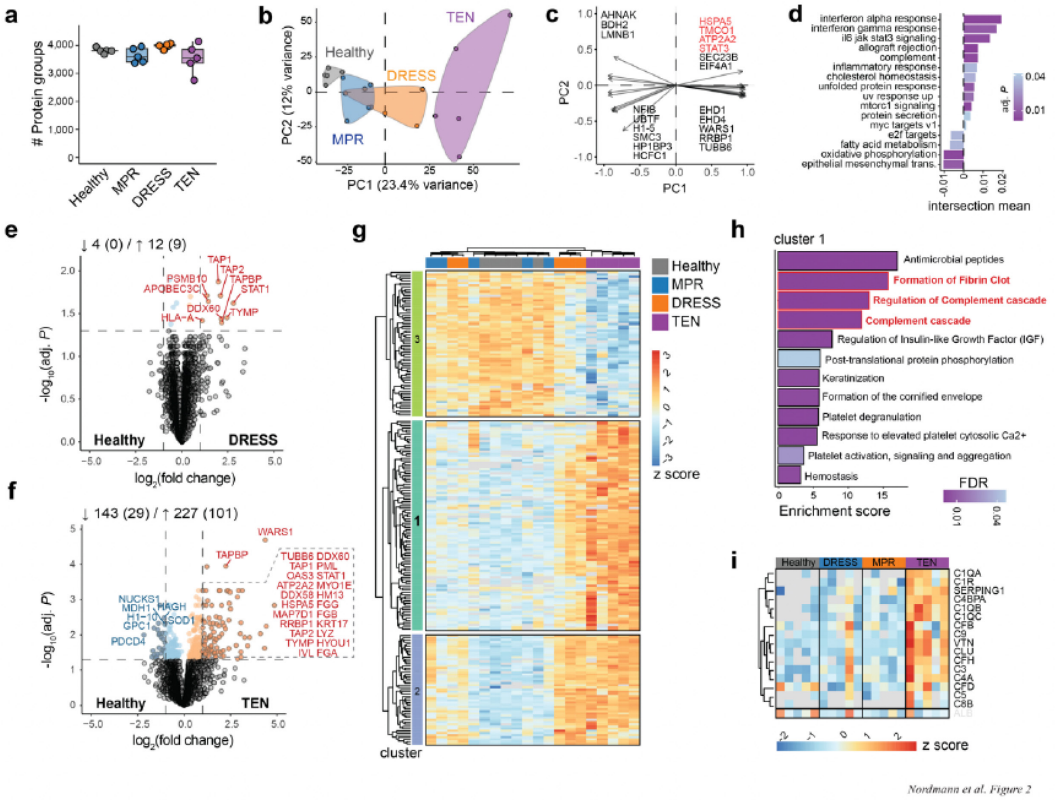
## The proteome of lesional keratinocytes in cutaneous drug reactions

Having acquired the cell-type resolved proteomes of our cohort, we performed comprehensive data analysis of lesional keratinocytes of the selected CADRs. We quantified a median of 3788 protein groups across all patients, covering a dynamic range of 4.2 orders of magnitude (Fig. 2a and Extended Data Fig. 2a). Data completeness and quantitative accuracy were excellent (median coefficient of variation (CV) of 17.4% per cohort, despite patient heterogeneity, Extended Data Fig. 2b-c). Keratinocyte proteomes clustered according to the type of drug reaction, with Principal Component 1 separating TEN from other CADRs and keratinocytes from healthy individuals (Fig. 2b). PC1 was enriched in proteins involved in calcium homeostasis and cellular stress response, including HSPA5 (heat-shock protein family A member 5), TMCO1 (transmembrane and coiled coil domains 1) and ATP2A2 (ATPase sarcoplasmic/endoplasmic reticulum $Ca^{2+}$ transporting 2; Fig. 2c). Interestingly, STAT3 (signal transducer and activator of transcription 3) was also amongst the main drivers of PC1, congruent with a dominant interferon signature in gene set enrichment analysis of PC1 (GSEA; Fig. 2c, d). Pairwise comparison of diseased to healthy proteomes revealed that the number of differentially expressed proteins (DEP) increased with the severity of the CADRs (Extended Data Fig. 2d). Thus, MPR, the mildest CADR, had no DEP when compared to healthy, whereas both DRESS and TEN exhibited significant changes. DRESS robustly upregulated multiple MHC class-I proteins in keratinocytes, suggesting that keratinocytes participate actively in culprit-drug associated antigen presentation (TAP1, TAP2, TAPBP, HLA-A; $\log_2$ fold change $> 1$, Fig. 2e).

In keratinocytes from TEN patients, the most severe CADR, there were 227 up - and 143 significantly down-regulated proteins, including those involved in antimicrobial response, such as lysozyme and clusterin (Fig. 2f). WARS1, one of the most strongly upregulated proteins ($\log_2$ fold change 4.36, *adj. P* $2.04 \times 10^{-5}$), has not previously been associated with TEN. This tryptophanyl-tRNA synthetase has a non-canonical role in activating the innate immune system as an endogenous TLR2/4 ligand, thereby triggering cytokine and chemokine production[27]. Expression of superoxide-dismutase (SOD) 1 was reduced by half. This is interesting because oxidative stress and reactive oxygen species (ROS) upregulate keratinocyte Fas-ligand expression, a key cytolytic molecule involved in keratinocyte apoptosis and epidermal detachment in TEN[17,28]. Unsupervised hierarchical clustering of ANOVA significant proteins between all three CADR types, showed a clear alignment for TEN and three distinct protein clusters, revealing a further degree of stratification (Fig. 2g). Overrepresentation analysis of

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

cluster 1, which represents proteins with highest intensity values in TEN, revealed multiple terms involved in the regulation of complement activation and blood clotting (Fig. 2h and Extended Data. Fig. 2e-f). In accordance, nearly all complement factors in our keratinocyte proteome dataset are uniquely upregulated in keratinocytes of TEN (Fig. 2i and Extended Data. Fig. 2g). Taken together, proteomes of keratinocytes derived from severe CADRs, DRESS and TEN, reveal distinct disease-associated protein expression signatures.




Nordmann et al. Figure 2

**Fig. 2 | a**. Number of proteins identified in keratinocytes across cohorts. Box plots show the median (center line) with interquartile range of 25% to 75% and 95% confidence interval. **b.** Principal Component (PCA) analysis, main drivers of PC1/2 separation (**c**) and PC1 gene set enrichment analysis (**d**). **e-f.** Differentially expressed proteins (DEPs) between the indicated condition and healthy. Colored dots are significant (BH adjusted $P < 0.05$, FDR 0.05), dashed vertical line represents $\log_2$ fold change of $\leq 1$ or $\geq 1$. **g.** Unsupervised hierarchical clustering with analysis of variance (ANOVA) significant proteins (adjusted $P < 0.05$, post-hoc Tukey HSD). Color indicates normalized intensities (z score) **h.** Overrepresentation analysis of ANOVA cluster 1, ordered according to enrichment score. Color indicates degree of significance. **i.** Semisupervised heatmap of complement factors in the indicated conditions, and albumin (at the bottom) to exclude serum contamination, color indicates normalized values (z score). N = 5 individuals per cohort, mean of biological duplicates per patient.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

## The proteome of lesional immune cells in cutaneous drug reactions

Cutaneous immune cell infiltration in TEN is thought to be sparse and scattered, in contrast to the dense and confined perivascular infiltrates in MPR and DRESS[29]. To understand underlying molecular differences in immune cells of the skin in CADRs, we assessed the in-depth proteome of immune cells (median of 3418 per patient, median CV 19.8% within cohorts, with high data completeness; Fig. 3a and Extended Data Fig. 3a-c). These proteomes again clustered according to disease phenotype and were separated from healthy controls, but in contrast to keratinocytes, Principal Component 1 of immune cells separated DRESS from other CADRs (Fig. 3b and Extended Data Fig. 3d). Gene set enrichment analysis of PC1 was prominently associated with targets of E2F and MYC (Extended Data Fig. 3e). Proteins driving this include RNA processing and splicing factors such as DDX39A (DExD-Box Helicase 39A), EIF4A3 (Eukaryotic Translation Initiation Factor 4A3) and SRSF6 (Serine and Arginine Rich Splicing Factor 6), indicating a highly proliferative state (Extended Data Fig. 3 f). PC1 and PC3 together revealed unique features of DRESS and TEN. In DRESS, multiple viral pathways terms are enriched, in line with the postulated role of latent viral reactivation in its pathogenesis (Fig. 3c). In contrast, a dominant interferon signature was evident in TEN, together with STAT1 as a main driver (Fig. 3c and Extended Data Fig. 3f). In direct proteomic comparison to healthy control, DRESS had the highest number of DEPs (n = 730), followed by TEN (n = 233) and MPR (n = 128; Fig. 3d - g). DEPs unique to TEN included GCA (Grancalcin, $\log_2$ fold change 1.56, q 2.38 x $10^{-2}$) and PDIA3 (Protein Disulfide Isomerase Family A Member 3, $\log_2$ fold change 0.642, q 2.51 x $10^{-2}$) (Fig. 3f), both not previously described in the context of TEN. Similar to its role in cancer[30], PDIA3 may modulate MHC-class I and subsequent $CD8^+$-driven cytotoxicity in TEN.

A minority of proteins were specifically identified within defined CADR subsets (Extended Data Fig. 3g). Of particular interest is histone methylase EZH2 (Enhancer of Zeste homolog), the catalytic member of *PRC* (Polycomb Repressive Complex) 2, detectable only in immune cells from DRESS patients (Fig. 3h). Immunofluorescence staining, as well as near-exclusive transcription in a data set of PBMC-derived lymphocyte cluster from a therapy resistant DRESS patient[31], confirmed our proteomic observation (Extended Data Fig. 3h - i). In fact, EZH2 fold change over healthy control was higher than the main targets reported in that study (STAT1, JAK3, IL2RG)[31]. While the role of EZH2 in cancer prompted the development of an FDA-approved inhibitor, its contribution to inflammatory and allergic diseases has not been extensively investigated[32].

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

IDO (Indoleamine 2, 3-dioxygenase 1) was only detected in severe CADRs and represents a further protein of interest with an already available, clinically approved inhibitor (Fig. 3h). Whether IDO1 has a detrimental role or dampens the cytotoxic reaction, as within the tumor microenvironment[33], requires further investigation. Although granzyme B has been described as a key mediator of cytotoxicity in TEN[21], our proteomics data showed that it is present at similar levels in non-cytotoxic DRESS (Fig. 3h).

Unsupervised hierarchical clustering of ANOVA significant proteins in immune cells resulted in three visually distinct protein clusters with perfect disease alignment (Fig. 3i). Overrepresentation analysis of Cluster 2, which mainly represents DRESS, confirmed the highly replicative features of infiltrating immune cells in this disease, illustrated by multiple terms for DNA replication (Extended Data Fig. 3j). Its most prominent term (Activation of ATR in response to replication stress) was dominated by the DNA replication factor MCM (Minichromosome Maintenance protein complex) and RFC (Replication Factor C; Extended Data Fig. 3k). Intriguingly, human MCM associates with EBV replication origin (oriP) for the propagation of the viral genome[34]. This potentially serves as a direct link to the viral pathogenesis of this disorder. Importantly, proteins in cluster 1, which dominantly represents TEN and are only mildly elevated in MPR and DRESS, were almost exclusively associated with type I and II interferon signaling (Fig. 3j). Given the urgent clinical need for a specific therapy for TEN, we further focused on the role of the interferon pathway in TEN.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

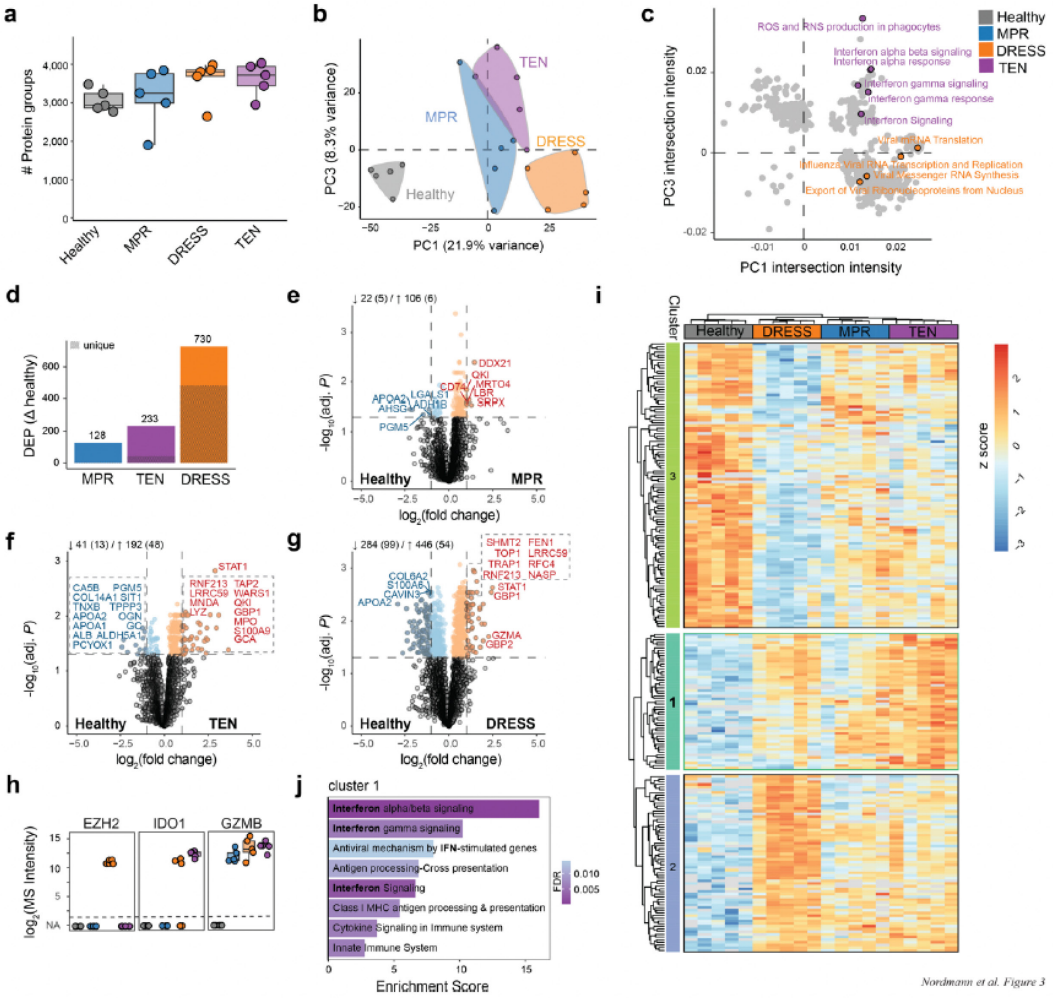



**Fig. 3| a.** Number of proteins identified in immune cells across all cohorts. Box plots show the median (center line) with interquartile range of 25% to 75% and 95% confidence interval. **b.** Principal Component (PCA) analysis of the proteomic results. **c.** Intersection of PC1 and PC3 gene set enrichment analysis. Terms including "viral" or "interferon" are highlighted. **d-g.** DEPs in the indicated conditions versus healthy and corresponding volcano plots. Shaded area within columns represents the proportion of unique proteins. In the volcano plots, color represents significance (*adj.* $P < 0.05$), dashed vertical line represents $\log_2$ fold change of $\leq 1$ or $\geq 1$. **h.** Selected proteins of interest, with a distinct identification across the cohorts **i.** Unsupervised hierarchical clustering of ANOVA significant proteins (*adj.* $P < 0.05$, post-hoc Tukey HSD). Color indicates normalized intensities (z score). **j.** Overrepresentation analysis of ANOVA cluster 1, ordered according to enrichment score. Color indicates degree of significance (FDR). N = 5 individuals per cohort, mean of biological duplicates per patient.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

## The JAK/STAT pathway is drastically activated in TEN

Severe inflammatory disorders are often characterized by feedback loops that exacerbate the immune response[35]. We therefore hypothesized that proteins significantly regulated in both keratinocytes and immune cells would identify therapeutically relevant pathways. We identified six such proteins (WARS1, STAT1, S100A9, LYZ, GBP1 and APOL2) that were all at least four-fold upregulated in both keratinocytes and immune cells of TEN compared to healthy control (Fig. 4a). Strikingly, all six proteins are part of the interferon pathway, responsible for signal transduction (STAT1) or triggered by it[36-40]. Interferons signal through the Janus Kinase / Signal Transducer and Activator of Transcription (JAK/STAT) and we examined the molecular components of this signaling cascade using our comprehensive cell-type resolved proteomic dataset (Fig. 4b and Extended Data Fig. 4a-b). STAT1 and STAT3 were upregulated in both immune cells and keratinocytes, while STAT2 was up in keratinocytes and STAT5 in immune cells. JAKs have low abundance and the only one we identified was JAK3 in TEN. Most downstream interferon-stimulated genes (ISG) were either strongly upregulated, illustrated by a 9-fold increase of GBP1, or only detectable in TEN (e.g. GBP1). This points to a strong upregulation of all components of the JAK-STAT pathway in TEN.

We confirmed substantial and broad JAK/STAT activation in TEN and also in Steven-Johnson-Syndrome (SJS)-TEN overlap by transcriptomic analysis in a second CADR cohort (N=20, Fig. 4c and Extended Data Fig. 3k). SJS-TEN overlap is defined by an affected body surface area of 10 – 30% in contrast to over 30% in TEN, with an otherwise identical phenotype. Of cytokines known to signal through the JAK/STAT pathway, interferon-gamma gene expression was the most strongly upregulated in TEN, exceeding that of cytokines such as TNF, known to be involved in TEN[41] (Fig. 4d). Visualization of STAT1 in FFPE tissue sections further confirmed its significant upregulation in TEN (Fig. 4e). Notably, the degree of single-cell nuclear STAT1 upregulation in TEN was similar to DRESS, in which therapeutic intervention with JAK-inhibitor proved to be successful[31]. To further establish STAT1 activation in TEN at the signaling level, we proceeded with histological analysis of phosphorylated STAT1. This revealed a profound degree of STAT1 phosphorylation in both skin-residing immune cells and keratinocytes of TEN patients (Fig. 4f). Collectively, these results suggest that activation of the JAK/STAT pathway is a key driver of TEN.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

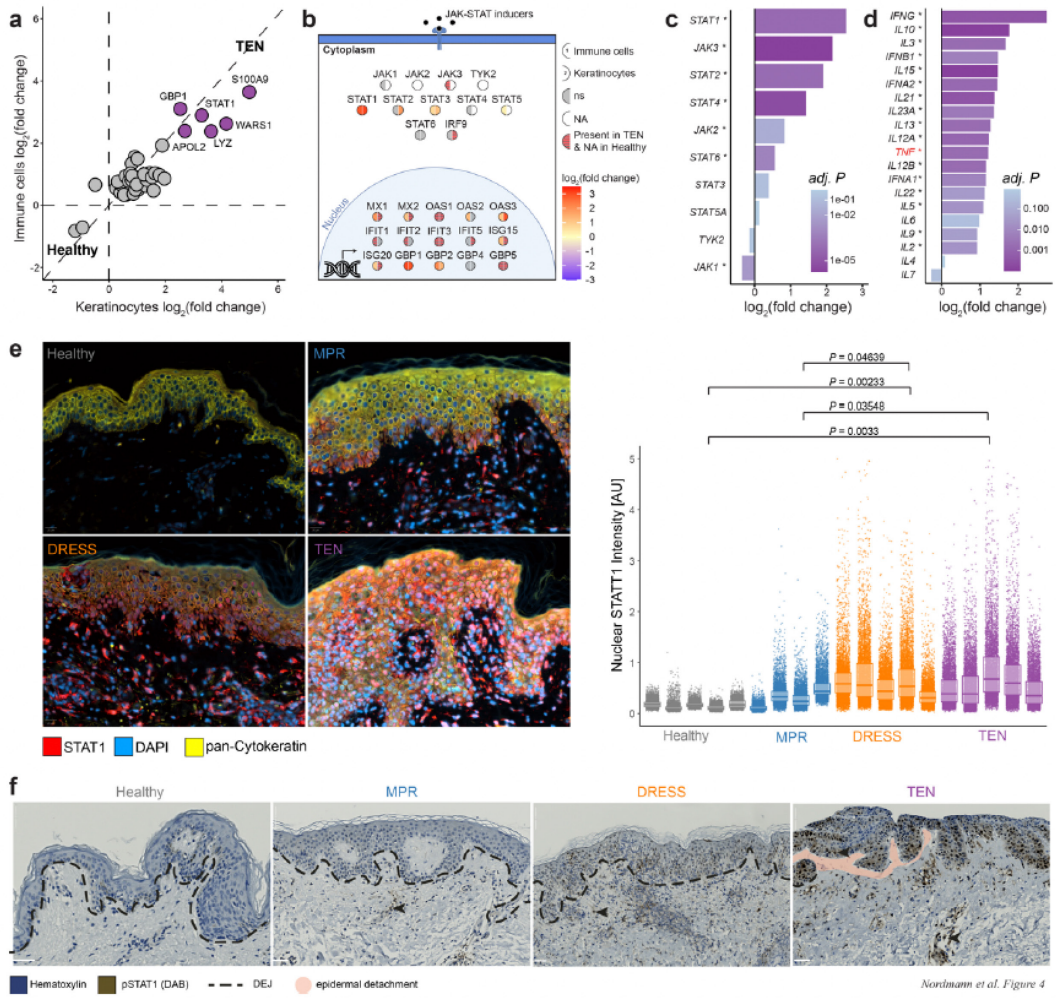



**Fig. 4 | a.** Significantly quantified keratinocyte (x-axis) and immune cell (y-axis) proteins and their corresponding $\log_2$ fold change compared to healthy. Purple color indicates proteins with $\log_2$ fold change > 2 in both cell types. Dashed vertical and horizontal lines indicate fold change = 0 and the dashed diagonal line represents equal regulation. **b.** Schematic illustration of quantified JAK/STAT pathway proteins in both cell types. Color represents $\log_2$ fold change compared to healthy, with direct comparison between immune cells (left semi-circle) and keratinocytes (right) of the indicated protein. Empty (semi-)circles are proteins with missing values. All colors except for grey are significantly regulated (*adj. P* < 0.05). **c.** mRNA expression levels in TEN compared to healthy for the indicated genes of the JAK/STAT pathway (validation cohort) **d.** mRNA expression change in TEN compared to healthy of genes signaling through the JAK/STAT pathway (validation cohort). Asterisks indicate significance (*adj. P* < 0.05). **e.** Representative immunofluorescent staining of FFPE tissue sections against Stat1 (red) and pan-Cytokeratin (yellow). Quantification of nuclear Stat1 intensity per cell, normalized to its own nuclear staining intensity, for all samples of the corresponding conditions. Box plots show the median (center line) with interquartile range of 25% to 75%. **f.** Representative phospho-STAT1 immunohistochemistry staining. Dashed lines represent the dermo-epidermal junction (DEJ). Epidermal detachment is highlighted with red (in

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

TEN). N = 5 individuals per cohort (TEN, DRESS, MPR and healthy) and mean of biological duplicates per patient, except for **c** and **d** where N = 10 per cohort (TEN, healthy).

## The JAK/STAT pathway is an actionable therapeutic target in preclinical models of TEN

To investigate the clinical significance of our experimental findings, we developed a novel *in-vitro* model that authentically replicates key aspects of cutaneous drug reactions (Fig. 5a). To this end, we cultivated primary keratinocytes from single hair follicles of TEN survivors, building on previous protocols[42]. Concurrently, we isolated patient-matched PBMCs and activated them using CD3/28. After five days, we fluorescently labeled the keratinocytes and co-cultured them with activated or resting PBMCs. We introduced a fluorescent DNA-binding dye to identify cells with compromised membrane integrity, in order to visualize, track and quantify the cytotoxic action of activated PBMCs towards keratinocytes using live-cell imaging. In this autologous co-culture model, keratinocytes were efficiently killed within 72 hours in the presence of activated PBMCs (Fig. 5b). Applying the pan-JAK inhibitor tofacitinib, an FDA and EMA approved drug, resulted in dose-dependent inhibition of keratinocyte cytotoxicity, whereas co-culturing of keratinocytes with resting PBMCs or exposure of keratinocytes to tofacitinib alone did not induce keratinocyte cell death (Fig. 5c).

Encouraged by these results, we next evaluated the efficacy of oral JAK inhibitors in established mouse models of TEN[43,44]. Loss of Inhibitor of Apoptosis proteins (IAPs) shifts TNF signaling towards apoptosis[45] and necroptosis[46]. Both cell death modalities contribute to the pathology of TEN[19,47]. In mice, subcutaneous (s.c.) injection of a small molecule IAP antagonist drug (smac-mimetic, SM) induced TEN-like cutaneous inflammation, epidermal detachment and histological features of TEN (Fig. 5d-f). Oral treatment with JAKi (tofacitinib 30 mg/kg or baricitinib 10 mg/kg) markedly reduced macroscopic disease severity on day 1 and 3 as quantified by a clinical assessment score composed of all major TEN hallmarks (Nikolsky sign, edema, rubor; Fig. 5e and Extended Data Fig. 5a). Similar to human TEN, a strong increase in pSTAT1 was detectable following SM injection and reduced disease severity through treatment with JAK inhibitors was accompanied by a significant reduction in cutaneous pSTAT1 levels (Fig. 5f, and Extended Data Fig. 5b). Average dermal thickness was elevated upon SM injection compared to vehicle and restored to normal levels upon JAKi (Fig 5g and Extended Data Fig. 5c). Further, JAKi also decreased keratinocyte cell death ($TUNEL^+$/cleaved caspase-3 $(CC3)^+$, Fig. 5f and Extended Data Fig. 5d,) and showed less

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

immune cell infiltration ($CD45^+$, Extended Data Fig. 5b). Interestingly, in both inhibitor groups CC3 positive cells were more frequently TUNEL negative indicating an earlier stage of the apoptotic program (Fig. 5f). This suggests, that the process of cell death is dampened and delayed as a result of JAK inhibition. Importantly, epidermal recovery following s.c. SM injection was accelerated in JAKi treated mice, demonstrated by the strong expression of the proliferation marker Ki67 in the basal layer, ruling out a detrimental effect of treatment on wound healing (Extended Data Fig. 5b).

Despite the striking similarities of the SM-induced murine model with human TEN, SM primarily engages downstream pathways which represent a potential limitation in translating these findings to the human context. To address this, we further tested the efficacy of JAK inhibitor baricitinib in an established humanized mouse model of TEN, in which patient PBMCs of a TEN survivor are injected intravenously into immunocompromised NOD/Shi-scid, IL-2Rγnull (NOG) mice (Fig. 5h). Following daily oral administration of the causative drug (acetaminophen) severe ocular conjunctivitis and cell death of conjunctival epithelium reminiscent of TEN occurred at day 14 (Fig. 5i – j and Extended Data Fig. 6a - b and d-e)[19]. In this second mouse model of TEN, baricitinib treatment reduced (clinical) ocular conjunctivitis and (histological) cell death of conjunctival epithelium as compared to vehicle treatment ($TUNEL^+/TUNEL^-$ ratio for baricitinib: 1.15% (mean) ± 1.63 (SD) and Vehicle: 13.45% (mean) ± 6.29 (SD); Fig. 5j and Extended Data Fig. 6a - g). Taken together, the *in vitro* and *in vivo* data clearly demonstrates the efficacy of JAKi in preclinical models of TEN.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

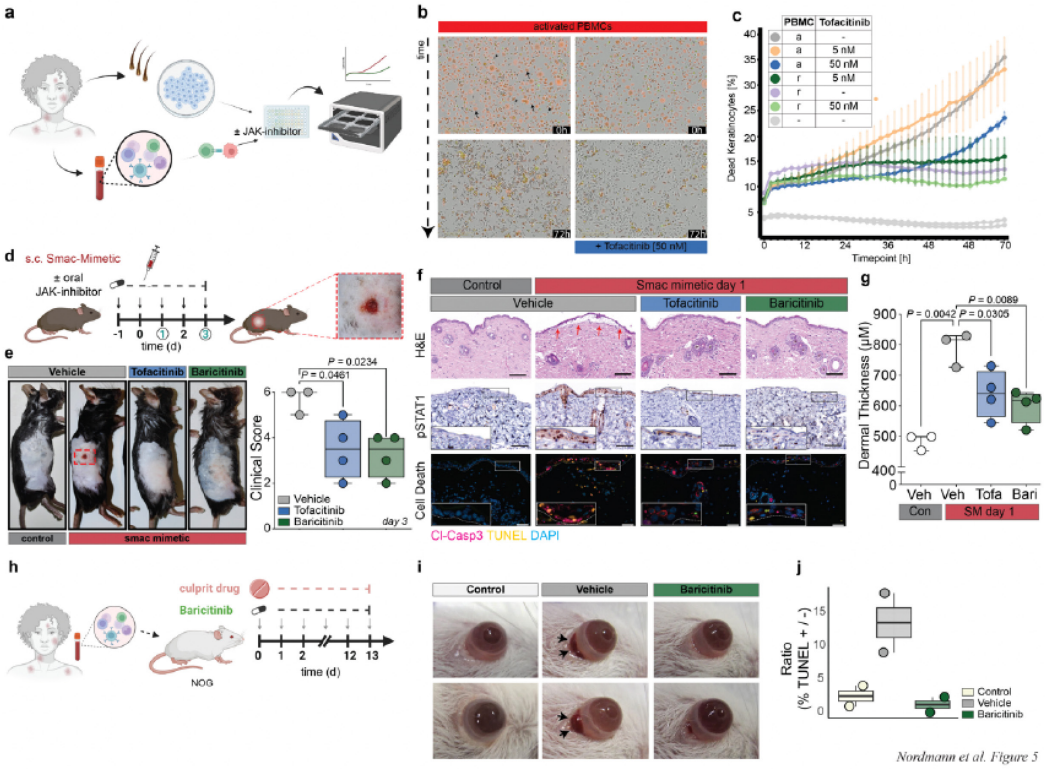



**Fig. 5 | a.** Live-cell imaging co-culture assay with keratinocytes and activated PBMCs from the same patient ± JAK inhibitor, to assess keratinocyte cell death **b.** Keratinocytes labeled with CellTracker Red (arrow) in co-culture with unlabeled PBMCs (arrowhead) at timepoint 0 and 72h, ± 50nM tofacitinib. **c.** Kinetic data of **b.** quantified over time (a = activated; r = resting; n=6 replicates per condition). **d.** Experimental protocol to assess efficacy of oral JAK inhibitor tofacitinib (30 mg/kg) or baricitinib (10 mg/kg) in combination with s.c. SMAC mimetics, which cause a TEN phenotype (red arrows: epidermal detachment). **e.** Representative macroscopic image and clinical assessment score to evaluate the macroscopic effect of treatment of the corresponding cohort on day three (n = 3 for vehicle and 4 for tofacitinib or baricitinib group). **f.** Representative H&E, pSTAT1 and cell death (magenta = cleaved caspase 3 / yellow = TUNEL / cyan = DAPI) microscopy image of the corresponding cohort. Inserts magnify indicated area of interest. Scale bars = 100µM. **g.** Average dermal thickness of the indicated conditions on day 1 (each data point is a mouse, N=15 measurements/mouse). **h.** Experimental protocol to assess efficacy of baricitinib (10 mg/kg) in a humanized mouse model of TEN, using patient PBMCs. **i.** Ocular reaction (black arrows) following daily culprit drug administration ± oral baricitinib (10 mg/kg). **j.** Ratio of TUNEL positive over TUNEL negative subepithelium of the corresponding cohort (n = 2 per condition) Box plots show the median (center line) with interquartile range of 25% to 75% and 95% confidence interval.

## Patients with toxic epidermal necrolysis recover after JAK inhibition

Based on our preclinical data and the urgent clinical need in this devastating disease, we treated four patients with TEN or SJS/TEN overlap with off-label JAKi treatment. One such patient

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

was receiving carboplatin, etoposide and the PD1-antagonist serplulimab for small-cell lung cancer (Fig. 6). After the third cycle of anti-cancer treatment, he developed TEN with epidermal detachment affecting 35% of his body surface and a severity-of-illness score (SCORTEN[48]) of 4, indicating a predicted mortality rate of 58.3%. Epidermal detachment progressed despite high-dose intravenous methylprednisolone (80 mg $d^{-1}$) and therefore we initiated JAK1 inhibitor rescue therapy with abrocitinib, an acknowledged drug in the field of dermatology and recently approved for treatment of atopic dermatitis[49]. Within two days, the disease stopped progressing and re-epithelialization was visible within four days, reaching near-completeness after 16 days (95%), after which the patient was discharged following full recovery from TEN. All four patients treated with JAKi survived at day 30 and no side effects occurred (Extended Data Fig. 7a - d). Notably, cutaneous pSTAT1 levels were drastically reduced in all patients following JAK-inhibitor treatment (Extended Data Fig. 7e). This clinical data shows the potential of repurposing approved JAK1 inhibitor abrocitinib or pan-JAK inhibitors such as tofacitinib for the treatment of TEN.

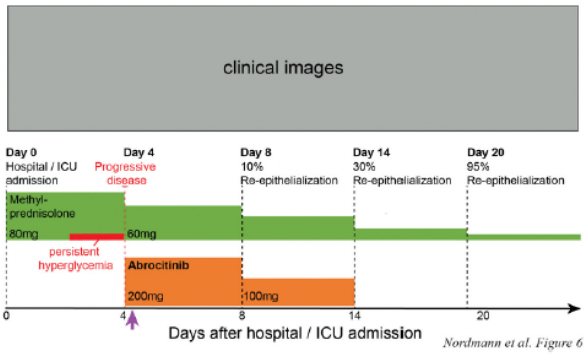



**Fig. 6** | Disease course under JAK1 inhibition (abrocitinib) in a severe case of toxic epidermal necrolysis. Arrow marks timepoint starting with abrocitinib.

## Discussion

The skin is the most frequently affected organ by adverse drug reactions, and these reactions occur in 1 – 3 % of all hospitalized patients[50]. Severe CADRs and, in particular, TEN have an elusive pathogenesis and cause substantial fatality[2]. Here, we used deep spatial proteomic profiling of keratinocytes and immune cells to analyze FFPE tissue sections from patients exhibiting three different types of CADR and healthy donors, and quantified ~5,000 proteins in both types of cells. This in-depth proteomic analysis across all major types of cutaneous drug reactions with cell-type resolution provided unique insights into the molecular mechanism of

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

each CADR and identified targets for therapeutic interventions of these disorders. Skin tissue is composed of many different cell types and Deep Visual Proteomics allowed a nuanced view of these cell-type specific proteomic changes, even in rare samples with sparse and scattered immune cell infiltration. For instance, the striking upregulation of WARS1 in keratinocytes and immune cells of TEN is of special interest, in light of the increasingly appreciated role of the innate immune system in this disease[51]. Both immune cells and keratinocyte carry functional TLR receptors, including TLR2 and 4[52]. Thus, our results may link the non-canonical role of WARS1 as an endogenous TLR2/4 ligand to the immune pathogenesis of TEN. Given its ability to trigger cytokine and chemokine production, it would be interesting to investigate if secreted WARS1 directly contributes to the damage occurring in TEN by initiating and exacerbating the immune response.

In skin-infiltrating immune cells of DRESS, the most dominant finding was the appearance of the histone methyltransferase EZH2, which we further confirmed by histology. EZH2 plays a key role in epigenetic gene regulation by modifying histone H3 at lysine 27 (H3K27me3). Functional consequences mainly attribute to the increased proliferation of cancer cells; however, recent studies suggest a more nuanced, quantitative effect on defining cellular states[53]. EZH2 is the third most common site of mutation (29%) in patients with idiopathic hypereosinophilic syndrome[54]. Intriguingly, systemic eosinophilia is a hallmark of DRESS and the primary culprit of organ damage, indicating a potential connection to EZH2 dysfunction[55]. In addition, E2F pathway enrichment detected in our proteomic dataset correlates with the overexpression of EZH2, which functions as a downstream effector in this signaling cascade[56]. Together, these findings strongly suggest that EZH2 and, by extension, epigenetic modifications, may play a significant part in the pathogenesis of DRESS. As EZH2 is inhibited by the FDA-approved drug tazemetostat, elucidating the precise role of EZH2 in DRESS could open new therapeutic avenues for this severe, multiorgan inflammatory CADRs[57].

The major finding of our study, is the drastic upregulation of the JAK/STAT pathway in TEN, which is not observed in less severe CADRs. Our results pinpoint this signaling pathway as the main driver of cutaneous inflammation, keratinocyte cytotoxicity and epidermal detachment. Our cell-type resolved findings indicate a more active role of keratinocytes than previously anticipated in driving the skin directed cytotoxic immune response. Many components of the JAK/STAT signaling cascade are upregulated in both immune cells and keratinocytes of TEN, highlighting the potential for detrimental self-amplifying nature of interferon signaling that could explain the extensive tissue damage[58].

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

To date, standard of care for TEN is limited to immediate cessation of the causative drug and best supportive care, usually in an intensive care unit. Unfortunately, under these standard of care conditions hospital mortality rates range from 15-20 percent, highlighting the urgent need for more effective treatments. Our spatial proteomics results nominated the JAK/STAT pathway as an actionable therapeutic target in TEN, and we provided *in vitro* and *in vivo* preclinical evidence that JAK inhibitors substantially reduced keratinocyte-directed cytotoxicity.

Importantly, we treated four patients with TEN or SJS-TEN overlap syndrome, two of which were previously unresponsive to high-dose systemic corticosteroids. All four showed a remarkable response to JAK inhibitors and were discharged in good health after treatment. Taken together, our data strongly suggest that JAK inhibition is feasible and beneficial in TEN. Given the high short-term risk of mortality associated with TEN and the short duration of specific therapy required, it appears that the potential benefits of short-term use of JAK inhibitor therapy, if confirmed in larger clinical trials, could outweigh the potential risks of long-term JAK inhibition observed in patients with chronic inflammatory disorders[59]. Due to strong association of developing TEN during immune checkpoint inhibition (OR= 4.33)[60], JAKi may be especially attractive to rapidly resume oncological treatment in case this adverse reaction develops. While our study with JAK-inhibition in a small number of patients cannot be considered conclusive in terms of efficacy and should, therefore, be interpreted with caution, its potential for the management of TEN, a life-threatening disease with no efficient treatment so far, should be taken into consideration.

These data pave the way for a clinical trial of early JAKi therapy in TEN with the potential to change the outcome of this lethal adverse reaction to common drugs. Our study is also a successful example of directly linking emerging cell-type resolved spatial omics technologies, especially spatial proteomics, with a new treatment modality that is of benefit to patients. We envision that similar approaches could be transformative in a broad set of inflammatory or oncological conditions by helping to target treatments more precisely.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

## Acknowledgements

We thank all members of the Proteomics and Signal Transduction group at the Max Planck Institute of Biochemistry, especially Dr. Maximilian T. Strauss for bioinformatic cell segmentation strategies, Katharina Zettl for superb immunofluorescence staining and Bianca Splettstoesser for cell culture. We thank Dr. Katrin Kerl-French for revalidation of all cases. Dr. Leonie Zeitler and Dr. Patricia P. Ogger provided technical assistance for the live-cell imaging experiments. The J2 / 3T3 cells were a generous gift from Prof. Hans-Dietmar Beer (University Hospital Zurich, Switzerland).

## Author contributions

T.M.N., L.E.F. and M.M. conceived the project. T.M.N., H.A. A.H., R.A., J.S., L.E.F. and M.M. designed the experimental framework. T.M.N. performed the proteomic experiments with help from L.S., A.M., M.Z. and M.T. H.A., M.C.T., N.S. and J.S., performed and analyzed the smac mimetic mouse model experiments. A.H., Y.S., H.K., H.H.N. and R.A. performed and analyzed the humanized mouse model experiments. P.Z., T.G. and C.J. performed and analyzed the JAK inhibition in patients. T.M.N., H.A., J.S., L.E.F. and M.M. designed the figures with help from F.A.R., M.T. and L.S.. T.M.N. performed the live-cell imaging experiments and analyses, with substantial experimental guidance from A.G. and P.J.M. Transcriptomics experiments and analyses were performed by P.C.S. and T.K.S., whereas L.S., A.S., A.M., F.A.R., M.Z., M.T., E.H.R. and A.M. provided intellectual input and helped interpret data. F.A., R.G., M.P.L. and S. I-H-O. provided intellectual input and helped construct the cohorts. T.M.N., L.E.F. and M.M. wrote the manuscript. All authors read, revised and approved the manuscript.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

## Funding

T.M.N. is supported by a Swiss National Science Foundation (SNSF) Early Postdoc Mobility (P2ZHP3-199648) and Postdoc Mobility Fellowship (P500PM_210917). M.T. is supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 874839 (ISLET). J.S., H.A. and N.S. are supported by NHMRC grants Investigator Grant (1195038) and through the Victorian State Government Operational Infrastructure Support and Australian Government NHMRC IRIISS (GNT9000719).

## Methods

### Patient biopsies

Skin tissue biopsies were obtained during routine histopathological diagnostic procedures at the Departments of Dermatology and Allergology of the University Hospital Zurich (proteomic cohort) and Ludwig Maximilian University of Munich (validation cohort). In the proteomic cohort, TEN patients had a minimal affected skin area of 30%. In the validation cohort, patients with SJS/TEN overlap were included (minimal affected skin >10%). All patients with DRESS had a RegiSCORE of 6-7 (definite diagnosis). In addition to the typical clinical features, the lymphocyte transformation test (LTT) or skin-tests were positive in all MPR patients. Histopathological diagnosis was re-validated in all cases by a board-certified dermatopathologist using a fresh H&E-stained tissue section. Baseline characteristics were statistically evaluated using Analysis of Variance (ANOVA) for numeric variables (age) and a Chi-squared test of independence for categorical variables (sex). All obtained disease biopsies were performed in the context of routine clinical workup for retrospective analysis with informed consent and ethical approval in place (Munich: 22-0342, 22-0343; Zurich: BASEC: Req-2021-00226 and 2017--00494; Fujian: MRCTA, ECFAH of FMU[2023]400. All experiments were performed in accordance with the Declaration of Helsinki.

### Treatment of patients with JAK inhibitors

Regular assessment and vigilant surveillance of the vital signs, hematological parameters and coagulation markers was performed throughout treatment. Patients with active infections were excluded. All patients were additionally treated according to current best supportive care guidelines[22]. Oral abrocitinib was administered at a regimen of 200mg daily for 5-7 days, and tapered to 100mg daily for 5-7 days. Similarly, oral tofacitinib was administered at a dosage of 10mg daily for 5-7 days, and then tapered to 5mg daily for 5-7 days. In cancer patients with

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

high SCORTEN, a reduction in JAKi dosage was be initiated after observing a certain degree of re-epithelialization. Treatment was approved by the local ethics committee and the institutional review board of the First Affiliated Hospital of Fujian Medical University (Fujian: MRCTA, ECFAH of FMU[2023]400), and the patient was provided written informed consent.

**Mice**

**Smac mimetic TEN mouse model:**

As previously published[43] male mice were injected subcutaneously with 100µl of 1mg/ml smac mimetic (SM) CompA (TetraLogic Pharmaceuticals) or vehicle (12% Captisol). JAK inhibitor (10 mg/kg Baricitinib or 30 mg/kg Tofacitinib; Selleckchem) or vehicle (20% Captisol; 50µl; Selleckchem) was administered by oral gavage twice daily starting one day before s.c. SM injection. Mice were sacrificed on either day one or day three post SM injection, the site was photographed, scored, and samples collected for ex-vivo analysis. We used a four-point ordinal scale (0-4) to assess the three clinical associated macroscopic criteria of TEN (Oedema, rubor and epidermal disruption (Nikolksy sign / lesion severity)) at day one or three post SM injection. They were assessed and scored as 0 (not. present), 1 (mild), 2 (moderate), or 3 (severe) and the scores were combined for an overall clinical score. Experiments were approved by the local ethics committee (WEHI AEC# 2022.009).

**Humanized mouse model of TEN:**

Immunocompromised NOD/Shi-scid, IL-2Rγnull (NOG) mice at 6 weeks of age were included in the experiment with n = 2 in each of Baricitinib treatment and both control groups. PBMCs ($2\times10^6$) from a patient who had recovered from SJS/TEN 1 year earlier were injected intravenously into the NOG mice, followed by oral administration of the causative drugs (acetaminophen, 1.5 mg/100 µl) at day one and once daily thereafter. The dosage used in this model was based on mg / kg body weight, converted from the normal adult human dose. Baricitinib (10 mg/kg) or vehicle (20% Captisol; 50 µl) was administered by oral gavage twice daily from day one. Body weight, ocular reactions or cutaneous changes were assessed daily. On day 14, the degree of conjunctivitis was evaluated after ocular dislocation under general anesthesia and scored (0 = no conjunctivitis, 1 = mild conjunctivitis, 2 = severe conjunctivitis). The ratio of the number of TUNEL-positive cells to the total conjunctival cell count in a 400x field was calculated for all mice. The patients received no systemic glucocorticoids at the time of PBMCs collection. Experiments were approved by the local ethics committee and the

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

institutional review board of Niigata University, and the patient was provided written informed consent.

**Membrane slide tissue mounting, immunofluorescence staining, imaging and laser microdissection**

2 µm PEN membrane slides (MicroDissect GmbH) were exposed to UV light (254 nm) for one hour and then coated with Vectabond (Vector laboratories; SP-1800-7) according to the manufacturers protocol. 3 µM thick tissue sections of each FFPE block were mounted on the pretreated membrane slides and dried overnight at 37°C. Immunofluorescence was performed according to our previously optimized protocol for membrane slides[61]. In brief, slides were heated to 56°C for 20min and immediately deparaffinized / rehydrated (Xylol 2x2 min, 100% EtOH 2x1 min, 90% EtOH 2x1 min, 75% EtOH 2x1 min, 30% EtOH 2x1 min, dd$H_20$ 2x1 min). Slides were then transferred to prewarmed glycerol-supplemented Antigen Retrieval buffer[61] (DAKO pH9 #S2367 + 10% Glycerol) at 88°C for 20 min, followed by a 20 min cooldown at RT. Slides were then washed in water and blocked with 5% BSA / PBS for 30 min, followed by overnight primary antibody incubation at 4°C in a humid staining chamber (1:100 mouse monoclonal CD45, DAKO M0701; 1:200 rabbit polyclonal CK Pan, DAKO Z0622). After washing in PBS, secondary antibodies (1:400 anti-mouse IgG A647, Invitrogen A32728; 1:400 anti-rabbit IgG A555, Invitrogen A32732) were incubated for 1 h at RT. After washing in PBS, sections were counterstained with SYTOX™ Green (1:400, Invitrogen S7020) for 10 min, washed again and coverslipped using Slowfade Diamond Antifade Mountant (Invitrogen, S36963). Sections were imaged on a Zeiss Axioscan at 20x magnification with a tile overlap of 10%. At an excitation wavelength of 553 nm and 631 nm, 100% laser intensity were used, while 50% was used in the 493 nm channel. Illumination time was adapted to the optimal spectral properties. Five z-stacks at an interval of 1.25 µm were acquired. Multi-scene images were then split into single scenes, z-stacks combined to a single plane by extended depth of focus (variance method, standard settings) and stitched, using the proprietary Zeiss Zen Imaging software. Images were imported as .czi files into the Biological Image Analysis Software (BIAS) with the packaged import tool[26]. Singe-cell segmentation was performed using deep neural network on the basis of pan-cytokeratin for keratinocytes and CD45 for immune cells at 1.0 input spatial scaling, 50% detection confidence and 30% contour confidence. Only contours between 30 µm$^2$ and 200 µm$^2$ were taken into consideration. After removal of duplicates at tile-overlapping regions, a supervised machine learning approach was used to remove false identifications and overlapping cell types. Contours were exported

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

together with three calibration points that were chosen at characteristic tissue positions. Contour outlines were simplified by removing 99% of data points. For keratinocytes, only every second shape was (randomly) chosen in order to prevent membrane instability while cutting multiple adjacent cells. Contour outlines were imported after reference point alignment, and shapes were cut by laser microdissection with the LMD7 (Leica) with a 63x objective in a semi-automated manner at the following settings: power 57, aperture 1-2, speed 23, middle pulse count 1, final pulse -3, head current 46 - 53%, pulse frequency 2600, offset 210. We collected 700 contours for each keratinocyte sample and 1000 for each immune cell sample. For each patient, biological duplicates were collected when possible, from the same slide at different regions. Dissected contours were collected directly into the same underlying low-binding 384-well plate (Eppendorf 0030129547), immune cells in the top half of the plate and keratinocytes in the bottom half, leaving an empty well between each patient. The plate was sealed with adhesive foil (Covaris), centrifuged at 1,000xg for 1 minute and then kept at RT until further processing.

**Sample preparation for downstream LC/MS**

Semi-automated sample preparation and digestion was performed in the collection plate using the Bravo pipetting robot (Agilent) as previously described[26]. For this, wells were washed with 28 µl of 100% acetonitrile and dried in a SpeedVac for 20 min at 45°C. The contours were then resuspended in 4 µl of 60mM triethylammonium bicarbonate (TEAB, Sigma) in MS-grade H2O, sealed with two adhesive foils and heated at 95°C for 60min in a 384-well thermal cycler (Eppendorf). 1 µl of 60% acetonitrile in 60 mm TEAB (12.5% final acetonitrile concentration) was added, followed by heating at 75°C for 60 min in a 384-well thermal cycler. Samples were then predigested with 4 ng of Trypsin for 4 h followed by overnight digestion with 6 ng LysC in a 384-well thermal cycler at 37°C. After 18 h the reaction was quenched with 1.5 µl 6% trifluoroacetic acid (1% final concentration). Samples were then manually transferred to single PCR tubes, dried in a SpeedVac for approx. 60 min at 45°C and stored at -20°C.

**LC-MS/MS analysis**

Peptides were resuspended in 4.1µl MS loading buffer (2% AcN (v/v) + 0.1% TFA (v/v) in MS-grade H20) immediately prior to measurement. All samples were stratified into cell-type and replicates, and further randomized using the "rand" function in Excel. An EASY nanoLC 1200 (Thermo Fisher Scientific) was coupled to a timsTOF SCP mass spectrometer (Bruker) via a nanoelectrospray ion source (Captive spray source, Bruker). Peptides were separated on

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

a 50 cm in-house packed HPLC column (75 um inner diameter, 1.9 um ReproSil-Pur C18-AQ silica beads (Dr. Maisch GmbH)), which was heated to 60°C by an in-house manufactured oven. A linear gradient of 120 min was ramped at a constant flow rate of 300 nl/min from 3 to 30% buffer B in 95 min, followed by an increase to 60% for 5 min, washed at 95% buffer B for 10 min and re-equilibration for 10 min at 5% buffer B (buffer A: 0.1% formic acid (FA) and 99.9% ddH2O; buffer B: 0.1% FA, 80% ACN, and 19.9% ddH2O). The timsTOF SCP mass spectrometer was operated in dia-PASEF mode using the standard 16 dia-PASEF scan acquisition scheme[62] with 4 IM steps per dia-PASEF scan, covering a m/z range from 400 to 1200 and ion mobility of 0.6 to 1.6 Vs $cm^{-2}$. All other settings were standard as described in[63].

**Raw data analysis with DIA-NN**

A library-free search was performed, using a DL predicted spectral library in DIA-NN (v1.8.0)[64]. Uniprot human databases UP000005640_9606 and UP000005640_9606_additional were used. MS raw files from immune cells and keratinocytes were searched separately in DIA-NN, apart from data shown in Figure 1c. Methionine oxidation was defined as variable modification and missed cleavages were limited to one. The precursor charge ranged from 2 to 4, precursor mass range was set to 300 to 1,800, and peptide length from 7 to 35. Mass and MS1 accuracy were set to 15, based on prior estimation. Isotopologues, and MBR were enabled and neural network classifier was set to single-pass mod. The '--relaxed-prot-inf' function was activated using the command line for more conservative protein grouping. Proteins were inferred from FASTA. Library generation was set as 'Smart profiling', 'RT-dependent' as cross-run normalization an 'Robust LC (high precision)' as quantification strategy.

**Preparation and subsequent transcriptomic analysis of FFPE tissue sections**

Ten-micrometer formalin-fixed, paraffin embedded human skin sections were cut from each sample and deparaffinized using ROTICLEAR® (Carl Roth GmbH + Ko. KG). RNA was isolated using RecoverAllTM Total NucleicAcid Isolation Kit (AM1975, Thermo Fisher Scientific) and concentrated using the Savant™ SpeedVac™ DNA 130 Integrated Vacuum Concentator (Thermo Fisher Scientific). Concentration was measured using Nanodrop One/Onec (Thermo Fisher Scientific). 150 µg of RNA from each sample was hybridized over night with unique probe pairs for 624 genes, including 594 genes from the nCounter® Immunology Panel, 15 internal reference genes, and 30 user defined genes using the Panel Plus option (NanoString Tecnhologies). Data were collected using the nCounter® SPRINTTM Profiler (NanoString Tecnhologies).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

**Bioinformatics data analysis on MS and transcriptomic data**

Bioinformatics data analysis was carried out using the R statistical computing environment version 4.0.2. For subsequent statistical analysis, protein or gene intensities were log2-transformed and average intensities of the biological replicates were computed for each patient and cell type. Aside from the unique protein numbers mentioned in Fig. 1c, only proteins that exhibited an identification rate of 70% or higher in at least one cohort were included, for each cell type independently. Box plots show the median (centre line) with interquartile range of 25% to 75% and 95% confidence interval. For Principal component analysis (PCA), missing values were imputed patient-wise based on a normal distribution (width of 0.3, downshift of 1.8). For the analysis of variance (ANOVA), the R stats package was used; variables were zero-centered and scaled prior to analysis. Subsequent multiple testing correction (*adj. P*) was conducted using the Benjamini-Hochberg (BH) method with a false discovery rate (FDR) cutoff of 5%; Post-hoc Tukey Honest Significant Differences (Tukey-HSD) were calculated at a confidence level of 0.95. For differential expression, the non-paired t-test was performed using the rstatix package with standard settings, followed by multiple comparison correction (BH, FDR 5%) from the stats package. For the comparison between TEN and healthy for selected proteins in both cell types (shown in Figure 4b), a non-paired t-test was performed and the $\log_2$ fold change value of TEN over healthy was visualized in the figure using a consistent range and color-coding for both cell types. The fill color of the (semi-circles) was manually adjusted using the "eyedropper" tool in Adobe illustrator. Proteomic overrepresentation analyses (ORA) and gene set enrichment analysis (GSEA) were performed with WebGestalt 2019[65] in the same R environment. The heatmaps were generated using the pheatmap package, with the data being zero-centered and scaled before display.

**Generation of hair follicle-derived keratinocytes for co-culture assay**

Hair follicles were placed on thin-Matrigel (Corning, #354234) coated T25 flasks. 1 ml of conditioned mouse embryonic fibroblast medium, from irradiated J2-3T3 cells, was carefully added. Medium was changed every 48 hours. When outgrowth of keratinocytes became visible, medium was aspirated and replaced with 2 ml Keratinocyte medium. When confluent, keratinocytes were detached using TrypLE™ Express Enzyme.

**Isolation of patient PBMCs for co-culture assay**

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

PBMCs were isolated using Ficol according to standard protocol. Cells were frozen in liquid nitrogen until further usage. Keratinocyte expansion and PBMC isolation were approved by the local ethics committee with written informed consent (23-0544).

**Live-cell imaging co-culture assay**

Patient PBMCs (2 x $10^5$ / well / 200 µl) were stimulated in U-shape 96-well plates in the presence or absence of CD3/28 (Gibco Dynabeads, #11161D) at a ratio of 1:1, in X-VIVO 15 (Lonza, #881026) for 5 days at 37 °C / 5% $CO^2$. On day 4 of PBMC stimulation, hair-follicle derived keratinocytes from the same patient were plated on 96-well plates. On day 5 of PBMC stimulation, keratinocytes were fluorescently labeled (Celltracker Red; Invitrogen #C34552) according to the manufacturer's recommendation. Activated or non-activated ("resting") PBMCs were then added to the keratinocytes at a ratio of 1:1, after magnetic removal of CD3/28 beads where necessary. Tofacitinib (CST, #14703) or control (DMSO) was added in the indicated concentration as well as Celltox Green (Promega, G8731, 1:8000 final dilution) to reach a final volume of 200 µl per well. Every 2 hours images were acquired using a live cell analysis system (Incucyte Sx3, Sartorius) for 72 hours at 37 °C / 5% $CO^2$. Quantification of dead keratinocytes ($Celltox^+Celltracker^+$ / total $Celltracker^+$) was performed using the proprietary analysis software (Incucyte, Sartorius).

**Immunostaining on regular glass slides**

3 µm FFPE tissue sections were mounted on regular superfrost-plus glass slides and exposed to high temperatures for improved tissue adherence. Subsequently, sections were deparaffinized and subjected to heat-induced antigen retrieval and blocking solution. The primary antibody was applied to the tissue section for one hour at room temperature or overnight at 4°C (pSTAT1 and cleaved caspase-3). For immunofluorescence staining, a species-specific secondary antibody coupled to a fluorophore was added for an additional hour, followed by a nuclear counterstain. Alternatively, immunohistochemical staining was conducted using horseradish peroxidase (HRP) coupled secondary antibodies, or signal amplified using biotinylated secondary antibodies followed by labelling with and Avidin-Biotin Conjugate (ABC) HRP Kit. HRP labelled samples were then color development with 3,3'-Diaminobenzidine (DAB) and Hematoxylin counterstain. TUNEL staining was performed according to the manufacturer's instruction prior to application of the anti-cleaved caspase-3 primary antibody for co-staining. H&E staining were performed according to standard protocol. Images were acquired on a Zeiss Axioscan 7 or a DP72 microscope and cellSens

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

imaging software (Olympus), or entire slides were scanned with a 3D Histech Virtual Slide Microscope (Olympus), viewed using CaseCenter (3D Histech, Hungary) and image captured on a MacBook Pro 13-inch retina display (Apple, CA, USA) (Ki67, CD45, H&E).

**Quantification of STAT1 expression in IF images**

3 µm FFPE tissue sections of all patients from the proteomic cohort were subjected to multiplex immunofluorescence staining for STAT1, pan-Cytokeratin and Hoechst. All stainings were then uniformly acquired as .czi files using the Zeiss Axioscan 7. Single-cell segmentation was performed within QuPath (v0.4.1) using the standard nuclear segmentation algorithm. After segmentation, all available features were extracted, including staining intensities for all channels and nuclei. Segmentation and feature extraction was automated with QuPath using the script editor, to ensure equality. In R (v. 4.0.2) 5000 nuclei and their corresponding measurements were then randomly chosen per patient and merged into a single file, excluding samples with fewer than 5000 cells (n = 1). To account for staining variation between slides, mean STAT1 fluorescence intensity was divided by its own mean Hoechst intensity on a per-cell basis. Statistical significance between cohorts was determined by a two-tailed t-test on mean normalized STAT1 values. Data is visualized using ggplot2.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

## References


1 Del Pozzo-Magana, B. R. & Liy-Wong, C. Drugs and the skin: A concise review of cutaneous adverse drug reactions. *Br J Clin Pharmacol* (2022). https://doi.org:10.1111/bcp.15490

2 Hoetzenecker, W. *et al.* Toxic epidermal necrolysis. *F1000Res* **5** (2016). https://doi.org:10.12688/f1000research.7574.1

3 Cho, Y. T., Yang, C. W. & Chu, C. Y. Drug Reaction with Eosinophilia and Systemic Symptoms (DRESS): An Interplay among Drugs, Viruses, and Immune System. *Int J Mol Sci* **18** (2017). https://doi.org:10.3390/ijms18061243

4 Lazarou, J., Pomeranz, B. H. & Corey, P. N. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* **279**, 1200-1205 (1998). https://doi.org:10.1001/jama.279.15.1200

5 Hsu, D. Y., Brieva, J., Silverberg, N. B. & Silverberg, J. I. Morbidity and Mortality of Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis in United States Adults. *J Invest Dermatol* **136**, 1387-1397 (2016). https://doi.org:10.1016/j.jid.2016.03.023

6 Sekula, P. *et al.* Comprehensive survival analysis of a cohort of patients with Stevens-Johnson syndrome and toxic epidermal necrolysis. *J Invest Dermatol* **133**, 1197-1204 (2013). https://doi.org:10.1038/jid.2012.510

7 Hung, S. I. *et al.* HLA-B*5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol. *Proc Natl Acad Sci U S A* **102**, 4134-4139 (2005). https://doi.org:10.1073/pnas.0409500102

8 Chung, W. H., Hung, S. I. & Chen, Y. T. Human leukocyte antigens and drug hypersensitivity. *Curr Opin Allergy Clin Immunol* 7, 317-323 (2007). https://doi.org:10.1097/ACI.0b013e3282370c5f

9 Chung, W. H. *et al.* Medical genetics: a marker for Stevens-Johnson syndrome. *Nature* **428**, 486 (2004). https://doi.org:10.1038/428486a

10 Ko, T. M. *et al.* Shared and restricted T-cell receptor use is crucial for carbamazepine-induced Stevens-Johnson syndrome. *J Allergy Clin Immunol* **128**, 1266-1276 e1211 (2011). https://doi.org:10.1016/j.jaci.2011.08.013

11 Nassif, A. *et al.* Toxic epidermal necrolysis: effector cells are drug-specific cytotoxic T cells. *J Allergy Clin Immunol* **114**, 1209-1215 (2004). https://doi.org:10.1016/j.jaci.2004.07.047

12 Le Cleach, L. *et al.* Blister fluid T lymphocytes during toxic epidermal necrolysis are functional cytotoxic cells which express human natural killer (NK) inhibitory receptors. *Clin Exp Immunol* **119**, 225-230 (2000). https://doi.org:10.1046/j.1365-2249.2000.01119.x

13 Friedmann, P. S., Strickland, I., Pirmohamed, M. & Park, B. K. Investigation of mechanisms in toxic epidermal necrolysis induced by carbamazepine. *Arch Dermatol* **130**, 598-604 (1994).

14 Villada, G., Roujeau, J. C., Clerici, T., Bourgault, I. & Revuz, J. Immunopathology of toxic epidermal necrolysis. Keratinocytes, HLA-DR expression, Langerhans cells, and mononuclear cells: an immunopathologic study of five cases. *Arch Dermatol* **128**, 50-53 (1992). https://doi.org:10.1001/archderm.128.1.50

15 Heng, M. C. & Allen, S. G. Efficacy of cyclophosphamide in toxic epidermal necrolysis. Clinical and pathophysiologic aspects. *J Am Acad Dermatol* **25**, 778-786 (1991). https://doi.org:10.1016/s0190-9622(08)80969-3

16 Correia, O., Delgado, L., Ramos, J. P., Resende, C. & Torrinha, J. A. Cutaneous T-cell recruitment in toxic epidermal necrolysis. Further evidence of CD8+ lymphocyte involvement. *Arch Dermatol* **129**, 466-468 (1993).

17 Viard, I. *et al.* Inhibition of toxic epidermal necrolysis by blockade of CD95 with human intravenous immunoglobulin. *Science* **282**, 490-493 (1998). https://doi.org:10.1126/science.282.5388.490

18 Paul, C. *et al.* Apoptosis as a mechanism of keratinocyte death in toxic epidermal necrolysis. *Br J Dermatol* **134**, 710-714 (1996). https://doi.org:10.1111/j.1365-2133.1996.tb06976.x

19 Saito, N. *et al.* An annexin A1-FPR1 interaction contributes to necroptosis of keratinocytes in severe cutaneous adverse drug reactions. *Sci Transl Med* **6**, 245ra295 (2014). https://doi.org:10.1126/scitranslmed.3008227

20 Viard-Leveugle, I. *et al.* TNF-alpha and IFN-gamma are potential inducers of Fas-mediated keratinocyte apoptosis through activation of inducible nitric oxide synthase in toxic epidermal necrolysis. *J Invest Dermatol* **133**, 489-498 (2013). https://doi.org:10.1038/jid.2012.330

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

21 Chung, W. H. *et al.* Granulysin is a key mediator for disseminated keratinocyte death in Stevens-Johnson syndrome and toxic epidermal necrolysis. *Nat Med* **14**, 1343-1350 (2008). https://doi.org:10.1038/nm.1884

22 Bruggen, M. C. *et al.* Supportive care in the acute phase of Stevens-Johnson syndrome and toxic epidermal necrolysis: an international, multidisciplinary Delphi-based consensus. *Br J Dermatol* **185**, 616-626 (2021). https://doi.org:10.1111/bjd.19893

23 Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* **24**, 494-515 (2023). https://doi.org:10.1038/s41576-023-00580-2

24 Mund, A., Brunner, A. D. & Mann, M. Unbiased spatial proteomics with single-cell resolution in tissues. *Mol Cell* **82**, 2335-2349 (2022). https://doi.org:10.1016/j.molcel.2022.05.022

25 Rosenberger, F. A. *et al.* Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome. *bioRxiv*, 2022.2012.2003.518957 (2022). https://doi.org:10.1101/2022.12.03.518957

26 Mund, A. *et al.* Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat Biotechnol* **40**, 1231-1240 (2022). https://doi.org:10.1038/s41587-022-01302-5

27 Nguyen, T. T. T. *et al.* Tryptophanyl-tRNA Synthetase 1 Signals Activate TREM-1 via TLR2 and TLR4. *Biomolecules* **10** (2020). https://doi.org:10.3390/biom10091283

28 Lerner, L. H., Qureshi, A. A., Reddy, B. V. & Lerner, E. A. Nitric oxide synthase in toxic epidermal necrolysis and Stevens-Johnson syndrome. *J Invest Dermatol* **114**, 196-199 (2000). https://doi.org:10.1046/j.1523-1747.2000.00816.x

29 Quinn, A. M. *et al.* Uncovering histologic criteria with prognostic significance in toxic epidermal necrolysis. *Arch Dermatol* **141**, 683-687 (2005). https://doi.org:10.1001/archderm.141.6.683

30 Ye, L. *et al.* In vivo CRISPR screening in CD8 T cells with AAV-Sleeping Beauty hybrid vectors identifies membrane targets for improving immunotherapy for glioblastoma. *Nat Biotechnol* **37**, 1302-1313 (2019). https://doi.org:10.1038/s41587-019-0246-4

31 Kim, D. *et al.* Targeted therapy guided by single-cell transcriptomic analysis in drug-induced hypersensitivity syndrome: a case report. *Nat Med* **26**, 236-243 (2020). https://doi.org:10.1038/s41591-019-0733-7

32 Keenan, C. R. *et al.* Polycomb repressive complex 2 is a critical mediator of allergic inflammation. *JCI Insight* **4** (2019). https://doi.org:10.1172/jci.insight.127745

33 Uyttenhove, C. *et al.* Evidence for a tumoral immune resistance mechanism based on tryptophan degradation by indoleamine 2,3-dioxygenase. *Nat Med* **9**, 1269-1274 (2003). https://doi.org:10.1038/nm934

34 Chaudhuri, B., Xu, H., Todorov, I., Dutta, A. & Yates, J. L. Human DNA replication initiation factors, ORC and MCM, associate with oriP of Epstein-Barr virus. *Proc Natl Acad Sci U S A* **98**, 10085-10089 (2001). https://doi.org:10.1073/pnas.181347998

35 Linkermann, A., Stockwell, B. R., Krautwald, S. & Anders, H. J. Regulated cell death and inflammation: an auto-amplification loop causes organ failure. *Nat Rev Immunol* **14**, 759-767 (2014). https://doi.org:10.1038/nri3743

36 Levy, D. E., Kessler, D. S., Pine, R., Reich, N. & Darnell, J. E., Jr. Interferon-induced nuclear factors that bind a shared promoter element correlate with positive and negative transcriptional control. *Genes Dev* **2**, 383-393 (1988). https://doi.org:10.1101/gad.2.4.383

37 Tretina, K., Park, E. S., Maminska, A. & MacMicking, J. D. Interferon-induced guanylate-binding proteins: Guardians of host defense in health and disease. *J Exp Med* **216**, 482-500 (2019). https://doi.org:10.1084/jem.20182031

38 Sung, Y., Yoon, I., Han, J. M. & Kim, S. Functional and pathologic association of aminoacyl-tRNA synthetases with cancer. *Exp Mol Med* **54**, 553-566 (2022). https://doi.org:10.1038/s12276-022-00765-5

39 Lewis, C. E., McCarthy, S. P., Lorenzen, J. & McGee, J. O. Differential effects of LPS, IFN-gamma and TNF alpha on the secretion of lysozyme by individual human mononuclear phagocytes: relationship to cell maturity. *Immunology* **69**, 402-408 (1990).

40 Liao, W. *et al.* A novel anti-apoptotic role for apolipoprotein L2 in IFN-gamma-induced cytotoxicity in human bronchial epithelial cells. *J Cell Physiol* **226**, 397-406 (2011). https://doi.org:10.1002/jcp.22345

41 Wang, F. *et al.* Diverse expression of TNF-alpha and CCL27 in serum and blister of Stevens-Johnson syndrome/toxic epidermal necrolysis. *Clin Transl Allergy* **8**, 12 (2018). https://doi.org:10.1186/s13601-018-0199-6

42 Wustner, L. S. *et al.* Generating iPSCs with a High-Efficient, Non-Invasive Method-An Improved Way to Cultivate Keratinocytes from Plucked Hair for Reprogramming. *Cells* **11** (2022). https://doi.org:10.3390/cells11121955

medRxiv preprint doi: https://doi.org/10.1101/2023.11.11.23295492; this version posted November 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

43 Anderton, H., Rickard, J. A., Varigos, G. A., Lalaoui, N. & Silke, J. Inhibitor of Apoptosis Proteins (IAPs) Limit RIPK1-Mediated Skin Inflammation. *J Invest Dermatol* **137**, 2371-2379 (2017). https://doi.org:10.1016/j.jid.2017.05.031

44 Saito, N. *et al.* Stevens-Johnson syndrome/toxic epidermal necrolysis mouse model generated by using PBMCs and the skin of patients. *J Allergy Clin Immunol* **131**, 434-441 e431-439 (2013). https://doi.org:10.1016/j.jaci.2012.09.014

45 Varfolomeev, E. *et al.* IAP antagonists induce autoubiquitination of c-IAPs, NF-kappaB activation, and TNFalpha-dependent apoptosis. *Cell* **131**, 669-681 (2007). https://doi.org:10.1016/j.cell.2007.10.030

46 Vasilikos, L., Spilgies, L. M., Knop, J. & Wong, W. W. Regulating the balance between necroptosis, apoptosis and inflammation by inhibitors of apoptosis proteins. *Immunol Cell Biol* **95**, 160-165 (2017). https://doi.org:10.1038/icb.2016.118

47 Chung, W. H. & Hung, S. I. Recent advances in the genetics and immunology of Stevens-Johnson syndrome and toxic epidermal necrosis. *J Dermatol Sci* **66**, 190-196 (2012). https://doi.org:10.1016/j.jdermsci.2012.04.002

48 Bastuji-Garin, S. *et al.* SCORTEN: a severity-of-illness score for toxic epidermal necrolysis. *J Invest Dermatol* **115**, 149-153 (2000). https://doi.org:10.1046/j.1523-1747.2000.00061.x

49 Bieber, T. *et al.* Abrocitinib versus Placebo or Dupilumab for Atopic Dermatitis. *N Engl J Med* **384**, 1101-1112 (2021). https://doi.org:10.1056/NEJMoa2019380

50 Zhang, C., Van, D. N., Hieu, C. & Craig, T. Drug-induced severe cutaneous adverse reactions: Determine the cause and prevention. *Ann Allergy Asthma Immunol* **123**, 483-487 (2019). https://doi.org:10.1016/j.anai.2019.08.004

51 Kinoshita, M. *et al.* Neutrophils initiate and exacerbate Stevens-Johnson syndrome and toxic epidermal necrolysis. *Sci Transl Med* **13** (2021). https://doi.org:10.1126/scitranslmed.aax2398

52 Miller, L. S. & Modlin, R. L. Human keratinocyte Toll-like receptors promote distinct immune responses. *J Invest Dermatol* **127**, 262-263 (2007). https://doi.org:10.1038/sj.jid.5700559

53 Dror, E. *et al.* Epigenetic dosage identifies two major and functionally distinct beta cell subtypes. *Cell Metab* **35**, 821-836 e827 (2023). https://doi.org:10.1016/j.cmet.2023.03.008

54 Wang, S. A. *et al.* Targeted next-generation sequencing identifies a subset of idiopathic hypereosinophilic syndrome with features similar to chronic eosinophilic leukemia, not otherwise specified. *Mod Pathol* **29**, 854-864 (2016). https://doi.org:10.1038/modpathol.2016.75

55 Musette, P. & Janela, B. New Insights into Drug Reaction with Eosinophilia and Systemic Symptoms Pathophysiology. *Front Med (Lausanne)* **4**, 179 (2017). https://doi.org:10.3389/fmed.2017.00179

56 Bracken, A. P. *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J* **22**, 5323-5335 (2003). https://doi.org:10.1093/emboj/cdg542

57 Julia, E. & Salles, G. EZH2 inhibition by tazemetostat: mechanisms of action, safety and efficacy in relapsed/refractory follicular lymphoma. *Future Oncol* **17**, 2127-2140 (2021). https://doi.org:10.2217/fon-2020-1244

58 Hall, J. C. & Rosen, A. Type I interferons: crucial participants in disease amplification in autoimmunity. *Nat Rev Rheumatol* **6**, 40-49 (2010). https://doi.org:10.1038/nrrheum.2009.237

59 Ytterberg, S. R. *et al.* Cardiovascular and Cancer Risk with Tofacitinib in Rheumatoid Arthritis. *N Engl J Med* **386**, 316-326 (2022). https://doi.org:10.1056/NEJMoa2109927

60 Zhu, J. *et al.* Stevens-Johnson syndrome/toxic epidermal necrolysis in patients treated with immune checkpoint inhibitors: A safety analysis of clinical trials and FDA pharmacovigilance database. *EClinicalMedicine* **37**, 100951 (2021). https://doi.org:10.1016/j.eclinm.2021.100951

61 Nordmann, T. M. *et al.* A standardized and reproducible workflow for membrane glass slides in routine histology and spatial proteomics. *Mol Cell Proteomics*, 100643 (2023). https://doi.org:10.1016/j.mcpro.2023.100643

62 Meier, F. *et al.* diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat Methods* **17**, 1229-1236 (2020). https://doi.org:10.1038/s41592-020-00998-0

63 Brunner, A. D. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol Syst Biol* **18**, e10798 (2022). https://doi.org:10.15252/msb.202110798

64 Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* **17**, 41-44 (2020). https://doi.org:10.1038/s41592-019-0638-x

65 Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **47**, W199-W205 (2019). https://doi.org:10.1093/nar/gkz401

## 3.4. Transfer of technology on the example of immunopeptidomics

### Article 11. IMBAS-MS discovers organ-specific HLA peptide patterns in plasma.

*Mol Cell Proteomics*, 2024 (23, 1, 100689); DOI: 10.1016/j.mcpro.2023.100689

Maria Wahle[1], Marvin Thielert[1], Maximilian Zwiebel[1], Patricia Skowronek[1], Wen-Feng Zeng[1], Matthias Mann[1]

*[1]Department Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany.*

Concepts and sensitivity of sc-proteomics technology offer a valuable platform for exploring other areas where limitations in material availability remains a key constraint. In this paper, we exemplified this concept in immunopeptidomics, although most of the lessons learned can be applied in similar circumstances with field-specific changes. The immunopeptidome typically refers to the membrane bound form of HLA protein-peptide complexes (mHLA) accessible from cells or tissue samples, while there are also peptides bound to soluble HLA receptors (sHLA) circulating in plasma. We developed a highly sensitive and automated workflow for HLA peptide analysis by MS, termed immunopeptidomics by biotinylated antibodies and streptavidin (IMBAS-MS). To note, IMBAS-MS is also suitable to detect mHLA by applying an additional lysis step prior to immunopeptidomic enrichment, although this is not investigated here.

To accommodate similar sensitivity increases as in sc-proteomics, we designed the workflow in a 96-well plate format applicable for automatic liquid handling to ensure reproducibility and throughput. A key aspect was the incorporation of biotinylated antibodies. These are captured by magnetic streptavidin beads, eliminating the need for chemical crosslinking of the antibody to the bead matrix. This allowed us to simplify the workflow steps, leading to a miniaturized one-pot solution and thereby eliminating transfer steps, while requiring only a final molecular weight filtering step. Furthermore, low flow gradients increased identifications by 50%. A 200 µL plasma sample yielded 3,500 immunopeptides in DDA and close to 1,500 peptides were still detectable from 50 µL, which is significantly more than in conventional workflows. We also evaluated different library-based DIA analysis strategies. Immunopeptides are dependent on the MHC type which is highly polymorphic between humans. Therefore, sample-specific libraries are beneficial for identification and quantification. However, empirically constructing those libraries requires much measurement time and sample. Our laboratory developed a deep-learning software package for proteomics called AlphaPeptDeep, which established the concept of transfer-learning for sample-specific

libraries. Using this approach in our HLA examples, one immunopeptidomics run identified 5,000 immunopeptides with directDIA-tuned sample-specific libraries compared to 3,800 with an experimental DDA library, making the acquisition of additional runs unnecessary.

Deep immunopeptide profiles were acquired with up to 13,000 HLA peptides using fractionated plasma from a 5 mL starting amount. When we evaluated the origin of the identified immunopeptides, we found that the majority of organs displayed the same median frequency in tissue proteome than in the immunopeptidomics dataset, suggesting that sHLA are representative of the tissue proteome. An exception was brain, which showed lower frequency in presentation. Possible reasons may be the removal by the blood-brain barrier or a less likely presentation because of lower protein turnover. Moreover, we found that the immunopeptidome is temporally stable qualitatively and quantitatively in single donors. This is important as it potentially suggests a general healthy baseline for HLA peptidomes that can serve as a reference point for disease state comparisons.

In summary, we established an immunopeptidomics workflow for sHLA analysis optimized using sc-proteomics concepts. IMBAS-MS holds the promise to reveal insights into physiological and pathological processes that are presented to the immune system. This may facilitate the identification of disease-specific immunopeptides from a non-invasive sample. Additionally, the streamlined workflow may enable disease monitoring and thus personalized medicine, which is highly relevant for clinical applications.

***Contribution:***

Co-authorship, second-author. Maria Wahle and Matthias Mann conceptualized the study. Maria Wahle conducted the project as first author and designed all experiments. I provided feedback and guidance across the study. Together with Maria Wahle, I discussed data, interpretation and figures. I revised and evaluated the manuscript.

MCP TECHNOLOGICAL INNOVATION AND RESOURCES

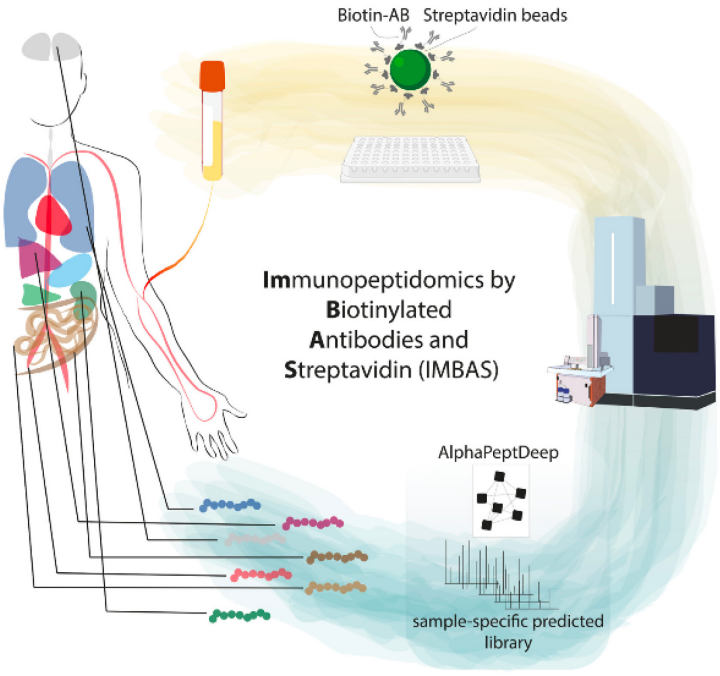
# IMBAS-MS Discovers Organ-Specific HLA Peptide Patterns in Plasma

**Authors**
Maria Wahle, Marvin Thielert, Maximilian Zwiebel, Patricia Skowronek, Wen-Feng Zeng, and Matthias Mann

**Correspondence**
mmann@biochem.mpg.de

**In Brief**
Soluble HLA (sHLA) peptides are thought to have great potential as biomarkers for different diseases such as cancer. However, efficient enrichment of sHLA peptides from body fluids like plasma has been a bottleneck, requiring 3 to 5 ml per person. Here, we present IMBAS-MS, an efficient and automated workflow for the enrichment of sHLA peptides from only 200 µl plasma. In combination with data-independent acquisition (DIA) and computational library-based analysis, we acquire deep immunopeptidomes from single measurements, elucidating its origins.

**Graphical Abstract**




**Highlights**

- Automated one-pot enrichment of soluble HLA peptides using magnetic beads (IMBAS).
- Personalized computational library generation for DIA analysis using AlphaPeptDeep.
- IMBAS-MS identifies more than 5000 HLA peptides form 200 µl of plasma.
- Soluble HLA peptides represent a variety of tissues including the brain.

2024, Mol Cell Proteomics *23(1)*, 100689
© 2023 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
https://doi.org/10.1016/j.mcpro.2023.100689

MCP TECHNOLOGICAL INNOVATION AND RESOURCES

# IMBAS-MS Discovers Organ-Specific HLA Peptide Patterns in Plasma

**Maria Wahle, Marvin Thielert, Maximilian Zwiebel, Patricia Skowronek, Wen-Feng Zeng, and Matthias Mann***

**Distinction of non-self from self is the major task of the immune system. Immunopeptidomics studies the peptide repertoire presented by the human leukocyte antigen (HLA) protein, usually on tissues. However, HLA peptides are also bound to plasma soluble HLA (sHLA), but little is known about their origin and potential for biomarker discovery in this readily available biofluid. Currently, immunopeptidomics is hampered by complex workflows and limited sensitivity, typically requiring several mL of plasma. Here, we take advantage of recent improvements in the throughput and sensitivity of mass spectrometry (MS)-based proteomics to develop a highly sensitive, automated, and economical workflow for HLA peptide analysis, termed Immunopeptidomics by Biotinylated Antibodies and Streptavidin (IMBAS). IMBAS-MS quantifies more than 5000 HLA class I peptides from only 200 µl of plasma, in just 30 min. Our technology revealed that the plasma immunopeptidome of healthy donors is remarkably stable throughout the year and strongly correlated between individuals with overlapping HLA types. Immunopeptides originating from diverse tissues, including the brain, are proportionately represented. We conclude that sHLAs are a promising avenue for immunology and potentially for precision oncology.**

The immune system relies on the human leukocyte antigen (HLA) peptide-protein complex to present immunopeptides derived from both endogenous and exogenous sources to circulating T-cells. These immunopeptides play a crucial role in immune surveillance, as they enable the elimination of abnormal or infected cells. Upon recognition of a peptide–HLA protein complex by cytotoxic T lymphocytes (CTLs), downstream cascades are activated, causing the presenting cell to undergo apoptosis. This biological principle is exploited in immunotherapeutic strategies such as CAR T-cell treatment or mRNA peptide vaccination (1). A crucial but challenging step is the identification of peptides specifically presented by tumor cells. Most efforts have focused on the enrichment of membrane-bound human leukocyte antigen (mHLA) receptors with their bound immunopeptides from tumor tissue. This is followed by mass spectrometric identification in search of tumor-specific antigens or neo-epitopes (2, 3).

In addition to the membrane-anchored HLA proteins, a soluble fraction (sHLA) can be enriched from blood and other body fluids (4). sHLAs are thought to arise by shedding, cleavage from the cell surface, or the expression of a splicing variant lacking the membrane anchoring domain (5). Although their function and exact release mechanisms remain unclear, it is known to change in a disease context (6). In the case of cancer, a disproportionate fraction of peptides in relation to the total tumor mass may originate from tumor tissue (4), representing a potential additional source for tumor antigens (7). Whereas the analysis of HLA peptides from mHLAs requires substantial tissue amounts from surgery, sHLAs are rather easily accessible through minimally invasive procedures like regular blood withdrawal without placing additional burden on patients. However, despite the clear potential for disease diagnosis and treatment monitoring in a clinical setup (8–10), there are only a few studies investigating sHLAs, and they typically only identified them from milliliters of plasma (4, 7, 8).

Beyond clinical applications, another attraction of the sHLA immunopeptidome is that it can serve as an unlimited source of native immunopeptides from diverse HLA backgrounds. As importantly, an extensive repertoire of sHLA peptides from a diversity of healthy donors could potentially serve as a resource to improve the general knowledge about peptide processing and presentability (10, 11). In contrast, due to limited analytical sensitivity and tissue accessibility mHLA data is currently restricted to a few hundred different alleles, with on average around 5000 up to 160,000 (HLA-A0201) associated MHC peptides identified by mass spectrometry (12).

In this study, we address the limitations of sHLA immunopeptidomics and characterize its nature over time in healthy donors. We describe an automated workflow for efficient 'one-pot' enrichment of HLA immunopeptides followed by ultra-high sensitive mass spectrometry (MS), termed IMBAS-MS for Immunopeptidomics by Biotinylated Antibodies and

From the Department Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany
*For correspondence: Matthias Mann, mmann@biochem.mpg.de.



© 2023 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.1016/j.mcpro.2023.100689

Streptavidin. Quantifying immunopeptides from a few hundred microliters showed that the immunopeptidome is stable in healthy individuals for over a year and exhibits high reproducibility between overlapping HLA types. IMBAS-MS allowed profiling of sHLA to a depth of over 10,000 peptides per person, which were broadly representative of different tissues.

## EXPERIMENTAL PROCEDURES

### *Plasma Collection from Healthy Donors*

Plasma was obtained by withdrawing blood from eight healthy donors (specified in supplemental Table S1) in EDTA tubes (BD Vacutainer K2E; REF 367525). The tubes were inverted three times and centrifuged at 2000*g* for 20 min at 4 °C. The plasma was separated, aliquoted, snap-frozen, and stored at −80 °C.

To determine the HLA types of healthy donors, genomic DNA was extracted from buccal swabs. Cotton swabs were used to obtain oral mucosa samples, placed in protease buffer (30 mM Tris-HCl, pH 8, 0.5% Tween-20, 0.5% IGEPAL CA-630), and proteins were digested with 100 µg proteinase K per sample for 12 min at 50 °C in a thermal shaker. The enzyme was inactivated at 75 °C for 30 min and genomic DNA was purified using the QIAamp DNA Micro Kit (Qiagen). HLA class I (HLA-A, HLA-B, HLA-C) and class II (DRB1, DQB1, DPB1) loci were amplified using the NGSgo-MX6-1 kit (GenDx), and multiplexed sequencing libraries were prepared using the Ultra II FS DNA Library Prep Kit for Illumina (NEB). Libraries were sequenced on a NovaSeq 6000 system (Illumina) in 150 bp paired-end mode and the genotype data were analyzed using the NGSengine software (GenDx). Blood was sampled from healthy donors, who provided written informed consent, with prior approval of the ethics committee of the Max Planck Society in accordance with the Declaration of Helsinki principles.

### *Affinity Purification of HLA Molecules*

Plasma samples were thawed on ice, diluted 1:2 with PBS (Gibco), and incubated overnight, shaking at 4 °C with variable amounts of biotinylated W6/32 antibody (custom produced by inVivo Bioscience). Each injection was processed in a separate well of a 96-well plate. Captured HLA molecules were enriched using magnetic streptavidin beads (ReSyn Bioscience) and washed first with 100 µl of 150 mM NaCl in 10 mM Tris pH 8.5, then 100 µl of 450 mM NaCl in 10 mM Tris pH 8.5 and finally 100 µl of 10 mM Tris pH 8.5 at 4 °C. The sHLA molecules were eluted from the beads using 150 µl elution buffer (200 mM glycine pH 2), transferred into prewetted 10 kDa MWCO plates (Millipore), and filtered at 4000*g* for 20 min. The flowthrough was directly loaded onto Evotips Pure following the recommended standard procedure. Briefly, Evotips were activated by 1-propanol, washed two times with 50 µl buffer B (99.9% ACN, 0.1% FA) and two times with 50 µl buffer A (99.9% ddH2O, 0.1% FA). 70 µl buffer A was briefly spined on the disks and sample elution was loaded by 80 s centrifugation. Evotips were then washed with 50 µl buffer A and stored with buffer A on top. All centrifugation steps were performed at 700*g* for 1 min. The whole protocol was performed in a semi-automated fashion using the Agilent Bravo liquid handling platform.

### *Immunopeptide Fractionation*

To acquire deep fractionated immunopeptidomes, 5 ml of plasma per individual was thawed at once, distributed in ten wells of a 96-deep-well plate, and processed with the same workflow as described above. The final elutions were combined in one single peptide pool without centrifugation. Fractionation was carried out on an AssayMAP Bravo Sample Prep Platform (Agilent), using the Fractionation v1.1 Protocol in the Protein Sample Prep Workbench v3.2.0 with standard settings. 5 µl C18 Cartridges (Agilent) were used as solid phase, six elution fractions were collected, using high-pH buffers with increasing acetonitrile concentrations (ammonium hydroxide solution, pH 10; 7, 12, 15, 23, 30, and 40% acetonitrile, respectively). The 40% acetonitrile buffer was used for priming and 7% for equilibration of the cartridges.

### *DDA and DIA LC-MS Acquisition*

Peptides were separated with the Evosep One LC system using predefined gradients as mentioned in each section. The majority of data was acquired using the Whisper40 method over a 15 cm Aurora Elite CSI column (AUR3-15075C18-CSI, IonOpticks) at 50 °C inside a nanoelectrospray ion source (Captive spray source, Bruker). The mobile phases were 0.1% FA in LC-MS grade water (buffer A) and 99.9% ACN with 0.1% FA (buffer B). For gradient testing, the 15SPD, 30SPD, and 60SPD method was used in combination with a 15 cm PepSep (150 um ID and 1.5 um bead size, Bruker) connected to a 10um ID ZDV emitter (Bruker). The LC system was coupled to a timsTOF Ultra instrument (Bruker).

When operated in dda-PASEF mode, a 10 PASEF/MSMS scan per topN acquisition method was used with a precursor signals intensity threshold at 500 arbitrary units. Standard polygon settings were adapted in the m/z-IM plane to exclude adverse ions, but include single-charge precursors based on their expected position in an m/z-IM plane. The mass spectrometer was operated in sensitivity mode with an accumulation and ramp time of 100 ms. Precursors were isolated with a 2 Th window below m/z 700 and 3 Th above and actively excluded for 0.4 min when reaching a target intensity threshold of 20,000 arbitrary units. A range from 100 to 1700 m/z and 0.6 to 1.6 Vs cm-2 was covered with collision energy from 20 eV at 0.6 Vs cm-2 ramped linearly to 59 eV at 1.6 Vs cm-2.

When operating in dia-PASEF mode, we used optimal dia-PASEF methods generated with our Python tool py_diAID (13). These dia-PASEF methods optimally cover the precursor cloud in m/z-IM plane, while being highly efficient with 1.17 s cycle time. We generated acquisition schemes specifically for dominant HLA types that cover up to 99.9% of all precursor species including singly charged ions, with 8 dia-PASEF scans, where each scan is divided into two ion mobility windows. The method covers precursors within 300 to 1200 Da. Other settings remained the same as for dda-PASEF.

### *Raw Data Analysis*

DDA data was analyzed using FragPipe 19.1 with the nonspecific HLA workflow against a human fasta containing 40,818 entries (UP000005640 + 50% decoy, downloaded 31.01.2022) (14–16). The precursor and fragment mass tolerance were kept as the default values of 20 ppm, cystinylation was set as a variable modification and results were filtered applying a peptide FDR of 1%. The quality of identified peptides was assessed using MHCVizPipe (v0.7.11) (17).

DIA data was analyzed using DIA-NN version 1.8.1 with standard settings searching against sample-specific predicted libraries generated using the AlphaPeptDeep package (https://github.com/MannLabs/alphapeptdeep) together with the peptdeep_hla (https://github.com/MannLabs/PeptDeep-HLA) DL model. Precursor and fragment mass tolerance were determined automatically for each run separately and ranged from 10 to 20 ppm. Sample specific immunopeptide lists (from DDA analysis or directDIA results generated using Spectronaut 17) were used to tune an immunopeptide deep learning model which reports a peptide list with a high likelihood of being a presented immunopeptide in this sample from an unspecific *in silico* digest of a human fasta. The peptide list of each individual is used for transfer learning to predict individual- or sample-specific spectral library. Library sizes are adjustable using a precision cutoff

(probability ≥ 0.7) in the peptdeep_hla DL model. Exact library sizes are detailed in supplemental Table S1. The 'report.tsv' table was used for further analysis.

### *Statistical Analysis*

All data analysis was performed using R. Unless stated differently, only peptides predicted to bind to any of the donors' HLA-types were used for downstream analysis. Peptides were defined to be binders as provided by MHCVizPipe interfacing NetMHCpan 4.1. Binder Frequency (BF) scores were reported as provided by MHCVizPipe. The BF score describes the fraction of peptides predicted to bind the provided HLA Alleles within the expected length range. The UpSet plots were generated using a custom function only displaying intersections larger than 10% of the smallest dataset

### *Experimental Design and Statistical Rational*

All experiments were done using human plasma obtained as described above. Altogether, the dataset including raw data files and search results was uploaded to MassIVE (see below). We used the same plasma batch for the benchmarking and technical evaluation. In brief, measurements with different gradients, input amounts, or from different donors were acquired in triplicates unless mentioned differently. The experimental design and statistical rational are described in the respective figure legends. Workflow replicates were acquired to evaluate reproducibility and quantitative accuracy.

## RESULTS

### *IMBAS-MS Design and Evaluation*

Earlier sHLA workflows have used several milliliters of plasma for enrichment, limiting throughput and applicability. We aimed to develop a workflow that is highly reproducible, sensitive, and allows for deep immunopeptidome coverage, without neglecting throughput and cost. To accommodate all these aspects, we optimized IMBAS (Immunopeptidomics by Biotinylated Antibodies and Streptavidin) in a 96-well format that could be processed in parallel. We automated the immunoaffinity enrichment on a Bravo liquid handling robot (Agilent) with less than 2 h of hands-on time for the entire procedure (Fig. 1*A*). The workflow was designed to be flexible, thus the enrichment can either be performed by hand or by any robot with a magnetic plate and a cooled plate station. IMBAS-MS is modular and although not demonstrated here, can directly be applied to cell lysates or biopsy samples by adding a homogenization and lysis step up front. The enrichment, washing, and elution steps take place within the same well, minimizing transfer steps and reducing sample loss due to plastic contact. A key aspect of IMBAS is the replacement of the commonly used ProteinA/G-IgG domain interaction between the antibody and bead matrix. To achieve this, we chose to use biotinylated antibodies which can be captured with streptavidin beads. The high specificity and stability of the streptavidin-biotin interaction allow to omit chemical crosslinking of the desired anti-HLA-antibody to the slurry upfront of the enrichment protocol, saving time and material while still eliminating a plasma preclearance step. Following the enrichment, the eluent is molecular-weight filtered and the resulting, separated peptides are loaded onto Evotips. In this way, an entire 96-well plate can be prepared and MS-data acquired within 3 days with minimal reagent preparation and cost. For technical details of the protocol see Experimental Methods.

To evaluate IMBAS-MS, we first identified and quantified immunopeptides from 200 µl plasma from the same donor at different HPLC flowrates and gradient lengths, from 21 min up to 88 min long gradients (termed 60–15 Samples Per Day (SPD)). The standard methods on the Evosep system have 1 µl/min down to 0.22 µl/min flow rates with throughputs of 15, 30, and 60 SPD. We reasoned that the recently introduced very low flow gradients of only 100 nl/min (Whisper gradients 20 or 40) that had substantially boosted sensitivity for single-cell analysis (18) would also be beneficial for HLA peptides. Indeed, the nanoflow gradients substantially outperformed the standard gradients with more than 3000 immunopeptide identifications in data-dependent acquisition (DDA) (Fig. 1*B*). The Whisper20 gradient identified only 10% more peptides than the Whisper20 gradient, at the cost of doubled measurement time (Fig. 1*B*). With a focus on maximizing depth and throughput, we chose the Whisper40 gradient (31 min length) for all subsequent experiments.

A key advantage of our workflow is that it needs much less plasma input than the milliliters used before. To investigate input requirements and tradeoffs, we enriched sHLA from only 10 µl up to 500 µl of plasma. Volumes from 10 µl to 100 µl plasma can be processed in a standard 96-well plate. They required only 1 µg of antibody for efficient enrichment and did not benefit from increasing the antibody amount 10-fold (Fig. 1*C*). For higher volumes—for instance 200 µl—higher antibody amounts boosted immunopeptide identifications about 20%. Over the entire tested input range, we identified from 500 to 4500 immunopeptides in data-dependent acquisition (DDA).

To investigate the purity of our immunopeptidomes, we inspected their length distribution, their calculated binder scores, and the presence of singly charged precursors. Identified peptides retained expected features such as a strong preference for nonameric peptides and a significant proportion of singly charged precursors (supplemental Fig. S1). The fraction of peptides with high binding scores (BF) within the expected length range was 0.9, further indicating high purity of the enriched and identified peptides (see Experimental Procedures).

IMBAS-MS also demonstrates high quantitative reproducibility between replicates at 500 µl (Pearson correlation of 0.97). Even with ten-fold reduced input, reproducibility is largely retained with a Pearson correlation of 0.81 (Fig. 1, *D* and *E*).

Based on the results above, in particular the purity and depth of the immunopeptide fraction, we chose a sample volume of 200 µl as an optimal combination for data quality, sample availability, ease of handling, and cost-effectiveness.

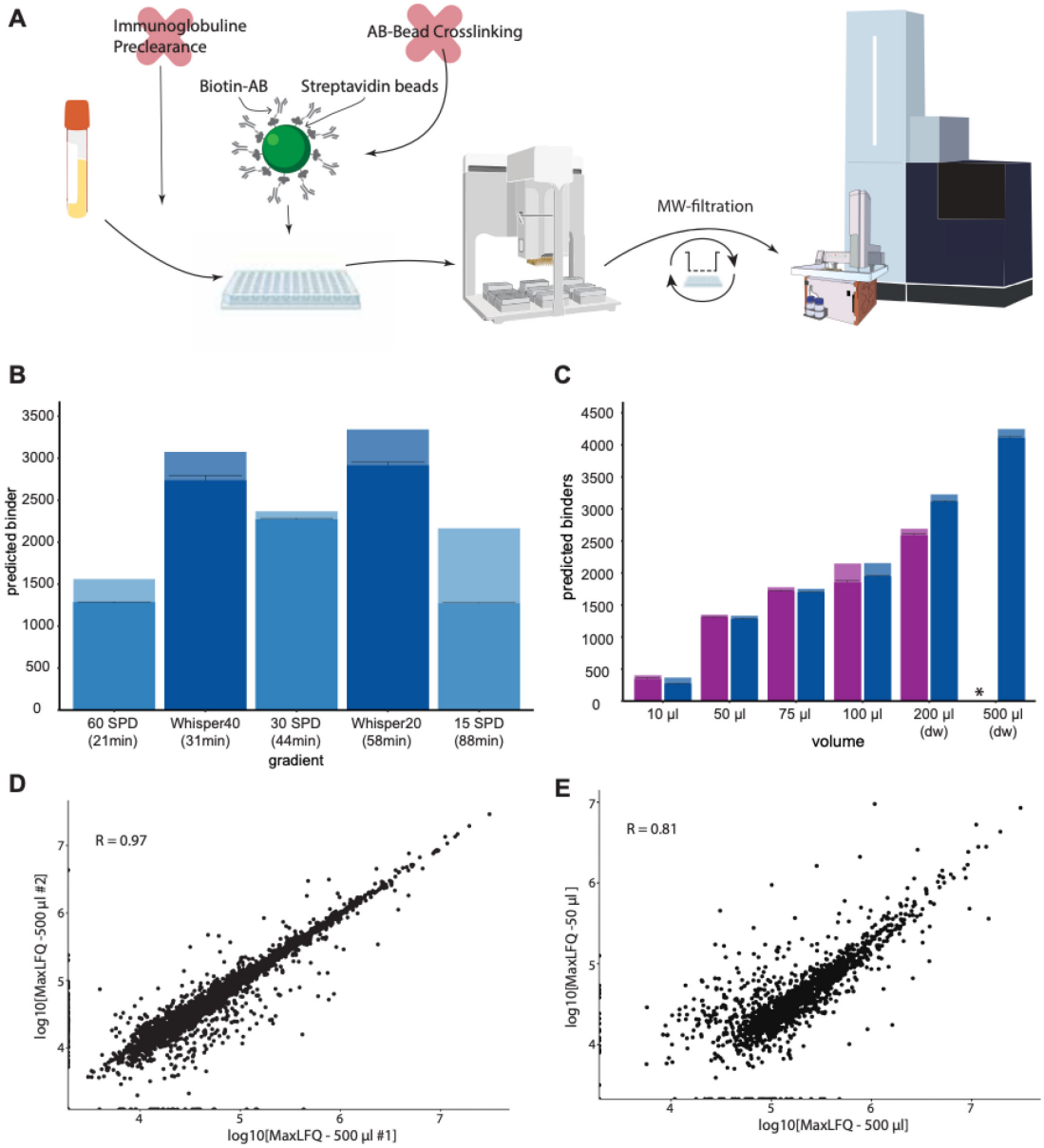



FIG. 1. **IMBAS-MS design and evaluation.** *A*, schematic representation of the sHLA IMBAS-MS workflow. Plasma is incubated with anti-HLA antibody (W6/32) and subjected to our automated bead-based enrichment workflow on a liquid handling robot (Agilent Bravo). Eluted peptides are loaded onto StageTips (Evotips) and measured by ultra-high sensitivity LC-MS/MS (Evosep and Bruker timsTOF Ultra). *B*, immunopeptide identifications from enrichment of 200 µl plasma using different gradient types and lengths. Low-flow gradients (Whisper40 and 20 on the Evosep, *dark blue*) show an increased sensitivity by identifying over 3000 immunopeptides. The total uniquely identified immunopeptides per triplicate (*light bar*) as well as the average and standard deviation (SD) per replicate (*black lines*) is plotted. *C*, evaluation of various input amounts using 1 ug (*purple*) or 10 ug (*blue*) of W6/32 antibody. The '*' indicates that this data point was not acquired due to plasma/antibody requirements. For unique peptide number determination see *B*); dw: deep-well plate. *D*, quantification correlation shows high reproducibility between workflow replicates using 500 µl plasma. *E*, Pearson correlation between 50 µl and 500 µl of the same plasma sample.

*Predicted Library-Based DIA for Immunopeptidomics*

Having evaluated IMBAS-MS with data-dependent acquisition (DDA) methods, we set out to couple it with data-independent acquisition (DIA) based mass spectrometry, which promises much greater depth and higher data completeness between experiments (19, 20). A major challenge for efficient and comprehensive analysis of DIA data is the selection or generation of a suitable spectral library. Three different strategies are commonly used: experimental libraries, typically acquired by DDA; pseudospectra-based libraries extracted by directDIA as introduced by DIA-Umpire (21) and implemented in Spectronaut; and libraries in which fragment intensities are predicted by deep learning (19, 22). In connection with the latter approach, we recently introduced a deep learning-based framework called AlphaPeptDeep, which predicts spectral libraries tailored for different MS platforms, only based on a database file of the proteome in FASTA format or just a peptide list as input (23). It contains the PeptDeep-HLA model which makes use of the inherent similarity of immunopeptides present within one person based on their HLA type. Given a preliminary list of identified peptides, this package then predicts a large subset of HLA peptides that are potentially present in this allelotype(s). Here, we compared three different modes of library generation in AlphaPeptDeep, purely experimental libraries and pseudospectra-extracted libraries (Fig. 2*A*).

First, we built an experimental DDA library using MSFragger (14–16) on the above-described dilution series files, which resulted in nearly 7000 identified precursors. Spectronaut internally builds a directDIA extracted list of peptides, in this case containing around 2000 identified precursors from the replicates searched in parallel. Using AlphaPeptDeep, we predicted the fragment intensities of the set of immunopeptides constituting the 'pan library', with around 385,000 precursors (MSV000084172;PXD004894). The PeptDeep-HLA model only needs about 1000 identified immunopeptides to learn how to extract potential immunopeptides from a FASTA for the allelotypes in question. For this, we compare two strategies, using either peptides identified in DDA experiments or from a directDIA search of the same file. On this basis, AlphaPeptDeep generated two large libraries (about 1 and 0.5 million precursors, respectively). Figure 2*B* compares the different DIA library sizes and their overlap.

With the exception of the directDIA strategy as currently implemented in Spectronaut, DIA always substantially outperformed DDA, as expected (Fig. 2*C*). Although a directDIA-based analysis strategy as a standalone solution is not able to outperform a DDA immunopeptidomics analysis, a combination of directDIA with AlphaPeptDeep increased the depth by up to 67% compared to the DDA experiment (Fig. 2*C*). Importantly, this increase in depth did not come at the expense of the quality of the data, as judged by the peptide length distribution and the binder scores which ranged from 0.9 for the experimental library to 0.98 for the DDA tuned library (Supplements). All three computational libraries, the pan library as well as the sample specific libraries outperform the experimental library while retaining the vast majority of peptides (Fig. 2*D*). Given the small difference between the results from predicting the library based on DDA data or extracted peptides by directDIA, we suggest that the latter strategy will be attractive for immunopeptidomics in the future as no additional DDA experiments are required any more.

We conclude that the combination of directDIA and AlphaPeptDeep enables us to acquire deep DIA-based immunopeptidomes from a single measurement of a sample.

*Deep Soluble Immunopeptidomics in Comparison to mHLA Immunopeptidomics*

Previous state-of-the-art reports used around 10 ml plasma per donor to reach a median depth of around 1000 unique immunopeptides with a maximum of 2500 for healthy donors (4, 7–9). With our sample-specific predicted spectral libraries and DIA acquisition IMBAS-MS, surpassed those results with just 2% of the input material (200 µl of plasma) (Fig. 3*A*). From one blood withdrawal that yielded 5 ml of plasma, we quantified up to 13,000 immunopeptides from six fractions per individual, six-fold higher than before (Fig. 3*A*). (Note that the previous studies employed Q Exactive instruments rather than the latest generation Bruker timsTOFs.)

In total, the experiment above covered nearly 40,000 unique immunopeptides from 28 different alleles in six donors (Fig. 3*B*, for distribution per person, see supplemental Fig. S2). This is a large number, as evidenced by the fact that in some cases (*e.g.* HLA-C0401) this data surpasses the number of epitopes reported by the community database Immune Epitope Database & Tools (IEDB) which for this allelotype contains around 5000 peptides identified by mass spectrometry compared to our 8000 (https://www.iedb.org/result_v3.php?cookie_id=13cd62).

Next, we used our in-depth data to assess whether the soluble immunopeptidome displayed similar physicochemical properties as those described for their membrane-bound equivalents.

We noted that immunopeptides identified from sHLA show a strong enrichment of nonameric peptides, similar to what is known for immunopeptides from membrane-bound equivalents (Fig. 3*C*).

Comparing the ratio of peptides which bind strongly (%Rank <0.5) to their cognate HLA protein to those that bind weakly (0.5< %Rank <2), we did not find significant differences in pairwise comparisons of HLA types present in our sHLA dataset or an mHLA dataset (2) (Fig. 3*D*). Given the easy accessibility of the sHLA peptidome by IMBAS-MS and its close correspondence to the mHLA peptidome, we conclude that plasma is a

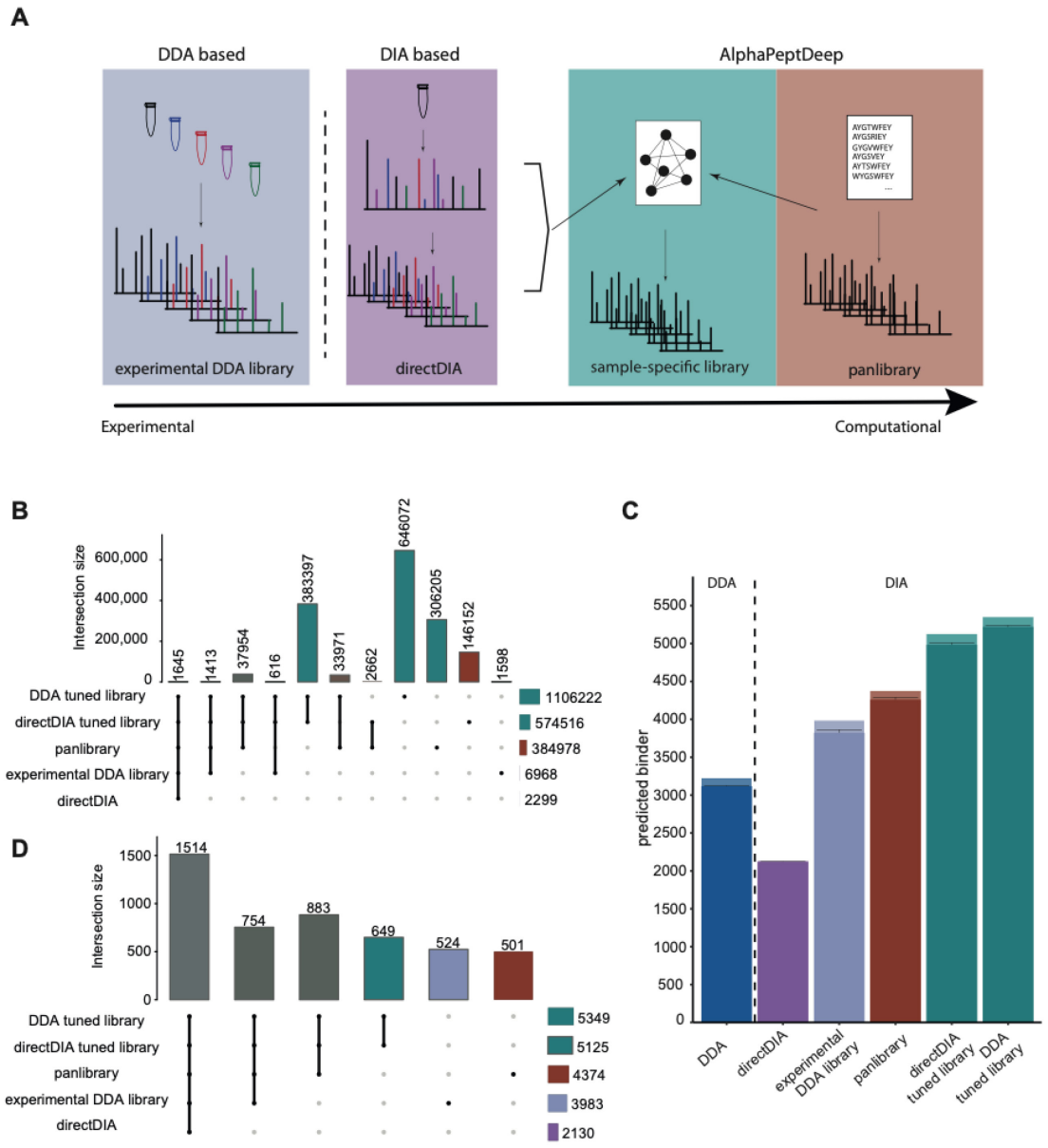



FIG. 2. **Evaluation of different strategies of predicted library-based DIA.** *A*, overview of the library generation process. *B*, overlap of experimental library, pseudospectra library (directDIA), pan library and AlphaPeptDeep predicted libraries. *C*, triplicate measurement of immunopeptides from 200 µl plasma analyzed using different DIA data analysis strategies. DDA identifications are shown for comparison (*left* of the *stippled vertical line*). Library matching strategies are described in the main text. The height of the bar represents overall uniquely identified immunopeptides per triplicate and the mean per run as a *horizontal black line* with standard deviation. *D*, overlap of identified peptides analyzed with different library strategies (*C*).

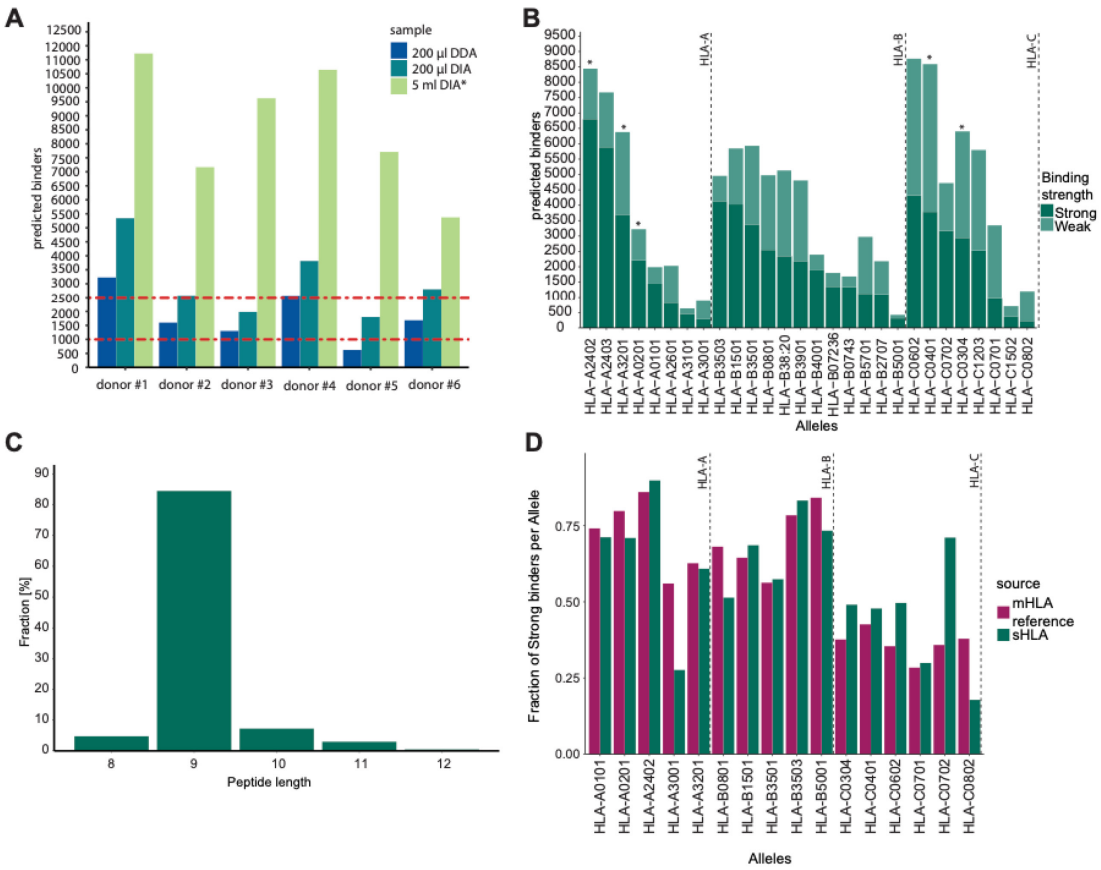



FIG. 3. **Deep soluble immunopeptidome retains properties described for mHLA based immunopeptidomes.** *A*, Immunopeptides identified by IMBAS-MS from 200 µl plasma by DDA (*blue*), DIA (*green*), or six separately analyzed fractions from 5 ml plasma (*light green*) from six healthy donors. For comparison the stippled red lines indicate the maximum and median identifications from 10 ml of 'non-cancerous plasma' (7).* The *asterisk* highlights that the 5 ml DIA runs were measured in six fractions. *B*, Strong or weak predicted binders from (*A*) across HLA-types. *Asterisks* mark types present in multiple people. *C*, length distribution of identified HLA class I peptides of all samples in (A). *D*, the fraction of strong binders for each allele present in our soluble HLA dataset from (*A*) compared to the fraction reported in a reference mHLA dataset (HLA-Ligandatlas (2)).

very attractive source to increase our general knowledge of immunopeptides not only in a disease background.

### *Tissue Origin of sHLA Peptides*

A fundamental and still an open question is to what degree each organ contributes to the sHLA peptidome and we reasoned that our deep and unbiased dataset on healthy donors could shed light on this. To infer the origin of the soluble immunopeptidome we compare our immunopeptidome data to a recent and deep proteome atlas of 29 healthy tissues (24). In that proteome dataset, each gene was classified into one of four groups, namely (i) expressed in all tissues, (ii) group enriched, (iii) tissue enhanced and (iv) tissue enriched. We transferred these classifications to our deep, fractionated sHLA peptidome to assign organ specificity to it. Next, we compared the frequency of group-enriched, tissue-enhanced, and tissue-enriched genes represented in the immunopeptidome to the frequency of those groups within the proteome of each tissue. Note that this assumes that proteins expressed in different organs have a similar chance of being presented by HLA proteins. That would make the fraction of genes assigned to each organ within the immunopeptidome a good estimate of the overall representation of that organ.

We observed that the median frequency of classified genes in our immunopeptidome dataset correlates well with the frequency of those within the organ proteome dataset ($R^2$ between 0.8 and 0.84). As an example, around 7.5% of all genes identified in the duodenum proteome were classified as group enriched (24). This is very close to the value of 8% of all genes in our soluble immunopeptidomes. As can be seen in Figure 4*A*, 'tissue enriched genes' are less frequent in the proteome and the immunopeptidome than genes belonging to the two less enriched groups.

A notable exception from the above general observation is enriched genes. They are represented at around 3% within the immunopeptidomes, while encompassing around 5% of the brain proteome (Fig. 4*A*). We speculate that sHLA-protein-peptide complexes may be partially filtered by the blood–brain barrier or that brain-specific genes are somewhat less likely to be presented, perhaps due to slow protein turnover. Another reason could be a lack of peptides originating from those proteins that have an affinity to one of the analyzed alleles in our dataset.

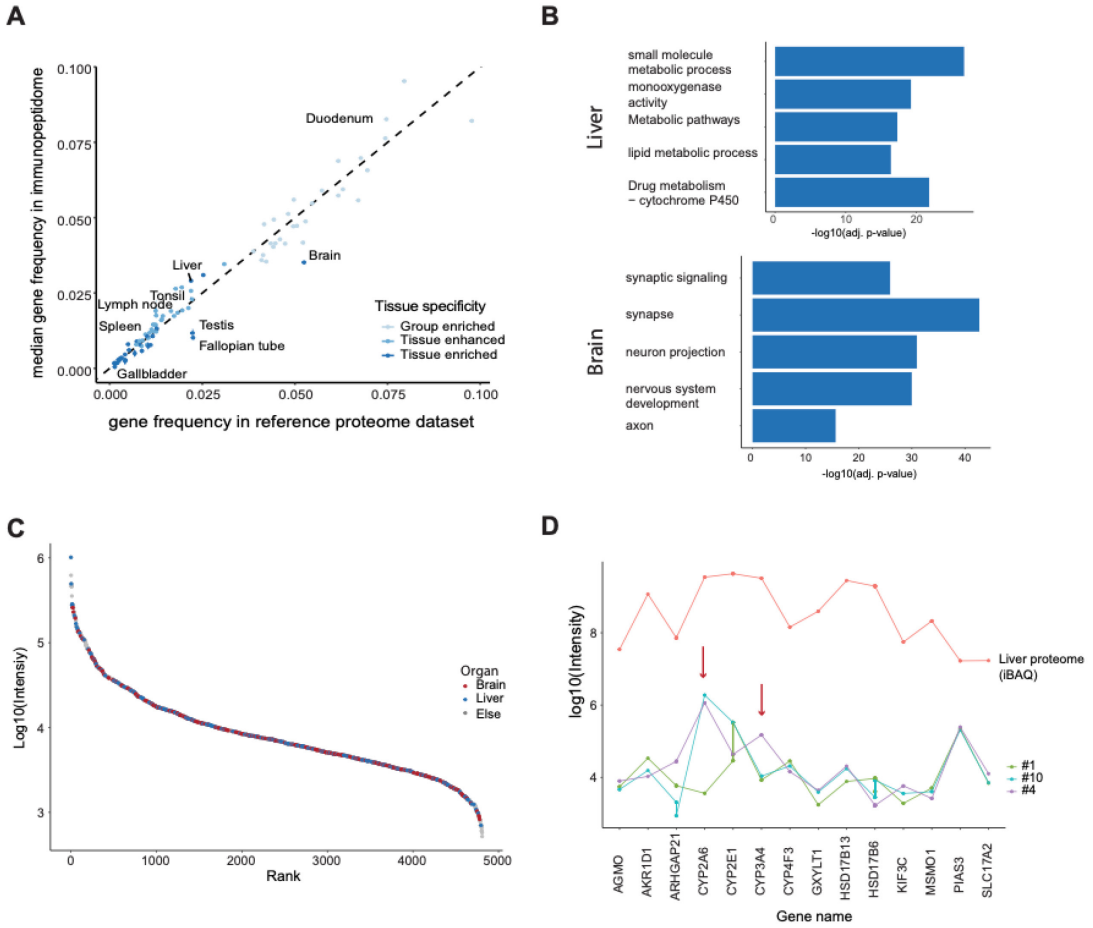



FIG. 4. **Tissue origin of immunopeptides.** *A*, comparison of the fraction of immunopeptides assigned to different organs as inferred from a reference proteome dataset (24). Genes and immunopeptides are classified as group enriched, tissue enhanced and tissue enriched depending on their degree of enrichment in the corresponding proteins reference proteome. The immunopeptidome dataset is the same as in Figure 3*A*. Points on or close to the diagonal indicate that the organ is equally represented in the peptidome and the proteome. *B*, GO-term enrichment of immunopeptides whose proteins were assigned as tissue enhanced in brain or liver from donor #1 (dataset from Fig. 3*A*, 200 µl DIA). Terms are representative of tissue specific functions. *C*, intensity rank plot of peptides derived from group enriched, tissue enhanced and tissue enriched genes colored by their respective organ assignment (dataset as in (*B*)). *D*, median intensity traces of selected liver tissue enriched immunopeptides for three donors. The iBAQ intensities for the corresponding genes from the reference proteome dataset are plotted for reference.

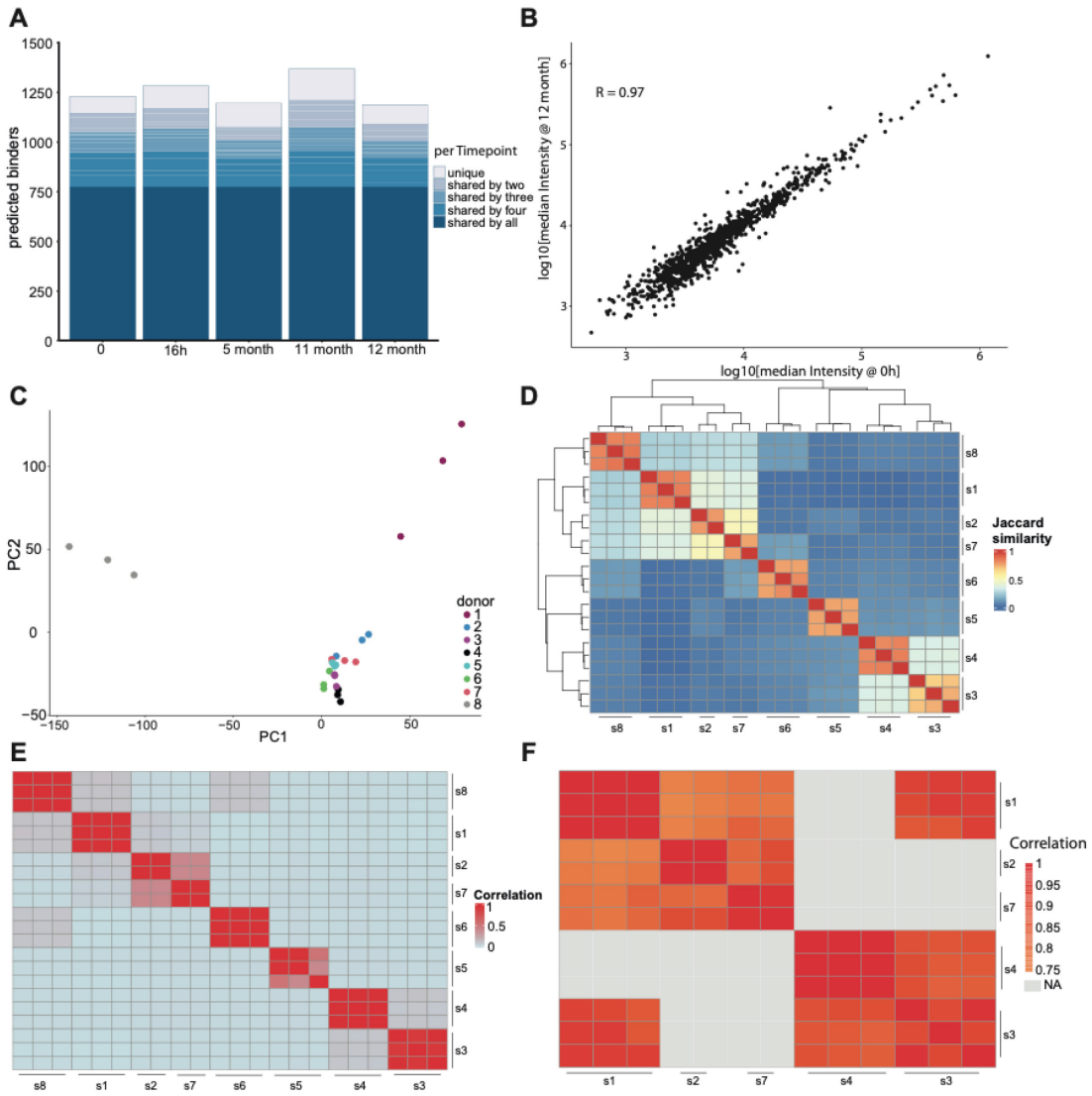



FIG. 5. **The plasma immunopeptidome is stable over time and quantitative reproducible between healthy controls.** *A*, immunopeptides identified by IMBAS-MS of a healthy donor over the course of a year. Two closely spaced time points at the start assess short term variation, and the 11-months time point is immediately post-COVID 19. Dark blue represents peptides shared between all timepoints and light blue represents peptides only measured at one time point. *B*, Pearson correlation of immunopeptide quantities between timepoint 0 and after 12 months. *C*, principal component analysis of immunopeptidomes from eight different healthy donors. Numbers refer to the different donors and colors represent the replicates. *D*, clustered heatmap of Jaccard similarities of immunopeptidomes between healthy donors (*B*). Note that 7 and 2 have only two replicates. *E*, unfiltered imputed Pearson correlation between healthy donors (*B*). The sample order was taken from the clusters built by Jaccard similarities in (*C*). *F*, filtered pairwise complete Pearson correlation of median intensities between donors showing a Jaccard similarity of more than 0.3 with at least one other donor (conditions without shared peptides are *grey*).

In addition, we assessed whether the immunopeptidome quantified from the measurement of only 200 µl has sufficient depth to infer organ-specific gene ontology enrichment terms. Indeed, applying the above gene classification strategy allows to discern organ function-specific gene sets represented by immunopeptides as illustrated for donor#1 for liver-enriched genes and brain-enriched genes (Fig. 4*B*).

Interestingly in view of the different sizes of the organs, on a quantitative level, peptides representing all organs are distributed equally over the peptide abundance range with no clear trend of specific organs being represented by more highly abundant species or vice versa. For example, both brain and liver-assigned immunopeptides cover all three orders of magnitude in peptide intensities (Fig. 4*C*, other organs see supplemental Fig. S2).

Among the set of tissue-enriched liver genes common to three donors with shared HLA types, the trend of immunopeptide intensities is generally similar (Fig. 4*D*). However, there are clear differences in the presentation of Cytochrome-P-Oxygenase 2A6 (CYP) and CYP3A4, which are involved in the metabolization of nicotine and pharmaceutical drugs, respectively (red arrows). This may be attributable to lifestyle differences or genetic differences between donors.

Our results demonstrate that the soluble immunopeptidome is overall representative of the organs constituting the body. They also suggest a considerable potential of plasma immunopeptidome analysis for studying system-wide changes in the human proteome and in providing novel insights into physiological and pathological processes that are presented to the immune system.

### *sHLA Immunopeptidomic Reproducability Over Time and Between Donors*

While the immunopeptidome is thought to change considerably upon disease, little is known about its stability in healthy persons over time. To address this fundamental question, we followed an initially healthy person over a year. We sampled plasma at and shortly after the initial time point (16 h apart) to gauge short-term biological variation, at the 5-month mark and at the end of the year. At about 11 months, the donor contracted COVID-19, and we sampled their immunopeptidome as soon as they were not positive any more (supplemental Table S1).

Throughout the entire time period more than half of all sHLA peptides were detectable and quantifiable, with 88 to 93% being shared in at least two time points (Fig. 5*A*). Remarkably, quantitative reproducibility over the entire year was very high (Pearson correlation of 0.97 between the first and the last time point (Fig. 5*B*)). The first two sampling points were only 16 h apart, also agreed very well with each other, suggesting that time of day did not have a large influence. Even the immunopeptidome shortly after a mild COVID-19 infection did not show large variations at a global scale. We note in passing that the impact of a COVID-19 infection on the immunopeptidome as well as the soluble immunopeptidome was studied in great detail here (25, 26).

Having established temporal stability of the sHLA peptidome in a single healthy donor, we next compared the immunopeptidomes of eight healthy donors (supplemental Table S1).

A Principal Component Analysis (PCA) clearly clustered workflow replicates of the same donor but next grouped donors by shared types or supertypes (Fig. 5*C*). Supporting this, a similar grouping emerged from pairwise Jaccard distances, which also revealed up to 50% overlap of identified peptide sequences between different donors with overlapping or similar presenting alleles (Fig. 5*D*).

Overall, the immunopeptidome of the different donors at best exhibited only a loose correlation (Fig. 5*E*). However, when selecting donors with a Jaccard similarity of more than 30%, the pairwise quantitative correlation significantly improved (Fig. 5*F*). Interestingly, donor 1 and donor 3 have a low Jaccard similarity (3%) between them—despite sharing one HLA-type (HLA-C0304); nonetheless, those 3% peptides show a high Pearson correlation.

These findings highlight the consistency and stability of the plasma immunopeptidome, further supporting its usefulness for insights into potential commonalities and variations among individuals.

## DISCUSSION AND OUTLOOK

Here we developed and applied IMBAS-MS, an improved approach to immunopeptidomics, with drastically enhanced sensitivity. This user-friendly and adaptable workflow replaces the traditional ProteinA/G affinity-based capture of anti-HLA antibodies (27) with a streptavidin-biotin one. This enables generic use of any biotinylated antibody, regardless of their immunoglobulin type, and greatly simplifies plasma-based immunopeptidomics by eliminating the need for plasma preclearance with its associated losses without introducing a time-consuming crosslinking step.

IMBAS-MS also eliminates nearly all hands-on time, in turn enabling the rapid preparation and acquisition of a large number of samples, which will be especially important in clinical environments. Without the need of specialized equipment as suggested in other automated immunopeptidomics platforms (28–32), IMBAS-MS can be used efficiently simply with a magnet and is thereby equally accessible to specialized immunopeptidomics laboratories as well as regular mass spectrometry-focused laboratories or MS-facilities. We expect IMBAS-MS to have the same advantages in tissue-based immunopeptidomics and we plan to explore this aspect in the future.

As part of our workflow, we have also implemented Data Independent Acquisition (DIA) to expand the depth of the

immunopeptidomic data. To tackle the challenge of creating a suitable search space for immunopeptidomics, we employed personalized HLA peptide libraries (23). This considerably reduces the number of potential 9mers to 12mers in a human FASTA to be searched, increasing the number of significant identifications. In contrast to other library generation strategies (33, 34), our approach eliminates the need for any upfront measurements and can be transferred between MS platforms. It also avoids building a library from DDA runs and could be adapted to supertype or study-specific libraries, potentially incorporating common post-translational modifications. This significantly reduces both the measurement time and material required as each sample serves as a base for its own predicted library and discovery based analysis at the same time.

As a next step, we envision combining IMBAS-MS with multiplexed DIA and in particular to use one of the channels as a reference channel (35, 36). Chemical labeling strategies allow for the combined measurement of otherwise separately acquired samples, where each label represents one possible channel. A reference channel could be used to decouple identification and quantification, improving immunopeptidomic depth, sensitivity and comparability between samples as it was shown for single-cell proteomics experiments (35, 36).

Our results highlight the potential diagnostic applicability beyond identifying cancer neoepitopes. They demonstrate the presence of very large numbers of immunopeptides in plasma samples, further supporting the notion of plasma as a valuable, non-invasive source of immunopeptides (4, 10, 37).

We observed that the immunopeptides found in plasma are mostly representative of the tissue proteome. However, brain-associated proteins were less represented and it would be interesting to investigate mechanisms of presentation of these sHLAs in the plasma. We also demonstrated the existence of a stable healthy plasma immunopeptidome, both quantitatively and qualitatively, across different healthy individuals, expanding on what was described earlier (4). This finding is highly relevant for clinical applications, as it suggests that a general baseline healthy immunopeptidome can be established. In turn, this could significantly facilitate the identification of disease-specific immunopeptide signatures and aid in the development of novel diagnostic markers and therapeutic strategies. Such an approach could extend the diagnostic potential of plasma immunopeptidome profiling within and beyond the search for neoepitopes in the context of cancer. This may provide insights into a wide range of pathological conditions that involve alterations in immune responses, such as autoimmune disorders, infectious diseases or inflammatory conditions. In this context, the minimal-invasive nature of plasma-based immunopeptidome profiling combined with the streamlined IMBAS-MS technology could enable a patient-friendly approach to disease monitoring and personalized medicine, facilitating earlier intervention and more effective treatment strategies. Clearly, future studies are needed to expand upon these exciting findings by investigating basic aspects of sHLA generation and presentation and the diagnostic capabilities of plasma immunopeptide signatures in specific disease states. Combined with the ongoing development of the underlying analytical technology, sHLA peptidomics may become an important addition to the arsenal of precision medicine.

## DATA AVAILABILITY

The raw mass spectrometry data have been deposited in the public proteomics repository MassIVE for reviewer access (MSV000092557). This data will be made public upon acceptance of the manuscript.

*Supplemental data*—This article contains supplemental data.

*Acknowledgments*—M. W., M. T. and M. Z. acknowledge support from the International Max Planck Research School for Life Sciences – IMPRS-LS. We are grateful for the sequencing support provided by the NGS Core Facility at the Max Planck Institute of Biochemistry. We are especially grateful for the blood donors. We thank our colleagues in the Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry and at the Center for Protein Research at Copenhagen University, for discussions and support. In particular, we thank I. Paron and T. Heymann for technical support and Medini Steger for scientific administration support.

*Funding and additional information*—This study was supported by the Max Planck Society for Advancement of Science, the European Union's Horizon 2020 research and innovation program under grant agreement No 874839 (ISLET) and by the Bavarian State Ministry of Health and Care through the research project DigiMed Bayern (www.digimed-bayern.de).

*Author contributions*—M. W., M. T., and M. M. conceptualization; M. W., M. T, and M. M. methodology; M. W., W.-F. Z., and M. Z. investigation; M. W. and M. T. formal analysis; M. W. and P. S. resources; M. W. and M. M. writing–original manuscript; M. W., M. T., W.-F. Z., P. S., and M. M. writing–review and editing.

*Conflict of interest*—M. M. is an indirect investor in Evosep Biosystems. All other authors declare that they have no conflicts of interest with the contents of this article.

*Abbreviations*—The abbreviations used are: ACN, acetonitrile; CAR T-cell, chimeric antigen receptor T-cell; CTL, cytotoxic T lymphocytes; CYP, Cytochrome-P-Oxygenase; dda, data-dependent acquisition; dia, data-independent acquisition; FA, formic acid; GO, Gene Ontology; (s/m)HLA, (soluble/membrane) human leukocyte antigen; IgG, Immunoglobuline

G; IEDB, Immune Epitope Database & Tools; IM, ion mobility; IMBAS, Streptavidin based immunopeptidomics workflow (Immunopeptidomics by Biotinylated Antibodies and Streptavidin); MeOH, methanol; MHC, major histocompatibility complex; PASEF, parallel accumulation – serial fragmentation; PBS, phosphate-buffered saline; py_diAID, Python package for Data-Independent Acquisition with an Automated Isolation Design; SPD, samples per day; TIMS, trapped ion mobility spectrometry.

Received August 22, 2023, and in revised form, November 24, 2023
Published, MCPRO Papers in Press, December 1, 2023, https://doi.org/10.1016/j.mcpro.2023.100689

REFERENCES

1. Nelde, A., Maringer, Y., Bilich, T., Salih, H. R., Roerden, M., Heitmann, J. S., *et al*. (2021) Immunopeptidomics-guided warehouse design for peptide-based immunotherapy in chronic lymphocytic leukemia. *Front. Immunol.* **12**, 705974
2. Marcu, A., Bichmann, L., Kuchenbecker, L., Kowalewski, D. J., Freudenmann, L. K., Backert, L., *et al*. (2021) HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. *J. Immunother. Cancer* **9**, e002071
3. Bassani-Sternberg, M., Bräunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., *et al*. (2016) Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* **7**, 13404
4. Bassani-Sternberg, M., Barnea, E., Beer, I., Avivi, I., Katz, T., and Admon, A. (2010) Soluble plasma HLA peptidome as a potential source for cancer biomarkers. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 18769–18776
5. Demaria, S., and Bushkin, Y. (2000) Soluble HLA proteins with bound peptides are released from the cell surface by the membrane metalloproteinase. *Hum. Immunol.* **61**, 1332–1338
6. Adamashvili, I., Kelley, R. E., Pressly, T., and McDonald, J. C. (2005) Soluble HLA: patterns of expression in normal subjects, autoimmune diseases, and transplant recipients. *Rheumatol. Int.* **25**, 491–500
7. Shraibman, B., Barnea, E., Kadosh, D. M., Haimovich, Y., Slobodin, G., Rosner, I., *et al*. (2019) Identification of tumor antigens among the HLA peptidomes of glioblastoma tumors and plasma. *Mol. Cell. Proteomics* **18**, 1255–1268
8. Ritz, D., Gloger, A., Neri, D., and Fugmann, T. (2017) Purification of soluble HLA class I complexes from human serum or plasma deliver high quality immuno peptidomes required for biomarker discovery. *Proteomics* **17**, 1600364
9. Ritz, D., Sani, E., Debiec, H., Ronco, P., Neri, D., and Fugmann, T. (2018) Membranal and blood-soluble HLA class II peptidome analyses using data-dependent and independent acquisition. *Proteomics* **18**, e1700246
10. Gfeller, D., and Bassani-Sternberg, M. (2018) Predicting antigen presentation-What could we learn from a million peptides? *Front. Immunol.* **9**, 1716
11. Tanuwidjaya, E., Schittenhelm, R. B., and Faridi, P. (2022) Soluble HLA peptidome: a new resource for cancer biomarkers. *Front. Oncol.* **12**, 1069635
12. Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., *et al*. (2019) The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343
13. Skowronek, P., Thielert, M., Voytik, E., Tanzer, M. C., Hansen, F. M., Willems, S., *et al*. (2022) Rapid and in-depth coverage of the (Phospho-) proteome with deep libraries and optimal window design for dia-PASEF. *Mol. Cell. Proteomics* **21**, 100279
14. Yu, F., Teo, G. C., Kong, A. T., Haynes, S. E., Avtonomov, D. M., Geiszler, D. J., and Nesvizhskii, A. I. (2020) Identification of modified peptides using localization-aware open search. *Nat. Commun.* **11**, 4065
15. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., and Nesvizhskii, A. I. (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513–520
16. Yu, F., Haynes, S. E., Teo, G. C., Avtonomov, D. M., Polasky, D. A., and Nesvizhskii, A. I. (2020) Fast quantitative analysis of timsTOF PASEF data with MSFragger and IonQuant. *Mol. Cell. Proteomics* **19**, 1575–1585
17. Kovalchik, K. A., Ma, Q., Wessling, L., Saab, F., Duquette, J. D., Kubiniok, P., *et al*. (2022) MhcVizPipe: a quality control software for rapid assessment of small- to large-scale immunopeptidome datasets. *Mol. Cell. Proteomics* **21**, 100178
18. [preprint] Brunner, A.-D., Thielert, M., Vasilopoulou, C. G., Ammar, C., Coscia, F., Mund, A., *et al*. (2020) Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *bioRxiv*. https://doi.org/10.1101/2020.12.22.423933
19. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., *et al*. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111.016717
20. Hansen, F. M., Tanzer, M. C., Brüning, F., Bludau, I., Stafford, C., Schulman, B. A., *et al*. (2021) Data-independent acquisition method for ubiquitinome analysis reveals regulation of circadian biology. *Nat. Commun.* **12**, 254
21. Tsou, C.-C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A. C., and Nesvizhskii, A. I. (2015) DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* **12**, 258–264
22. Gessulat, S., Schmidt, T., Zolg, D. P., Samaras, P., Schnatbaum, K., Zerweck, J., *et al*. (2019) Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **16**, 509–518
23. [preprint] Zeng, W.-F., Zhou, X. X., Willems, S., Ammar, C., Wahle, M., Bludau, I., *et al*. (2022) AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics. *bioRxiv*. https://doi.org/10.1101/2022.07.14.499992
24. Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D. P., *et al*. (2019) A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503
25. Nelde, A., Rieth, J., Roerden, M., Dubbelaar, M. L., Hoenisch Gravel, N., Bauer, J., *et al*. (2022) Increased soluble HLA in COVID-19 present a disease-related, diverse immunopeptidome associated with T cell immunity. *iScience* **25**, 105643
26. Weingarten-Gabbay, S., Klaeger, S., Sarkizova, S., Pearlman, L. R., Chen, D. Y., Gallagher, K. M. E., *et al*. (2021) Profiling SARS-CoV-2 HLA-I peptidome reveals T cell epitopes from out-of-frame ORFs. *Cell* **184**, 3962–3980.e17
27. Purcell, A. W., Ramarathinam, S. H., and Ternette, N. (2019) Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics. *Nat. Protoc.* **14**, 1687–1707
28. Chong, C., Marino, F., Pak, H., Racle, J., Daniel, R. T., Müller, M., *et al*. (2018) High-throughput and sensitive immunopeptidomics platform reveals profound interferon γ-mediated remodeling of the human leukocyte antigen (HLA) ligandome. *Mol. Cell. Proteomics* **17**, 533–548
29. Feola, S., Haapala, M., Peltonen, K., Capasso, C., Martins, B., Antignani, G., *et al*. (2021) PeptiCHIP: a microfluidic platform for tumor antigen landscape identification. *ACS Nano* **15**, 15992–16010
30. Phulphagar, K. M., Ctortecka, C., Jacome, A. S. V., Klaeger, S., Verzani, E. K., Hernandez, G. M., *et al*. (2023) Sensitive, high-throughput HLA-I and HLA-II immunopeptidomics using parallel accumulation-serial fragmentation mass spectrometry. *Mol. Cell. Proteomics* **22**, 100563
31. Li, X., Pak, H. S., Huber, F., Michaux, J., Taillandier-Coindard, M., Altimiras, E. R., and Bassani-Sternberg, M. (2023) A microfluidics-enabled automated workflow of sample preparation for MS-based immunopeptidomics. *Cell Rep. Methods* **3**, 100479
32. Zhang, L., McAlpine, P. L., Heberling, M. L., and Elias, J. E. (2021) Automated ligand purification platform accelerates immunopeptidome analysis by mass spectrometry. *J. Proteome Res.* **20**, 393–408
33. Pak, H., Michaux, J., Huber, F., Chong, C., Stevenson, B. J., Müller, M., *et al*. (2021) Sensitive immunopeptidomics by leveraging available large-scale multi-HLA spectral libraries, data-independent acquisition, and MS/MS prediction. *Mol. Cell. Proteomics* **20**, 100080
34. Xin, L., Qiao, R., Chen, X., Tran, H., Pan, S., Rabinoviz, S., *et al*. (2022) A streamlined platform for analyzing tera-scale DDA and DIA mass spectrometry data enables highly sensitive immunopeptidomics. *Nat. Commun.* https://doi.org/10.1038/s41467-022-30867-7

35. Derks, J., Leduc, A., Wallmann, G., Huffman, R. G., Willetts, M., Khan, S., *et al*. (2023) Increasing the throughput of sensitive proteomics by plexDIA. *Nat. Biotechnol.* **41**, 50–59
36. [preprint] Thielert, M., Itang, C. E. M., Ammar, C., Schober, F. A., Bludau, I., Skowronek, P., *et al*. (2022) Robust dimethyl-based multiplex-DIA workflow doubles single-cell proteome depth *via* a reference channel. *boRxiv*. https://doi.org/10.1101/2022.12.02.518917
37. [preprint] Ouspenskaia, T., Law, T., Clauser, K. R., Klaeger, S., Sarkizova, S., Aguet, F., *et al*. (2020) Thousands of novel unannotated proteins expand the MHC I immunopeptidome in cancer. *boRxiv*. https://doi.org/10.1101/2020.02.12.945840

## 3.5. Pancreatic islet biology

### Article 12. Cell-type resolved proteome dissection of mouse endocrine pancreatic cell types

*Manuscript draft*, 2024

Marvin Thielert[1], Adrian Villalba[2], Maria Wahle[1], Andreas-David Brunner[1,3], Alexis Fouque[2], Chloé Lourenco[2], Masaya Oshima[2], Latif Rachdi[2], …, Matthias Mann[1,#]

*[1]Max Planck Institute of Biochemistry, Martinsried, Germany. [2]Department of Endocrinology, Metabolism and Diabetes, Institute Cochin, INSERM, University Paris Descartes, Paris, France. [3]Now: Boehringer Ingelheim Pharma GmbH & Co. KG, Drug Discovery Sciences, Biberach an der Riss, Germany. [#]Corresponding author: Matthias Mann, mmann@biochem.mpg.de*

During my PhD studies, I have been part of a European consortium called ISLET, which aims to develop advanced cell therapy for diabetes for type 1 diabetes (T1D). Within the consortium, I took part in and managed the proteomics part with the ultimate goal to improve a potential therapeutic product. Many collaborative projects have evolved from this network and are also part of my PhD studies, but mostly not part of this thesis.

One of the projects aimed at a comprehensive understanding of adult mouse pancreatic islet cell types of alpha, beta and delta. Protein-level data for these cell types is limited due to their scarcity and the challenges involved in their purification. We employed our sensitive MS approach and combined it with a specialized FACS-sorting approach developed by our collaborators of cell type populations based on the combination of the surface markers CD24, CD49f, and CD71. We quantified over 6,000 proteins in each endocrine cell type from less than 25,000 cells. Our proteomic analysis revealed known and novel differences between the cell types with bona fide markers such as insulin, glucagon and somatostatin enriched in their respective cell types. This also led to the discovery of new protein markers for cell sorting, notably Sirpa was predominantly detected in beta cells, as well as Gpr158. We further investigated cell type specific changes upon IFNγ, which mimics the inflammatory environment in T1D. Our findings showed global changes in all cell types, suggesting that all cell types are inflamed to some degree. We also found cell type specific proteomic changes upon IFNγ stimulation that had previously not been detected by transcriptomic approaches.

In summary, this work will enhance the understanding of the different cell types in pancreatic islet and their contribution to the inflammation status, which can be transferred to advance the therapeutic product of ISLET.

***Contribution:***

First-authorship. Matthias Mann and I conceptualized the study. I conducted the project, performed experiments and designed figures. I acquired and analyzed the proteomic data. I wrote the manuscript draft together with Adrian Villalba Felipe and Matthias Mann. This work is part of the European consortium ISLET, a collaboration of many laboratories.

# Cell-type resolved proteome dissection of mouse endocrine pancreatic cell types

**Author list**

Marvin Thielert[1], Adrian Villalba[2], Maria Wahle[1], Andreas-David Brunner[1,3], Alexis Fouque[2], Chloé Lourenco[2], Masaya Oshima[2], Latif Rachdi[2], …, Matthias Mann[1,#]

**Affiliations**

1 Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany.
2 Department of Endocrinology, Metabolism and Diabetes, Institute Cochin, INSERM, University Paris Descartes, Paris, France
3 New address: Boehringer Ingelheim Pharma GmbH & Co. KG, Drug Discovery Sciences, Biberach an der Riss, Germany

[#]Corresponding author: Matthias Mann, mmann@biochem.mpg.de

**Running title:** Proteome of pancreatic endocrine cell types

**Keywords:** Cell-type resolved proteomics / pancreatic islets / endocrine / atlas

**Abstract**

**Islets of Langerhans are micro-organs scattered throughout the pancreas. They are composed of endocrine cells responsible for synthesizing, storing and secreting pancreatic hormones such as insulin (beta cells), glucagon (alpha cells) and somatostatin (delta cells). While their transcriptome has been analyzed in depth at cell type level, far less is known about the proteome of alpha, beta and delta cell populations. This is partly due to the scarcity of such endocrine cell populations and the difficulty to purify them. We reasoned that recent developments in high sensitivity mass spectrometry should now allow the acquisition of proteomes from such low input material, enabling cell-type resolved proteome signatures in pancreatic islets. Here, we combined such highly sensitive mass spectrometry approaches with our previously developed strategy to sort mouse alpha, beta, and delta cell populations. In this way, we obtained a proteomic depth of more than 6,000 proteins for every endocrine cell population. We discovered new proteins enriched in specific cell types, including Sirpa, a plasma membrane protein that we validated as a new marker to sort beta cell subpopulations and Gpr158 which is a new beta cell specific receptor. We also performed deep proteomics profiling upon interferon gamma (IFNγ) treatment, a condition that mimics inflammation observed in type 1 diabetes. Our results demonstrate that all three endocrine cell types were inflamed upon stimulation and revealed cell-type specific proteome changes beyond those seen in transcriptomics approaches.**

**Introduction**

Diabetes is a group of metabolic disorders characterized by chronic hyperglycaemia and caused by the destruction or dysfunction of insulin-producing beta cells. These endocrine cells are clustered in the pancreas in small micro-organs, named islets of Langerhans[1]. Adult mouse islets are composed of 75% insulin-producing beta cells, 15% glucagon-producing alpha cells and 10% somatostatin-producing delta cells[2,3]. The transcriptomic landscape of islet cells has now been extensively characterized at the single cell level[4–6], resulting in atlases identifying cell specific signatures for each pancreatic endocrine cell type under basal and disease conditions[7]. Inter-species comparisons have also been explored at the transcriptome level[8].

The dynamic proteome is thought to be a very informative reflection of the biological function of cells[9]. However, our understanding of the proteomic profiles of alpha, beta and delta cells remains limited[10,11]. Previous proteomic analyses have focused primarily on entire pancreatic islets, with limited attempts to separate the different endocrine populations. This is due to the difficulties in purifying these distinct cell populations and the limitations of mass spectrometer (MS)-based proteomics, which has struggled with the sensitivity required to especially the less abundant alpha and delta cells[12,13].

To address these challenges, we here build upon a strategy to isolate highly enriched populations of alpha, beta, and delta cells using a combination of surface markers[14] and combine this sorting approach with recent developments in deep proteomic profiling. We employ high sensitivity MS of trapped-ion mobility time of flight (timsTOF) instruments that allow the acquisition of proteomes from low input material, making unbiased characterization on protein level possible[15]. Furthermore, the combination with data-independent acquisition (DIA) strategies enhances data completeness, ensuring that proteome measurements are less prone to stochastic sampling of high abundant proteins like hormones[16].

Altogether, this approach has enabled us to build the first proteomic atlas of mouse adult alpha, beta, and delta cells. It consists of over 6,000 protein groups in each endocrine cell type, an unprecedented depth of proteomic data. Additionally, our approach facilitated the detection of dynamic proteomic changes, such as the fingerprint of IFNγ, enabling exploration of proteomic alterations under inflammatory conditions. These detailed datasets provide a valuable resource for advancing research into the molecular mechanisms underlying islet biology and related diseases[17,18].

## Results

### Mouse alpha, beta, and delta cells can be resolved through FACS and proteomic profiling

While global proteomic profiling has previously been described even for single whole mouse pancreatic islets[19], these analyses were dominated by beta cell proteins due to their abundance. In contrast, individual proteomic profiles of alpha, beta, and delta cell subpopulations have remained elusive. To address this gap, we prepared islets from 12-week-old C57BL/6 mice, cultured them overnight and dispersed them into single cells. We then employed our previously described FACS-based approach on alpha, beta, and delta cell populations[14] to flow sort with antibodies against CD24 and CD71 in the EpCam+ epithelial fraction after excluding non-pancreatic lineages and dead cells (CD31+, CD45+, CD235a+ and propidium iodide+) (Fig. 1a, Supp. Fig. 1). To verify the purity of our sorted cell populations, we analyzed the mRNA expression of endocrine, acinar, and ductal markers by qRT-PCR. As expected, sorted alpha, beta, and delta cell subsets were consistently enriched in Gcg, Ins1, and Sst, respectively (Supp. Fig. 2). Furthermore, acinar (Amy and Cpa1) and ductal (Spp1) markers were expressed at extremely low levels in all three endocrine cell subsets (Supp. Fig. 2). Following cell sorting, we performed proteomics on the isolated alpha, beta and delta cell populations (Fig. 1a). Starting from only 25,000 alpha and beta cells, and 14,500 delta cells, we detected more than 6,000 proteins per endocrine subset (6,471±42, 6,493±3, and 6,283±106 protein groups, for alpha, beta, and delta cells, respectively) (Fig. 1b). Principal component analysis (PCA) revealed that alpha, beta and delta cell populations form distinct clusters (Fig. 1c). To further validate our approach, we analysed the abundance of bona fide markers for each endocrine cell type. As anticipated, sorted alpha cell subsets were consistently enriched in Gcg and in the transcription factor Arx, beta cells were enriched in Ins1, Ins2, Glp1R and Ucn3 compared to alpha and delta cells and delta cells were enriched in Sst and Rbp4 (Fig. 1d).

We extended our analysis to investigate the effects of inflammation by performing similar proteomic analyses on alpha, beta and delta cell populations sorted from islets treated in vitro with IFNγ, which mimics inflammation observed in type 1 diabetes. We detected similar numbers of proteins in all three subsets (6,488±35, 6,488±9, and 6,396±47 protein groups, for alpha, beta, and delta cells, respectively) and treated samples cluster close to their respective cell type (Fig. 1b and c).

We conclude that our combined strategy of FACS and proteome profiling allows unbiased quantitative analysis of the proteome of alpha, beta, and delta cell populations in control conditions and upon IFNγ treatment, providing a powerful foundation to investigate islet cell biology.

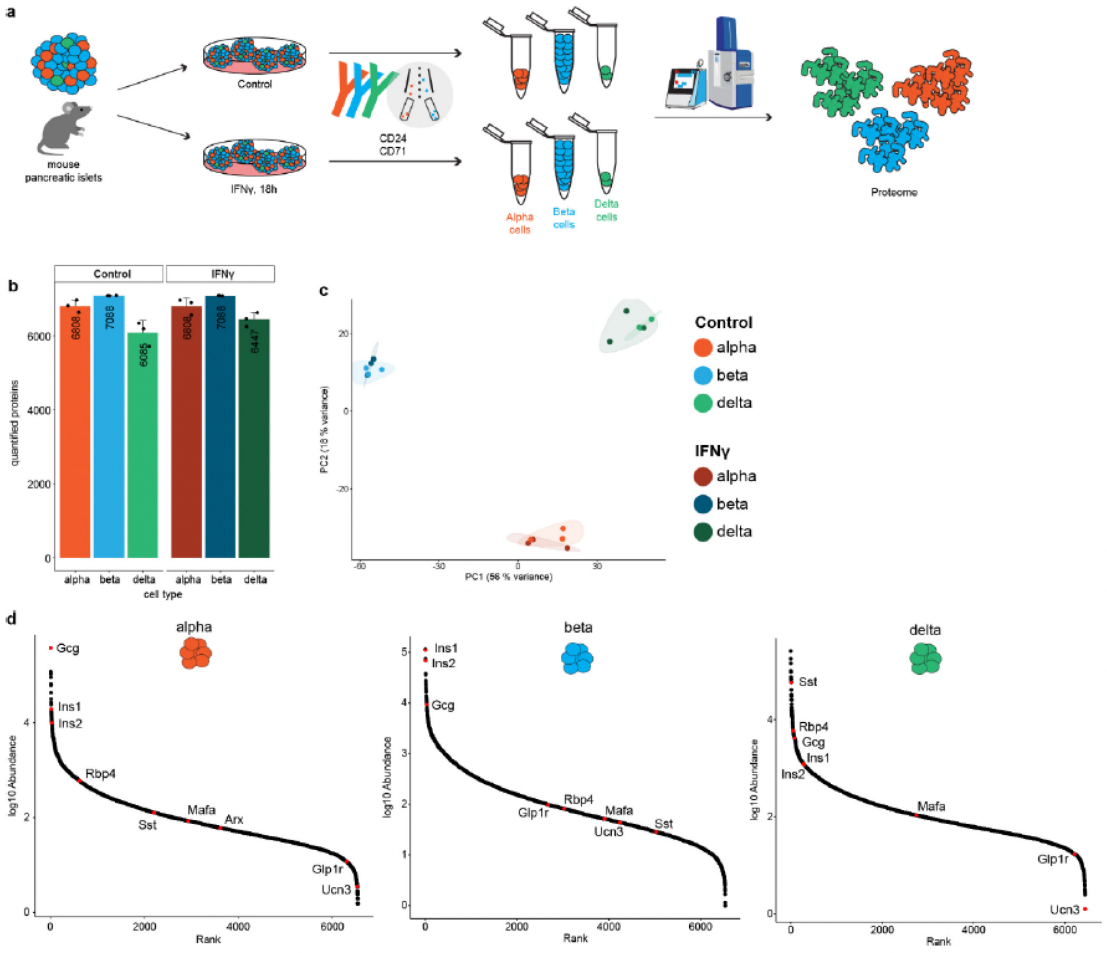



**Figure 1 - Cell type-resolved proteomic profiling of sorted mouse adult pancreatic endocrine cells.** (a) Schematic representation outlining the experimental workflow. (b) Quantification of the total number of proteins per cell type in control and IFNγ-treated conditions. (c) Principal Component Analysis (PCA) applied to the proteomic profiles of alpha, beta and delta cells in control and IFNγ-treated conditions. (d) Log-rank graphs depicting the total protein quantification in each endocrine cell type, including bona fide markers for each cell type. Biological replicates n = 3 per group.

### Unravelling cell-type resolved proteome profiles for mouse alpha, beta and delta cells

To identify proteins enriched in alpha, beta, or delta cells, we conducted comparative analyses between these cell types (Fig. 2a-c). Alpha cells were enriched in established markers like Gcg, Ttr and Pou3f4 (Fig. 2a-b) and beta cells in Ins2, Nkx6.1, Pdx1, Iapp, Wfs1 (a protein localized to the endoplasmic reticulum), Ucn3, Glp1r, Slc2a2, Pppr1a, and Pcsk1 (Fig. 2a and c). Interestingly, Pdx1, Iapp, Wfs1 and Pcsk1 were also enriched in delta cells compared to alpha cells (Fig. 2b). As expected, delta cells were specifically enriched in Sst and Rbp4, further validating our approach (Fig. 2b-c).

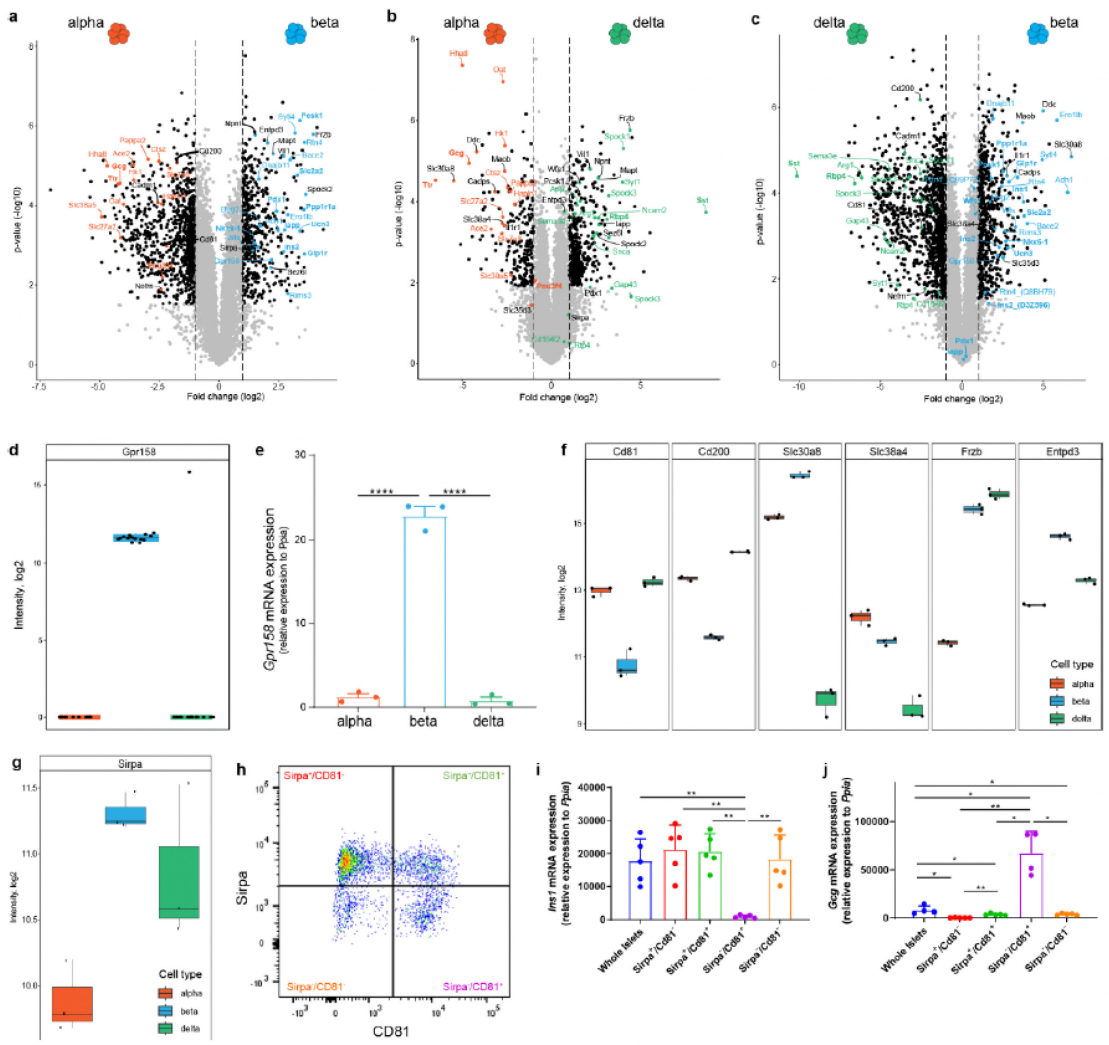



**Figure 2 - Proteomic profile of mouse alpha, beta and delta cells.**

(a-c) Volcano plots showing the differential enrichment of proteins in alpha vs beta vs delta cells. (d) Protein levels of Gpr158 protein in alpha, beta and delta cells from proteomic measurement. (e) Relative mRNA expression levels of *Gpr158* in alpha, beta and delta cells. (f) Boxplots presenting the abundance of distinct proteins enriched in alpha, beta and delta cells. (g) Boxplot presenting the levels of Sirpa protein in alpha, beta and delta cells. Each biological replicates n = 3 per group. (h) Representative flow cytometry plot of dispersed mouse islets. CD81 and Sirpa expression was analyzed on $CD24^-$ cells after excluding dead cells (propidium $iodide^+$ cells) and hematopoietic and endothelial cells (CD45-, TER119-, or CD31-positive cells). (i-j) Relative mRNA expression levels of endocrine cell markers *Ins1* (i) and *Gcg* (j) in whole islets and $Sirpa^+CD81^-$, $Sirpa^+CD81^+$, $Sirpa^-CD81^+$, $Sirpa^-CD81^-$ cells. n= 4-5 independent experiments.

Our analysis revealed cell type-specific enrichment of several little studied or unreported proteins. Alpha cells were enriched in transporter proteins, including the fatty acid transporter Slc27a2 and the neutral amino acid transporters Slc7a8, and Slc38a5. They also exhibited higher levels of

membrane proteins such as Ace2 (a carboxypeptidase involved in the cleavage of angiotensin and – interestingly - being one of the two co-receptor for the COVID-19 virus cell entry[20]), and Hapln1 (an extracellular matrix component) as well as enzymes including Hk1 (hexokinase 1), Ctsz (a protease), Oat (an aminotransferase), Hhatl (an acetyltransferase), and Pappa2 (a metalloprotease involved in the cleavage of Igfbp5) (Fig. 2a-b).

Beta cells demonstrated enrichment in membrane proteins like Emilin1 (involved in cell-cell contacts), Rtn4, Rims3 and Sytl4 (three regulators of vesicle exocytosis), Wls (a Wnt receptor), Bace2 (an enzyme that cleaves APP). They also showed higher levels of Adh1 (an alcohol dehydrogenase enzyme), Bag2 and Dnajb11 (chaperone proteins), Ero1lb (an enzyme involved in protein folding) (Fig. 2a and c). Notably, Gpr158, a recently deorphanized receptor[21], was exclusively detected in the beta cell fraction (Fig. 2a, c and d). We further validated this finding by qRT-PCR, confirming Gpr158 mRNA expression only in beta cells (Fig. 2e).

Delta cells exhibited enrichment in Gap43 and Ncam2 (membrane proteins regulating cell-cell contacts), Spock1, Spock3, and Sema3e (cell-extracellular matrix interactions), Cd164l2 and App (2 membrane proteins), Arg1 (an enzyme involved in arginine hydrolysis), and Syt1, Snca, and Rtp4 (that regulate vesicle exocytosis and membrane trafficking) (Fig. 2b and c).

We also identified several little studied proteins enriched in two of the three cell types. Alpha and delta cell subsets were enriched in Cadm1 (a cell adhesion molecule), Cd81 and Cd200 (2 cell surface molecules), and Nefm (a cytoskeleton component) (Fig. 2a, c, f). Alpha and beta cell subsets were enriched in Slc30a8 (a zinc transporter), Slc38a4 (an amino acid transporter), Slc35d3 (a nucleotide-sugar transporter), Cadps (a regulator of vesicle exocytosis), Ddc (DOPA decarboxylase), Maob (an oxidase located in the mitochondrial membrane), and Il1r1 (interleukin receptor) (Fig. 2b, c, f). Beta and delta cells showed enrichment for proteins such as Frzb (a Wnt-binding protein), Sez6l (a membrane receptor), Entpd3 (a nucleotidase), Vil1 (an actin-binding protein), Spock2 (a regulating cell-cell contacts), Npnt (a ligand of integrins) and Mapt (cytoskeleton components) (Fig. 2a, b, f).

Of particular interest was Sirpa (signal regulatory protein A), enriched in both beta and delta cells (Fig. 2a, c, g). This cell surface protein has been suggested as a regulator of beta cell viability[22]. To further investigate Sirpa's expression, we analysed dispersed islet cells from 12-week-old mice using flow cytometry with antibodies directed against Cd24, Cd81 and Sirpa. After excluding the CD24high population that contains delta cells[14], we identified four subsets based on Cd81 and Sirpa expression (Fig. 2h). A RT-PCR analysis revealed the expression of Gcg and Ins in Sirpa+ subsets (Fig. 2i) and Gcg in a Sirpa- fraction (Sirpa-/Cd81+, Fig. 2j), supporting our proteomic analysis. Interestingly, we also observed a small subset of CD81-Sirpa- cells (Fig. 2h) expressing

Ins (Fig. 2i) but not Gcg (Fig. 2j), indicating that the majority of the beta cells express Sirpa on their surface, while a minor beta cell subpopulation does not.

**IFNγ treatment induces distinct proteomic changes in pancreatic endocrine cell populations**

PCA revealed that IFNγ-treated alpha, beta, and delta cell populations could be distinguished from untreated cells along principal component 3 (Fig. 3a). Analysis of variance (ANOVA) identified significant differences in 67% of all proteins across all conditions (cell type and upon IFNγ treatment) (Supp. Fig. 3). However, the vast majority of them (97%) remained unaltered upon IFNγ treatment alone.

Proteomic profiling indicated that exposure to IFNγ can modulate the expression of some proteins in each cell subset. Most of them are induced upon IFNγ exposure, and very few downregulated, as previously observed in our transcriptomic data. We observed significant overlap of the proteins induced by IFNγ across the three endocrine cell subsets (Fig. 3b), a common responsiveness to this treatment of all three cell types. Key upregulated proteins in all three cell types included Igtp (an IFN-induced gtpase), Stat1 and Stat2 (transcription factors of the IFN pathway), and Tap1(associated protein to class I MHC processing) (Fig. 3c).

Intriguingly, we also identified cell type-specific protein enrichments following IFNγ treatment (Fig. 3d-f): alpha cells: Gbp8 (another IFN-induced GTPase), Ryr2 (an intracellular $Ca^{2+}$ channel), and Rbms1 (an RNA-binding protein) (Fig. 3d); beta cells: Cxcl10 (a cytokine), Lama2 (a laminin subunit), Fau (a ubiquitin-like protein), and Mfsd8 (an orphan membrane transporter) (Fig. 3e); delta cells: Trim41 and Rnf115 (two E3 ligases related to innate immunity), Elf5 (a transcription factor), and Tuft1 (a regulator of the mTOR pathway) (Fig. 3f)

Thus, our proteomic profiling successfully captured the fingerprint of IFNγ-induced changes in mouse alpha, beta, and delta cells, providing a novel approach to explore the dynamic proteome in response to inflammatory stimuli.

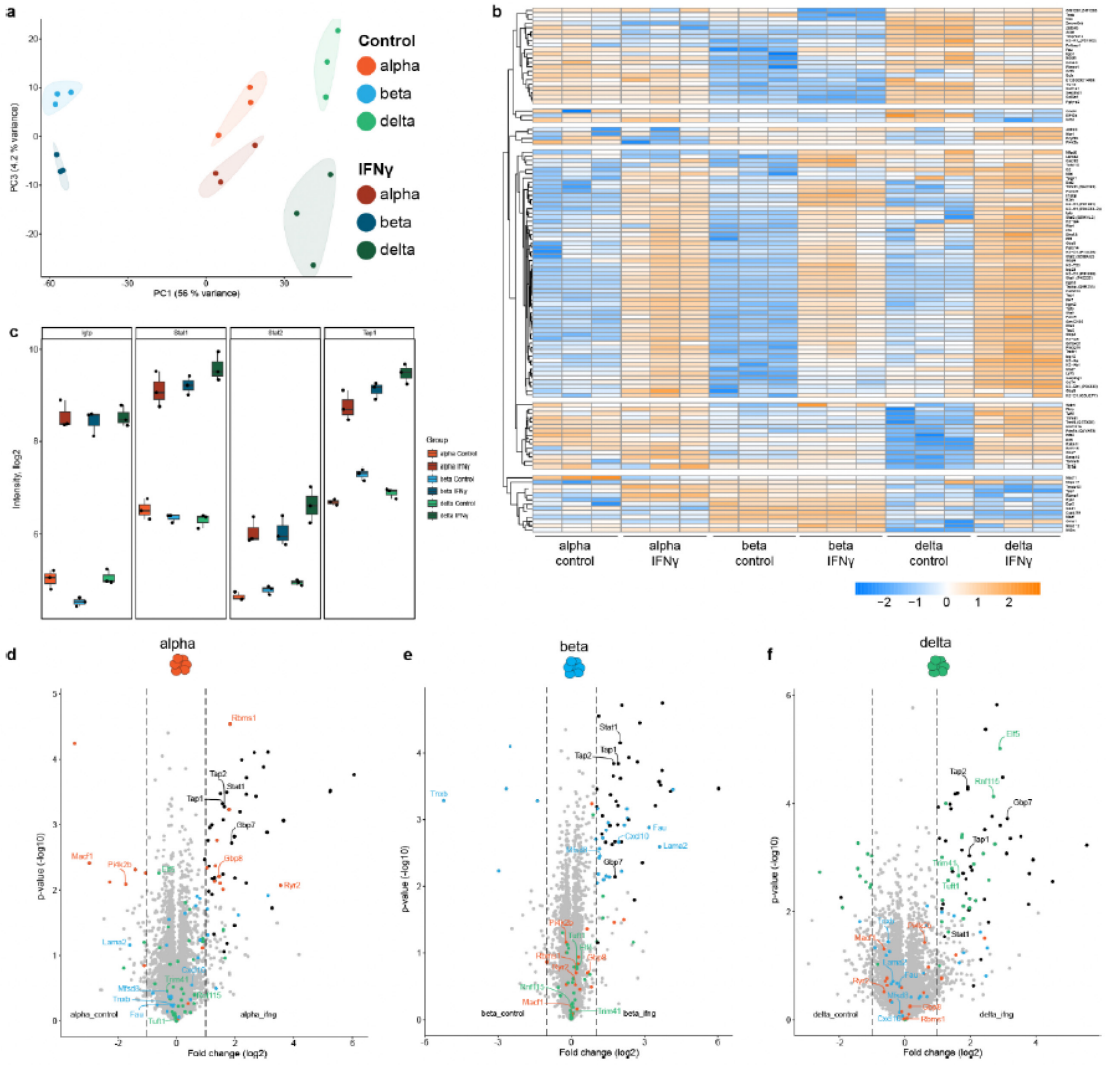



**Figure 3 - Proteomic profile of alpha, beta and delta cells upon IFNγ exposure.**

(a) Principal Component Analysis (PCA) of the proteomic profiles of alpha, beta and delta cells in control and IFNγ-treated conditions. IFNγ-treated conditions are separated by principal component 3. (b) Heatmap of significantly differently expressed proteins upon IFNγ treatment comparing the proteomic profiles of alpha, beta and delta cells in control and IFNγ-treated conditions. Enriched proteins are clustered into six categories by hierarchical clustering. (c) Boxplots presenting the abundance of various proteins associated with inflammation in control and IFNγ-treated alpha, beta and delta cells. Biological replicates n = 3 per group. (d-f) Volcano plots illustrating the differential enrichment of proteins between control and IFNγ-treated alpha (d), beta (e) and delta cells (f).

**Discussion**

In this manuscript draft, we present a proteomic data set derived from alpha, beta and delta cell populations of mouse adult pancreatic islets. Mouse pancreatic islets are composed of alpha, beta and delta with beta cells comprising about 75% of the islet population. Previous studies have provided a beta-centric view of the islet proteome, without cell types specific distinctions[10,11], limiting our understanding of alpha and delta cell proteomes. Our cell type-specific approach addresses this gap, offering insights into the unique proteomes of these less abundant cell populations[14,17].

FACS followed by proteomics has been used in the past in other tissues, for instance to construct an in-depth quantitative proteomic atlas of human skin melanocytes, endothelial and immune cells[23]. In another study, endothelial and immune cells were sorted from the gut of patients with Crohn's disease[24]. However, these studies used orders of magnitude larger amount of starting tissues. In contrast, our FACS-based approach followed by deep proteomics was readily applicable to mouse pancreatic endocrine cells, which comprise only 2% in the pancreas. Our approach captured the dynamic proteome of alpha, beta and delta cells upon IFNγ exposure, which mimics inflammatory conditions and opens new avenues for exploring the specific proteome of alpha, beta and delta cell populations in other conditions like type 2 diabetes or aging. Additionally, our workflow is suitable for other FACS sortable cell populations with limited sample amount.

Our analysis revealed novel proteins expressed in specific endocrine cell populations. For instance, we detected Gpr158 specifically in beta cells, a finding that aligns with mRNA expression data. This discovery opens new avenues for studying GPCR function in beta cells, particularly in relation to insulin secretion and cell survival[25]. Due to the lack of properly validated Gpr158 antibodies, this protein had not been analyzed in pancreatic beta cells and this specific expression opens new avenues to study the role of this new GPCR in beta cells. Whether its recently described ligands osteocalcin[26] and glycine[21] in the central nervous system signal on beta cells to modulate insulin secretion or cell survival will be the topic of future studies. Gpr158 can also now be added to the list of beta cell-specific proteins that are also expressed in neurons, such as the transcription factor NKX6.1[27], Glutamic Acid decarboxylase (GAD1) and the membrane glycoprotein BACE2[28]. We also identified new pancreatic endocrine cell surface markers, such as Sirpa, a membrane protein that acts on neighboring cells through interactions with CD47[29]. Here, we demonstrate that Sirpa is highly enriched in beta but also delta cells. Sirpa has recently been shown to enhance beta cell viability[22] and our results suggest potential protective roles in delta cells as well. Finally, our sensitive approach allowed for deep proteome profiling of the less abundant delta cells, revealing enrichment in proteins regulating vesicle

exocytosis[30] (Syt1, Snca) and neuronal-like features (Gap43, Ncam2, Spock1, Spock3, Sema3e), which may explain their unique morphology[31].

In conclusion, our results highlight the importance of cell type resolved proteomic analysis in pancreatic islets and provide deeper insights into their proteomic signatures. They offer avenues for further targeted research into the functional roles and interconnections of the presented proteins and lay the foundation for future studies in islet biology, diabetes pathogenesis, and potential therapeutic targets. Moreover, our methodology demonstrates the feasibility of deep proteomic profiling from limited biological samples, with broad implications for biomedical research. Our approach could be extended to translating these findings to human islet cells, and advancing single-cell proteomic approaches[32–34], as well as studying cell heterogeneity and the interplay of those cell types. These efforts will further enhance our understanding of pancreatic endocrine cell biology and its implications for human health and disease.

**Data availability**

All mass spectrometry raw data, libraries, and outputs from each particular search engine analyzed in this study will be deposited to the ProteomeXchange Consortium via the PRIDE partner repository upon submission.

**Acknowledgement**

This study was supported by the Max Planck Society for Advancement of Science (MM), European Union's Horizon 2020 research and innovation program under grant agreement No. 874839 (ISLET) (MM and RS), and the Laboratoire d'Excellence consortium Revive (Investissements d'Avenir ANR-10-LABX-73-01) (RS). AV is supported by postdoctoral grant from Laboratoire d'Excellence consortium Revive. MT and MW acknowledge the International Max Planck Research School for Life Sciences – IMPRS-LS. We thank our colleagues in the Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry and at the Center for Protein Research at Copenhagen University, for discussions and support. Open Access funding enabled and organized by Projekt DEAL.

**Author information**

MT and MM conceptualized and designed the study. MT, AV, AF, CL, MO, and LR performed experiments. MT, AV, MW, ADB, and MM analyzed the data. MT, AV and MM wrote the manuscript.

**Ethics declaration**

The authors declare no competing interests.

## Material and Methods

### Animals

Twelve weeks old C57BL/6JRj male mice were purchased from Janvier Labs (Saint Berthevin, France). Mice were kept on a 12:12 light-dark cycle and provided with water and food ad libitum. They were euthanized by cervical dislocation. The animal studies complied with the ARRIVE guidelines and were conducted accordance with the EU Directive 2010/63/EU for animal experiments and with regard to specific INSERM guidelines. All the experiments were approved by the Ethical Committee of Paris Cité University and the French Ministry of Higher Education and Research (#16376-2017122210502504).

### Isolation, culture and treatment of mouse pancreatic islets

Mouse islet isolation was performed as described[17]. Briefly, type V collagenase (#C9263, Sigma Aldrich, St. Louis, MO, USA) was injected into the common bile duct while the ampulla of Vater was clamped. Inflated pancreas was collected and digested for 20 min at 37°C. Digested pancreas was resuspended in HBSS (#14025, Thermo Fisher Scientific,Waltham, MA, USA) and islets were then handpicked. They were cultured for 17h in RPMI 1640 (#61870-010, Thermo Fisher Scientific) with 10% fetal calf serum (FCS CVFSVF00-01, Eurobio, Les Ulis, France) and penicillin/streptomycin (#15140122, ThermoFisher Scientific). Next, islets were treated with or without IFNγ (5ng/ml) for 20h (#485-MI-100, R&D systems, Minneapolis, MN, USA).

### Flow cytometry and sorting

Islets were dispersed in single cell suspensions using the Neural Tissue Dissociation Kit (#130-092-628, Miltenyi Biotec, Bergisch Gladbach, Germany). Cell surface staining was performed as described[17]. Briefly, dispersed islet cells were centrifuged and resuspended in FACS medium (HBSS 10% Fetal Calf Serum) with antibodies for 15 min at 4°C in the dark. The following antibodies were used: CD31 (1:100, #102522, Biolegend), CD45 (1:100, #103132, Biolegend), CD235a (1:100, #349105, Biolegend), CD24 (1:100, #101840, Biolegend), CD71 (1:100, #113806, Biolegend), CD49f (1:200, #313612, Biolegend), CD81 (1:100, #740060, BD Biosciences), SIRPA (1:100, #144008, Biolegend), EPCAM (1:100, #118210, Biolegend). Then, cells were rinsed and resuspended in FACS medium (HBSS 10% Fetal Calf Serum) with propidium iodide (1/4,000, #P4864, Sigma Aldrich). For each antibody, the optimal dilution was determined by titration. Cell sorting was carried using a FACS Aria III (BD Biosciences) and data were analysed using FlowJo™ Software 10.6.1(RRID:SCR_008520, BD Bioscience). FACS-sorted cells were washed twice with cold PBS, centrifuged at 500 g and flash-frozen until further processing.

### RNA extraction, reverse transcription, and qPCR

RNA extraction was performed using the RNeasy Micro kit (#74004, Qiagen) according to the

instruction of the manufacturer on cells sorted in RLT buffer (#74004, Qiagen, Hilden, Germany). cDNAs were obtained using Maxima First Strand cDNA synthesis Kit (#K1642, Thermo Fisher Scientific) according to the protocol of the manufacturer. qPCR was performed with Power SYBR Green Master Mix (#4367659 Thermo Fisher Scientific) on a QuantStudio 3 machine (Thermo Fisher Scientific). The relative gene expression was assessed according to the $2^{-\Delta Ct}$ x100 (expressed as % of Ppia gene expression).

**Sample preparation for proteomic profiling**

Samples were prepared as described[35]. Briefly, cell pellets were re-suspended and lysed in 2% sodium deoxycholate (SDC), 100mM Tris, pH 8.5. For reduction and alkylation of disulfide bonds 10mM TCEP (tris(2-carboxyethyl)phosphine) and 40mM chloroacetamide were added to the buffer. Samples were heated for 10 min at 95°C while shaking. Protein lysate was sonicated for 10 min at maximum energy in 30 sec on and off (Bioruptor, Diagenode, Seraing, Belgium). Trypsin (Sigma-Aldrich) and LysC (FUJIFILM Wako) were added for protein digestion in a 1:25 enzyme:protein ratio. Proteins were digested at 37°C overnight while shaking (at 800 rpm). Peptides were acidified by 1% trifluoroacetic acid (TFA) in isopropanol and purified by SDB-RPS (styrene divinylbenzene-reversed phase sulfonate) cleanup. Peptides were washed twice with 1% TFA/isopropanol and twice with 1% TFA/dd$H_2O$ before elution with 5% $NH_3$ in 80% acetonitrile (ACN). Dried eluates were reconstituted in water with 2% ACN and 0.1% TFA.

**Mass spectrometric acquisition by LC-MS**

The reconstituted samples were measured with an EASY-nLC 1200 (Thermo Fisher Scientific) nano-flow liquid chromatography system (nLC). A home packed column (50 cm, 75µm i.d.) with a pulled emitter tip, packed with 1.9 µm C18-coated porous silica beads (Dr. Maisch, Ammerbuch-Entringen, Germany) was used at 60°C and with a constant flow of 300 nL/min. Mobile phases A and B were water with 0.1% formic acid (v/v) and 80/20/0.1% ACN/water/formic acid(v/v/vol), respectively. In 120-min gradient, peptides were separated with a linear gradient from 3 to 30% buffer B within 95 min, followed by an increase to 60% buffer B within 5 min, washed at 95% within 10 min and re-equilibrated for 10 min. The nLC was coupled to a trapped ion mobility spectrometry quadrupole time of flight mass spectrometer (timsTOF Pro 2, Bruker Daltonics, Bremen, Germany) with a nano-electrospray ion source (CaptiveSpray, Bruker Daltonics). The mass spectrometer was operated in dia-PASEF mode using the 16 dia-PASEF scan acquisition scheme (standard scheme)[36]. The method covered a m/z range from 400 to 1200 and ion mobility of 0.6 to 1.6 Vs $cm^{-2}$. All other settings were standard as described[32]. Briefly, ion accumulation and ramp time in the dual TIMS analyzer was set to 100 ms each and the collision energy was increased linearly as a function of mobility starting from 20 eV at 1/K0 = 0.6 Vs $cm^{-2}$ to 59 eV at 1/K0 = 1.6 Vs $cm^{-2}$.

### Raw data analysis with DIA-NN for proteomic analysis

Mass spectrometry raw data for proteomic analysis were searched using the DIA-NN software[37,38] (version 1.8.1) using the library-free search and the mouse Uniprot databases (UP000000589_10090 with isoforms, June 24, 2015). In short, a deep-learning module, match-between-runs (MBR) and heuristic protein inference ('--relaxed-prot-inf') was enabled. N-terminal methionine excision and carbamidomethylation were set as fixed modifications, 'IDs, RT & IM profiling' was used for library generation, 'robust LC (high accuracy)' for quantification and 'Global' for cross-run normalization. One miss cleavage was allowed. The false discovery rate (FDR) was controlled at less than 1% for peptide spectrum matches and protein group identifications. The 'pg_matrix.tsv' output file was used for further data analysis.

### Proteomics downstream data analysis

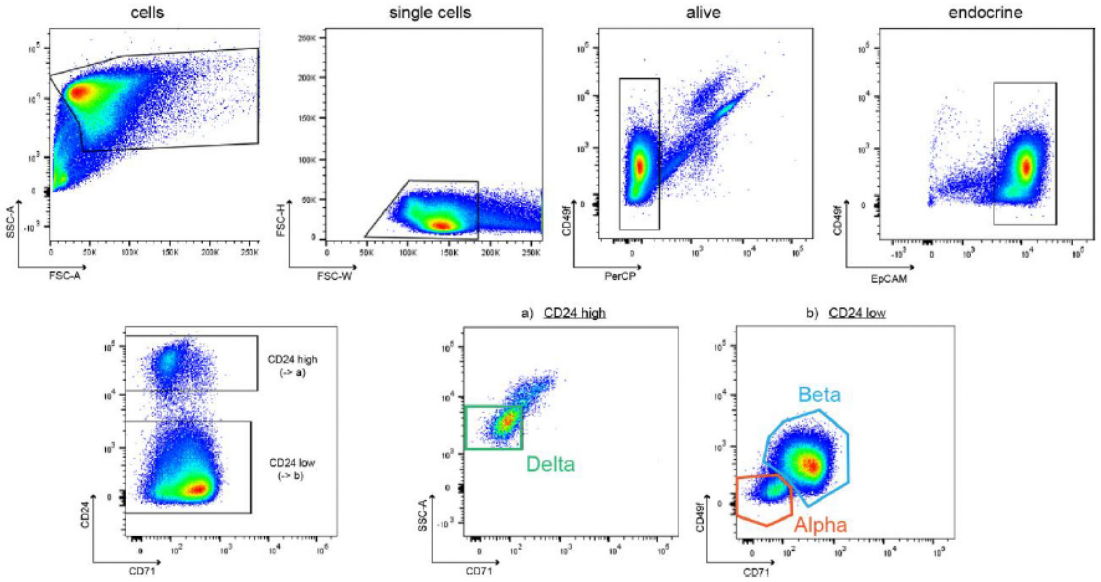
Bioinformatic analysis and visualization was performed in either Python (Jupyter Notebook), Perseus[39] (v1.6.7.0) or the R statistical computing environment (version 2023.09.1).

### References


1. Marshall, S. M. The pancreas in health and in diabetes. *Diabetologia* **63**, 1962–1965 (2020).
2. Lickert, H. & Kaestner, K. H. Islet biology. *Mol. Metab.* **6**, vi (2017).
3. Shih, H. P., Wang, A. & Sander, M. Pancreas Organogenesis: From Lineage Determination to Morphogenesis. *Annu. Rev. Cell Dev. Biol.* **29**, 81–105 (2013).
4. Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst.* **3**, 346-360.e4 (2016).
5. Xie, G. *et al.* NKX2-2 based nuclei sorting on frozen human archival pancreas enables the enrichment of islet endocrine populations for single-nucleus RNA sequencing. *BMC Genomics* **25**, (2024).
6. Kang, R. B. *et al.* Single-nucleus RNA sequencing of human pancreatic islets identifies novel gene sets and distinguishes β-cell subpopulations with dynamic transcriptome profiles. *Genome Med.* **15**, (2023).
7. Carrano, A. C., Mulas, F., Zeng, C. & Sander, M. Interrogating islets in health and disease with single-cell technologies. *Mol. Metab.* **6**, 991–1001 (2017).
8. Tarifeño-Saldivia, E. *et al.* Transcriptome analysis of pancreatic cells across distant species highlights novel important regulator genes. *BMC Biol.* **15**, (2017).
9. Aebersold, R. & Mann, M. *Mass-spectrometric exploration of proteome structure and function*. *Nature* vol. 537 347–355 (Nature Publishing Group, 2016).
10. Mitok, K. A. *et al.* Islet proteomics reveals genetic variation in dopamine production resulting in altered insulin secretion. *J. Biol. Chem.* **293**, 5860–5877 (2018).
11. Brownrigg, G. P. *et al.* Sex differences in islet stress responses support female β cell resilience. *Mol. Metab.* **69**, (2023).
12. Sacco, F. *et al.* Glucose-regulated and drug-perturbed phosphoproteome reveals molecular mechanisms controlling insulin secretion. *Nat. Commun.* **7**, 1–13 (2016).
13. Sacco, F. *et al.* Phosphoproteomics Reveals the GSK3-PDX1 Axis as a Key Pathogenic Signaling Node in Diabetic Islets. *Cell Metab.* **29**, 1422-1432.e3 (2019).
14. Berthault, C., Staels, W. & Scharfmann, R. Purification of pancreatic endocrine subsets reveals increased iron metabolism in beta cells. *Mol. Metab.* **42**, 101060 (2020).
15. Meier, F., Park, M. A. & Mann, M. Trapped Ion Mobility Spectrometry and Parallel Accumulation-Serial Fragmentation in Proteomics Authors. *Mol Cell Proteomics* **20**, 100138 (2021).
16. Hansen, F. M. *et al.* Data-independent acquisition method for ubiquitinome analysis reveals regulation of circadian biology. *Nat. Commun. 2021 121* **12**, 1–13 (2021).
17. De Burghgrave, M. *et al.* Pancreatic Islet Cells Response to IFNγ Relies on Their Spatial Location within an Islet. *Cells* **12**, (2022).

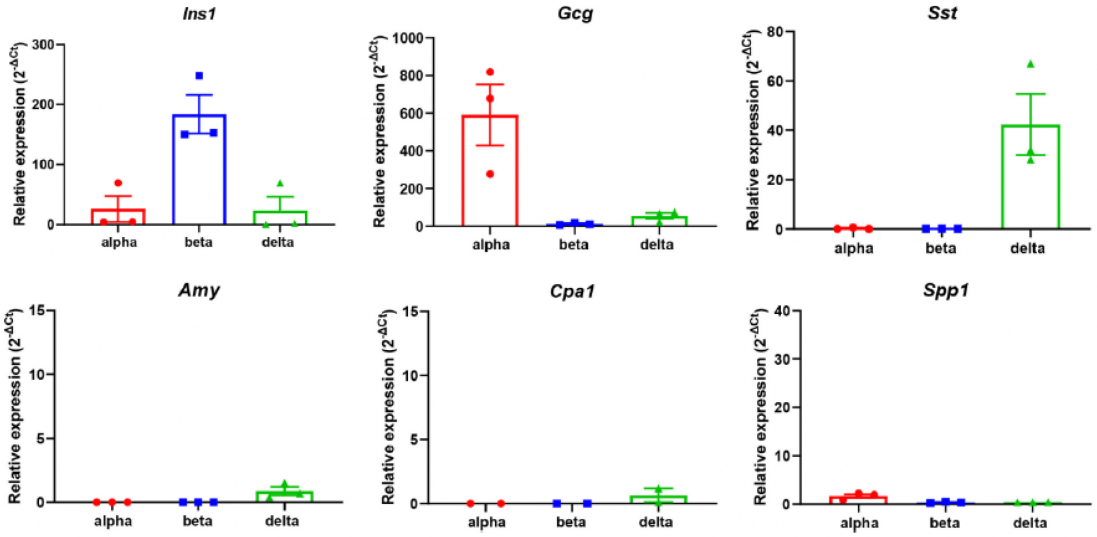
18. Coomans de Brachène, A. *et al.* Interferons are key cytokines acting on pancreatic islets in type 1 diabetes. *Diabetologia* **67**, 908–927 (2024).
19. Waanders, L. F. *et al.* Quantitative proteomic analysis of single pancreatic islets. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 18902–18907 (2009).
20. Tai, W. *et al.* Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell. Mol. Immunol. 2020 176* **17**, 613–620 (2020).
21. Laboute, T. *et al.* Orphan receptor GPR158 serves as a metabotropic glycine receptor: mGlyR. *Science* **379**, 1352–1358 (2023).
22. Leslie, K. A., Russell, M. A., Taniguchi, K., Richardson, S. J. & Morgan, N. G. The transcription factor STAT6 plays a critical role in promoting beta cell viability and is depleted in islets of individuals with type 1 diabetes. *Diabetologia* **62**, 87–98 (2019).
23. Dyring-Andersen, B. *et al.* Spatially and cell-type resolved quantitative proteomic atlas of healthy human skin. *Nat. Commun.* **11**, 1–14 (2020).
24. Alfredsson, J. *et al.* Isobaric labeling-based quantitative proteomics of FACS-purified immune cells and epithelial cells from the intestine of Crohn's disease patients reveals proteome changes of potential importance in disease pathogenesis. *Proteomics* **23**, (2023).
25. Yadav, V. K., Berger, J. M., Singh, P., Nagarajan, P. & Karsenty, G. Embryonic osteocalcin signaling determines lifelong adrenal steroidogenesis and homeostasis in the mouse. *J. Clin. Invest.* **132**, (2022).
26. Khrimian, L. *et al.* Gpr158 mediates osteocalcin's regulation of cognition. *J. Exp. Med.* **214**, 2859–2873 (2017).
27. Prakash, N. *et al.* Nkx6-1 controls the identity and fate of red nucleus and oculomotor neurons in the mouse midbrain. *Development* **136**, 2545–2555 (2009).
28. Esterházy, D. *et al.* Bace2 is a β cell-enriched protease that regulates pancreatic β cell function and mass. *Cell Metab.* **14**, 365–377 (2011).
29. Morrissey, M. A., Kern, N. & Vale, R. D. CD47 Ligation Repositions the Inhibitory Receptor SIRPA to Suppress Integrin Activation and Phagocytosis. *Immunity* **53**, 290-302.e6 (2020).
30. Shin, O. H. Exocytosis and synaptic vesicle function. *Compr. Physiol.* **4**, 149–175 (2014).
31. Rorsman, P. & Huising, M. O. The somatostatin-secreting pancreatic δ-cell in health and disease. *Nat. Rev. Endocrinol.* **14**, 404–414 (2018).
32. Brunner, A.-D. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **18**, e10798 (2022).
33. Rosenberger, F. A., Thielert, M. & Mann, M. Making single-cell proteomics biologically relevant. *Nat. Methods* **20**, 320–323 (2023).
34. Thielert, M. *et al.* Robust dimethyl-based multiplex-DIA doubles single-cell proteome depth via a reference channel. *Mol. Syst. Biol.* **19**, 2022.12.02.518917 (2023).
35. Wigger, L. *et al.* Multi-omics profiling of living human pancreatic islet donors reveals heterogeneous beta cell trajectories towards type 2 diabetes. *Nat. Metab.* **3**, 1017–1031 (2021).
36. Meier, F. *et al.* diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat. Methods* **17**, 1229–1236 (2020).
37. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).
38. Demichev, V. *et al.* dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nat. Commun.* **13**, 1–8 (2022).
39. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

## Supplemental Figures



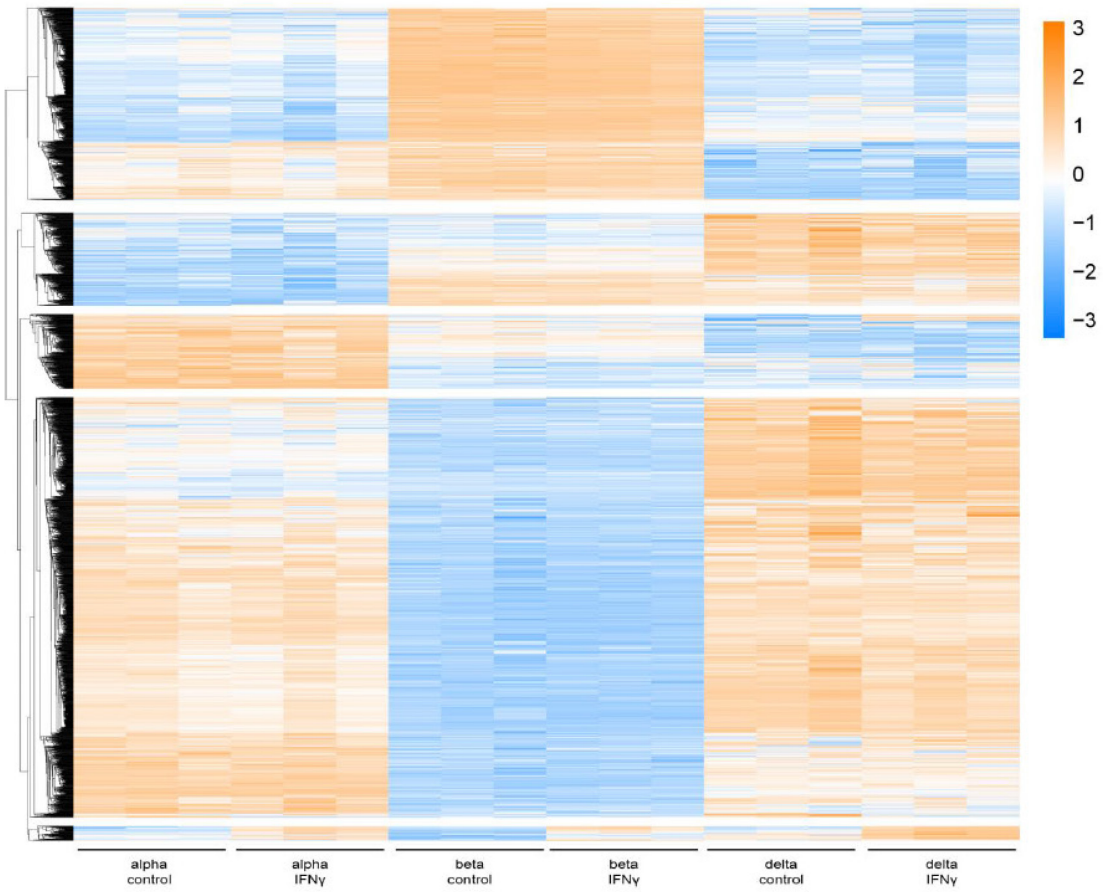

**Supplemental figure 1 – Sorting strategy.**

Cells were gated on single cells and filtered for alive and endocrine cells by negative selection on propidium iodide and positive selection of Epcam. Delta cells were extracted by a CD24 high gate while alpha and beta were separated by CD71 in the CD24 low fraction.




**Supplemental figure 2 – Quality control markers by qPCR.**

Quality control of sorted alpha, beta and delta cell populations analyzed by RT-qPCR of bonafide markers for alpha (*Gcg*), beta (*Ins1*), delta (*Sst*), acinar (*Amy* and *Cpa1*) and ductal (*Spp1*) cells.




**Supplemental figure 3 – Global protein profiles.**

Heat map of protein expression as z-score per protein across all samples. ANOVA significant ($p<0.01$) proteins are displayed in five categories according to hierarchical clustering. 67% of identified proteins changed across any condition. The majority of clusters is cell type-driven, while only one cluster (less than 3% of proteins) differ especially upon IFNγ treatment.

# 4. Discussion and Outlook

Coming from the unit of life – the cell – to the molecular drivers of function – the proteins – the goal of understanding gene expression is to analyze single cells on omics and spatial levels. MS-based proteomics has made great efforts and achievements over the last decades, constantly pushing its boundaries to reach the next levels. This unfolds fundamentally new possibilities and has just recently opened up the single-cell and spatial levels to the proteomics field. This progress is demonstrated by the identification of less than 1,000 in the beginning to now about 5,000 protein groups[103,128,190] and has been made possible by multiple technological advances and key principles. I have contributed to some of these in this thesis, including method development on scan modes and mass spectrometer themselves, as well as multiplexing of sc-proteomics. Also, we here demonstrate first single-cell spatial applications. With the new instrumentation of timsTOF Ultra from Bruker and Orbitrap Astral from Thermo Fisher, there is set to be continuing competition and progress by the mass spectrometry companies. Furthermore, there is great interest in biological applications that are becoming possible with sc-proteomics.

## 4.1. Improvements in data-acquisition

While new instrumentation is important, the DIA scan mode is crucial for sc-proteomics to enable the analysis of deep and quantitative accurate proteomes of single cells with high data completeness. In this thesis, I have described two major contributions for dia-PASEF with optimal method design (**Article 1**) and synchro-PASEF (**Article 2**). Both strategies enabled short scan cycle times and high duty cycle, resulting in a more efficient usage of ions, thus improving sensitivity. Additionally, short cycle times come with a higher sampling rate and enable higher quantitative accuracy. Conversely, these methods allow longer maximum injection times or accumulation ramps in TIMS, and thus open up potential for increasing sensitivity. Furthermore, improved search algorithms will influence DIA method design by their enhanced ability to deconvolute complex spectra. With this, the complexity of spectra, which are still feasible to be analyzed, may increase and even approaches that fragment the full m/z range may become viable[201]. The ultimate goal would be a scan mode with a single isolation window, which would open up the potential of flexibility for high sensitivity and high throughput of short gradients without compromising coverage and quantitative accuracy.

## 4.2. Single-cell proteomics

In this thesis, we enabled the analysis of single cells in a label-free manner based on our previous scan mode designs (**Article 3**). We identified 1,500 protein groups on average and up to 2,000 protein groups in individual cells. By the introduction of a reference channel in mDIA, this increased to 2,400 protein groups and up to 4,000 protein groups (**Article 4**). As mentioned in our perspective (**Article 5**), the latest preprints on the newly introduced Orbitrap Astral report up to 5,000 protein groups for single cells or single-cell HeLa equivalent amounts in combination with the cellenOne instrument for sample preparation[128,190]. The single-cell field will continue to push the boundaries with new technologies and increased sensitivity. In particular the Orbitrap Astral is not yet fully optimized with regards to ion usage and experimental design. This will enable the application of sc-proteomics to a broader audience, and in this context, standardization is important to reach its full capabilities for the community.

We established that the proteome is stable compared to the more stochastic transcriptome and confirmed this at a relatively high proteomic depth (**Article 3** and **Article 4**). Supporting this concept, sc-proteomics displays lower variability of gene expression which suggests that sc-proteomics may not need the throughput of scRNA-seq for biological relevance. To calculate needed sample sizes in case the standard deviation scales quadratically[286], we note that sc-proteomics with a fifth smaller CV would only need 400 cells, compared to 10,000 cells in scRNA-seq. Therefore sc-proteomics may already be a tool for biological applications, which is already indicated by the results of this thesis (**Article 3**, **Article 4** and **Article 7**). However, these studies are just the beginning of a plethora of biological applications, which will ultimately greatly enhance our biological understanding by sc-proteomics.

Despite these promising achievements, sc-proteomics is still in its early development. scRNA-seq has already developed powerful tools to deal with data interpretation such as cell type characterization as well as data correction for normalization, missing values and batch correction[68,287,288]. sc-proteomics should adapt many of these tools and methods. One interesting approach is RNA velocity, a time derivative of the gene expression state. It estimates the dynamic changes from a static snapshot measurement by calculating the ratio of spliced RNA versus unspliced RNA[69]. Transferring this concept to protein velocity, the dynamic change is calculated by the ratio of protein against spliced RNA[289]. This is one reason why it would be highly valuable to measure sc-proteomics and scRNA-seq from the same single cell.

Importantly, the technological advances of sc-proteomics can be transferred to other intriguing fields which are limited by sample amounts. In this thesis, this has been exemplary shown on immunopeptidomics (**Article 11**). Other peptidomic approaches would gain similarly. Furthermore, PTMs often have low stoichiometries and are thus sensitivity limited. Phosphorylation is especially interesting, since this PTM modulates many signaling pathways which are often dysregulated in diseases such as cancer or diabetes. Until recently, over 100 µg of protein material was needed (about a million cells) to measure deep phospho proteomic profiles. Some studies have already shown the promise of significantly lowering the input material[290,291]. However, more efforts are needed to realize low input phospho proteomics.

### 4.3. Multiplexed-DIA and the reference channel concept

Especially mDIA could enhance phospho proteomics. We introduced the reference channel concept by dimethyl labeling and multiplexed-DIA (mDIA) (**Article 4**). One channel consists of higher input amount and is representative of the samples in the other channels – the target channels. Due to non-isobaric labeling, the channels are isolated from each other which results in decoupling of identification and quantification. This implies the transfer of identifications from the reference channel to the target channels (with appropriate FDR) and the quantification by ratio estimation between reference channel and each target channel.

In phospho proteomics, the reference channel concept may help on multiple levels. First of all, it would benefit from the sensitivity transfer as shown in single cells. Second, when processed together, the reference channel could help to minimize losses at and after the phosphopeptide enrichment by saturating the plastic surfaces. Intriguingly, the reference channel might enable the phospho proteomic analysis of non-enriched samples. In using a phospho-enriched reference channel, low stoichiometric phosphopeptides in the target channels would be enhanced by the reference channel to retrieve their signals.

Another aspect of the reference channel is how to best use it in a normalization strategy. The reference channel is common to all measurement and can therefore be used to normalize each run with respect to the others. This will help to overcome batch effects in large clinical cohorts, e.g., plasma or urine for biomarker discovery. However, even more important, this allows for comparability across instruments, laboratories and time, which would simplify integration of different cohorts, which is still a major challenge[292]. Additionally, the reference channel concept allows the integration of retrospective

measurements, making longitudinal measurements viable, and supports continued data integration and interpretation, which is crucial in clinical routine.

Finally, the reference channel concept also allows for absolute quantification post measurement. Once a reference channel is benchmarked via its absolute quantification – either prior to sample measurements or even retrospectively – the absolute values can be referenced to any measured sample. This would enable the use of spike-in standards for absolute quantification and seamless integration into the clinical routine.

In summary, we have introduced the reference channel concept with multiplexed-DIA (**Article 4**), which allows enhanced sensitivity, normalization and absolute quantification in sc-proteomics, but implies many advantages for clinical settings as well.

### 4.4. The promise of spatial proteomics

The human body is a spatial arrangement of many different cell types, which coordinate their function and activity. In this, cellular heterogeneity of single cells is just one aspect. Resolving cellular heterogeneity at the spatial level is fundamental to understand the molecular mechanism and biology of humans in health and disease. Our laboratory has recently developed Deep Visual Proteomics (DVP) to disentangle spatial proteomic measurements by combining high-content imaging, AI-guided laser microdissection and ultra-high sensitive MS. Despite recent advances, spatial proteomics still faces several challenges. First, antibody-based staining allows the focused analysis of a particular research question by the characterization of specific markers. However, an unbiased analysis of universal markers might be useful to discover novel cellular morphologies. In this context, multiplexed imaging such as CODEX boosts the visual information for broader cell selection and characterization. To enable the analysis of clinical cohorts, robust and reproducible fluorescence staining on membrane slides is crucial for efficient tissue handling and to avoid losing rare clinical samples. In this context, we optimized the staining procedure by adding glycerol in the antigen retrieval process to sustain the integrity of the membrane in a reproducible manner (**Article 6**).

Spatial proteomics will further advance with increased sensitivity and throughput enabled by sc-proteomics. In this context, we connected mDIA with DVP to benefit from its advantages on the spatial level (**Article 4**). Currently, this combination enables four times the throughput on the MS and seven-fold at the LCM side because of its sensitivity and multiplexing ability, thereby also strengthen the spatial resolution through smaller cell numbers. Additionally, we already demonstrated the feasibility of single-cell DVP

(scDVP) (**Article 7**). scDVP is a combination of mDIA and DVP and analyzes a single contour at a time. We demonstrated the power of scDVP on the liver zonation along the path of portal to central vein. Despite its advantages, scDVP is limited in throughput and proteomic depth, thus the decision of using scDVP over DVP depends on the research question. Overall, spatial proteomics will benefit from sc-proteomics driven MS sensitivity improvements including the new Orbitrap Astral instrument.

DVP enables spatial proteomics in two dimensions. However, tissues and the human body are a three-dimensional arrangement. Therefore, for special cases it would be advantageous to analyze this full complexity. In this context, we have co-developed DISCO-MS which enables the spatial analysis in the three-dimensional space and intact tissues (**Article 8**). We accomplished this combination with tissue clearing, advanced imaging and deep-learning techniques. In the future, DVP could be enhanced to three-dimensions by stacking consecutive sections[293,294]. This approach combined with algorithms of single-cell matching would allow to extract a whole single cell from several sections instead of extracting single, incomplete shapes in single slices.

By its nature, DVP pools cell types and states which may not completely cover the full heterogeneity of single cells. However, the pooling of phenotypical identical cells on small regions of high similarity implies a ‘biological enrichment’, which focusses and still captures the main variability of functional and biological interest. Therefore, DVP offers straightforward biological interpretation while enabling a time-efficient workflow and great proteomic depth. In this thesis, biological meaningful results are demonstrated with unraveling mechanisms in ovarian cancer of serous borderline tumors (**Article 9**). We showed mechanistically important protein differences, such as NOVA2 which might enable the treatment by inhibition of metastasis and invasion and thus inhibit the switch in malignancy of this tumor. In addition, we have exemplified the power of DVP in cutaneous drug reactions. Here, spatial proteomics provided a molecular mechanism that extended all the way to the treatment of patients (**Article 10**). Those examples prove that these proteomic tools can pave the way towards personalized medicine to find relevant pathways in the individual case and decide on specific treatment strategies accordingly. Additionally, spatial proteomics and clinical proteomics have the opportunity to monitor clinical patients over time. Further interesting biological questions to be addressed are different malignancies in melanoma, long studies brain diseases such as Alzheimer and Parkinson’s and metabolic diseases in the liver or pancreas. Methodologically, spatial proteomics has been driven by deep-learning algorithms for image analysis such as cell segmentation, cell classification and recognition of patterns. The biological analysis will be enhanced by tools for data analysis which allow easy

integration of visual and proteomic information. With this and advanced deep-learning, spatial proteomics will be able to better integrate visual and proteomic information.

The DVP workflow is modular, open format and widely applicable allowing the use on its own or in combination with other technologies. Consecutive sections or enhanced sample preparation techniques make the application of multiomics feasible, such as combinations with metabolomics like SpaceM[295] and lipidomics[296] as well as already established techniques of transcriptomics, epigenomics[297,298] and genomics. Spatially-resolved proteogenomics would enable the analysis of a plethora of genetic diseases such as cancer.

## 4.5. A vision for single-cell and spatial proteomics

Although we are only at the beginning of the journey for sc-proteomics and spatial proteomics, a multitude of biological questions can already be addressed from the new experimental angle outlined in this thesis. sc-proteomics will further enhance and open up opportunities for biological applications which may revolutionize our understanding of biology. Specifically, spatial proteomics can understand the connection of cells and disentangle the fundamental concept of biomolecular gradients. Both sc-proteomics and DVP workflows are modular and open formats allowing the use on its own or in combination with other omic techniques.

In conclusion, this thesis presented technological advances in sc-proteomics and spatial proteomics with clinical and biological relevant applications, demonstrating the power of both tools, and their combination. For the future, this underscores the potential of further developments to enhance system-wide biological and clinical studies for understanding the functional mechanisms in health and disease, and paves the way for precision medicine.

# 5. Bibliography

## 5.1. References


1. Virchow, R. *Die Einheitsbestrebungen in der wissenschaftlichen Medicin*. (Reimer, 1849).
2. Schleiden, M. J. *Schleiden: Beitrage zur Phytogenesis - Google Scholar*.
3. Schwann, T. *Mikroskopische Untersuchungen über die Uebereinstimmung in der Struktur und dem Wachsthum der Thiere und Pflanzen*. (Sander, 1839).
4. Hooke, R. *Micrographia: Or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses*. (J Marthyn & J. Allestry, 1665).
5. van Leewenhoeck, A. Observations, communicated to the publisher by Mr. Antony van Leewenhoeck, in a dutch letter of the 9th Octob. 1676. here English'd: concerning little animals by him observed in rain-well-sea- and snow water; as also in water wherein pepper had lain infus. *Philos. Trans. R. Soc. London* **12**, 821–831 (1677).
6. Dutrochet, H. *Recherches anatomiques et physiologiques sur la structure intime des animaux ...*
7. Grzybowski, A. & Pietrzak, K. Robert Remak (1815-1865): Discoverer of the fungal character of dermatophytoses. *Clin. Dermatol.* **31**, 802–805 (2013).
8. Darwin, C. On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life. *Libr. Congr. Washington, D.C. 20540 USA* (1859).
9. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nat. 1953 1714356* **171**, 737–738 (1953).
10. Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644 (2020).
11. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).
12. Cobb, M. 60 years ago, Francis Crick changed the logic of biology. *PLoS Biol.* **15**, (2017).
13. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467 (1977).
14. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269**, 496–512 (1995).
15. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546–567 (1996).
16. Waterston, R. & Sulston, J. The genome of Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 10836–10840 (1995).
17. Adams, M. D. *et al.* The genome sequence of Drosophila melanogaster. *Science* **287**, 2185–2195 (2000).
18. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19096–19101 (2009).
19. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Mol. Cell* **58**, 586–597 (2015).
20. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351 (2016).
21. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
22. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nat. Publ. Gr.* (2016) doi:10.1038/nrg.2015.16.
23. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
24. Abdellah, Z. *et al.* Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
25. Craig Venter, J. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
26. Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
27. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
28. Velculescu, V. E. *et al.* Characterization of the Yeast Transcriptome. *Cell* **88**, 243–251 (1997).
29. Panni, S., Lovering, R. C., Porras, P. & Orchard, S. Non-coding RNA regulatory networks. *Biochim. Biophys. acta. Gene Regul. Mech.* **1863**, (2020).
30. Koch, L. Adding another dimension to gene regulation. *Nat. Rev. Genet.* **16**, 563–563 (2015).
31. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome.

*Nat. Rev. Genet.* **20**, 207–220 (2019).

32. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15**, 163–175 (2014).
33. Jumaa, H. & Nielsen, P. J. The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.* **16**, 5077 (1997).
34. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38**, D750 (2010).
35. Volpe, P. & Eremenko-Volpe, T. Quantitative studies on cell proteins in suspension cultures. *Eur. J. Biochem.* **12**, 195–200 (1970).
36. Beck, M. *et al.* The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, (2011).
37. Muñoz, J. & Heck, A. J. R. From the Human Genome to the Human Proteome. *Angew. Chemie Int. Ed.* **53**, 10864–10866 (2014).
38. Ponomarenko, E. A. *et al.* The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.* **2016**, (2016).
39. Smith, L. M. *et al.* Proteoform: a single term describing protein complexity. *Nat. Methods* **10**, 186–187 (2013).
40. Aebersold, R. *et al.* How many human proteoforms are there? *Nat. Chem. Biol.* **14**, 206–214 (2018).
41. Olsen, J. V. *et al.* Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).
42. Choudhary, C. & Mann, M. Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.* **11**, 427–439 (2010).
43. Barrios-Rodiles, M. *et al.* High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* **307**, 1621–1625 (2005).
44. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).
45. Doll, S. *et al.* Rapid proteomic analysis for solid tumors reveals LSD1 as a drug target in an end-stage cancer patient. *Mol. Oncol.* **12**, 1296–1307 (2018).
46. McConnell, M. J. *et al.* Mosaic copy number variation in human neurons. *Science* **342**, 632–637 (2013).
47. Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160 (2014).
48. Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
49. Sneddon, T. P., Li, P. & Edmunds, S. C. GigaDB: Announcing the GigaScience database. *Gigascience* **1**, (2012).
50. Tsompana, M. & Buck, M. J. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin* **7**, (2014).
51. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
52. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
53. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).
54. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
55. Schaum, N. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
56. Elmentaite, R., Domínguez Conde, C., Yang, L. & Teichmann, S. A. Single-cell atlases: shared and tissue-specific cell types across human organs. *Nat. Rev. Genet.* **23**, 395–410 (2022).
57. Method of the Year 2013. *Nat. Methods* **11**, 1–1 (2013).
58. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
59. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
60. Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
61. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-

seq3. *Nat. Biotechnol.* **38**, 708–714 (2020).

62. Natarajan, K. N. Single-cell tagged reverse transcription (STRT-Seq). *Methods Mol. Biol.* **1979**, 133–153 (2019).
63. Izar, B. *et al.* A single-cell landscape of high-grade serous ovarian cancer. *Nat. Med.* **26**, 1271–1279 (2020).
64. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
65. Jones, R. C. *et al.* The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science (80-. ).* **376**, (2022).
66. Suo, C. *et al.* IMMUNOLOGY Mapping the developing human immune system across organs. *Science (80-. ).* **376**, (2022).
67. Eraslan, G. *et al.* Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science* **376**, (2022).
68. Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A. & Theis, F. J. A test metric for assessing single-cell RNA-seq batch correction. *Nat. Methods* **16**, 43–49 (2019).
69. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
70. Fischer, D. S. *et al.* Inferring population dynamics from single-cell RNA-sequencing time series data. *Nat. Biotechnol.* **37**, 461–468 (2019).
71. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
72. Subkhankulova, T., Gilchrist, M. J. & Livesey, F. J. Modelling and measuring single cell RNA expression levels find considerable transcriptional differences among phenotypically identical cells. *BMC Genomics* **9**, 1–12 (2008).
73. Eberwine, J., Sul, J. Y., Bartfai, T. & Kim, J. The promise of single-cell sequencing. *Nat. Methods* **11**, 25–27 (2014).
74. Carter, M. G. *et al.* Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray. *Genome Biol.* **6**, 1–12 (2005).
75. Kawasaki, E. S. Microarrays and the gene expression profile of a single cell. *Ann. N. Y. Acad. Sci.* **1020**, 92–100 (2004).
76. Marguerat, S. *et al.* Quantitative Analysis of Fission Yeast Transcriptomes and Proteomes in Proliferating and Quiescent Cells. *Cell* **151**, 671–683 (2012).
77. Chung, H. *et al.* Joint single-cell measurements of nuclear proteins and RNA in vivo. *Nat. Methods* **18**, 1204–1212 (2021).
78. Brunner, A. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **18**, 2020.12.22.423933 (2022).
79. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).
80. Kennedy-Darling, J. *et al.* Highly multiplexed tissue imaging using repeated oligonucleotide exchange reaction. *Eur. J. Immunol.* **51**, 1262–1277 (2021).
81. Black, S. *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat. Protoc.* **16**, 3802–3835 (2021).
82. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
83. Peterson, V. M. *et al.* Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939 (2017).
84. Chen, A. F. *et al.* NEAT-seq: simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells. *Nat. Methods* **19**, 547–553 (2022).
85. Gerlach, J. P. *et al.* Combined quantification of intracellular (phospho-)proteins and transcriptomics from fixed single cells. *Sci. Rep.* **9**, 1–10 (2019).
86. Wik, L. *et al.* Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis. *Mol. Cell. Proteomics* **20**, 100168 (2021).
87. Gold, L. *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS One* **5**, e15004 (2010).
88. Fredriksson, S. *et al.* Protein detection using proximity-dependent DNA ligation assays. *Nat. Biotechnol.* **20**, 473–477 (2002).
89. Schueder, F. *et al.* Super-Resolution Spatial Proximity Detection with Proximity-PAINT. *Angew. Chemie Int. Ed.* **60**, 716–720 (2021).
90. Keren, L. *et al.* MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure. *Sci. Adv.* **5**, (2019).

91. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422 (2014).
92. Bodenmiller, B. *et al.* Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat. Biotechnol.* **30**, 858–867 (2012).
93. Bodenmiller, B. Multiplexed Epitope-Based Tissue Imaging for Discovery and Healthcare Applications. *Cell Syst.* **2**, 225–238 (2016).
94. Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
95. Wagner, J. *et al.* A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. *Cell* **177**, 1330-1345.e18 (2019).
96. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620 (2020).
97. Wiener, J., Kokotek, D., Rosowski, S., Lickert, H. & Meier, M. Preparation of single- and double-oligonucleotide antibody conjugates and their application for protein analytics. *Sci. Rep.* **10**, 1–11 (2020).
98. Aebersold, R. & Mann, M. *Mass-spectrometric exploration of proteome structure and function*. *Nature* vol. 537 347–355 (Nature Publishing Group, 2016).
99. Bekker-Jensen, D. B. *et al.* An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* **4**, 587-599.e4 (2017).
100. Meier, F., Geyer, P. E., Virreira Winter, S., Cox, J. & Mann, M. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat. Methods* **15**, 440–448 (2018).
101. Muntel, J. *et al.* Comparison of Protein Quantification in a Complex Background by DIA and TMT Workflows with Fixed Instrument Time. *J. Proteome Res.* **18**, 1340–1351 (2019).
102. Rosenberger, F. A., Thielert, M. & Mann, M. Making single-cell proteomics biologically relevant. *Nat. Methods* **20**, 320–323 (2023).
103. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, (2018).
104. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71 (1989).
105. Mann, M., Meng, C. K. & Fenn, J. B. Interpreting Mass Spectra of Multiply Charged Ions. *Anal. Chem.* **61**, 1702–1708 (1989).
106. Mann, M. & Fenn, J. The ever expanding scope of electrospray mass spectrometry—a 30 year journey. *Nat. Commun.* **10**, 1–3 (2019).
107. Kafader, J. O. *et al.* Measurement of Individual Ions Sharply Increases the Resolution of Orbitrap Mass Spectra of Proteins. *Anal. Chem.* **91**, 2776–2783 (2019).
108. Tran, J. C. *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254–258 (2011).
109. Pandeswari, P. B. & Sabareesh, V. Middle-down approach: a choice to sequence and characterize proteins/proteomes by mass spectrometry. *RSC Adv.* **9**, 313 (2018).
110. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
111. Nesvizhskii, A. I. & Aebersold, R. Interpretation of Shotgun Proteomic Data. *Mol. Cell. Proteomics* **4**, 1419–1440 (2005).
112. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
113. Batth, T. S. *et al.* Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation. *Mol. Cell. Proteomics* **18**, 1027 (2019).
114. Hailemariam, M. *et al.* S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics. *J. Proteome Res.* **17**, 2917–2924 (2018).
115. Müller, T. *et al.* Automated sample preparation with SP 3 for low-input clinical proteomics . *Mol. Syst. Biol.* **16**, 9111 (2020).
116. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
117. Chang, Y. H. *et al.* New mass-spectrometry-compatible degradable surfactant for tissue proteomics. *J. Proteome Res.* **14**, 1587–1599 (2015).
118. Liu, J. *et al.* High-sensitivity N-glycoproteomic analysis of mouse brain tissue by protein extraction with a mild detergent of N-dodecyl β-D-maltoside. *Anal. Chem.* **87**, 2054–2057 (2015).

119. Tsai, C. F. *et al.* Surfactant-assisted one-pot sample preparation for label-free single-cell proteomics. *Commun. Biol.* **4**, 1–12 (2021).

120. Liang, Y. *et al.* Fully automated sample processing and analysis workflow for low-input proteome profiling. *Anal. Chem.* **93**, 1658–1666 (2021).

121. Schoof, E. M. *et al.* Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nat. Commun.* **12**, 1–15 (2021).

122. Specht, H. *et al.* Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* **22**, (2021).

123. Gebreyesus, S. T. *et al.* Streamlined single-cell proteomics by an integrated microfluidic chip and data-independent acquisition mass spectrometry. *Nat. Commun.* **13**, 1–13 (2022).

124. Woo, J. *et al.* High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. *Nat. Commun.* **12**, 1–11 (2021).

125. Zhu, Y. *et al.* Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. *Nat. Commun.* **9**, 1–10 (2018).

126. Matzinger, M., Müller, E., Dürnberger, G., Pichler, P. & Mechtler, K. Robust and Easy-to-Use One-Pot Workflow for Label-Free Single-Cell Proteomics. *Anal. Chem.* **95**, 4435–4445 (2023).

127. Ctortecka, C. *et al.* An Automated Nanowell-Array Workflow for Quantitative Multiplexed Single-Cell Proteomics Sample Preparation at High Sensitivity. *Mol. Cell. Proteomics* **22**, 100665 (2023).

128. Ye, Z. *et al.* High-throughput and scalable single cell proteomics identifies over 5000 proteins per cell. *bioRxiv* 2023.11.27.568953 (2023) doi:10.1101/2023.11.27.568953.

129. Claudia Ctortecka, A. *et al.* Automated single-cell proteomics providing sufficient proteome depth to study complex biology beyond cell type classifications. *bioRxiv* 2024.01.20.576369 (2024) doi:10.1101/2024.01.20.576369.

130. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B. & Aebersold, R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat. Methods* **4**, 231–237 (2007).

131. Andersson, L. & Porath, J. Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. *Anal. Biochem.* **154**, 250–254 (1986).

132. Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B. & Heck, A. J. R. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* **76**, 3935–3943 (2004).

133. Russell, J. D. & Murphy, S. Agilent AssayMAP Bravo Technology Enables Reproducible Automated Phosphopeptide Enrichment from Complex Mixtures Using High-Capacity Fe(III)-NTA Cartridges Application Note Automated Peptide Sample Preparation for LC/MS.

134. Molnar, I. & Horvath, C. Reverse-phase chromatography of polar biological substances: separation of catechol compounds by high-performance liquid chromatography. *Clin. Chem.* **22**, 1497–1502 (1976).

135. van Deemter, J. J., Zuiderweg, F. J. & Klinkenberg, A. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chem. Eng. Sci.* **50**, 3869–3882 (1995).

136. Gritti, F. & Guiochon, G. The van Deemter equation: assumptions, limits, and adjustment to modern high performance liquid chromatography. *J. Chromatogr. A* **1302**, 1–13 (2013).

137. Bian, Y. *et al.* Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS. *Nat. Commun.* **11**, (2020).

138. Kuster, B. *et al.* Robust Microflow LC-MS/MS for Proteome Analysis: 38 000 Runs and Counting. *Anal. Chem.* **93**, 3686–3690 (2021).

139. Zheng, R. *et al.* A High-Sensitivity Low-Nanoflow LC-MS Configuration for High-Throughput Sample-Limited Proteomics. *Anal. Chem.* **95**, 18673–18678 (2023).

140. Webber, K. G. I. *et al.* Label-Free Profiling of up to 200 Single-Cell Proteomes per Day Using a Dual-Column Nanoflow Liquid Chromatography Platform. *Anal. Chem.* **94**, 6017–6025 (2022).

141. Cong, Y. *et al.* Improved Single-Cell Proteome Coverage Using Narrow-Bore Packed NanoLC Columns and Ultrasensitive Mass Spectrometry. *Anal. Chem.* **92**, 2665–2671 (2020).

142. Marginean, I., Tang, K., Smith, R. D. & Kelly, R. T. Picoelectrospray ionization mass spectrometry using narrow-bore chemically etched emitters. *J. Am. Soc. Mass Spectrom.* **25**, 30–36 (2014).

143. Xiang, P. *et al.* Picoflow Liquid Chromatography-Mass Spectrometry for Ultrasensitive Bottom-Up Proteomics Using 2-µm-i.d. Open Tubular Columns. *Anal. Chem.* **92**, 4711–4715 (2020).

144. Kreimer, S. *et al.* Parallelization with Dual-Trap Single-Column Configuration Maximizes Throughput of Proteomic Analysis. *Anal. Chem.* **94**, 12452–12460 (2022).

145. Hahne, H. *et al.* DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat. Methods* **10**, 989–991 (2013).

146. Battellino, T., Ogata, K., Spicer, V., Ishihama, Y. & Krokhin, O. Acetic Acid Ion Pairing Additive for

Reversed-Phase HPLC Improves Detection Sensitivity in Bottom-up Proteomics Compared to Formic Acid. *J. Proteome Res.* **22**, 272–278 (2023).

147. Orsburn, B. C. Acetic acid is a superior acidifier for sub-nanogram and single cell proteomic studies. *bioRxiv* 2023.08.01.551522 (2023) doi:10.1101/2023.08.01.551522.
148. Bache, N. *et al.* A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Mol. Cell. Proteomics* **17**, 2284–2296 (2018).
149. Rappsilber, J., Ishihama, Y. & Mann, M. Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
150. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324 (2014).
151. Binai, N. A. *et al.* Rapid analyses of proteomes and interactomes using an integrated solid-phase extraction-liquid chromatography-MS/ms system. *J. Proteome Res.* **14**, 977–985 (2015).
152. Stejskal, K., Potěšil, D. & Zdráhal, Z. Suppression of peptide sample losses in autosampler vials. *J. Proteome Res.* **12**, 3057–3062 (2013).
153. Gritti, F. & Gilar, M. Impact of frit dispersion on gradient performance in high-throughput liquid chromatography. *J. Chromatogr. A* **1591**, 110–119 (2019).
154. Yue, G., Luo, Q., Zhang, J., Wu, S. L. & Karger, B. L. Ultratrace LC/MS proteomic analysis using 10-microm-i.d. Porous layer open tubular poly(styrene-divinylbenzene) capillary columns. *Anal. Chem.* **79**, 938–946 (2007).
155. Li, S. *et al.* An Integrated Platform for Isolation, Processing, and Mass Spectrometry-based Proteomic Profiling of Rare Cells in Whole Blood. *Mol. Cell. Proteomics* **14**, 1672–1683 (2015).
156. De Malsche, W. *et al.* Pressure-driven reverse-phase liquid chromatography separations in ordered nonporous pillar array columns. *Anal. Chem.* **79**, 5915–5926 (2007).
157. De Malsche, W. *et al.* Capillary liquid chromatography separations using non-porous pillar array columns. *J. Chromatogr. A* **1230**, 41–47 (2012).
158. Stadlmann, J. *et al.* Improved Sensitivity in Low-Input Proteomics Using Micropillar Array-Based Chromatography. *Anal. Chem.* **91**, (2019).
159. Petrosius[1], V. *et al.* Enhancing single-cell proteomics through tailored Data-Independent Acquisition and micropillar array-based chromatography. *bioRxiv* 2022.11.29.518366 (2022) doi:10.1101/2022.11.29.518366.
160. Stewart, H. I. *et al.* Parallelized Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis. *Anal. Chem.* **95**, 15656–15664 (2023).
161. Zhang, Z., Qu, Y. & Dovichi, N. J. Capillary zone electrophoresis-mass spectrometry for bottom-up proteomics. *TrAC Trends Anal. Chem.* **108**, 23–37 (2018).
162. Amenson-Lamar, E. A., Sun, L., Zhang, Z., Bohn, P. W. & Dovichi, N. J. Detection of 1 zmol injection of angiotensin using capillary zone electrophoresis coupled to a Q-Exactive HF mass spectrometer with an electrokinetically pumped sheath-flow electrospray interface. *Talanta* **204**, 70 (2019).
163. Kelly, R. T. Single-cell Proteomics: Progress and Prospects. *Mol. Cell. Proteomics* **19**, 1739–1748 (2020).
164. Sun, L., Zhu, G., Zhang, Z., Mou, S. & Dovichi, N. J. Third-generation electrokinetically pumped sheath-flow nanospray interface with improved stability and sensitivity for automated capillary zone electrophoresis-mass spectrometry analysis of complex proteome digests. *J. Proteome Res.* **14**, 2312–2321 (2015).
165. Page, J. S., Kelly, R. T., Tang, K. & Smith, R. D. Ionization and Transmission Efficiency in an Electrospray Ionization–Mass Spectrometry Interface. *J. Am. Soc. Mass Spectrom.* **18**, 1582–1590 (2007).
166. Page, J. S., Tang, K., Kelly, R. T. & Smith, R. D. Subambient pressure ionization with nanoelectrospray source and interface for improved sensitivity in mass spectrometry. *Anal. Chem.* **80**, 1800–1805 (2008).
167. Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *https://doi.org/10.1146/annurev-anchem-071114-040325* **8**, 61–80 (2015).
168. Cox, J. & Mann, M. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. *https://doi.org/10.1146/annurev-biochem-061308-093216* **80**, 273–299 (2011).
169. Dawson, P. H. Quadrupole mass analyzers: Performance, design and some recent applications. *Mass Spectrom. Rev.* **5**, 1–37 (1986).
170. Douglas, D. J. Linear quadrupoles in mass spectrometry. *Mass Spectrom. Rev.* **28**, 937–960 (2009).
171. Bowers, W. D., Delbert, S. S., Hunter, R. L. & Mclver, R. T. Fragmentation of Oligopeptide Ions

Using Ultraviolet Laser Radiation and Fourier Transform Mass Spectrometry. *J. Am. Chem. Soc.* **106**, 7288–7289 (1984).

172. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9528–9533 (2004).
173. Mikesh, L. M. *et al.* The utility of ETD mass spectrometry in proteomic analysis. *Biochim. Biophys. Acta* **1764**, 1811–1822 (2006).
174. McAlister, G. C., Phanstiel, D., Good, D. M., Berggren, W. T. & Coon, J. J. Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer. *Anal. Chem.* **79**, 3525–3534 (2007).
175. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601–601 (1984).
176. Bekker-Jensen, D. B. *et al.* A compact quadrupole-orbitrap mass spectrometer with FAIMS interface improves proteome coverage in short LC gradients. *Mol. Cell. Proteomics* **19**, 716–729 (2020).
177. Guzman, U. H. *et al.* Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition. *Nat. Biotechnol. 2024* 1–12 (2024) doi:10.1038/s41587-023-02099-7.
178. Stewart, H. *et al.* A Conjoined Rectilinear Collision Cell and Pulsed Extraction Ion Trap with Auxiliary DC Electrodes. *J. Am. Soc. Mass Spectrom.* (2023) doi:10.1021/jasms.3c00311.
179. Heil, L. R. *et al.* Evaluating the Performance of the Astral Mass Analyzer for Quantitative Proteomics Using Data-Independent Acquisition. *J. Proteome Res.* **22**, 3290–3300 (2023).
180. Meier, F. *et al.* Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol. Cell. Proteomics* **17**, 2534–2545 (2018).
181. Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal. Chem.* **72**, 1156–1162 (2000).
182. Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **40**, 430–443 (2005).
183. Savaryn, J. P., Toby, T. K. & Kelleher, N. L. A Researcher's Guide to Mass Spectrometry-Based Proteomics. *Proteomics* **16**, 2435 (2016).
184. Kelstrup, C. D. *et al.* Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *J. Proteome Res.* **17**, 727–738 (2018).
185. Phlairaharn, T. *et al.* Optimizing Linear Ion-Trap Data-Independent Acquisition toward Single-Cell Proteomics. *Anal. Chem.* **95**, 9881–9891 (2023).
186. Beck, S. *et al.* The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics. *Mol. Cell. Proteomics* **14**, 2014–2029 (2015).
187. Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M. & Muddiman, D. C. Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal. Chem.* **83**, 5442–5446 (2011).
188. Petrosius, V. *et al.* Evaluating the capabilities of the Astral mass analyzer for single-cell proteomics. *bioRxiv* 2023.06.06.543943 (2023) doi:10.1101/2023.06.06.543943.
189. Ye, Z. *et al.* One-Tip enables comprehensive proteome coverage in minimal cells and single zygotes. *bioRxiv* 2023.08.10.552756 (2023) doi:10.1101/2023.08.10.552756.
190. Bubis, J. A. *et al.* Challenging the Astral TM mass analyzer-going beyond 5200 proteins per single-cell at unseen quantitative 1 accuracy to study cellular heterogeneity. doi:10.1101/2024.02.01.578358.
191. Meier, F. *et al.* Parallel accumulation-serial fragmentation (PASEF): Multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J. Proteome Res.* **14**, 5378–5387 (2015).
192. Dodds, J. N. & Baker, E. S. Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *J. Am. Soc. Mass Spectrom.* **30**, 2185–2195 (2019).
193. Winter, D. L., Wilkins, M. R. & Donald, W. A. Differential Ion Mobility–Mass Spectrometry for Detailed Analysis of the Proteome. *Trends Biotechnol.* **37**, 198–213 (2019).
194. Meier, F., Park, M. A. & Mann, M. Trapped Ion Mobility Spectrometry and Parallel Accumulation-Serial Fragmentation in Proteomics Authors. *Mol Cell Proteomics* **20**, 100138 (2021).
195. Furtwängler, B. *et al.* Real-Time Search-Assisted Acquisition on a Tribrid Mass Spectrometer Improves Coverage in Multiplexed Single-Cell Proteomics. *Mol. Cell. Proteomics* **21**, (2022).
196. Huffman, R. G. *et al.* Prioritized mass spectrometry increases the depth, sensitivity and data completeness of single-cell proteomics. *Nat. Methods* **20**, 714–722 (2023).
197. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1**, 39–

45 (2004).

198. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, (2012).
199. Röst, H. L., Malmström, L. & Aebersold, R. Reproducible quantitative proteotype data matrices for systems biology. *Mol. Biol. Cell* **26**, 3926–3931 (2015).
200. Meier, F. *et al.* diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nat. Methods* **17**, 1229–1236 (2020).
201. Szyrwiel, L., Sinn, L., Ralser, M. & Demichev, V. Slice-PASEF: fragmenting all ions for maximum sensitivity in proteomics. *bioRxiv* **10**, 514544 (2022).
202. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).
203. Huang, T., Wang, J., Yu, W. & He, Z. Protein inference: a review. *Brief. Bioinform.* **13**, 586–614 (2012).
204. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).
205. Demichev, V. *et al.* dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nat. Commun.* **13**, 1–8 (2022).
206. Bruderer, R. *et al.* Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. *Mol. Cell. Proteomics* **14**, 1400–1410 (2015).
207. Zeng, W. F. *et al.* AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics. *Nat. Commun.* **13**, 1–14 (2022).
208. Gessulat, S. *et al.* Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **16**, 509–518 (2019).
209. Mann, M., Kumar, C., Zeng, W. F. & Strauss, M. T. Artificial intelligence for proteomics and biomarker discovery. *Cell Syst.* **12**, 759–770 (2021).
210. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
211. Krismer, E., Bludau, I., Strauss, M. T. & Mann, M. AlphaPeptStats: an open-source Python package for automated and scalable statistical analysis of mass spectrometry-based proteomics. *Bioinformatics* **39**, (2023).
212. Voytik, E. *et al.* AlphaViz: Visualization and validation of critical proteomics data directly at the raw data level. *bioRxiv* 2022.07.12.499676 (2022) doi:10.1101/2022.07.12.499676.
213. Voytik, E. *et al.* AlphaMap: an open-source Python package for the visual annotation of proteomics data with sequence-specific knowledge. *Bioinformatics* **38**, 849–852 (2022).
214. Willems, S., Voytik, E., Skowronek, P., Strauss, M. T. & Mann, M. AlphaTims: Indexing Trapped Ion Mobility Spectrometry–TOF Data for Fast and Easy Accession and Visualization. *Mol. Cell. Proteomics* **20**, 100149 (2021).
215. Bantscheff, M., Lemeer, S., Savitski, M. M. & Kuster, B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* **404**, 939–965 (2012).
216. Thielert, M. *et al.* Robust dimethyl-based multiplex-DIA doubles single-cell proteome depth via a reference channel. *Mol. Syst. Biol.* **19**, 2022.12.02.518917 (2023).
217. Petrosius, V. *et al.* Exploration of cell state heterogeneity using single-cell proteomics through sensitivity-tailored data-independent acquisition. *Nat. Commun.* **14**, (2023).
218. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003).
219. Li, J. *et al.* TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nat. Methods* **17**, (2020).
220. Cheung, T. K. *et al.* Defining the carrier proteome limit for single-cell proteomics. *Nat. Methods* **18**, 76–83 (2021).
221. Ye, Z., Batth, T. S., Rüther, P. & Olsen, J. V. A deeper look at carrier proteome effects for single-cell proteomics. *Commun. Biol.* **5**, (2022).
222. Derks, J. *et al.* Increasing the throughput of sensitive proteomics by plexDIA. *Nat. Biotechnol.* **41**, 50–59 (2023).
223. Čuklina, J. *et al.* Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial. *Mol. Syst. Biol.* **17**, (2021).

224. Menon, G. & Krishnan, J. Spatial localisation meets biomolecular networks. *Nat. Commun.* **12**, 1–21 (2021).

225. Rehfeld, A., Nylander, M. & Karnov, K. Compendium of Histology. *Compend. Histol.* (2017) doi:10.1007/978-3-319-41873-5.

226. Weinberg, R. A. *The Biology of Cancer*. (W.W. Norton & Company, 2013). doi:10.1201/9780429258794.

227. Kouchmeshky, A. & McCaffery, P. Use of fixatives for immunohistochemistry and their application for detection of retinoic acid synthesizing enzymes in the central nervous system. *Methods Enzymol.* **637**, 119–150 (2020).

228. Kocsmár, É. *et al.* Proteome alterations in human autopsy tissues in relation to time after death. *Cell. Mol. Life Sci.* **80**, (2023).

229. Scicchitano, M. S., Dalmas, D. A., Boyce, R. W., Thomas, H. C. & Frazier, K. S. Protein Extraction of Formalin-fixed, Paraffin-embedded Tissue Enables Robust Proteomic Profiles by Mass Spectrometry. *J. Histochem. Cytochem.* **57**, 849 (2009).

230. Coscia, F. *et al.* A streamlined mass spectrometry–based proteomics workflow for large-scale FFPE tissue analysis. *J. Pathol.* **251**, 100–112 (2020).

231. Craven, R. A. *et al.* Proteomic analysis of formalin-fixed paraffin-embedded renal tissue samples by label-free MS: assessment of overall technical variability and the impact of block age. *Proteomics. Clin. Appl.* **7**, 273–282 (2013).

232. Tayri-Wilk, T. *et al.* Mass spectrometry reveals the chemistry of formaldehyde cross-linking in structured proteins. *Nat. Commun.* **11**, (2020).

233. Zhu, Y. *et al.* High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification. *Mol. Oncol.* **13**, 2305–2328 (2019).

234. Bressan, D., Battistoni, G. & Hannon, G. J. The dawn of spatial omics. *Science* **381**, (2023).

235. Hickey, J. W. *et al.* Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging. *Nat. Methods* **19**, 284–295 (2021).

236. Lewis, S. M. *et al.* Spatial omics and multiplexed imaging to explore cancer biology. *Nat. Methods* **18**, 997–1012 (2021).

237. Zunder, E. R. *et al.* Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. *Nat. Protoc.* **10**, 316–333 (2015).

238. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, (2015).

239. Lundberg, E. & Borner, G. H. H. Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302 (2019).

240. The Human Protein Atlas. https://www.proteinatlas.org/.

241. Cornett, D. S., Reyzer, M. L., Chaurand, P. & Caprioli, R. M. MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat. Methods* **4**, 828–833 (2007).

242. Aichler, M. & Walch, A. MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice. *Lab. Investig.* **95**, 422–431 (2015).

243. Piehowski, P. D. *et al.* Automated mass spectrometry imaging of over 2000 proteins from tissue sections at 100-µm spatial resolution. *Nat. Commun.* **11**, (2020).

244. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat. Methods* **18**, 9–14 (2021).

245. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, (2015).

246. Merritt, C. R. *et al.* Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat. Biotechnol.* **38**, 586–599 (2020).

247. Mund, A. *et al.* Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* **40**, 1231–1240 (2022).

248. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods 2022 192* **19**, 171–178 (2022).

249. Coscia, F. *et al.* Multi-level Proteomics Identifies CT45 as a Chemosensitivity Mediator and Immunotherapy Target in Ovarian Cancer. *Cell* **175**, 159-170.e16 (2018).

250. Eckert, M. A. *et al.* Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts. *Nature* **569**, 723–728 (2019).

251. Dyring-Andersen, B. *et al.* Spatially and cell-type resolved quantitative proteomic atlas of healthy human skin. *Nat. Commun.* **11**, 1–14 (2020).

252. Doll, S. *et al.* Region and cell-type resolved quantitative proteomic map of the human heart. *Nat. Commun.* **8**, 1–13 (2017).

253. Wang, D. *et al.* A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, (2019).

254. Jiang, L. *et al.* A Quantitative Proteome Map of the Human Body. *Cell* **183**, 269-283.e19 (2020).
255. Niu, L. *et al.* Dynamic human liver proteome atlas reveals functional insights into disease pathways. *Mol. Syst. Biol.* **18**, (2022).
256. Cai, X. *et al.* High-throughput proteomic sample preparation using pressure cycling technology. *Nat. Protoc.* **17**, 2307–2325 (2022).
257. Wiśniewski, J. R., Duś, K. & Mann, M. Proteomic workflow for analysis of archival formalin-fixed and paraffin-embedded clinical samples to a depth of 10 000 proteins. *Proteomics. Clin. Appl.* **7**, 225–233 (2013).
258. Kennedy, J. J. *et al.* Optimized protocol for quantitative multiple reaction monitoring-based proteomic analysis of formalin-fixed, paraffin embedded tissues. *J. Proteome Res.* **15**, 2717 (2016).
259. Virreira Winter, S. *et al.* Urinary proteome profiling for stratifying patients with familial Parkinson's disease. *EMBO Mol. Med.* **13**, (2021).
260. Lichtman, J. S., Marcobal, A., Sonnenburg, J. L. & Elias, J. E. Host-centric proteomics of stool: a novel strategy focused on intestinal responses to the gut microbiota. *Mol. Cell. Proteomics* **12**, 3310–3318 (2013).
261. Geyer, P. E. *et al.* Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* **2**, 185–195 (2016).
262. Bader, J. M. *et al.* Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease. *Mol. Syst. Biol.* **16**, (2020).
263. Bader, J. M., Albrecht, V. & Mann, M. MS-Based Proteomics of Body Fluids: The End of the Beginning. *Mol. Cell. Proteomics* **22**, 100577 (2023).
264. Hortin, G. L. & Sviridov, D. The dynamic range problem in the analysis of the plasma proteome. *J. Proteomics* **73**, 629–636 (2010).
265. Doll, S., Gnad, F. & Mann, M. The Case for Proteomics and Phospho-Proteomics in Personalized Cancer Medicine. *Proteomics. Clin. Appl.* **13**, (2019).
266. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl.* **44**, 7342–7372 (2005).
267. Pawson, T. & Scott, J. D. Protein phosphorylation in signaling – 50 years and counting. *Trends Biochem. Sci.* **30**, 286–290 (2005).
268. Bekker-Jensen, D. B. *et al.* Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat. Commun.* **11**, 1–12 (2020).
269. Ostasiewicz, P., Zielinska, D. F., Mann, M. & Wisniewski, J. R. Proteome, phosphoproteome, and N-glycoproteome are quantitatively preserved in formalin-fixed paraffin-embedded tissue and analyzable by high-resolution mass spectrometry. *J. Proteome Res.* **9**, 3688–3700 (2010).
270. Wakabayashi, M. *et al.* Phosphoproteome analysis of formalin-fixed and paraffin-embedded tissue sections mounted on microscope slides. *J. Proteome Res.* **13**, 915–924 (2014).
271. Friedrich, C. *et al.* Comprehensive micro-scaled proteome and phosphoproteome characterization of archived retrospective cancer repositories. *Nat. Commun.* **12**, (2021).
272. Jassim, A., Rahrmann, E. P., Simons, B. D. & Gilbertson, R. J. Cancers make their own luck: theories of cancer origins. *Nat. Rev. Cancer* **23**, 710–724 (2023).
273. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
274. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
275. Mani, D. R. *et al.* Cancer proteogenomics: current impact and future prospects. *Nat. Rev. Cancer* **22**, 298–313 (2022).
276. Rodriguez, H., Zenklusen, J. C., Staudt, L. M., Doroshow, J. H. & Lowy, D. R. The next horizon in precision oncology: Proteogenomics to inform cancer diagnosis and treatment. *Cell* **184**, 1661–1670 (2021).
277. Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**, 541–550 (2018).
278. Cho, N. H. *et al.* IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res. Clin. Pract.* **138**, 271–281 (2018).
279. Association, A. D. 2. Classification and diagnosis of diabetes: Standards of medical care in diabetesd2019. *Diabetes Care* **42**, S13–S28 (2019).
280. Bensellam, M., Jonas, J. C. & Laybutt, D. R. Mechanisms of β;-cell dedifferentiation in diabetes: Recent findings and future research directions. *J. Endocrinol.* **236**, R109–R143 (2018).
281. Ahlqvist, E. *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol.* **6**, 361–369 (2018).
282. Mahaddalkar, P. U. *et al.* Generation of pancreatic β cells from CD177+ anterior definitive

endoderm. *Nat. Biotechnol.* **38**, 1061–1072 (2020).

283. Shapiro, I. E. & Bassani-Sternberg, M. The impact of immunopeptidomics: From basic research to clinical implementation. *Semin. Immunol.* **66**, 101727 (2023).
284. Wahle, M. *et al.* IMBAS-MS Discovers Organ-Specific HLA Peptide Patterns in Plasma. *Mol. Cell. Proteomics* **23**, 100689 (2024).
285. Phulphagar, K. M. *et al.* Sensitive, High-Throughput HLA-I and HLA-II Immunopeptidomics Using Parallel Accumulation-Serial Fragmentation Mass Spectrometry. *Mol. Cell. Proteomics* **22**, (2023).
286. Gupta, K. K., Attri, J. P., Singh, A., Kaur, H. & Kaur, G. Basic concepts for sample size calculation: Critical step for any clinical trials! *Saudi J. Anaesth.* **10**, 328 (2016).
287. Ianevski, A., Giri, A. K. & Aittokallio, T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nat. Commun. 2022 131* **13**, 1–10 (2022).
288. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, 8746 (2019).
289. Gorin, G., Svensson, V. & Pachter, L. Protein velocity and acceleration from single-cell multiomics experiments. *Genome Biol.* **21**, 1–6 (2020).
290. Oliinyk, D., Will, A., Schneidmadel, F. R., Humphrey, S. J. & Meier, F. µPhos: a scalable and sensitive platform for functional phosphoproteomics. *bioRxiv* 2023.04.04.535617 (2023) doi:10.1101/2023.04.04.535617.
291. Bortel, P. *et al.* Systematic optimization of automated phosphopeptide enrichment for high-sensitivity phosphoproteomics. *bioRxiv* 2023.11.23.568418 (2023) doi:10.1101/2023.11.23.568418.
292. Karayel, O. *et al.* Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease. *Cell Reports Med.* **3**, 100661 (2022).
293. Bhatia, H. S. *et al.* Spatial proteomics in three-dimensional intact specimens. *Cell* **185**, 5040-5058.e19 (2022).
294. Kiemen, A. L. *et al.* CODA: quantitative 3D reconstruction of large tissues at cellular resolution. *Nat. Methods* **19**, 1490–1499 (2022).
295. Rappez, L. *et al.* SpaceM reveals metabolic states of single cells. *Nat. Methods* **18**, 799–805 (2021).
296. Tsugawa, H. *et al.* A lipidome atlas in MS-DIAL 4. *Nat. Biotechnol.* **38**, 1159–1163 (2020).
297. Deng, Y. *et al.* Spatial profiling of chromatin accessibility in mouse and human tissues. *Nat. 2022 6097926* **609**, 375–383 (2022).
298. Zhang, D. *et al.* Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature* **616**, 113–122 (2023).

## 5.2. List of figures

Figure 1: Overview of the gene expression pathway. ..... 2
Figure 2: Schematic of bottom-up MS-based proteomics workflow. ..... 8
Figure 3: Different fragmentation pattern of the peptide backbone for MS2 spectra. ..... 15
Figure 4: Ion path of different mass spectrometers and their MS analyzer principles. ..... 17
Figure 5: Ion mobility principles of TIMS and FAIMS. ..... 20
Figure 6: Different data acquisition strategies as an overview. ..... 21
Figure 7: Schematics of different quantification strategies in proteomic data. ..... 25
Figure 8: Overview of the Deep Visual Proteomics (DVP) technology. ..... 29

# 6. Supplements

## Article 13. Large-scale proteomics across neurological conditions uncovers biomarker panels and targets in multiple sclerosis

Submitted to *Cell*, 2023

Jakob Maximilian Bader[1,$], Christine Makarov[2,$], Sabrina Richter[3], Maximilian Thomas Strauss[4], Friederike Held[2], Patricia Skowronek[1], Marvin Thielert[1], Achim Berthele[2], Wen-Feng Zeng[1], Constantin Ammar[1], Isabell Bludau[1], Benjamin Schubert[3], Fabian Theis[3], Christiane Gasperi[2,§], Bernhard Hemmer[2,5,§], Matthias Mann[1,4,6,§]

*[1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [2]Clinical Neuroimmunology Unit Klinikum rechts der Isar University Hospital, Technical University of Munich, Munich, Germany. [3]Institute of Computational Biology, Helmholtz Center Munich, Munich, Germany. [4]NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. [5]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.*

*[$]these authors contributed equally.*

*[§]these authors contributed equally.*

Classical clinical samples, such as body fluids, are frequently used for biomarker discovery. Cerebrospinal fluid (CSF) in particular is known for its pivotal role in clinical diagnostics of neuropathological diseases. To facilitate a large-scale proteomics analysis of 5,000 CSF samples, we generated a high-throughput workflow capable of quantifying around 1,500 proteins per sample. We conducted analysis across a wide range of neurological disorders with special emphasis on multiple sclerosis. This study outlined pathological processes and prioritized prospective drug targets for multiple sclerosis. Interestingly, we discovered an unexpected link of multiple sclerosis to atypical memory B cells and neutrophils, extending our understanding beyond the current focus on T and typical B cells.

***Contribution:***

Co-authorship. Christiane Gasperi, Bernhard Hemmer and Matthias Mann conceptualized the study. Jakob Bader conducted the project as first author. Together with Jakob Bader, I established the proteomic workflow and performed the MS measurements.

## Article 14. AlphaViz: Visualization and validation of critical proteomics data directly at the raw data level.

Preprint, *bioRxiv*, 2022 (Jul, 13); DOI: 10.1101/2022.07.12.499676

Eugenia Voytik[1], Patricia Skowronek[1], Wen-Feng Zeng[1], Maria C. Tanzer[1], Andreas-David Brunner[1], Marvin Thielert[1], Maximilian T. Strauss[2], Sander Willems[1], Matthias Mann[1,2]

*[1]Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany. [2]NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark.*

MS-based proteomics identifies and quantifies thousands of proteins and modified peptides. However, only a small subset is usually considered for in-depth downstream analysis. The biological conclusion is sometimes even drawn from several few proteins. With the integration of automated processing workflows, the validation of biologically or clinically significant results by accessing the foundational raw data rarely occurs within a study. Existing tools are often not integrated into the overall analysis and are not easily adaptable to incorporate new methodologies. To address these issues, we developed AlphaViz, an open-source Python package that overlays the output from common search algorithm workflows on raw data, thereby enabling easy visualization and validation of protein and peptide identifications. Leveraging recent advancements in deep learning-assisted prediction of experimental peptide properties, AlphaViz facilitates manual comparison of the expected versus observed peptide result.

In summary, AlphaViz allows fast validation or invalidation of peptide identifications in MS raw data, irrespective of the score assigned by search algorithms. We encourage that validation of critical proteomics results will become a standardized feature in MS-based proteomics with tools like AlphaViz. Sc-proteomics would also benefit from this evaluation.

### *Contribution:*

Co-authorship. Eugenia Voytik and Matthias Mann conceptualized the study. I contributed ideas for the concept and visualization in AlphaViz. Additionally, I provided feedback to the graphical user interface and functionalities of AlphaViz. Together with the other co-authors, I revised the manuscript.

# 7. Acknowledgement

I am immensely grateful to all people who contributed in any way to my PhD studies, making my journey truly remarkable and invaluable.

**Prof. Matthias Mann,** first of all, I want to express my gratitude to my PhD supervisor for the tremendous support. It is an honor to experience your mentorship and dedication to science. Thank you for fostering such a vibrant and innovative environment and pushing the best out of us resulting in a steep learning curve.

**Thesis advisory committee**, thank you for the support, time investment and engagement into my PhD studies, all guidance and valuable feedback along my path shaped parts of this thesis.

**timsTOFfice**, Maria, Cathrine, Patricia, Erika and Andreas, I am very thankful for all the discussions and fun time we had in the office and beyond, especially with all the plants and the long-lasting Ouzo bottle. Thank you for welcoming me in the beginning and making it hard to ever leave.

**Mann lab members,** I am thankful to Mario, Igor and Tim for technical support, the whole Admin team including Medini, Ute and Alison; Florian, Georg, Thierry, Constantin, Isabell, Feng, Caro, Marc and Andreas for all the fruitful discussion, projects and fun in the laboratory. Unfortunately, the lines and words are too short to express all my gratitude to all members. Thank you to everyone in the department for being a unique and collaborative team and creating such an inspiring atmosphere.

**Copenhagen crew,** thank you to Andreas, Fabian, Max, Sonja and the whole team for the great time during my visits. The Christmas party was one highlight which I will not forget.

**Colleagues from Bruker and EvoSep team,** thank you for all your support and great collaborative effort over the full PhD study.

**Collaboration partner around the world,** I have been fortunate to work in a collaborative network and experience great science. Thank you for sharing your expertise and guidance throughout my PhD study and in the scientific projects.

**TTT team and BC et al.**, Lisa, Max and Sophia, some BC et al. papers have become reality, however the TTT team goes on as well as our deep friendship. Thank you for all the fun times as friends and colleagues.

**All friends and Volleyball Stammtisch**, thank you for making my time outside of the lab really enjoyable. If I needed support, I was always welcomed.

**Family**, especially my parents Lissi and Gerhard and brother Philipp, thank you for all your constant encouragement, help and support in my life and throughout my education. I would not have been able to accomplish this path without your support and advice. I am very blessed to have you as parents and brother.

**Toni**, last but not least, my love. Thank you for the never-ending support and joy you bring to my life.