

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES DER  
FAKULTÄT FÜR CHEMIE UND PHARMAZIE DER LUDWIG-  
MAXIMILIANS-UNIVERSITÄT MÜNCHEN



# Ovarian Cancer Proteomics

Fabian Coscia

aus  
Ebersberg, Deutschland

2016



## 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, 25.08.2016

—  
Fabian Coscia  
—

Dissertation eingereicht am 26.08.2016

1. Gutachter: Prof. Dr. Matthias Mann
2. Gutachter: Prof. Dr. Ernst Lengyel

Mündliche Prüfung am 30.09.2016

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Ovarian cancer	1
1.1.1	Pathogenesis of high-grade serous ovarian cancer	1
1.1.2	The proteogenomic landscape of HGSOc	5
1.1.3	Prioritized HGSOc research goals	7
1.2	The Proteome	11
1.3	Mass spectrometry-based quantitative proteomics	11
1.4	Clinical cancer proteomics	20
1.4.1	Cancer proteomics: A brief history	20
1.4.2	Cancer tissue proteomics	22
1.4.3	Biomarker discovery	25
1.4.4	Cancer antigen discovery	30
1.5	Aims of the thesis	<b>35</b>
<b>2</b>	<b>Results</b>	<b>36</b>
2.1	Manuscript 1: Deep proteomic profiling of ovarian cancer models	36
2.1.1	Project aim and summary	36
2.1.2	Contribution	38
2.1.3	Publication	39
2.2	Manuscript 2 (in preparation): Proteomic characterization of chemotherapy sensitive ovarian cancer	53
2.2.1	Project aim and summary	53
2.2.2	Contribution	56
2.2.3	Publication	57
2.3	Manuscript 3 (in preparation): Compartment-resolved proteomics of ovarian cancer progression	87
2.3.1	Project aim and summary	87
2.3.2	Contribution	89
2.3.3	Publication	90
<b>3</b>	<b>Discussion and Outlook</b>	<b>121</b>
3.1	The modern proteomic toolbox for the study of ovarian cancer	121
3.2	The age of large-scale proteomics: implications for biomedicine	131
	<b>Abbreviations</b>	<b>133</b>
	<b>References</b>	<b>134</b>
	<b>Acknowledgments</b>	<b>148</b>
	<b>Curriculum Vitae</b>	<b>149</b>

# Summary

---

High-grade serous ovarian cancer (HGSOC) is the most commonly diagnosed (70%) and aggressive epithelial ovarian cancer (OvCa) subtype; only 30% of the patients diagnosed with advanced stage disease live beyond 5 years after initial diagnosis, making it the most lethal gynecologic cancer. In contrast to many other cancers, the prognosis for HGSOC is unchanged in recent decades, despite extensive translational and clinical research. To a large degree, this is linked to the fact that HGSOC pathogenesis has remained an enigma such that there is currently no clear consensus even about the exact cell-of-origin. Nonetheless, recent 'omics' technologies have started to shed new light on the molecular landscape of HGSOC, which may not only help to better understand its biology, but also to uncover new therapeutic strategies and stratify patients for treatment. In particular, the field of quantitative proteomics has reached a state that now enables system-wide analysis of protein expression in any biological system. Complemented by the study of protein-protein interactions, post-translational modifications or protein dynamics, modern proteomic technologies are extremely versatile for the study of proteins in various settings, such as in the context of HGSOC biology.

All three projects described in this thesis were conducted in a very productive interdisciplinary collaboration with the Ernst Lengyel ovarian cancer laboratory at the University of Chicago.

In the *first* project, I aimed to characterize 30 frequently used ovarian cancer cell lines for their suitability to represent HGSOC at the proteomic level. It was shown recently that the most often used HGSOC cell lines do not resemble human HGSOC tissue at the global genomic level (12). Conversely, some rarely used cell lines were found to be more suitable to model HGSOC, at least based on their genomic make-up. However, it was still unknown to what extent these reported genomic features were reflected at the protein level. We hypothesized that a streamlined and in-depth proteomic workflow may refine the previously suggested cell line classification and define the proteomic HGSOC cell line landscape. By quantitatively analyzing > 10,000

proteins, we found for the first time that the 'good' HGSOC cell lines grouped into two distinct proteome classes. Integration of primary cells and tumor tissues further revealed that the grouping of the proteomes likely reflects a cell-of-origin based classification, an interesting observation in light of the previously hypothesized dualistic precursor model for HGSOC (13). This large inventory of proteins across cell lines additionally provides a useful resource for the OvCa community and can be accessed via the MaxQB database at <http://maxqb.biochem.mpg.de/mxldb/project/list>.

In the **second** and principal project, I conducted a discovery proteomics based study with the goal to improve understanding of chemotherapy sensitivity in advanced stage HGSOC, the largest histologic subgroup of epithelial ovarian cancers (70%). Up to 80-90% of patients with metastatic HGSOC develop a chemoresistant state despite a favorable initial response to conventional carboplatin/taxol chemotherapy (14), leading to the above-mentioned low 5-year survival rate of only 30%. However, a small number of patients with metastatic disease (17%) respond remarkably well to chemotherapy and remain free of relapse for more than a decade after primary treatment (15). The molecular details for these exceptional responses are still unknown despite the extensive genomic characterization of HGSOC (16, 17). In a first of a kind clinical proteomics study, we analyzed 25 HGSOC patient biobank samples from responders and non-responders to a depth of in total ~ 9,000 proteins. Integrating the clinical survival data into the analysis revealed that chemotherapy outcome is generally not apparent from the global proteome pattern. Instead, I found the expression of a single protein, cancer-testis antigen 45 (CT45), to be an independent prognostic indicator for long-term survival. We validated this finding in a larger patient cohort and subsequently showed an active function of CT45 in mediating chemotherapy sensitivity based on *in vitro* and *in vivo* models. Extending these findings, my work revealed mechanistic insights of the biology of CT45 mediated chemotherapy sensitivity. It interferes with protein phosphatase 4 (PP4) dependent signaling on chromatin during the DNA damage response (DDR). Furthermore, we found evidence of an additional axis of anti-cancer immune response directed against CT45, involving T-cell mediated cytotoxicity and likely potentiating the effect on long-term survival.

In the *third* project of this thesis, I established a highly sensitive ‘micro-proteomic’ workflow, which enabled the analysis of as few as 5,000 laser-microdissected tumor cells from FFPE biobank material to an unprecedented depth of up to 5,000 proteins from single-run (4h) analysis. This made it possible to address important clinical questions directly related to HGSOC biology: How does the proteome change during HGSOC progression from early pre-neoplastic lesion to late metastasis? What is the contribution of the tumor microenvironment (TME) to ovarian cancer progression? For the first time, we analyzed the proteomic progression of HGSOC in eleven patients diagnosed with advanced stage disease. Interestingly, whereas our results did not reveal commonly observed metastatic changes in the tumor proteome, we identified a conserved stromal protein signature in the tumor microenvironment, which was linked to metastasis to the omentum. Follow-up experiments by my collaboration partners at the University of Chicago (Lengyel group) highlighted the importance of the identified stromal proteins in mediating the aggressive nature of metastatic HGSOC. These observations may open up new avenues for HGSOC treatment in future.

*To my family.*

# 1 Introduction

## 1.1 Ovarian cancer

Ovarian cancer (OvCa) is an aggressive disease and the most lethal gynecologic malignancy: Only 46 % of patients survive five or more years, which accounts for 14,180 deaths in the US annually (18) and 29,770 deaths in the European Union (19-21). The high mortality is in part due to presentation at late stage (only 15% are diagnosed with loco-regional stage I/II disease). Furthermore, the response to the standard platinum-based chemotherapy has a limited durability, with 20% of patients chemo refractory and 30% chemo resistant, meaning they have recurrence in less than six months. Despite extensive biomedical research efforts in the past decades, uncertainties about the exact origin and pathogenesis of this disease remain, making OvCa one of the least understood cancers.

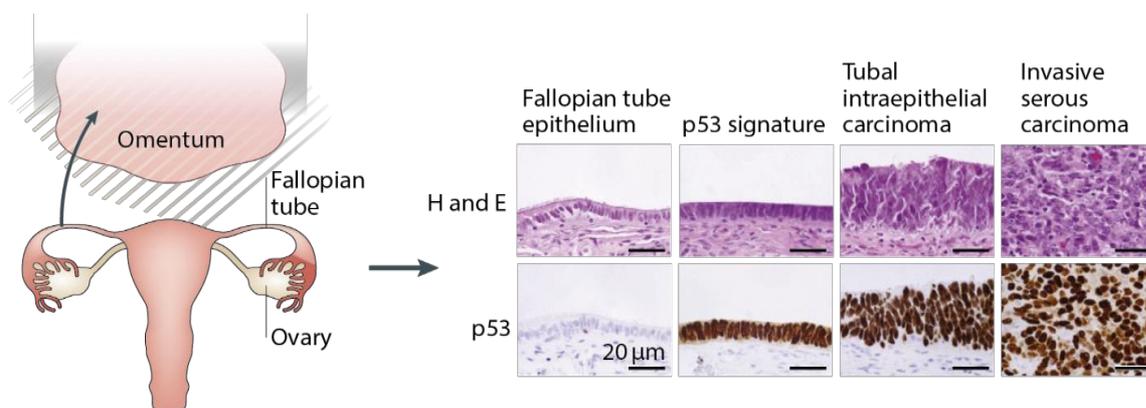
Major histological subtypes for epithelial OvCa, which are associated with different prognosis and pathogenesis, include serous, endometrioid, clear-cell and mucinous cancers. For this thesis, the term 'ovarian cancer' is used as a synonym for ovarian, fallopian tube, and peritoneal cancer.

### 1.1.1 Pathogenesis of high-grade serous ovarian cancer

High-grade serous ovarian cancer (HGSOC) represents the most commonly diagnosed (70%) and aggressive subtype of epithelial OvCa with a 5-year survival rate of only 30% for metastatic, advanced stage disease (22). According to the classification of the International Federation of Gynecology and Obstetrics (FIGO), 75% of patients are initially diagnosed with stage III or IV disease, highlighting the importance of research dedicated to early ovarian cancer detection and prevention (18).

Traditionally, HGSOC was believed to originate exclusively in the ovaries as the ovaries are almost invariably transformed. There is experimental and pathological evidence for the ovarian surface epithelium (OSE) as a possible site of origin as OSE-derived inclusion cysts can undergo tubal metaplasia and full transformation (23-25). Moreover, OSE is a multipotent epithelium that may give rise to HGSOC (23).

However, the recent paradigm-shifting discovery of a potential precursor lesion, serous tubal intraepithelial carcinoma (STIC), in the tubal fimbria of BRCA mutation carriers and HGSOC patients has questioned the traditional view of HGSOC origin from the OSE (14, 26-33). These studies revealed that the pre-neoplastic STIC lesions were associated with secretory cells, *TP53* mutations and up-regulation of DNA damage markers (e.g.  $\gamma$ H2AX) (34-36). Interestingly, somatic *TP53* mutations in STICs and corresponding HGSOCs were found to be identical in 27 of 29 patients with ovarian cancer suggesting a clonal relationship between STICs and HGSOCs (37). Today, HGSOC is thought to 1) originate primarily in the secretory cells of the distal fallopian tube, 2) evolve to STIC, 3) fully transform to invasive fallopian tube cancer, and 4) regularly metastasize to ovarian and/or extra-ovarian tissues such as the omentum (Fig. 1). Consequently, HGSOC of the ovary may represent a first metastatic spread from the fallopian tube to the ovary. However, some HGSOCs develop without any detectable fallopian tube involvement, suggesting a different site of origin for these cancers, such as the OSE or another cell type in the ovary. It is noteworthy that both hypotheses are supported by animal models (38-41). While neither hypothesis on the cell of origin of OvCa has unequivocally been proven correct, it is generally accepted that the high prevalence of STICs (27) (70% for ovarian or peritoneal HGSOC), indicates that the majority of HGSOC may indeed originate in the fallopian tube.



**Figure 1: Model of high-grade serous ovarian cancer progression in the fallopian tube.** Left: High-grade serous ovarian cancer (HGSOC) is thought to originate in the secretory cells of the fallopian tube (FTEC) and often metastasizes to omental tissue. Right: Different stages of HGSOC progression in the fallopian tube. Immunohistochemical staining of p53 mutant signatures reveals HGSOC initiation in STIC lesions. Tumor cell expansion leads to the development of invasive carcinoma. Adapted and modified from (1).

Due to ineffective screening methods, the majority of HGSOC cases (75%) is diagnosed with advanced stage disease when the tumor has already spread and metastasized to the peritoneal cavity. Interestingly, HGSOC disseminates primarily to the adipocyte-rich omentum, which the tumor can utilize as energy source for rapid growth in the peritoneal cavity (42). Following detachment from the primary tumor site such as the fallopian tube or ovary, which usually involves an epithelial-to-mesenchymal transition (43), E-cadherin loss (44) and a more invasive phenotype (45), the tumor cells are thought to reach the mesothelium-lined omentum and peritoneum via passive transport through the peritoneal fluid (46). Tumor cell binding to the mesothelium, a single cell layer of mesothelial cells attached to a basement membrane, is orchestrated through binding via  $\beta_1$ -integrins (47) as well as CD44 (48). Once bound, ovarian carcinoma cells can induce a cascade of proteolytic (49, 50) and pro-apoptotic (51) activities to invade deeper into the sub-mesothelial layers. Only little is known about what happens after successful ovarian carcinoma implantation. However, it has been speculated that the tumor induces proangiogenic signaling to build *de-novo* vasculature surrounding the site of metastasis (46).

The tumor microenvironment (TME) plays a key role during all phases of tumorigenesis (52), influences chemotherapy response (53) and patient outcome (54, 55). Several cell types of the microenvironment interact with tumor cells and modulate the aggressiveness of their behavior. For cancers such as HGSOC, the TME cell types include cancer-associated fibroblasts (CAFs), tumor-associated macrophages (TAMs), endothelial cells, adipocytes, mesothelial cells as well as various immune cell types. Of note, HGSOC was one of the first reported examples for a positive correlation between prolonged patient survival and the number of intraepithelial tumor-infiltrating lymphocytes (TILs) (56).

In light of the recent success in applying immune checkpoint inhibition strategies for cancer immunotherapy treatment of melanoma or non-small cell lung cancer (57), it is tempting to speculate that HGSOC might also be vulnerable to these combination treatments. However, a recent clinical phase II study of an immune checkpoint inhibitor against PD-L1, although not associated with safety issues, had minimal overall efficiency (58). Future studies are needed to dissect why only a small subset of patients showed complete and durable responses. Interestingly, in another study, patients with good neoadjuvant chemotherapy response had a lower number of intratumoral FOXP3<sup>+</sup> regulatory T-cells (Tregs) (59), suggesting a functional link between these

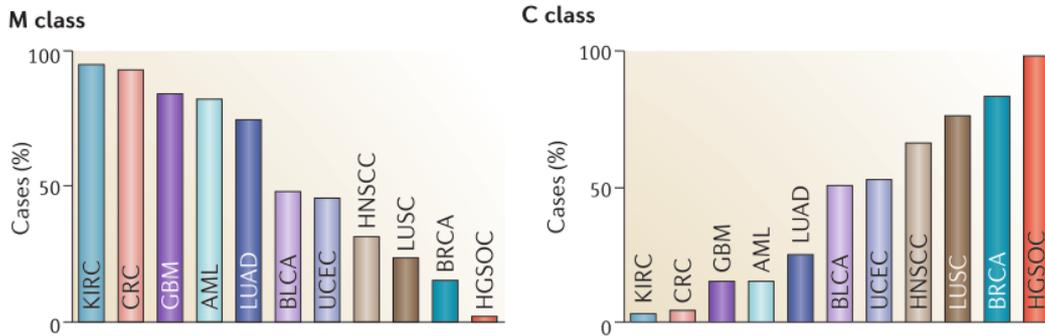
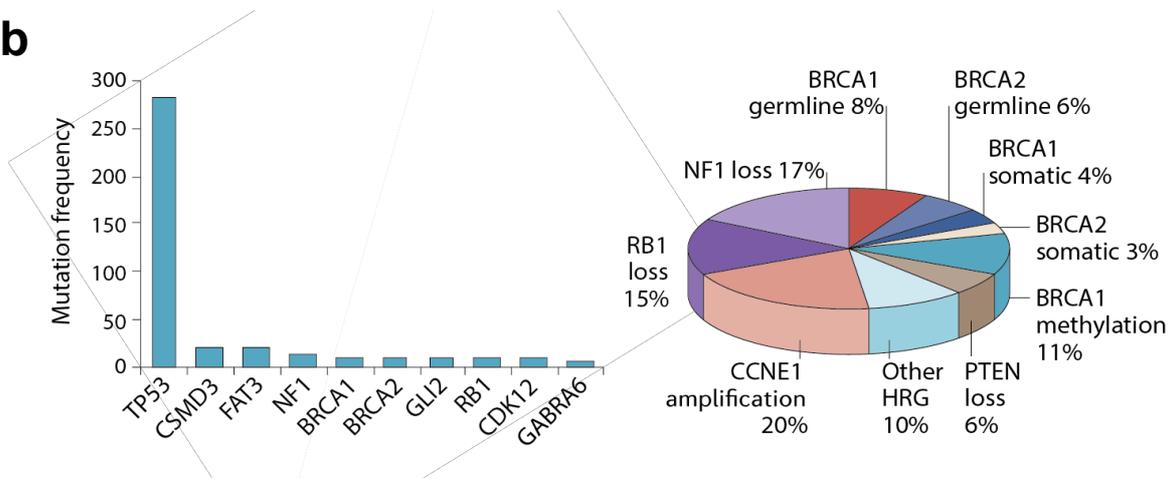
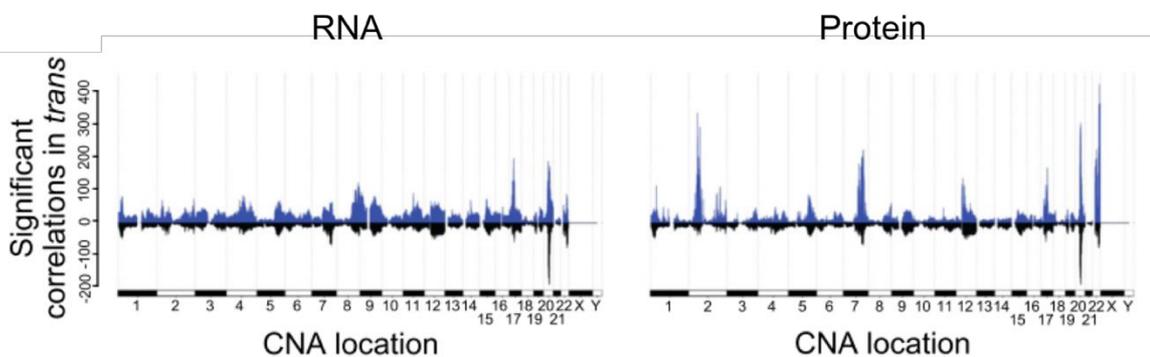
cells and chemotherapy. Moreover, neoadjuvant chemotherapy increased the T-cell specific expression of PD-L1 indicating that immune checkpoint inhibition in combination with chemotherapy may offer a promising new therapeutic approach for HGSOC treatment.

Tumor antigens such as ERBB2, TP53, CTAG1B, MSLN, NY-ESO-1, MAGE melanoma antigen family members or TERT (60) have already shown promise to elicit an immune response against HGSOC. Clearly, treating HGSOC with immunotherapy approaches is very promising and this thesis directly contributes to this concept (Results section 2.2).

### 1.1.2 The proteogenomic landscape of HGSOC

Genetically, HGSOC is characterized by ubiquitous somatic *TP53* mutations and genetic instability (16). In addition to *TP53* mutational inactivation, there are only very few known additional recurrent mutations, including *BRCA1*, *BRCA2*, *RB1*, *NF1*, *FAT3*, *CSMD3*, *GABRA6* and *CDK12*, as well as *BRIP1* and *PALB2* (61). Instead, there are widespread DNA copy number alterations (CNA), which together with the low number of recurrent mutations make HGSOC a 'C class' tumor (62) (Fig. 2). This suggests an early breakdown of DNA repair pathways, ultimately leading to pronounced chromosomal instability. Interestingly, functional loss of *TP53* is not sufficient to drive the full transformation from STIC to invasive fallopian tube cancer in animal models (63) indicating that additional, yet unidentified molecular events are crucial for this step in progression. (61). The most commonly observed chromosomal alterations lead to *CCNE1* amplification (20%), *NF1* loss (17%) and *RB1* loss (15%). Based on genomic pathway analyses, it was suggested that roughly half of all HGSOC tumors are defective in homologous recombination (64).

Owing to its high degree of structural chromosomal alterations and low mutational profile, HGSOC should represent an ideal cancer type for proteomic or proteogenomic approaches, since all proteins in the amplicons are investigated together (65). A very recent report from the CPTAC consortium investigated 169 HGSOC tissue samples at the proteomic level and related it to genomic information such as copy number alterations (5). A low overall correlation between mRNA and protein levels (Spearman  $r = 0.45$ ) was reported and several copy number alterations with *in trans* effects on protein abundance (independent of the genomic locus) showed no changes of mRNA levels. This contradicts reports in other cancers such as colorectal (66) or breast cancer (67) and, assuming correct proteomic quantification, highlights the unique molecular landscape of HGSOC. The authors identified several genomic copy number alterations that directly translated to changes in protein abundance and used these findings to stratify patients into different survival groups. These proteogenomic alterations included fractions of chromosomes 2, 7, 20 and 22, which caused changes in abundance of more than 200 proteins (Fig. 2c). Further analysis revealed that many of these proteins were involved in cell movement/adhesion and immune system related functions, indicating important biological roles of these processes during HGSOC progression.

**a****b****c**

**Figure 2: HGSOc is characterized by genomic copy number alterations rather than mutations. a.** HGSOc is classified as a ‘C class’ tumor and features the highest number of recurrent copy number alterations across human cancers. ‘M class’ tumors (left) are dominated by mutations. For HGSOc, only few recurrent mutations have been identified (1). **b.** Overview of recurrent mutations and genomic alterations in HGSOc (1). **c.** Copy number alterations (CNA) associated with RNA abundance and protein abundance along the genome. Adapted from (5).

### 1.1.3 Prioritized HGSOC research goals

#### 1.1.3.1 *The need for accurate cellular models*

The identification of accurate HGSOC models for preclinical investigations was declared as one of the key missions at the Ovarian Cancer Action's international research meeting (Helene Harris Memorial Trust (HHMT)) in January 2015 (14). Since most pre-clinical research relies on cellular models for *in vitro* and *in vivo* studies, this raises the important question of how closely they resemble the tissue being studied.

It has recently become apparent that a considerable amount of research in the past was based on cellular models that do not resemble HGSOC at the global genomic level (12). This was caused in part because no reliable genomic tools were available to distinguish between representative and poor cellular models, as well as to identify cross-contamination or an incorrect assignment of the underlying tissue of origin. With the development of "omics" tools that identify models which represent human disease, a variety of urgent, HGSOC related, questions, such as characterizing distinct HGSOC subtypes reflective of their cell-of-origins or understanding mechanisms of drug resistance and/or sensitivity at the molecular level, as well as tumor biology can be tackled.

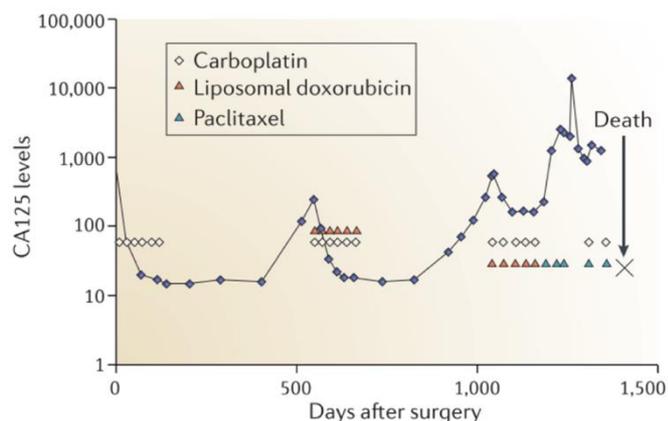
The concept of characterizing cell lines for pre-clinical research is not novel. Historically, first reports in the ovarian cancer field date back to the late 70s and 80s when individual, newly established OvCa cell lines were described according to morphology, histology, karyotype, and growth characteristics (68-74). This led to the assignments of the tissue of origin for many of the OvCa cell lines used today. However, due to a lack of standardized and validated methods for the systematic comparison of cell lines, the results of these investigations were often questionable. More advanced and large-scale analyses had to wait until the 21<sup>st</sup> century when major breakthroughs in 'omics' technologies had occurred. Since then, multiple studies have been published almost all focusing on genomic or transcriptomic-based methods (12, 75-77). Most importantly, in conjunction with cell line data obtained from the Cancer Cell Line Encyclopedia (CCLE) consortium (78), a recent study evaluated a panel of 47 ovarian cancer cell lines by means of copy number alterations, mutation frequency, and gene expression data (12). Surprisingly, the authors found that the most commonly

used cell lines did not resemble HGSOC at the genomic level. This study has also provided a rationale for the use of a set of newly identified HGSOC cell lines. Based on these findings, three recent reports addressed functional properties of the 'good' HGSOC cell lines (79-81). Unexpectedly, these were less capable to form *in vivo* tumors than the previously used ones and displayed a larger functional heterogeneity in e.g. proliferation, migration, invasion and the expression of EMT markers.

Before the subsequently described studies, it remained elusive to what extent the previously reported cell line features are represented at the protein level. The proteomic profiling of OvCa cell lines conducted in this thesis (Results section 2.1) complements and significantly expands the above-mentioned recent studies and findings.

### 1.1.3.2 Understanding chemotherapy efficacy

HGSOC treatment involves first surgery, which often requires surgical debulking, followed by adjuvant chemotherapy with a combination of carboplatin and paclitaxel. Although most HGSOC patients initially respond well to first-line chemotherapy, 80-90% of those diagnosed with advanced stage disease develop chemoresistant disease. This usually happens through several cycles of relapse which can be monitored by serum CA-125 levels (Fig. 3). Much research has concentrated on understanding the multifaceted nature of chemoresistant HGSOC by using a variety of genomic and bioinformatic tools, however, so far only little has really translated into the clinic. Therefore, despite the tremendous amount of work by oncologists, clinicians and life-scientists all around the world, HGSOC mortality has only changed minimally in the last decades. Conversely, the molecular landscape of chemotherapy sensitive HGSOC has not yet been studied in great detail. A recent large-scale ovarian cancer study found that 17% of patients with advanced stage disease survived beyond 10 years after initial diagnosis (15). This suggests the presence of yet unidentified, predisposing molecular features promoting chemotherapy sensitivity and long-term survival. It is of great clinical importance to understand these 'exceptional responders', as they may shed light on the problem of very widespread chemotherapy resistance.



**Figure 3: Treatment and disease course of HGSOC.** Following initial response to carboplatin based chemotherapy, the majority of HGSOC patients develops a chemoresistant state over multiple treatment/relapse cycles as indicated by CA125 levels (in units per millilitre) over time (1).

### *1.1.3.3 Understanding disease progression*

To date, only little is known about the proteomic progression of HGSOC and the development of its complex microenvironment from early serous tubal intra-epithelial carcinoma (STIC) lesions in the fallopian tube to omental metastases. As outlined in the HHMT 2015 meeting (14), more sophisticated and integrated 'omics' approaches are urgently needed to dissect the multitude of molecular changes taking place in the tumor.

There is a current lack of streamlined and robust proteomics methods to analyze 'low-input' samples, such as those obtained from laser-capture microdissection (LCM), coupled to the unavailability of tumor tissue from multiple tumor sites. Therefore, the previous large-scale proteomic analysis of HGSOC was based on whole-tissue specimens predominantly obtained from invasive ovarian tumors (5). LCM approaches are superior to whole-tissue analysis in regard of the finely-resolved quantitative read-out in the presence of cell type heterogeneity in the HGSOC environment. This scientific challenge could be addressed using improved sample processing workflows with the goal to analyze minute amounts of tissue. Together with well annotated biobanks this will allow further mechanistic studies at the proteomic level. These will not only address the degree and nature of proteomic alterations linked to disease progression, but might also reveal novel disease-relevant driver proteins.

## 1.2 The Proteome

The central dogma of molecular biology states that genetic information is encoded in the form of DNA in the nucleus, which can be read and transcribed into transportable messenger RNA (mRNA) molecules and ultimately translated into functional proteins (82). Today this sequential flow of information can be analyzed in a very comprehensive and system-wide manner for genome and transcriptome analyses using next-generation sequencing approaches. Mass spectrometry (MS) based proteomics is the technology of choice to study the entirety and nature of all expressed proteins in the cell, the so called proteome (83). Whereas the genome encodes a defined number of genes (~20.000) (84, 85), the level of complexity increases substantially for the number of transcripts, which are dynamically expressed in different cell types and time points and which can be differentially spliced. The proteome is much more complex still, due to the existence of diverse protein modifications or protein processing (86). Furthermore, protein-protein interactions as well as spatio-temporal protein dynamics add yet another level of protein complexity. Consequently, proteomics is a multidisciplinary research field comprised of a plethora of approaches and methodologies to analyze and characterize proteins under various biological or biomedical conditions.

## 1.3 Mass spectrometry-based quantitative proteomics

Over the last decades, mass spectrometry (MS) has evolved into the key technology for the study of proteins, the functional entities of most biological processes taking place inside and outside of the cell. Mostly due to very significant technological advances in MS itself, availability of powerful solutions for sample preparation, as well as sophisticated computational workflows to analyze gigabytes of raw data files, MS-based proteomics in principle allows converting every protein-centric biological experiment into a systems biology approach. Historic milestones for the field of MS-based proteomics include the development of the soft ionization techniques matrix-assisted laser desorption / ionization (MALDI) (87, 88) and electrospray ionization (ESI) (89). In the latter technique, which shared the Chemistry Nobel Prize in 2002, kilovolt potentials between the end of the capillary column and the entrance of the

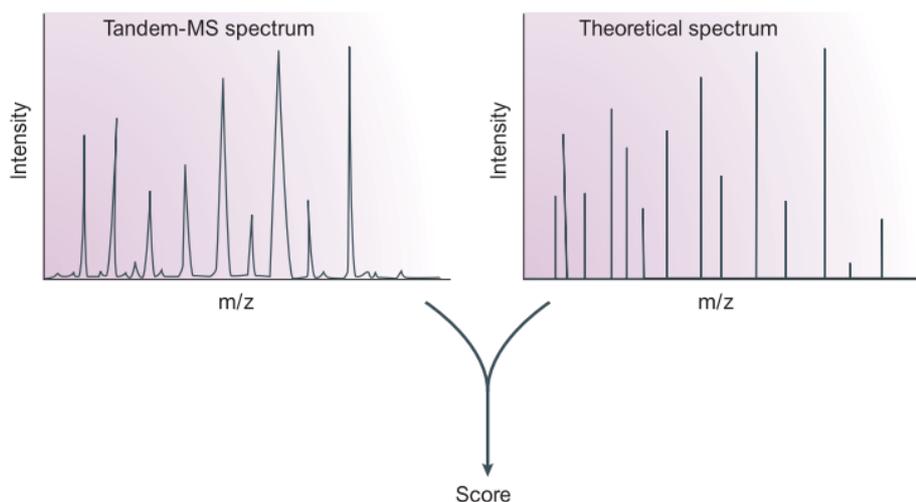
mass spectrometer generate a spray of charged droplets. Intact proteins or peptides are then directly ionized out of the liquid phase into the gas phase.

Conceptually, there are two strategies for MS-based proteomics: In top-down proteomics, intact proteins are analyzed by the MS, which, due to the complexity of the mass spectra generated from individual intact proteins, involves extensive up-front protein purification. Top-down proteomics is therefore mainly used for recombinant protein characterization such as therapeutic antibodies, protein degradation products, and proteoforms, including the exact combination of post-translational modifications (PTMs) on them (90). Small proteins are more amenable to MS-analysis than large ones, due to better protein purification and fewer charge states, and top down may have advantages in this case.

In bottom-up or 'shotgun' proteomics (Fig. 5), proteins are first digested to peptides by using sequence specific proteases such as trypsin and are then analyzed individually by the mass spectrometer. This is the method of choice for the study of proteins, protein complexes and entire cellular proteomes. An underlying reason for this is that MS of peptides is much more sensitive and easy than MS of intact proteins and lends itself to straightforward protein identification. Furthermore, many proteins such as insoluble or 'sticky' proteins are not be amenable to purification and analysis but can readily be analyzed at the peptide level by shotgun proteomics (91).

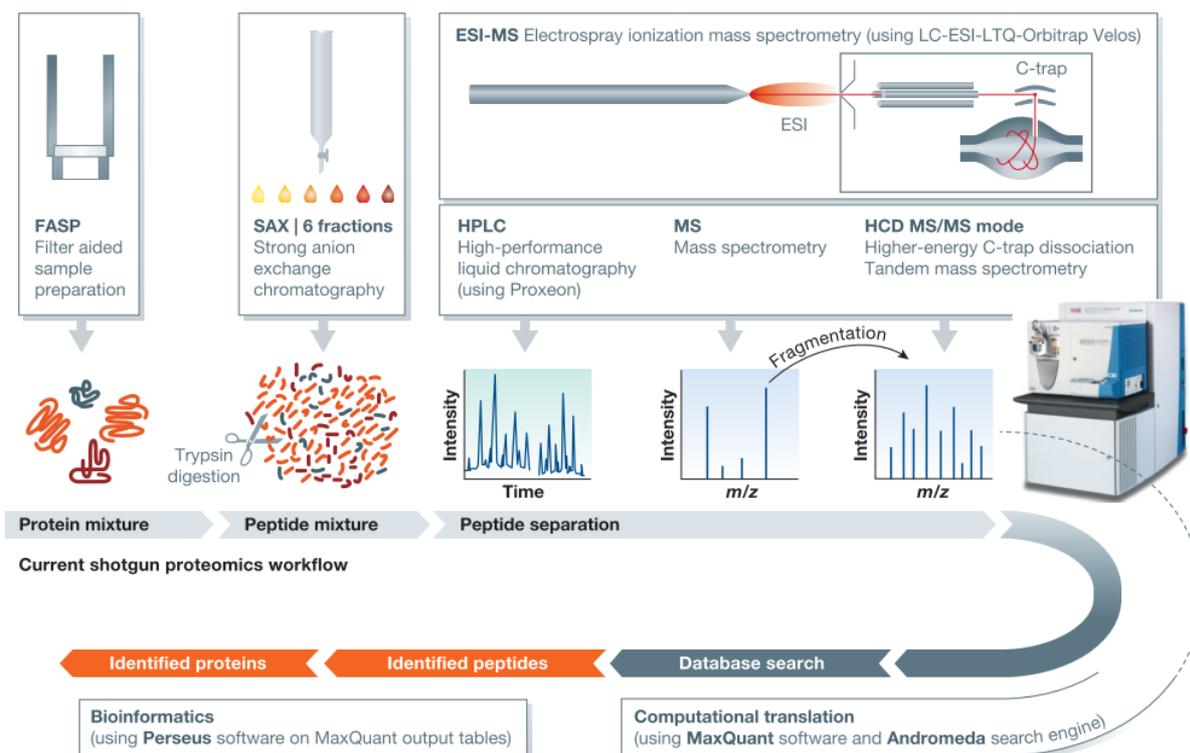
Complex peptide mixtures require sufficient upfront chromatographic separation to ensure that as many peptide species as possible can be subjected to mass spectrometric analysis. The combination of liquid chromatography (LC) coupled 'on-line' to ESI mass spectrometry has emerged as a very powerful platform to analyze complex samples (92). In LC-MS, chromatographic separation of peptides is based on their different strengths of hydrophobic interaction with a stationary phase, typically C<sub>18</sub>-silica phase. Peptide elution then occurs in a time dependent manner by an increasing concentration of an organic solvent such as acetonitrile. Eluting peptide species are then ionized via ESI and directly transferred to the vacuum of the mass spectrometer. The mass spectrometer continuously scans the peptide mass range, recording the masses and intensities of the eluting peptides. For peptide identification, tandem MS (or MS/MS) is employed, in which specific peptides are selected, isolated and fragmented by collision with neutral gas molecules such as nitrogen, argon or helium. A pre-defined number of the most intense peptides per MS1 scan (also termed survey or full scan) is selected and subsequently fragmented by collision-induced

dissociation (CID) or higher-energy collisional-dissociation (HCD). This strategy is called data-dependent acquisition and produces alternating cycles of full scans and fragmentation scans throughout the entire chromatographic retention time. CID or HCD fragmentation primarily break peptide bonds, resulting in b-ion (N-terminal part) or y-ions (C-terminal part). In current bottom up proteomics, hybrid Orbitrap systems have become very popular and are used exclusively in this thesis. They typically include a selection quadrupole, which is coupled to a collision cell and a 'C-trap' for injection of the ions into the Orbitrap analyzer itself (93). In a typical shotgun proteomics experiment, hundreds of thousands of tandem mass spectra are recorded in this way. Correctly assigning them to their calculated spectra obtained from *in silico* digestion of the corresponding peptide sequence is a challenging task in the computational downstream analysis. MaxQuant is a widely used software package for the analysis of shotgun proteomics data developed by our group (94). Peptides (MS1 features) are identified by first assigning them at parts-per-billion (ppb) mass accuracy to compatible peptides retrieved from a proteome reference file, and secondly by using the probability-based Andromeda search engine (95) to stringently match the observed fragment ions (MS/MS information) of each peptide to the theoretical fragment ions. A scoring function then identifies the best database hit among all submitted comparisons (Fig. 4).



**Figure 4: Peptide identification in tandem mass spectrometry.** Measured tandem mass spectra (left) and theoretical mass spectra retrieved from the *in silico* digestion of a reference proteome (right) are matched to each other for the probability score based peptide identification process. Adapted and modified (91).

The resulting peptide spectral matches (PSM) are finally assembled back into proteins. Due to the inherent fact that not all identified tryptic peptides are unique to one protein, the concept of protein groups is widely used in the field of shotgun proteomics. In order to deal with this 'protein inference' problem, proteins are assigned to the same protein group if there are only shared peptides among the proteins within one group. This usually involves splice variants from the same genomic locus or proteins sharing similar domains.

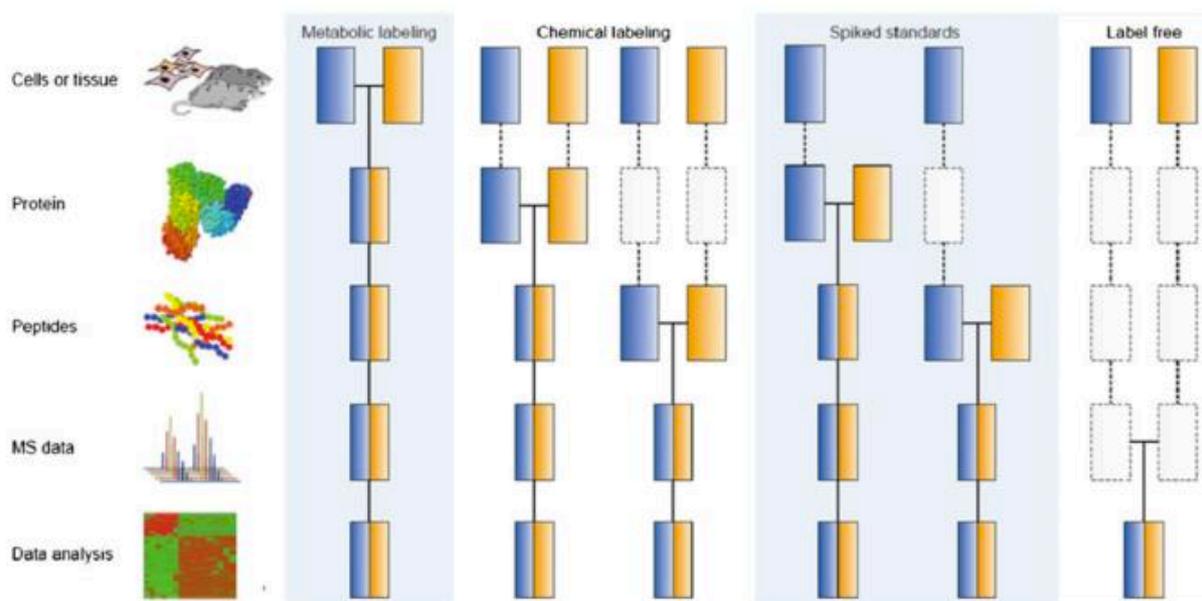


**Figure 5: A representative shotgun proteomics workflow.**

Following homogenization and lysis of cells or tissues, extracted proteins are digested with sequence specific enzymes such as trypsin and separated via high performance liquid chromatography (HPLC). Electrospray-ionized peptides are then analyzed by the mass spectrometer in data-dependent mode. A pre-defined number of the most intense peptides per full-scan are then subjected to fragmentation via higher-energy collisional dissociation (HCD). The resulting fragment ions contain peptide sequence information and are then used to search a reference database to identify the measured peptides. Peptides are then assembled into proteins using bioinformatic tools (Maxquant), based on their published sequences. Finally, downstream bioinformatic analyses include descriptive and inferential statistics for data interpretation. Adapted (8).

Mass spectrometry itself is not inherently a quantitative methodology. This is due to the fact that different analytes can strongly vary in their physico-chemical properties such as size, charge or hydrophobicity, which can ultimately lead to different mass spectrometric signals (4). Consequently, relative quantitative information can only be retrieved by comparing identical peptide species across samples. Peptide signals from different experimental conditions are compared to each other without any a priori information about the absolute protein concentration in the sample. To this end, various strategies have been developed over the years, each of them having unique strengths and weaknesses. They can conceptually be divided into label-free or stable isotope labeling approaches and the most commonly used approaches summarized in Figure 6.

For stable isotope labeling, specific mass shifts are introduced according to the heavy and light stable isotope labels employed. This can be achieved either metabolically, chemically or enzymatically during protein digestion. Quantitative information is extracted by comparing the intensities of the light to the corresponding heavy peptides. In metabolic labeling such as stable isotope labeling with amino acids in cell culture (SILAC) (96), heavy isotopes are introduced at the earliest possible experimental time point through growth media or food containing the heavy isotopes.



**Figure 6: Labeling strategies for quantitative proteomics and their accuracy.** Different colors (blue and yellow) represent different experimental conditions. Horizontal lines show when the samples are combined. Dashed lines indicate stages of potential experimental variation. Adapted from (2-4).

Consequently, metabolic labeling is by far the most accurate method as variation caused by different sample handling can be neglected. In SILAC, a combination of heavy arginines and lysines with tryptic digestion is usually used in order to generate peptide species that contain at minimum one heavy amino acid (except for the last, C-terminal peptide). This is crucial as non-labeled peptides are not distinguishable from the light sample. Therefore, SILAC strictly depends on successful label incorporation, which can substantially vary between different cell lines and might require several rounds of cell divisions. Primary cells or clinical samples such as tumor tissues can generally not be labeled with SILAC amino acids. To address this limitation, multiple labeled cell lines can be combined and used as a surrogate reference termed super-SILAC mix (97). These cell lines should ideally closely resemble the tissue being studied in order to provide a reference to as many endogenous proteins as possible. However, as cell lines may not fully reflect the entire proteomic composition of tissues, e.g. due to the presence of multiple cell types in the tissue microenvironment, or tissue-specific extracellular matrix proteins, super-SILAC quantification – despite its accurate and robust quantification – inevitably suffers from less quantitative depth compared to label-free approaches. Furthermore, the presence of different label states increases the MS1 level complexity.

In chemical labeling approaches such as for example isobaric Tag for Relative and Absolute Quantitation (iTRAQ) (98) or tandem mass tags (TMT) (99), isobaric tags are introduced chemically by the covalent binding to primary amines on the peptides. Due to the identical masses of the used isobaric tags, which is achieved by balancing groups, peptides derived from samples labeled with different isobaric reagents are indistinguishable at the MS1 level. This is advantageous because, in contrast to metabolic labeling, the MS1 level complexity does not increase. Upon fragmentation, distinct reporter ion species are then generated, which provide relative peptide quantification. Current versions of TMT allow analyzing up to ten samples at once in the same mass spectrometric run (TMT 10-plex) (100). This is a very versatile approach for the comparison of a large number of samples, for example in discovery based clinical proteomics projects that aim at screening large patient cohorts over a short time period. Disadvantages include more complex fragmentation spectra, chemical side reactions and mainly the co-fragmentation of peptides with similar elution profiles, which inherently result in an underestimation of the actual

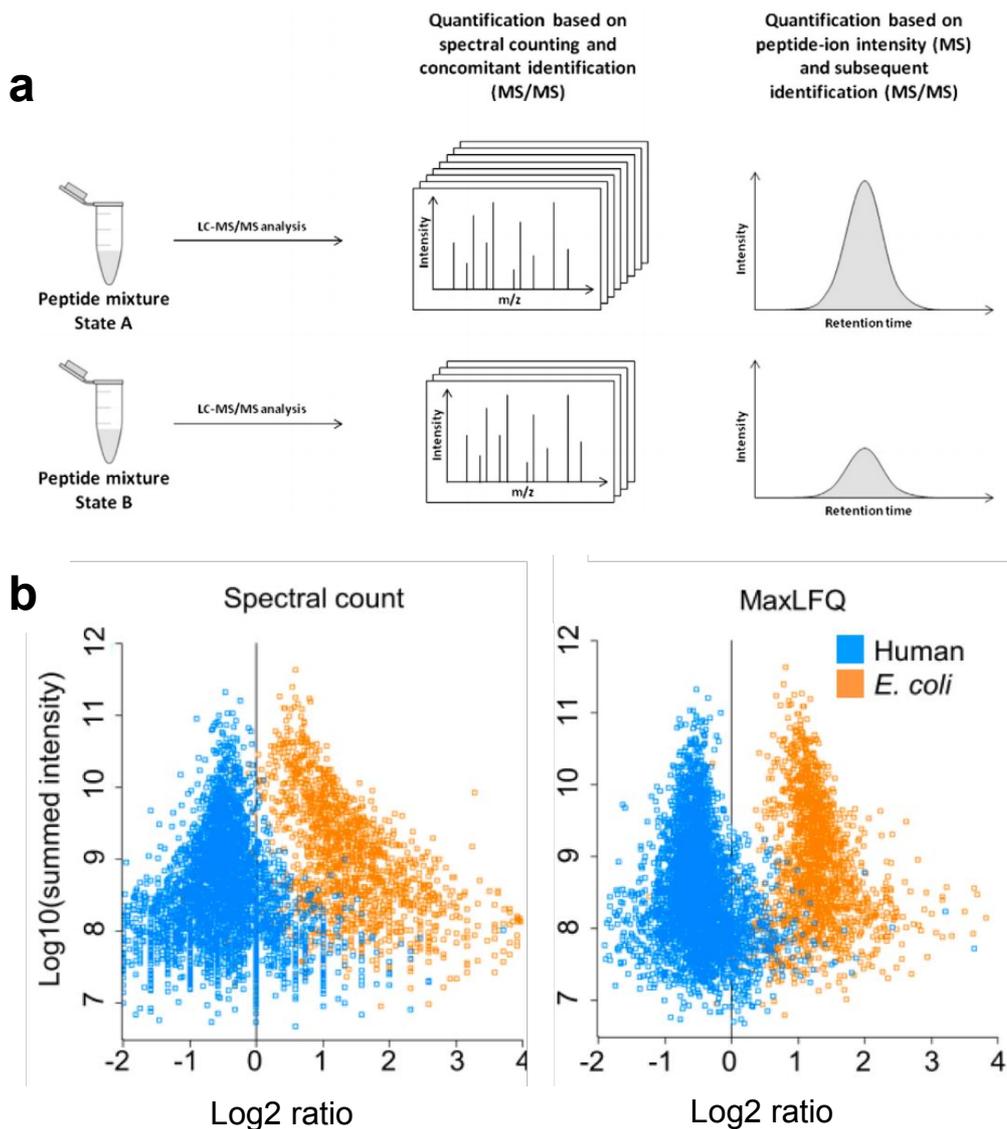
protein/peptide abundance ratios ('ratio compression' (101)), unless specialized methods are used (85).

The method used exclusively in this thesis is label-free quantification (LFQ). Here, the intensities of identical peptide species from multiple mass spectrometric runs are directly compared to each other without any isotope labeling. Consequently, LFQ can principally be applied to any type and number of biological samples of interest. As illustrated in Fig. 7a, LFQ can be further subdivided into spectral counting or intensity based approaches (102). Due to higher quantitative accuracy, intensity-based approaches such as the MaxLFQ algorithm (103) of the MaxQuant software (94), are superior to spectral counting approaches as these simply correlate the number of peptide spectral matches for a given protein to its abundance (102). Spectral counting can be especially problematic for the quantification of small or low abundance proteins that may not provide enough data points (spectral counts) for precise quantification (Fig. 7b).

Due to the completely separate sample handling in label-free quantification, experimental variability can theoretically be introduced at all stages in the proteomic workflow, potentially reducing the quantitative accuracy of LFQ workflows. However, advanced software has now been developed, which is capable of dealing with technical variations introduced during sample processing or MS measurement (103). This makes label-free quantification a very attractive and cost-effective alternative to label-based approaches (although at the cost of increased replicates). Post-processing of proteomic raw files - apart from peak picking or noise reduction - generally includes normalization algorithms to adapt retention times and MS intensities (10, 103). In the context of studies that aim at comparing a large number of samples at great proteomic depth, large dynamic range and reasonably high quantification accuracy, intensity-based LFQ has gained strong momentum in recent years. In particular, MaxLFQ, which is a part of the MaxQuant bioinformatic suite, incorporates very sophisticated re-normalization algorithms and the maximum number of pairwise comparisons (103) and has been used throughout this study. Remarkably, MaxLFQ enabled the time-resolved quantification of the *in vivo* insulin signaling pathway covering >10,000 phosphorylation sites without any sample pre-fractionation (104). The LFQ workflow was highly reproducible and, as experimentally assessed, only marginally contributed to the total sample variability. In fact, biological differences between tissues or variations in LC-

MS measurements caused higher variability. In this way, the study accurately quantified phosphorylation changes on timescales of less than 1 min and from fold changes as small as 2-fold.

In summary, intensity-based LFQ workflows are now cost-effective, reproducible, streamlined, and universally applicable. The high level of quantitative accuracy and precision, which can be achieved today, has made LFQ a true and valid alternative to label-based approaches. Apart from relative quantification, the summed LFQ values of the peptides identifying each protein can also be used as a rough proxy for the absolute amount of protein.



**Figure 7: Label-free quantification strategies and their accuracies.**  
**a.** The two types of label-free quantification strategies are spectral counting and peptide intensity-based quantification. Quantification based on spectral counting (middle panel) compares the number of MS2 scans recorded for each sample in order to estimate relative peptide/protein abundances. Right panel: peptide intensity-based quantification relies on the comparison of peak intensities along the peak retention time. **b.** Intensity-based quantification approaches (e.g. the MaxLFQ algorithm) are more accurate than spectral counting approaches. Log ratios of *E.coli* (orange)/human (blue) 3:1 vs 1:1 samples are plotted against the estimated absolute protein abundance (summed peptide intensities (log<sub>10</sub>) of the 1:1 sample). Note that quantification ratios are less accurate for spectral counting compared to MaxLFQ as determined by the larger ratio scattering in the lower intensity (y-axis) region. Adapted and modified from (10).

## 1.4 Clinical cancer proteomics

The first mass spectrometric clinical test was recently approved by the federal drug administration (FDA) in the United States, detecting a large number of pathogenic yeast and bacteria strains (105, 106). The simultaneous and timely identification of almost 200 different strains with a single device provides a significant advance in the field as conventional methods require culturing of microorganisms, which can take several days and may be problematic for severe infections where the correct clinical treatment is time-sensitive. This example clearly shows that current proteomic technology has the capacity to improve or even replace current clinical standard tests. For diseases such as cancer, this may open the door for proteomics-based personalized healthcare to guide patient diagnosis, treatment and clinical care.

### 1.4.1 Cancer proteomics: A brief history

First attempts to profile protein expression in a disease context such as cancer already date back to more than half a century ago when Fine and Creyssel reported in 1959 abnormal serum levels of globulin among patients with myeloma and macroglobulinemia (107). This observation was based on starch gel electrophoresis (108), a technique that – although revolutionary at the time - allowed the detection of only a few protein bands. More complex proteome analyses became feasible with the development of isoelectric focusing and polyacrylamide gels in the 1960s, which subsequently led to the development of two-dimensional polyacrylamide gel electrophoresis (2D PAGE). Because hundreds of protein spots could be visualized simultaneously, the era of 2D PAGE can be seen as the beginning of protein-centric systems biology. 2D PAGE was then extensively applied to a variety of cancer diseases. For example, one study reported the identification of acute lymphoblastic leukemia cell subgroups based on the analysis of 413 protein spots, which also led to the discovery of putative markers for specific B or T cell lineages (109). However, despite its wide application, 2D PAGE has remained a cumbersome methodology for the detection and identification of low abundant proteins due to the limited dynamic range of detection and unreliable reproducibility. In addition, protein identification via ‘Edman degradation’ was laborious and no longer adequate for the ever increasing number of detectable protein bands or spots. These unsurmountable hurdles strongly

increased the interest for alternative analytical methods such as mass spectrometry; however, its application had to wait for major technological breakthroughs in the field, namely the development of the soft ionization techniques MALDI and ESI techniques for peptide or protein ionization. Early applications of MS for protein combined 2D PAGE to first separate proteins by their molecular mass and isoelectric point and secondly to identify them via specific peptide mass maps obtained by the mass spectrometer (110). This was a paradigm-shifting development due to the rapid, sensitive and systematic nature of protein identification. That said, gel-based protein separation eventually was superseded by more direct approaches. In particular, ESI MS in combination with sample pre-fractionation and isotopic labeling has replaced that technology and has been used extensively for the large-scale analysis of proteins derived from biological samples such as cell lines, tissues or body fluids for many years (111).

Today, MS-based proteomics is coming ever closer to its goal of detecting all expressed proteins of complex samples such as cancer cell lines. This is due to great technological improvements of the current generation of mass spectrometers, as well as refined sample preparation and computational workflows (112). For the field of cancer proteomics, this has opened up a new era of systems biology to decipher signaling pathways, identify novel protein signatures related to tumor initiation and progression, or to discover new biomarker candidates. However, despite the great level of excitement and enthusiasm that current technologies provide, pre-mature claims from the early 2000s have affected the reputation of the cancer proteomics field. In 2002, a study was published that reported the unambiguous detection of early ovarian cancer based on unidentified proteomic patterns obtained by surface-enhanced laser desorption and ionization (SELDI) (113). This diagnostic test was reported to have an unprecedented diagnostic sensitivity (100%) and specificity (95%), which consequently gained great attention in the scientific and non-scientific media. Shortly after publication, doubts about effectiveness of the test accumulated (114) and it turned out that the obtained results were not reproducible due to artifacts from different sample processing (115, 116). Notably, even noise regions were sufficient to distinguish OvCa patients from normal individuals in the test samples, suggesting that systematic differences in sample handling strongly contributed to these results. The lessons from this study should guide the current development of cancer proteomic approaches, caution researchers with respect to biomarker claims and motivate them

to rigorously avoid previous pitfalls. For example, the use of biologically meaningful data such as protein expression levels provides stronger scientific rationale for the stratification of patients than the mere presence or absence of dubious unassigned mass spectrometric peaks. Furthermore, to put the potential biomarkers into a biological perspective, modern proteomic technologies allow the unambiguous and system-wide analysis of protein-protein interactions (117) or signaling networks (118), aiding the study of protein function on the molecular level. Integrating these additional approaches may lead to more promising new cancer discoveries in the future.

In light of the improved proteome coverage, reduced biases and increased quantitative precision of modern MS-based proteomics technologies (92), the integration of whole-genome or transcriptome sequencing data is currently one of the most important goals for the field of cancer proteomics. This could allow to address alterations in the flow of genetic information from the genome to the proteome (119). Based on the efforts of 'The Cancer Genome Atlas' (TCGA) project, which sequenced, characterized and catalogued cancer-specific alterations for several human cancer types (120), the National Cancer Institute launched the 'Clinical Proteomic Tumor Analysis Consortium' (CPTAC) to systematically identify disease-relevant proteins and their underlying biological pathways. For example, CPTAC's proteogenomic analyses of colorectal cancer revealed novel proteomic tumor subtypes associated with clinical outcome, similar to their studies in breast (67) and ovarian cancer (5). Interestingly, protein levels in tumor tissues could not be reliably predicted based on DNA or RNA-bases methods alone, emphasizing the urgent need to expand proteogenomic efforts.

#### 1.4.2 Cancer tissue proteomics

Clinical cancer proteomics can be subdivided by the sample of interest - generally into tissue or body fluid analysis. Cancer tissue proteomics, as also conducted in this thesis, usually aims at the global, ideally proteome-wide, quantification of proteins expressed in the sample of interest to systematically address proteomic changes between e.g. healthy and diseased states. In particular, studies tailored to retrospectively analyze the proteome under certain clinical conditions, such as in the context of a drug response, are among the most prominent applications of cancer tissue proteomics. These approaches provide a historical snapshot of the protein

repertoire in the diseased tissue, which can be combined with knowledge about subsequent patient outcome events. This combination of reliable clinical data with proteomics information, could be a powerful tool to uncover new drug targets or disease-specific biomarkers. For example, in a recent lymphoma study conducted by our laboratory, a combination of tissue proteomics and machine-learning based classification reliably segregated patients with diffuse large B-cell lymphoma according to the cell-of-origin (121). Interestingly, this also revealed that a cell line derived signature can be used to stratify patients.

As most cancer tissue specimens are archived in the form of formalin-fixed and paraffin-embedded (FFPE) material, we have developed techniques to reverse formalin crosslinking of proteins to ensure deep and quantitative proteomic profiling to a depth of up to 10,000 proteins (122, 123). This allowed the systematic dissection of the proteomic composition of biobank-derived tumor cells in comparison to matched normal tissues. We found that cell-surface and nuclear proteins showed the strongest expression differences, suggesting de-regulated signaling pathways in the cancer cells. Furthermore, proteomic differences between primary and metastatic colorectal cancer were of minor nature indicating that no extensive proteome remodeling is required for the metastatic spread to distant organs (124). Interestingly, this observation was also made in breast cancer recently (125).

The utility of FFPE material for proteomic analyses may not drastically depend on storage time as even 28 years old samples have already been used successfully (126). Moreover, post-translational modifications such as phosphorylation and N-glycosylation can also be analyzed in FFPE material (127).

With respect to sample collection and processing, cancer tissue proteomics can be performed based on whole-tissue samples or homogeneous cell populations isolated after careful microscopic inspection of the area of interest via laser-capture microdissection (LCM) (128). This is especially advantageous for the analysis of distinct cell types present in the heterogeneous tumor tissue environment such as cancer cells, fibroblasts, immune cells or endothelial cells. Indeed, LCM has been widely applied to identify new biomarker candidates in a variety of diseases such as prostate (129), breast (130) and ovarian cancer (131).

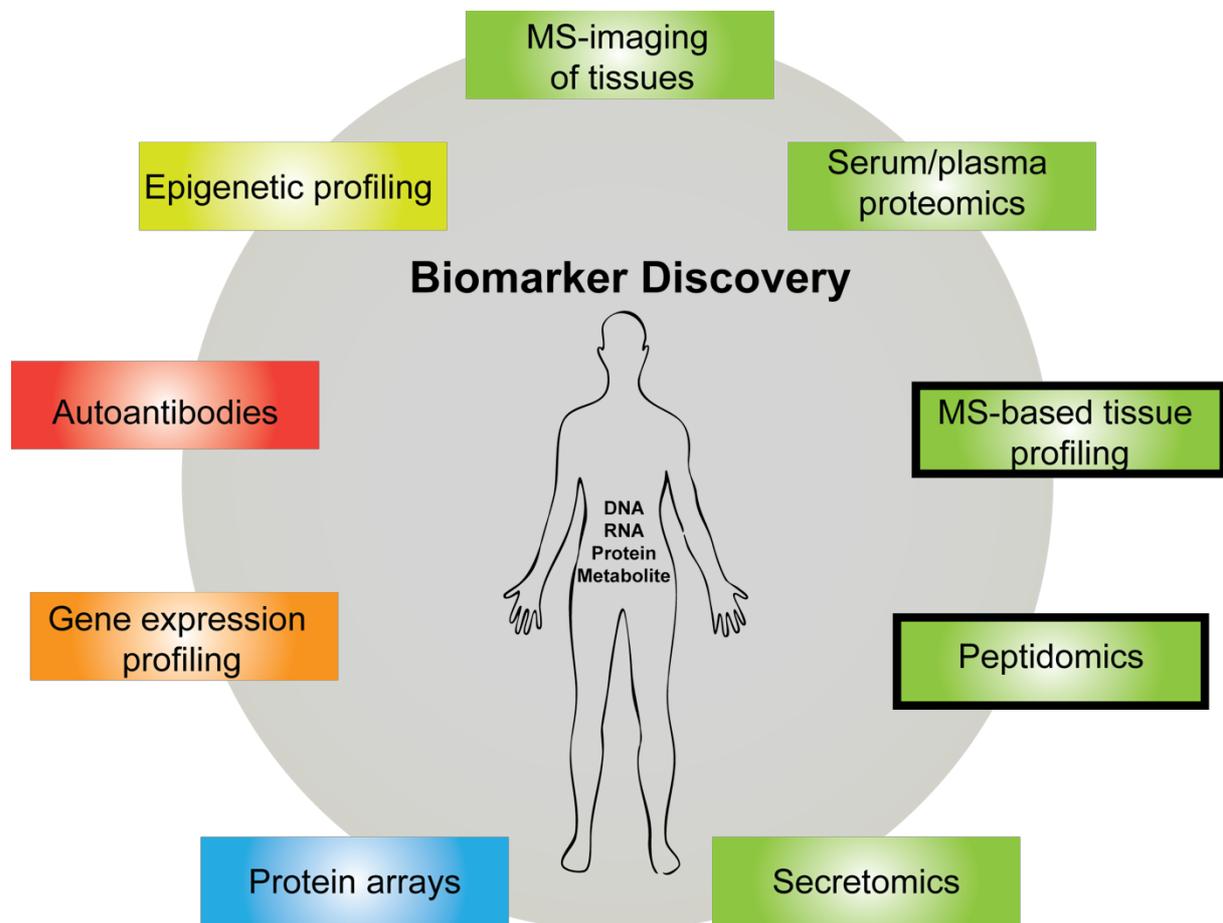
Due to low sample quantities, LCM workflows are inherently more challenging and require optimized sample processing to minimize any potential sample loss. This can be aided either by using filter devices (123) or MS compatible compounds for tissue lysis and protease digestion such as trifluoroethanol (132), which is fully volatile and thus enables simple buffer exchange via vacuum centrifugation without any additional sample loss. In this way, as few as 1.000 microdissected cells have been analyzed already in 2011, although only to a depth of a few hundred proteins (132). Clearly further technological improvements in mass resolution, accuracy and sensitivity are needed to substantially increase the proteome coverage of these 'micro-proteomic' approaches.

### 1.4.3 Biomarker discovery

#### 1.4.3.1 Types and technology

According to the National Cancer Institute, a biomarker is defined as:

*“A biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease. It may be used to see how well the body responds to a treatment for a disease or condition “.*



**Figure 8: Common strategies for biomarker discovery.**

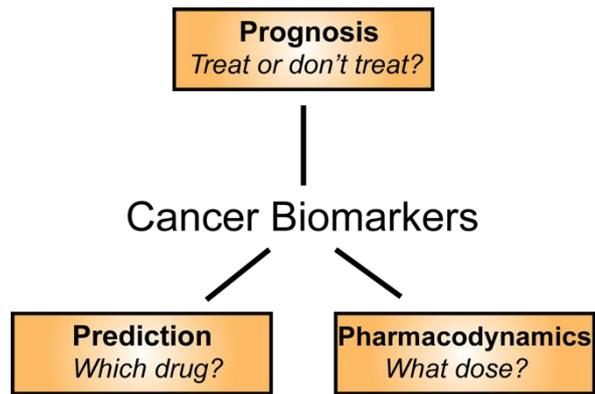
Green boxes indicate the involvement of mass spectrometry. The two marked green boxes indicate methodologies that were used in this thesis (Results sections 2.2 and 2.3).

In respect to personalized medicine and the development of highly effective anticancer drugs, biomarkers can be grouped into prognostic, predictive or pharmacodynamic types (Fig. 9) (7).

Prognostic biomarkers such as the gene expression based Mamma Print (Agendia), Oncotype DX (Genomic Health) or the H/I test (AviaraDx), are used to predict the clinical outcomes of breast cancer patients and to stratify these patients according to

their likelihood of benefiting from a particular therapeutic intervention. Prognostic biomarker studies therefore usually aim at analyzing treatment naive tissue or body fluid specimens collected during the initial surgical procedure, usually shortly after diagnosis. In contrast, a predictive biomarker may indicate the probability of success for a specific treatment. A classic example of a predictive biomarker is the genomic amplification of the *ERBB2* gene in breast cancer which indicates the clinical efficacy of the *ERBB2* blocking antibody trastuzumab (Herceptin). Furthermore, leukemia patients carrying the PML–RARA translocation respond remarkably well to all-trans retinoic acid (ATRA) therapy. The third group of ‘pharmacodynamic biomarkers’ is used to monitor treatment outcome. They are important to assess the optimal drug dose or to define the best route of administration. The identification of pharmacodynamic biomarkers is challenging for solid tumors due to the limited availability of tissues from multiple time points during the treatment course.

Molecular biomarkers include DNA or RNA molecules, as well as proteins and metabolites. Consequently, various analytical platforms and strategies have been developed over the years for biomarker discovery (see Fig. 8). As proteins represent the functional molecular machineries of the human body and are actively involved in disease onset, progression and/or resolution, they are principally well suited as molecular biomarkers to diagnose and monitor patients or to guide treatment and assess therapeutic outcomes (133). However, the discovery of new protein biomarkers is much more challenging than that of DNA or RNA-based ones. This is due to the comparative immaturity of MS-based proteomics and to the large dynamic range of



**Figure 9: Biomarker types. Adapted and modified from (7).**

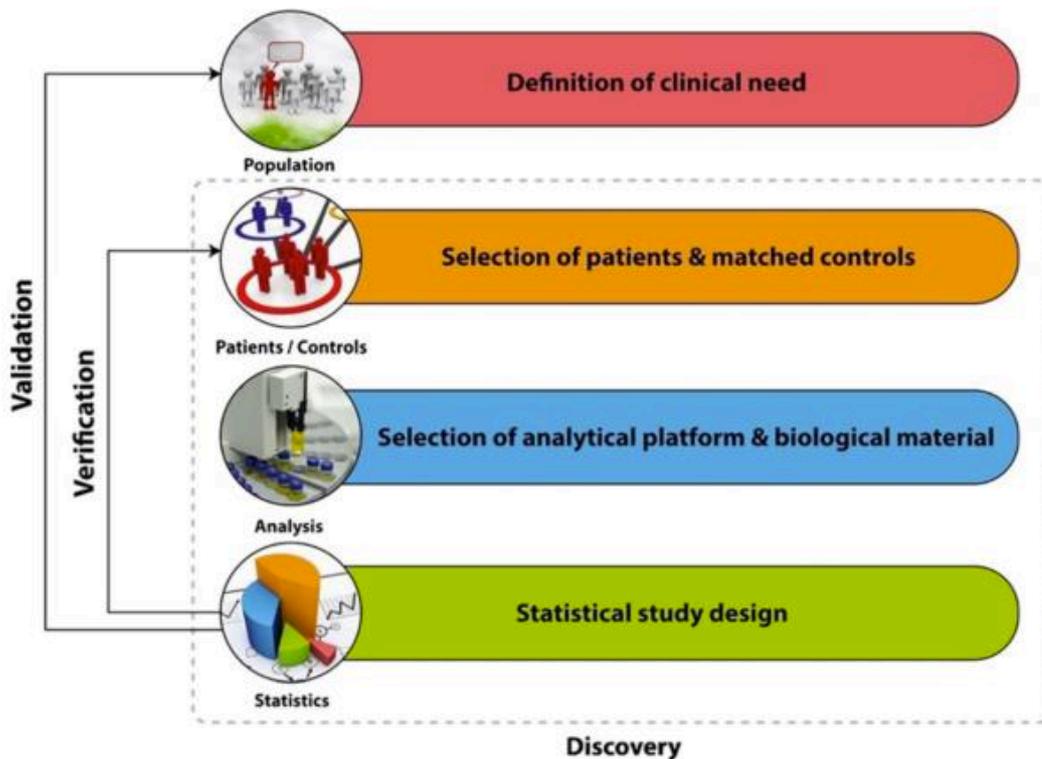
protein concentrations found in tissues and body fluids. In human plasma for example, protein concentrations are distributed over more than 10 orders of magnitude (134), an overwhelmingly large span compared to the dynamic range of detection of roughly 6 orders of magnitude for a typical state-of-the-art shotgun proteomics measurement (92). Multiple strategies have attempted to overcome the dynamic range problem such as antibody-based depletion of the most abundant plasma proteins or extensive sample pre-fractionation. However, these methods usually come at the cost of reproducibility and throughput, two very important aspects in the design of a clinical proteomics study. This has prompted our group to develop a highly reproducible single-run plasma proteomics workflow, employing short 20 min gradients, which is capable of robustly quantifying hundreds of proteins, among them more than 40 FDA-approved biomarkers (135). Once MS-based proteomics develops more powerful and reliable workflows to characterize the plasma proteome in greater depth, it may uncover novel biomarkers or protein patterns in an unbiased fashion.

The classical approaches for the discovery of protein biomarkers are immunoaffinity-based, as there is currently no PCR equivalent available for proteins to amplify signals of low abundant analytes. Depending on the assay, they can detect protein biomarkers in the low nanogram/milliliter range from blood samples (136). However, by their nature they are 'biased' in the sense that biomarker candidates have to be selected a priori.

#### *1.4.3.2 Biomarker study design*

Great care has to be taken in designing and performing discovery proteomics based biomarker studies, as large datasets inevitably produce a high number of false-positive results without suitable statistical validation. These wrong 'hits' may be eliminated by proper design of verification and validation studies. A recent search of SciFinder Scholar™ resulted in roughly 20,000 'hits' for the terms protein or peptide biomarker (137). However, this stands in stark contrast to the actual number of roughly one FDA approved biomarker per year since 1998 (136). Although the long and laborious route from the discovery phase to the clinical application is one reason for the low number of new biomarker tests, it is also apparent that flaws during primary study design or realization inevitably prevent the initial discovery of meaningful candidates.

Figure 10 summarizes the different phases of a typical biomarker study. In this thesis, we applied discovery-based clinical proteomics to two ovarian cancer projects, partly following such an outline (Results sections 2.2 and 2.3).



**Figure 10: Phases of biomarker study design (6).**

The definition of a distinct clinical problem or need is the starting point of any hypothesis-driven discovery proteomics study. The research question should be as simple as possible to simplify later data interpretation. A well-defined homogenous patient cohort should be selected according to clinico-pathological characteristics such as disease subtype, stage or medication as well as socio-demographical background like gender, age and body mass index. The size of the discovery cohort has to be chosen with respect to the expected inter-patient variability as well as assay variability. Typical numbers for discovery proteomics based studies are in the range of 10-50 patients, however, this may strongly vary depending on the clinical question, sample type and analytical platform as well as available resources. Moreover, analytical strategies are usually defined based on the exact clinical question and sample type of interest. As an example from this thesis (Results section 2.3), a highly sensitive single-run mass spectrometric workflow needed to be developed and adapted, which was

capable of analyzing low-input samples (~ 5.000 – 25.000 cancer cells) derived from laser-capture microdissection. If sufficient material is available, pre-fractionation steps or PTM enrichment strategies can also be applied to tissue or body fluid samples prior to MS analysis (123, 138, 139).

The selection of the appropriate sample source is a very critical aspect of the study design. For body fluid analyses, samples can generally be categorized according to their proximity to the tissue of interest (6). Proximal fluids such as cerebrospinal fluid (CSF) or ascites are ideal sources for disease-specific biomarkers due to likely high local analyte concentrations (136), however, their availability is usually limited due to the invasive nature of collection. The second category is comprised of non-invasive body fluids such as urine or plasma, which are available in large quantities. However, disease-specific biomarkers present in these samples may be difficult to detect due to strong dilution effects in addition to the large dynamic range in plasma. With respect to sample collection, care has to be taken regarding any pre-analytical variability introduced by different sample handling or storage conditions (140). The use of approved standard protocols for sample collection and handling reduces variability (141).

Downstream data analysis should incorporate stringent statistical testing to identify significant 'true' hits in the dataset. Especially for large-scale 'omics' datasets, Bonferroni correction, Benjamini-Hochberg or permutation-based false-discovery rates, need to be applied in addition to the primary statistical tests to account for multiple-hypothesis testing. The outcome is usually a list of biomarker candidates, which need to be further verified in a larger and independent patient cohort. In this context, antibody-based methods such as tissue microarrays (TMA) or ELISA are frequently used to screen large patient cohorts in a candidate-driven, targeted manner. Furthermore, targeted, quantitative MS such as multiple reaction monitoring (MRN) in combination with isotope-labeled standards can be applied for biomarker verification, with the promise of high accuracy, sensitivity, and reproducibility (142).

Finally, for clinical approval, 'validation studies' may need to be performed in which thousands of samples are analyzed to assess clinical performance (i.e. sensitivity and specificity) of the biomarker test.

In summary, the proper design of biomarker discovery studies is extremely challenging and does not follow a 'one-fits-all' rule. Instead, each phase has to be adapted to the end goal addressing a precise clinical need.

#### 1.4.4 Cancer antigen discovery

The immune system is extremely efficient in recognizing tumor-related antigens presented on human leukocyte antigens (HLA) complexes, which can ultimately result in T-cell mediated tumor rejection and long-term protection. This has led to the design of highly effective immunotherapies to treat diverse of human cancers (57).

Generally, cancer immunotherapy comprises a variety of strategies to activate the patient's immune system against the diseased, neoplastic tissue. Among them, checkpoint inhibition strategies, which aim at breaking the tumor's intrinsic ability to suppress tumor-directed T-cell activation, have recently shown remarkable clinical results in e.g. advanced melanoma (143) and an ever-expanding list of other neoplasms. This has prompted the field to uncover the identity and nature of the underlying tumor-derived antigens to develop more personalized cancer immunotherapies in form of e.g. cancer vaccines or adoptive T-cell transfers. The combination of these personalized approaches with checkpoint inhibition therapies and/or standard conventional therapies, such as chemotherapy, may unleash the full potential of cancer immunotherapy in the near future.

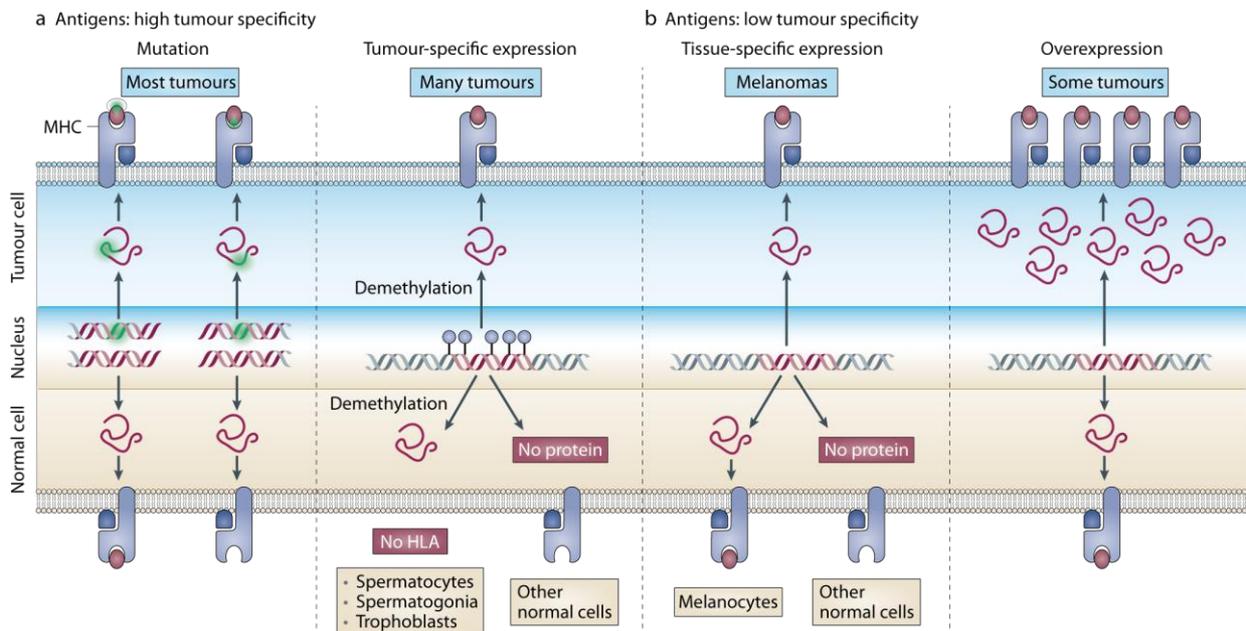
Tumor antigens can be grouped into tumor associated antigens (TAAs), tumor-specific antigens (TSAs) and cancer germline/cancer-testis antigens (CTAs) (144). TAAs are often differentiation antigens or overexpressed oncogenes, for example Wilms tumor 1 (WT1) (145) or Her2/neu (146). Due to the fact that TAAs are also expressed in normal tissues, their immunogenicity strongly depends on overexpression in the tumor to overcome immunological tolerance mechanisms (144). Given their frequent expression, cancer immunotherapy strategies targeting TAAs are principally applicable to a large number of patients. However, the immunoreactivity can be associated with severe side effects caused by the destruction of normal tissue (147).

TSAs, in contrast, represent antigens with high tumoral specificity such as neo-antigens derived from somatic non-synonymous mutations or chromosomal rearrangements, which are usually absent in the germline genome. The altered amino acid sequence can be presented on HLA complexes on the cell surface and finally be detected as foreign by T-cells. Therefore, the presence of TSAs usually correlates with

the mutational status of the tumor. Lung carcinomas, melanomas and colorectal carcinomas are prime examples for tumors with high mutational loads and presumably large neo-antigen repertoires (11, 148). One challenge of TSAs is that mutated antigens are rarely shared between patients, which limits the design of effective cancer vaccines applicable to a large number of patients. Nevertheless, patient-specific TSAs may be used in future to individually treat patients according to their available tumor antigen repertoire, an ultimate instance of personalized medicine. In particular, a combination of mass spectrometry, next generation sequencing (NextGenSeq) and bioinformatics was used recently to discover novel mutated neo-antigens that, when administered as vaccines, provided therapeutic protection in a mouse model (149). This provides a proof of principle that a personalized immunotherapy approach targeting patient-specific neo-epitopes might also work in a clinical context.

The last group of antigens is comprised of cancer germline/cancer-testis antigens. Due to their frequent and tumor specific expression, CTAs have been increasingly recognized as attractive cancer immunotherapy targets. CTAs are normally expressed in immune-privileged organs such as testis, fetal ovaries, and trophoblasts (150). In cancer tissues, CTAs can be activated at an advanced disease stage due to demethylation of CpG islands in promotor regions. According to the genomic locus, CTAs are grouped into an X chromosome-coded class (CTA-X) and a non-X chromosome-coded class (non-CTA-X) (150). CTAs make up an estimated half of all tumor-specific antigens that elicit spontaneous T-cell responses in cancer patients (11). The CTA database (<http://www.cta.lncc.br/>) currently lists 158 CT antigen families. Among them, the MAGE-A gene family and NY-ESO-1 are the most prominent CTA-X members, which are currently tested as potential cancer vaccines in the clinic. Of particular note, NY-ESO-1-specific T-cells were successfully used for an adoptive T-cell transfer in myeloma patients with promising clinical responses in 16 of 20 patients (151). To extend the list of expressed CTAs in various cancer types, much research has been dedicated to the discovery of frequently expressed CTAs and their underlying immunogenic HLA peptides. Based on RT-PCR data, CTAs are commonly expressed in lung cancer, hepatocellular carcinoma, bladder cancer, ovarian cancer, and melanoma (150). Taking advantage of to their known regulation by DNA methylation, a recent study employed a DNA de-methylation strategy to activate CTA expression with the goal of identifying CTA-derived HLA-I peptides by mass

spectrometry (152). This approach may furthermore represent a very attractive therapeutic approach for cancers with low CTA expression such as renal cancer, colon cancer, gastric cancer or leukemia/lymphoma. Indeed, treatment with 5-aza-2'-deoxycytidine, a drug causing global DNA demethylation, activated NY-ESO-1 expression, which in turn promoted T-cell mediated tumor cell killing (153). In addition, treatment with decitabine (5-aza-2'-deoxycytidine) was recently shown to be synergistic with immune checkpoint blockage therapy in a murine ovarian cancer mouse model (154), underlining the great potential of CTAs for immunotherapeutic interventions to treat cancer.



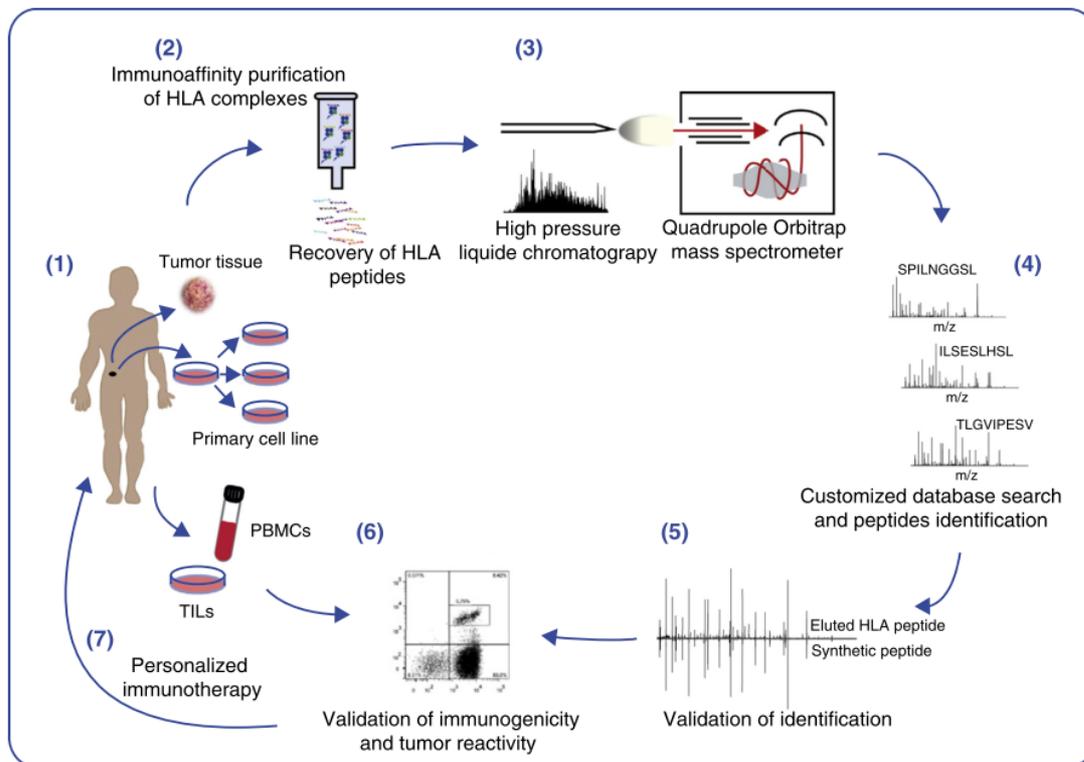
**Figure 11: Human tumor antigen classes.**

**a.** Tumor-specific antigens and their origins are shown. As a result of genomic mutations, mutant variants of peptides can be presented on HLA complexes on the cell surface. In addition, mutations can also enable HLA binding of non-binding peptides. Cancer-testis antigens are tumor-specific due to their non-physiological up-regulation via DNA demethylation in the tumor. Expression in germline cells does not lead to presentation due to the lack of HLA receptor molecules. **b.** Antigens with low tumor specificity and their origins are shown. Genes that are expressed tissue-specifically (e.g. melanocyte-specific) or at unnatural high levels (e.g. *ERBB2*) can also elicit anti-tumor responses (11).

Historically, the first T-cell recognized cancer antigen, MAGEA1, was discovered in 1991 using a genetic screening approach of a melanoma DNA library (155). In 1995 the SEREX method (156) (serological analysis of recombinant cDNA expression libraries) was established, which used a recombinantly expressed cDNA tumor library to identify specific serum antibody epitopes. Multiple CTAs were identified this way, for example NY-ESO-1 (157) and SSX2 (158). However, today in the age of whole-genome analysis and systems biology, new cancer-specific antigens are primarily identified by a combination of NextGenSeq approaches, such as whole-exome-sequencing, and bioinformatic *in silico* predictions for HLA binding (159-161). This strategy has been widely applied for cancer antigen discovery, however, as intracellular processing and antigen immunogenicity is generally not taken into account, many reported peptide hits are irrelevant *in vivo*. Consequently, immense experimental follow-up work is needed to filter for 'true' and T-cell reactive candidates. For example, in a recent study only half (54%) of the a priori known antigens were also recovered within the top three hits from an *in silico* prediction of a large number of peptides (162), highlighting the necessity for alternative approaches that identify HLA peptides in a completely unbiased and more direct fashion.

Mass spectrometry should in principle be a very powerful tool to directly measure and identify HLA peptides (HLAp) isolated from cell lines, tissue or plasma-derived HLA complexes (163-165). This approach is termed 'MS-based immunopeptidomics' and includes immunoaffinity purification of HLA complexes and subsequent isolation of the entirety of bound HLA peptides. These are then separated and analyzed by LC-MS. For HLA complex purification, pan-HLA class I and pan-HLA class II antibodies are used, however, allele-specific antibodies have also been reported in this context (166). Beginnings with just a few peptides more than 20 years ago (167), this today allows the simultaneous identification of thousands of HLA-I peptides in a single mass spectrometric run, without the use of peptide pre-fractionation. This has led to the identification of multiple cancer-related antigens in a large number of cancer diseases such as melanoma, glioblastoma, renal cell carcinoma, ovarian cancer, hepatocellular carcinoma and leukemia (163). Even neo-antigens resulting from somatic mutations can readily be detected by integrating NextGenSeq data of the analyzed sample (149). MS-based immunopeptidomics further allows discovering post-translationally modified HLAp, which might be exclusive to the tumor and therefore offer an additional level of immunogenic targets for immunotherapy. Validation approaches for the identified

HLAp include the design and measurement of synthetic versions of the candidate peptides, whose spectra should exactly match those of the experimentally observed ones. In addition, functional follow-up work is needed to test T-cell reactivity.



**Figure 12: MS-based immunopeptidomics workflow for antigen discovery and personalized immunotherapy.**

Antigen discovery by MS-based immunopeptidomics for personalized cancer immunotherapy. Immunoaffinity-based HLA complex purification from cancer tissues or cell lines. Isolated HLA-binding peptides are measured by HPLC-MS/MS, followed by database matching. In specialized laboratories this now allows to simultaneously identify thousands of HLA-I peptides in a single MS run. Selected candidate peptides are validated by obtaining mass spectra from synthetic peptide versions as well as through T-cell assays to determine HLA binding and immunogenicity. Downstream immunotherapeutic applications include the development of personalized immunotherapies such as peptide vaccines or adoptive T-cell transfers (9).

## 1.5 Aims of the thesis

In this thesis, state-of-the-art MS-based proteomics methodologies were developed and applied to investigate high-grade serous ovarian cancer (HGSOC) at the proteomic level. I completed three projects addressing important clinical questions related to HGSOC. All three projects were conducted in an exceptionally productive and fruitful collaboration with the Ernst Lengyel ovarian cancer laboratory at the Department of Obstetrics and Gynecology, of the University of Chicago.

The **first** project addresses an important question for any HGSOC-related pre-clinical investigation: which ovarian cancer cell lines are ‘good’ models to study HGSOC? We reasoned that an integrated and streamlined proteomic workflow applied to frequently employed ovarian cancer cell lines would offer a very attractive methodology to identify suitable cell line models by comparing their underlying proteomic profiles to those of the actual HGSOC tissues and to corresponding primary cells. Our single-run, label-free workflow identified distinct HGSOC cell line proteomes and refined a previously reported genomics-based cell line classification.

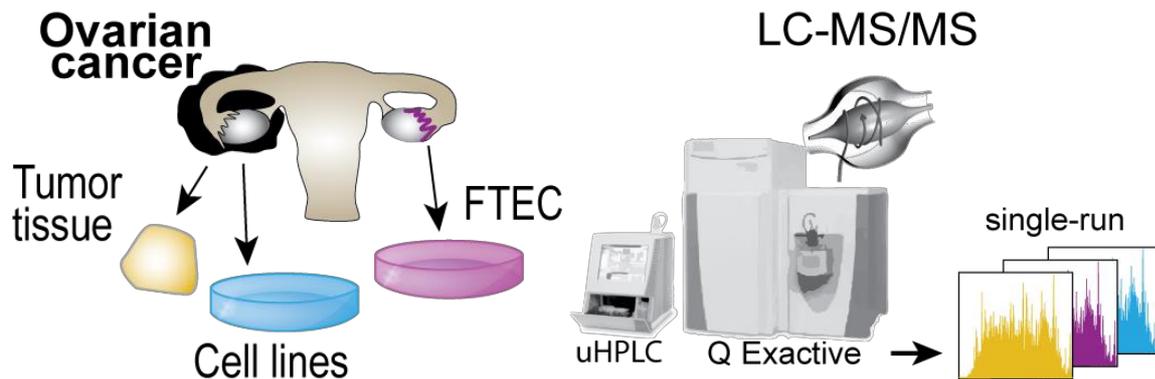
For the **second** project, I used the developed single-run workflow to retrospectively compare the proteomes of HGSOC patients with differential survival times following standard carboplatin/taxol chemotherapy. The in-depth analysis yielded a total of 9,000 proteins obtained from FFPE biobank samples and highlighted the power of discovery-based clinical proteomics to identify novel protein biomarker candidates linked to chemotherapy efficacy. Complemented by a variety of techniques including interaction proteomics and immunopeptidomics, we further gained functional and mechanistic insights into the biology of chemotherapy sensitive HGSOC. The results are currently being evaluated for potential use in a clinical setting in the future.

The **third** project of this thesis likewise focused on the analysis of FFPE biobank specimens and constitutes the first proteomic analysis of HGSOC progression. Towards this goal we developed and adapted a novel and highly sensitive and streamlined proteomic workflow to analyze as few as 5,000 tumor cells. This allowed us investigating the proteome of the tumor and its adjacent microenvironment in the context of HGSOC progression. Our results revealed novel insights into ovarian cancer biology and highlighted the importance of the tumor microenvironment for ovarian cancer progression. Current work addresses the potential clinical utility of our findings.

## 2 Results

### 2.1 Manuscript 1: Deep proteomic profiling of ovarian cancer models

#### 2.1.1 Project aim and summary



The large heterogeneity of ovarian cancer is reflected in the fact that the histopathological origin of the most frequently used OvCa cell lines is not completely clear (12). Research dedicated to the identification of bona-fide HGSOC cell lines has consequently gained strong interest in the OvCa community (22, 168). In a recent genomics-based study, HGSOC cell lines with high genetic similarity to HGSOC tumor tissues were identified, however, so far a proteomics-based classification comparing cell lines and tumors has not been undertaken. We therefore undertook such an integrated and streamlined MS-based proteomics approach, reasoning that it would represent a powerful technology for the identification of ‘good’ HGSOC cell lines.

For the first time, we acquired quantitative proteomic data (>10.000 proteins) for 30 ovarian cancer cell lines, primary cells and tumor tissues. This revealed a high proteomic similarity in terms of expressed proteins (77%). However, our data clearly showed large protein expression level differences for the analyzed cell lines indicating a pronounced quantitative proteome heterogeneity. This observation prompted us to investigate whether our data allowed the identification of distinct proteomic cell line subgroups.

Extending previous genomics-based findings (12), we identified two distinct HGSOC types of cell line proteomes, in addition to a group of clear-cell OvCa cell lines. The integration of immortalized ovarian surface epithelial cells (IOSE) and fallopian-tube epithelial cells (FTEC) further suggested a possible dual cell-of-origin based classification for the newly identified HGSOC cell line proteomes. Based on a novel 67-protein signature that we derived, we were then able to validate our findings in two independent datasets, among them a recent study of 84 HGSOC tumor proteomes. This showed that patients with an IOSE-like profile had a worse prognosis than those with a FTEC-like profile and suggests a previously overlooked simple stratification according to the cell-of-origin.

As a proof-of-concept, we additionally showed that our proteomic dataset can be used to select the most appropriate cell lines for various pre-clinical investigations. This delineated a strong up-regulation of several 'vitamin A pathway' proteins in a subgroup of cell lines with HGSOC-like genomic and proteomic features. For example, the transporter protein CRABP2, which promotes anti-oncogenic all-trans retinoic acid (ATRA) signaling (169), was highly expressed in these cells. We showed that our quantitative data correctly predicted the anticipated tumor-suppressive response to ATRA that we measured in our system.

Our proteomic dataset has been made available for the OvCa community on a user-friendly website: <http://maxqb.biochem.mpg.de/mxldb/project/list>.

### 2.1.2 Contribution

Ernst Lengyel conceived the initial idea for the project and proposed it to Matthias Mann and myself in 2013. Matthias Mann, Ernst Lengyel (University of Chicago), Karen Watters (University of Chicago), and I incubated and developed this project. Matthias Mann and Ernst Lengyel supervised me during all phases of the project. My contribution included the conceptual study design, proteomic sample preparation, proteomic data acquisition, as well as data analysis and interpretation. Karen Watters conducted all functional ATRA experiments and validation experiments. I designed the majority of the figures and tables in the publication and wrote the manuscript together with Karen Watters, Ernst Lengyel and Matthias Mann.

## 2.1.3 Publication



### ARTICLE

Received 23 Feb 2016 | Accepted 19 Jul 2016 | Published 26 Aug 2016

DOI: 10.1038/ncomms12645

OPEN

# Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status

F. Coscia<sup>1\*</sup>, K.M. Watters<sup>2,\*</sup>, M. Curtis<sup>2</sup>, M.A. Eckert<sup>2</sup>, C.Y. Chiang<sup>2</sup>, S. Tyanova<sup>1</sup>, A. Montag<sup>3</sup>, R.R. Lastra<sup>3</sup>, E. Lengyel<sup>2,\*\*</sup> & M. Mann<sup>1,\*\*</sup>

A cell line representative of human high-grade serous ovarian cancer (HGSOC) should not only resemble its tumour of origin at the molecular level, but also demonstrate functional utility in pre-clinical investigations. Here, we report the integrated proteomic analysis of 26 ovarian cancer cell lines, HGSOC tumours, immortalized ovarian surface epithelial cells and fallopian tube epithelial cells via a single-run mass spectrometric workflow. The in-depth quantification of >10,000 proteins results in three distinct cell line categories: epithelial (group I), clear cell (group II) and mesenchymal (group III). We identify a 67-protein cell line signature, which separates our entire proteomic data set, as well as a confirmatory publicly available CPTAC/TCGA tumour proteome data set, into a predominantly epithelial and mesenchymal HGSOC tumour cluster. This proteomics-based epithelial/mesenchymal stratification of cell lines and human tumours indicates a possible origin of HGSOC either from the fallopian tube or from the ovarian surface epithelium.

<sup>1</sup>Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. <sup>2</sup>Department of Obstetrics and Gynecology, Section of Gynecologic Oncology, University of Chicago, Chicago, Illinois 60637, USA. <sup>3</sup>Department of Pathology, University of Chicago Medicine, Chicago, Illinois 60637, USA. \*These authors contributed equally to this work. \*\*These authors jointly supervised this work. Correspondence and requests for materials should be addressed to E.L. (email: elengyel@uchicago.edu) or to M.M. (email: mmann@biochem.mpg.de).

Invasive ovarian cancer (OvCa) is a highly heterogeneous disease divided into four major histologic subtypes, namely serous, endometrioid, mucinous and clear cell OvCa. High-grade serous ovarian cancer (HGSOC) is the most common (70%) and aggressive subtype and is primarily responsible for the low survival rate<sup>1</sup>. Until recently, HGSOC was thought to originate exclusively in the ovaries, as tumours almost invariably involve the ovary. However, the discovery of a possible precursor lesion, serous tubal intraepithelial carcinoma (STIC), in the fallopian tube fimbria of BRCA-mutation carriers, as well as in HGSOC patients, provides strong evidence for the fallopian tube fimbria as the probable site of origin<sup>2–5</sup>. HGSOC is characterized by ubiquitous somatic TP53 mutations<sup>6</sup> and genetic instability<sup>7</sup>, and frequently evolves to a chemo-resistant state. From a molecular standpoint, it is classified into sub-groups based on characteristic gene expression signatures—differentiated, immunoreactive, proliferative and mesenchymal<sup>7,8</sup>.

Given that it is difficult to perform mechanistic studies with primary tissue, the necessity for cellular models for *in vitro* and *in vivo* experiments is apparent. However, these models should be as representative of the tumour as possible, as there is little clinical utility for experimental data obtained in cell lines that do not reflect the disease being studied; these results might be, at best, misleading and, at worst, harmful to patients. The time that has elapsed since many OvCa cell lines were established (some were created more than 30 years ago), coupled with the risk of switching or cross-contamination when propagated for a long time, and the only recent introduction of genomic ‘fingerprinting’ techniques, has led to the incorrect assignment of the tissue origin of many OvCa cell lines<sup>9</sup>. The recent establishment of The Cancer Genome Atlas (TCGA)<sup>7</sup> has opened the door for researchers to begin to address these uncertainties and should allow selection of the most representative cell lines on the basis of genomic and transcriptomic information. A number of recent studies have integrated these HGSOC genomic characteristics into their assessment of suitable cell lines, to better understand OvCa biology and find novel treatment targets. Domcke *et al.*<sup>9</sup> evaluated data from the Cancer Cell Line Encyclopedia (CCLE)<sup>10</sup> by comparing a panel of 47 OvCa cell lines to HGSOC tissue data available through the TCGA consortium<sup>7</sup> by means of copy number alterations (CNA), mutation frequency and gene expression data. The authors cautioned against the use of some of the more commonly used OvCa cell lines due to their poor overall resemblance to HGSOC in patients at the genomic level. Mitra *et al.*<sup>11</sup> and Elias *et al.*<sup>12</sup> recently highlighted some of the limitations of the newly described HGSOC cell lines in pre-clinical studies, including their limited and inconsistent ability to form tumours in immuno-deficient mice. Interestingly, the cells with less genetic resemblance to the TCGA tumours had a metastatic pattern that was very similar to that of human HGSOC (for example, disseminated abdominal tumour nodules and omental involvement, but no extra-abdominal metastasis)<sup>11,12</sup>.

It is as yet unknown to what extent the molecular characteristics reported by all these studies are represented at the protein level. Genomic and transcriptomic level analyses do not necessarily reflect the phenotype-defining proteomic profile. So far, there has been no direct and in-depth proteomic comparison between OvCa cellular models and tumour tissues. However, given that proteins represent the functional and phenotype-defining units of a cell, a quantitative proteomics approach should be superior to gene expression-driven comparisons alone. We therefore hypothesized that an integrated and streamlined mass-spectrometry (MS)-based proteomics approach<sup>13–15</sup> is a promising methodology for molecular subtype characterization. We here demonstrate that

directly integrating proteomic profiles from cell lines, tumour tissues and primary cells adds a highly informative level to the evaluation of OvCa cellular model systems. A cell line-derived 67-protein signature classifies OvCa tumours potentially arising in the ovarian surface epithelium (OSE) or fallopian tube epithelial cells (FTECs) and predicts functional properties of these cells. We also provide a user-friendly resource of the quantitative protein expression of 30 ovarian cell lines.

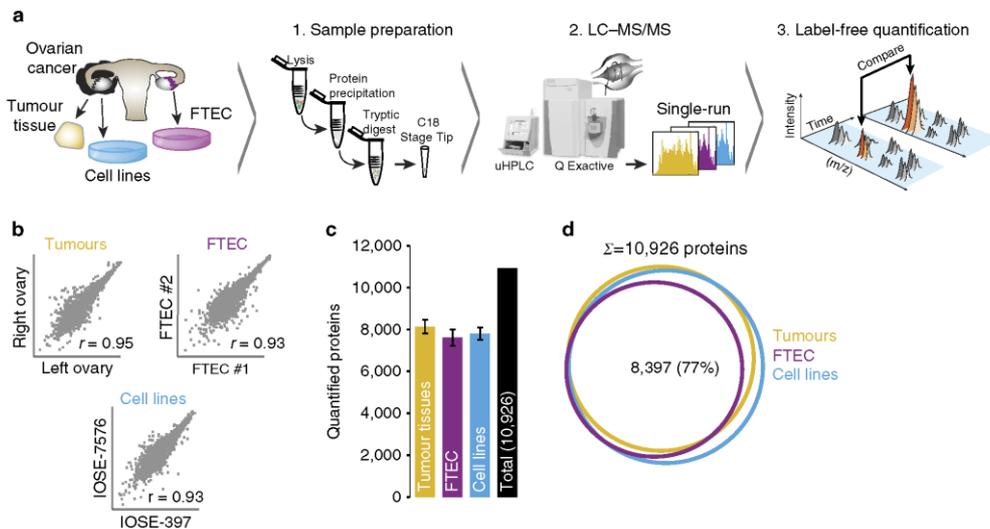
## Results

**Deep single-run proteomics of OvCa tissues and cell lines.** To the extent that proteomics has reached a reasonable depth of quantification, this has usually involved multi-step workflows with extensive fractionation and correspondingly long measurement times. Since we sought to compare a large number of OvCa proteomes, we instead adapted and applied a recently described method based on a single-run workflow<sup>15,16</sup>. Briefly, we performed tryptic digestion of the entire proteome followed by chromatographic separation using relatively long (4 h) HPLC gradients coupled to online mass spectrometric analysis on high-resolution quadrupole Orbitrap mass spectrometers (see the ‘Methods’ section). Using this workflow, we quantified the proteomes of eight HGSOC tumour tissues, 30 cell lines (26 OvCa; two cervical cancer; two immortalized ovarian surface epithelial (IOSE)) and three primary FTEC isolates (Fig. 1a).

In total, our analysis resulted in 229,004 unique peptide sequences corresponding to 11,070 distinct protein groups at a peptide and protein false-discovery rate (FDR) of less than 1% (refs 17,18; see the ‘Methods’ section). This remarkable depth, considering the absence of any fractionation step, was accompanied by a median protein sequence coverage of 43.6% by the identified peptides. Label-free protein quantification (LFQ) using the MaxLFQ algorithm<sup>19</sup> resulted in a median depth of 7,828 protein quantifications per single measurement. The LFQ values were highly reproducible as cell line replicates had median Pearson correlation coefficients of 0.95 (Supplementary Fig. 1a), whereas comparison of the two different IOSE cell lines and FTECs from two different healthy donors had only somewhat lower correlations of 0.93 each. Interestingly, HGSOC tumours from the left and right ovaries from a single patient featured a very high correlation of 0.95 at the proteome level (see also below; Fig. 1b). These results demonstrate high experimental reproducibility and cell-type homogeneity.

When we required quantification in at least two of three cell line replicates, we obtained a filtered data set of, in total, 10,926 proteins containing an average of 7,810 protein quantifications per cell line, 8,143 per HGSOC tumour tissue and 7,609 for the primary FTEC replicates (Fig. 1c). A total of 8,397 were present in all three filtered data sets (77%; Fig. 1d). More than 99% of the FTEC proteins were also detected in OvCa cell lines or tumour tissues and, likewise, nearly all proteins detectable in the tumour tissue at our depth of analysis were also found in the cell line proteomes (97%). The few unique proteins in the tumour data set were enriched for plasma proteins and these were filtered out in subsequent analyses (see the ‘Methods’ section). There was even a high representation of the OSE proteome as represented by IOSE cell lines in the combined cancer cell line proteome (99%) and a very high overlap with the FTECs (91%). Furthermore, protein intensities were comparably distributed between these systems, as were the percentages of total protein mass attributable to major cellular compartments (Supplementary Fig. 1b, Supplementary Fig. 1c).

**Protein expression is heterogeneous in OvCa cell lines.** We first analysed commonalities and differences in protein expression



**Figure 1 | Deep single-run proteomics of cell lines and human ovarian cancer tissue.** (a) Summary of the shotgun proteomics workflow for OvCa cellular models and high-grade serous ovarian cancer (HGSOC) tissues. Following lysis, protein purification, and tryptic digest, peptides were separated by ultra-high performance liquid chromatography and measured in single runs using a quadrupole Orbitrap mass spectrometer. Label-free proteome quantification was performed using the MaxQuant software environment. (b) Workflow reproducibility for cell lines ( $n=30$ ), HGSOC tissues ( $n=8$ ) and primary FTEC cells ( $n=3$ ). Pearson correlations ( $r$ ) were calculated for biological replicates of cell lines, primary FTEC isolates from different healthy donors, and ovarian tumour tissues from both ovaries from a woman with HGSOC. (c) Average number of quantified proteins from each sample type. Error bars represent standard deviations. (d) Number of proteins common to FTECs, HGSOC tumours and cell lines. FTEC, fallopian tube epithelial cell; LC-MS/MS, liquid chromatography tandem mass spectrometry;  $m/z$ , mass-to-charge ratio; uHPLC, ultra-high performance liquid chromatography.

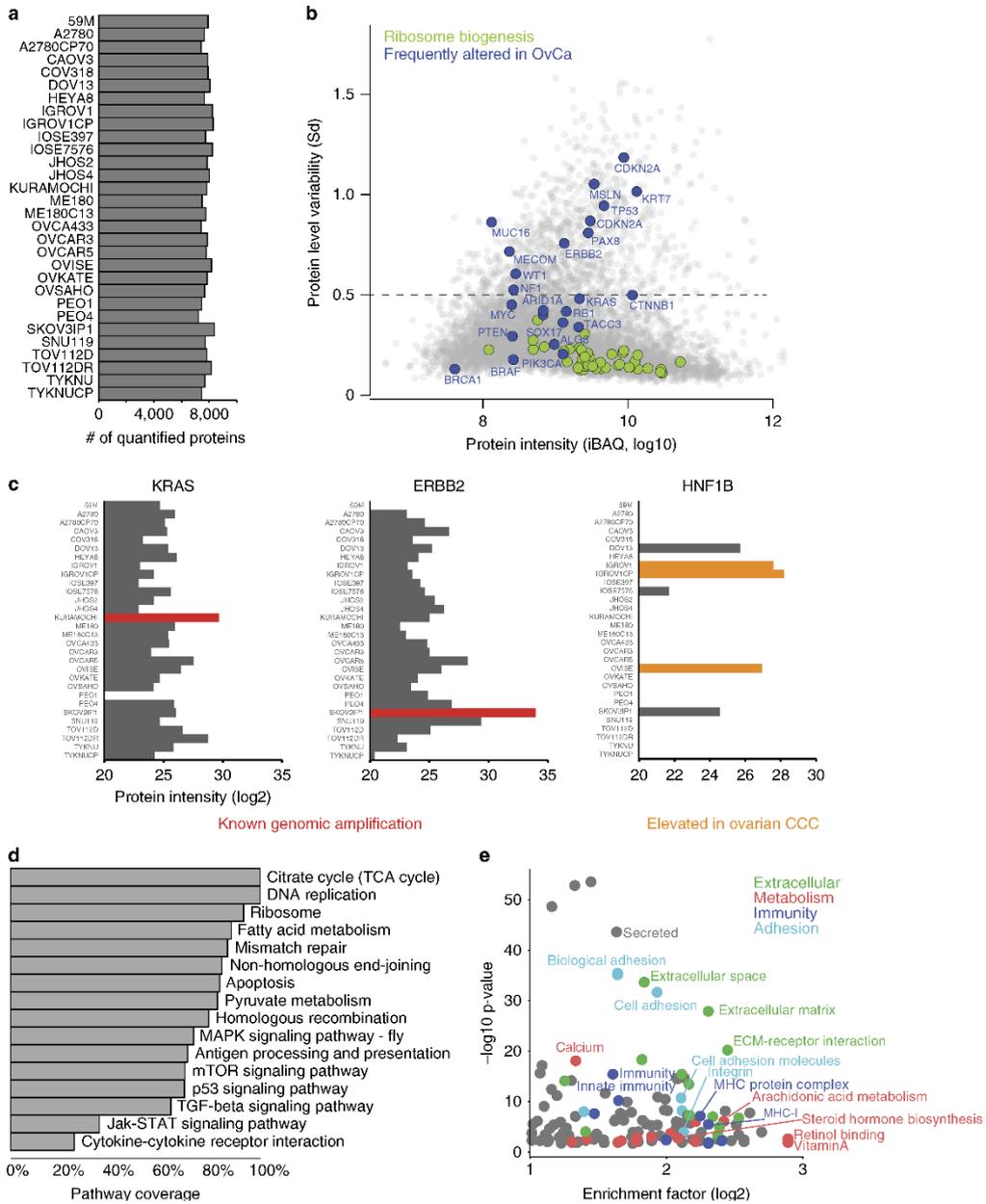
across 30 cell lines. Average filtered quantification depth and expression levels from the triplicate analyses of each cell line are shown in Fig. 2a and Supplementary Data 1. We asked which proteins and pathways were constant or differentially regulated across the cell lines and, to that end, calculated the protein level variability (Sd of logarithmized protein intensity) and plotted it against the estimated absolute protein abundance (sum of the intensity-based absolute quantification [iBAQ]<sup>20</sup> values calculated by MaxQuant, Fig. 2b). Proteins involved in household functions, such as ribosome biogenesis, were stably expressed across cell lines (median Sd < 0.18), whereas proteins known to be frequently lost, amplified or over-expressed in diverse OvCa subtypes<sup>7,9,21</sup> showed strong expression differences (shown in blue in Fig. 2b). To investigate this in more detail, we focused on known cell line-specific genomic alterations, such as the amplification of the receptor tyrosine kinase *ERBB2* in SKOV3IP1 or the proto-oncogene *KRAS* in KURAMOCHI<sup>9</sup>. Indeed, our data clearly reflected the amplification of these two genes in the expression profiles of those cell lines. Our data also indicated higher expression of *KRAS* in the carboplatin-resistant TOV112DR, compared with its parental cell line (Fig. 2c). Of note, the proteomic approach can detect protein upregulation regardless of the mechanism responsible. For instance, the transcription factor hepatocyte nuclear factor 1-beta (*HNF1B*), often over-expressed in clear cell ovarian cancer (CCC)<sup>22</sup> showed the strongest expression in the CCC OVISe cell line and the IGROV1 cell line, which, according to its genomic profile, may be of clear cell or endometrioid origin<sup>9</sup>. Furthermore, as shown previously, the proteomic data allow investigation of the expression status of all proteins in the amplicon<sup>23</sup>.

The complete data set covered a large proportion of the KEGG-annotated members of major biological processes and cancer-related signalling pathways, such as all 35 members of the DNA replication pathway and >70% of the p53 pathway (Fig. 2d). Pathway enrichment analysis across the cell lines identified the biological pathways that were most differentially regulated (Fig. 2e). For the quantified proteins that varied the most (Sd > 0.5, 13% of all proteins; above dashed line in Fig. 2b), significantly enriched annotations revealed major differences in the expression of proteins related to the categories 'Extracellular', 'Metabolism', 'Immunity' and 'Adhesion', suggesting pronounced cell line heterogeneity for a variety of biological processes (Fig. 2e, Supplementary Data 2).

**A discriminating 67-protein cell line signature.** Unsupervised hierarchical clustering based on the expression of 8,487 distinct proteins quantified in at least 10 of 30 cell lines resulted in three main groups (Fig. 3a). Group I cell lines comprised OVKATE, SNU119, JHOS4, OVCAR3, COV318, OVSAHO, KURAMOCHI, CAO3, OVCA433, JHOS2 and the cell line pair PEO1/4, which was derived from primary and recurrent tumours of the same patient and clustered together in the dendrogram. Much of group I consisted of cell lines that were previously reported to likely represent HGSOC cell lines based on features of their genomic profiles such as *TP53* mutational status, mutation frequency and DNA copy number alterations<sup>9</sup>. Group II contained OVCAR5, the cervical cancer cell line pair ME180/C13, the CCC cell line OVISe and the IGROV1 and SKOV3IP1 cell lines; the latter two lines have been previously described as hypermutated and 'unlikely HGSOC' cell lines<sup>9</sup>. In group III, the two IOSE cell

lines, IOSE-397 and IOSE-7576, clustered closely together suggesting that proteomic profiles accurately reflected tissue of origin and/or subtype-specific molecular signatures (Fig. 3a). The IOSE cell lines and the OvCa cell lines HEYA8, DOV13 and 59M had very high proteome similarities (Pearson correlation

0.85–0.92) and they grouped together in a sub-cluster of group III. This group also contained the TOV112D and A2780 cell line pairs, previously described as endometrioid in origin<sup>9,24,25</sup>. In each of their respective groups, the corresponding members of the isogenic cell line pairs were all located directly next to each other



4

in the clustering analysis. Noticeably absent from group I, and instead clustered within group III, based on their proteomic signature, were the TYKNU and 59M cell lines, both of which display high genomic similarity to HGSOc tumours<sup>9</sup>. The clustering of these two cell lines within group III indicated that certain discriminating features were detectable only at the protein level, and suggested the presence of two distinct HGSOc proteomes.

Independent of the unsupervised hierarchical clustering, we used a principal component analysis (PCA) on the basis of whole-proteome levels, which confirmed the presence of the three main cell line groups (Fig. 3b). Figure 3c depicts the proteins driving the segregation into the three groups: group I proteins included known HGSOc markers such as PAX8, MSLN, KRT7 and MUC16 (CA-125); group II, which contained the OVISe CCC cell line, showed the highest expression of AKR1C1 and HNF1B, known CCC-associated proteins<sup>22,26</sup>; while group III drivers included the mesenchymal proteins HMOX1, VIMENTIN, FN1 and ITGA5 (ref. 7), a protein detected in ~20% of serous OvCa tumours<sup>27</sup>.

With the aim of determining a small group of proteins with the strongest discriminating power between the groups, we used feature selection in combination with Support Vector Machines (SVMs) classification, as previously described<sup>28,29</sup>. This identified a set of 67 proteins, which included many known OvCa markers as well as novel ones (Supplementary Fig. 2, Supplementary Table 1).

#### Integrative analysis of HGSOc tissue and cell line proteomes.

Given the distinct proteome characteristics of the cell line groups, we reasoned that group I cell lines would most closely resemble HGSOc human cancer tissue. To first assess the quantitative robustness of our workflow for tumour tissue analysis, we analysed eight HGSOc tumour proteomes from five patients. This revealed that the proteomes were very similar between bilateral ovarian tumours from the same patient (mean Pearson correlation 0.95), compared with the inter-patient variation (Pearson correlation 0.72–0.87; Supplementary Fig. 3a). Consistently, tumours from the same patient tightly grouped together in the PCA (Fig. 4a). Adding the cell line proteomes to the PCA indicated a slight gravitation of the group I cell lines towards the HGSOc tumours and the FTECs in component 1 (Fig. 4b), suggesting underlying proteomic similarities with both the OvCa tumours and the primary FTECs. Of note, group III cell lines, which included the two IOSE cell lines, clustered furthest away from both the HGSOc and FTEC proteomes on component 1, but closest to the HGSOc tumours on component 2. Further analysis of component 2 revealed that it represented epithelial/mesenchymal protein levels (Supplementary Fig. 3b). Application of the discriminating 67-protein cell line signature to the PCA

and hierarchical clustering analyses then resulted in three groups and two main clusters (Fig. 4c,d, Supplementary Fig. 3c). The first cluster was exclusively composed of the group I cell lines, with seven of the eight HGSOc tumours and the primary FTEC samples. In particular, the cell lines COV318, KURAMOCHI and OVSAHO clustered closest to the tumour samples, whereas the PEO1 cell line clustered with the FTEC isolates. The second cluster contained groups II and III cell lines, organized into distinct sub-clusters, and HGSOc-5, which clustered closest to the group III cell lines.

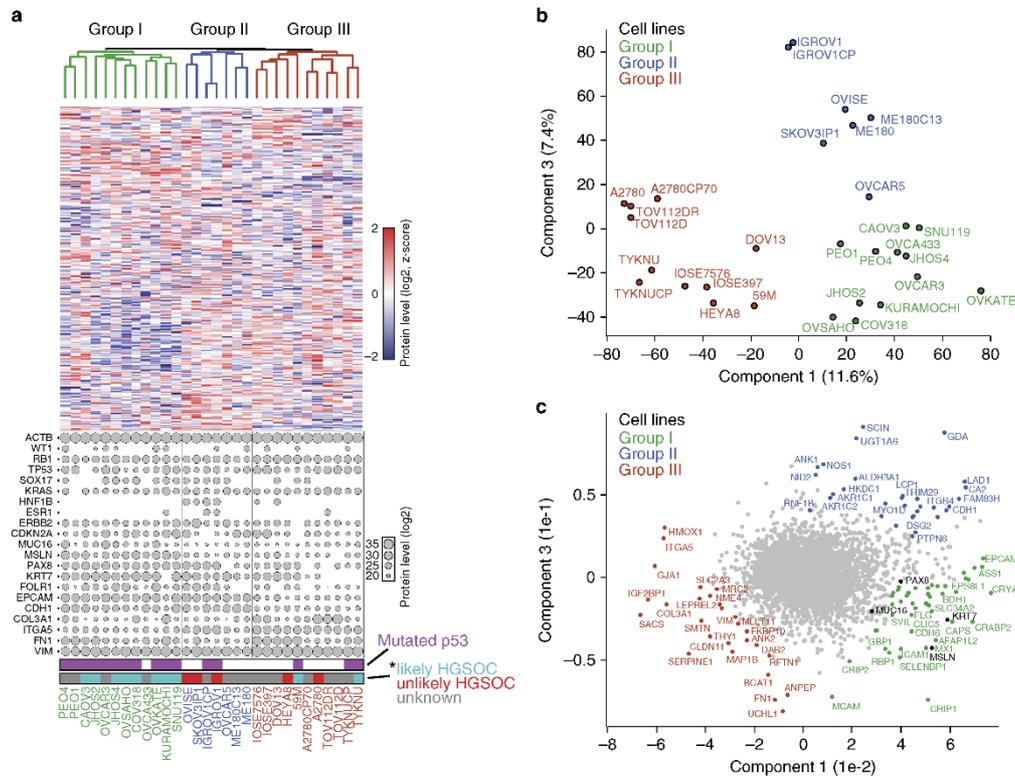
Known epithelial HGSOc markers such as MSLN, PAX8 and KRT7 were higher in the FTEC isolates, HGSOc tumours and group I cell lines compared with groups II and III cell lines (including IOSEs; Fig. 4d). Other proteins with a similar trend in expression included proteins such as the retinoic acid transporter CRABP2, a protein important in cancer-relevant arginine metabolism, ASS1 (ref. 30), and the p53 target gene CRYAB<sup>31</sup>. Conversely, groups II and III cell lines had higher expression of ITGA5, a protein detected in ~20% of serous OvCa tumours<sup>27</sup>. Interestingly, while HGSOc-5 expressed similar levels of the above-mentioned epithelial proteins, it had higher levels of ITGA5, indicating that ITGA5 contributed to its cluster location with group III cell lines.

Cluster-specific protein levels were validated for selected proteins: cluster 1, CRABP2 and PAX8; cluster 2, ITGA5. We confirmed cell type-specific expression of CRABP2 in human HGSOc tumours and normal OSE using immunofluorescence (Fig. 4e), as well as in normal FTECs, fallopian tube HGSOc and omentum HGSOc (Supplementary Fig. 4a). The absence of CRABP2 staining in normal OSE was consistent with its low protein levels in the proteomic profiles of the IOSE cell lines. Furthermore, we confirmed the levels of PAX8, CRABP2 and ITGA5 by western blot (Supplementary Fig. 4b).

#### The signature separates HGSOc from TCGA into two groups.

The above analyses established two very robust, distinct clusters based on the 67-protein signature, with cluster 1 containing the group I cell lines, which had an FTEC-type profile and harboured mostly epithelial proteins and cluster 2 containing the group III cell lines, which had an IOSE-type profile with mostly mesenchymal proteins. To validate our finding in a larger patient cohort, we made use of the publicly available data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC)<sup>32,33</sup>. The PNNL proteomics data set was downloaded and we applied our 67-protein signature to the 84 tumour proteomes alone, and then in conjunction with our own proteome data. This confirmed the presence of two main tumour clusters, which we denoted TCGA-A and TCGA-B, and which were enriched for group I/epithelial and group III/mesenchymal proteins, respectively (Fig. 5a, Supplementary Fig. 5). The group I cell

**Figure 2 | Proteomic analysis reveals proteome diversity across frequently used OvCa cell lines.** (a) Total number of quantified proteins in 26 OvCa cell lines, two immortalized ovarian surface epithelial cell lines (IOSE) and two cervical cancer cell lines ( $n = 3$  measurements for each cell line). A median depth of 7,812 quantified proteins across all the samples was obtained. (b) Quantification of proteins with low and high standard deviation (s.d.) across all the 30 cell lines identified constantly or variably expressed proteins. Protein level variability (s.d.,  $\log_{10}$ ) and protein intensity (sum of the intensity-based absolute quantification (iBAQ) values,  $\log_{10}$ , calculated by MaxQuant) were compared for each protein in all cell lines. An s.d. cut-off of 0.5 was used to identify proteins with the highest variability (13% of total proteins; s.d.  $> 0.5$ ; dashed line). Proteins frequently altered in OvCa are highlighted in blue. Proteins associated with ribosome biogenesis (Gene Ontology Biological Process), shown in green, display small differences in expression across cell lines. CDKN2A is depicted twice, representative of two different isoforms. (c) Known genomic alterations are captured at the protein level. The relative protein intensities (MaxLFQ) for proteins whose genes have known amplification events in OvCa cell lines are depicted in red. HNF1B, an ovarian CCC marker, is shown in orange. (d) Coverage of cancer-related KEGG pathways. KEGG annotations, including cancer-related pathways, were applied to the data. Percentages of pathway coverage in the indicated cancer-related KEGG categories are shown. Pathways involved in DNA replication and DNA repair, as well as metabolic annotations such as the citric acid cycle or fatty acid metabolism were almost completely covered ( $> 80\%$ ). Signalling pathways such as mTOR, p53 or TGF- $\beta$  signalling are largely covered by  $> 60\%$ . (e) Pathway enrichment analysis using Fisher's exact test (Benjamini-Hochberg false discovery rate  $< 2\%$ ) was performed for the proteins with the highest expression variability (proteins above dashed line in b).



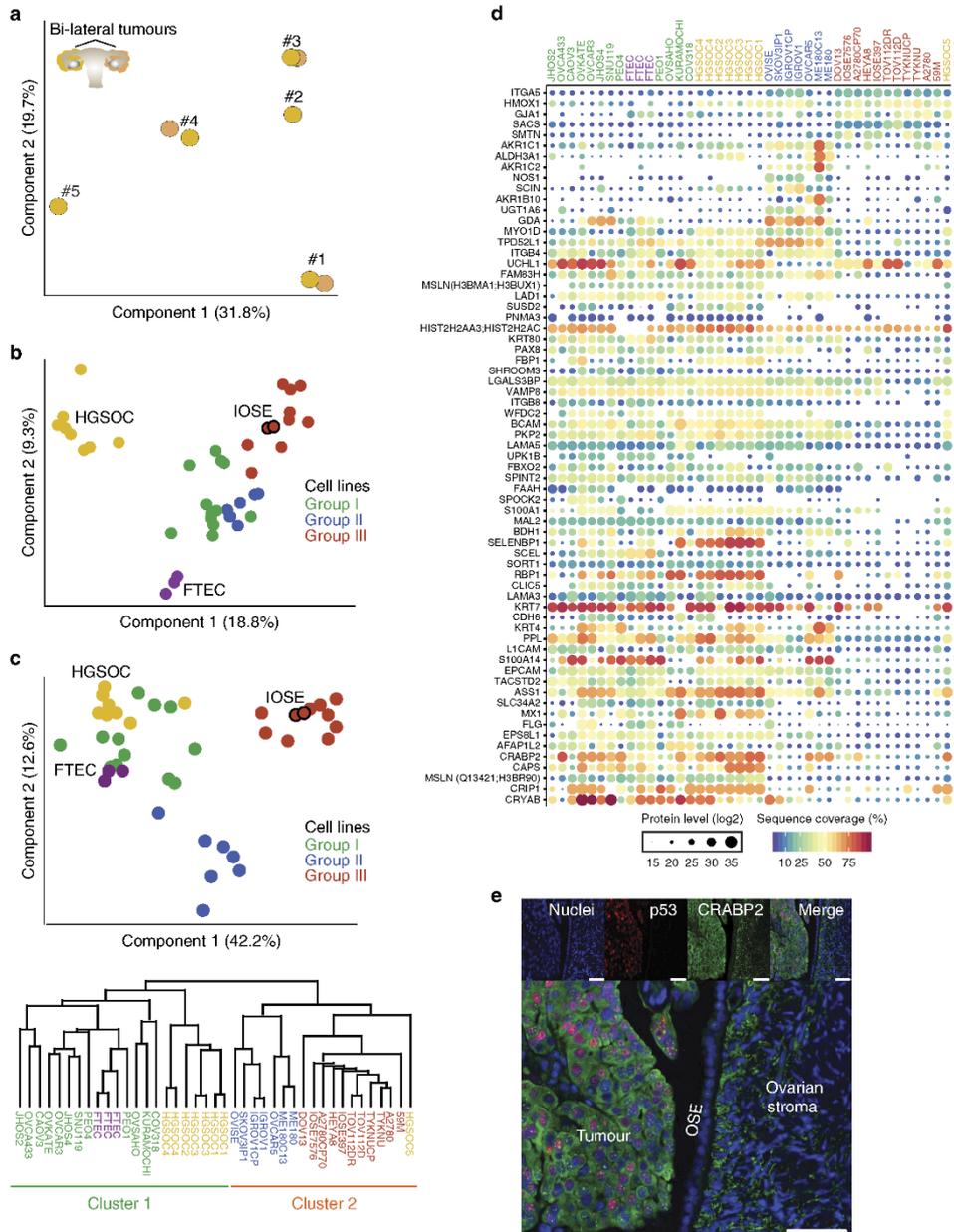
**Figure 3 | Proteomic clustering reveals three distinct cell line groups.** (a) Unsupervised hierarchical clustering was carried out on the basis of the relative expression of 8,487 proteins, delineating three major cell line groups (group I, green; group II, blue; group III, red). The relative expression levels of 15 frequently altered proteins in OvCa are plotted for each of the 30 cell lines. Levels of known OvCa-related proteins (VIM, FN1, EPCAM, CDH1 and CDKN2A) are also plotted. Relative protein levels are depicted by circle size. Relative ACTB levels are included for reference. 'likely HGSOc' (teal) and 'unlikely HGSOc' (red) refer to the descriptions of these cell lines based on their genomic profiles as reported by Domcke *et al.*<sup>9</sup>; 'unknown' cell lines in grey were not analysed in that study. (b) The presence of three major cell line groups was confirmed by principal component analysis (PCA) of the 30 cell lines. (c) Proteins driving the PCA separation. The driver proteins from the three groups are identified on the PCA. Known HGSOc proteins MUC16, PAX8, KRT7 and MSLN are highlighted in black.

lines and the FTECs clustered with the TCGA-A tumours, and the group III cell lines and the IOSEs clustered with the TCGA-B tumours. TCGA-A tumours showed higher expression of most group I/FTEC epithelial proteins such as KRT7, MSLN, CDH6, ASS1 and EPCAM, while TCGA-B tumours had higher expression of the group III/IOSE mesenchymal proteins ITGA5, HMOX1, SMTN and GJA1 (refs 7,34; Fig. 5b). Kaplan–Meier analysis of the overall survival of the two TCGA groups revealed that patients in the TCGA-B group had a significantly lower overall survival than those in the TCGA-A group ( $P=0.0048$ ; Fig. 5c). We then utilized a publically available messenger RNA (mRNA) data set that had identified two different primary ovarian cancer cell line clusters. In that study, the gene expression profiles of these clusters correlated with different survival outcome in patients<sup>35</sup>. Figure 5d shows that mRNA levels of our group I and group III proteins were higher in the good survival- and poor survival-associated sets, respectively. Thus, the two cell-type clusters established in our integrated cell line, primary cell and tumour proteome analyses are independently validated by the TCGA and Ince<sup>35</sup> data sets.

**Utility of the proteomic resource in functional assay design.** A cell line representative of human HGSOc should not only resemble its tumour of origin at the molecular level, but more importantly, also demonstrate functional utility in pre-clinical investigations. To exploit our global protein expression data, we first determined all proteins that were significantly different between the cluster 1 and cluster 2 cell lines identified in our study and then bioinformatically determined annotation terms that were statistically enriched in either cluster<sup>36</sup> (Fig. 6a,b). Interestingly, this revealed that the vitamin A and retinal binding pathways were the two most enriched pathways in the group I cell lines compared with groups II and III cell lines. In contrast, groups II and III cell lines expressed higher levels of proteins associated with cell proliferation (mitosis, DNA replication, Fig. 6a), in line with a previous study<sup>25</sup>. The volcano plot demonstrates the protein expression fold change between the two clusters. As expected, members of the 67-protein signature showed strong expression differences (Fig. 6b), and the levels of the retinoic acid pathway proteins differed by up to 30-fold. Mean expression of the five vitamin A

pathway proteins likewise showed a large difference between group I and groups II and III cell lines, however, no pronounced differences were present between FTECs, HGSOCS and group I cell lines (Supplementary Fig. 6a).

On the basis of the results of this pathway analysis, we hypothesized that group I cell lines would be more susceptible to the known CRABP2-mediated anti-oncogenic effect of all-*trans* retinoic acid (ATRA)<sup>37</sup>, the predominant physiological form of retinoic acid, than groups II and III cell lines. This was supported



by the inverse correlation between CRABP2 and FABP5, an intracellular lipid binding protein that binds to retinoic acid in the presence of a low CRABP2/FABP5 ratio<sup>38</sup>. The majority of groups II and III cell lines had a low CRABP2/FABP5 ratio (Fig. 6c). To test the hypothesized differential response to ATRA, we treated representative groups I, II and III cell lines with ATRA for 7 days and measured proliferation. Cell proliferation of group I cell lines was inhibited by  $26 \pm 2\%$  revealing that ATRA induced an anti-oncogenic effect in these cell lines (Fig. 6d). The only exception was the group I KURAMOCHI cell line; however, in these cells, ATRA treatment induced a more differentiated phenotype with long spindle-like protrusions (Supplementary Fig. 6b) consistent with the ATRA-induced differentiation previously reported in other cancers<sup>39,40</sup>. As predicted, ATRA had no inhibitory effect on cell proliferation in the tested groups II and III cell lines; in fact, ATRA-treated HEYA8 and 59M cells showed an increase in proliferation compared with their non-treated counterparts (Fig. 6d).

### Discussion

It is evident by now that the most commonly used HGSOc cell lines are among those with the least genomic similarity to HGSOc tumours<sup>9,11,12</sup>. Nevertheless, these 'bad' cell lines still closely recapitulate the metastatic distribution of human HGSOc in pre-clinical animal models<sup>11</sup>.

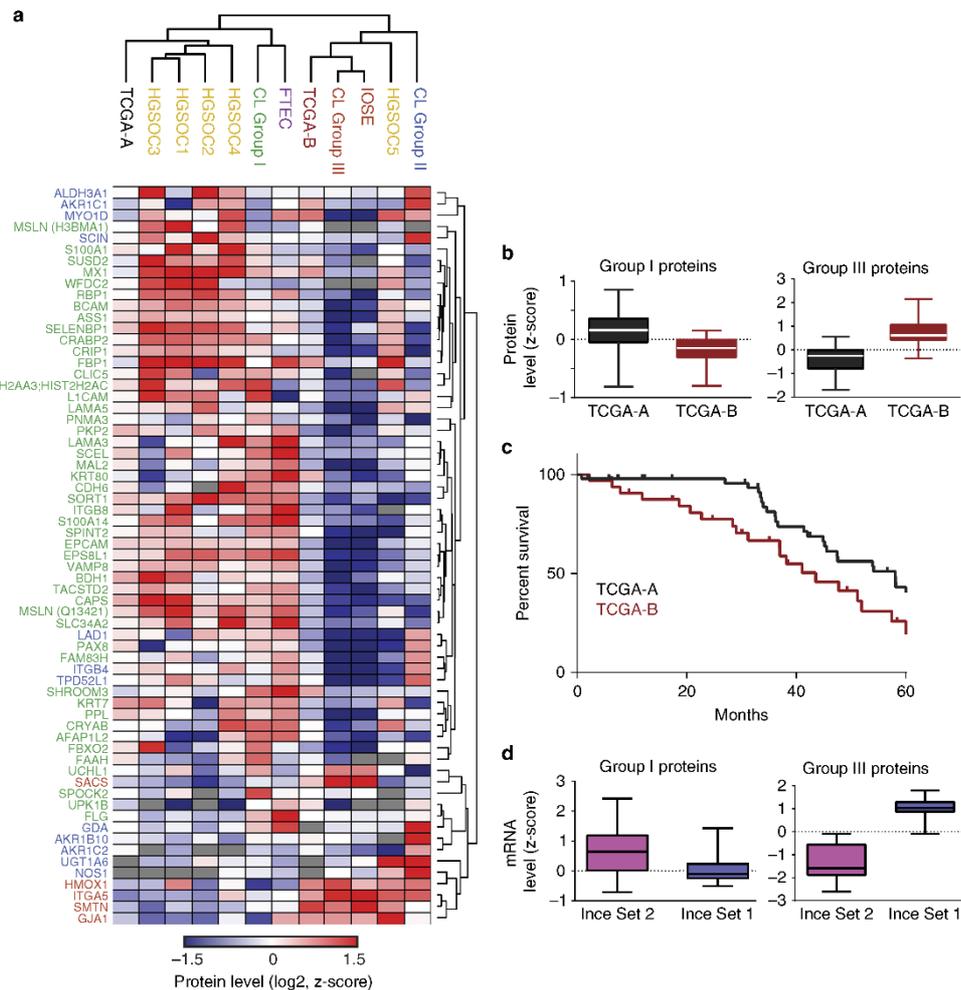
Here we used MS-based proteomics to offer a more phenotype-associated characterisation of OvCa cell lines than that of genomic and gene expression characterization alone. Our streamlined, single-run workflow allowed us to characterize a large number of cancer proteomes in a relatively short time (4 h). Importantly, the absence of fractionation did not unduly compromise depth of coverage of the proteome as we detected ~8,000 proteins in cell lines, primary cells and tumours, with a total number of ~11,000. To put this in perspective, a recent proteomic study of cancer tissues using a different platform consistently detected 2,000 proteins<sup>41</sup> and our own laboratory previously achieved the numbers detected here only after using fractionation techniques, which greatly add to measurement time and sample requirements<sup>29</sup>. On the basis of the broad and quantitative coverage, the 30 cell lines unambiguously separated into three specific cell line groups: group I, containing a number of HGSOc cell lines, expressed higher levels of many known epithelial HGSOc proteins; group II, containing the CCC cell line OVISE, expressed a number of CCC-associated proteins<sup>26,27</sup>; and group III, containing the IOSE cell lines, expressed relatively lower levels of known epithelial HGSOc markers and higher levels of several mesenchymal proteins. A novel, maximally discriminating 67-protein signature faithfully segregated the groups and contained interesting group-enriched proteins in

addition to known OvCa markers. Of note, while group II cell lines were all cultured in DMEM, the culture media for group I and group III cell lines, both of which contain HGSOc cell lines, consisted of a variety of different media, suggesting that the signature is not dependent on the culture media.

We integrated the cell line proteomes with our primary FTEC and HGSOc tumour data and, in a second step, with the CPTAC proteomic data from 84 patients. Application of the 67-protein cell line signature divided the entire data set into two core clusters, clearly placing the primary FTEC isolates in the epithelial cluster and the IOSEs in the other, more mesenchymal, cluster. The FTEC cluster expressed high levels of known HGSOc proteins such as MUC16 (CA-125), PAX8 and MSLN. It also revealed novel markers for FTEC-derived HGSOc cell lines such as CRABP2 and ASS1, which are highly expressed in serous OvCa compared with the clear cell, endometrioid and mucinous subtypes<sup>30,42</sup>; CRYAB, a p53 target gene<sup>31</sup>, which is associated with patient outcome in serous, but not non-serous, OvCa<sup>43</sup>; and CAPS and MX1, which have not been previously described as potential markers of HGSOc (see summary in Supplementary Table 1). The defining feature of the IOSE cluster was the high expression of a small set of mesenchymal proteins: GJA1, HMOX1, ITGA5, SMTN and SACS<sup>7,34</sup>. GJA1 facilitates cell adhesion, invasion and metastasis in a number of other cancers<sup>44-46</sup>; its high expression in the IOSE cluster indicates that it may play a similar role in these cell lines and tumours.  $\alpha_5$ -integrin (ITGA5), which we previously reported to be an important mediator of early OvCa metastasis<sup>47</sup>, was a strong discriminator between the FTEC and IOSE clusters, with high expression in the latter.  $\alpha_5$ -integrin is regulated by the epithelial differentiation marker E-cadherin (CDH1; ref. 48) and the absence of CDH1 is a predictor of poor survival in OvCa patients<sup>49</sup>. Although CDH1 was not in the discriminating 67-protein signature, its expression was drastically lower in group III cell lines compared to group I cell lines (Fig. 3a). In addition, OSE lacks, or inconsistently expresses, CDH1 (ref. 50). This suggests that the ITGA5/CDH1 axis may be an important distinguishing feature of our newly defined HGSOc sub-groups.

Group II cell lines were the least HGSOc-like cell lines in this study and, interestingly, the ITGA5/CDH1 axis does not appear to apply to this group. In general, group II lines express high levels of epithelial proteins such as CDH1 and EPCAM, which contributes to their clustering with the group I (epithelial) cell lines in component 1 of the cell line PCA (Fig. 3b,c). However, they also express high levels of ITGA5, which may contribute to their clustering with group III in the integrated tumour and cell line analysis (Fig. 4d). At least four of the seven cell lines in group II contain an ARID1A mutation<sup>9</sup>; previous studies have reported an association between these mutations and the transformation of

**Figure 4 | Integrative analysis of HGSOc tissue and cell line proteomes.** (a) Proteomic clustering of eight HGSOc tissue specimens. PCA was performed on the ovarian tumour specimens ( $n=8$ ) based on the proteomic profiles to evaluate inter-patient heterogeneity and intra-patient homogeneity. Component 1 and component 2 account for 51.5% of the total data variation. (b) PCA clustering of cell lines ( $n=30$ ), FTEC ( $n=3$ ) and HGSOc tissue proteomes ( $n=8$ ). Groups I (green), II (blue) and III (red) are as defined in Fig. 3. The juxtaposition of the IOSE cell lines, red circles outlined in black, relative to the rest of the group III cell lines is indicated. Component 1 and component 2 account for 28.1% of the total data variation. (c) PCA segregation and clustering of all samples based on the 67-protein signature. The juxtaposition of the IOSE cell lines, red circles outlined in black, relative to the rest of the group III cell lines is indicated. Component 1 and component 2 account for 54.8% of the total data variation. The dendrogram below the PCA summarizes the hierarchical clustering analysis of all samples based on the 67-protein signature. Two main clusters were obtained based on this signature (detailed in Supplementary Fig. 3c). (d) Relative levels of the 67 proteins used for PCA clustering in c. Relative protein levels (MaxLFQ intensities, log<sub>2</sub>) are depicted by circle size. Colours indicate protein sequence coverage per sample. MSLN is depicted twice, representative of two different isoforms (shown in parentheses). (e) Immunofluorescence staining for CRABP2 and p53 in formalin-fixed paraffin-embedded (FFPE) sections of normal OSE and ovary HGSOc tumour. FFPE sections were stained with an anti-CRABP2 antibody and an anti-p53 antibody, and detected with Alexa Fluor 488- and Alexa Fluor 647-labelled antibodies, respectively. Merged images on the bottom show invasive HGSOc, normal OSE and normal ovarian stroma in the same frame. Scale bar, 50  $\mu$ m.

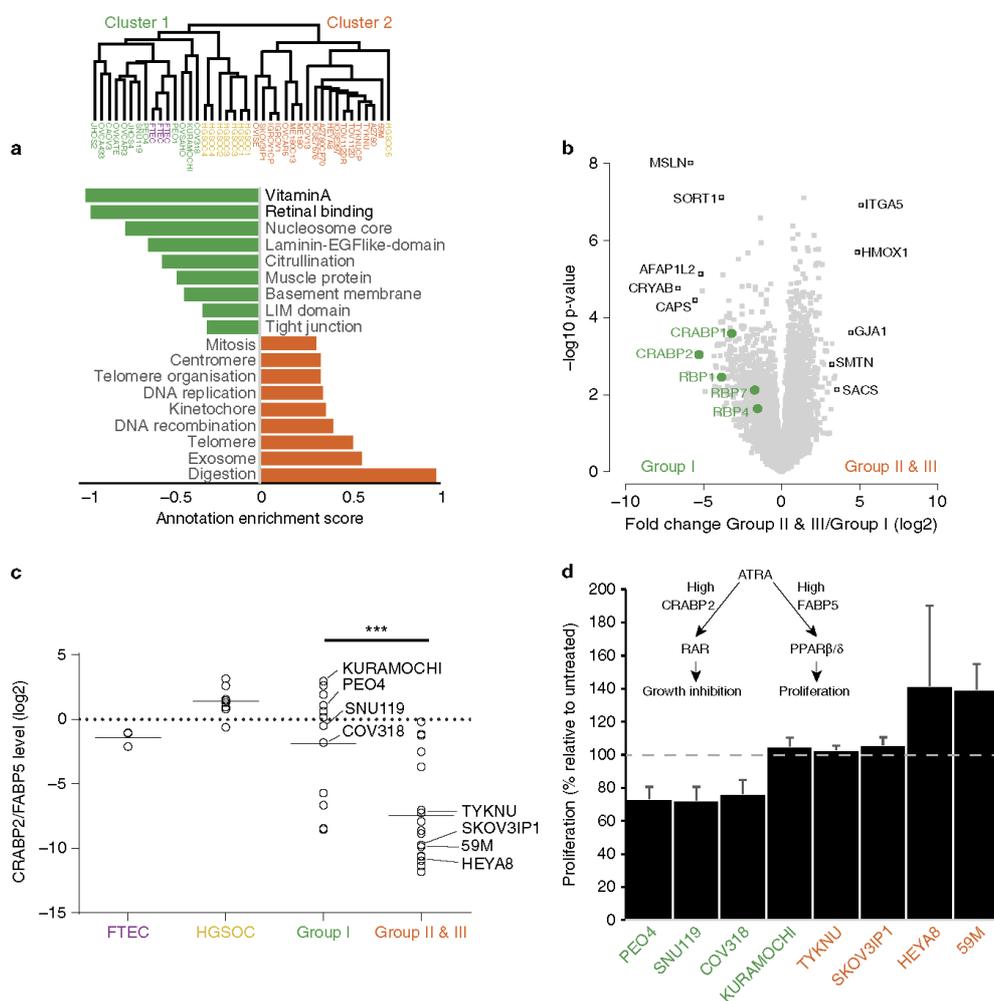


**Figure 5 | Clustering of the TCGA tumours based on the 67-protein cell line signature.** Hierarchical clustering summary of 84 TCGA tumours with the cell lines, FTECs, IOSEs and HGSOC-1 to -5. **(a)** Hierarchical clustering based on the 67-protein signature was applied to the publicly available proteomic profiles of 84 TCGA tumours, and our own data set comprises the cell lines, FTECs, IOSEs and HGSOC-1 to -5 (Supplementary Fig. 5). The average protein levels for each sample group are shown. **(b)** Group I and group III proteins are higher in TCGA-A and TCGA-B tumours, respectively. Z-scored protein levels of group I and group III proteins were plotted as box plots for TCGA-A and TCGA-B tumours. **(c)** TCGA-B tumours (maroon line) are associated with a poorer overall survival compared with TCGA-A tumours (black line). Kaplan-Meier overall survival curves were plotted for patients in TCGA-A or TCGA-B. Survival associated with TCGA-B was significantly lower than that associated with TCGA-A (43.5 months versus 58 months, Mantel-Cox  $P$  value = 0.0048). **(d)** The mRNA levels of group I and group III proteins were analysed in a publicly available mRNA data set containing two different primary ovarian cancer cell line clusters<sup>35</sup>. Box plots show higher median mRNA levels of group I proteins in the favourable (Ince Set 2) and relatively lower in the unfavourable (Ince Set 1) survival-associated sets, while the converse is true for group III proteins.

endometriosis into ovarian CCC<sup>51</sup>. In addition, the high levels of CCC proteins, AKR1C1 and HNF1B<sup>22,26</sup>, that are low or undetected in the other cell line groups, is further evidence of a potential CCC background for most of group II cell lines.

A 219 microRNA (miRNA)-associated mesenchymal gene signature separated OvCa cases in different data sets into two subtypes: an integrated epithelial subtype and an integrated

mesenchymal subtype<sup>34</sup>. In line with this, the proteomes of group I HGSOC cell lines, which clustered with the FTECs, contained a mix of epithelial and mesenchymal proteins, while those of group III, which contained the IOSE cell lines, were predominantly mesenchymal (high VIM, low CDH1 and EPCAM, Fig. 3a). This is also consistent with data from a study showing that the gene expression profile of the mesenchymal HGSOC subtype is similar



**Figure 6 | Proteomic profiles predict functional cell line properties.** (a) Pairwise comparison of enriched annotations for group I (cluster 1) and groups II and III (cluster 2) cell lines. Pathway enrichment analysis was calculated on the basis of the protein expression fold change between group I and groups II and III cell lines. Green and orange bars denote the strongest enriched pathways (Benjamini-Hochberg FDR < 0.02) in group I and groups II and III cell lines, respectively. Annotation enrichment position score, between  $-1$  and  $1$ , indicates the centre of the protein distribution for each significant category, relative to the overall distribution of values. Dendrogram shows sample clustering, based on the 67-protein signature. (b) Volcano plot of the pairwise comparison between group I and groups II and III cell line proteomes. Expression fold changes ( $t$ -test difference,  $\log_2$ ) were calculated and plotted against the  $t$ -test  $P$  value ( $-\log_{10}$ ). Vitamin A pathway-associated proteins are highlighted in green. Their position on the left side of the plot indicates their higher expression in group I cell lines ( $P = 0.003$ ). Strongest outlier proteins for both groups are marked in black. (c) Relative protein levels of the retinoic acid transporter proteins CRABP2 and FABP5 were compared between FTECs, HGSOC, group I cell lines and groups II and III cell lines. The CRABP2/FABP5 ratios are shown. Groups II and III cell lines show significantly lower levels of CRABP2 expression relative to FABP5 (two-sided  $t$ -test,  $P$  value < 0.001). The cell lines used in subsequent proliferation experiments (d) are indicated. (d) ATRA treatment induces growth arrest or differentiation in group I cell lines. Group I (green) and groups II and III (orange) cell lines were treated with  $7 \mu\text{M}$  ATRA daily for 7 days. On day 8, MTT assays were performed. Results are plotted as the proliferation rate of ATRA-treated cells relative to that of untreated cells. Values < 100% indicate growth arrest; values > 100% indicate growth promotion. Error bars represent standard errors for two to four replicates per cell line. A simplistic model of the ATRA pathway is shown. ATRA, all-trans retinoic acid; PPAR, peroxisome proliferator-activated receptor; RAR, retinoic acid receptor.

to that of normal ovarian tissue<sup>52</sup>. Our separation of HGSOc tumours into two groups on the basis of the 67-protein signature was solidly confirmed in a validation set from the CPTAC as it grouped their 84 HGSOc tumours into two main clusters, reflective of group I and group III cell line proteomes, also providing further evidence for a dualistic precursor model of HGSOc. Importantly, there is a clear survival difference between the two TCGA groups, with the TCGA-B (mesenchymal) patients demonstrating a significantly worse overall survival than that of the TCGA-A (epithelial) patients. In addition, the juxtaposition of our tumour specimen HGSOc-5 with group III cell lines, the IOSEs and TCGA-B in the integrated analysis strongly indicates that this tumour represents the mesenchymal HGSOc subtype<sup>7,52</sup>.

Clustering of group I cell lines and HGSOc-1 to -4 tumours with the FTECs provides support for the fallopian tube epithelium as their cell of origin, while the clustering of group III cell lines and HGSOc-5 with the IOSEs, and their lower expression of epithelial proteins, suggests that they may be OSE-derived. This suggests the potential presence of an inherent level of stratification in HGSOc tumours based on their protein expression and cell of origin. Regarding the precursor cell of HGSOc, there are convincing arguments for both FTECs and OSE<sup>3,4,53,54</sup>. Supporters of the 'fallopian tube theory' argue that the HGSOc protein PAX8 is a marker of the tubal epithelium<sup>55</sup> and that STICs in the fallopian tube fimbria of BRCA1-mutation carriers<sup>4</sup> and HGSOc patients<sup>3</sup> represent part of the serous carcinogenic sequence; supporters of the 'ovarian theory' argue that OSE expresses multiple stem cell markers<sup>56</sup>, and that OSE-lined inclusion cysts can undergo tubal metaplasia<sup>56,57</sup> followed by transition to carcinoma<sup>58</sup>. There is support for both hypotheses from animal models<sup>54,59</sup>. However, aside from the investigation of STICs, distinct biologic, clinical or molecular features capable of categorical differentiation between the two have not yet been identified<sup>60</sup>. Our proteomics results now suggest a potential new level of simple stratification in HGSOc into mesenchymal and epithelial subtype HGSOc. However, this is, as yet, only hypothesis-generating; more detailed molecular, immuno-histochemical and clinico-pathologic studies will be necessary to substantiate or reject this concept.

The underlying differences between the cell lines in the two different clusters were also reflected in the biological pathways associated with their respective proteins. The vitamin A and retinol pathways were highly enriched in group I cell lines, with more than 10-fold expression level differences. Retinoic acid has been used successfully in the treatment of acute promyelocytic leukemia and neuroblastoma and its limited success in other cancers may be due to different cellular responses. On the basis of the proteomic profiling alone, we correctly predicted that group I cell lines would be sensitive to the anti-proliferative, differentiating effects of ATRA, while groups II and III cell lines would be either ATRA-resistant, or responsive to its pro-oncogenic effects. Future investigation into these different mechanisms of action of, and cellular response to, ATRA in group I and group III HGSOc cell lines may inform the investigation of retinoic acid in OvCa.

In summary, the integrated level of cell line proteomic profiling introduced here provides the OvCa research community with an additional resource to select the most appropriate model for their research. The described 67-protein signature may shed more light on the cell of origin and respective driver proteins of HGSOc and contribute to the further investigation of important clinical problems such as chemotherapy resistance. Apart from opening the door for systematic and routine interrogation of proteome-wide differences in cancer models, our streamlined and high-sensitivity proteomics workflow will be especially attractive in *in vivo* contexts, where only small numbers of cells are available.

## Methods

**Patient samples.** All patient samples were collected at the University of Chicago Medical Center, with approval from the Institutional Review Board. All the patients provided informed consent.

**Frozen tumour samples.** Ovarian tumours from five chemo-naïve patients with serous-papillary high-grade, FIGO Stage IIIB/C, OvCa were collected by EL under an IRB-approved protocol at the University of Chicago during the primary debulking surgery or laparoscopy before neo-adjuvant treatment. In some cases, tumours were collected from both ovaries ( $n = 3$  patients). Tumours were immediately snap frozen and stored at  $-80^{\circ}\text{C}$  until sample processing for MS analysis. Serous histology was confirmed by two gynaecologic pathologists (A.M., R.R.L.).

**Primary FTEC isolation.** Fallopian tubes were removed from three patients with benign gynaecological conditions not affecting the fallopian tube. Primary fallopian tube secretory epithelial cells (FTECs) were isolated<sup>61</sup> and cultured<sup>62</sup> as previously described.

**Cell lines.** Thirty cell lines were included in this study; their sources and respective media are detailed in Supplementary Data 3. All the cell lines were genotyped to confirm their authenticity; cell lines were authenticated by using the commercial service CellCheck (IDEXX Bioresearch). The samples were confirmed to be of human origin and no mammalian inter-species contamination was detected. The alleles for nine short tandem repeat markers were determined and the results were compared with the profiles from DSMZ, ATCC, JCRB and RIKEN short tandem repeat databases. All the cell lines were mycoplasma-negative. The cells were grown under recommended culture conditions and the samples were collected from three consecutive passages for  $n = 3$  replicates for each line.

**Sample preparation for MS analysis.** Cell lysis was performed in lysis buffer (4% SDS, 10 mM HEPES pH 8.0) at  $99^{\circ}\text{C}$  for 10 min and by 15 min sonication (level 5, Bioruptor, Diagenode). HGSOc tissues were first homogenized in lysis buffer using an Ultra Turbax blender. Proteins in the lysate were reduced with 10 mM DTT for 30 min and alkylated with 55 mM iodoacetamide for an additional 30 min. Remaining SDS detergent was removed by acetone precipitation. Briefly, acetone ( $-20^{\circ}\text{C}$ ) was added to 100  $\mu\text{g}$  of proteins to a final concentration of 80% v/v and the proteins were precipitated overnight at  $-20^{\circ}\text{C}$ . After centrifugation (15 min,  $4^{\circ}\text{C}$ , 16,000 g), the detergent-containing supernatant was removed and the protein pellet was washed with 80% acetone ( $-20^{\circ}\text{C}$ ). Protein pellets were then resolved in 100  $\mu\text{l}$  6 M urea/2 M thiourea (in 10 mM HEPES pH 8.0) and digested with 1  $\mu\text{g}$  of LysC for 3 h at room temperature. After adding four volumes of 50 mM ammonium bicarbonate, 1  $\mu\text{g}$  trypsin was added and tryptic digestion carried out overnight. The next day, digestion was stopped by adding 1% TFA. Peptides were finally desalted on C18 StageTips and kept at  $-20^{\circ}\text{C}$  until MS analysis.

**Liquid chromatography-MS analysis.** MS analysis was performed using Quadrupole Orbitrap mass spectrometers<sup>63,64</sup> (Q Exactive and Q Exactive HF, Thermo Fisher Scientific, Rockford, IL, USA) coupled to an EASY-nLC 1000 HPLC system (Thermo Fisher Scientific) via a nano electrospray source. Columns (75  $\mu\text{m}$  inner diameter, 50 cm length) were in-house packed with 1.9  $\mu\text{m}$  C<sub>18</sub> particles (Dr Maisch GmbH, Germany). Peptides were separated over a 250 min gradient from 2% to 60% (5 min to 5%, 180 min to 25%, 45 min to 35%, 20 min to 60%) in buffer B (80% acetonitrile, 0.5% formic acid) at 200 nl min<sup>-1</sup>. The column temperature was constantly set to  $50^{\circ}\text{C}$  by using an in-house-made column oven. The survey scans (300 to 1,650  $m/z$ ) were acquired with a resolution of 70,000 (60,000 for Q Exactive HF), at  $m/z$  200. A top-five method was used to select up to the five most abundant precursor ions with a charge  $\geq 2$ . Selected precursor ions were subjected to high-energy collisional dissociation fragmentation at a normalized collision energy of 25 (27 for Q Exactive HF), an isolation window of 2.2 Th (1.4 Th for Q Exactive HF) and a resolution of 17,500 at  $m/z$  200 (15,000 for Q Exactive HF). For survey scans, ion injection times were set to 20 ms (target value 3E6) and 120 ms (target value 1E5) for MS/MS scans. Dynamic exclusion of sequenced peptides was set to 30 s. Data were acquired using Xcalibur software (Thermo Scientific).

**MS data analysis.** MS raw files were analysed with MaxQuant software<sup>18</sup> (version 1.5.0.38). MS/MS-based peptide identification was carried out with the Andromeda search engine in MaxQuant<sup>17</sup>. Briefly, Andromeda uses a target-decoy approach to identify peptides and proteins at an FDR  $< 1\%$ . As a forward database, the human UniProtKB database (Oct 2014) was used. A reverse database for the decoy search was generated automatically in MaxQuant. Enzyme specificity was set to 'Trypsin', and a minimum number of seven amino acids were required for peptide identification. Default settings were used for variable and fixed modifications (variable modification, acetylation (N terminus) and methionine oxidation; fixed modification, carbamidomethylation). Proteins and protein isoforms that could not be discriminated by unique peptides were grouped into protein groups<sup>18</sup>. For label-free protein quantification, the MaxLFQ algorithm was used as part of the MaxQuant environment<sup>19</sup>. Briefly, quantitative information was retrieved on the

basis of high-resolution three-dimensional peptide profiles in mass-to-charge, retention time and intensity space. The algorithm first calculated pairwise protein ratios by taking the median of all pairwise peptide ratios per protein. Only shared identical peptides were considered for each pairwise comparison. A minimum number of one ratio count was required for each pairwise comparison. To retrieve quantitative information for all possible sample comparisons, a least-squares analysis was used to reconstruct the relative abundance profile for each protein. This step preserved the total summed intensity for a protein over all the samples. To maximize the number of quantification events across samples, we enabled the 'Match Between Runs' option in MaxQuant, which allowed the quantification of high-resolution MS1 features that were not identified in each single measurement.

Isobaric tag for relative and absolute quantification (iTRAQ)-based TCGA proteome raw files (PNNL study) generated by the Clinical Proteomic Tumor Analysis Consortium (NCI/NIH) were downloaded from the CPTAC data portal (<https://cptac-data-portal.georgetown.edu/cptacPublic/>) and analysed with MaxQuant. Logarithmic reporter intensities were normalized against the control reporter channel (channel 117, pooled sample of 84 TCGA ovarian tumour tissue samples) and each sample median normalized before data analysis.

**Statistical analysis.** All statistical and bioinformatics analyses were performed using the freely available software Perseus<sup>55</sup> (as part of the MaxQuant environment) or the R framework<sup>56</sup>. Proteins identified only by site modification or found in the decoy reverse database were not considered for data analysis. For the analysis of LQ cell line data, we first filtered out proteins that were only quantified in one of three replicates and took the average expression per cell line for the remaining protein quantifications. To calculate protein level variability across cell lines (standard deviation of MaxLQ intensity, Fig. 2b) we required a minimum of 10 quantified values. Pathway enrichment analysis for categorical data (Fig. 2e) was performed based on a Fisher's exact test with a Benjamini-Hochberg FDR threshold of 0.02. GOBP, GOCC, KEGG and Uniprot Keyword annotations were used for enrichment analysis and required a minimum category size of at least four proteins. For numerical data such as protein expression fold change between two groups (Fig. 6b) we used a one-dimensional enrichment analysis<sup>36</sup> with a Benjamini-Hochberg FDR threshold of 0.02. For hierarchical clustering of cell lines (Fig. 3a), a minimum of 10 valid values (one-third of all samples) was required. Missing values were imputed on the basis of a normal distribution (width = 0.15, down-shift = 1.8). MaxLQ intensities were first z-scored and the samples clustered according to Spearman rank correlations as a distance measure for column and row clustering.

We used the classification framework implemented in Perseus and, in particular, SVMs combined with feature selection methods to identify a subset of proteins that act as strong discriminators between the three cell line groups. This should allow deeper insight into the underlying biological processes, while keeping the selected subset small enough to allow follow-up studies (Supplementary Fig. 2, Fig. 4c,d).

The SVMs implementation in Perseus is an adaptation of the well-established and commonly used LIBSVM library<sup>67</sup>, which uses sequential minimal optimization to solve the quadratic problems during model training. We used a One-versus-All implementation of SVMs classification, which resulted in three distinct ranked lists of proteins—one for each of the cell line groups. Feature selection was embedded in a cross-validation procedure to avoid overfitting and random sampling using 85% of the data for training and the other 15% for testing; this was repeated 250 times. Proteins were ranked by the *P* value computed using a modified test statistic<sup>68</sup> with an  $s_0 = 4$  parameter. For the final list of 67 discriminating proteins, the top-ranked ones from each of the three ranked lists that offered a good tradeoff between minimal sets and the smallest error rates were combined (Group I, 53 proteins; Group II, 10 proteins; Group III, 10 proteins; with six proteins overlapping). This feature selection method resembles that of a domain expert (biologist) selecting a small subset of proteins that can conveniently be followed up. Furthermore, the required large differences between the groups should allow more robust classification in a clinical setup. The method is conceptually analogous to the standard analysis of variance test and, in fact, performing analysis of variance identifies the selected proteins as significantly differentially expressed. However, using feature selection avoids the need for an FDR cutoff and allows for selecting smaller subsets.

For PCA of cell lines, primary cells and HGSOc tissues (Fig. 4b), we first filtered out the roughly 200 most abundant plasma proteins<sup>69</sup>. A minimum number of 30 valid values out of 41 was required, resulting in 6,649 proteins. Missing values were imputed as described above.

For the clustering of iTRAQ and label-free data (Fig. 5a,b, Supplementary Fig. 5), z-scoring was performed group-wise for iTRAQ and label-free data.

For pairwise comparison of proteomes (Fig. 6b), a two-sided *t*-test statistic was used, including a permutation-based FDR of 5% and an  $s_0$  value<sup>68</sup> of 2.

**ATRA treatment and proliferation assay.** The following cell lines were treated (or untreated) with 7  $\mu$ M ATRA daily for 7 days in 96-well plates: SNU119, TYKNU, KURAMOCHI, HEYA8, PEO4, 59M, SKOV3IP1 and COV318. On day 8, MTT assays were carried out. Proliferation rates were calculated for ATRA-treated and untreated cells and plotted as percentage relative to untreated cells.

**Western blot.** The cells were grown in six-well plates and treated with 7  $\mu$ M ATRA daily for 7 days. The cell lysates were collected in RIPA buffer and 20  $\mu$ g protein was electrophoresed on 4–20% resolving gels. The following antibodies were used in western blots: anti-rabbit ITGA5 (1/1,000 dilution, #sc-10729, Santa Cruz Biotechnology, Dallas, TX, USA), anti-rabbit PAX8 (1/1,000 dilution, #9857, Cell Signaling Technology, Danvers, MA, USA), anti-rabbit CRABP2 (1/3,000 dilution, #PA5-27451, Thermo Fisher Scientific) and anti-rabbit GAPDH (1/1,000 dilution, #2118, Cell Signaling Technology).

**Immunofluorescence.** Formalin-fixed paraffin-embedded HGSOc and normal tissue sections (5  $\mu$ m) were deparaffinized in xylene and rehydrated in graded alcohol solutions. One FFPE section of normal (OSE and FTTC) or tumour (omental tumour and fallopian tube tumour) tissue was used for immunofluorescence. Antigen retrieval was carried out in 0.01 M sodium citrate buffer with 0.05% Tween 20 (pH 6.0) for 30 min at 95 °C. Briefly, slides were washed 3  $\times$  with PBS, blocked with 10% goat serum in phosphate-buffered saline/0.05% Tween (PBST), and incubated overnight with anti-rabbit CRABP2 antibody (1:200 dilution, #PA5-27451, Thermo Fisher Scientific) or anti-mouse p53 (1:200 dilution, panoptic, #OP42, Millipore, Billerica, MA, USA). Following five PBST washes, sections were incubated with Alexa Fluor 488 goat anti-rabbit IgG (1:500 dilution, #A11008, Thermo Fisher Scientific) or Alexa Fluor 647 goat anti-mouse IgG (1:500 dilution, #A21235, Thermo Fisher Scientific) and Hoechst (1/200 dilution, #H1399, Thermo Fisher Scientific) for 1 h, washed 5  $\times$  with PBST, 2  $\times$  with PBS, and mounted with Prolong Gold Antifade Reagent (P36934, Molecular Probes, Thermo Fisher Scientific). The images were obtained with a Zeiss 510 LSM confocal microscope, using the LSM 510 software. As a negative control, the primary antibody was omitted for one section in each set of samples.

**Data availability.** The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org/cgi/GetDataset>) via the PRIDE<sup>70</sup> partner repository with the data set identifier PXD003668. The data set is also accessible via the user-friendly MaxQB database (<http://maxqb.biochem.mpg.de/mxdb/project/list>). All other data supporting the findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request.

## References

- Koonings, P. P., Campbell, K., Mishell, Jr D. R. & Grimes, D. A. Relative frequency of primary ovarian neoplasms: a 10-year review. *Obstet. Gynecol.* **74**, 921–926 (1989).
- Callahan, M. J. *et al.* Primary fallopian tube malignancies in BRCA-positive women undergoing surgery for ovarian cancer risk reduction. *J. Clin. Oncol.* **25**, 3985–3990 (2007).
- Kindelberger, D. W. *et al.* Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: evidence for a causal relationship. *Am. J. Surg. Pathol.* **31**, 161–169 (2007).
- Piek, J. M. *et al.* Dysplastic changes in prophylactically removed fallopian tubes of women predisposed to developing ovarian cancer. *J. Pathol.* **195**, 451–456 (2001).
- Bowtell, D. D. *et al.* Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat. Rev. Cancer* **15**, 668–679 (2015).
- Marks, J. R. *et al.* Overexpression and mutation of p53 in epithelial ovarian cancer. *Cancer Res.* **51**, 2979–2984 (1991).
- Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
- Tothill, R. W. *et al.* Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin. Cancer Res.* **14**, 5198–5208 (2008).
- Domcke, S., Sinha, R., Levine, D. A., Sander, C. & Schultz, N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat. Commun.* **4**, 2126 (2013).
- Barretina, J. *et al.* The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
- Mitra, A. K. *et al.* *In vivo* tumor growth of high-grade serous ovarian cancer cell lines. *Gynecol. Oncol.* **138**, 372–377 (2015).
- Elias, K. M. *et al.* Beyond genomics: critical evaluation of cell line utility for ovarian cancer research. *Gynecol. Oncol.* **139**, 97–103 (2015).
- Beck, M. *et al.* The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, 549 (2011).
- Burkard, T. R. *et al.* Initial characterization of the human central proteome. *BMC Syst. Biol.* **5**, 17 (2011).
- Mann, M., Kulak, N. A., Nagaraj, N. & Cox, J. The coming age of complete, accurate, and ubiquitous proteomes. *Mol. Cell* **49**, 583–590 (2013).
- Thakur, S. S. *et al.* Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol. Cell. Proteomics* **10**, M110 003699 (2011).

17. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
18. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
19. 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).
20. Schwanhaussner, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
21. Ross, J. S. *et al.* Comprehensive genomic profiling of epithelial ovarian cancer by next generation sequencing-based diagnostic assay reveals new routes to targeted therapies. *Gynecol. Oncol.* **130**, 554–559 (2013).
22. Tsuchiya, A. *et al.* Expression profiling in ovarian clear cell carcinoma: identification of hepatocyte nuclear factor-1 beta as a molecular marker and a possible molecular target for therapy of ovarian clear cell carcinoma. *Am. J. Pathol.* **163**, 2503–2512 (2003).
23. Geiger, T., Cox, J. & Mann, M. Proteomic changes resulting from gene copy number variations in cancer cells. *PLoS Genet.* **6**, e1001090 (2010).
24. Anglesio, M. S. *et al.* Type-specific cell line models for type-specific ovarian cancer research. *PLoS ONE* **8**, e72162 (2013).
25. Beaufort, C. M. *et al.* Ovarian Cancer Cell Line Panel (OCCP): clinical importance of *in vitro* morphological subtypes. *PLoS ONE* **9**, e103988 (2014).
26. Schaner, M. E. *et al.* Gene expression patterns in ovarian carcinomas. *Mol. Biol. Cell* **14**, 4376–4386 (2003).
27. Ohyagi-Hara, C. *et al.* miR-92a inhibits peritoneal dissemination of ovarian cancer cells by inhibiting integrin alpha5 expression. *Am. J. Pathol.* **182**, 1876–1889 (2013).
28. Deeb, S. J. *et al.* Machine learning-based classification of diffuse large B-cell lymphoma patients by their protein expression profiles. *Mol. Cell. Proteomics* **14**, 2947–2960 (2015).
29. Tyanova, S. *et al.* Proteomic maps of breast cancer subtypes. *Nat. Commun.* **7**, 10259 (2016).
30. Cheon, D. J. *et al.* Differential expression of argininosuccinate synthetase in serous and non-serous ovarian carcinomas. *J. Pathol. Clin. Res.* **1**, 41–53 (2015).
31. Watanabe, G. *et al.* alpha-B-crystallin: a novel p53-target gene required for p53-dependent apoptosis. *Cancer Sci.* **100**, 2368–2375 (2009).
32. Ellis, M. J. *et al.* Connecting genomic alterations to cancer biology with proteomics: the NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discov.* **3**, 1108–1112 (2013).
33. Edwards, N. J. *et al.* The CPTAC data portal: a resource for cancer proteomics research. *J. Proteome Res.* **14**, 2707–2713 (2015).
34. Yang, D. *et al.* Integrated analyses identify a master microRNA regulatory network for the mesenchymal subtype in serous ovarian cancer. *Cancer Cell* **23**, 186–199 (2013).
35. Ince, T. A. *et al.* Characterization of twenty-five ovarian tumour cell lines that phenocopy primary tumours. *Nat. Commun.* **6**, 7419 (2015).
36. Cox, J. & Mann, M. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **13** (Suppl 16): S12 (2012).
37. Schug, T. T., Berry, D. C., Shaw, N. S., Travis, S. N. & Noy, N. Opposing effects of retinoic acid on cell growth result from alternate activation of two different nuclear receptors. *Cell* **129**, 723–733 (2007).
38. Schug, T. T., Berry, D. C., Shaw, N. S., Travis, S. N. & Noy, N. Dual transcriptional activities underlie opposing effects of retinoic acid on cell survival. *Cell* **129**, 723–733 (2007).
39. Yoshida, H. *et al.* Accelerated degradation of PML-retinoic acid receptor alpha (PML-RARA) oncoprotein by all-trans-retinoic acid in acute promyelocytic leukemia: possible role of the proteasome pathway. *Cancer Res.* **56**, 2945–2948 (1996).
40. Sidell, N. Retinoic acid-induced growth inhibition and morphologic differentiation of human neuroblastoma cells *in vitro*. *J. Natl Cancer Inst.* **68**, 589–596 (1982).
41. Guo, T. *et al.* Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.* **21**, 407–413 (2015).
42. Toyama, A. *et al.* Proteomic characterization of ovarian cancers identifying annexin-A4, phosphoserine aminotransferase, cellular retinoic acid-binding protein 2, and serpin B5 as histology-specific biomarkers. *Cancer Sci.* **103**, 747–755 (2012).
43. Volkmann, J. *et al.* High expression of crystallin alphaB represents an independent molecular marker for unfavourable ovarian cancer patient outcome and impairs TRAIL- and cisplatin-induced apoptosis in human ovarian cancer cells. *Int. J. Cancer* **132**, 2820–2832 (2013).
44. Lin, J. H. *et al.* Connexin 43 enhances the adhesivity and mediates the invasion of malignant glioma cells. *J. Neurosci.* **22**, 4302–4311 (2002).
45. Kapoor, P. *et al.* Breast cancer metastatic potential: correlation with increased heterotypic gap junctional intercellular communication between breast cancer cells and osteoblastic cells. *Int. J. Cancer* **111**, 693–697 (2004).
46. Elzarrad, M. K. *et al.* Connexin-43 upregulation in micrometastases and tumor vasculature and its role in tumor cell attachment to pulmonary endothelium. *BMC Med.* **6**, 20 (2008).
47. Kenny, H. A. *et al.* Mesothelial cells promote early ovarian cancer metastasis through fibronectin secretion. *J. Clin. Invest.* **124**, 4614–4628 (2014).
48. Sawada, K. *et al.* Loss of E-cadherin promotes ovarian cancer metastasis via alpha 5-integrin, which is a therapeutic target. *Cancer Res.* **68**, 2329–2339 (2008).
49. Darai, E. *et al.* Expression of cadherins in benign, borderline, and malignant ovarian epithelial tumors: a clinicopathologic study of 60 cases. *Hum. Pathol.* **28**, 922–928 (1997).
50. Maines-Bandiera, S. L. & Auersperg, N. Increased E-cadherin expression in ovarian surface epithelium: an early step in metaplasia and dysplasia? *Int. J. Gynecol. Pathol.* **16**, 250–255 (1997).
51. Wiegand, K. C. *et al.* ARID1A mutations in endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.* **363**, 1532–1543 (2010).
52. Verhaak, R. G. *et al.* Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J. Clin. Invest.* **123**, 517–525 (2013).
53. Auersperg, N. The origin of ovarian cancers—hypotheses and controversies. *Front. Biosci. (Schol Ed)* **5**, 709–719 (2013).
54. Flesken-Nikitin, A., Choi, K. C., Eng, J. P., Schmidt, E. N. & Nikitin, A. Y. Induction of carcinogenesis by concurrent inactivation of p53 and Rb1 in the mouse ovarian surface epithelium. *Cancer Res.* **63**, 3459–3463 (2003).
55. Bowen, N. J. *et al.* Emerging roles for PAX8 in ovarian cancer and endosalpingeal development. *Gynecol. Oncol.* **104**, 331–337 (2007).
56. Bowen, N. J. *et al.* Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells. *BMC Med.* **7**, 21 (2009).
57. Auersperg, N., Woo, M. M. & Gilks, C. B. The origin of ovarian carcinomas: a developmental view. *Gynecol. Oncol.* **110**, 452–454 (2008).
58. Pothuri, B. *et al.* Genetic analysis of the early natural history of epithelial ovarian carcinoma. *PLoS ONE* **5**, e10358 (2010).
59. Kim, J. *et al.* High-grade serous ovarian cancer arises from fallopian tube in a mouse model. *Proc. Natl Acad. Sci. USA* **109**, 3921–3926 (2012).
60. Lengyel, E., Fleming, S., McEwen, K. A., Montag, A. & Temkin, S. M. Serial sectioning of the fallopian tube allows for improved identification of primary fallopian tube carcinoma. *Gynecol. Oncol.* **129**, 120–123 (2013).
61. Karst, A. M. & Drapkin, R. Primary culture and immortalization of human fallopian tube secretory epithelial cells. *Nat. Protoc.* **7**, 1755–1764 (2012).
62. Merritt, M. A. *et al.* Gene expression signature of normal cell-of-origin predicts ovarian tumor outcomes. *PLoS ONE* **8**, e80314 (2013).
63. Michalski, A. *et al.* Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* **10**, M111.011015 (2011).
64. Scheltema, R. A. *et al.* The Q Exactive HF, a benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol. Cell. Proteomics* **13**, 3698–3708 (2014).
65. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (pro)teomics data. *Nat. Methods*. doi: 10.1038/nmeth.3901 (2016).
66. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2011).
67. Chang, C. C. & Lin, C. J. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 27 (2011).
68. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
69. Geyer, P. E. *et al.* Plasma proteome profiling to assess human health and disease. *Cell Syst.* **2**, 185–195 (2016).
70. Viscaino, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447–D456 (2016).

#### Acknowledgements

We thank Korbinian Mayr, Igor Paron and Gabriele Sowa for their assistance in mass spectrometric analysis. This work was supported by the Max-Planck Society for the Advancement of Science, the Körber Foundation (Körber European Science Prize), a National Cancer Institute grant (NCI-R01 CA111882—E.L.) and a Colleen's Dream Foundation award (K.M.W.).

#### Author contributions

F.C. acquired and interpreted the proteomics data, developed the concept and edited the paper. K.M.W. wrote the manuscript, developed the concept, collected the samples and

carried out the ATRA experiments, western blots and immunofluorescence. M.C. and C.Y.C. collected the samples. M.A.E. isolated the primary cells. S.T. performed the statistical analysis. A.M. and R.R.L. helped with the histological analysis and provided advice. M.M. and E.L. together designed and supervised the study, developed the concept and edited the paper. All the authors have approved the final version.

#### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Coscia, F. *et al.* Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. *Nat. Commun.* 7:12645 doi: 10.1038/ncomms12645 (2016).

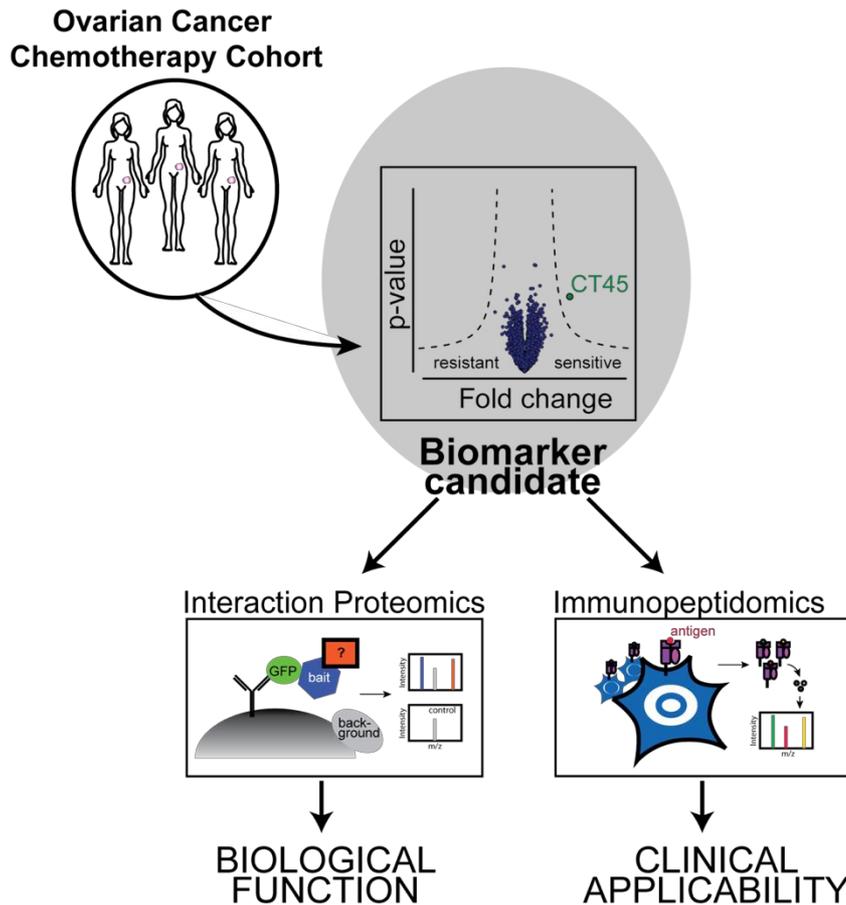


This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016

## 2.2 Manuscript 2 (in preparation): Proteomic characterization of chemotherapy sensitive ovarian cancer

### 2.2.1 Project aim and summary



Treatment for HGSOC involves a surgical debulking procedure followed by neoadjuvant carboplatin/paclitaxel combination chemotherapy. Unfortunately, 80-90% of patients diagnosed with advanced stage disease develop a chemoresistant state independent of any favorable initial response. Conversely, 17% of patients with advanced stage disease show long-term survival beyond 10 years after initial diagnosis (15). To date, it is still unclear what distinguishes these long-term survivors from the majority of patients with poor clinical outcome. The strong survival differences between these patient subgroups suggest the existence of yet unidentified, predisposing molecular mechanisms that actively contribute to chemotherapy sensitivity and promote long-term protection from any tumor relapse. The objective of

this study was therefore to apply state-of-the-art discovery-based proteomics to clinical biobank (FFPE) specimens obtained from advanced-stage HGSOC patients to identify discriminating proteins linked to the above mentioned different chemotherapy outcomes.

Our single-run proteomic workflow led to the quantification of in total ~ 9,000 proteins obtained from 25 biobank specimens; an unprecedented proteomic depth considering the absence of any sample pre-fractionation. Out of all quantitatively compared proteins, we identified a single differentially expressed protein that was highly up-regulated in the majority of the long-term survivors. To our knowledge, this is the first time that MS-based proteomics unambiguously discovered a clear molecular marker for cancer treatment success, and interestingly it manifested not as a global pattern but as a unique, differentially expressed entity. The protein turned out to be unknown except having been picked up in a large group of differentially expressed mRNAs where it has been termed cancer-testis antigen 45, CT45 (170) and neither the literature nor its sequence associate it to any known molecular function. The finding that CT45 was differentially expressed was then confirmed in a larger patient cohort (~150 cases) using tissue microarrays (TMA). This revealed that platinum resistance commenced after 363 days when CT45 staining was weak to strong *versus* 153.5 days for patients with no detectable expression of CT45.

Due to the tumor-specific expression of cancer-testis antigens and their recognized capacities to mount anti-cancer immune responses (150), we addressed if the favorable patient outcome was linked to T-cell mediated tumor rejection. We used an immunopeptidomics approach, which we applied to cell lines and tumor tissues to screen for CT45-derived HLA class I peptides. This indeed led to the first identification of CT45-derived HLA-I peptides. Our work further showed that these antigens are capable of activating patient-derived cytotoxic T-cells and to promote cancer cell killing *in vitro*. Based on this evidence, our results indeed suggested that the immune system may contribute to the favorable outcome of CT45+ patients. However, we were unable to ascertain a general correlation between the number of cytotoxic T-cells and CT45 expression in our initial discovery cohort. This indicated a strong immune inhibitory microenvironment in omental HGSOC tissues, in line with a recent report (59).

As no clear biological function has been reported for CT45, we used an interaction proteomics approach in the hope of gaining functional insights through ‘guilt by association’. Indeed, we found that CT45 interacted with the DNA damage-linked protein phosphatase 4 (PP4) complex on chromatin and that this caused an inhibitory phenotype in the context of PP4-mediated DNA damage signaling. We further showed that ovarian cancer cells expressing CT45 were more susceptible to carboplatin-based chemotherapy *in vitro* and *in vivo*, ultimately leading to apoptosis. Based on our functional studies, I developed a model in which cells expressing CT45 run into stalled chromatin relaxation after DNA damage repair, resulting from inhibited PP4-mediated KAP1/TRIM28 de-phosphorylation.

We conclude that CT45 may represent an ideal prognostic biomarker for chemotherapy sensitivity and long-term survival in HGSOV and possibly in other cancers. The identified HLA-I peptides represent promising candidate antigens for cancer immunotherapy approaches in the future or may help to stratify patients with differential responses to immune checkpoint blockage therapies. Patients with high tumoral levels of CT45 may represent ideal subjects for immune checkpoint therapies as T-cells would be exposed to strong antigens presented on the cancer cells.

## 2.2.2 Contribution

Ernst Lengyel suggested the comparison of responders and non-responders to chemotherapy and proposed this to Matthias Mann and myself in 2012. Matthias Mann, Ernst Lengyel (University of Chicago), Marion Curtis (University of Chicago), and I initiated this project, which was a shared effort between both laboratories over a period of 4 years starting end of 2012. Matthias Mann and Ernst Lengyel supervised me during all phases of the project. My work started with an adaptation and optimization of the proteomic sample preparation for clinical FFPE material, which soon allowed to us measure relatively deep (6.000 - 8.000 proteins) biobank proteomes in single-run (4h) analysis. I acquired all proteomic measurements of the study, analyzed and interpreted the data and identified CT45 as a marker of chemotherapy sensitivity. My interactomics work suggested an important functional role in DNA damage signaling and my HLA immunopeptidomics results showed that it may be a cancer immunotherapy target. Together with Marion Curtis and Bradley Ashcroft from University of Chicago, I then worked on follow-up experiments to gain functional insights. Marion Curtis collected and processed all biobank specimens of the study, conducted all animal experiments and performed the comet assays and other functional assays. Jaikumar Duraiswamy performed all T-cell based experiments. I designed most of the figures in the publication and wrote the manuscript together with Marion Curtis (Univ. of Chicago). Ernst Lengyel and Matthias Mann edited the manuscript.

### 2.2.3 Publication

## **Proteomics identifies CT45 as a mediator of chemosensitivity and immunotherapy target in ovarian cancer**

Fabian Coscia<sup>1</sup>, Ernst Lengyel<sup>2,5</sup>, Jaikumar Duraiswamy<sup>3</sup>, Bradley Ashcroft<sup>2</sup>, Michal Bassani-Sternberg<sup>1†</sup>, Michael Wierer<sup>1</sup>, Alyssa Johnson<sup>2</sup>, Kristen Wroblewski<sup>4</sup>, Hans Schreiber<sup>3</sup>, Anthony Montag<sup>3</sup>, S. Diane Yamada<sup>2</sup>, Matthias Mann<sup>1,5</sup>, and Marion Curtis<sup>2</sup>

(\*These authors jointly supervised this work)

<sup>1</sup>Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried 82152, Germany

<sup>2</sup>Department of Obstetrics and Gynecology/Section of Gynecologic Oncology, University of Chicago, Chicago, Illinois 60637, USA

<sup>3</sup>Department of Pathology, University of Chicago, Chicago, Illinois 60637, USA

<sup>4</sup>Department of Public Health Sciences, University of Chicago, Chicago, Illinois 60637, USA

<sup>†</sup>Present address: Department of Oncology, UNIL/CHUV, Ludwig Cancer Research Center, Epalinges 1066, Switzerland

<sup>5</sup>To whom correspondence should be addressed

## **ABSTRACT**

Most high-grade serous ovarian cancer (HGSOC) patients will develop resistance to platinum-based chemotherapy but a subset (15%) will remain disease-free for over a decade. To discover drivers of long-term survival following chemotherapy, we analyzed the proteomes of 25 platinum resistant and sensitive HGSOC patients to a depth of over 9,000 proteins. We identified cancer/testis antigen 45 (CT45) as an independent prognostic factor for prolonged disease-free survival. Immunopeptidomics discovered several CT45 derived HLA class I peptides capable of activating patient-derived cytotoxic T cells. Interaction proteomics identified a direct interaction with members of the protein phosphatase 4 (PP4) complex linking CT45 to the DNA damage response. CT45 mediates chemosensitivity by impeding PP4-dependent KAP-1 dephosphorylation following DNA damage leaving the chromatin in an open conformation and making it susceptible to carboplatin toxicity. Thus, CT45 is a novel regulator of chemosensitivity and is a potential target for immunotherapy.

## INTRODUCTION

HGSOC accounts for the majority of ovarian cancer fatalities (~70%) with the high mortality rate mainly driven by late diagnosis and frequently acquired resistance to chemotherapy (14). However, one in six patients remains disease-free for more than ten years after initial diagnosis with advanced stage disease (III or IV) treated with tumor debulking surgery following by chemotherapy treatment (15). This suggests the presence of tumor-intrinsic, pre-disposing features that influence chemotherapy outcome and promote long-term survival in this subset of patients. The molecular mechanisms which explain why some patients are long-term survivors have not been illuminated.

Cancer is predominantly studied at the genetic level, but proteomics has long been a promising complementary technology to gain insight into this disease at the protein level. Currently, there are few if any examples where this technology has identified clinically relevant biomarkers of the disease (136, 171). However, proteomics technology has improved tremendously in recent years and advances in proteomic sample preparation and computational workflows now enable very deep and quantitative profiling of cellular systems, including formalin-fixed and paraffin-embedded (FFPE) biobank material, to a depth and quantitative accuracy that could not be achieved previously (67, 123, 172). We therefore reasoned that a system-wide quantitative proteomics approach using state-of-the-art mass-spectrometry applied to clinical biobank specimens would offer a promising strategy to uncover driver proteins responsible for chemotherapy response and improved survival.

## RESULTS

### Identification of CT45 by shotgun proteomics

We performed quantitative proteomics on FFPE tumor samples isolated from 25 chemotherapy-naïve patients with advanced stage HGSOE (Fig. 1a, Extended Data Fig. 1, Supplementary Tables 1 and 2). Since patients with advanced stage disease generally have the worst prognosis (15), we selected metastatic tumors from the omentum for proteomic analysis. We first assessed global proteomic differences between chemoresistant (N=11, median DFS = 190 days) and chemosensitive (N=14, median DFS=1160 days) patients in our cohort by adapting a recently described and highly sensitive label-free proteomic workflow capable of accurately quantifying a large portion of the cellular proteome (103, 173). We stringently identified and quantified more than 9,000 proteins from low  $\mu\text{g}$  input archival samples in single-run measurements in the MaxQuant environment (94, 103) (1% FDR at protein and peptide levels; Extended Data, Fig. 1b). The dynamic range of protein signals spanned more than six orders of magnitude (Extended Data Fig. 1c). Pearson R values between all specimens were consistently above 0.77 (mean 0.88) and the correlation was 0.95 between independently prepared tissue of the same tumor (Extended Data Fig. 1d). Based on the quantitative levels of 8,190 proteins after data filtering (Methods, Extended Data Fig. 1b), we observed few overall proteome changes across patients. However, the comparison of chemosensitive and chemoresistant patients identified cancer/testis antigen 45 (CT45), as significantly higher expressed in chemosensitive patients (FDR < 5%, Fig. 1b). Furthermore, CT45 protein levels strongly correlated with disease-free survival time (Extended Data Fig. 2a). CT45 is comprised of 10 distinct but highly similar genes, which make them nearly identical at the protein level (amino acid identity >98%, Extended Data Fig. 2b). Immunohistochemistry for CT45 confirmed the relative expression of CT45 in serial sections of the tumors used in the proteomic cohort and showed localization to the nucleus and in some cases to the nucleolus (Fig. 1c).

To validate the proteomic findings in a larger patient cohort, we stained tissue microarrays (TMA) to analyze CT45 expression in over 200 cases of primary and metastatic ovarian cancer (Extended Data Fig. 3a). CT45 protein levels were assessed by a gynecologic pathologist blinded to patient outcomes. CT45 was only rarely expressed in cases of endometrioid, clear cell, or mucinous ovarian cancer subtypes compared to serous papillary - in line with a previous study of mRNA expression (174) - and the protein was expressed at identical levels between primary and metastatic tumors from the same HGSOE patients. Consistent with our discovery

cohort, CT45 expression correlated with chemosensitivity in 124 patients with advanced stage HGSOC (FIGO 2014 stage IIIb or higher) ( $P = 0.005$ , **Extended Data Fig. 3a**). Furthermore, patients with high CT45 expression (staining 1+) had prolonged disease-free survival compared to patients with no expression of CT45 ( $P = 0.02$ , 363 days *versus* 153.5 disease free days; **Fig. 1d**) and there was a trend towards longer overall survival ( $P = 0.09$ , **Extended Data Fig. 3b**). Thus, CT45 expression is an independent prognostic indicator for advanced stage HGSOC.

### **CT45 is an independent prognostic indicator**

To validate the proteomic findings in a larger patient cohort, we stained tissue microarrays (TMA) to analyze CT45 expression in over 200 cases of primary and metastatic ovarian cancer (**Extended Data Fig. 3**). CT45 protein levels were quantified by a gynecologic pathologist who was blinded to the patients' outcomes. CT45 expression was not as common in endometrioid, clear cell, mucinous ovarian cancer subtypes compared to serous papillary, in line with large-scale mRNA expression data (174). There was no significant difference in protein expression between primary and metastatic tumours from the same HGSOC patients ( $P = 0.61$ ). However, consistent with our discovery cohort, CT45 expression correlated with chemosensitivity in 124 patients with advanced stage disease (FIGO IIIb or higher) (**Extended Data Fig. 3**). Kaplan-Meier survival analysis revealed that advanced stage (FIGO IIIb or higher) HGSOC patients with high CT45 expression had a lower risk of recurrence with median days to platinum resistance of 363 days (staining 1+) *versus* 153.5 days for patients with no expression of CT45 ( $P = 0.02$ , **Fig. 1e**). A trend towards longer overall survival was also evident ( $P = 0.09$ , **Fig. 1f**). We further validated our proteomic findings using a larger and independent public dataset generated by The Cancer Genome Atlas (TCGA) consortium (64). Using RNA-Seq data of 305 HGSOC cases, we calculated CT45 expression from the accumulated expression of the detected 6 gene family members (CT45A1-A6). Grouping patients into a CT45 high group (top 25%,  $N=76$ ) and CT45 low group (lower 75%,  $N=229$ ) resulted in a significantly longer overall survival for the CT45 high group ( $P = 0.01$ , **Fig. 1g**). Thus CT45 expression is an independent prognostic indicator for advanced stage HGSOC.

## CT45 is a native tumor antigen

Despite previous reports on CT45 expression in several tumor types (175-178), its role as a tumor antigen has not been elucidated. To investigate whether CT45-specific peptides are bound and presented on human leukocyte antigen (HLA) class I complexes on cancer cell lines, we used an approach coupling immunopeptidomics to mass spectrometry(163) with a cell line (59M) expressing high endogenous levels of CT45 (**Extended Data Fig. 4a and 4b**). Of 6,413 identified HLA-I peptides (**Extended Data Fig. 4c, Supplementary Table 3**), five derived from CT45: AVDPETVFK, GVQGPTAVR, GVQGPTAVRK, VQGPTAVRK, QGPTAVRK; all of which matched the binding motifs of the A-03:01 and A-11:01 HLA-I receptor (**Extended Data Fig. 4d, 4e, and 4f**). We further confirmed the identity of these peptides using synthetic versions and tandem mass spectrometry (**Extended Data Fig. 4g**). *In-silico* HLA epitope prediction revealed weak and strong binding affinity of AVDPETVFK to A-03:01 or A-11:01, respectively, and low or no binding affinity for the remaining four peptides (**Fig. 2a**).

CT45 expression is regulated by DNA methylation (174) and patients treated with demethylating agents show an altered immune response that correlates with upregulation of cancer/testis antigens as well as genes involved in immunomodulatory pathways (179, 180). Treatment with 5-aza-2'-deoxycytidine (DAC) strongly activated CT45 expression in SKOV3ip1 ovarian cancer cells (**Fig. 2b, Extended Data Fig. 5a**) along with the CT antigens NY-ESO-1, MAGEA4 and SSX2, in line with recent publications (179, 180). Using the same immunopeptidomics strategy, we identified three additional, but sequence related CT45 peptides matching to the measured peptide motifs of the alleles A-03:01 and/or A-68:01 (**Fig. 2c, Extended Data Fig. 5b, 5c, and 5d, Supplementary Table 3**).

To address if the identified epitopes function as targets for CD8<sup>+</sup> T cells, we stimulated patient derived A-11:01 CD8<sup>+</sup> T cells collected from a CT45 positive tumor with the two CT45 peptides with the best binding affinity predictions (AVDPETVFK and GVQGPTAVRK). Both CT45 peptides, but not a control peptide, induced T cell activation as assessed by Ki67 and intracellular IFN $\gamma$  staining of CD8<sup>+</sup> T cells (**Fig. 2d**). Similar results were obtained using A-03:01 CD8<sup>+</sup> T cells using three peptides identified after DAC treatment (**Extended Data Fig. 5e**). The A-11:01 CD8<sup>+</sup> T cells reacted with one of two CT45 tetramers (A-11) containing the peptide, AVDPETVFK, which had the highest predicted affinity for A-11:01, while A-03:01 CD8<sup>+</sup> T cells reacted with the tetramer (A-03) containing the GVQPTAVRK peptide (**Fig. 2e**). Co-culture of the CT45<sup>+</sup> 59M cell line with peptide stimulated A-11:01 CD8<sup>+</sup> tumor-infiltrating lymphocytes (TIL) targeting two different CT45 peptides lysed the cancer cells in a

dose-dependent manner (**Fig. 2f**). These data demonstrate that CT45 is an endogenously processed and presented antigen recognized and targeted by patient-derived CD8<sup>+</sup> T cells.

### **CT45 is a functional mediator of chemosensitivity**

In addition to DAC's ability to modulate the immune response, it also shows synergism with platinum agents in ovarian cancer therapy (181, 182). DAC sensitized SKOV3ip1 ovarian cancer cells to carboplatin (**Extended Data Fig. 6a**) and intriguingly, proteomic analysis on DAC treated cells revealed that CT45 was among the top ten DAC induced proteins (**Fig. 2b**). Overexpression of CT45 in an ovarian cancer cell line reduced the number of colonies in a colony formation assay (**Fig. 3a**) but did not affect proliferation or the cell cycle (**Extended Data Fig. 6c** and **6d**). Addition of carboplatin reduced proliferation and colony formation in CT45 expressing cells (**Extended Data Fig. 6c** and **Fig. 3a**). Similar effects were observed in a second serous ovarian cancer cell line, OVKATE (**Extended Data Fig. 6b** and **6e**). Treatment with carboplatin significantly reduced growth of CT45 expressing subcutaneous tumors in an immunodeficient mouse as compared to the untreated CT45<sup>+</sup> tumors while having no effect on the vector control cell line (**Fig. 3b**). CT45 expression also significantly reduced overall tumor growth as compared to the vector control (**Fig. 3b**). This data reveals a functional link between CT45 expression and carboplatin chemosensitivity and highlights the potential tumor suppressive capacities of CT45 in ovarian cancer.

To understand the mechanism of CT45-mediated chemosensitivity, we mined a recent quantitative human interactome study published by our laboratory (117). Interestingly, we found that CT45 interacts with several members of the evolutionarily conserved protein phosphatase 4 (PP4) complex. The PP4 interaction was confirmed in two ovarian cancer cell lines (OVCAR-5 and COV318) expressing Flag or V5-tagged CT45, respectively. The regulatory subunits PP4R3 $\alpha$ , PP4R3 $\beta$ , and PP4R2 as well as the catalytic subunit, PP4C, of PP4 were all highly significantly enriched in CT45 immunoprecipitates ( $p < 0.001$ ; **Fig. 3c** and **3d**, **Extended Data Fig. 6f**) pointing to their specific and direct interaction with CT45. Since PP4 deficiency impedes the DNA damage response (DDR) (183-187) and causes hypersensitivity to platinum compounds (188), we next investigated if CT45 expression influenced the level of DNA damage in ovarian cancer cells following treatment with carboplatin. Indeed, CT45-expressing cancer cells showed increased levels of the DNA damage marker  $\gamma$ H2AX and cleaved caspase-3 after carboplatin exposure (**Fig. 3e**). In addition, more DNA damage was present in CT45-expressing cells as demonstrated by longer tail moments in a comet assay (**Fig. 3f**).

### **CT45 promotes chromatin relaxation mediated by KAP1 phosphorylation**

PP4 plays an important role in the DNA damage response by dephosphorylating several critical proteins including 53BP1 (186),  $\gamma$ H2AX (183, 189), RPA2 (184), and KAP-1/TRIM28 (190, 191). Using a candidate based approach we found that carboplatin treatment induced phosphorylation of KAP-1 at the ATM target site Ser824 (192). Phosphorylation was increased in CT45-expressing cells as compared to vector control cells following DNA damage (**Fig. 4a**). Consistent with these results, phosphorylated KAP-1 was pan-nuclear, peaking at day 3 following carboplatin treatment in both cell lines (**Fig. 4b**). In CT45-expressing cells p-KAP-1 peaked at a higher level and remained elevated on day 5 and day 7, while the p-KAP-1 signal dissipated to baseline levels by day 7 in control cells. We observed similar trends for  $\gamma$ H2AX. KAP-1 is a building block of heterochromatin important for chromatin condensation (193) and to understand whether CT45 is also associated with chromatin, we analyzed both ectopically (OVCAR5-CT45) and endogenously (59M) CT45 expressing cells using a biochemical chromatin segregation assay(194). CT45 was strongly chromatin-enriched and associated with nuclease resistant heterochromatin, similar to its PP4 interaction partners, as well as the known heterochromatin linked proteins HDAC2 and KAP-1 (**Fig. 4c** and **Extended Data Fig. 7a**). One possibility for how CT45 may inhibit KAP-1 dephosphorylation is by altering KAP-1 retention on the chromatin (195), but we did not find that either CT45 or carboplatin altered the chromatin enrichment of KAP-1 or PP4 (**Extended Data Fig. 7b**).

Furthermore, chromatin-immunoprecipitation coupled to mass spectrometry (ChIP-MS)(196) confirmed an association of CT45 with heterochromatin as indicated by the co-enrichment of KAP1 and HDAC2 in addition to the PP4 complex (**Fig. 4d**). Pathway analysis of the CT45 interactome revealed a strong enrichment of heterochromatic and DDR linked proteins (**Fig. 4e**). Reciprocal ChIP-MS using KAP-1 as a bait showed a clear enrichment of CT45, independently confirming its association with chromatin (**Extended Data Fig. 7d** and **7e**). The interaction of CT45 with heterochromatin-associated PP4 was not affected by carboplatin (**Extended Data Fig. 7c**).

Since KAP-1 phosphorylation at S824 induces global chromatin decondensation following DNA damage (191, 192, 194), we next asked if the KAP-1 Ser-824 phosphorylation differences caused changes in chromatin structure. Following carboplatin treatment, cells expressing CT45 had larger nuclei on day 5 suggesting an increased level of relaxed chromatin (**Fig. 4f**). Structural chromatin changes were interrogated using a micrococcal nuclease (MNase) assay (197) at several time points following carboplatin exposure. After five days, carboplatin treatment induced global chromatin relaxation in both vector control and the CT45

overexpressing cell line. Strikingly, CT45 induced a higher proportion of nuclease-accessible chromatin compared to vector control cells (**Fig. 4g** and **4h**) indicating that the chromatin is more accessible to platinum-induced chromatin damage. Furthermore, whereas the control cells completely restored chromatin compaction at day 7, chromatin in CT45 expressing cells remained in a relaxed conformation suggesting that CT45 impedes the KAP-1 mediated restoration of chromatin compaction through its binding to PP4 during recovery from DNA damage (**Fig 4i**).

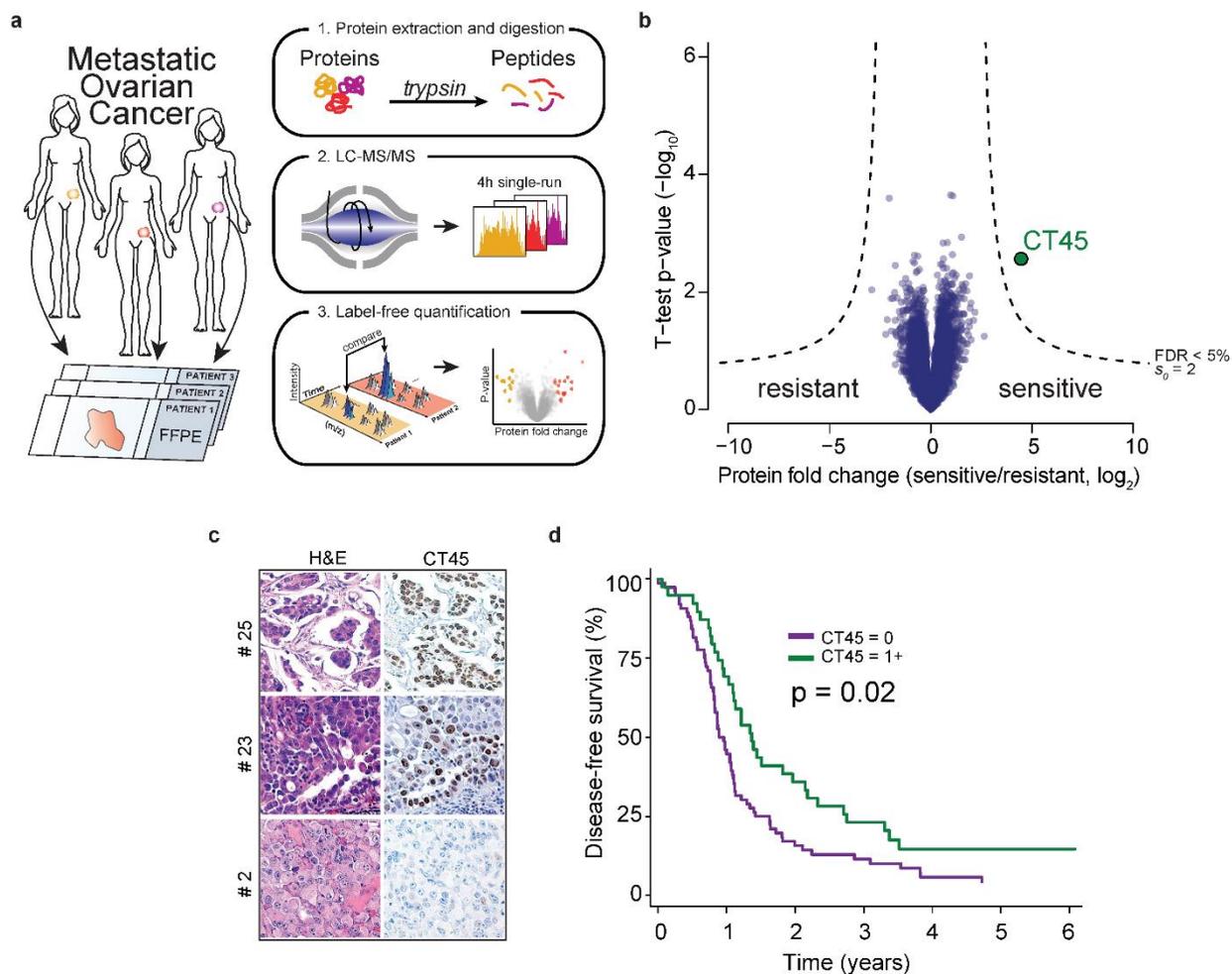
## DISCUSSION

Until recently, proteomic technology was not amenable to in-depth proteome analysis from FFPE tissue without extensive sample pre-fractionation, which requires relatively large sample amounts and measurement times. Using a high sensitivity, single-run workflow on minute amounts of FFPE tissue, we are now able to quantify over 9,000 proteins, which was crucial in the identification of CT45 as an independent prognostic factor for patients with advanced stage HGSOV. The ability to successfully use minute amounts of starting material from biobank samples as demonstrated here provides validation for the application of MS-based proteomics to a variety of important clinical questions and in particular to the long elusive identification of clinically relevant cancer biomarkers.

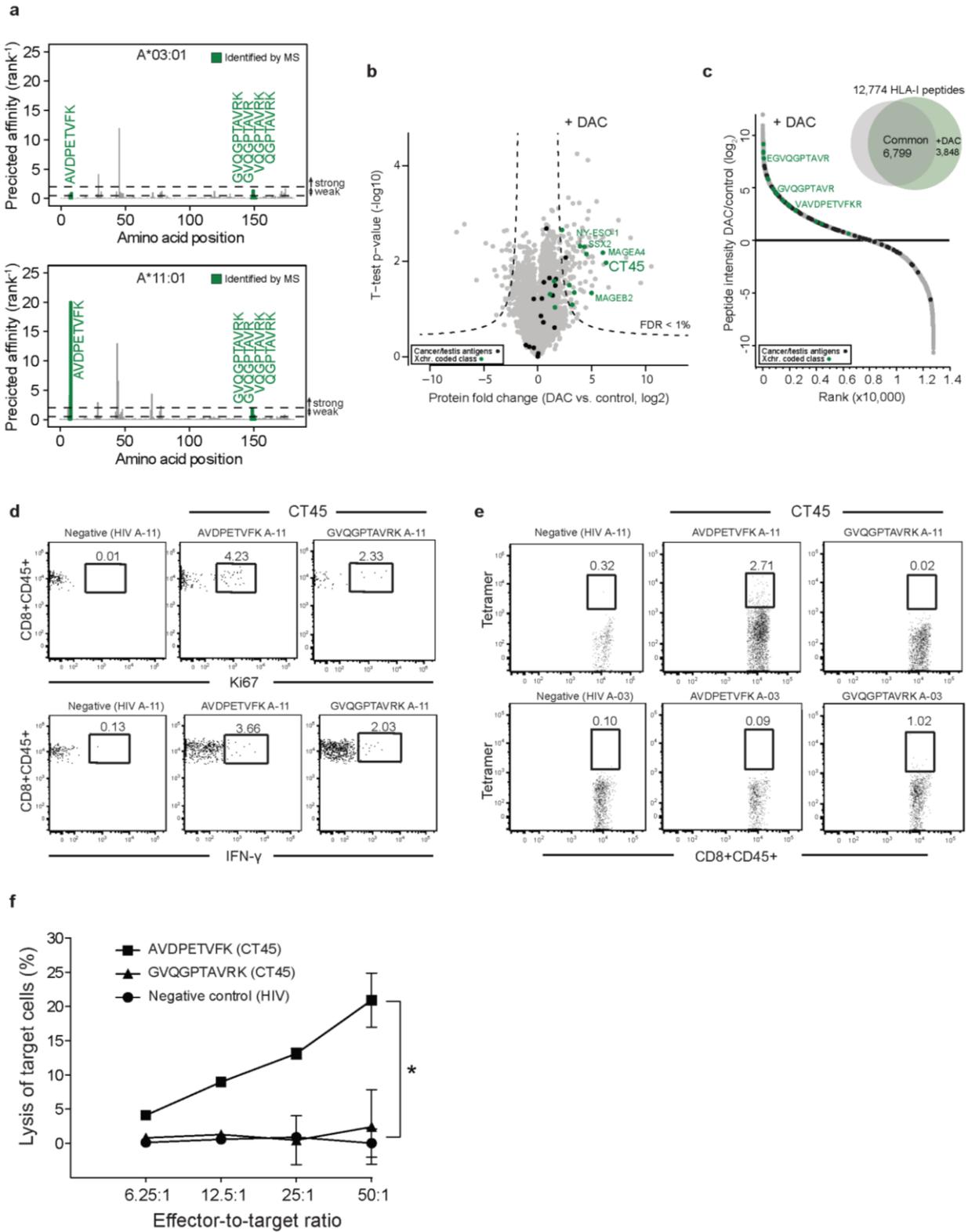
Mutational load correlates with clinical response to CTLA-4 blockade in melanoma (198) and to PD-1 inhibition in colorectal cancers with mismatch-repair deficiencies (199). Therefore, most immunotherapy approaches have focused on cancers with a high mutational load and the presence of mutant antigens. Notably, ovarian cancer patients treated with immune checkpoint inhibitors have not experienced the impressive response rates evident in some other cancers (200). This has been attributed to the low mutational burden present in ovarian tumors which are characterized by copy number alterations (62, 201). In our discovery cohort several patients showed long term survival following treatment with surgery and chemotherapy. Two of these patients are positive for the HLA class I alleles A-03:01 or A-11:01 which bind the CT45 derived peptides we discovered using immunopeptidomics (**Fig. 2**). We hypothesize that CT45 may serve as a non-mutant tumor rejection antigen that provides long-term protection to patients through activation of cytotoxic T cells which inhibit cancer growth. Based on our studies we could envision a treatment approach where CT45 expression is activated on tumor cells (e.g. by DAC) and targeted by T cells engineered to recognize the CT45 peptide:HLA-I complex.

Using interaction proteomics, we found that CT45 directly interacts with the evolutionarily conserved protein phosphatase 4 complex, mediating sensitivity to the DNA damaging agent carboplatin by impeding dephosphorylation of KAP-1. Phosphorylated KAP-1 provides a global signal to transiently relax chromatin in order for DNA damage repair to occur (193, 202). Our data suggest that CT45 prevents chromatin associated PP4 from dephosphorylating KAP-1 and thereby suspends chromatin in a relaxed state (**Fig. 4i**). Over time this leads to further accumulation of platinum induced DNA damage as more of the chromatin is exposed and ultimately result in cell death. KAP-1 was originally described as a transcriptional corepressor (203) and was shown to assist in the inactivation of p53 through an interaction with MDM2 leading to repression of p53 transcriptional activity (204). Since most HGSOC tumors have mutant p53 it is unknown what if any affect CT45 may have on the transcriptional activity of p53 mediated by prolonged KAP-1 phosphorylation.

Overall, our data suggest a model (**Fig. 4i**) whereby CT45 functions as both a cell intrinsic mediator of chemosensitivity by impeding KAP-1 mediated chromatin condensation during recovery from DNA damage and as an antigen for CD8+ T cells. Both these findings may be clinically relevant. We propose that CT45 expression could be activated in tumors lacking CT45 by treating with demethylating agents to improve efficacy of chemotherapy both during first line therapy and with recurrent disease. Furthermore, immunotherapy targeting CT45 either alone or in combination with platinum-based chemotherapy could be an effective strategy for the treatment of advanced stage ovarian cancer.



**Figure 1. Proteomics identifies CT45 expression to predict long-term survival in HGSO**  
**a**, Summary of the shotgun proteomics workflow applied to FFPE biobank specimens from ovarian cancer patients. Following tissue lysis and homogenization, purified proteins were digested and analyzed in single-run HPLC-MS using a Q Exactive mass spectrometer. Data were analyzed and quantified in MaxQuant(94, 103). **b**, Volcano plot of chemotherapy resistant *versus* sensitive patient proteomes. Expression fold changes are plotted against the t-test p-value. Dashed lines indicate the significance threshold (FDR < 0.05,  $s_0 = 2$ ). CT45 is highlighted in green. **c**, Immunohistochemistry for CT45 and corresponding H&E staining in serial sections of tumor from 3 representative patients. **d**, Kaplan-Meier survival analysis based on CT45 staining scores from ovarian cancer TMAs for disease-free survival. Advanced stage HGSO patients comparing a staining score of 0 (N=82) versus 1+ (N=42).



**Figure 2. CT45 is a native tumor antigen**

**a**, Predicted binding affinities (NetMHC4.0) for HLA class I peptides of CT45 with a length of 8-11 amino acids. Affinities are plotted on the y-axis as % rank<sup>-1</sup>. Weak affinity cut-off: % rank < 2, high affinity cut-off: % rank < 0.5. Peptides identified by MS are highlighted in green.

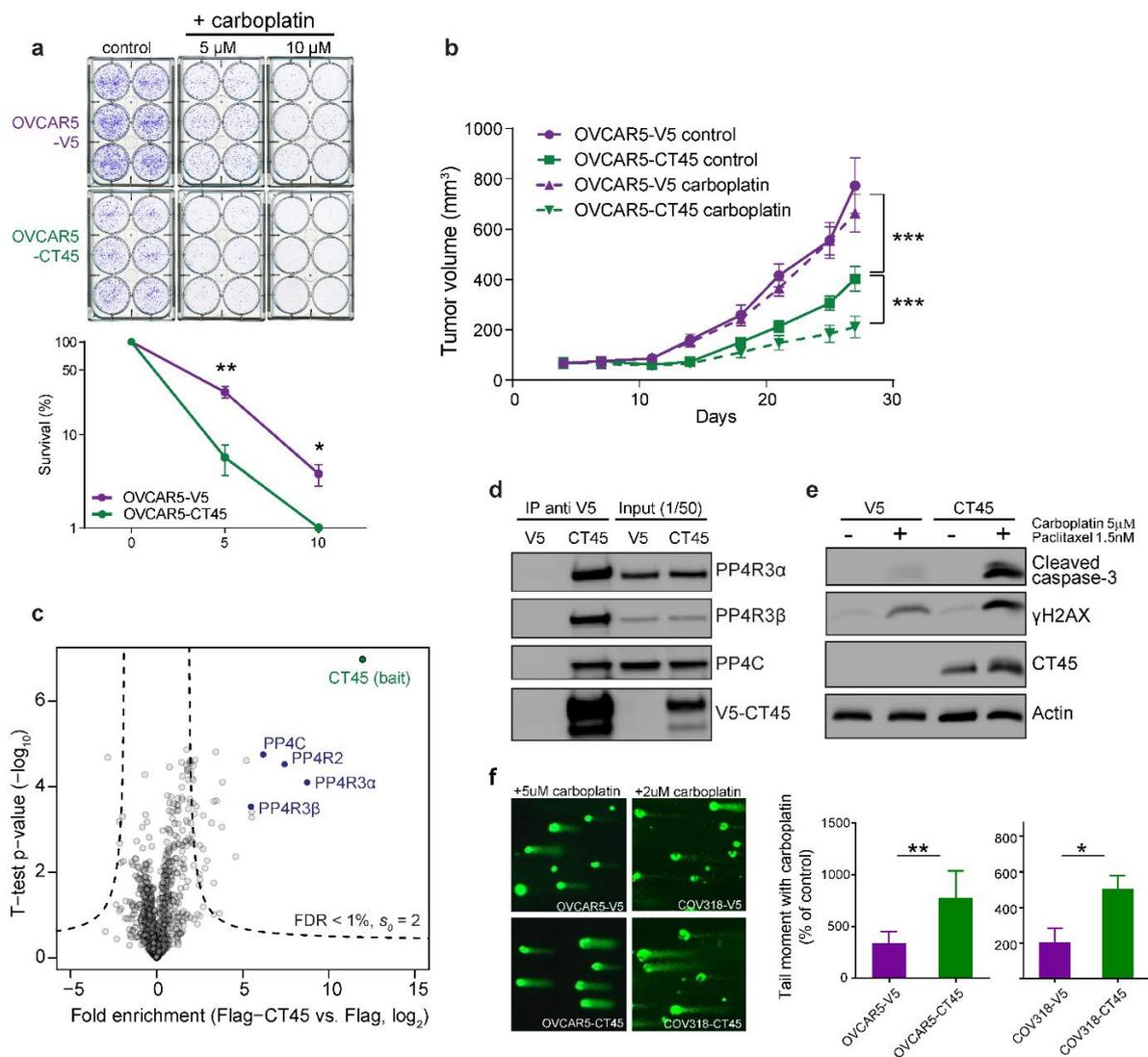
**b**, Volcano plot of the proteomic comparison between 5-aza-2'-deoxycytidine (DAC) treated and control SKOV3ip1 ovarian cancer cells. Protein fold change (t-test difference, log<sub>2</sub>) is plotted against the t-test p-value (-log<sub>10</sub>). Significance thresholds are indicated by dashes lines (FDR < 0.01). Cancer-testis antigens including CT45 are highlighted.

**c**, HLA-I peptide intensity ratio from immunopeptidomics of DAC treated versus control SKOV3iP1 ovarian cancer cells, plotted against the ranked peptide ratio. CT45 peptides are highlighted in green.

**d**, Staining for Ki-67 and IFN-γ of CD8+ T cells (A-11:01) after stimulation with two CT45 peptides (AVDPETVFK and GVQGPTAVRK) or one HIV negative control peptide analyzed with flow cytometry.

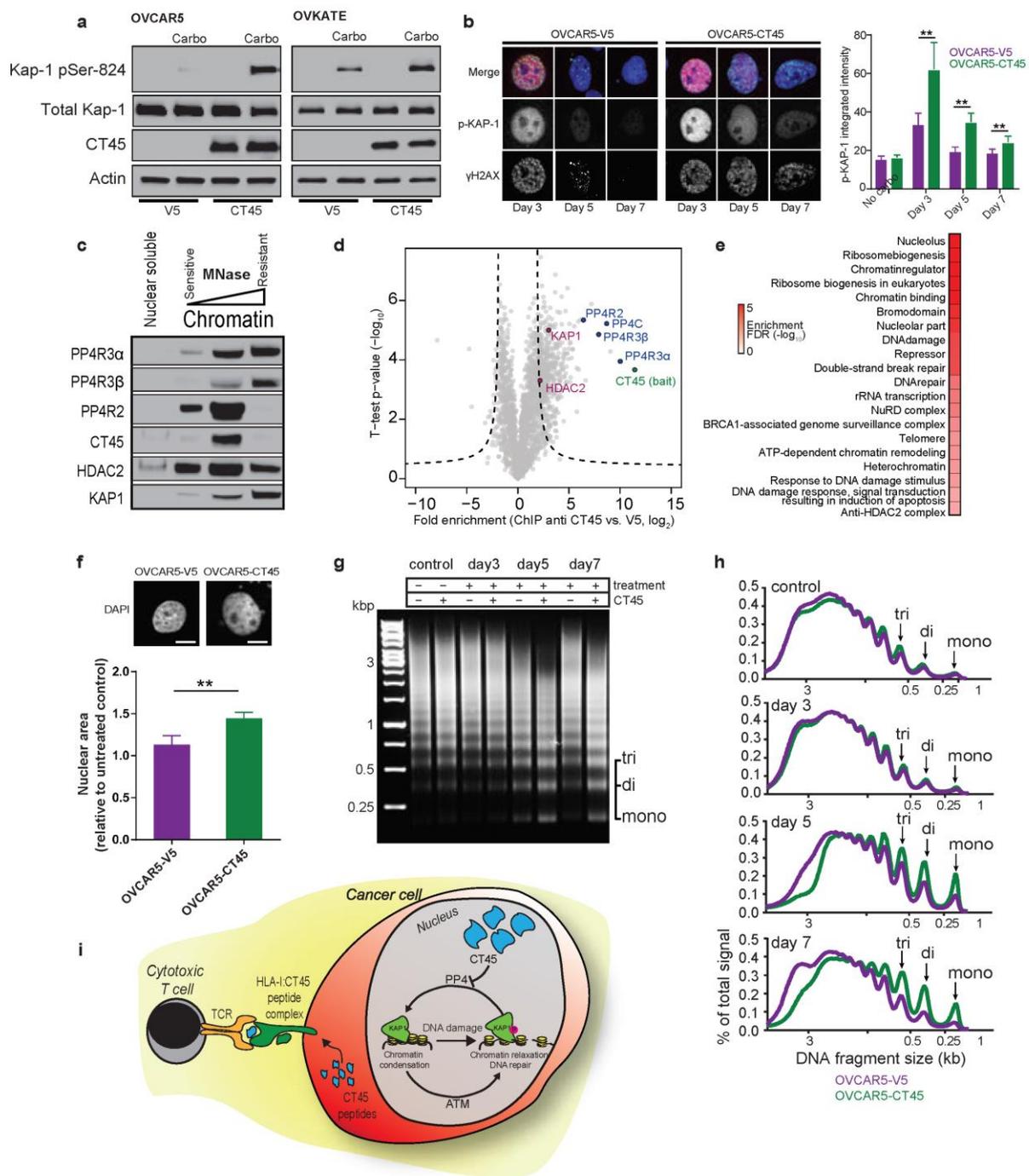
**e**, Tetramer staining of A-11:01 or A-03:01 CD8+ T cells with two CT45 tetramers (AVDPETVFK and GVQGPTAVRK) and one HIV negative control tetramer analyzed with flow cytometry.

**f**, Lysis of HLA-A11:01 positive 59M cell line by CD8+ effector T cells (A-11:01) at indicated effector:target ratios using a chromium release assay. Data are means ± s.d. from two independent experiments.



### Figure 3. CT45 mediates chemotherapy sensitivity

**a**, Clonogenic survival assay of the ovarian cancer cell line OVCAR5 stably overexpressing CT45 (green) or control vector (purple) after carboplatin (5 $\mu$ M) treatment. Dots represent mean values from three independent experiments. Error bars show s.e.m. for each group. Representative images are shown above bars. **b**, Growth of OVCAR5-V5 (control plasmid) and OVCAR5-V5-CT45 tumors (N=5-8) over time during treatment with carboplatin (20mg/kg). Data are means  $\pm$  s.e.m. for each group. **c**, Interaction proteomics screen in OVCAR5 cells stably overexpressing FLAG-tagged CT45. Protein enrichment (t-test difference) was calculated over the corresponding control cell line (FLAG tag alone) and plotted against the t-test p-value (-log<sub>10</sub>). Dashed lines indicate significance thresholds. The bait protein CT45 (green) and members of the PP4 complex (blue) are highlighted. Results represent 3 replicates per experiment group P < 0.01. **d**, Western blot of immunoprecipitated V5-tagged CT45 protein. The co-enriched members of the protein phosphatase 4 complex, PP4R3 $\alpha$ , PP4R3 $\beta$  and PP4C, are shown. **e**, Western blot of cleaved caspase-3,  $\gamma$ H2AX, and CT45 following treatment with carboplatin (5 $\mu$ M) and paclitaxel (1.5nM). Day 5. **f**, Tail moment following carboplatin treatment at day 5 using a comet assay. Data are means  $\pm$  s.e.m. of 4 independent replicates (right panel). Representative images of comet assays are shown. OVCAR5 cells were treated with 5 $\mu$ M carboplatin and COV318 were treated with 2 $\mu$ M carboplatin.



**Figure 4. CT45 interferes with chromatin dynamics after DNA damage**

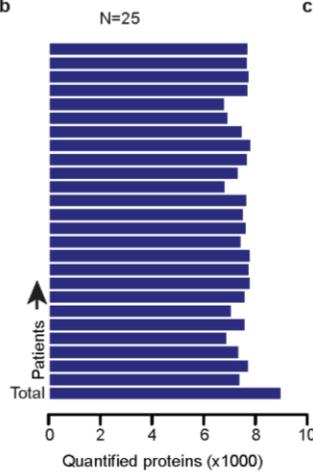
**a**, Western blot analysis of OVCAR5 and OVKATE cell line pairs treated with carboplatin (5 $\mu$ M). Day 3. **b**, Immunofluorescence staining of KAP1-S824 (red)  $\gamma$ H2AX (green) and DAPI (blue) in OVCAR5-V5 and OVCAR5-V5-CT45 cell line pair with or without carboplatin treatment (5 $\mu$ M) at day 3, 5, and 7. Data are means  $\pm$  s.e.m. of 5 independent replicates (right panel). Representative cells are shown left. **c**, CT45 positive 59M cells were collected and processed for a chromatin segregation assay as detailed in the Methods. PP4 complex members, KAP1, and HDAC2 were detected by western blot. **d**, Volcano plot of chromatin-immunoprecipitation mass spectrometry (ChIP-MS) results for the V5-tag in OVCAR5-V5-CT45 vs. OVCAR5-V5 cell line. Fold enrichment of V5-tagged CT45 over control cell line

(V5 tag alone) is plotted against the t-test p-value ( $-\log_{10}$ ). Dashed lines indicate significance thresholds ( $FDR < 0.01$ ,  $s_0 = 2$ ). **e**, Pathway enrichment analysis of proteins significantly co-enriched with CT45 (right side of volcano plot in panel d). Most significant pathways are ranked by enrichment false-discovery rate. **f**, Fold change in nuclear sizes of OVCAR5-V5-CT45 and OVCAR5-V5 cells at day 5 with carboplatin ( $5\mu\text{M}$ ) treatment normalized to untreated cells. Representative DAPI images are shown in upper panel. Bar plots represent nuclear sizes from 5 independent experiments. Scale bar= $10\mu\text{M}$ . Data are means  $\pm$  s.e.m. **g**, **h** OVCAR5-V5-CT45 and OVCAR5-V5 cells were collected and processed for the micrococcal nuclease (Mnase) assay as described in the Methods. Results are shown for untreated control samples at day 3 and day 3, 5 and 7 with carboplatin ( $5\mu\text{M}$ ) treatment. DNA size profiles are illustrated in **h**. **i**, Proposed model of CT45 mediating chemosensitivity and long-term survival in metastatic HGSOC.

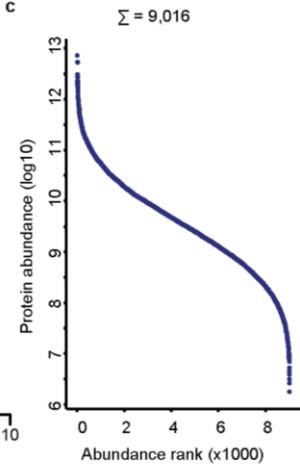
a

	Chemotherapy Status				P value	All cases
	Resistant		Sensitive			
	No.	%	No.	%		
<b>Stage</b>					1.00	
3	8	72.7	10	71.4		18
4	3	27.3	4	28.6		7
<b>Primary Site</b>					0.0850	
Fallopian Tube	4	36.4	6	42.8		10
Ovary	7	63.6	4	28.6		11
Peritoneum	0	0	4	28.6		4
<b>Residual Disease</b>					0.0531	
Larger than 1 cm	8	72.7	5	35.7		13
Smaller than 1 cm	2	18.2	9	64.3		11
Unknown	1	9.1	0	0		1
<b>Grade</b>					0.4867	
2	0	0	2	14.3		2
3	11	100	12	85.7		23
<b>Chemo Type</b>						
Adjuvant/1st line	11	100	14	100		
<b>Chemo Class</b>					1.00	
Platinum only	0	0	1	7.1		1
Taxane/Platinum	11	100	13	92.9		24
<b>Progression free survival</b>					< 0.0001	
Median days (range)	190 (113-391)		1160.5 (552-3229)			
<b>Overall survival</b>					< 0.0001	
Median days (range)	448 (228-1489)		1550.5 (967-3229)			
<b>Age</b>					0.0908	
Mean age (range)	60.0 (47-78)		67.1 (51-86)			

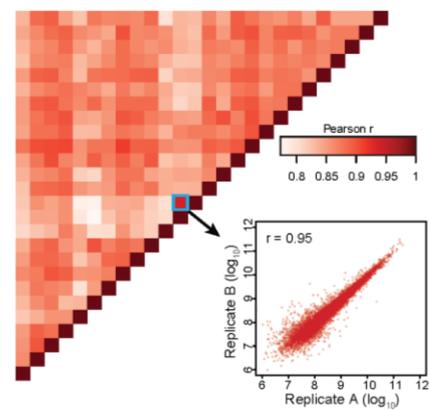
b



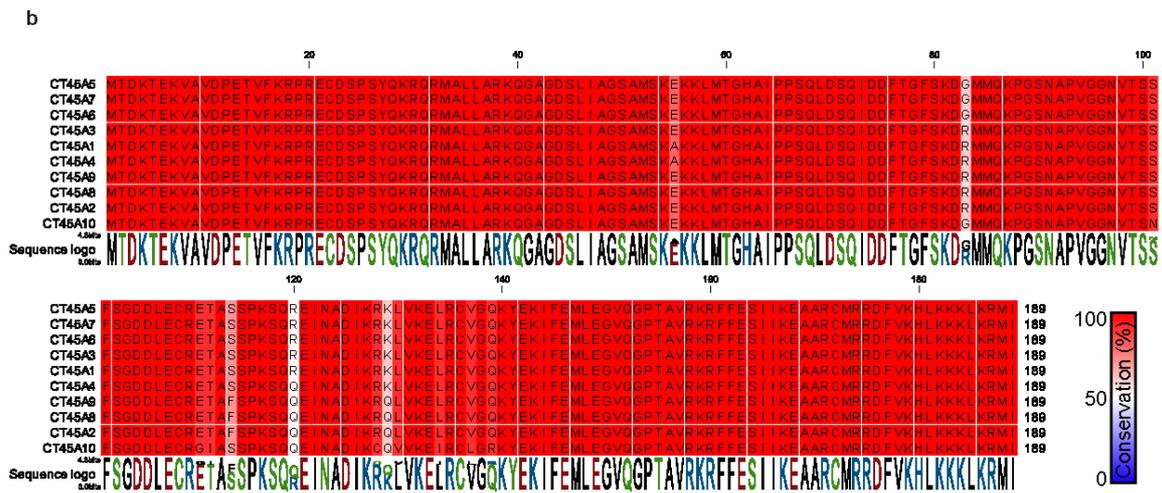
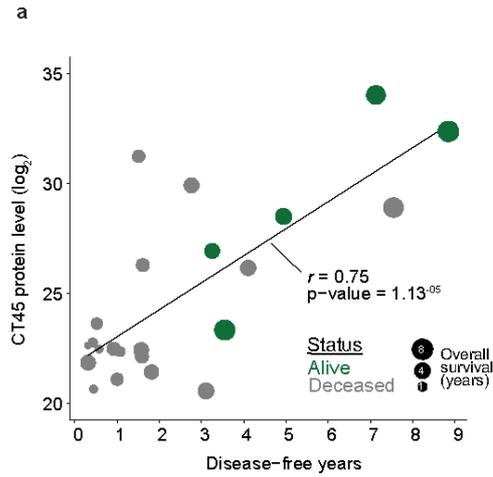
c



d



**Extended Data Figure 1. Patient characteristics of proteomic cohort and proteomic depth per specimen.** **a**, Clinico-pathologic parameters for the resistant and sensitive patients analyzed by proteomics. **b**, Proteomic depth per patient. Samples were measured as technical duplicate. **c**, Dynamic range of protein abundance of all quantified proteins. **d**, Correlation matrix of all measured samples based on Pearson correlation values. The proteomic comparison of independently prepared tissue of the same tumor is highlighted.

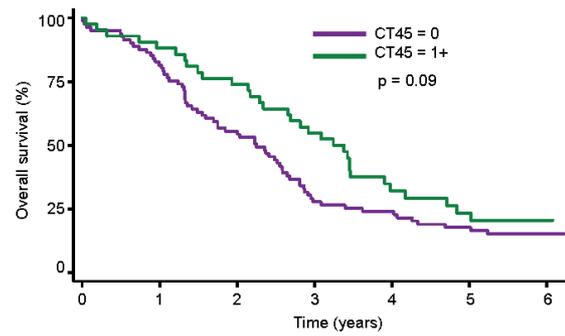


**Extended Data Figure 2. CT45 gene homology and staining validation.** **a**, Correlation analysis of disease-free survival time versus CT45 protein expression ( $\log_2$ ) for 25 HGSOC patients. Patients highlighted in green were alive at the time of analysis. **b**, Amino acid alignment of the 10 members of the CT45 gene family. Color shows level of conservation with red being 100% and blue being 0%.

a

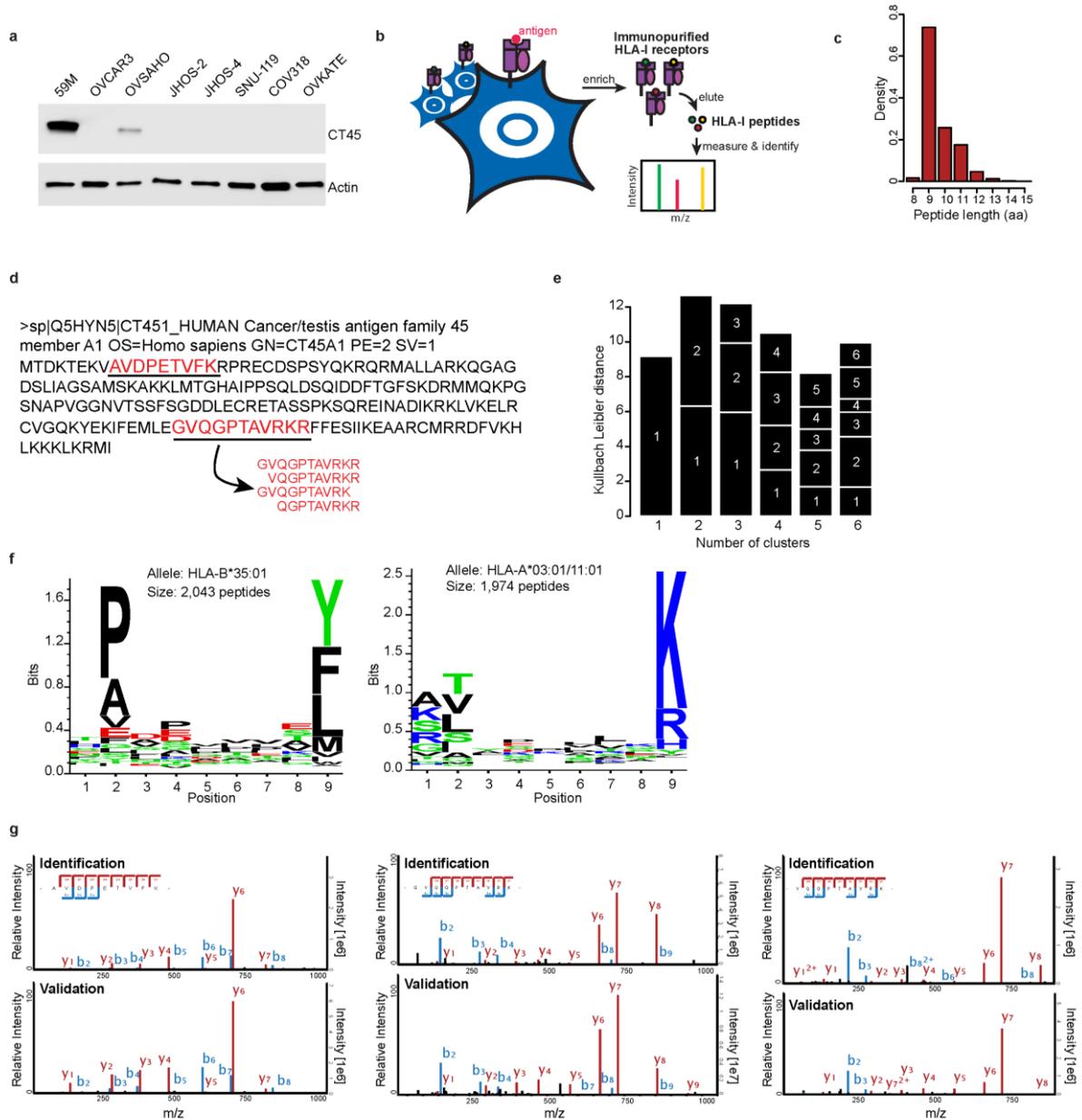
	CT45 Status				P value
	0 (N=82)		1+ (N=42)		
	No.	%	No.	%	
<b>Stage</b>					0.72
3	56	68.29	30	71.43	
4	26	31.71	12	28.57	
<b>Primary Site</b>					0.41
Fallopian Tube	10	12.2	9	21.43	
Ovary	62	75.61	29	69.05	
Pentoneum	10	12.2	4	9.52	
<b>Residual Disease</b>					0.10
Larger than 1 cm	47	57.32	17	40.48	
Smaller than 1 cm	34	41.46	25	59.52	
Unknown	1	1.22	0	0	
<b>Grade</b>					0.59
1	1	1.22	0	0	
2	22	26.83	8	19.05	
3	59	71.95	34	80.95	
<b>Chemo Type</b>					0.67
Adjuvant/1st line	64	78.05	35	83.33	
Neoadjuvant	13	15.85	4	9.52	
None	5	6.1	3	7.14	
<b>Chemo Class</b>					0.19
Other	4	4.88	6	14.29	
Taxane/Platinum	73	89.02	33	78.57	
None	5	6.1	3	7.14	
<b>Chemoresistance</b>					0.005
resistant	46	56.1	10	23.81	
intermediate	11	13.41	9	21.43	
sensitive	19	23.17	18	42.86	
N/A	6	7.32	5	11.9	
<b>Mean age (range)</b>	62	(33-94)	58	(33-82)	0.13

b

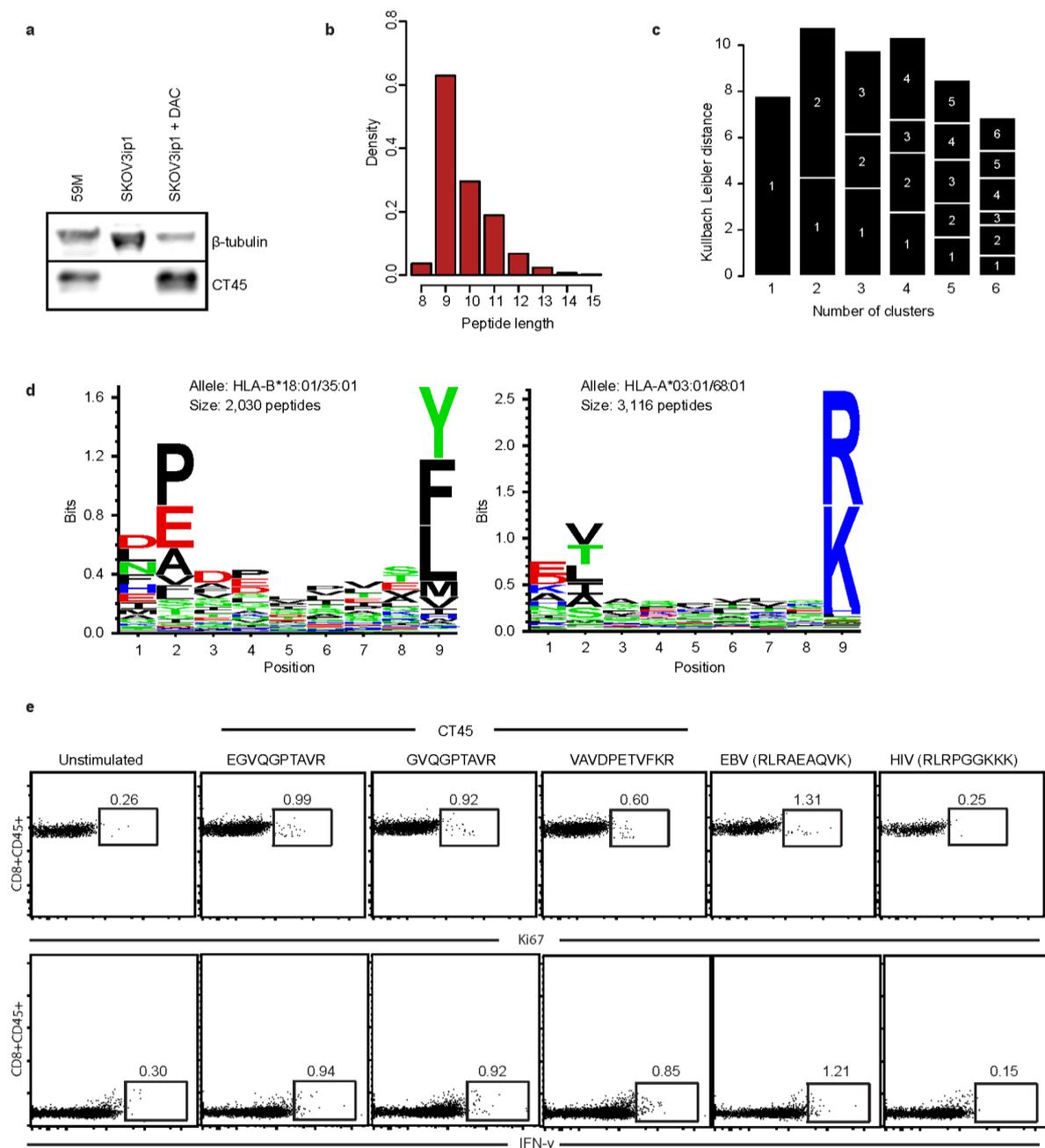


### Extended Data Figure 3. Patient characteristics of tissue microarray validation cohort.

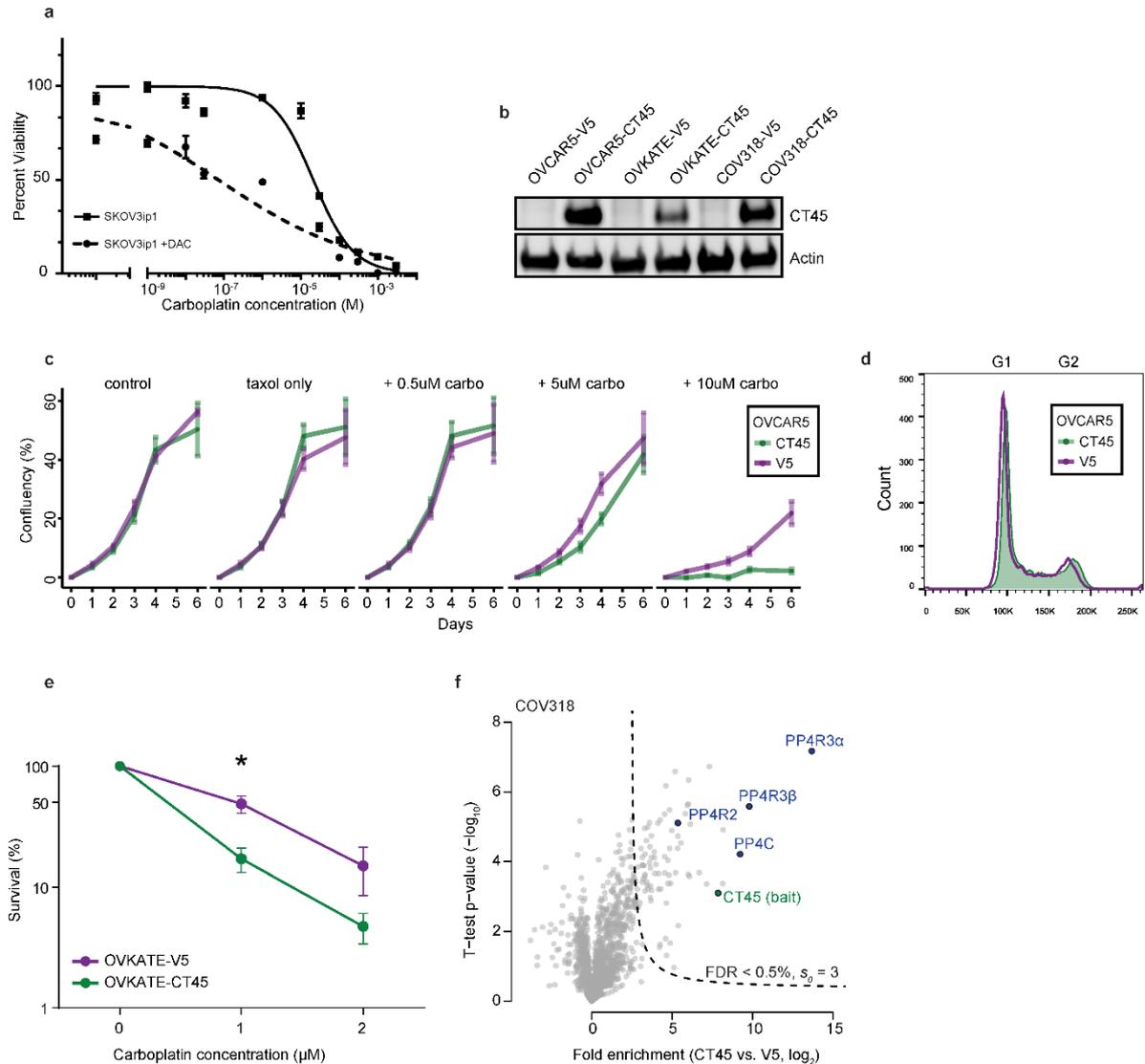
**a**, Clinico-pathologic parameters for patients with advanced stage serous papillary ovarian cancer included in the TMA validation cohort. Tumors were stained for CT45 and the staining was scored 0-3. **b**, Kaplan-Meier survival analysis for overall survival based on CT45 staining scores from ovarian cancer TMAs. Advanced stage HGSOC patients comparing a staining score of 0 (N=82) versus 1+ (N=42) are compared.



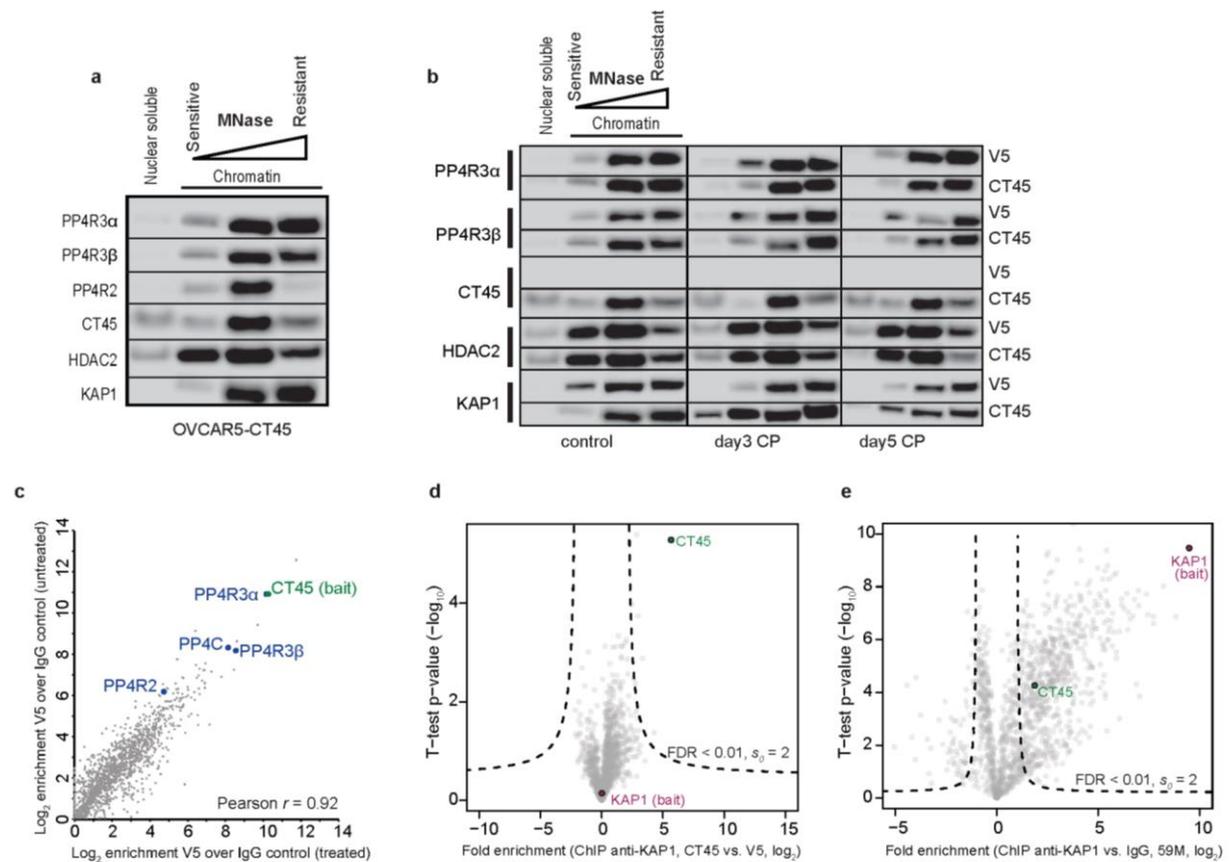
**Extended Data Figure 4. HLA-I peptidomics reveals presentation of CT45-derived peptides.** **a**, Western blot of CT45 protein expression in 8 ovarian cancer cell lines. **b**, Schematic of immuno-peptidomics strategy to identify HLA-I binding peptides. HLA-I receptors with peptide complexes are pulled down with an antibody specific for HLA-I. Peptides are eluted from the HLA complex and identified by mass spectrometry. **c**, Histogram of the identified peptide lengths from the 59M cell line (length 8-11 amino acids) consistent with peptides that bind to HLA-I receptors. **d**, CT45A1 protein sequence. Identified HLA-I peptides are highlighted in red. **e**, Identified consensus clusters based on the GibbsCluster -1.0 tool for all identified 9-mer peptides (4,017) **f**, Binding motifs. HLA-I consensus binding motifs for the two largest clusters in **e**. The A11:01 motif is similar to A3:01. **g**, Comparison of MS/MS scans from experimentally identified HLA-I peptides (upper panel) and synthetic versions of the same peptides (lower panel). Y (red) and b-ions (blue) are shown.



**Extended Data Figure 5. CT45-derived peptides are presented after DAC treatment.** **a**, Western blot of CT45 protein expression in SKOV3ip1 (HLA-A\*03:01, 68:01) following DAC treatment. 59M serves as positive control. **b**, Histogram of the identified peptide lengths from the DAC treated SKOV3ip1 cell line. **c**, Identified consensus clusters based on the GibbsCluster tool for all identified 9-mer peptides (5,146). **d**, HLA-I consensus binding motifs for the two largest clusters identified in **c**. The A68:01 motif is similar to A3:01. **e**, Stimulation of CD8<sup>+</sup> tumor infiltrating HLA-A\*03:01<sup>+</sup> T cells isolated from ascites with CT45-derived peptides. Staining for Ki-67 and IFN- $\gamma$  of CD8<sup>+</sup> T cells (A-03:01) after stimulation with 3 CT45 peptides (EGVQGPTAVR, GVQGPTAVR, and VAVDPETVFKR) or an EBV positive control or HIV negative control peptide analyzed with flow cytometry.



**Extended Data Figure 6. CT45 sensitizes cancer cells to carboplatin.** **a**, MTT of SKOV3ip1 cells with or without 500nM DAC treated with increasing doses of carboplatin. Cells were pretreated with 500nM DAC for 72hr then cultured for 4 days and treated with carboplatin on day 7. The MTT was performed after 72hr with carboplatin treatment. **b**, Western blot of CT45 expression in ovarian cancer cell lines transduced with a CT45A5 lentivirus. **c**, Proliferation of OVCAR5 control (V5) or CT45 expressing cells with increasing concentrations of carboplatin. **d**, Cell cycle analysis of OVCAR5 control (V5) or CT45 expressing cells stained with propidium iodide. **e**, Clonogenic survival assay of the ovarian cancer cell line OVKATE stably overexpressing CT45 (green) or control (purple). Dots represent mean values from three independent experiments. Error bars show s.e.m. for each group. **f**, Interaction proteomics screen in the ovarian cancer cell line COV318 stably over-expressing V5- tagged CT45. Protein enrichment (t-test difference) is calculated over the corresponding control cell line (V5 tag alone) and plotted against the t-test p-value ( $-\log_{10}$ ). Dashed lines indicate significance thresholds ( $p < 0.005$ ,  $s_0 = 3$ ). The bait protein CT45 (green) and members of the protein phosphatase 4 complex (blue) are highlighted. Results represent 3 replicates per experiment group.



**Extended Data Figure 7. CT45 is chromatin-bound independent of DNA damage.** **a**, Chromatin fractionation of the OVCAR5-V5-CT45 cell line. **b**, Chromatin fractionation of the OVCAR5-V5-CT45 cell line carboplatin **c**, ChIP-MS results targeting V5-tagged CT45 in the OVCAR5-V5-CT45 cell line  $\pm$  carboplatin. Fold enrichment is calculated over an IgG control antibody. The bait protein CT45 (green) and members of the protein phosphatase 4 complex (blue) are highlighted. Results represent 3 replicates per experiment group. **d**, **e**, ChIP-MS results targeting KAP1 in the OVCAR5 cell line pair (**d**), and the 59M cell line (**e**). Protein enrichment (t-test difference) is calculated over the corresponding control (V5 control cell line for **d** or an IgG control antibody for **e**), and plotted against the t-test p-value ( $-\log_{10}$ ).

### **Acknowledgements**

We thank Ani Solaki for assistance with tail-vein injections and Shawn Pan for assistance with the comet assay (University of Chicago). We also thank Dirk Wischniewski for laboratory support and scientific discussions and Korbinian Mayr, Igor Paron, and Gabriele Sowa for their assistance in mass-spectrometric analysis (Max Planck Institute). We further acknowledge the Human Tissue Resource Center for immunohistochemistry funded by the University of Chicago Cancer Center Support Grant (P30CA014599), the NIH Tetramer Core Facility for supplying CT45 tetramers and Hans-Jürgen Heidebrecht (Kiel, Germany) for CT45 antibody. This work was supported by the National Cancer Institute grant CA111882 (EL), a Harris Family Foundation award (SDY), the Körber Foundation (Körber European Science Prize), the Max-Planck Society for the Advancement of Science (MM), and the Ovarian Cancer Research Fund Alliance (Ann Schreiber Mentored Investigator Award) (MC).

### **Author contributions**

F.C., M.C., and B.A. performed the majority of experiments. F.C., M.M., E.L., and M.C. conceived and designed experiments. F.C. performed mass spectrometry and data analysis. J.D. performed immune function experiments. M.B-S. assisted with immunopeptidomics experiments and data analysis. M.W. assisted with ChIP-MS experiments. A.J. performed experiments. A.M. scored tissue microarrays. K.W. performed statistical analysis. S.D.Y. and E.L. collected ovarian cancer tissues and clinico-pathologic patient information. M.C., F.C., M.M., and E.L. wrote and edited the manuscript. F.C., M.C., M.M., and E.L. directed the study.

### **Author information**

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org/cgi/GetDataset>) via the PRIDE partner repository with the dataset identifier PXD005315 (Username: reviewer10144@ebi.ac.uk, Password: R4mKRdtJ). All other data supporting the findings of this study are available within the article and its extended data files or from the corresponding author upon reasonable request. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). For authors F.C., E.L., M.M., and M.C., the University of Chicago and the Max Planck Society has jointly filed a patent application related to this technology. Correspondence and requests for materials should be addressed to MM ([mmann@biochem.mpg.de](mailto:mmann@biochem.mpg.de)) or E.L. ([elengyel@uchicago.edu](mailto:elengyel@uchicago.edu)).

## METHODS

### Cell lines and reagents

SKOV3ip1 (from Dr. Gordon Mills, M.D. Anderson Cancer Center, Houston, TX), OVCAR5 (UCSF) and COV318 (from Dr. Gottfried Konecny, UCLA) were cultured in DMEM, 10% FBS. 59M (ECACC) was cultured in DMEM, 10% FBS supplemented with 10 $\mu$ g/ml bovine insulin (Sigma, MO). OVKATE (from Dr. Gottfried Konecny, UCLA) were cultured in RPMI-1640, 10% FBS. All cell lines were tested for mycoplasma and authenticated using a commercial service (CellCheck, IDEXX Bioresearch). Growth factor reduced Matrigel was from BD Biosciences (Rockville, MD). pLX304 (205) was a gift from Dr. David Root (Addgene plasmid #25890). pLX304-CT45A5 was acquired from DNASU Plasmid Repository (206-208) (clone HsCD00446210). 3xFLAG-CT45A1 was synthesized (Eurofins) and cloned into pcDNA3.1 (Invitrogen). The Ki-CT45-2 antibody (used for WB and IMF) was a kind gift of Hans-Jürgen Heidebrecht. W6/32 monoclonal antibodies were purified from the growth medium of HB95 cells that were grown in CELLline CL-350 flask (Wilson Wolf Manufacturing Corporation, Minnesota) using Protein-A Sepharose (Invitrogen, CA). Antibodies acquired from Cell Signaling Technology were:  $\gamma$ H2AX (9718, rabbit), Cleaved Caspase-3 (9661), anti-rabbit IgG-HRP (#7074), and anti-mouse IgG-HRP (#7076), normal rabbit IgG (#2729). Antibodies acquired from Bethyl laboratories were: PP4C (A300-835A), PP4R3 $\beta$  (A300-842A), KAP1 (A300-274A), pS824-KAP1 (A300-767A). Other antibodies used were: actin (Sigma, #A5441), anti-CT45A antibody (Sigma, SAB1301842),  $\gamma$ H2AX (mouse, Thermo Scientific, #MA1-2022), and V5 (Life Technologies, #MA5-15253), PP4R2 (Atlas antibodies, HPA034695), PP4R3 $\alpha$  (Atlas antibodies, HPA002568). HLA-I types of cell lines were determined using high-resolution genotyping (Center for Human Genetics and Laboratory Medicine, Martinsried).

### FFPE tissue preparation for MS analysis

Tumors were collected from patients undergoing primary debulking surgery by a gynecologic oncologist at the University of Chicago Hospital, Department of Obstetrics and Gynecology, Section of Gynecologic Oncology. Informed consent was obtained before surgery and the study was approved by the IRB of the University of Chicago. FFPE biobank specimens (5 serial sections, 10 $\mu$ m thick) were first deparaffinized as previously described (123). Areas containing 70% or more tumor were macrodissected from the slide using a scalpel blade. Lysis was then carried out in 4% SDS, 10 mM Hepes pH 8.0 at 99 $^{\circ}$ C for 60 min and by 15 min sonication (level 5, Bioruptor, Diagenode). Proteins in the cleared lysate (16,000 g, 10 min) were reduced with 10 mM DTT for 30 min and alkylated with 55 mM iodoacetamide for an additional 30 min. 100  $\mu$ g of proteins were purified from SDS by acetone precipitation and the protein pellet resolved in 100  $\mu$ l 6 M urea/2 M thiourea (in 10 mM Hepes pH 8.0). LysC digestion was carried out with 1  $\mu$ g of LysC for 3 h at room temperature. After adding 4 volumes of 50 mM ammonium bicarbonate buffer, 1  $\mu$ g trypsin was added for tryptic digestion overnight. The next day, digestion was stopped by adding 1% TFA. Peptides were finally desalted on C18 StageTips and kept at -20 $^{\circ}$ C until MS analysis. The majority of samples were injected twice for MS analysis.

### Liquid Chromatography (LC)-MS analysis of FFPE samples

For LC-MS analysis, a Q Exactive(209) (Thermo Fisher Scientific) mass spectrometer was used coupled on-line to an EASY-nLC 1000 HPLC system (Thermo Fisher Scientific). Desalted peptides were separated on in-house packed C<sub>18</sub> columns (75  $\mu$ m inner diameter, 50 cm length, 1.9  $\mu$ m particles, Dr. Maisch GmbH, Germany) in a 250-min gradient from 2% to 60% in buffer B (80% acetonitrile, 0.5% formic acid) at 200 nl/min. Mass spectra were acquired in data-dependent mode. Briefly, each survey scan (range 300 to 1,650  $m/z$ , resolution of 70,000 at  $m/z$

200, maximum injection time 20 ms, ion target value of 3E6) was followed by high-energy collisional dissociation based fragmentation (HCD) of the 5 most abundant isotope patterns with a charge  $\geq 2$  (normalized collision energy of 25, an isolation window of 2.2  $m/z$ , resolution of 17,500, maximum injection time 120 ms, ion target value of 1E5). Dynamic exclusion of sequenced peptides was set to 45 s. All data was acquired using Xcalibur software (Thermo Scientific).

### **Data analysis of proteomic raw files**

MS raw files were processed with the MaxQuant software (94) (version 1.5.3.15). The integrated Andromeda search engine (95) was used for peptide and protein identification at an FDR of less than 1%. The human UniProtKB database (August 2015) was used as forward database and the automatically generated reverse database for the decoy search. 'Trypsin' was set as the enzyme specificity. We required a minimum number of 7 amino acids for the peptide identification process. Proteins that could not be discriminated by unique peptides were assigned to the same protein group (94). Label-free protein quantification was performed using the MaxLFQ (103) algorithm (MaxQuant environment). Briefly, quantification was based on extracted high-resolution 3D peptide features in mass-to-charge, retention time and intensity space. Only common peptides were used for pair-wise ratio calculations. Protein ratios were then determined based on median peptide ratios. We required a minimum peptide ratio count of 1 to report a quantitative read-out and averaged the results from duplicate measurements of the same sample. The 'Match Between Runs' feature of MaxQuant was enabled to transfer peptide identifications across runs based on high mass accuracy and normalized retention times. Prior to data analysis, proteins, which were found as reverse hits or only identified by site-modification, were filtered out.

### **Tissue microarray**

Tissue microarrays (TMAs) were deparaffinized and rehydrated through xylenes and serial dilutions of EtOH to deionized water. They were incubated in antigen retrieval buffer (Tris-EDTA, pH 9, S2367, DAKO) and heated in steamer at over 97 °C for 20 minutes. Tissue sections were incubated in a humidity chamber with CT45A antibody (1:200, Sigma, SAB1301842) for 1h at room temperature. The antigen-antibody binding was detected by Bond Polymer Refine Detection (DS9800, Leica Microsystems). Tissue sections were briefly immersed in hematoxylin for counterstaining and were covered with cover glasses. The stained TMAs were scored by an expert pathologist on a scale from 0-3. Data acquisition and analysis were blinded. Tumors were collected from patients undergoing primary debulking surgery by a gynecologic oncologist at the University of Chicago Hospital, Department of Obstetrics and Gynecology, Section of Gynecologic Oncology. Informed consent was obtained before surgery and the study was approved by the IRB of the University of Chicago.

### **HLA-I peptidomics**

SKOV3ip1 cells were treated with 500nM 5-aza-2'-deoxycytidine (DAC) (Sigma, MO) for 3 days with DAC refreshed every 24 hours. After treatment cells were cultured an additional 4 days without DAC and collected for HLA-I purification at day 7. 59M cells were cultured under normal conditions prior to collection. HLA peptide purification, mass spectrometric analysis and data analysis were performed as previously described (163). We used the GibbsCluster-1.0 Server tool (210) to perform Gibbs clustering analysis of all identified 9-mer HLA-I peptides as input using the default settings for 1-6 clusters. We compared the resulting motifs to the known and predicted motifs of the HLA-I alleles using the MHC motif viewer (211). Binding motifs were plotted using the Seq2Logo tool (212). The NetMHC 4.0 algorithm (213) was used

to model binding affinity of the identified peptides using default affinity thresholds for peptide binding (% rank 0.5 for strong binders and 2 for weak binders).

### **T cell peptide stimulation**

Tumors and ascites were collected from ovarian cancer patients undergoing primary debulking surgery at the University of Chicago. TILs were thawed and resuspended in RPMI, 10% FCS (or human serum). Cells were adjusted to  $2 \times 10^6$ /ml in 24-well plates and cultured for 7-10 days in the presence or absence of  $1 \mu\text{g/ml}$  peptide. After 3 days, 6IU/ml of rhIL-2 was added to the culture. Flow cytometric analysis of cultured cells was performed.

**Intracellular cytokine staining:** *In vitro* expanded TILs were restimulated at the end of culture with the relevant peptide ( $100 \text{ ng/ml}$ ) in the presence of Brefeldin A (BD Biosciences) for about 5-6 hours at  $37^\circ\text{C}$  in a 5%  $\text{CO}_2$  incubator. The cells were washed once with FACS buffer and stained with surface makers (anti-CD3, -CD8, live/dead stain) for 30 min on ice. Then the cells were fixed and permeabilization with the Cytofix/Cytoperm kit (BD Biosciences), and stained for cytokines using either anti-IFN- $\gamma$  antibody. After washing, the cells were analyzed on a flow cytometer.

**Tetramer staining:** *In vitro* expanded TILs were stained with PE-labeled MHC Class I tetramers along with surface staining antibodies (anti-CD3, -CD8, live/dead stain) for 1 hour on ice. MHC class-I tetramers carrying CT45 antigen (GVQGPTAVRK AVDPETVFK) or HIV peptides (RLRPGGKKK or QVPLRPMTYK) were used. After washing, the cells were analyzed on a flow cytometer.

**$^{51}\text{Cr}$  release assays:** The 59M tumor cell line (target cells) was labeled with  $100 \mu\text{Ci}$   $^{51}\text{Cr}$  at  $37^\circ\text{C}$  for 1 hour. Target cells were then washed three times in PBS, resuspended in culture medium at  $1 \times 10^5$  viable cells/ml and  $100 \mu\text{l}$  was added per well of a 96-well U-bottom plate. *In vitro* expanded TILs (effector cells) were washed twice in culture medium and added to targets at the given ratios. Plates were shortly centrifuged to settle cells, and incubated at  $37^\circ\text{C}$  in a 5%  $\text{CO}_2$  incubator for 5 hours after which time the supernatants were harvested, transferred to small tubes and counted using a Liquid Scintillation Counter. Spontaneous  $^{51}\text{Cr}$  release was evaluated in target cells incubated with medium alone. Maximal  $^{51}\text{Cr}$  release was measured in target cells incubated with zap solution at a final concentration of 1% (v/v). Percent specific lysis was calculated as (experimental - spontaneous lysis/maximal - spontaneous lysis) times 100.

### **Clonogenic survival assay**

Cells were treated with different carboplatin concentrations for 48-72 hours. 2000 cells per 6-well were then plated for 7-9 days in drug-free medium. Grown colonies were fixed and stained with 1% formaldehyde, 1% methanol and 0.05% Crystal Violet for 20 min. Colony numbers were then counted with the ColonyArea ImageJ plugin and plotted as percent of control on a logarithmic scale.

### **Western blot analysis**

Cells were treated with indicated drugs for 3 days and then media was changed. On indicated day of collection, both adherent and non-adherent cells were collected in RIPA buffer (25mM Tris-HCl pH 7.6, 150mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with protease inhibitor cocktail (Thermo Scientific) and phosphatase inhibitor cocktail (Sigma). Lysates were incubated for 30 min on ice, sonicated and cleared by

centrifugation (15min, 14 000 rpm, 4°C). The quantity of protein was determined by BCA reagent (Pierce). The extracts were analyzed by SDS-PAGE on a 4-20% gradient gel (Bio-Rad).

### **Comet assay**

Cells were treated with carboplatin for 3 days and then incubated an additional 2 days without drug. The comet assay was performed on day 5 as previously described(214). Briefly, 2.5e4 cells/ condition were resuspended in 70uL 0.5% low melting point agarose (LMPA) at 37C and plated on a glass slide precoated with 1% agarose in PBS. Samples were allowed to solidify at 4C for 20 minutes with a square cover glass. The cover class was gently removed and a second 70uL layer of 0.5% LMPA was applied and again allowed to solidify. The cover glass was removed and the slides immersed in Comet Lysis solution (2.5M NaCl, 100mM EDTA, 10mM Tris, 0.015% Triton X-100, pH 10) for one hour at 4C and from this point forward protected from light. After lysis, the slides were equilibrated for 20 minutes in Comet Electrophoresis Buffer (0.3N NaOH, 1mM EDTA, pH 10) at 4C, and then run at 25V for 20 minutes. Then they were incubated at room temp in Comet Neutralization Buffer (0.4M Tris, pH 7.5) for five minutes twice, then in ddH2O for three minutes, stained with a 1:10,000 dilution of SYBR Gold in ddH2O, and finally mounted with a glass cover slip for imaging. Images were taken at 10X using a Zeiss AxioObserver A.1. At least 100 cells were quantified/ sample using the software OpenComet (215). Data shown is the mean  $\pm$  s.e.m. of 4 biological repeats.

### **Mouse experiments**

Five million OVCAR5-V5 or OVCAR5-V5-CT45 cells suspended in a 1:2 solution of serum-free media to growth-factor reduced matrigel were injected subcutaneously into the right and left flanks respectively, of 8 weeks old female athymic nude mice. After 5 days, treatment was administered through the tail vein 1 time/week at 20mg/kg. Sterile water served as the control treatment. Tumor growth was measured every 2-3 days using calipers until the tumor neared 1 cm<sup>3</sup> and was measured daily. Once the tumor reached 1 cm<sup>3</sup> the mouse was sacrificed. 4 mice were removed from the study early due to ulcerations of the skin. All experiments were approved by the Institutional Animal Care and Use Committee of the University of Chicago.

### **Affinity Purification and Mass Spectrometry**

Affinity purification coupled to mass spectrometry (AP-MS) was performed as previously described (216). For immunoprecipitation of cell lines ectopically expressing N-terminally tagged CT45A1 (3xFLAG tag) or C-terminally tagged CT45A5 (V5 tag), 30 ul of ANTI-FLAG M2 Affinity Gel (Sigma) or ANTI-V5 Affinity Gel (Sigma, CLONE V5-10), respectively, was used and incubated with 1 mg of total lysate overnight at 4°C. After three washing steps in washing buffer (150mM NaCl, 50mM Tris (pH 7.5), 5% glycerol, 0.05% IGPAL-CA-630) and three washing steps in a buffer containing 150mM NaCl, 50mM Tris (pH 7.5), 5% glycerol, elution was carried out by a partial on-bead digest. Peptides were finally desalted with C18 StageTips prior to MS analysis.

### **Chromatin experiments**

For chromatin immunoprecipitation coupled to mass spectrometry (ChIP-MS), freshly harvested cells were crosslinked with 1% formaldehyde for 10 min in PBS. Cells were lysed in IP Buffer (50 mM Tris-HCl (pH 8), 100 mM NaCl, 5mM EDTA (pH 8), 0.3% SDS, 1.7% Triton-X-100, supplemented with EDTA-free protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktail (Roche)) and chromatin sonicated to an average size of 200-400 bp. 1 mg of total cell lysate was incubated with 3  $\mu$ g of respective antibody overnight at 4°C under constant rotation. The next day, 30  $\mu$ l of protein G-coupled agarose beads (Cell signaling technology) were added and incubated for 3h at 4°C under constant rotation. Antibody-bait

complexes were then washed three times with low salt wash buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1% Triton), once with high salt wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 1% Triton) and twice with TBS. Elution was carried out by a partial on-bead digest as recently described(216). Peptides were finally desalted with C18 StageTips prior to MS analysis.

Chromatin fractionation was performed as described previously(194). Soluble proteins and proteins bound to open or compacted chromatin were separated by stepwise increasing the salt and nuclease (MNase) treatment. 1E7 cells were washed with PBS and 1 ml low salt buffer (LSB:10 mM HEPES [pH 7.4], 25 mM KCl, 10 mM NaCl, 1mM MgCl<sub>2</sub>, 0.1 mM EDTA). Pelleted cells were then resuspended in six times the packed cell volume (PCV) of LSB supplemented with protease and phosphatase inhibitor cocktail (Roche). After snap freezing in liquid nitrogen, samples were quickly thawed and immediately centrifuged (10 min at 10,000 rpm). The pellet was resuspended in a volume of high-salt buffer (HSB: 50 mM Tris-HCl [pH 8.0], 5% [v/v] glycerol, 1 mM EDTA, 10 mM MgCl<sub>2</sub>, 400 mM KCl, supplemented with protease and phosphatase inhibitor cocktail), equal to 0.25 V of LSB. After centrifugation at 10,000 rpm (supernatant = nucleoplasmic fraction), the pellet was resuspended in a volume nuclease buffer containing 10 U/ml MNase (NEB) and incubated at 37°C for 10 min and centrifuged for 5 min at 10,000 rpm (supernatant = chromatin fraction 1). The pellet was then resuspended in the same volume of nuclease buffer containing 100 U/ml MNase and incubated another 45 min at 37°C before an equal V of solubilization buffer (nuclease buffer + 2% [v/v] NP-40, 2% [v/v] Triton X-100, 600 mM NaCl) was added. After brief vortexing, samples were centrifuged (5 min, 10,000 rpm) and the supernatant collected (chromatin fraction 2). Finally, the pellet was resuspended in a volume of solubilization buffer and an equal V of denaturing buffer (50 mM Tris [pH 6.8], 1% [v/v] SDS, 100 mM DTT, 10% glycerol), briefly sonicated, boiled for 5 min, and centrifuged for 5 min at 10,000 rpm (supernatant = chromatin fraction 3).

Chromatin relaxation was assayed using the micrococcal nuclease (MNase) assay as previously described (197), with a few modifications. Briefly, nuclei from 1E6 cells were extracted with 300 µl ice-cold lysis buffer (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 3 mM MgCl<sub>2</sub> and 0.4% IGPAL-CA-630) on ice for 5 min. After centrifugation (2,000g for 5 min at 4°C), nuclei were washed two times in lysis buffer and one time in 500 µl digestion buffer (0.32M sucrose, 50mM Tris-HCl (pH7.5), 4mM MgCl<sub>2</sub>, 1mM CaCl<sub>2</sub>). Nuclei were then digested in digestion buffer with 50 gel units MNase (NEB) for 9 min at 27°C in a final volume of 100 µl. The reaction was stopped by adding a final concentration of 1% SDS and 15mM EGTA. Genomic DNA was purified and separated by gel electrophoresis (1.2% agarose).

### **Immunofluorescence**

Cells were treated with 0 or 5 µM carboplatin on glass chamber slides. After 72hr, media was changed. At designated timepoints, slides were fixed for 10min at room temperature with 4% paraformaldehyde, cells were washed and permeabilized with PBS/0.1% Triton-X for 15min and then blocked for 1hr in blocking buffer (PBS/0.1%Triton-X/0.05% BSA/0.05% goat serum). Cells were washed three times and then were then incubated with primary antibody overnight diluted in blocking buffer at 4°C. Following three washes, cells were incubated with secondary antibody (diluted in blocking buffer) for 1hr at room temperature and then with Hoechst 33342 for 2min. Slides were mounted with ProLong Gold Antifade. Slides were imaged using a Zeiss LSM 510 microscope. Image analysis was performed using CellProfiler (217). A minimum of 100 cells/sample were analyzed. Data shown is the mean  $\pm$  s.e.m. of 4-5 biological repeats.

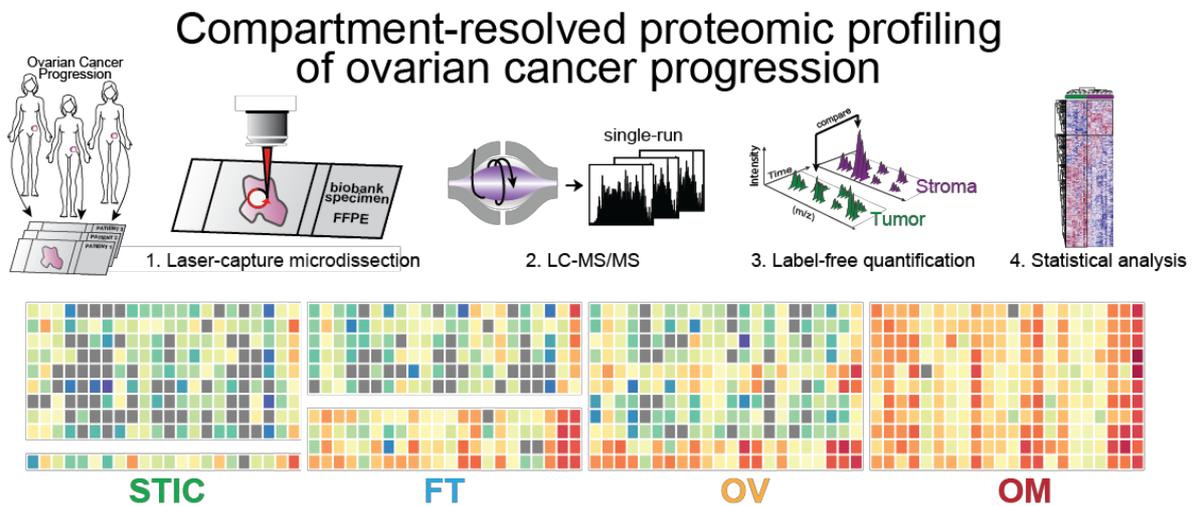
## Statistical analysis

All statistical and bioinformatics analyses were done using the freely available software Perseus (218) (MaxQuant environment), R framework, Stata Version 14 (Stata Corp., College Station, TX) or GraphPad Prism (GraphPad). For pairwise proteomic comparisons, we used a 2-sided t-test statistic including a permutation-based false discovery rate (FDR) of 1% (5% for Fig. 1b after filtering for at least 10 out of 25 valid values) and an  $s_0$  value (219) of 2. Missing values were imputed based on a normal distribution (width = 0.15; downshift = 1.8). For analysis of clinicopathological data, comparisons between groups were performed using chi-squared or Fisher's exact tests for categorical variables and Wilcoxon rank-sum tests for continuous variables. Overall survival and disease-free survival was compared between groups using the log-rank test. The association between CT45 levels and disease-free days was assessed using Pearson correlation. Pathway enrichment analysis (Fig. 4e) was performed based on a Fisher exact test with a Benjamini-Hochberg FDR cutoff of 0.02. GOBP, CORUM and Uniprot Keyword annotations were used for the analysis.

Sample sizes were determined based on previous experience with the individual experiment except for animal studies where power calculations were used. With the exception of the tissue microarray analysis, no randomization or blinding was done for data acquisition or assessment of outcome. The mean and the standard error of the mean (s.e.m) indicating variance are reported for all graphs. For experiments making one comparison, data was analyzed using a two-tailed Mann Whitney U test to account for non-normal distribution of the data. For experiments with more than one comparison, One-Way ANOVA with Tukey's multiple comparisons post-test was used. Before applying ANOVA, we first tested whether the variation was similar among the groups using the Bartlett's test. Where the standard deviations were significantly different, a  $\log_2$  transformation was applied to the data before analysis. Differences were considered significant if  $p < 0.05$ .

## 2.3 Manuscript 3 (in preparation): Compartment-resolved proteomics of ovarian cancer progression

### 2.3.1 Project aim and summary



HGSOC has been characterized at the genomic, transcriptomic and proteomic level (5, 16, 17). This has usually meant analyzing whole tumor tissues taken primarily from the ovaries, which are almost inevitably transformed. However, due to the large variety of different cell types present in the tumor microenvironment, as well as due to contributions from the extracellular matrix, data obtained from whole-tissue analyses inherently result in a mixture of different cell type specific read outs, which may limit data interpretation or at worst lead to incorrect conclusions. Consequently, whole-tissue analysis usually requires further orthogonal validation to ensure the validity of the identified features. Not surprisingly, known HGSOC subtypes, namely 'differentiated', 'immunoreactive', 'proliferative', and 'mesenchymal' (220, 221), are strongly linked to features representing the complex HGSOC microenvironment. However, given the availability of efficient and streamlined approaches tailored to analyze only the cell type of interest, more accurate results may be obtained than that of classical whole-tissue analyses. Laser-capture microdissection (LCM) allows isolating distinct cell types present in the tissue of interest, which increases the sample homogeneity and accuracy of the analytical read-out. We therefore wished to employ LCM coupled to state-of-the-art proteomic analysis to study the compartment-resolved proteomic composition and progression of HGSOC at high quantitative resolution.

In total, we collected 107 archival (FFPE) biobank samples obtained from 11 patients diagnosed with late stage HGSOC. This for the first time allowed to study the proteomic progression from early neoplastic STIC lesions (see introduction 1.1.1), through primary invasive fallopian tube and ovarian cancer, to omental metastasis, at a depth of in total ~ 7.000 proteins. Surprisingly, we found no shared metastatic proteome changes in the tumor compartments. There was a remarkably high proteomic correlation between primary and metastatic tumors from the same patients, similar to previous observations in studies of colorectal (124) and luminal breast cancer (125). However, by additionally analyzing the proteomic changes in the adjacent tumor microenvironment (stromal compartment), we identified a highly conserved stromal signature of HGSOC metastasis to the omentum. We validated our findings by a candidate-driven follow-up approach, which provided new functional insights into the complex metabolic and epigenetic interplay between HGSOC and its associated microenvironment. Of particular note, Nicotinamide N-Methyltransferase (NNMT), a metabolic enzyme that is associated with reduced S-adenosyl methionine (SAM) levels in cancer cells (222) and with loss of repressive H3K27me3 chromatin marks in differentiating human embryonic stem cells (223), was highly and ubiquitously expressed in the stroma of all omental metastasis specimens. This was further confirmed in a larger patient cohort by using tissue microarray analysis (TMA) encompassing over 200 matched primary (fallopian tube or ovary) and metastatic omental sites. NNMT expression in primary tumor sites was less frequent (35%), however, these patients showed significantly lower disease-free and overall survival time compared to the NNMT stromal negative group. Further functional experiments revealed an important role of NNMT in HGSOC progression as a metabolic regulator for cancer-associated fibroblasts. Interestingly, NNMT abrogation leads to a reversal of the CAF phenotype *in vitro* and was associated with less pronounced tumorigenicity in an orthotopic mouse model. We plan future pre-clinical work to address whether selective inhibition of the enzymatic activity of NNMT could be used in a clinical setup, for example in combination with standard chemotherapy.

### 2.3.2 Contribution

Ernst Lengyel proposed to investigate the ovarian cancer microenvironment by proteomics to Matthias Mann and myself in October 2012. Matthias Mann, Ernst Lengyel (University of Chicago), Mark Eckert (University of Chicago), and I initiated this project. I developed and optimized a highly sensitive and streamlined sample preparation method for the analysis of low input samples (few thousand laser-microdissected tumor cells) obtained from FFPE biobank material. I then acquired all proteomic measurements of the study, analyzed and interpreted the data. This identified a conserved stromal protein signature of OvCa metastasis to the omentum. Mark Eckert (Univ. of Chicago) collected all tissue samples and worked on the follow-up experiments. I designed all proteomics related figures and tables, and wrote the manuscript together with Mark Eckert, Ernst Lengyel and Matthias Mann.

### 2.3.3 Publication

## **Compartment-resolved proteomics reveals NNMT as a master metabolic regulator of cancer associated fibroblasts**

Mark A. Eckert<sup>1,\*</sup>, Fabian Coscia<sup>2,\*</sup>, Agnieszka Chryplewicz<sup>1</sup>, Jae Won Chang<sup>3</sup>, Shawn Pan<sup>1</sup>, Samantha M. Tienda<sup>1</sup>, Ricardo R. Lastra<sup>4</sup>, Marion Curtis<sup>1</sup>, S. Diane Yamada<sup>1</sup>, Ruth Perets<sup>5</sup>, Stephanie M. McGregor<sup>4</sup>, Raymond E. Moellering<sup>3</sup>, Matthias Mann<sup>2</sup>, Ernst Lengyel<sup>1</sup>

<sup>1</sup>Department of Obstetrics and Gynecology/Section of Gynecologic Oncology, University of Chicago, Chicago, Illinois 60637, USA.

<sup>2</sup>Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried 82152, Germany.

<sup>3</sup>Department of Chemistry, University of Chicago, Chicago, Illinois, 60637, USA.

<sup>4</sup>Department of Pathology, University of Chicago, Chicago, Illinois 60637, USA.

<sup>5</sup> Division of Oncology, Clinical Research Institute at Rambam, Rambam Health Care Campus, Haifa, 31096, Israel.

\*These authors contributed equally to the work.

## ABSTRACT

High grade serous carcinoma (HGSC) has a poor prognosis primarily due to its early dissemination throughout the abdominal cavity. Although genomic and proteomic approaches have provided snapshots of the proteogenomics of ovarian cancer (OvCa) (224, 225), a systematic examination of both the tumor and stromal compartments is critical to understanding OvCa metastasis. We therefore developed a label-free proteomic workflow to analyze as few as 5,000 microdissected cells from each compartment. The tumor proteome was stable during the progression from *in situ* lesions to metastatic disease; however, the stroma was characterized by a highly conserved proteomic signature associated with metastasis. This signature prominently included the methyltransferase enzyme nicotinamide N-methyltransferase (NNMT). Functionally, stromal NNMT expression was necessary and sufficient for multiple functional aspects of the cancer associated fibroblast (CAF) phenotype, including the expression of CAF markers and the secretion of cytokines and oncogenic extracellular matrix. Stromal NNMT supported OvCa migration, proliferation, and *in vivo* growth and metastasis. Mechanistically, expression of NNMT in CAFs led to a depletion of S-adenosyl methionine (SAM) and a reduction in histone methylation associated with gene expression changes in the tumor stroma. This work supports the use of ultra-low input proteomics to identify candidate drivers of disease phenotypes and reveals that NNMT is a central, metabolic regulator of CAF differentiation and cancer progression in the stroma.

## INTRODUCTION

While almost all serous OvCa harbor *TP53* mutations (37), other recurrent mutations are rare and serous cancers are generally characterized by chromosomal instability (224, 226). HGSC has a relatively high proportion of stroma, but little is known about how interactions between the cancer cells and the surrounding extracellular microenvironment regulate tumor growth. Several reports describe the proteome of human ovarian cancer (173, 225, 227, 228) but none have differentiated between proteins in the stroma and the epithelial components of the tumor. Given that the stroma has a tumor-supporting role and co-evolves with the epithelial compartment during progression and metastasis (229-232), we set out to evaluate the proteome of both compartments in a systematic fashion. We reasoned that developments in mass spectrometry (MS)-based proteomics (172), particularly in ultra-high sensitivity analysis (233), could be combined with microdissection technology to obtain a more accurate and finely-resolved picture of cancer progression.

## RESULTS

To elucidate the proteomic changes underlying OvCa progression in both the tumor and stroma, we identified a cohort of 11 patients which supplied access to HGSC tissue representing serous tubal *in situ* carcinoma (STIC), invasive fallopian tube (FT) lesions, invasive ovarian (Ov) lesions, and omental (Om) metastases (**Supplementary Table S1; Fig. 1a**). All tissues were collected prospectively during initial debulking surgery and all patients were chemotherapy naïve. For each patient and every anatomic site, both tumor and stromal compartments were microdissected and proteins extracted using an optimized high sensitivity, label-free proteomic workflow for low-input samples. We applied an MS1-based quantification method that enables quantification even when many of the peptide signals are insufficient for fragmentation (**Methods**). In total, we obtained 7,515 unique protein quantifications from 107 analyzed samples, both tumor and benign, at a protein and peptide false-discovery rate (FDR) of less than 1% (94). A median of 5,080 and 4,511 proteins were quantified per tumor or stromal sample, respectively, and at similar dynamic ranges (**Supplementary Table S2, Fig. 1b, and Extended Data Fig. 1a**). Independently microdissected and processed specimens had excellent reproducibility (Pearson  $r = 0.986$ ) (**Extended Data Fig. 1b**) and unsupervised hierarchical clustering robustly segregated tumor and stromal proteomes (**Extended Data Fig. 1c**). Tumor compartments were enriched for known markers of HGSC (e.g., PAX8, EPCAM, mucin-16, and folate receptor alpha) while stromal compartments were characterized by expression of ECM components and activated fibroblast markers (e.g., collagens, fibronectin (FN1), smooth muscle actin (SMA), and myosins) (**Supplementary Table S2, Fig. 1c and Extended Data Fig. 1d**). Pathway enrichment analyses revealed an over-representation of pathways involved in DNA replication and epithelial differentiation in the tumor compartment (**Extended Data Fig. 2a**) while the stromal compartment was dominated by extracellular matrix and transforming-growth factor beta (TGF- $\beta$ ) signaling pathways (**Extended Data Fig. 2b**).

We hypothesized that proteins differentially expressed between anatomic sites would reveal functional effectors in both tumor and stromal cells associated with disease progression and metastasis. However, unsupervised hierarchical clustering of tumor compartments revealed no conserved proteomic signatures that correlated with anatomic sites, and only one protein (FABP4; upregulated in omental metastases and expressed at the tumor-stromal interface) was differentially expressed in any anatomic site to the depth of our analysis (FDR < 0.01, **Fig. 1d and Extended Data Fig. 3a-c**). Indeed, the tumor compartment was characterized by patient-

specific protein signatures (ANOVA FDR < 0.01, **Extended Data Fig. 3a, c**) that likely reflect the molecular heterogeneity of HGSC between different patients (224, 225). In contrast, clustering of stromal samples led to a clear grouping of all omental samples, suggesting a conserved stromal response associated with metastasis across all patients (**Supplementary Table S3** and **Extended Data Fig. 3c**). Moreover, a comparison between the four anatomic sites (STIC, FT, OV, and OM) identified 128 proteins significantly differentially expressed in these stromal compartments (**Extended Data Fig. 3a**). Interestingly, the omental metastasis group contributed most to this list of significant proteins. To identify the drivers with the strongest expression differences between primary and metastatic stroma, we compared all primary (FT or OV) and metastatic (OM) stromal sites in a pairwise fashion. This resulted in a set of 60 proteins which were universally up- or downregulated in all omental metastases (FDR < 0.01, **Fig. 1d**).

The stromal signature consisted of 21 up- and 39 downregulated proteins, including FAP, LOX, TNC, and VCAN, which are proteins known to have tumor-supporting roles in the stroma (**Fig. 1e** and **Extended Data Fig. 3d**). Downregulated proteins included negative regulators of TGF $\beta$  such as LTBP4 (234) and SDPR (235). The 21 upregulated proteins that make up the stromal signature were all highly enriched in the mesenchymal TCGA subtype (**Extended Data Fig. 3e**) (5). Other proteins, such as ENPP1, PYCR1, and COPZ2, had no previously described roles in the tumor stroma or the biology of cancer associated fibroblasts (CAFs). Due to its biochemical activity and roles in epigenetic regulation, upregulation of nicotinamide N-methyltransferase (NNMT) in the omental stroma was an interesting target. NNMT transfers a reactive methyl group from S-adenosyl methionine (SAM) to nicotinamide (NA) to generate S-adenosyl homocysteine (SAH) and the metabolically inert product 1-methyl nicotinamide (1-MNA) (**Fig. 2a**). This activity generates a methyl sink in the form of 1-MNA, that leads to a depletion of SAM and a reduction in the global methylation potential of the cell (222, 236). NNMT-mediated SAM depletion leads to an attenuation of histone and protein methylation in cancer cells, adipocytes, and embryonic stem cells (222, 223, 237-239).

NNMT was universally elevated in the metastatic stroma of OvCa patients, as assessed by proteomics and IHC (**Fig. 2b-c** and **Extended Data Fig. 4a**), and was significantly increased in transformed omental tissue compared to benign omental stroma (**Fig. 2c** and **Extended Data Fig. 4b**). A TMA encompassing over 200 matched primary (FT or Ov) and metastatic OvCa samples confirmed that NNMT protein expression was primarily localized in the stroma of both omental and peritoneal metastases (**Fig. 2c-d**). Tumor expression of NNMT did not vary significantly by anatomic site (**Extended Data Fig. 4c**). In both syngeneic (240) and

autochthonous (241) mouse models of HGSC, NNMT was highly expressed in the stroma of omental metastases (**Extended Data Fig. 4d**). High stromal NNMT was also observed in breast and colon cancer stroma, suggesting that stromal NNMT expression may be a general feature of CAFs in multiple cancer types (**Extended Data Fig. 4e**).

CAFs are differentiated from normal fibroblasts by expression of CAF markers, secretion of oncogenic ECM components, production of pro-tumorigenic cytokines, cytoskeletal rearrangements associated with SMA expression, and increased ability to contract collagenous matrices (230). Knockdown of NNMT in OvCa CAFs led to a reversion of cell morphology to one that more closely resembled normal omental fibroblasts (NOFs) (**Fig. 2e** and **Extended Data Fig. 5a-c**). Knockdown or overexpression of NNMT led to significant perturbation of its activity, as assessed by 1-MNA production using MS (**Fig. 2f**). CAF markers, including SMA and fibronectin, were decreased upon NNMT knockdown and increased with its overexpression (**Fig. 2g** and **Extended Data Fig. 5c-e**). Analysis of TMAs also revealed that high stromal NNMT expression correlated with increased fibronectin levels (**Extended Data Fig. 5f**). Functionally, NNMT knockdown attenuated CAF collagen contractility (**Fig. 2h**) and globally affected CAF gene expression (**Supplementary Table S4** and **Extended Data Fig. 5g**). In particular, gene set enrichment analysis (GSEA) demonstrated a robust and significant enrichment of genes upregulated in the proteomic signature of metastasis and the TCGA mesenchymal subtype, including COMP, FN1, and COL11A1 (**Fig. 2i** and **Extended Data Fig. 5h**).

Because NNMT regulated multiple aspects of the CAF phenotype and significantly altered expression of numerous genes, we hypothesized that high stromal NNMT expression drives gene expression through hypomethylation of DNA, RNA, or histones via attenuation of the SAM/SAH ratio (*i.e.*, methylation potential of the cell) (222). Indeed, methylation potential was directly regulated by NNMT expression, as assessed by mass spectrometry. NNMT knockdown increased the SAM/SAH ratio more than 2.5-fold, while NNMT overexpression led to a greater than 5-fold decrease (**Fig. 3a**). Since global DNA methylation was not significantly increased upon NNMT knockdown (**Extended Data Fig. 6a**), we hypothesized that NNMT-mediated histone hypomethylation drove the CAF phenotype.

Histone H3 lysine methylation and acetylation are tightly coupled to transcriptional activity (242-244). To understand how the observed alterations in methylation potential impact histone methylation, we performed targeted proteomics of histone H3, using a multi-reaction monitoring approach to quantify relative levels of histone methylation (243). NNMT knockdown led to an increase in histone methylation at residues associated with transcriptional

regulation (243, 244), including a significant increase in H3K4 and H3K27 trimethylation (me<sub>3</sub>; **Fig. 3b**). Immunoblotting confirmed that H3K27 trimethylation was perturbed upon knockdown or overexpression of NNMT (**Fig. 3c** and **Extended Data Fig. 6b**). Cartilage oligomeric matrix protein (COMP), an extracellular matrix protein (232), was the most upregulated protein in all metastatic stroma samples of the omentum when compared to the corresponding primary sites (FT and OV) (**Fig. 1d** and **Extended Data Fig. 3c**). COMP is highly expressed in the stroma of omental metastases (**Fig. 3d**) and tightly regulated by NNMT expression (**Fig. 3e**). We therefore investigated histone methylation (H3K27me<sub>3</sub>) at the COMP promoter as a proof-of-concept. Indeed, knockdown of NNMT led to a significant increase in H3K27me<sub>3</sub> occupancy at the COMP promoter, as assessed by chromatin immunoprecipitation (**Fig. 3f-g**). Treatment of CAFs expressing shNNMT with the EZH2 histone methyltransferase inhibitor, DZNep, or the general histone methyltransferase inhibitor, 3DZA, was sufficient to revert the CAF phenotype by restoring expression of CAF markers and promoting collagen contractility (**Fig. 3g-I** and **Extended Data Fig. 6c**) (230).

Functionally, CAFs support and accelerate tumor growth, progression, and metastasis (230, 245, 246). Stromal NNMT regulates the expression of several pro-tumorigenic cytokines (**Supplementary Table S4**) including IL-8, CXCL10, and CCL5 (**Extended Data Fig. 7a**). Overexpression of NNMT in normal fibroblasts promoted cancer cell proliferation in response to conditioned media; conversely knockdown of NNMT in CAFs attenuated cancer cell proliferation (**Fig. 4a**). Extending these findings, knockdown of NNMT in CAFs inhibited cancer cell chemotaxis (**Fig. 4b**). To understand if stromal NNMT activity is required for tumor progression *in vivo*, we co-injected HGSC cells expressing luciferase with CAFs expressing shCtrl or shNNMT constructs. Knockdown of NNMT specifically in the stromal compartment reduced *in vivo* proliferation and overall tumor burden (**Fig. 4c**). To determine if NNMT is sufficient to promote ovarian cancer progression, we utilized an orthotopic model in which ID8 mouse ovarian cancer cells are co-cultured with conditioned media from fibroblasts expressing control or NNMT overexpression constructs for 48 hours before intraperitoneal injection (**Fig. 4d**). Tumor seeding of the omentum was significantly increased when ID8 cells were pre-treated with conditioned media from fibroblasts overexpressing NNMT (**Fig. 4d**).

Although NNMT was primarily expressed in the stroma of omental metastases, a subset of patients had high stromal expression of NNMT in primary sites (37%; ovary and fallopian tube) (**Fig. 2e**). The TMA used to validate stromal NNMT expression was also used to evaluate the prognostic role of NNMT in chemo-naïve, high grade serous cancer (247). We found that elevated stromal NNMT at primary sites was associated with a significantly worse recurrence-

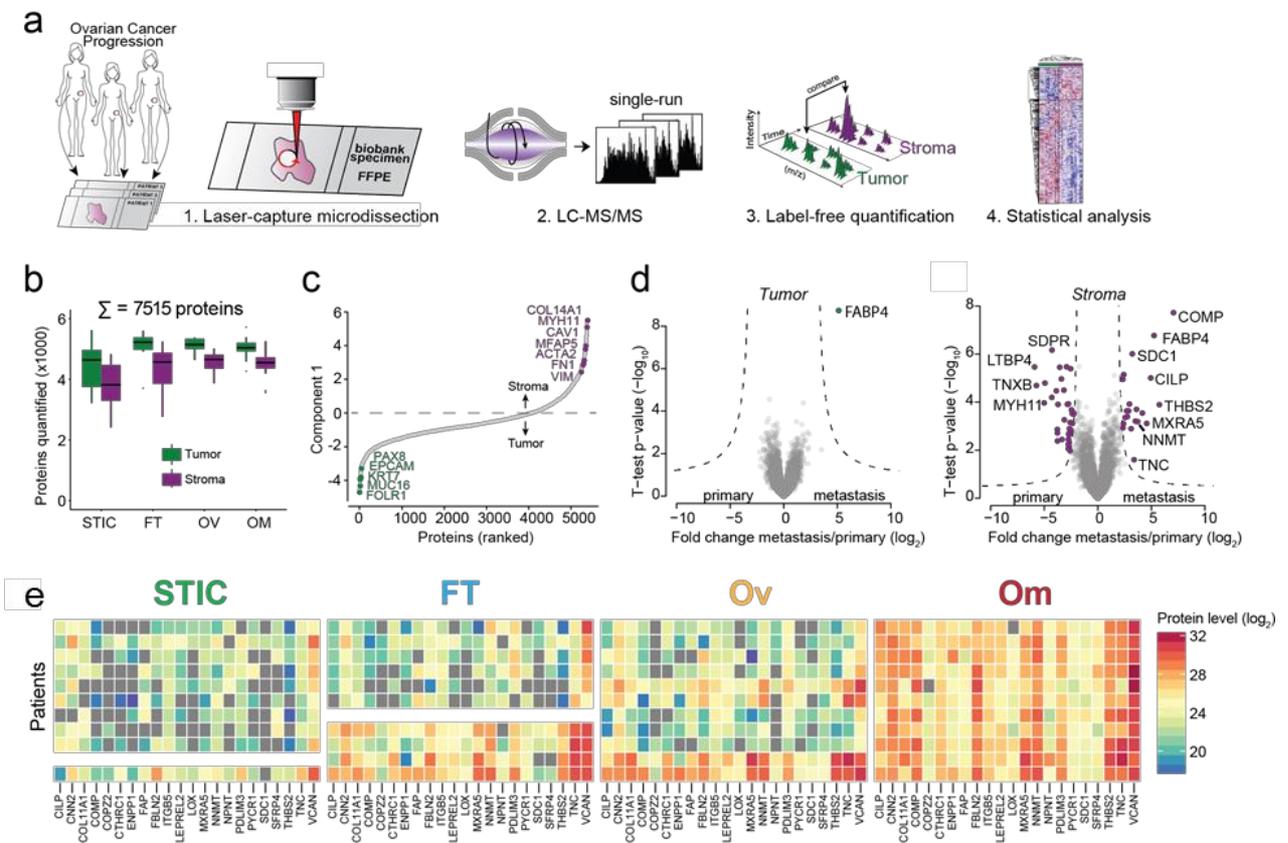
free (median survival of 349 versus 598 days) and overall survival (737 versus 1489 days; **Supplementary Table S5, Fig. 4e, and Extended Data Fig. 7b**). In contrast, expression of NNMT in the tumor compartment was not predictive of survival or recurrence (**Extended Data Fig. 7c-d**). High stromal NNMT expression was significantly associated with platinum resistance in our patient cohort (**Supplementary Table S5**).

## DISCUSSION

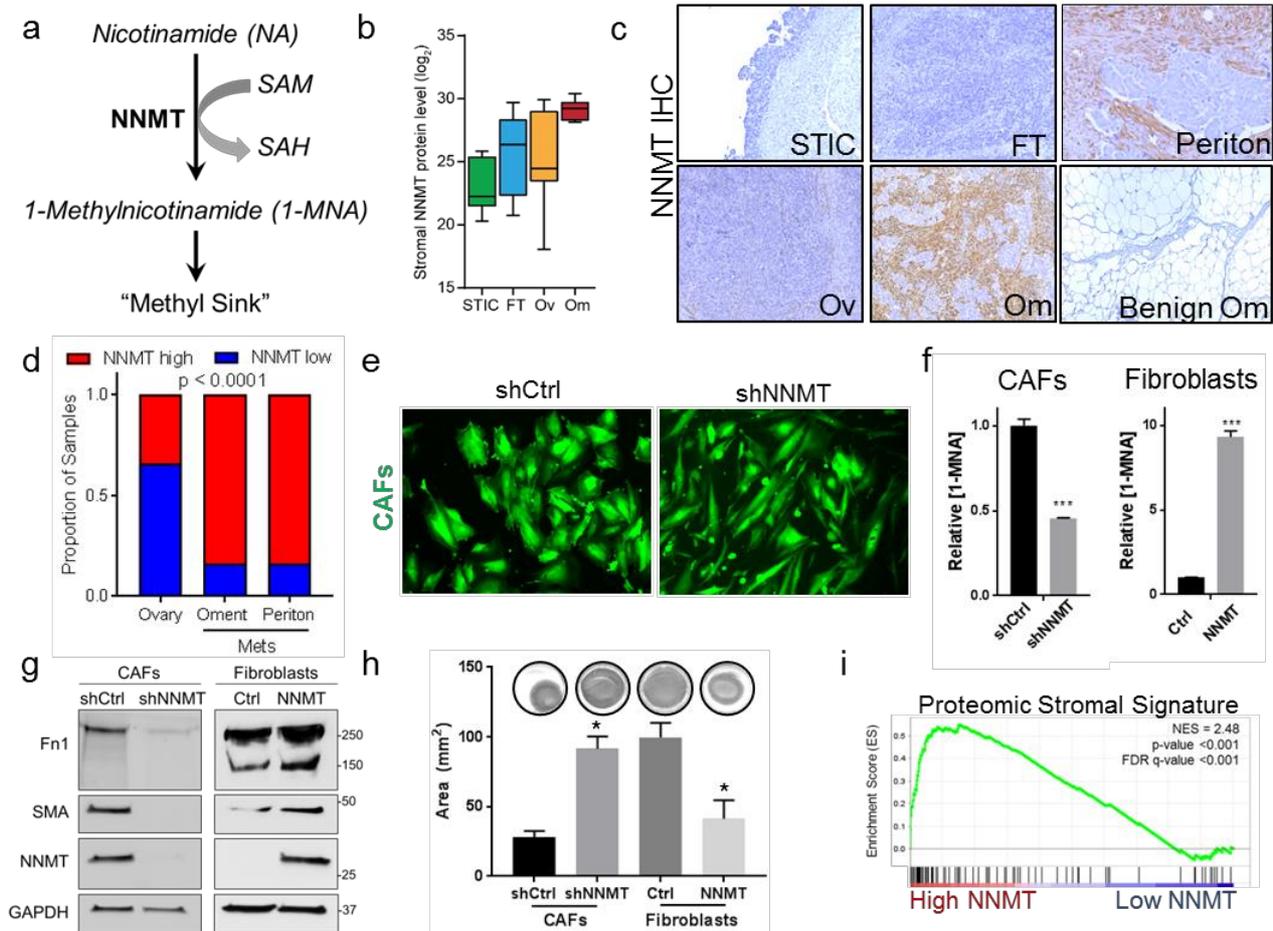
In summary, by using laser-capture microdissection combined with an optimized high-sensitivity proteomic pipeline we demonstrated the quantification of up to 5,000 unique proteins from as few as 5,000 cancer cells. This approach enabled compartment-resolved proteomic analysis of both tumor and stromal compartments across the HGSC progression series from STIC to metastatic tumors and revealed a metastatic stromal signature. Our results emphasize the molecular heterogeneity of ovarian cancer while also revealing that tumor proteomes within individual patients are comparatively stable during progression. In a recent genomic study of a subset of the patient cohort described here, we observed similar genomic alterations across all anatomic sites, including STIC, within each patient (248). Of note, we did not observe a proteomic signature that differentiated STIC from advanced cancers, suggesting STIC already possess the molecular aberrations at both the genomic and proteomic levels associated with a fully developed cancer. Despite the marked genetic and proteomic heterogeneity of epithelial OvCa across patients, the stromal proteome was remarkably uniform and characterized by high NNMT expression.

NNMT metabolically reprograms the epigenome of the stroma to promote OvCa progression in a manner that co-opts NNMT-dependent processes that occur during embryonic stem cell priming (223), metabolic syndrome (237), and tumor cell aggressiveness (**Fig 4f**) (222). The specificity of the observed methylation patterns, marked by H3K27 and H3K4 hypomethylation, is likely due to differences in the affinity of different methyltransferase enzymes for SAM (222, 249, 250). Although CAF gene expression can be regulated by chromatin modifiers and DNA methylation (251, 252), metabolically-defined histone methylation plays a central role in defining the pro-tumorigenic role of the stroma. Inhibition of NNMT activity by knockdown led to a reversion of the CAF phenotype, suggesting stromal methyltransferase activities can be targeted to normalize the tumor stroma. This study suggests

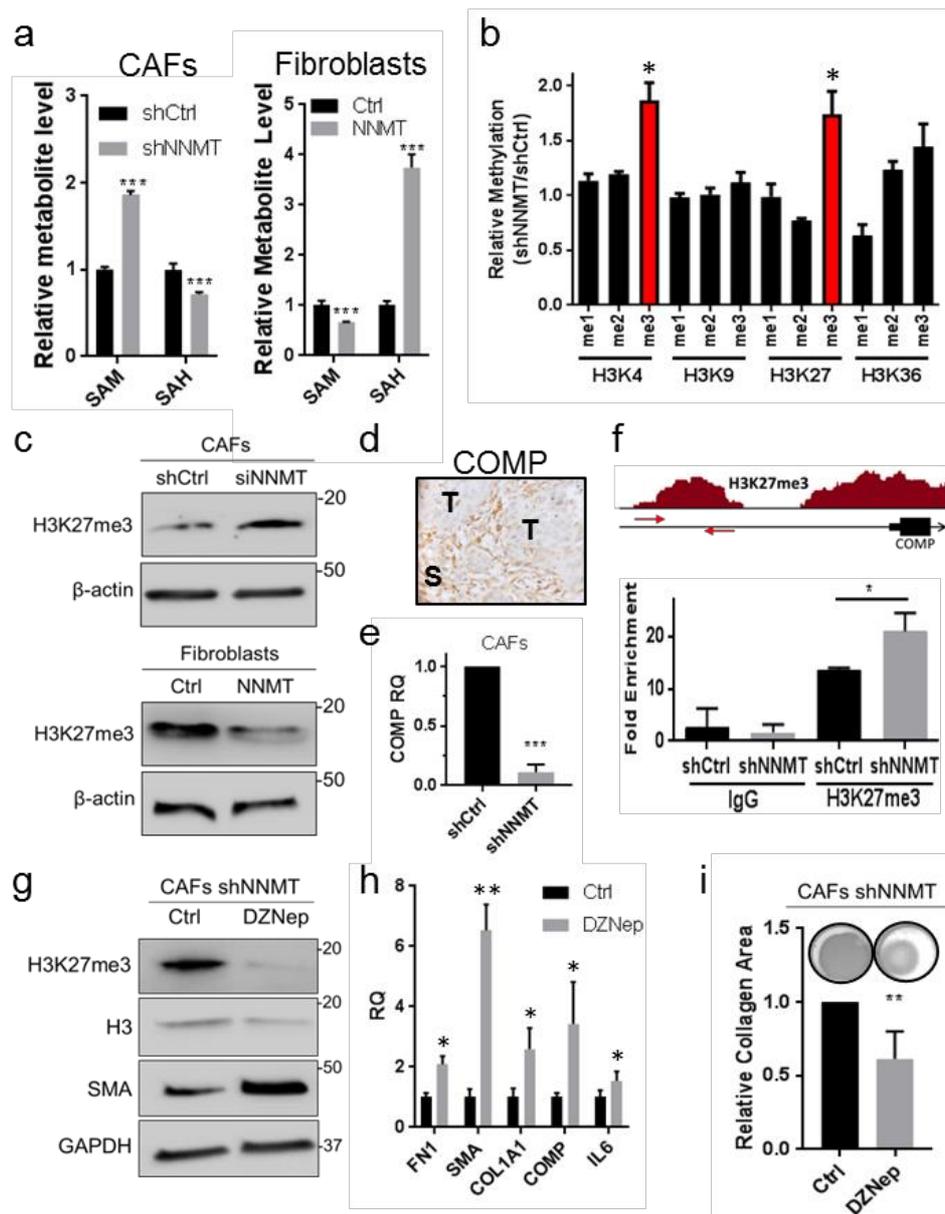
that epigenetic targets in the stroma may play key roles in the response to inhibitors targeting regulators of the epigenome undergoing clinical testing.



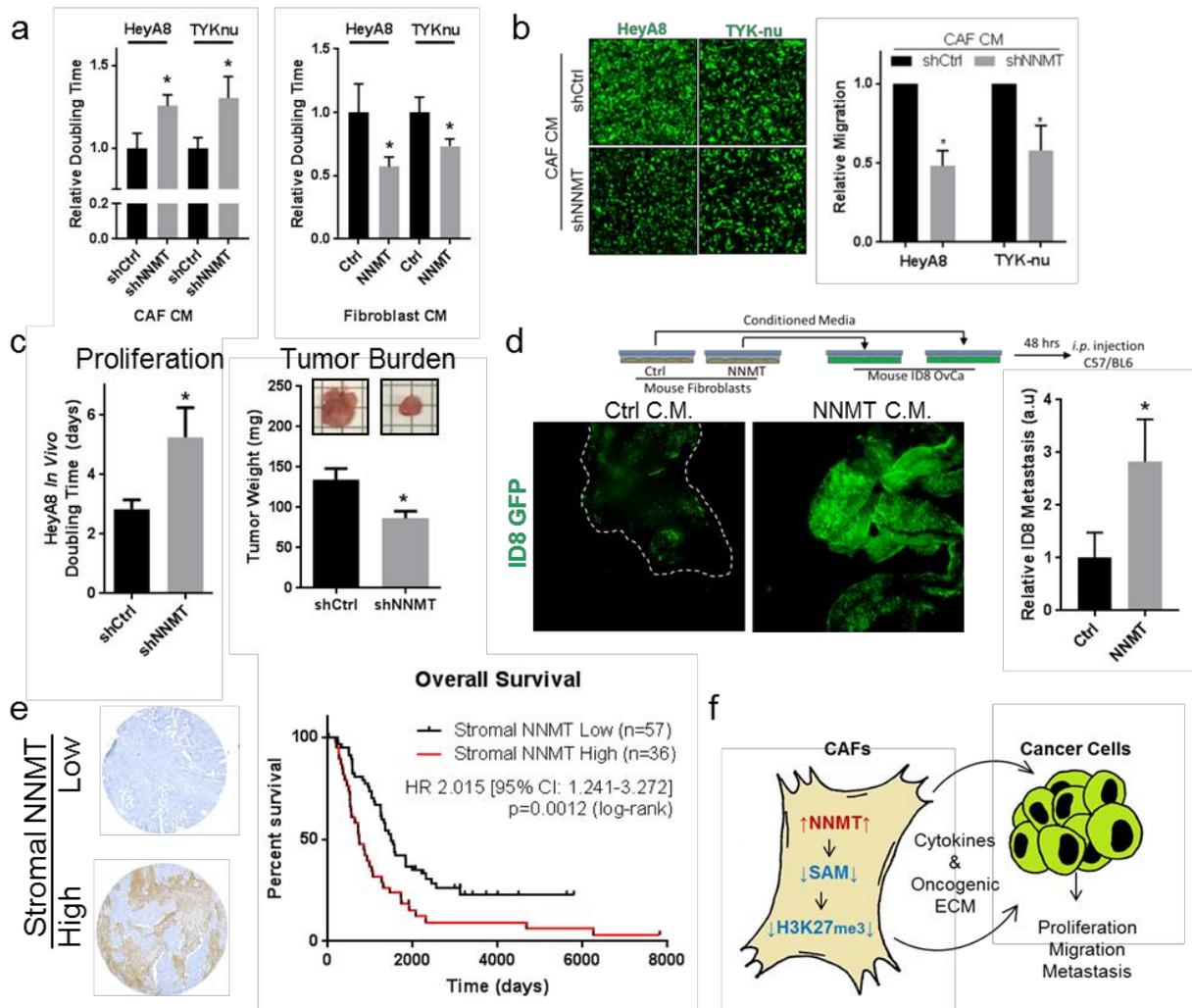
**Fig.1: Compartment-resolved proteomics of ovarian cancer progression reveal a stromal signature of HGSC metastasis.** (a) Tumor and stromal compartments were microdissected from an ovarian cancer progression series (serous tubal intraepithelial carcinoma, STIC; invasive fallopian tube, FT; ovarian lesions, Ov; and omental metastases, Om) and label-free, quantitative shotgun proteomics performed to identify proteins differentially expressed in tumor and stromal compartments across all anatomic sites. (b) Number of unique proteins quantified by MaxLFQ in each anatomic compartment. (c) Ranking of proteins by expression in tumor compartment versus stromal compartment identifies established markers characterizing tumor (green) and stromal (purple) components of the tumor. (d) Volcano plots comparing omental metastases to primary sites (FT and Ov) in tumor (left) and stromal (right) compartments. Significantly differentially expressed proteins are highlighted in green (tumor) or purple (stroma),  $n = 11$  patients. (e) Heatmap of proteins upregulated in omental stromal signature of metastasis (bottom axis) across all patients (rows) and anatomic sites (STIC, FT, Ov, and Om). Warmer colors are higher expression. Missing values are grey; missing samples are white.



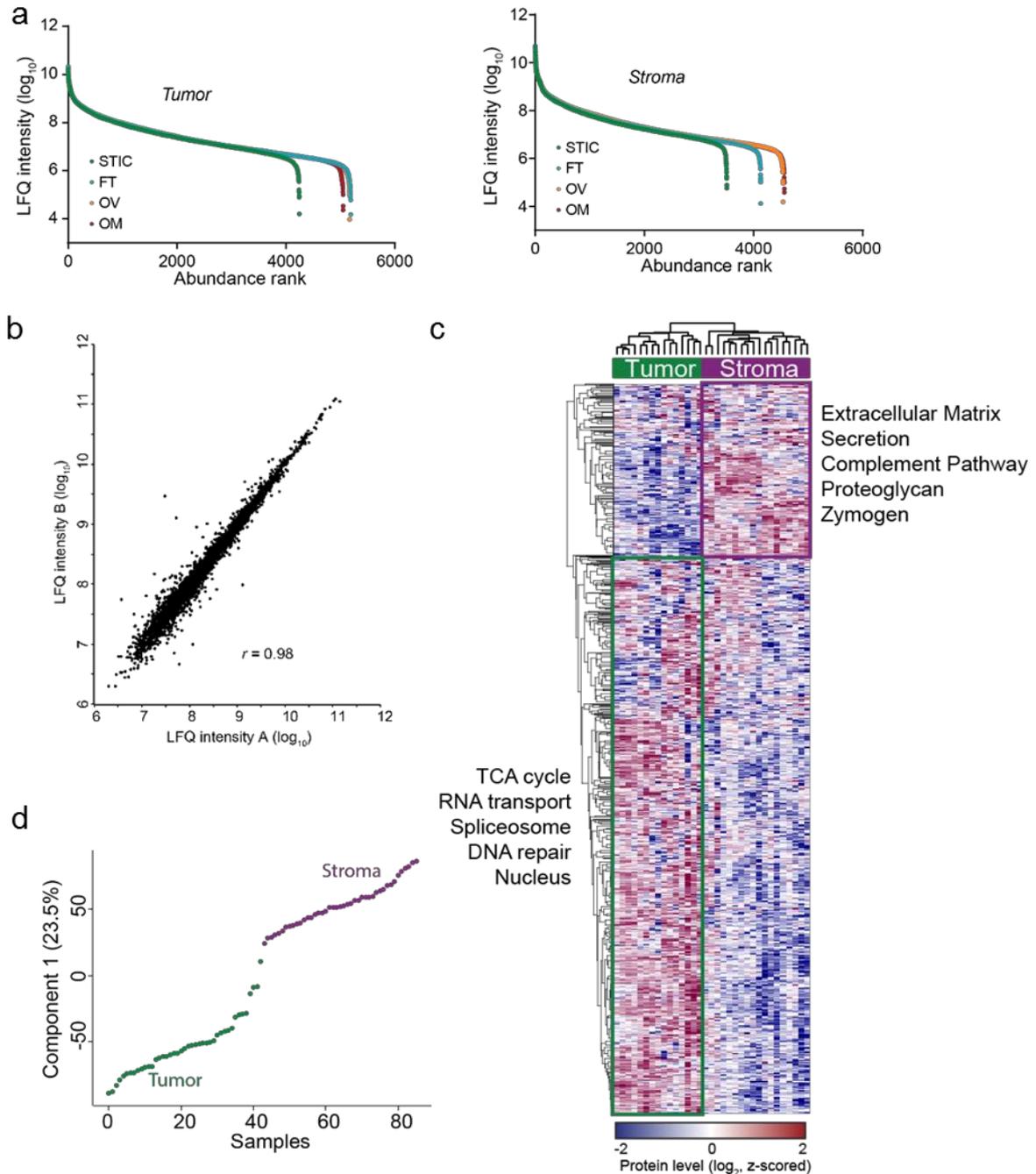
**Fig. 2: NNMT is upregulated in the stroma of HGSC metastases and regulates the CAF phenotype.** (a) NNMT catalyzes the transfer of a reactive methyl group from S-adenosyl-L-methionine (SAM) to nicotinamide (NA), generating S-adenosyl-L-homocysteine (SAH) and the metabolically inert product 1-methylnicotinamide (1-MNA), thus depleting intracellular SAM and reducing methylation potential. (b) Stromal NNMT expression from quantitative proteomic analyses reveals significant upregulation in omental metastases,  $n = 11$  patients. (c) Representative NNMT immunohistochemistry (IHC) confirms elevated expression of NNMT in omental and peritoneal (Periton) metastases and stromal specificity. NNMT is not expressed in the benign omentum. (d) Stromal NNMT expression is elevated in omental (Om) and peritoneal metastases compared to ovarian sites. (e) Morphology of CAFs expressing indicated constructs (green; GFP). (f) Production of 1-MNA as assessed with HPLC-MS is attenuated upon knockdown and enhanced upon overexpression of NNMT,  $n = 3$  biological replicates. (g) Knockdown of NNMT in CAFs attenuates expression of CAF markers (fibronectin, Fn1; smooth muscle actin, SMA), while NNMT overexpression in normal fibroblasts increases expression of the markers. (h) Knockdown of NNMT in CAFs reduces collagen contractility; overexpression in normal fibroblasts (3T3) increases contractility,  $n = 3$  biological replicates. (i) Gene set enrichment analysis reveals that genes regulated by NNMT in CAFs are enriched for components of the stromal signature of metastasis. All error bars are standard error of mean (SEM). \* $p < 0.05$ ; \*\*\* $p < 0.001$ .



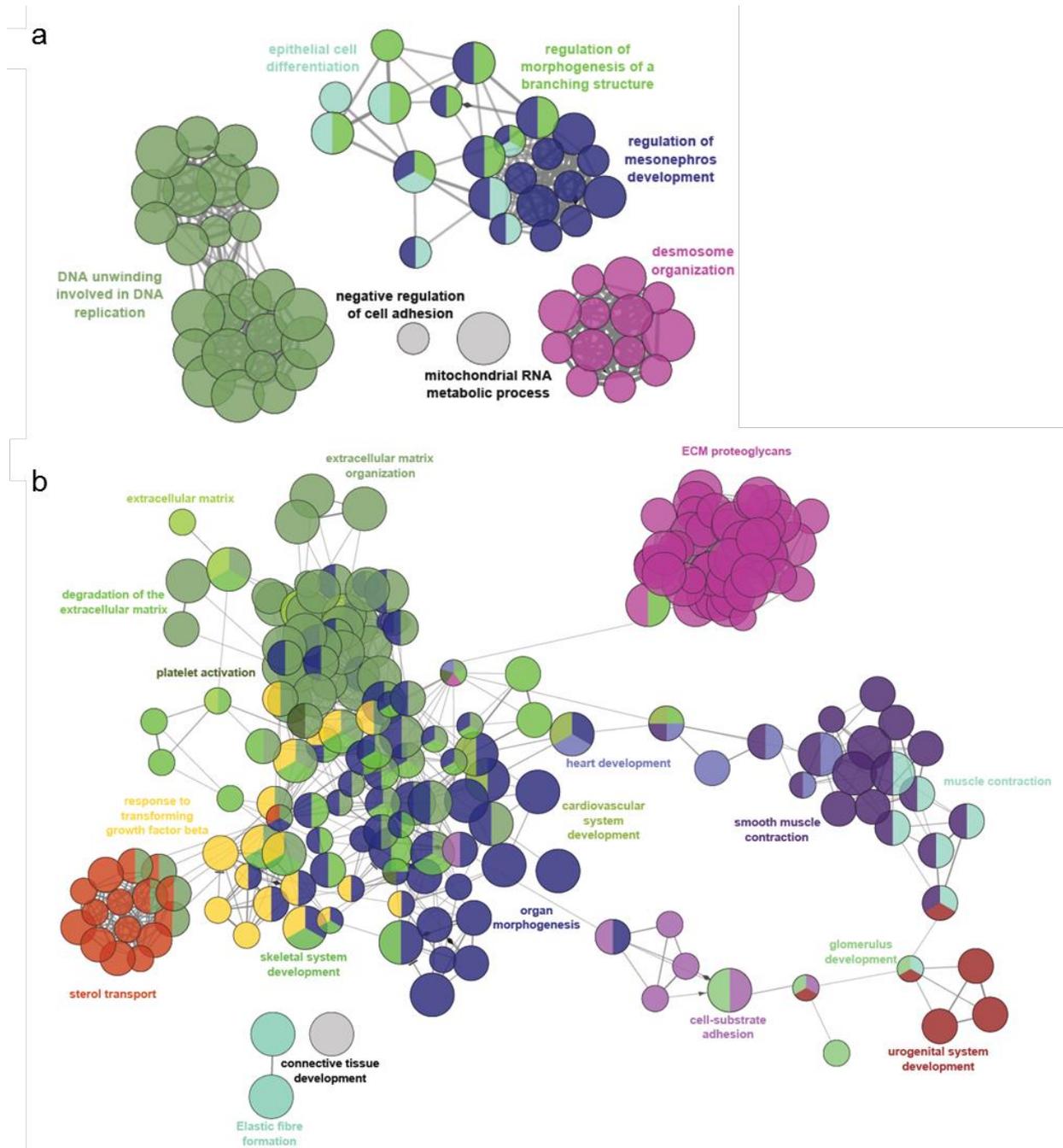
**Fig. 3: NNMT regulates histone methylation to drive the CAF phenotype.** (a) HPLC-MS relative quantification of SAM and SAH levels of CAFs (left) and normal fibroblasts (right; 3T3, 293T, and normal omental fibroblasts) expressing the indicated constructs,  $n = 3$  biological replicates. (b) Quantitative proteomics of histone H3 methylation in CAFs expressing shCtrl or shNNMT constructs following immunoprecipitation. me1=mono-methylation; me2=dimethylation; me3=trimethylation. Significantly altered histone marks are highlighted in red.  $n = 3$  technical replicates. (c) Immunoblotting of H3K27me3 in fibroblasts overexpressing NNMT (top) or CAFs transfected with the indicated siRNAs. (d) Representative immunohistochemistry of COMP in an omental metastasis. T=tumor; S=stroma. (e) qRT-PCR analysis of COMP expression in CAFs expressing the indicated constructs,  $n = 3$  biological replicates. (f) Schematic of H3K27me3 peaks (maroon) and primers targeting COMP promoter region (red) in relation to COMP transcriptional start site. Relative H3K27me3 enrichment as determined with qPCR in CAFs expressing shCtrl or shNNMT constructs,  $n = 2$  biological replicates. (g) Immunoblot and (h) qRT-PCR of CAF markers after treatment with the EZH2 histone methyltransferase inhibitor DZNep,  $n = 3$  biological replicates. (i) Collagen contractility of CAFs expressing shNNMT construct treated with DZNep,  $n = 3$  biological replicates. All error bars are SEM. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .



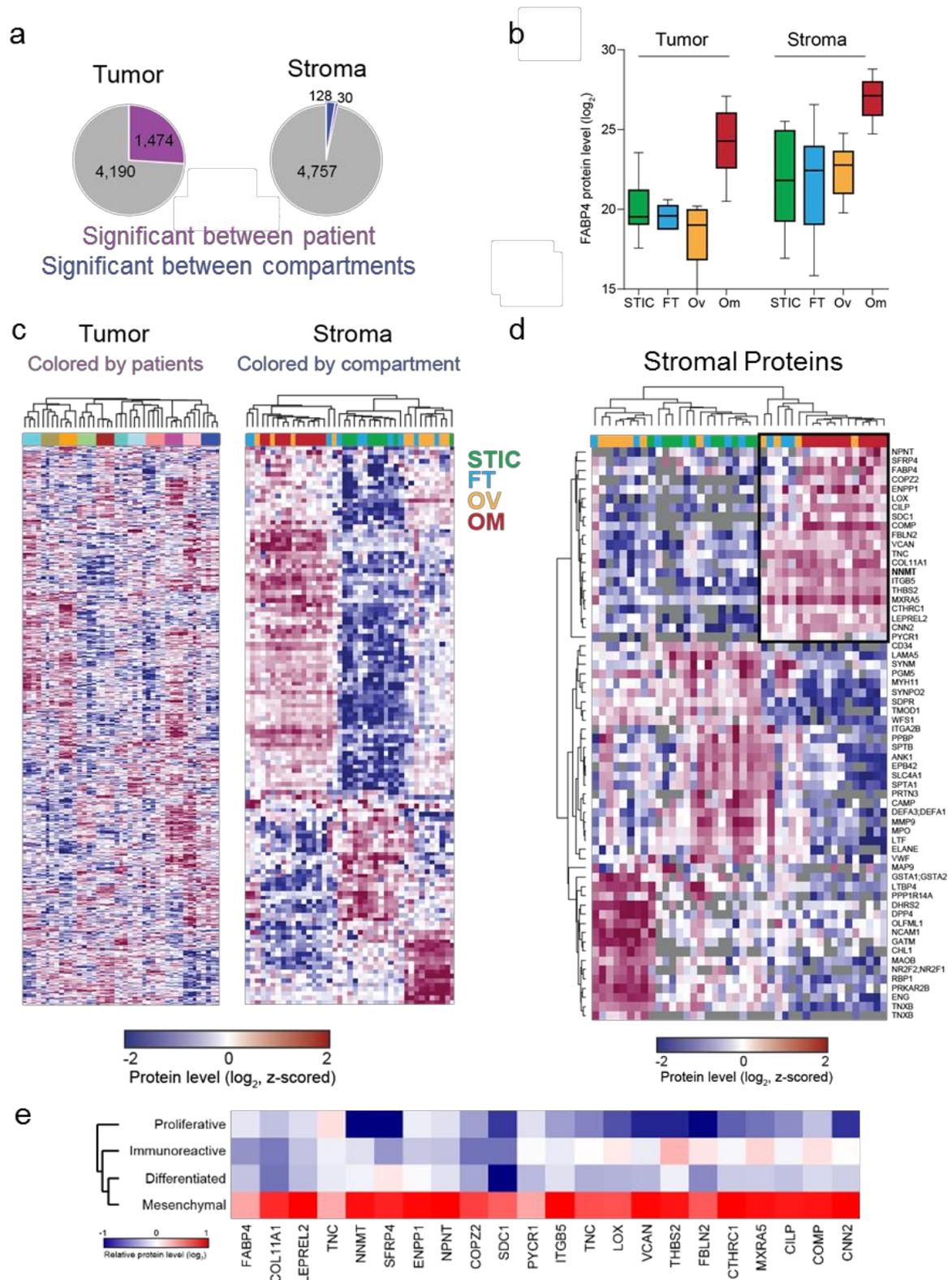
**Fig. 4: Stromal NNMT supports HGSC progression and is associated with a poor prognosis.** (a) Proliferation (doubling time) of HeyA8 and TYKnu OvCa cells following treatment with the indicated conditioned media (CM). Proliferation rate increases (doubling time decreases) with NNMT overexpression and decreases (doubling time increases) upon knockdown,  $n = 3$  biological replicates. (b) Representative images (left) and quantification (right) of chemotaxis of the indicated OvCa cells in response to conditioned media from CAFs expressing the shCtrl or shNNMT constructs,  $n = 3$  biological replicates. (c) *In vivo* proliferation and total tumor burden of luciferase-labeled HeyA8 cells co-injected with CAFs expressing shCtrl or shNNMT constructs. 9 tumors per group. (d) Schematic of experimental design (top). Representative images and quantification of omental adhesion of luciferase/GFP-labeled ID8 mouse OvCa cells treated with conditioned media from fibroblasts expressing the indicated constructs,  $n = 7$  mice per group. (e) A TMA assembled from 92 patients with chemo-naïve HGSC and matched primary and metastatic tumors was stained for NNMT. Representative images of ovarian tumors expressing low (top) or high (bottom) levels of stromal NNMT (left). Kaplan-Meier overall survival curves for patients with low (black) or high (red) stromal expression of NNMT in ovarian sites (right). (f) Stromal NNMT drives ovarian cancer progression by metabolic regulation of histone methylation which causes epigenetic and transcriptional changes in the stroma that promote OvCa proliferation, migration, and metastasis. All error bars are SEM. \* $p < 0.05$ .



**Extended Data Fig. 1: Quantitative proteomics of low-input samples.** (a) MaxLFQ label-free quantitation values and dynamic range are similar across all anatomic sites and in both tumor (left) and stroma (right) samples. (b) Example of experimental replicates of microdissection, protein extraction, and quantitative proteomics with Pearson correlation of  $>0.95$ . (c) Unsupervised hierarchical clustering of all proteomic samples leads to clustering of tumor and stromal samples characterized by proteomic signatures associated with the indicated pathways. (d) One-dimensional principal component analysis of all tumor and stromal samples. Component 1 accounts for 23.5% of the total data variation.

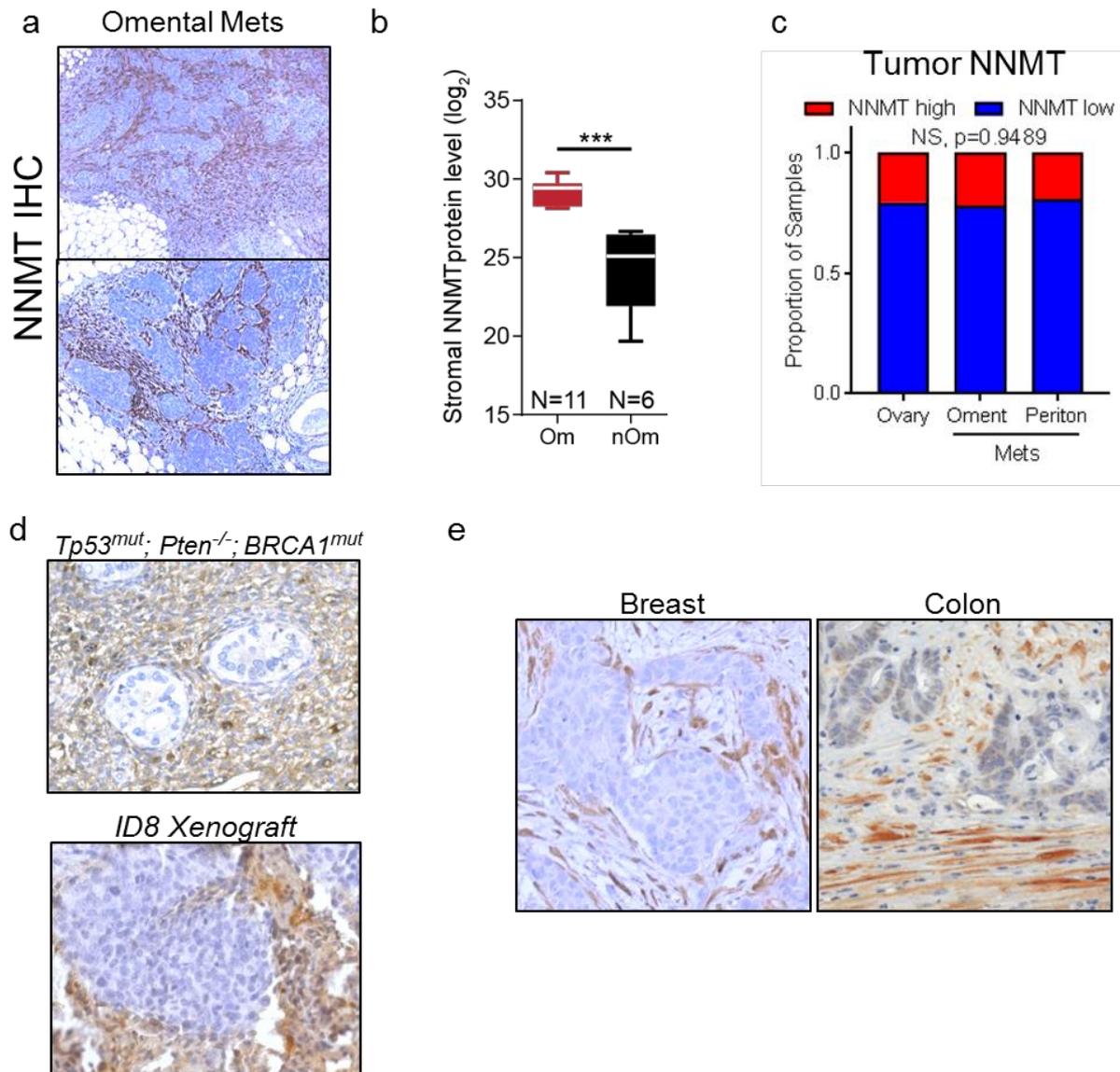


**Extended Data Fig. 2: Gene network analysis of HGSC progression proteomics.** Gene network analysis of (a) all tumor compartments and (b) all stromal compartments in the progression series (Fig. 1a). Pathways in more than one network are filled with more than one color. Size of circles correlates with number of genes in the pathway; pathways will shared genes are connected with lines.

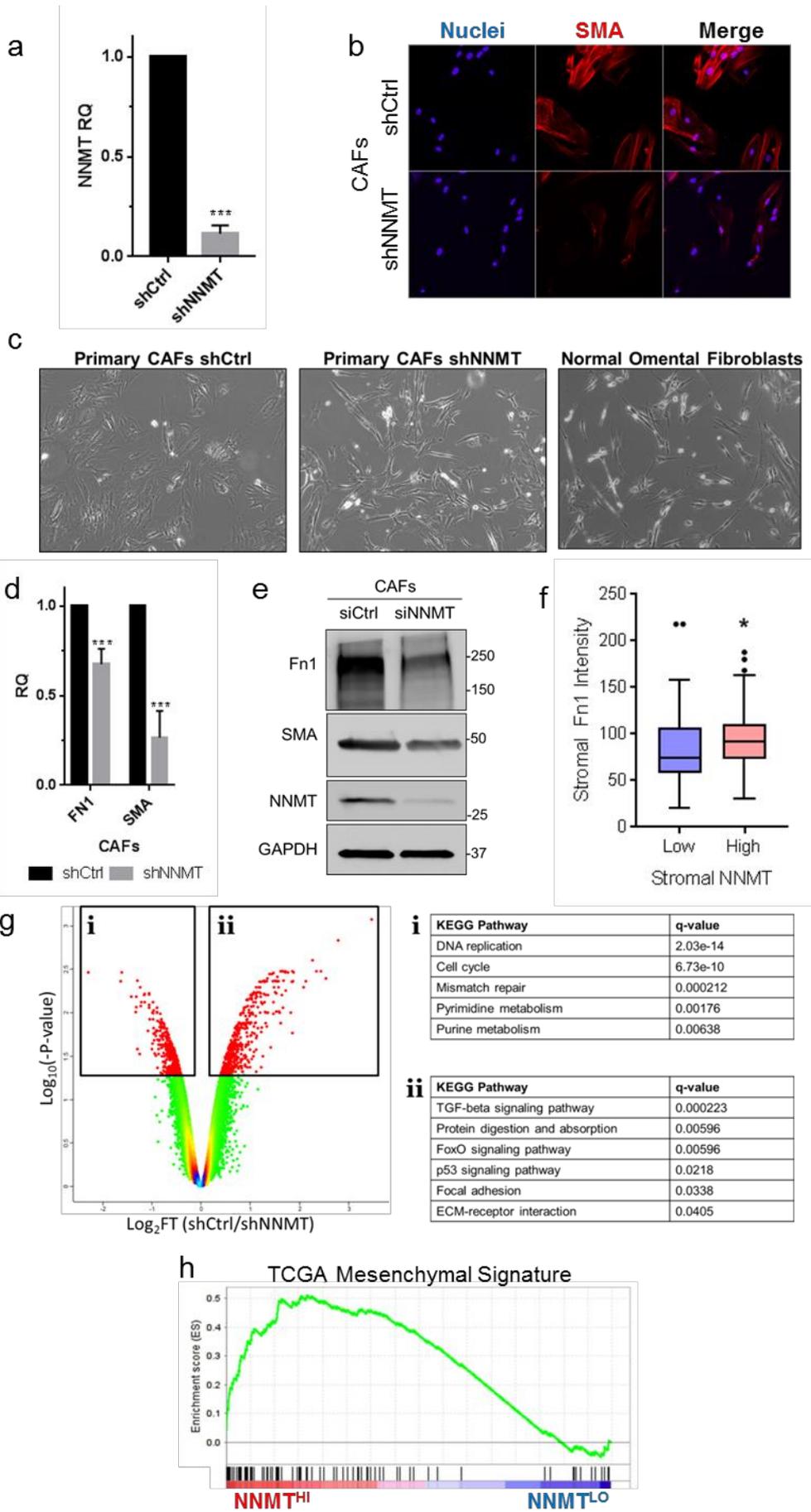


**Extended Data Fig. 3: HGSC progression is characterized by patient-specific signatures in the tumor compartment and site-specific signatures in the stroma.** (a) Left and right panels show proportions of all proteins that are significantly different by patient (purple) or anatomic site (blue). 1,474 proteins are differentially expressed in the tumor compartment between patients while only 30 stromal proteins are significantly different between patients. In respect to the compartment (tumor/stroma, blue) one protein is different in the tumor compartment (FABP4) while 128 proteins are differentially expressed in the stroma. (b) Boxplot of FABP4 expression in tumor (left) and stromal (right) compartments during progression. (c) Unsupervised hierarchical clustering of tumor (left) and

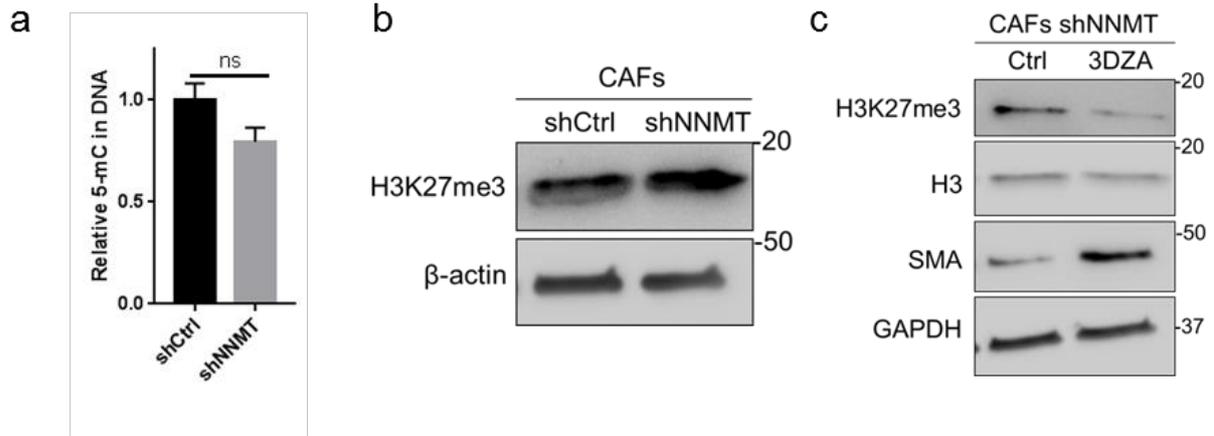
stroma (right) proteins reveals patient-specific clustering in the tumor compartment while the stromal samples cluster by anatomic site. For example, all STIC samples (green) across all patients cluster together. **(d)** Unsupervised hierarchical clustering of differentially expressed stromal proteins (**Fig. 1d**) reveal anatomic site-specific clusters, including a core signature of 21 proteins consistently upregulated in the stroma of omental metastases (**box**). **(e)** Expression of 21 protein signature in the TCGA subtypes reveals enhanced expression in mesenchymal subtype.



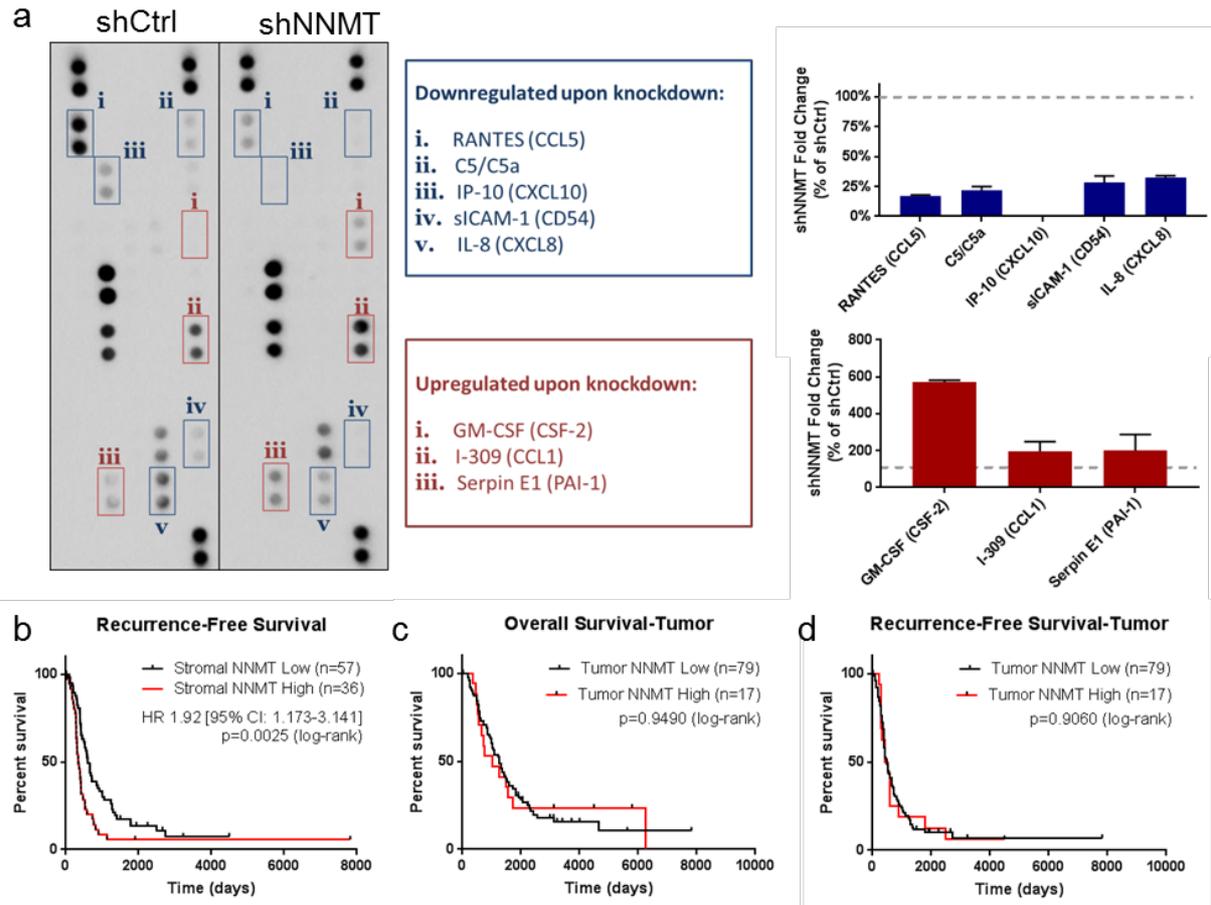
**Extended Data Fig. 4: NNMT is highly expressed in the stroma of ovarian cancers.** (a) Representative IHC of NNMT in omental metastases from individual patients reveals strong expression restricted to the stromal compartment. (b) Quantitative proteomics of the stroma of omental metastases ( $n = 11$  patients) and normal omental tissue (nOm;  $n = 6$  patients) find significantly elevated expression of NNMT in the metastatic stroma. (c) Quantification of tumoral NNMT staining in TMA analysis. (d) Representative IHC of NNMT in the stroma of metastases in an autochthonous model of ovarian cancer (top, PAX8:TP53<sup>mut</sup>;PTEN<sup>-/-</sup>;BRCA1<sup>mut</sup>) and a syngeneic model (bottom, ID8 intraperitoneal xenograft). (e) NNMT is expressed in the stroma of breast and colon cancers. \*\*\* $p < 0.001$ .



**Extended Data Fig. 5: NNMT promotes acquisition and maintenance of the CAF phenotype.** (a) qRT-PCR for NNMT in CAFs expressing the indicated constructs,  $n = 3$  biological replicates. (b) Immunofluorescence analysis of smooth-muscle actin (SMA) reveals attenuation of SMA stress fibers upon knockdown of NNMT in CAFs. (c) Representative brightfield images of normal omental fibroblasts and primary CAFs expressing shCtrl and shNNMT constructs reveal a reversion of CAF morphology to more closely resemble normal omental fibroblasts upon knockdown of NNMT. (d) Relative mRNA ( $n = 3$  biological replicates) and (e) protein expression of fibronectin and SMA in CAFs expressing the indicated constructs. (f) Elevated stromal fibronectin is associated with NNMT expression. (g) Volcano plot of genes significantly regulated by NNMT upon knockdown in CAFs. Significantly differentially expressed genes are highlighted in red. Over-represented KEGG pathways are highlighted. (h) GSEA analysis of genes regulated by knockdown of NNMT in the TCGA mesenchymal signature. Error bars are 95% confidence intervals.



**Extended Data Fig. 6: NNMT does not regulate DNA methylation and a general methyltransferase inhibitor is sufficient to induce expression of CAF markers.** (a) Quantification of 5-methylcytosine (5-mC) in genomic DNA extracted from CAFs expressing the indicated constructs,  $n = 3$  biological replicates. (b) Treatment of CAFs expressing and shNNMT construct with the general methyltransferase inhibitor 3-DZA reduces H3K27 trimethylation and increases SMA expression. Error bars are SEM; ns = not significant.



**Extended Data Fig. 7: NNMT regulates pro-tumorigenic cytokines and is associated with a poor clinical outcome.** (a) Representative images and quantification of a cytokine array of CAFs expressing the indicated constructs. Genes downregulated upon knockdown of NNMT are highlighted in blue, those increased in red. 100% relative to shCtrl is highlighted with dashed grey line ( $n = 2$  technical duplicates). (b) Kaplan-Meier curve of recurrence-free survival of patients with low (black) or high (red) stromal NNMT expression in primary sites. NNMT expression in the tumor compartment at primary sites was not associated with (c) overall or (d) recurrence-free survival. Error bars are SEM.

## **METHODS.**

### **Patient samples.**

All patients included in the study underwent primary debulking surgery at the University of Chicago and were diagnosed with metastatic HGSO (Supplementary Table S1). All human tissue samples were collected with informed consent under University of Chicago Institutional Review Board-approved protocols and in accordance with the Declaration of Helsinki.

### **Laser-capture microdissection.**

Samples were prepared as previously described (248). Formalin-fixed, paraffin-embedded (FFPE) specimens were sectioned with a microtome (10  $\mu\text{m}$  sections) and mounted on Leica PEN-membrane MembraneSlides (2  $\mu\text{m}$ ). Slides were deparaffinized with xylene and rehydrated through graded alcohols and water. Sections were stained with Mayer's hematoxylin (Sigma) and dehydrated through graded alcohols and xylene. Tumor and stromal tissues were dissected with a Leica LMD 6500 laser microdissection system and tumor and stromal samples collected in 0.5 ml tubes. Depending on the FFPE specimen, an area of  $1\text{-}5\times 10^6 \mu\text{m}^2$  was collected (approximately 5-25,000 cells, as derived from dissected area  $\times$  slide thickness / average mammalian cell volume of  $2,000 \mu\text{m}^3$ , BNID 100434).

### **FFPE tissue preparation for MS analysis.**

FFPE tissue lysis was adapted from previously described methods (132). Microdissected tissue was collected by pipetting 50  $\mu\text{l}$  of lysis buffer (50% 2,2,2-trifluoroethanol (TFE) and 5 mM dithiothreitol in 25 mM  $\text{NH}_4\text{HCO}_3$  buffer) and centrifuged for 5 min at maximum speed to collect all remaining tissue. Samples were boiled at  $95^\circ\text{C}$  for 30 min and briefly centrifuged every 5-10 min. Subsequently, samples were sonicated for 10 min (level 5, Bioruptor Plus, Diagenode). After centrifugation for 10 min at maximum speed, supernatants were transferred to new tubes and alkylated with 20 mM iodoacetamide for 30 min in the dark. Samples were vacuum-concentrated for 30 min at  $45^\circ\text{C}$  (until approximately 10  $\mu\text{l}$  buffer remained). 40  $\mu\text{l}$  of fresh digestion buffer (10% TFE and trypsin in 50 mM  $\text{NH}_4\text{HCO}_3$ ) was added in a trypsin (Promega) to total protein ratio of approximately 1:50. Digestion was carried out at  $37^\circ\text{C}$  overnight and digestion stopped by addition of 1% TFA. Samples were vacuum-evaporated to a volume of 5-10  $\mu\text{l}$  and resuspended in 100  $\mu\text{l}$  0.2% TFA and 2% acetonitrile prior to desalting with C18 StageTips. Samples were stored at  $-20^\circ\text{C}$  until MS analysis.

### **Liquid chromatography (LC)-mass spectrometry (MS) analysis.**

Quadrupole Orbitrap mass spectrometers (209, 253) (Q Exactive and Q Exactive HF, Thermo Fisher Scientific) coupled to an EASY-nLC 1000 HPLC system (Thermo Fisher Scientific) via a nano-electrospray source were operated in data dependent mode for LC-MS analysis of peptides. LC columns (75  $\mu\text{m}$  inner diameter, 50 cm length) were packed in-house with C18 particles (1.9  $\mu\text{m}$ , Dr. Maisch GmbH, Germany). Peptides were separated with a 250 min HPLC gradient from 2% to 60% in buffer B (80% acetonitrile, 0.5% formic acid) at a flow-rate of 200 nl/min. We employed a resolution of 70,000 at  $m/z$  200 (60,000 for Q Exactive HF) for survey scans. The scan range was set to 300 to 1,650  $m/z$ . Up to the 3 most abundant MS1 features (charge  $\geq 2$ ) were selected for high-energy collisional dissociation fragmentation at a resolution of 17,500 at  $m/z$  200 (15,000 for Q Exactive HF). Dynamic exclusion of sequenced peptides was set to 45 s. Ion injection times and ion target values were set to 20 ms and  $3 \times 10^6$  for the survey scans and 220 ms and  $1 \times 10^5$  for the MS/MS scans, respectively. Data was acquired using Xcalibur software (Thermo Scientific).

### **Data analysis of proteomic raw files.**

MS raw files were processed with MaxQuant (94) (version 1.5.3.15). The Andromeda search engine(95) of MaxQuant was used for peptide and protein identification. Andromeda is a probability-based search engine that employs a target-decoy approach to identify peptides and proteins at a false-discovery rate (FDR) of less than 1%. The UniProtKB database release (August 2015) was used as the forward database. MaxQuant automatically generated a reverse decoy database based on the provided forward database. Proteins which could not be distinguished based on the identified peptides were grouped into protein groups (94). The MaxLFQ(103) algorithm was used for label-free proteome quantification. MaxLFQ is a peptide intensity-based algorithm that makes use of high-resolution 3D peptide features in mass-to-charge, retention time and intensity space.

For each protein, peptide ratios were calculated in a pairwise manner and combined with protein ratios by calculating the median of all peptide ratios. Only the exact same peptide species was considered for each pairwise calculation. A least-squares analysis was then used to reconstruct the relative protein abundance across samples, which preserves the total summed intensity for a protein over all samples. A minimum of one ratio count for each pairwise comparison was required. The “Match Between Runs” feature of MaxQuant was enabled to match high-resolution MS1 features between runs.

### **Proteomic bioinformatic analyses.**

All statistical and bioinformatics analyses were done using Perseus (218) (MaxQuant environment) or R framework. For pairwise proteomic comparisons, we used a 2-sided t-test statistic including a permutation-based FDR of 1% and an  $s0$  value (219) of 2. Missing values were imputed based on a normal distribution (width = 0.15; downshift = 1.8). For pathway enrichment analyses, we used the Cytoscape plugin ClueGO. Enriched Gene Ontology annotations (GOMF) for the top 100 tumor or stromal proteins were calculated over the entire protein list and corrected by applying a Benjamini-Hochberg FDR of 5%.

### **Tissue immunohistochemistry & immunofluorescence.**

FFPE tissue specimens were cut at 10  $\mu\text{m}$ , deparaffinized in xylene, and rehydrated through graded ethanols. Following heat-mediated, sodium citrate antigen retrieval (10 mM sodium citrate, 0.05% Tween 20, pH 6), slides were stained with anti-NNMT (Santa Cruz G-4; 1:200) or anti-COMP (Thermo Fisher Scientific MA1-20221; 1:100) antibodies and processed with the VECTASTAIN Elite ABC HRP kit and DAB Substrate Kit (Vector Laboratories). Slides were counterstained with hematoxylin and dehydrated through graded alcohols and xylene. For immunofluorescence, slides were stained with fluorescently-labeled secondary antibodies (1:500; Thermo Fisher) and Hoechst 33258 (1:200; Molecular Probes). Confocal microscopy was performed with a Zeiss LSM510 and images processed with Image J 1.50j.

### **Tissue microarray analysis.**

Assembly and construction of the tissue microarray have been previously described(254). NNMT immunohistochemical reactivity was scored without knowledge of clinical outcome by two experienced pathologists (S.M. and R.L.). Each sample was scored based on the percentage of positive cells in each compartment (0, no staining; 1, <30%; 2, 30–50%; 3,  $\geq$ 50%); staining intensity was similar across all samples. Expression was considered “low” if the staining intensity was 0 or 1 and “high” if 2 or 3. Analysis was limited to the ovarian compartment of chemotherapy naïve patients with high grade serous pathology (n=97 patients; **Supplementary Table S5**). Kaplan-Meier survival curves and statistical analyses of overall and progression-free survival were performed with GraphPad Prism 7 using the Mantel-Cox (log-rank) test.

### **Isolation of primary cells.**

CAFs were isolated from transformed omental tissue of patients with HGSOC. Normal omental fibroblasts (NOFs) were isolated from omental tissue from female patients undergoing surgery for benign conditions. CAFs and NOFs were isolated and validated as previously described(255). Tissues were thoroughly rinsed with PBS before mincing and digestion (12-18 hr) with collagenase (3 mg/ml) and hyaluronidase (0.5 mg/ml) in 10% fetal bovine serum (FBS) in DMEM. Primary NOFs or CAFs adhered to tissue culture plastic within 24 hr.

### **Cell lines.**

HeyA8 (Gordon Mills, MD Anderson, Houston, TX), 3T3 (American Type Culture Collection), 293T (Lucy Godley, University of Chicago, Chicago, IL), and primary and immortalized CAF cells were cultured in DMEM supplemented with 10% FBS, MEM vitamins, MEM non-essential amino acids, penicillin, and streptomycin. TYKnu cells (Gottfried Koneczny, University of California, Los Angeles) were cultured in MEM Alpha supplemented with 10% FBS, MEM vitamins, and MEM non-essential amino acids. ID8 cells were grown in DMEM supplemented with 4% FBS, Insulin-Transferrin-Selenium (5mg/ml; Gibco). All cells were cultured at 37 °C in a humidified incubator at 5% CO<sub>2</sub>. For experiments, CAFs and fibroblasts (3T3 and 293T cells) were grown in 10 μM methionine media. For some experiments, cells were treated with 1 μM DZNep or 1 μM 3DZA or vehicle (DMSO) control for 72 hr. All cell lines were *Mycoplasma*-negative and authenticated (IDEXX BioResearch short tandem repeat marker profiling).

### **Plasmids and expression.**

Short hairpin RNA (shRNA) oligonucleotides were designed using the shRNA Designer from BiOSETTIA (using accession number NM\_006169) and cloned into the pLV-hU6-CMV-Green backbone using the manufacturer's protocol (**Supplementary Table S7**). CAFs were infected with non-targeting shCtrl or shNNMT constructs; 3T3 normal fibroblasts were infected with pLenti6 empty vector control or pLX304-NNMT overexpression vector (HsCD00442343; DNAsu.org). CAFs were immortalized with pBABE-neo-hTERT (1774; Addgene). To produce lentivirus, 293T cells were seeded at  $1 \times 10^6$  cells per 6 cm dish in DMEM supplemented with 10% FBS. After 18 hr, cells were transfected with 1 μg expression vector, 0.9 μg pCMV-dR8.2 packaging vector (8455; Addgene), and 0.1 μg pCMV-VSV-G expression vector (8454; Addgene) using Lipofectamine 2000. Viral supernatant was harvested at 48 and 72 hr post-transfection, filtered through a 0.8 μm filter, and added to recipient cells with 4 μg/ml polybrene

for 8-12 hr before selection with puromycin (2 µg/ml) or blasticidin (1 µg/ml). For transient transfections, cells were seeded in a 6-well dish and transfected with 5 nmoles siRNA pools (GE Dharmacon) using Lipofectamine 2000 (Thermo Fisher) in OptiMEM Reduced Serum Media (Thermo Fisher).

### **Targeted LC-MS/MS metabolomics.**

Cells were scraped into ice-cold PBS and isolated by centrifugation at 1,400 rcf at 4°C. Cell pellets were resuspended in 300 µl of an 80:20 mixture of MeOH/H<sub>2</sub>O. Internal deuterated standards, 10 nmol *d*<sub>3</sub>-serine, were added to the extraction solution for sample normalization. The mixture was sonicated (Fisher Scientific FB-505) for 5 s followed by a 10 min centrifugation at 16,000 rcf. The supernatant was collected and dried under N<sub>2</sub> gas. The dried metabolites were resuspended in 30 µl of an 80:20 mixture of MeOH/H<sub>2</sub>O. Resuspended metabolites were separated by hydrophilic interaction chromatography with a Luna-NH<sub>2</sub> column (5 µm, 100 Å, 50 × 4.6 mm, Phenomenex). Mobile phase A was composed of 100% CH<sub>3</sub>CN, and mobile phase B was composed of 95:5 (v/v) H<sub>2</sub>O:CH<sub>3</sub>CN. For positive mode analysis, mobile phase A and B were supplemented with 0.1% formic acid. The gradient started with 0% B for 5 min and increased linearly to 100% B over 15 min with a flow rate of 0.4 ml/min, followed by an isocratic gradient of 100% B for 10 min at 0.4 ml/min. Then, the column was equilibrated with 0% B for 5 min at 0.4 ml/min. Targeted MS/MS analysis was performed on an Agilent triple quadrupole LC-MS/MS instrument (Agilent Technologies 6460 QQQ). The capillary voltage was set to 4.0 kV. The drying gas temperature was 350 °C, the drying gas flow rate was 10 L/min, and the nebulizer pressure was 45 psi. The mass spectrometer was run in MRM mode. The following MS transitions were used to measure the indicated metabolites: NA (*m/z* 123.05 → 80.05), 1-MNA (*m/z* 137 → 94), SAM (*m/z* 399 → 250), SAH (*m/z* 385 → 136), and *d*<sub>3</sub>-Serine (*m/z* 109.07 → 63.1). Relative metabolite abundance was quantified by integrated peak area for the given MRM-transition. Data presented are representative of three independent biological experiments each containing three technical replicates for a given condition.

### **Immunoblots and cytokine array.**

Cells were lysed in SDS lysis buffer containing 4% SDS and 10 mM HEPES, pH 8.5. Proteins were separated by SDS-PAGE on a 4-20% gel and transferred to a nitrocellulose membrane. The membrane was blocked with 5% non-fat dry milk (NFDM) in Tris-buffered saline with Tween 20 (TBST) for 30 min at RT and probed overnight with primary antibodies in 2% bovine

serum albumin in TBST at 4°C (**Supplementary Table S6**). After washing with TBST, the membrane was incubated with secondary antibodies (Thermo Fisher Scientific) conjugated to horseradish peroxidase at 1:5,000 dilution in 5% NFDN/TBST for 1 hr at RT. Proteins were visualized using Clarity Western ECL Substrate (Bio-Rad) or SuperSignal West Femto Substrate (Fisher Scientific). For the cytokine array, cells were seeded at  $3 \times 10^6$  cells per 15 cm plate and grown for 72 hr in growth media. Conditioned media was collected, filtered through a 0.22  $\mu\text{m}$  filter, and processed with the Proteome Profiler Human Cytokine Array Kit, Panel A (R&D Systems) following the manufacturer's recommendations. Images were quantified using ImageJ. Full-length immunoblots are available in Supplementary Data.

### **RT-qPCR.**

Total RNA was isolated and treated with DNase using RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. Reverse transcription of 2  $\mu\text{g}$  total RNA was carried out using the High Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific). qPCR was performed with TaqMan probes and TaqMan Fast Advanced Master Mix or custom primers (IDT) and Fast SYBR Green Master Mix on an Applied Biosystems StepOnePlus Real-Time PCR System and analyzed using the  $2^{-\Delta\Delta\text{Ct}}$  method (**Supplementary Table S7**).

### **Collagen contractility assay.**

Indicated cells were trypsinized and diluted to  $2 \times 10^5$  cells/ml in growth media. 400  $\mu\text{l}$  of diluted cells were mixed with 200  $\mu\text{l}$  of rat tail collagen (5 mg/ml; BD Biosciences) and neutralized with 5  $\mu\text{l}$  1 N NaOH and transferred to a 24-well plate. After 20 min of incubation, collagen gel was freed from the edges with a pipette tip and 600  $\mu\text{l}$  of media added to the well. Collagen gels were incubated at 37 °C for 24 hr before fixation (4% paraformaldehyde) and staining with Eosin-Y (Sigma) before imaging to quantify collagen gel area with ImageJ 1.50j.

### **Proliferation assays.**

Ovarian cancer cells were seeded at 1,000 cells per well in a 96-well plate and allowed to adhere for 24 hr before addition of conditioned media (72 hr) from the indicated cells. After 24 hr and 48 hr, nuclei were visualized by addition of Hoechst 33258 (1:5000) and wells fluorescently imaged with a Zeiss Axiovert Observer.A1 to extract cell number and doubling time. Images were analyzed with ImageJ 1.50j.

### **Migration assays.**

40,000 cells in serum-free media were added to the top chamber of an 8.0  $\mu\text{m}$  PET cell culture insert (Falcon) with conditioned media (72 hr) from indicated cells in the bottom reservoir. After 12 (HeyA8) or 18 (TYK-nu) hr cells were fixed with 4% PFA in PBS, representative fluorescent images collected with a Zeiss Axiovert Observer.A1, stained with 0.1% crystal violet, washed extensively with PBS, and dried. Crystal violet was released with 10% acetic acid and the absorbency measured at 520 nm to quantify relative migration.

### **Gene expression analysis.**

RNA was collected from CAFs expressing shCtrl or shNNMT constructs (RNeasy Mini Kit, Qiagen) and 100 ng of total RNA analyzed on Illumina HumanHT-12 v4 Expression BeadChips (two biological replicates per group). cDNA labeling and hybridization were performed at the University of Chicago Functional Genomics core facility using protocols as suggested by the manufacturer. Data was processed with Illumina GenomeStudio GSGX 1.9.0 and in R (<http://www.r-project.org>) with the *limma* Bioconductor package and *p* values adjusted using Benjamini-Hochberg correction for multiple testing. All analyses were performed at the probe level. Annotation of significant (corrected  $p > 0.05$ ) genes with KEGG pathways was performed with ConsensusPathDB-human (Release 32). Gene set enrichment analysis was performed with GSEA v2.2.0.

### **DNA methylation quantification.**

Genomic DNA was extracted from indicated cells using the Wizard Genomic DNA Purification Kit (Promega). DNA methylation (5-methylcytosine) was quantified with the MethylFlash Global DNA Methylation (5-mC) ELISA colorimetric kit (Epigentek) and read with a Molecular Devices SpectraMax i3 plate reader per manufacturer's protocol.

### **Histone modification proteomics.**

Nuclei were isolated using gentle detergent treatment (0.3% NP-40 in NIB-250 buffer) of cells and centrifugation at 0.6 *rcf* and washed with NIB-250 buffer. Histones were acid-extracted and derivatized with propionic anhydride both prior to and following trypsin. Propionylated histone peptides were resuspended in 50  $\mu\text{L}$  water with 1% TFA and 3  $\mu\text{L}$  were injected in 3 technical replicates on nanoLC/triple quadrupole MS which consisted of a Dionex UltiMate 3000 coupled to a ThermoFisher Scientific TSQ Quantum triple quadrupole mass spectrometer. Buffer A was 100% LC-MS grade water with 0.1% formic acid and buffer B was 100% ACN.

The propionylated peptides were loaded onto an in-house packed C18 trapping column (4 cm × 150 μm; Magic AQ C18, 3 μm, 200 Å -Michrom) for 10 min at a flow-rate of 2.5 μL/min in 0.1% TFA loading buffer. The peptides were separated by a gradient from 1 to 35% buffer B from 5 to 45 min. The analytical column was a 10 cm × 75 μm PicoChip (1PCH7515-105H253-NV New Objective) consisting of the same C18 material as the trapping column. The triple quadrupole settings were as follows: collision gas pressure of 1.5 mTorr; Q1 peak width of 0.7 (FWHM); cycle time of 3 s; skimmer offset of 10 V; electrospray voltage of 2.5kV. SRM mass spectrometer transitions were developed as described previously (256). Data were analyzed using Skyline software (v3.5; MacCoss Lab, University of Washington) with Savitzky–Golay smoothing of peaks. Automatic peak assignment and retention times were checked manually.

### **Chromatin immunoprecipitation.**

For chromatin immunoprecipitation (ChIP), freshly harvested cells were crosslinked with 1% formaldehyde for 10 min in PBS. Cells were lysed in IP Buffer (50 mM Tris-HCl (pH 8), 100 mM NaCl, 5mM EDTA (pH 8), 0.3% SDS, 1.7% Triton-X-100, supplemented with EDTA-free protease inhibitor cocktail (Roche) and phosphatase inhibitor cocktail (Roche)) and chromatin sonicated (Biorupter Plus, Diagenode) to an average size of 200-400 bp. 1 mg of total cell lysate was incubated with 3 μg of respective antibody overnight at 4°C under constant rotation. After 12 hr, 30 μl of protein G-coupled agarose beads (Cell Signaling Technology) were added and incubated for 3 hr at 4°C under constant rotation. Antibody-bait complexes were then washed and the bound chromatin eluted and reverse-crosslinked overnight. Samples were digested with RNase A for 2 hr at 37°C and DNA purified.

### **Xenograft and syngeneic models.**

All animal experiments were approved by the University of Chicago Institutional Animal Care and Use Committee. Animals were not randomized and were not excluded. Sample size was based on pilot experiments and previous experience with the models. For the xenograft model, female nude mice (6-week old; Harlan) were subcutaneously injected in the left and right flanks with 100,000 HeyA8-Luciferase cells mixed with 200,000 CAFs expressing shCtrl or shNNMT constructs in 50% Matrigel (BD Biosciences) diluted with serum-free DMEM. Tumor burden was assessed 7 and 14 days after injection using a Xenogen IVIS 200. Luciferase imaging was performed via intraperitoneal injection of 100 μl D-luciferin (30 mg/ml in PBS; 0.22 μm sterile filtered) 10 min prior to imaging. Images were analyzed with Living Image 4.4 software. Mice

were sacrificed with isoflurane and tumors dissected to assess total tumor burden under blinded conditions.

For the syngeneic model, ID8-luciferase/GFP cells were pre-treated with conditioned media (72 hr) from 3T3 cells overexpressing NNMT or a control construct for 48 hr. Animals were not randomized and were not excluded. 4 million pre-treated ID8 cells were injected intraperitoneally in 7 female C57BL/6 mice (6 weeks old; Harlan) and allowed to implant for 14 hr before mice were sacrificed and the omentums harvested. Following imaging of the omentums (GFP), tissues were lysed with luciferase assay lysis buffer (Promega) and luciferase signal detected using a Lumat LB 9507 luminometer (Berthold Technologies). Analysis of tumor burden was conducted under blinded conditions.

### **Statistical analyses.**

Statistical analyses were carried out using Perseus, Excel, R, and GraphPad Prism 7.01. Data are reported as mean  $\pm$  standard error of mean (SEM), unless otherwise noted in figure legends. Number and type of replicates are indicated in the legends of the corresponding figures. Sample size was based on pilot experiments or previous experience with the models. *P* values are reported in the figure legends and were calculated with unpaired Student's *t*-test for two groups or analysis of variation (ANOVA) with Tukey correction for experiments with more than two groups. For survival analyses, significance was assessed with log-rank (Mantel-Cox) test and hazard ratios calculated with log-rank test. *P* values less than 0.05 were considered significant (95% confidence interval).

### **Data availability.**

Proteomic data is available online at

<http://maxqb.biochem.mpg.de/mxdb/project/show/9373012627500> (username: review2; password: 5kcGES) and have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD006396>) via the PRIDE partner repository with the dataset identifier PXD006396 (username: [reviewer40116@ebi.ac.uk](mailto:reviewer40116@ebi.ac.uk); password: XsCmGwHJ). Microarray data is available in the Gene Expression Omnibus (GEO) database (series GSE98267) and can be accessed at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=wfvvgwkejtqdjol&acc=GSE98267>.

The MaxQuant quantitative proteomics software package and Perseus data analysis software

environment are freely available: <http://www.biochem.mpg.de/5111733/software>.  
Supplementary data is available online.

### **ACKNOWLEDGEMENTS.**

We thank the Harris Family Foundation for their support. We thank Drs. H.A. Kenny, K. Watters, and A. Mukherjee from the University of Chicago ovarian cancer laboratory for discussions. We are thankful to Gail Isenberg, University of Chicago, for editing the manuscript. This work was supported by a Marsha Rivkin Foundation award (M.A.E.), National Cancer Institute (NCI) grant CA111882 (E.L.), the Harris Family Foundation (S.D.Y.), NCI grant CA175399 (R.E.M.), V Foundation for Cancer Research (R.E.M.), and University of Chicago Cancer Center Support Grant P30CA014599, the Max-Planck Society for the Advancement of Science, the Körber Foundation (Körber European Science Prize).

### **AUTHOR CONTRIBUTIONS.**

The study was conceived by M.A.E. and E.L. Proteomic studies were performed by F.C. under supervision of M.M. Experiments were designed by M.A.E., F.C., and R.E.M. Experiments were performed by M.A.E., F.C., A.C., S.P., S.M.T., J.W.C., and M.C. Data was analyzed by M.A.E., F.C., and J.W.C. Human tissues were collected by S.D.Y. and E.L. Mouse tissues were provided by R.P. Tissue microarrays were interpreted and scored by S.S.M. and R.R.L. Paper was written by M.A.E. and E.L. Paper was edited by M.A.E., F.C., M.M., and E.L. Figures were prepared by M.A.E. and F.C. All authors reviewed and provided feedback on the manuscript.

## 3 Discussion and Outlook

### 3.1 The modern proteomic toolbox for the study of ovarian cancer

The poor clinical outcome of HGSOC is linked to a variety of factors such as a scarce understanding of HGSOC origin and pathogenesis, late diagnosis at an advanced metastatic stage and frequently acquired chemotherapy resistance. Furthermore, a substantial amount of pre-clinical research has been performed on inadequate cellular models.

Modern expression proteomics is powerful, comprehensive and comprised of a multitude of techniques to study protein expression, function, regulation, spatio-temporal dynamics or interactions. Remarkably, each of these can be achieved today in single shotgun proteomics runs. Even global proteome studies that aim at quantifying ideally the entirety of all expressed proteins in a single cellular system have reached a nearly complete detection of all expressed proteins (92, 112). This greatly impacts on systems biology-based projects that aim at obtaining a global proteomic portrait of the sample of interest.

In *manuscript 1* of this thesis, I addressed proteomic similarities and differences across 30 frequently used ovarian cancer cell lines. I showed that a single-run (4h) workflow can be used to detect and quantify roughly 8,000 proteins across the measured cell lines. These numbers could only be achieved previously by extensive protein or peptide pre-fractionation, which inevitably increases the analysis time and decreases the throughput. This was the first study to apply the single-run workflow to a large number of cell lines related to a single disease context. Our data revealed that state-of-the-art proteomics can be used to identify distinct cell line groups resembling HGSOC tissue at the proteomic level. This is of great relevance for the OvCa community as it allows selecting the most appropriate cell lines for any pre-clinical investigation. Especially in conjunction with the previous genomic characterization of ovarian cancer cell lines (12), the data will be useful for finding the most accurate cellular models for a variety of research questions. As a proof of concept, we used the proteomic resource to predict the treatment response to all-trans-retinoic acid (ATRA). Two receptors have been linked to differential cellular ATRA responses, CRABP2, which mediates the anti-oncogenic signaling of ATRA, and FABP5, which induces pro-

oncogenic signaling (169). The presence of a high CRABP2/FABP5 ratio was previously shown to promote apoptosis in cancer cells following treatment with ATRA, whereas higher FABP5 levels caused an opposing effect by mediating cell survival (169). Our data indeed revealed that the group of cell lines with high expression levels of CRABP2, compared to FABP5, decreased proliferation or changed morphology to a more differentiated phenotype. This clearly indicated the beneficial, anti-oncogenic ATRA response. Conversely, we observed ATRA-induced proliferation in cell lines that showed low CRABP2 expression, clearly reflecting the reported FABP5-mediated pro-oncogenic signaling pathway. With respect to future clinical trials, these observations may help to select patients potentially benefiting from ATRA-based therapy.

The identified cell line groups further revealed a potential cell-of-origin based classification into OSE-derived and FTEC-derived. This is of particular importance for HGSOC as currently there is no clear consensus on the cellular origin of HGSOC. This is largely due to the fact that both hypotheses, the OSE origin and FTEC origin, have been supported by animal models or clinico-pathological observations (23, 27, 28, 38, 40, 257, 258). Consequently, both hypotheses may be correct thus leading to a proposed third dualistic precursor model (13). Of note, the gene expression profile of normal ovarian tissue was found to be similar to the mesenchymal HGSOC subtype (259). This supports our new proteomics-based classification, into predominantly mesenchymal (potentially OSE-derived) and epithelial (potentially FTEC-derived). Interestingly, the differential expression of EMT markers in the 'good' HGSOC cell lines has already been reported recently (81), however, it remained unclear what these observations actually reflected. Based on our findings, the expression differences of the EMT markers may potentially represent two different cellular origins of serous OvCa (OSE and FTEC). Consistent with my results, a recent miRNA study likewise identified two distinct HGSOC subtypes, an integrated epithelial subtype and an integrated mesenchymal subtype (220). The study further identified that a specific miRNA, miR-506, promoted CDH1 (E-cadherin) expression, prevented TGF $\beta$ -induced epithelial-mesenchymal transition and correlated with good clinical prognosis. Indeed, our data clearly showed higher expression of CDH1 in the likely FTEC-derived cell lines, suggesting that, at least to some extent, miRNA regulatory networks led to the differential protein expression in the two identified HGSOC subtypes. Moreover, we likewise observed that the epithelial HGSOC subtype correlated with better prognosis

as assessed by the comparison of our results to two additional proteomic and genomic datasets.

In summary, *manuscript 1* sheds new light onto HGSOC origin and provides a useful database for the selection of bona-fide HGSOC cell lines, as well as candidate marker proteins to stratify patients into a mesenchymal or epithelial subtype. This might be of prognostic relevance as the mesenchymal subtype is linked to unfavorable clinical outcome (16, 221). The manuscript was published in the August issue of Nature Communications with me as first author.

**Manuscript 2** illustrates the power of modern discovery-based proteomics to investigate proteins under various biological aspects. We initially employed a single-run shotgun proteomics workflow to discover proteins associated with chemotherapy efficacy in late-stage HGSOC. In this retrospective study, our near-comprehensive proteomic dataset (~ 9,000 quantified proteins) identified the cancer-testis antigen 45 (CT45) as significantly up-regulated in the group of patients with favorable response to carboplatin/paclitaxel chemotherapy. Due to the strong correlation between CT45 expression and disease-free survival (Pearson  $r = 0.75$ ), this suggested an active role of CT45 in mediating chemotherapy sensitivity.

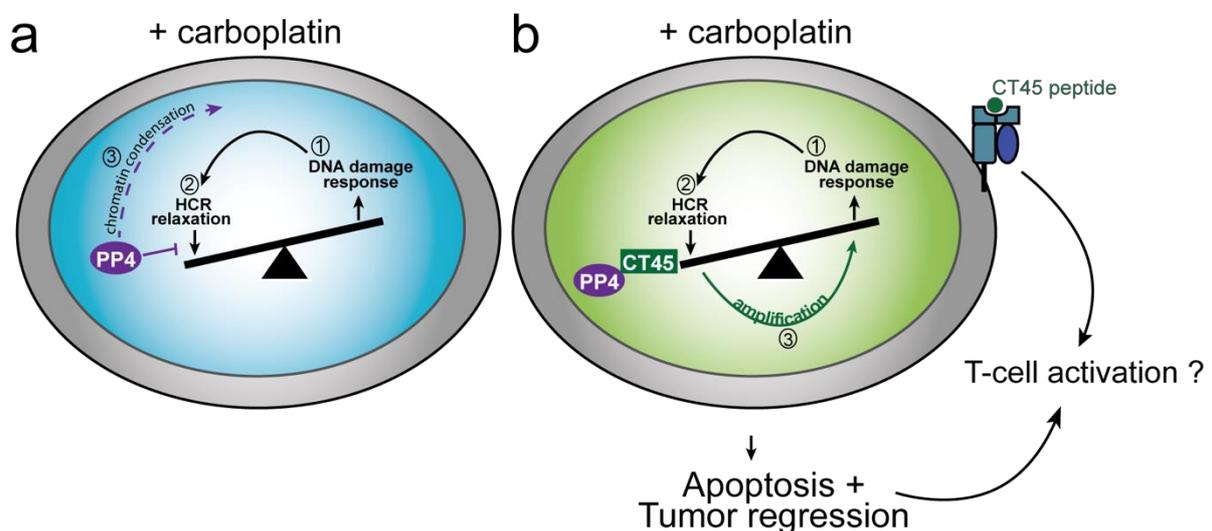
CT45 belongs to the group of cancer-testis antigens that are considered as attractive cancer immunotherapy targets due to their tumor specificity and broad applicability (see section 1.4.4). We used an immunopeptidomics approach to identify CT45 derived HLA class I specific peptides presented on the surface of CT45<sup>+</sup> ovarian cancer cells. Our follow-up work showed that these antigens were indeed *in vivo* T-cell targets that promoted proliferation and activation of patient-derived CD8<sup>+</sup> T-cells, and ultimately mediated cancer cell killing. However, we did not observe an overall correlation between the number of CD8<sup>+</sup> T-cells and CT45 expression in our discovery cohort of 25 patients. Assuming an effective T-cell response against CT45<sup>+</sup> tumors *in vivo*, this stands in contrast to the observed correlation between intratumoral T-cells and clinical outcome of advanced ovarian carcinoma after surgical debulking and chemotherapy (56). However, there are several possible explanations for this finding. Firstly, we did not identify CT45-derived HLA-I peptides of HLA alleles other than A\*03:01, A\*11:01 or A\*68:01. For example, two HLA-A\*02:01 CT45<sup>+</sup> cell lines were also screened, which did not result in an identification of additional CT45-derived HLA-

I peptides. The best *in silico* predicted A\*02:01 specific peptides did not result in T-cell activation. Therefore, CT45<sup>+</sup> tumors without the HLA-A\*03:01, A\*11:01 or A\*68:01 alleles may not induce a T-cell response against CT45, potentially due to the lack of presented peptides. One of the patients with the best clinical outcome in our cohort (DFS > 7 years, patient is alive) was positive for allele A\*11:01, which was predicted to have the highest affinity to our identified CT45 peptides. Additionally, this patient had high CT45 protein expression in the tumor. It is tempting to speculate that the observed long-term protection from tumor recurrence was facilitated by a strong T-cell memory response in addition to the observed chemotherapy sensitivity, and that this is mediated by CT45.

Secondly, in contrast to previous studies, we analyzed metastatic omental tumor tissues, which may be associated with a strong immunosuppressive environment inhibiting T-cell activation and expansion. In support of this, the number of immunosuppressive FOXP3<sup>+</sup> intratumoral T-cells was recently found to be predictive of good neoadjuvant chemotherapy (NACT) outcome, whereas the number of CD8<sup>+</sup> T-cells remained unchanged (59). This finding is of relevance for future studies addressing the link between CT45 expression, chemotherapy efficacy and the role of the immune system.

An interaction proteomics approach helped to discover the function of CT45 by identifying the protein phosphatase 4 (PP4) complex as strongly interacting with CT45. PP4 is critical for the repair of DNA lesions after exogenous DNA damage (183, 185, 187, 189, 260). Our subsequent work consequently focused on the role of PP4 and its related targets such as KAP1/TRIM28, in the context of DNA damage signaling. Briefly, we found that, similar to the PP4 members, CT45 was chromatin associated and perturbed PP4 mediated DNA signaling during heterochromatic DNA damage repair. CT45 expression inhibited PP4 mediated chromatin re-condensation via KAP1-S824 de-phosphorylation after DNA damage repair. This may prolong the DNA damage signal and ultimately lead to apoptosis (Fig. 13). Interestingly, knockdown of heterochromatic proteins, for example KAP1, or induced chromatin de-condensation have been linked to DNA damage sensitivity (192, 261). Our model now suggests that the selective inhibition of KAP1 de-phosphorylation after DNA damage might be a promising therapeutic strategy to pursue in combination with conventional chemotherapy. In such a scenario, ATM induced global chromatin relaxation, which is

required for the repair of heterochromatic DNA damage foci (202), would be prolonged. This may further sensitize cancer cells to more DNA damage due to the vulnerability of open chromatin to form carboplatin-DNA crosslinks. Furthermore, CT45-derived peptides binding to and perturbing PP4/KAP1 signaling during the DNA damage response could be designed. An interaction screen could be used to map the exact PP4 interaction site of CT45. These peptides could be further selected for their ability to mediate chemotherapy sensitivity. This might be especially advantageous in the context of a CT45 negative tumor, ideally in combination with neoadjuvant chemotherapy.

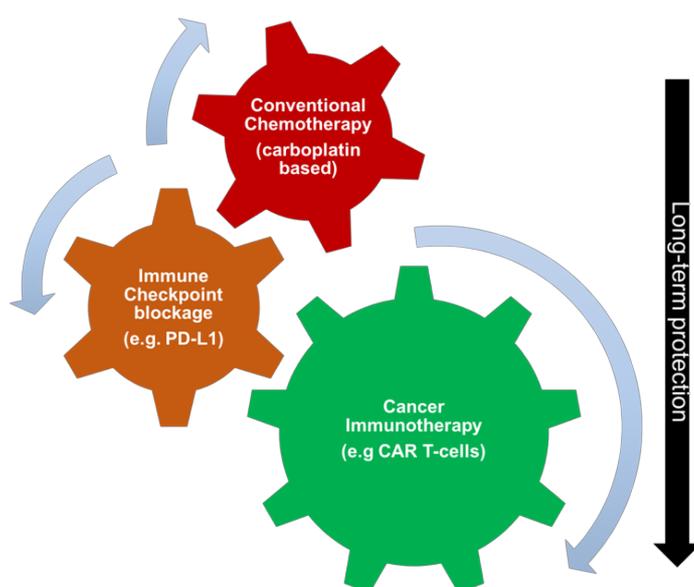


**Figure 13: Model of CT45 mediated chemotherapy sensitivity.**

**a.** Following DNA damage in heterochromatin (HCR), ATM induced chromatin relaxation occurs to facilitate DNA damage repair. PP4 restores heterochromatin to end the DDR. **b.** CT45 perturbs PP4 from restoring heterochromatin, leading to a maintained DNA damage signal and increased signal amplification in case DNA damaging drugs are still present. This may ultimately result in apoptosis and tumor shrinkage.

Based on our results and those of other studies, an attractive treatment strategy for HGSOc may be the combination of conventional chemotherapy, immune checkpoint blockage and personalized cancer immunotherapy (Fig. 14). This concept can be illustrated based on the results of *manuscript 2*.

Before first-line chemotherapy, tumor tissue taken from the initial debulking surgery could be used to evaluate CT45 expression, expression of immune checkpoint molecules such as PD-L1 or CTLA-4, and the presence of intratumoral T-cells. In parallel, blood and ascitic fluid would be collected to isolate PBMCs and CD8<sup>+</sup> T-cells, respectively. Following conventional chemotherapy, checkpoint immunotherapy may be used to eliminate the remaining tumor cells via T-cell mediated killing. The reported up-regulation of PD-L1 on T-cells through chemotherapy may provide a rational to start checkpoint blockage directly after chemotherapy (59). Chemotherapy induced apoptosis may further lead to T-cell activation through the presentation of tumor derived antigens on HLA-I complexes of antigen presenting cells (APCs). To account for immune cell depletion through chemotherapy, T-cells isolated from ascites may be expanded *in vitro* and infused back to the patient. Finally, CD8<sup>+</sup> T-cells derived from PBMC differentiation could also be used to generate CAR T-cells (262) for adoptive T-cell transfer targeting CT45-derived HLA-I peptides. The continuous administration and monitoring of CAR T-cells may promote long-term protection against any tumor recurrence. With respect to CT45 negative tumors, drugs like decitabine (5-aza-2'-deoxycytidine) could be evaluated in combination with standard chemotherapy to activate CT45 antigen expression and presentation (152, 179). Interestingly, decitabine treatment was shown to result in up-regulation PD-L1 expression in cancer cells further indicating its combination with checkpoint blockage therapies (263).



**Figure 14: Beyond chemotherapy: advanced possibilities for ovarian cancer treatment**

Despite the uncertainties with regards to HGSOC origin, there is little doubt about the fact that advanced-stage HGSOC almost inevitably includes metastasis to the omentum, an adipocyte-rich tissue located in the peritoneal cavity (42). Roughly 75% of HGSOC patients present at an advanced disease stage when the cancer has already spread from the primary fallopian tube or ovarian sites to the omentum. Consequently, research dedicated to understanding HGSOC progression and metastasis at the molecular level is of great clinical importance. The goal of *manuscript 3* was therefore to analyze the complete proteomic progression of HGSOC in a discovery cohort of eleven patients with late-stage disease. I established a novel, sensitive workflow to analyze proteins derived from minute amounts of laser-microdissected clinical biobank material (FFPE). We extended the scope of this study by separately analyzing changes in the tumor and its associated microenvironment. Such a study could not be conducted so far due to great challenges in sample collection, processing and proteomic analysis. Interestingly and surprisingly, we found that the tumor proteomes did not substantially change during progression to metastasis, indicating that the large genomic alterations frequently observed in HGSOC are already present at the pre-neoplastic level of STICs. In contrast to the tumor compartment, our data showed a distinct metastatic protein signature derived from the omental tumor microenvironment. Follow-up work by my collaborators confirmed the proteomic findings and discovered the functional importance of the proteins associated with omental metastasis. Out of these, we identified together Nicotinamide N-Methyltransferase (NNMT) as a major regulator of cancer-associated-fibroblasts by metabolically reprogramming normal fibroblasts. This knowledge might be useful for the design of specific inhibitors against metastatic HGSOC. Furthermore, it sheds additional light on the tumor microenvironment as a promising target for HGSOC treatment.

The proteomic findings described in *manuscript 3* might further have general implications for the current understanding of HGSOC and its related subtypes. Based on whole-tissue gene-expression analysis, HGSOC can be grouped into four subtypes termed differentiated, immunoreactive, proliferative, and mesenchymal (220, 221). However, it is a well-known that these subtypes are primarily a reflection of discrete microenvironmental features present in the analyzed whole-tumor tissues. Due to the absence of spatio-temporally resolution of previous analyses, proteomic changes

during progression had not be addressed so far. Consequently, existing subtype assignment may be imprecise since it is based on assessment of single samples. Our data revealed a conserved stromal signature of HGSOC in the omentum, however, in addition we found that the signature was not unique to metastasis and but also presented in some HGSOC tissues isolated from the fallopian tube or ovary. A comparison to previous datasets showed that the identified stromal proteins clearly represented markers of the known mesenchymal HGSOC subtype, which is associated with poor survival. This strongly highlights the importance of compartment-resolved protein analysis to obtain a more refined molecular profile. To put this in perspective, had the study be limited to solely omental tumors for subtype definition instead of ovarian tumors, one would have found predominantly one mesenchymal HGSOC subtype. Alternatively, the analysis of tumors from the FT or ovary of the same patient may have resulted in a different subtype assignment. This possibility is supported by the highly dynamic nature of the tumor microenvironment (264), e.g. by the continuously changing population of different cell types and restructuring of the extracellular matrix. In contrast to the stroma, our analysis clearly showed a remarkably stable tumor proteome throughout HGSOC progression. Interestingly, similar results were obtained for colorectal (124) and luminal breast cancer (125), pointing to a general proteomic feature across cancers. This knowledge may provide a rationale for the search for distinct HGSOC subtypes based on tumor cell derived protein signatures in addition to the known microenvironmental subtypes. Indeed, a recent large-scale characterization of 169 HGSOC cases stratified patients into different survival groups based on proteogenomic features, derived from DNA copy number alterations (5). Due to the known genomic stability of the tumor microenvironment, these features were in all likelihood tumor cell derived. Based on our results, tumor cell derived signatures might be more conserved across tumor sites than stroma derived signatures. This suggests alternative definitions of HGSOC subtypes.

The fact that the identified stromal protein signature, linked to omental metastasis, was conserved across all analyzed patients highlights the unique metastatic niche of the omentum. Mesothelial cells for example, covering omental tissues as a single cell layer, have been implicated in OvCa metastasis because they secrete large amounts of fibronectin, which in turn promotes tumor cell adhesion, invasion and proliferation (49). Indeed, fibronectin (FN1) was among the proteins highly expressed in all stromal

samples of omental metastasis, indicating a strong contribution of mesothelial cells to the protein signature. In regard of the observed stromal signature in invasive tumors of the fallopian-tube and ovary, it is tempting to speculate that these may represent a transformation of the ovarian surface epithelium (OSE) - a mesothelium - which can ultimately give rise to a mesenchymal-type tumor resembling tumors of omental metastasis. Consequently, the absence of the mesenchymal protein signature in FT or ovarian tumors may indicate an FT origin with primary metastasis from the FT to the ovary. This raises the question about the origin of tumors, which show a mesenchymal-type FT tumor but no mesenchymal signature in the ovarian tumor site. Our data revealed that two of eleven patients may belong to this group (Fig. 15). One possibility is that some mesenchymal-like FT cancers are linked to a potential stromal origin as previously described in a mouse model (40). Alternatively, the original FT derived tumor spread to the ovary in these cases, involving an additional epithelial-to-mesenchymal transition exclusively in the original FT tumor, but not in the ovarian site. These site-specific differences may be due to divergent clonal propagation of tumor cells as recently reported for HGSOC (265, 266). Our results in *manuscript 1* indeed showed that a minority of the likely FT-derived cell lines such as KURAMOCHI and COV318 showed mesenchymal characteristics (COL3A1<sup>high</sup>, CDH1<sup>low</sup>), similar to the likely OSE-derived cell lines, despite their clear proteomic assignment to the group of likely FT-derived, PAX8<sup>+</sup> cell lines. Interestingly, both cell lines were originally derived from ascites which may explain their mesenchymal features due to EMT during omental metastasis (46). However, our patient derived proteomic data did not reveal significant changes in the tumor compartment during progression raising the question about whether an EMT program in the tumor cells indeed occurred. We instead found strong pro-metastatic changes in the tumor-associated stroma (e.g. high FN1, TNC, VCAN, NNMT, COL11A1 and THBS2). This is in support by a recent study in colorectal cancer, which linked the expression of a poor-prognosis, mesenchymal gene signature to cancer-associated fibroblasts, as opposed to tumor cells (55). Of note, these observations do not exclude that tumor cells undergo EMT. The fact that we did not find a clear EMT related proteomic signature in the tumor compartment throughout progression simply argues for an apparently small proportion of tumor cells undergoing EMT, for example at the invasion front.

In conclusion, mesenchymal-type of cancers should not be inherently linked to tumor cell features, but instead considered to be strongly influenced by stromal contributions.



### 3.2 The age of large-scale proteomics: implications for biomedicine

The possibility to analyze the entirety of all expressed proteins in a given cellular context, the proteome, has been a source of fascination for a considerable time. Today, 28 years after the development of Matrix-assisted laser desorption / ionization (MALDI) (87, 88) and electrospray ionization (ESI) (89), the two techniques that are the cornerstones of modern MS-based proteomics, the field of proteomics has reached a remarkable level of maturation. It is now possible to routinely decode nearly complete cellular proteomes such as those from yeast (267) or to quantify 10,000 distinct protein phosphorylation events in a single 4h measurement without any upfront protein or peptide separation (104). For more complex proteomes such as those of mammalian cells, 10,000 proteins have already been reported (112, 268). Comparison to transcriptomic data revealed that roughly 10,000 – 12,000 different protein coding loci are likely expressed in an average mammalian cell line, demonstrating that state-of-the-art proteomic technology can detect and quantify nearly complete mammalian proteomes (defined by at least one protein from every genomic locus (92)). From a biomedical point-of-view, this strongly aids our current efforts at dealing with a variety of diseases by the identification of novel disease driver proteins or drug targets. For example, a single-run phospho-proteomics workflow recently allowed the quantification of 9,000 phosphorylation sites per sample in Parkinson's disease cellular models. This led to the discovery of novel bona-fide substrates for LRRK2, a kinase which is frequently mutated in a subset of Parkinson patients. These findings may directly aid the clinical application of specific LRRK2 kinase inhibitors and they already shed new light on the biology of Parkinson's disease. One of the largest challenges in the future will be to take full advantage of the spectacular journey MS-based proteomics has already made, specifically by demonstrating its uniqueness and versatility in important clinical settings. Additional technological advances such as improving mass spectrometric acquisition speed, dynamic range of detection, mass accuracy, sensitivity and mass resolution are already in development. The routine, fast, and accurate analysis of complete proteomes may be realistic in the near future. Specifically ion-mobility-separation (IMS) techniques show great promise due to the additional orthogonal nature of the separation based on the analyte size, shape, and charge (269). Fully integrated into the data dependent acquisition (DDA) workflow of

shotgun proteomics approaches, this combination might be especially advantageous for the rapid analysis of complete proteomes, e.g. in a discovery-based clinical context. It is already certain that MS-based proteomics will make a central contribution to all aspects of translational research and clinical practice.

## Abbreviations

APC:	Antigen presenting cell
ATRA:	All-trans retinoic acid
AUC:	Area under the curve
CAF:	Cancer-associated fibroblast
CCLE:	The Cancer Cell Line Encyclopedia
CID:	Collision-induced dissociation
CPTAC:	Clinical Proteomic Tumor Analysis Consortium
CSF:	Cerebrospinal fluid
CT45:	Cancer-testis antigen 45
CTA:	Cancer-testis antigen
CTA-X:	X chromosome coded class of cancer-testis antigens
DDA:	Data dependent acquisition
ESI:	Electrospray ionization
FDA:	Federal drug administration
FFPE:	Formalin-fixed and paraffin-embedded
FTEC:	Fallopian-tube epithelial cells
HCD:	Higher-energy collisional-dissociation
HGSOC:	High-grade serous ovarian cancer
HHMT:	Helene Harris Memorial Trust
HLA:	Human leukocyte antigen
HLAp:	HLA peptide
IMS:	Ion-mobility separation
IOSE:	Immortalized ovarian surface epithelial cells
ITRAQ:	Isobaric Tag for Relative and Absolute Quantitation
LCM:	Laser-capture microdissection
LC:	Liquid chromatography
LFQ:	Label-free quantification
MALDI:	Matrix-assisted laser desorption / ionization
MRN:	Multiple reaction monitoring
mRNA:	Messenger RNA
MS:	Mass spectrometry
NACT:	Neoadjuvant chemotherapy
NextGenSeq:	Next generation sequencing
NNMT:	Nicotinamide N-Methyltransferase
non-CTA-X:	Non-X chromosome coded class of cancer-testis antigens
OvCa:	Ovarian cancer
PP4:	Protein phosphatase 4
ppb:	Parts-per-billion
PSM:	Peptide spectral match
PTM:	Post-translational modification
SELDI:	Surface-enhanced laser desorption and ionization
SILAC:	Stable isotope labeling with amino acids in cell culture
STIC:	Serous tubal intraepithelial carcinoma
TAA:	Tumor-associated antigen
TCGA:	The Cancer Genome Atlas
TIL:	Tumor-infiltrating lymphocyte
TMA:	Tissue microarray
TMT:	Tandem mass tag
Treg:	Regulatory T-cell
TSA:	Tumor-specific antigen
XIC:	Extracted ion chromatogram (XIC)

## References

1. Bowtell DD, *et al.* (2015) Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat Rev Cancer* 15(11):668-679.
2. Ong SE & Mann M (2005) Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* 1(5):252-262.
3. Bantscheff M & Kuster B (2012) Quantitative mass spectrometry in proteomics. *Anal Bioanal Chem* 404(4):937-938.
4. Bantscheff M, Schirle M, Sweetman G, Rick J, & Kuster B (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* 389(4):1017-1031.
5. Zhang H, *et al.* (2016) Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell*.
6. Frantzi M, Bhat A, & Latosinska A (2014) Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker development. *Clin Transl Med* 3(1):7.
7. Sawyers CL (2008) The cancer biomarker problem. *Nature* 452(7187):548-552.
8. Mann M (2012) Proteomics for biomedicine: a half-completed journey. *EMBO Molecular Medicine* 4(2):75-77.
9. Bassani-Sternberg M & Coukos G (2016) Mass spectrometry-based antigen discovery for cancer immunotherapy. *Curr Opin Immunol* 41:9-17.
10. Megger DA, Bracht T, Meyer HE, & Sitek B (2013) Label-free quantification in clinical proteomics. *Biochim Biophys Acta* 1834(8):1581-1590.
11. Coulie PG, Van den Eynde BJ, van der Bruggen P, & Boon T (2014) Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy. *Nat Rev Cancer* 14(2):135-146.
12. Domcke S, Sinha R, Levine DA, Sander C, & Schultz N (2013) Evaluating cell lines as tumour models by comparison of genomic profiles. *Nature communications* 4:2126.
13. Auersperg N (2013) The origin of ovarian cancers--hypotheses and controversies. *Front Biosci (Schol Ed)* 5:709-719.
14. Bowtell DD, *et al.* (2015) Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat Rev Cancer* 15(11):668-679.
15. Cress RD, Chen YS, Morris CR, Petersen M, & Leiserowitz GS (2015) Characteristics of Long-Term Survivors of Epithelial Ovarian Cancer. *Obstet Gynecol* 126(3):491-497.
16. Cancer Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. *Nature* 474(7353):609-615.
17. Patch A-M, *et al.* (2015) Whole-genome characterization of chemoresistant ovarian cancer. *Nature* 521(7553):489-494.
18. Siegel RL, Miller KD, & Jemal A (2015) Cancer statistics, 2015. *CA Cancer J Clin* 65(1):5-29.
19. Allemani C, *et al.* (2015) Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* 385(9972):977-1010.
20. De Angelis R, *et al.* (2014) Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE--5-a population-based study. *Lancet Oncol* 15(1):23-34.

21. Ferlay J, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. (2014) GLOBOCAN 2012 v1.1, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 *International Agency for Research on Cancer*.
22. Vaughan S, *et al.* (2011) Rethinking ovarian cancer: recommendations for improving outcomes. *Nat Rev Cancer* 11(10):719-725.
23. Bowen NJ, *et al.* (2009) Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells. *BMC Med Genomics* 2:71.
24. Auersperg N, Woo MM, & Gilks CB (2008) The origin of ovarian carcinomas: a developmental view. *Gynecol Oncol* 110(3):452-454.
25. Pothuri B, *et al.* (2010) Genetic analysis of the early natural history of epithelial ovarian carcinoma. *PLoS One* 5(4):e10358.
26. Callahan MJ, *et al.* (2007) Primary fallopian tube malignancies in BRCA-positive women undergoing surgery for ovarian cancer risk reduction. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 25(25):3985-3990.
27. Kindelberger DW, *et al.* (2007) Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: evidence for a causal relationship. *Am J Surg Pathol* 31(2):161-169.
28. Piek JM, *et al.* (2001) Dysplastic changes in prophylactically removed fallopian tubes of women predisposed to developing ovarian cancer. *The Journal of pathology* 195(4):451-456.
29. Hu CY, Taymor ML, & Hertig AT (1950) Primary carcinoma of the fallopian tube. *Am J Obstet Gynecol* 59(1):58-67, illust.
30. Bannatyne P & Russell P (1981) Early adenocarcinoma of the fallopian tubes. A case for multifocal tumorigenesis. *Diagn Gynecol Obstet* 3(1):49-60.
31. Woolas R, *et al.* (1994) What is the true incidence of primary fallopian tube carcinoma? *Int J Gynecol Cancer* 4(6):384-388.
32. Alvarando-Cabrero I, Navani SS, Young RH, & Scully RE (1997) Tumors of the fimbriated end of the fallopian tube: A clinicopathologic analysis of 20 cases, including nine carcinomas. *International Journal of Gynecological Pathology* 16:189-196.
33. Alvarando-Cabrero I, Young RH, Vamvakas EC, & Scully RE (1999) Carcinoma of the fallopian tube: a clinicopathological study of 105 cases with observations on staging and prognostic factors. *Gynecologic Oncology* 72:367-379.
34. Crum CP, Miron A, Ince TA, Muto M, Kindelberger DW (2007) The distal fallopian tube: a new model for pelvic serous carcinogenesis. *Curr Opin Obstet Gynecol* 19(1):3-9.
35. Salvador S, *et al.* (2008) Chromosomal instability in fallopian tube precursor lesions of serous carcinoma and frequent monoclonality of synchronous ovarian and fallopian tube mucosal serous carcinoma. *Gynecol Oncol* 110(3):408-417.
36. Lee Y, *et al.* (2007) A candidate precursor to serous carcinoma that originates in the distal fallopian tube. *Journal of Pathology* 211(1):26-35.
37. Kuhn E, *et al.* (2012) TP53 mutations in serous tubal intraepithelial carcinoma and concurrent pelvic high-grade serous carcinoma-evidence supporting the clonal relationship of the two lesions. *Journal of Pathology* 226:421-426.

38. Flesken-Nikitin A, Choi KC, Eng JP, Shmidt EN, & Nikitin AY (2003) Induction of carcinogenesis by concurrent inactivation of p53 and Rb1 in the mouse ovarian surface epithelium. *Cancer Res* 63(13):3459-3463.
39. Connolly DC, *et al.* (2003) Female mice chimeric for expression of the simian virus 40 TAg under control of the MISIR promoter develop epithelial ovarian cancer. *Cancer Res* 63(6):1389-1397.
40. Kim J, *et al.* (2012) High-grade serous ovarian cancer arises from fallopian tube in a mouse model. *Proc Natl Acad Sci U S A* 109(10):3921-3926.
41. Kim J, Coffey DM, Ma L, & Matzuk MM (2015) The ovary is an alternative site of origin for high-grade serous ovarian cancer in mice. *Endocrinology* 156(6):1975-1981.
42. Nieman KM, *et al.* (2011) Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nat Med* 17(11):1498-1503.
43. Kalluri R & Weinberg RA (2009) The basics of epithelial-mesenchymal transition. *J Clin Invest* 119(6):1420-1428.
44. Veatch AL, Carson LF, & Ramakrishnan S (1994) Differential expression of the cell-cell adhesion molecule E-cadherin in ascites and solid human ovarian tumor cells. *Int J Cancer* 58(3):393-399.
45. Darai E, *et al.* (1997) Expression of cadherins in benign, borderline, and malignant ovarian epithelial tumors: a clinicopathologic study of 60 cases. *Hum Pathol* 28(8):922-928.
46. Lengyel E (2010) Ovarian cancer development and metastasis. *American Journal of Pathology* 177(3):1053-1064.
47. Strobel T & Cannistra SA (1999) Beta1-integrins partly mediate binding of ovarian cancer cells to peritoneal mesothelium in vitro. *Gynecol Oncol* 73(3):362-367.
48. Cannistra SA, *et al.* (1993) Binding of ovarian cancer cells to peritoneal mesothelium in vitro is partly mediated by CD44H. *Cancer Res* 53(16):3830-3838.
49. Kenny HA, *et al.* (2014) Mesothelial cells promote early ovarian cancer metastasis through fibronectin secretion. *J Clin Invest* 124(10):4614-4628.
50. Kenny HA & Lengyel E (2009) MMP-2 functions as an early response protein in ovarian cancer metastasis. *Cell cycle* 8(5):683-688.
51. Heath RM, Jayne DG, O'Leary R, Morrison EE, & Guillou PJ (2004) Tumour-induced apoptosis in human mesothelial cells: a mechanism of peritoneal invasion by Fas Ligand/Fas interaction. *British journal of cancer* 90(7):1437-1442.
52. Schauer IG, Sood AK, Mok S, & Liu J (2011) Cancer-associated fibroblasts and their putative role in potentiating the initiation and development of epithelial ovarian cancer. *Neoplasia* 13(5):393-405.
53. Wang W, *et al.* (2016) Effector T Cells Abrogate Stroma-Mediated Chemoresistance in Ovarian Cancer. *Cell* 165(5):1092-1105.
54. Farmer P, *et al.* (2009) A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nat Med* 15(1):68-74.
55. Calon A, *et al.* (2015) Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet* 47(4):320-329.
56. Zhang L, *et al.* (2003) Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *N Engl J Med* 348(3):203-213.
57. Chen DS & Mellman I (2013) Oncology meets immunology: the cancer-immunity cycle. *Immunity* 39(1):1-10.

58. Hamanishi J, *et al.* (2015) Safety and Antitumor Activity of Anti-PD-1 Antibody, Nivolumab, in Patients With Platinum-Resistant Ovarian Cancer. *J Clin Oncol* 33(34):4015-4022.
59. Bohm S, *et al.* (2016) Neoadjuvant Chemotherapy Modulates the Immune Microenvironment in Metastases of Tubo-Ovarian High-Grade Serous Carcinoma. *Clin Cancer Res* 22(12):3025-3036.
60. Kandalafi LE, Powell DJ, Jr., Singh N, & Coukos G (2011) Immunotherapy for ovarian cancer: What's next? *Journal of Clinical Oncology* 29(7):925-933.
61. Norquist BM, *et al.* (2015) Inherited Mutations in Women With Ovarian Carcinoma. *JAMA Oncol*:1-9.
62. Ciriello G, *et al.* (2013) Emerging landscape of oncogenic signatures across human cancers. *Nature Genetics* 45(10):1127-1133.
63. Perets R, *et al.* (2013) Transformation of the fallopian tube secretory epithelium leads to high-grade serous ovarian cancer in Brca;Tp53;Pten models. *Cancer cell* 24(6):751-765.
64. Network TCGAR (2011) Integrated genomic analyses of ovarian carcinoma. *Nature* 474(7353):609-615.
65. Geiger T, Cox J, & Mann M (2010) Proteomic changes resulting from gene copy number variations in cancer cells. *PLoS Genet* 6(9):e1001090.
66. Zhang B, *et al.* (2014) Proteogenomic characterization of human colon and rectal cancer. *Nature* 513(7518):382-387.
67. Mertins P, *et al.* (2016) Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 534(7605):55-62.
68. Freedman RS, Pihl E, Kusyk C, Gallager HS, & Rutledge F (1978) Characterization of an ovarian carcinoma cell line. *Cancer* 42(5):2352-2359.
69. Crickard K, Crickard U, & Yoonessi M (1983) Human ovarian carcinoma cells maintained on extracellular matrix versus plastic. *Cancer Research* 43:2762-2767.
70. Holzel F, Somon WE, Albrecht M, Hansel M, & Dietel M (1983) Cell lines derived from human ovarian carcinomas: Growth stimulation by gonadotropic and steroid hormones. *Journal of the National Cancer Institute* 70:839-845.
71. Benard J, *et al.* (1985) Characterization of a human ovarian adenocarcinoma line, IGROV1, in tissue culture and in nude mice. *Cancer Research* 45:4970-4979.
72. Buick RN, Pullano R, & Trent JM (1985) Comparative properties of five human ovarian adenocarcinoma cell lines. *Cancer Research* 45:3668-3676.
73. Pirker R, *et al.* (1985) Characterization of immunotoxins active against ovarian cancer cell lines. *Immunotox agasinst ovarian cancer* 76:1261-1267.
74. Allen HJ, Porter C, Gamarra M, Piver MS, & Johnson EA (1987) Isolation and morphologic characterization of human ovarian carcinoma cell clusters present in effusions. *Exp Cell Biology* 55(4):194-208.
75. Ince TA, *et al.* (2015) Characterization of twenty-five ovarian tumour cell lines that phenocopy primary tumours. *Nature communications* 6:7419.
76. Beaufort CM, *et al.* (2014) Ovarian Cancer Cell Line Panel (OCCP): clinical importance of in vitro morphological subtypes. *PLoS One* 9(9):e103988.
77. Urzua U, Best L, & Munroe DJ (2010) Microarray proteomic analysis discriminates tumorigenic mouse ovarian surface epithelial cells of divergent aggressive potential. *Mol Biosyst* 6(12):2521-2528.
78. Barretina J, *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483(7391):603-607.

79. Mitra AK, *et al.* (2015) In vivo tumor growth of high-grade serous ovarian cancer cell lines. *Gynecol Oncol*.
80. Elias KM, *et al.* (2015) Beyond genomics: critical evaluation of cell line utility for ovarian cancer research. *Gynecol Oncol* 139(1):97-103.
81. Haley J, *et al.* (2016) Functional characterization of a panel of high-grade serous ovarian cancer cell lines as representative experimental models of the disease. *Oncotarget*.
82. Crick F (1970) Central dogma of molecular biology. *Nature* 227(5258):561-563.
83. Aebersold R & Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422(6928):198-207.
84. International Human Genome Sequencing C (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431(7011):931-945.
85. Ting L, Rad R, Gygi SP, & Haas W (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods* 8(11):937-940.
86. Altelaar AF, Munoz J, & Heck AJ (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* 14(1):35-48.
87. Tanaka K. WH, Ido Y., Akita S., Yoshida Y., Yoshida T. (1988) Protein and polymer analysis up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2(8):151.
88. Karas M & Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60(20):2299-2301.
89. Fenn JB, Mann M, Meng CK, Wong SF, & Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246(4926):64-71.
90. Smith LM, Kelleher NL, & Consortium for Top Down P (2013) Proteoform: a single term describing protein complexity. *Nat Methods* 10(3):186-187.
91. Steen H & Mann M (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 5(9):699-711.
92. Mann M, Kulak NA, Nagaraj N, & Cox J (2013) The coming age of complete, accurate, and ubiquitous proteomes. *Molecular cell* 49(4):583-590.
93. Zubarev RA & Makarov A (2013) Orbitrap mass spectrometry. *Anal Chem* 85(11):5288-5296.
94. 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(12):1367-1372.
95. Cox J, *et al.* (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* 10(4):1794-1805.
96. Ong SE, *et al.* (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1(5):376-386.
97. Geiger T, Cox J, Ostasiewicz P, Wisniewski JR, & Mann M (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature Methods* 7(5):383-385.
98. Ross PL, *et al.* (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3(12):1154-1169.
99. Thompson A, *et al.* (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 75(8):1895-1904.

100. Murphy JP, Everley RA, Coloff JL, & Gygi SP (2014) Combining amine metabolomics and quantitative proteomics of cancer cells using derivatization with isobaric tags. *Anal Chem* 86(7):3585-3593.
101. Mertins P, *et al.* (2012) iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. *Mol Cell Proteomics* 11(6):M111 014423.
102. Bantscheff M, Lemeer S, Savitski MM, & Kuster B (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* 404(4):939-965.
103. Cox J, *et al.* (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 13(9):2513-2526.
104. Humphrey SJ, Azimifar SB, & Mann M (2015) High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nat Biotechnol* 33(9):990-995.
105. Bizzini A, Durussel C, Bille J, Greub G, & Prod'hom G (2010) Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory. *J Clin Microbiol* 48(5):1549-1554.
106. Saffert RT, *et al.* (2011) Comparison of Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometer to BD Phoenix automated microbiology system for identification of gram-negative bacilli. *J Clin Microbiol* 49(3):887-892.
107. Fine JM & Creyssel R (1959) Starch gel electrophoresis studies on abnormal proteins in myeloma and macroglobulinaemia. *Nature* 183(4658):392.
108. Longsworth LG, Shedlovsky T, & Macinnes DA (1939) Electrophoretic Patterns of Normal and Pathological Human Blood Serum and Plasma. *J Exp Med* 70(4):399-413.
109. Hanash SM, Baier LJ, McCurry L, & Schwartz SA (1986) Lineage-related polypeptide markers in acute lymphoblastic leukemia detected by two-dimensional gel electrophoresis. *Proc Natl Acad Sci U S A* 83(3):807-811.
110. Rasmussen HH, Mortz E, Mann M, Roepstorff P, & Celis JE (1994) Identification of transformation sensitive proteins recorded in human two-dimensional gel protein databases by mass spectrometric peptide mapping alone and in combination with microsequencing. *Electrophoresis* 15(3-4):406-416.
111. Mann M, Hendrickson RC, & Pandey A (2001) Analysis of proteins and proteomes by mass spectrometry. *Annual review of biochemistry* 70:437-473.
112. Nagaraj N, *et al.* (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* 7:548.
113. Petricoin EF, *et al.* (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359(9306):572-577.
114. Anonymous (2004) Proteomic diagnostics tested. *Nature* 429(6991):487.
115. Baggerly KA, Morris JS, & Coombes KR (2004) Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* 20(5):777-785.
116. Karsan A, *et al.* (2005) Analytical and preanalytical biases in serum proteomic pattern analysis for breast cancer diagnosis. *Clin Chem* 51(8):1525-1528.
117. Hein MY, *et al.* (2015) A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* 163(3):712-723.

118. Choudhary C & Mann M (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol* 11(6):427-439.
119. Alfaro JA, Sinha A, Kislinger T, & Boutros PC (2014) Onco-proteogenomics: cancer proteomics joins forces with genomics. *Nat Methods* 11(11):1107-1113.
120. Cancer Genome Atlas Research N, *et al.* (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45(10):1113-1120.
121. Deeb SJ, *et al.* (2015) Machine Learning-based Classification of Diffuse Large B-cell Lymphoma Patients by Their Protein Expression Profiles. *Mol Cell Proteomics* 14(11):2947-2960.
122. Wisniewski JR, Dus K, & Mann M (2013) Proteomic workflow for analysis of archival formalin-fixed and paraffin-embedded clinical samples to a depth of 10 000 proteins. *Proteomics Clin Appl* 7(3-4):225-233.
123. Wisniewski JR, Ostasiewicz P, & Mann M (2011) High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J Proteome Res* 10(7):3040-3049.
124. Wisniewski JR, *et al.* (2012) Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol Syst Biol* 8:611.
125. Pozniak Y, *et al.* (2016) System-wide Clinical Proteomics of Breast Cancer Reveals Global Remodeling of Tissue Homeostasis. *Cell Syst* 2(3):172-184.
126. Balgley BM, *et al.* (2009) Evaluation of archival time on shotgun proteomics of formalin-fixed and paraffin-embedded tissues. *J Proteome Res* 8(2):917-925.
127. Ostasiewicz P, Zielinska DF, Mann M, & Wisniewski JR (2010) 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(7):3688-3700.
128. Emmert-Buck MR, *et al.* (1996) Laser capture microdissection. *Science* 274(5289):998-1001.
129. Ornstein DK, *et al.* (2000) Proteomic analysis of laser capture microdissected human prostate cancer and in vitro prostate cell lines. *Electrophoresis* 21(11):2235-2242.
130. Hill JJ, *et al.* (2011) Identification of vascular breast tumor markers by laser capture microdissection and label-free LC-MS. *J Proteome Res* 10(5):2479-2493.
131. Jones MB, *et al.* (2002) Proteomic analysis and identification of new biomarkers and therapeutic targets for invasive ovarian cancer. *Proteomics* 2(1):76-84.
132. Roulhac PL, *et al.* (2011) Microproteomics: quantitative proteomic profiling of small numbers of laser-captured cells. *Cold Spring Harb Protoc* 2011(2):pdb prot5573.
133. Etzioni R, *et al.* (2003) The case for early detection. *Nat Rev Cancer* 3(4):243-252.
134. Anderson NL & Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 1(11):845-867.
135. Geyer Philipp E, *et al.* (2016) Plasma proteome profiling to assess human health and disease. *Cell Systems* 2(3):185-195.
136. Rifai N, Gillette MA, & Carr SA (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* 24(8):971-983.

137. Parker CE, Pearson TW, Anderson NL, & Borchers CH (2010) Mass-spectrometry-based clinical proteomics--a review and prospective. *Analyst* 135(8):1830-1838.
138. Tyanova S, *et al.* (2016) Proteomic maps of breast cancer subtypes. *Nature communications* 7:10259.
139. Boersema PJ, Geiger T, Wisniewski JR, & Mann M (2013) Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Mol Cell Proteomics* 12(1):158-171.
140. Good DM, *et al.* (2007) Body fluid proteomics for biomarker discovery: lessons from the past hold the key to success in the future. *J Proteome Res* 6(12):4549-4555.
141. Mischak H, *et al.* (2010) Comprehensive human urine standards for comparability and standardization in clinical proteome analysis. *Proteomics Clin Appl* 4(4):464-478.
142. Yin HR, *et al.* (2013) Hyperplex-MRM: a hybrid multiple reaction monitoring method using mTRAQ/iTRAQ labeling for multiplex absolute quantification of human colorectal cancer biomarker. *J Proteome Res* 12(9):3912-3919.
143. Schadendorf D, *et al.* (2015) Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. *J Clin Oncol* 33(17):1889-1894.
144. Gubin MM, Artyomov MN, Mardis ER, & Schreiber RD (2015) Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J Clin Invest* 125(9):3413-3421.
145. Ohminami H, Yasukawa M, & Fujita S (2000) HLA class I-restricted lysis of leukemia cells by a CD8(+) cytotoxic T-lymphocyte clone specific for WT1 peptide. *Blood* 95(1):286-293.
146. Kawashima I, *et al.* (1999) Identification of HLA-A3-restricted cytotoxic T lymphocyte epitopes from carcinoembryonic antigen and HER-2/neu by primary in vitro immunization with peptide-pulsed dendritic cells. *Cancer Res* 59(2):431-435.
147. Parkhurst MR, *et al.* (2011) T cells targeting carcinoembryonic antigen can mediate regression of metastatic colorectal cancer but induce severe transient colitis. *Mol Ther* 19(3):620-626.
148. Li B, *et al.* (2016) Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat Genet* 48(7):725-732.
149. Yadav M, *et al.* (2014) Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* 515(7528):572-576.
150. Simpson AJ, Caballero OL, Jungbluth A, Chen YT, & Old LJ (2005) Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer* 5(8):615-625.
151. Odunsi K, *et al.* (2014) Epigenetic potentiation of NY-ESO-1 vaccine therapy in human ovarian cancer. *Cancer Immunol Res* 2(1):37-49.
152. Shraibman B, Melamed Kadosh D, Barnea E, & Admon A (2016) HLA peptides derived from tumor antigens induced by inhibition of DNA methylation for development of drug-facilitated immunotherapy. *Mol Cell Proteomics*.
153. Klar AS, Gopinadh J, Kleber S, Wadle A, & Renner C (2015) Treatment with 5-Aza-2'-Deoxycytidine Induces Expression of NY-ESO-1 and Facilitates Cytotoxic T Lymphocyte-Mediated Tumor Cell Killing. *PLoS One* 10(10):e0139221.

154. Wang L, *et al.* (2015) Decitabine Enhances Lymphocyte Migration and Function and Synergizes with CTLA-4 Blockade in a Murine Ovarian Cancer Model. *Cancer Immunol Res* 3(9):1030-1041.
155. van der Bruggen P, *et al.* (1991) A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* 254(5038):1643-1647.
156. Sahin U, *et al.* (1995) Human neoplasms elicit multiple specific immune responses in the autologous host. *Proc Natl Acad Sci U S A* 92(25):11810-11813.
157. Chen YT, *et al.* (1997) A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. *Proc Natl Acad Sci U S A* 94(5):1914-1918.
158. Smith HA & McNeel DG (2010) The SSX family of cancer-testis antigens as target proteins for tumor therapy. *Clin Dev Immunol* 2010:150591.
159. Matsushita H, *et al.* (2012) Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* 482(7385):400-404.
160. Castle JC, *et al.* (2012) Exploiting the mutanome for tumor vaccination. *Cancer Res* 72(5):1081-1091.
161. Cai A, *et al.* (2012) Mutated BCR-ABL generates immunogenic T-cell epitopes in CML patients. *Clin Cancer Res* 18(20):5761-5772.
162. Andreatta M, Nielsen M (2015) Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*.
163. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, & Mann M (2015) Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics* 14(3):658-673.
164. Dutoit V, *et al.* (2012) Exploiting the glioblastoma peptidome to discover novel tumour-associated antigens for immunotherapy. *Brain* 135(Pt 4):1042-1054.
165. Bassani-Sternberg M, *et al.* (2010) Soluble plasma HLA peptidome as a potential source for cancer biomarkers. *Proc Natl Acad Sci U S A* 107(44):18769-18776.
166. Barouch D, *et al.* (1995) HLA-A2 subtypes are functionally distinct in peptide binding and presentation. *J Exp Med* 182(6):1847-1856.
167. Hunt D, *et al.* (1992) Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* 255(5049):1261-1263.
168. Berns EM & Bowtell DD (2012) The changing view of high-grade serous ovarian cancer. *Cancer Res* 72(11):2701-2704.
169. Schug TT, Berry DC, Shaw NS, Travis SN, & Noy N (2007) Opposing effects of retinoic acid on cell growth result from alternate activation of two different nuclear receptors. *Cell* 129(4):723-733.
170. Chen YT, *et al.* (2005) Identification of cancer/testis-antigen genes by massively parallel signature sequencing. *Proc Natl Acad Sci U S A* 102(22):7940-7945.
171. Anderson NL, Ptolemy AS, & Rifai N (2013) The riddle of protein diagnostics: future bleak or bright? *Clin Chem* 59(1):194-197.
172. Aebersold R & Mann M (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* 537(7620):347-355.
173. Coscia F, *et al.* (2016) Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. *Nat. Commun.* 7:12645.

174. Zhang W, *et al.* (2015) DNA hypomethylation-mediated activation of Cancer/Testis Antigen 45 (CT45) genes is associated with disease progression and reduced survival in epithelial ovarian cancer. *Epigenetics* 10(8):736-748.
175. Rudolph P, *et al.* (1999) Ki-A10, a germ cell nuclear antigen retained in a subset of germ cell-derived tumors. *Am J Pathol* 154(3):795-803.
176. Heidebrecht HJ, *et al.* (2006) Characterization and expression of CT45 in Hodgkin's lymphoma. *Clin Cancer Res* 12(16):4804-4811.
177. Andrade VC, *et al.* (2009) Frequency and prognostic relevance of cancer testis antigen 45 expression in multiple myeloma. *Experimental hematology* 37(4):446-449.
178. Chen YT, *et al.* (2009) Cancer/testis antigen CT45: analysis of mRNA and protein expression in human cancer. *Int J Cancer* 124(12):2893-2898.
179. Li H, *et al.* (2014) Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. *Oncotarget* 5(3):587-598.
180. Chiappinelli KB, *et al.* (2015) Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* 162(5):974-986.
181. Plumb JA, Strathdee G, Sludden J, Kaye SB, & Brown R (2000) Reversal of Drug Resistance in Human Tumor Xenografts by 2'-Deoxy-5-azacytidine-induced Demethylation of the hMLH1 Gene Promoter. *Cancer Research* 60(21):6039-6044.
182. Matei D, *et al.* (2012) Epigenetic resensitization to platinum in ovarian cancer. *Cancer Res* 72(9):2197-2205.
183. Nakada S, Chen GI, Gingras AC, & Durocher D (2008) PP4 is a gamma H2AX phosphatase required for recovery from the DNA damage checkpoint. *EMBO reports* 9(10):1019-1026.
184. Lee DH, *et al.* (2010) A PP4 phosphatase complex dephosphorylates RPA2 to facilitate DNA repair via homologous recombination. *Nature structural & molecular biology* 17(3):365-372.
185. Lee DH, *et al.* (2012) *Phosphoproteomic analysis reveals that PP4 dephosphorylates KAP-1 impacting the DNA damage response* pp 2403-2415.
186. Lee DH, *et al.* (2014) Dephosphorylation enables the recruitment of 53BP1 to double-strand DNA breaks. *Molecular cell* 54(3):512-525.
187. Liu J, *et al.* (2014) Protein phosphatase PP4 is involved in NHEJ-mediated repair of DNA double-strand breaks. *Cell Cycle* 11(14):2643-2649.
188. Gingras AC, *et al.* (2005) A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol Cell Proteomics* 4(11):1725-1740.
189. Chowdhury D, *et al.* (2008) A PP4-phosphatase complex dephosphorylates gamma-H2AX generated during DNA replication. *Molecular cell* 31(1):33-46.
190. Shaltiel IA, *et al.* (2014) Distinct phosphatases antagonize the p53 response in different phases of the cell cycle. *Proceedings of the National Academy of Sciences* 111(20):7313-7318.
191. Lee DH, *et al.* (2012) *Phosphoproteomic analysis reveals that PP4 dephosphorylates KAP-1 impacting the DNA damage response.* *The EMBO Journal* 31(10):2403-2415.

192. Ziv Y, *et al.* (2006) Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. *Nature cell biology* 8(8):870-876.
193. Goodarzi AA, Jeggo P, & Lobrich M (2010) The influence of heterochromatin on DNA double strand break repair: Getting the strong, silent type to relax. *DNA Repair* 9(12):1273-1282.
194. Goodarzi AA, *et al.* (2008) ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Molecular cell* 31(2):167-177.
195. Kalousi A, *et al.* (2015) The Nuclear Oncogene SET Controls DNA Repair by KAP1 and HP1 Retention to Chromatin. *Cell Reports* 11(1):149-163.
196. Wierer M & Mann M (2016) Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. *Human molecular genetics* 25(R2):R106-R114.
197. Li L, *et al.* (2016) SIRT7 is a histone desuccinylase that functionally links to chromatin compaction and genome stability. *Nature communications* 7:12235.
198. Snyder A, *et al.* (2014) Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *New England Journal of Medicine* 371(23):2189-2199.
199. Le DT, *et al.* (2015) PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *New England Journal of Medicine* 372(26):2509-2520.
200. Hamanishi J, *et al.* (2015) Safety and Antitumor Activity of Anti-PD-1 Antibody, Nivolumab, in Patients With Platinum-Resistant Ovarian Cancer. *Journal of Clinical Oncology* 33(34):4015-4022.
201. Eckert MA, *et al.* (2016) Genomics of ovarian cancer progression reveals diverse metastatic trajectories including intraepithelial metastasis to the fallopian tube. *Cancer Discov.*
202. Cann KL & Dellaire G (2011) Heterochromatin and the DNA damage response: the need to relax. *Biochem Cell Biol* 89(1):45-60.
203. Friedman JR, *et al.* (1996) KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes & Development* 10(16):2067-2078.
204. Wang C, *et al.* (2005) MDM2 interaction with nuclear corepressor KAP1 contributes to p53 inactivation. *The EMBO Journal* 24(18):3279-3290.
205. Yang X, *et al.* (2011) A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods* 8(8):659-661.
206. Cormier CY, *et al.* (2010) Protein Structure Initiative Material Repository: an open shared public resource of structural genomics plasmids for the biological community. *Nucleic Acids Res.* 38(Database issue):D743-749.
207. Cormier CY, *et al.* (2011) PSI:BiologY-materials repository: a biologist's resource for protein expression plasmids. *J Struct Funct Genomics* 12(2):55-62.
208. Seiler CY, *et al.* (2014) DNASU plasmid and PSI:BiologY-Materials repositories: resources to accelerate biological research. *Nucleic Acids Res.* 42(Database issue):D1253-1260.
209. Michalski A, *et al.* (2011) Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* 10(9):M111.011015.
210. Andreatta M, Lund O, & Nielsen M (2013) Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics* 29(1):8-14.

211. Rapin N, Hoof I, Lund O, & Nielsen M (2010) The MHC motif viewer: a visualization tool for MHC binding motifs. *Current protocols in immunology* Chapter 18:Unit 18 17.
212. Thomsen MC & Nielsen M (2012) Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res* 40(Web Server issue):W281-287.
213. Andreatta M & Nielsen M (2016) Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 32(4):511-517.
214. Nandhakumar S, *et al.* (2011) Evaluation of DNA damage using single-cell gel electrophoresis (Comet Assay). *Journal of Pharmacology & Pharmacotherapeutics* 2(2):107-111.
215. Gyori BM, Venkatachalam G, Thiagarajan PS, Hsu D, & Clement M-V (2014) OpenComet: An automated tool for comet assay image analysis. *Redox Biology* 2:457-465.
216. Hubner NC & Mann M (2011) Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). *Methods* 53(4):453-459.
217. Carpenter AE, *et al.* (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology* 7(10):R100.
218. Tyanova S, *et al.* (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*.
219. Tusher VG, Tibshirani R, & Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98(9):5116-5121.
220. Yang D, *et al.* (2013) Integrated analyses identify a master microRNA regulatory network for the mesenchymal subtype in serous ovarian cancer. *Cancer cell* 23(2):186-199.
221. Tothill RW, *et al.* (2008) Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical cancer research : an official journal of the American Association for Cancer Research* 14(16):5198-5208.
222. Ulanovskaya OA, Zuhl AM, & Cravatt BF (2013) NNMT promotes epigenetic remodeling in cancer by creating a metabolic methylation sink. *Nat Chem Biol* 9(5):300-306.
223. Sperber H, *et al.* (2015) The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. *Nature cell biology* 17(12):1523-1535.
224. The Cancer Genome Atlas N (2011) Integrated genomic analyses of ovarian carcinoma. *Nature* 474:609-615.
225. Zhang H, *et al.* (2016) Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell*.
226. Alexandrov LB, *et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463):415-421.
227. Francavilla C, *et al.* (2017) Phosphoproteomics of primary cells reveals druggable kinase signatures in ovarian cancer. *Cell reports* 18(13):3242-3256.
228. Carey MS, *et al.* (2010) Functional proteomic analysis of advanced serous ovarian cancer using reverse phase protein array: TGF- $\beta$  pathway signaling indicates response to primary chemotherapy. *Clin. Cancer Res.* 16(10):2852-2860.

229. Polyak K, Haviv I, & Campbell IG (2009) Co-evolution of tumor cells and their microenvironment. *Trends Genet.* 25:30.
230. Kalluri R (2016) The biology and function of fibroblasts in cancer. *Nat. Rev. Cancer* 16(9):582-598.
231. Martinez-Outschoorn UE, Sotgia F, & Lisanti MP (2012) Power Surge: Supporting cells "fuel" cancer cell metabolism. *Cell Metab.* 15(1):4-5.
232. Leung CS, *et al.* (2014) Calcium-dependent FAK/CREB/TNNC1 signalling mediates the effect of stromal MFAP5 on ovarian cancer metastatic potential. *Nat. Commun.* 5:5092.
233. Altelaar AF & Heck AJ (2012) Trends in ultrasensitive proteomics. *Current opinion in chemical biology* 16(1-2):206-213.
234. Koli K, *et al.* (2004) Disruption of LTBP-4 function reduces TGF-beta activation and enhances BMP-4 signaling in the lung. *J. Cell Biol.* 167(1):123-133.
235. Tian Y, *et al.* (2016) Serum deprivation response inhibits breast cancer progression by blocking transforming growth factor-beta signaling. *Cancer Sci.* 107(3):274-280.
236. Aksoy S, Szumlanski CL, & Weinshilboum RM (1994) Human liver nicotinamide N-methyltransferase. cDNA cloning, expression, and biochemical characterization. *J. Biol. Chem.* 269(20):14835-14840.
237. Kraus D, *et al.* (2014) Nicotinamide N-methyltransferase knockdown protects against diet-induced obesity. *Nature* 508(7495):258-262.
238. Hong S, *et al.* (2015) Nicotinamide N-methyltransferase regulates hepatic nutrient metabolism through Sirt1 protein stabilization. *Nat. Med.* 21(8):887-894.
239. Palanichamy K, *et al.* (2016) NNMT Silencing Activates Tumor Suppressor PP2A, Inactivates Oncogenic STKs, and Inhibits Tumor Forming Ability. *Clin. Cancer Res.*
240. Chiyoda T, *et al.* (2017) Loss of BRCA1 in the Cells of Origin of Ovarian Cancer Induces Glycolysis: A Window of Opportunity for Ovarian Cancer Chemoprevention. *Cancer Prev. Res. (Phila.)* 10(4):255-266.
241. Perets R, *et al.* (2013) Transformation of the fallopian tube secretory epithelium leads to high-grade serous ovarian cancer in Brca; Tp53; Pten models. *Cancer Cell* 24:751-765.
242. Alam H, Gu B, & Lee MG (2015) Histone methylation modifiers in cellular signaling pathways. *Cell. Mol. Life Sci.* 72(23):4577-4592.
243. Barski A, *et al.* (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129(4):823-837.
244. Bernstein BE, *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125(2):315-326.
245. Orimo A, *et al.* (2005) Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL 12 secretion. *Cell* 121:335-348.
246. Olumi AF, *et al.* (1999) Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res.* 59:5002-5011.
247. Kenny HA, *et al.* (2011) Targeting the urokinase plasminogen activator receptor inhibits ovarian cancer metastasis. *Clin. Cancer Res.* 17(3):459-471.
248. Eckert MA, *et al.* (2016) Genomics of ovarian cancer progression reveals diverse metastatic trajectories including intraepithelial metastasis to the fallopian tube. *Cancer Discov.* 6(12):1342-1351.

249. Mentch SJ & Locasale JW (2015) One-carbon metabolism and epigenetics: understanding the specificity. *Ann. N. Y. Acad. Sci.*
250. Ye C, Sutter BM, Wang Y, Kuang Z, & Tu BP (2017) A metabolic function for phospholipid and histone methylation. *Mol. Cell.*
251. Hu M, *et al.* (2005) Distinct epigenetic changes in the stromal cells of breast cancers. *Nat. Genet.* 37:899-905.
252. Zong Y, *et al.* (2012) Stromal epigenetic dysregulation is sufficient to initiate mouse prostate cancer via paracrine Wnt signaling. *Proc. Natl. Acad. Sci. U. S. A.* 109(50):E3395-3404.
253. Scheltema RA, *et al.* (2014) The Q Exactive HF, a benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol Cell Proteomics* 13(12):3698-3708.
254. Kenny HA, *et al.* (2014) Mesothelial cells promote early ovarian cancer metastasis through fibronectin secretion. *J. Clin. Invest.* 124(10):4614-4628.
255. Mitra AK, *et al.* (2012) MicroRNAs reprogram normal fibroblasts into cancer-associated fibroblasts in ovarian cancer. *Cancer Discov.* 2(12):1100-1108.
256. Zheng Y, *et al.* (2012) Total kinetic analysis reveals how combinatorial methylation patterns are established on lysines 27 and 36 of histone H3. *Proc. Natl. Acad. Sci. U. S. A.* 109(34):13549-13554.
257. Bowen NJ, *et al.* (2007) Emerging roles for PAX8 in ovarian cancer and endosalpingeal development. *Gynecol Oncol* 104(2):331-337.
258. Lengyel E, Fleming S, McEwen KA, Montag A, & Temkin SM (2013) Serial sectioning of the fallopian tube allows for improved identification of primary fallopian tube carcinoma. *Gynecol Oncol* 129(1):120-123.
259. Verhaak RG, *et al.* (2013) Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *The Journal of clinical investigation* 123(1):517-525.
260. Pfeifer GP (2012) Protein phosphatase PP4: role in dephosphorylation of KAP1 and DNA strand break repair. *Cell cycle* 11(14):2590-2591.
261. Kim MS, *et al.* (2003) Inhibition of histone deacetylase increases cytotoxicity to anticancer drugs targeting DNA. *Cancer Res* 63(21):7291-7300.
262. Lipowska-Bhalla G, Gilham DE, Hawkins RE, & Rothwell DG (2012) Targeted immunotherapy of cancer with CAR T cells: achievements and challenges. *Cancer immunology, immunotherapy : CII* 61(7):953-962.
263. Wrangle J, *et al.* (2013) Alterations of immune response of Non-Small Cell Lung Cancer with Azacytidine. *Oncotarget* 4(11):2067-2079.
264. Hanahan D & Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144(5):646-674.
265. Lee JY, *et al.* (2015) Tumor evolution and intratumor heterogeneity of an epithelial ovarian cancer investigated using next-generation sequencing. *BMC cancer* 15:85.
266. McPherson A, *et al.* (2016) Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat Genet* 48(7):758-767.
267. Nagaraj N, *et al.* (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol Cell Proteomics* 11(3):M111 013722.
268. Beck M, *et al.* (2011) The quantitative proteome of a human cell line. *Mol Syst Biol* 7:549.
269. Kanu AB, Dwivedi P, Tam M, Matz L, & Hill HH, Jr. (2008) Ion mobility-mass spectrometry. *J Mass Spectrom* 43(1):1-22.

## Acknowledgments

First of all, I would like to deeply thank Prof. Dr. Matthias Mann that I could work in such an excellent international laboratory with so many great and inspiring people and scientists from all around the world. Matthias, thanks for all the great support, your trust and patience and all the scientific freedom you gave me. I feel much honored that I could be part of your team and work on such an important clinical project. Also, your visionary and positive way of thinking has strongly influenced me as scientist.

I am very thankful to have worked with the best collaborator I can imagine. Prof. Dr. Ernst Lengyel has made my PhD project a fascinating journey between science and daily clinical practice. The way you connect research with patient care was outstanding to me and gave me so many important insights beyond the usual lab work. Thank you for all your support over the years, the endless skype and email sessions for data discussion, and all you did and organized for me during my Chicago visit. You are the best example for me how one can combine excellent clinical work with innovative and successful research. In addition, many thanks to Dr. Marion Curtis, Dr. Mark Eckert, Dr. Karen Watters, and Bradley Ashcroft for their fantastic work, all the scientific input and the great support during my visit. I fully enjoyed working with you on such important projects.

I would also like to thank Dr. Paul Boersema for his excellent supervision during the first few months in the lab. Without help from additional people, the last years would have been much more difficult. In particular, many thanks to Dirk Wischnewski for his great laboratory support and all the important discussions, Korbinian Mayr, Dr. Igor Paron and Gabriele Sowa for their mass spec assistance, Alison Dalfovo, Theresa Schneider, Mario Oroshi and Helen de los Santos for all the important administrative and laboratory support. Furthermore, thanks to Prof. Dr. Jürgen Cox and Dr. Stefka Tyanova for all the discussions and bioinformatic support. All my office mates, in particular Dr. Markus Räschle and Dr. Michael Wierer for all the scientific input and the fantastic atmosphere. I thank Dr. Felix Meissner for being part of my TAC committee and all the scientific input.

Prof. Dr. Jürgen Cox, Dr. Dietmar Martin, Dr. Franz Herzog, and Prof. Dr. Karl-Peter Hopfner for being part of my thesis committee.

I also want to thank the IMPRS coordination office for giving me the opportunity to work in such a fantastic scientific environment with so many creative young minds. It has been an amazing time.

I also want to thank my family and friends for all the support. My father for always being proud and encouraging and the ever-increasing interest in science ☺. I finally thank my girlfriend Trixi for the great seven years. Thanks for being there by my side.

# Curriculum Vitae

---

## PERSONAL INFORMATION

Name	Fabian Coscia
Date of birth	1987/02/08
Place of birth	Ebersberg, Germany
Nationality	German

## EDUCATION

06/2015 – 09/2015      **International graduate student** at the University of Chicago, Lengyel ovarian cancer laboratory, Department of Obstetrics and Gynecology/Section of Gynecologic Oncology.

Since 09/2012      **PhD candidate** in the laboratory of Matthias Mann, Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany.  
Title: *Ovarian cancer proteomics*

10/2009 – 12/2011      **Master of Science Studies** in Molecular Biotechnology at the Technical University of Munich.  
Focus: Molecular Medicine, Molecular Microbiology, Protein Biochemistry & Engineering and Biotechnology of higher organisms.

**Master's Thesis** in Molecular Oncology at the Ludwig-Maximilian-University of Munich in the laboratory of Olivier Gires.  
Title: *Functional characterization of the tumour-associated antigen EpCAM in murine embryonic stem cells.*

10/2006 – 10/2009      **Bachelor of Science Studies** in Molecular Biotechnology at the Technical University of Munich.

## SCIENTIFIC CONFERENCES

- 2016 **MaxQuant Summer School for Computational Proteomics**, Oxford, United Kingdom (oral presentation).
- 2015 **International Meeting of the German Society for Cell Biology**, Cologne, Germany (poster presentation).
- 8th Munich Life Science Symposium, Interact**, Munich, Germany (poster presentation)
- 2014 **MaxQuant Summer School for Computational Proteomics**, Bethesda, United States (oral presentation).
- American Association for Cancer Research**, Washington D.C., United States
- M4 Biomarker Conference**, Munich, Germany
- European Association for Cancer Research**, Munich, Germany
- 2013 **MaxQuant Summer School for Computational Proteomics**, Martinsried, Germany (oral presentation).
- AACR: Advances in Ovarian Cancer Research**, Miami, United States

## TEACHING

- 2016 **MaxQuant Summer School for Computational Proteomics**, Oxford, United Kingdom, 1-week practical course.
- Advanced Proteomics Methods course**, MPI of Biochemistry, Munich, Germany, 2-day workshop
- 2015 **Advanced Proteomics Methods course**, MPI of Biochemistry, Munich, Germany, 2-day workshop
- 2014 **MaxQuant Summer School for Computational Proteomics**, Bethesda, United States, 1-week practical course.
- Practical course in biochemistry**, 1-week supervision, LMU Munich.
- 2013 **MaxQuant Summer School for Computational Proteomics**, Munich, Germany, 1-week practical course.
- Practical course in biochemistry**, 1-week supervision, LMU Munich.

## FELLOWSHIPS

Since 09/2012

**International Max Planck Research School for Molecular and Cellular Life Sciences: From Biology to Medicine (IMPRS-LS)**

## SCIENTIFIC PUBLICATIONS (\*equal contribution)

**Coscia F\***, Watters KM\*, Curtis M, Eckert MA, Chiang CY, Tyanova S, Montag A, Lastra RR, Lengyel E, Mann M. Integrative proteomic profiling of ovarian cancer cell lines reveals precursor-cell associated proteins and functional status. **Nat. Commun.** 7:12645 doi: 10.1038/ncomms12645 (2016).

Hachmeister, M., Bobowski, K. D., Hogg, S., Dislich, B., Fukumori, A., Eggert, C., Mack, B., Kremling, H., Sarrach, S., **Coscia, F.**, Zimmermann, W., Steiner, H., Lichtenthaler, S. F., Gires, O. Regulated intramembrane proteolysis and degradation of murine epithelial cell adhesion molecule mEpCAM. **PLoS One** 8, e71836, doi:10.1371/journal.pone.0071836 (2013).

Pelz, C., Steininger, S., Weiss, C., **Coscia, F.** & Vogelmann, R. A novel inhibitory domain of *Helicobacter pylori* protein CagA reduces CagA effects on host cell biology. **J Biol Chem** 286, 8999-9008, doi:10.1074/jbc.M110.166504 (2011).