From nanometers to centimeters: Imaging across spatial scales with smart computer-aided microscopy

Dissertation der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

David Hörl

München, 17. Februar 2020

Diese Dissertation wurde angefertigt
unter der Leitung von Prof. Dr. Heinrich Leonhardt
im Bereich Humanbiologie und BioImaging
an der Ludwig-Maximilians-Universität München

Erstgutachter: Prof. Dr. Heinrich Leonhardt
Zweitgutachter: Prof. Dr. Thorben Cordes
Datum der Abgabe: 17. Februar 2020
Datum der mündlichen Prüfung: 13. Mai 2020

Erklärung

Ich versichere hiermit an Eides statt, dass meine Dissertation selbständig und ohne unerlaubte Hilfsmittel angefertigt worden ist. Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt. Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

München, den 17. Februar 2020

David Hörl

Abstract

Microscopes have been an invaluable tool throughout the history of the life sciences, as they allow researchers to observe the miniscule details of living systems in space and time. However, modern biology studies complex and non-obvious phenotypes and their distributions in populations and thus requires that microscopes evolve from visual aids for anecdotal observation into instruments for objective and quantitative measurements. To this end, many cutting-edge developments in microscopy are fuelled by innovations in the computational processing of the generated images. Computational tools can be applied in the early stages of an experiment, where they allow for reconstruction of images with higher resolution and contrast or more colors compared to raw data. In the final analysis stage, state-of-the-art image analysis pipelines seek to extract interpretable and humanly tractable information from the high-dimensional space of images.

In the work presented in this thesis, I performed super-resolution microscopy and wrote image analysis pipelines to derive quantitative information about multiple biological processes. I contributed to studies on the regulation of DNMT1 by implementing machine learning-based segmentation of replication sites in images and performed quantitative statistical analysis of the recruitment of multiple DNMT1 mutants. To study the spatiotemporal distribution of DNA damage response I performed STED microscopy and could provide a lower bound on the size of the elementary spatial units of DNA repair. In this project, I also wrote image anal-

ysis pipelines and performed statistical analysis to show a decoupling of DNA density and heterochromatin marks during repair. More on the experimental side, I helped in the establishment of a protocol for many-fold color multiplexing by iterative labelling of diverse structures via DNA hybridization. Turning from small scale details to the distribution of phenotypes in a population, I wrote a reusable pipeline for fitting models of cell cycle stage distribution and inhibition curves to high-throughput measurements to quickly quantify the effects of innovative antiproliferative antibody-drug-conjugates.

The main focus of the thesis is BigStitcher, a tool for the management and alignment of terabyte-sized image datasets. Such enormous datasets are nowadays generated routinely with light-sheet microscopy and sample preparation techniques such as clearing or expansion. Their sheer size, high dimensionality and unique optical properties poses a serious bottleneck for researchers and requires specialized processing tools, as the images often do not fit into the main memory of most computers. BigStitcher primarily allows for fast registration of such many-dimensional datasets on conventional hardware using optimized multi-resolution alignment algorithms. The software can also correct a variety of aberrations such as fixed-pattern noise, chromatic shifts and even complex sample-induced distortions. A defining feature of BigStitcher, as well as the various image analysis scripts developed in this work is their interactivity. A central goal was to leverage the user's expertise at key moments and bring innovations from the big data world to the lab with its smaller and much more diverse datasets without replacing scientists with automated black-box pipelines. To this end, BigStitcher was implemented as a user-friendly plug-in for the open source image processing platform Fiji and provides the users with a nearly instantaneous preview of the aligned images and opportunities for manual control of all processing steps. With its powerful features and ease-of-use, BigStitcher paves the way to the routine application of light-sheet microscopy and other methods producing equally large datasets.

Contents

Chapter 1

Introduction

A defining feature of life as we know it is the formation and upholding of intricate, non-random molecular structures in space and time. Life seems to be able to, in contrast to physical principles at first sight, perpetuate order over long time periods by *feeding on negative entropy* in its environment, as famously described by Erwin Schrödinger in his seminal book *What is life?* [1]. How the complex spatiotemporal patterns of life are formed from molecules obeying physical and chemical laws can be seen as the main question studied by modern biology over more than a century.

Many key principles of molecular biology, like the structure of deoxyribonucleic acid (DNA), the genetic code or the mechanisms of gene expression have been unraveled in the decades since Schrödinger posed his famous central question. But while the main mechanics of life might seem resolved at first glance, there are many layers of epigenetic regulation, e.g. non-coding ribonucleic acids (RNAs) or DNA and histone modifications, with myriads of interacting molecules on top of them. Progress is made continuously, but, as more and more factors in the systems of life are discovered, it is safe to assume that it will be a long time still until biologists can answer with confidence the simple question of what life is.

Nowadays, a relatively complete census of the molecules present in a biological system can easily be established by biochemical methods, like (R/D)NA sequencing or mass spectroscopy, albeit normally at the expense of spatial and temporal information. Clever chemical tricks allow researchers to also encode this information into the molecules of a system, for example by cross-linking of spatially close molecules, making it possible to reconstruct it by computational means as is done in chromatin immunoprecipitation-eequencing (ChIP-Seq), chromatin conformation capture or DNA microscopy, for example [2–4]. The rapid advances in omics technologies over the last years have reduced the size of systems studied by biochemical methods to single cells and it is safe to assume that human ingenuity will continue to push the boundaries of what information can be extracted from omics data.

But there is also a second line of investigation that has accompanied the science of biology since its inception, which consists of methods that aim to produce an image of the systems of life in space and time as perceived by us humans. For example, microscopes made the first observations of eukaryotic cells and microbes passible in the seventeenth century and X-ray crystallography and (cryo-)electron microscopy (EM) made it possible to "see" the large biomolecules contributing to life in the twentieth century. Improvements in microscope instrumentation and sample preparation continue to be made, with the main goal of producing a *better* representation of biological reality. Better might mean many things: improved spatial or temporal resolution, but also specificity to a subset of components of the system under study. The results of imaging experiments often appear more natural to scientists: since humans are visual animals, microscopy images seem more compatible with the human mind than list of tens of thousands of relative expression levels derived from an RNA-Seq experiment, for example. Whereas it is obvious that the high-dimensional results of omics must be compressed by computational statistics into a low-dimensional representation before arriving at an interpretable

result, microscopy-based experiments often result in the qualitative interpretation by a trained expert. As more and more of "simple" questions about phenotypes that are obvious to humans are answered, the need for data analysis to infer information from imaging experiments increases, for example when features of an entire population of cells need to be studied that might be ignored by an experimenter subjectively searching for "representative" cells. Like omics data, images can be seen as a high-dimensional representation of reality that can be compressed into a manageable number of quantitative measurements by computational means.

Modern microscopy-based science can be seen as a two-step process (figure 1.1): first, to provide an as-faithful-as-possible image of reality and secondly, to extract meaningful scientific insights from it. From the beginning, but even more as one goes to later stages, computational tools are enabling experiments: computational methods can, for example, improve spatial resolution, e.g. through structured illumination or single molecule localization, or allow increased spectral multiplexing [5] during image acquisition but become indispensable during later stages of the pipeline. While the technical challenges to image interpretation have long been how to preserve and record an image of the specimen, for example by drawing or photography, digital image detectors have opened the way to interpreting images as just a collection of numbers and image analysis pipelines that derive quantitative measurements from them.

1.0.1 A brief history of microscopy

The development of biology into a modern science from the 17^{th} century onwards happened in parallel to and was often fueled by novel developments in microscopy. Indeed, the moment the English naturalist Robert Hooke observed cork under a self-built microscope in 1665 and saw regular structures that he termed "cells" can be seen as the inception of cell biology. Likewise, the Dutch lens maker Antoni

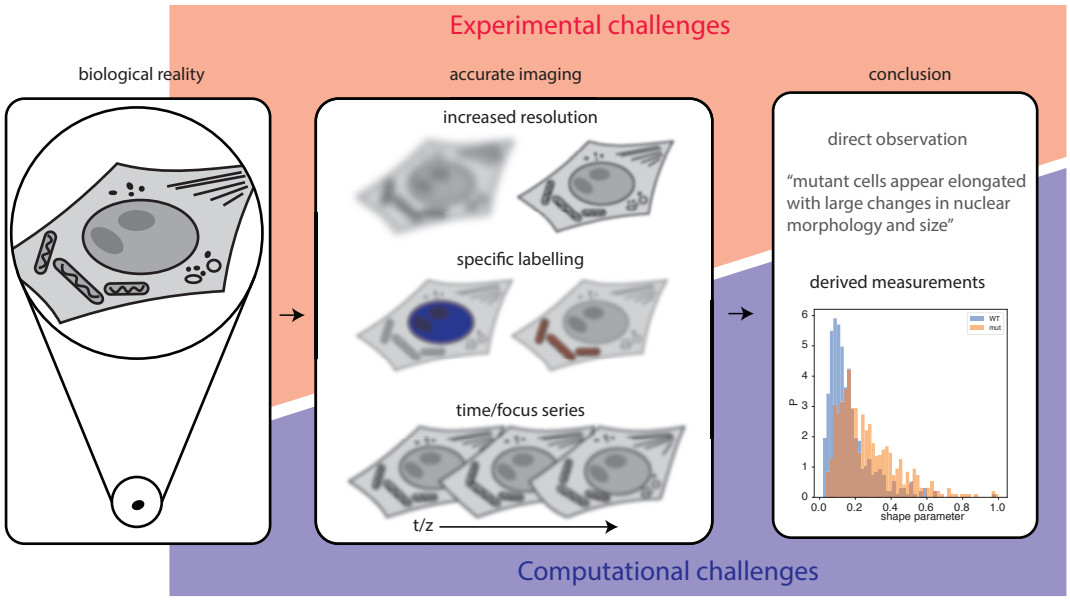


Figure 1.1: Simplified workflow and challenges in image-based biological science. The process can be seen as two steps: first, the goal is to produce an as faithful as possible image (middle) of the microscopic biological reality (left), e.g. an image with high spatial resolution, molecular specificity or time series of images of living specimen. While a scientific conclusion (right) might be drawn by an expert via direct observation, complex phenotypes often require quantitative measurements to be derived from the images before a conclusion can be reached. All steps in the experimental pipeline can be improved by experimental (microscope instrumentation or sample preparation) or computational (mostly software, though specialized hardware, e.g. GPUs, might be necessary in some cases) measures, with the latter becoming more important in later steps of the pipeline.

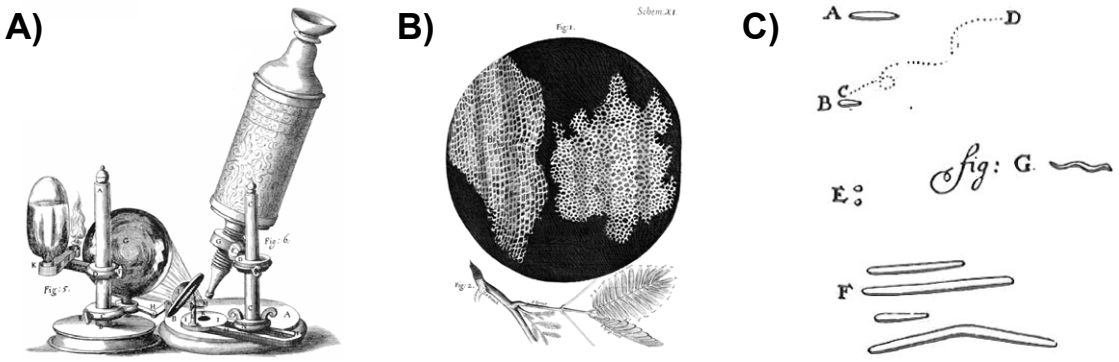


Figure 1.2: Microscopic inventions and discoveries in the seventeenth century. A) shows a drawing of an early microscope constructed by Robert Hooke. B) is a drawing of cells in cork, first observed by Hooke. C) is a drawing of microorganisms, first observed by Leeuwenhoek. All images are in the public domain and were taken from [7, 8]. The figure also appears in this form in [9].

van Leeuwenhoek, who first observed microorganisms ("animalcules") in the 1670s, could be considered the first forefather of microbiology (figure 1.2) [6].

While the usefulness of microscopes was widely accepted and saw continuous improvement, both in technical aspects, but also in sample fixation and staining methods, it remained a rather informal "art of lens making" for quite some time.

The 19$^{\text{th}}$ century saw the elevation of microscope building to a formal, theory-backed process, with descriptions of the physical limits of light microscopy by Ernst Abbe in 1873, followed by Helmholtz and Lord Rayleigh. Microscope instrumentation of the time already approached those theoretical limits. Based on increasing knowledge of the physics behind optics, new contrast methods like phase contrast or differential interference contrast (DIC) were developed in the 20th century. An avenue towards imaging structures at the molecular scale came with the development of X-ray crystallography of biomolecules and EM, far surpassing the limits of light microscopes. However, light microscopy has kept its importance due to developments in specific labelling of molecules by fluorescent markers, for ex-

ample recombinant fluorescent proteins [10] that allow for specific labelling in live samples, and indicators for various biological properties. Innovations into digital recording of images have also turned images into quantitative measurements and allow for computer-based image processing and analysis, which continues to fuel developments and novel insights in- and outside of biological research.

1.0.2 A microscope is only as good as its sample

The first step in microscopy-based experiments is to generate an as-accurate-as-possible image of the system under study. Taking an accurate image might actually mean many things: an obvious task is to record images at the highest possible spatial resolution and with high contrast, i.e. high signal-to-noise ratio (SNR), but to make sense of the myriads of molecules in a biological system, a further criterion for a "good image" is also specificity - it is mainly in this task that light microscopy excels. While EM enables molecular (or even atomic) resolution, the imaging of specific molecules within a highly complex biological environment remains a challenge for this technique.

Developments in light microscopy instrumentation were accompanied by improvements in staining and labelling techniques from the very beginning, as early microscopes operated in bright-field mode in which contrast is formed by absorption of light in the sample. To increase contrast in systems such as translucent eukaryotic cells or microbes and specifically label various structures, a multitude of staining protocols were developed. Famous examples include the Gram stain which allowed classification of bacteria into Gram-positive and Gram-negative species with or without a thick cell wall that retains the dye Crystal Violet, respectively. Another example that still sees extensive use in histology today is the Hematoxylin & Eosin (H&E) stain. Most staining protocols are not compatible with living cells, however, and require fixation prior to labelling. Thus, physical ways of achieving contrast in a microscopy image were developed that do not interfere with the

biological system, such as darkfield microscopy, that achieves contrast by recording only light that is scattered in the sample, polarization microscopy and phase contrast or DIC microscopy, which generate contrast from the refractive index differences within the sample.

In recent decades, this specificity is most often achieved through fluorescence microscopy, in which only light from specifically labelled structures is collected, and a variety of target-specific labeling strategies, like antibody labeling, fluorescent fusion proteins or specific molecular interactions (e.g. the DNA-binding fluorophore 4',6-diamidino-2-phenylindole (DAPI), or fluorescent labels conjugated to the actin-binding fungal toxin phalloidin). One can also label DNA oligomers and use them to selectively attach a label to their complementary sequence in a cellular system. This basic idea of fluorescence in-situ hybridization (FISH) provides a way towards microscopic genotyping and allowed for detailed study of the spatial organization of nuclei [11], but can also be extended to other applications such as the labeling and quantification of mRNAs [12, 13] and as a building block for sequential multiplexed labeling (see 1.2.1).

The discovery of fluorescent proteins [14] and their introduction as fusion proteins into new biological systems via molecular biology complements traditional fluorescent labeling with its live-cell compatibility and high specificity [15]. The importance of fluorescent proteins as a tool for specific imaging in live biological systems was honored by the awarding of the 2008 Nobel Prize in chemistry [16] to Osamo Shimomura, Martin Chalfie and Roger Y. Tsien, respectively for the discovery, adoption in molecular biology and characterization and modification of the green fluorescent protein (GFP), the prototypical fluorescent protein. Even though they added a groundbreaking new tool to the toolbox of molecular biology, fluorescent proteins are typically less bright and photostable than synthetic fluorophores, so their adaptation to tasks typically reserved for conventional flu-

orophores, such as, for example, specific labeling of single genomic loci remains challenging [17, 18]. As an alternative, proteins such as the HALO or SNAP tags [19, 20], which bind free dye molecules added to the medium, can be employed for bright staining of specifically tagged proteins in living cells.

While fluorescent labelling allows for high specificity, the number of fluorophores that can be imaged simultaneously is limited to about a handful by their spectral overlap. If one wants to specifically image many molecular structures, a possible solution is to physically remove labels from the sample and replace them by new ones, thus allowing the sequential imaging of many targets [21].

1.0.3 The current status and future of microscopy

Microscopy has been an invaluable tool in biology over the past few centuries from the scientific revolution onwards. Yet, until recently, microscopes have remained high-performance visual aids and were rarely seen as data collection instruments. Biochemical methods can nowadays often provide detailed molecular insight into phenomena that were previously reserved for microscopic investigation. Researchers naturally treat the highly specific results from such experiments as ”just data”, whereas micrographs are often taken as snapshots of the physical world and considered a result as-is.

Modern Omics has pushed into the singe-cell domain, but subcellular details often remain a task for light or (cryo-)EM, or just light microscopy in the case of live dynamics of cells. Thus, microscopy is set to continue to be an indispensable tool for biology. To reconcile the high-level overview that conventional micrographs provide with the intricate molecular processes studied in modern biology, microscopy will need to evolve in parallel to the questions studied. The improvements can come from optics and microscope instrumentation, but also from sample preparation and specific labelling and, most importantly, from computational tools that

both increase the amount of information that can be extracted from a sample and, on the other hand, help to condense that high-dimensional information into quantitative results understandable by humans. The coming years will hopefully see an integration of multi-modality data from both microscopy and biochemical sources to reach a heightened understanding of diverse biological processes and structures [22].

1.1 Modern microscopy

1.1.1 Fluorescence microscopy basics

As mentioned above, the majority of modern microscopy techniques rely on fluorescent labels to specifically image particular structures of interest. Furthermore, fluorescent labels are essential for many advanced techniques such as super-resolution microscopy.

Fluorescence is a process in which a molecule, called a *fluorophore*, is excited by absorbing a photon of a specific wavelength (i.e. energy). The molecule transitions to an excited singlet state, in which some energy is typically lost as heat to the environment. Shortly thereafter, the molecule returns to the ground state by emitting a photon again (figure 1.3). Due to energy loss as heat, the emitted photon will typically be of longer wavelength than the exciting photon (as phenomenon called Stokes shift). All the possible energy differences between the ground state and excited states determine at which wavelengths it can absorb photons (the absorption spectrum) and likewise, the emission spectrum is determined by the possible energy differences from excited states to the ground states.

Apart from the possibility of very specific labelling, a further advantage of fluorescence microscopy is the relatively easy construction of a microscope optimized for fluorescence (figure 1.4). Light is emitted from a source such as a laser or light emitting diode (LED), reflected and focused onto the sample along the optical axis (z) by the objective, where fluorescence emission is excited. The emission light, which has a longer wavelength, is collected again by the objective and is transmitted through the main dichroic and a tube lens onto a detection device such as a digital camera or eyepiece. Unlike transmitted light microscopy, fluorescence microscope can use the same high-quality objective for both excitation and detection.

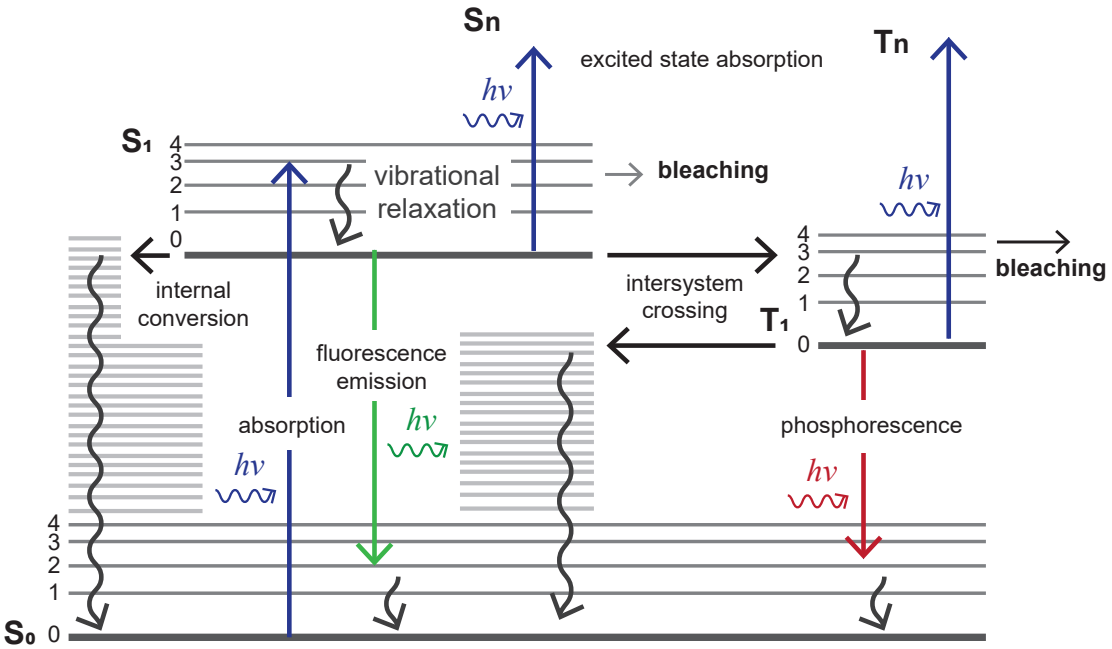


Figure 1.3: Jablonski energy level diagram of some of the possible transitions involved in fluorescence. A molecule in the lowest electric singlet state S_0 absorbs a photon and is pushed into the first excited singlet state S_1 (or typically a vibrational energy sub-state). After quick vibrational relaxation (giving off energy as heat into the environment), the molecule can fall back into the ground state through emission of a photon. From S_1, the molecule can also move to an excited triplet state T_1 via intersystem crossing. From this (usually long-lived) state, the molecule can move back to S_0 via phosphorescence emission, but in excited states, it may also react with molecules in its environment, resulting in permanent bleaching of the fluorophore. Non-radiative relaxation from excited states to vibrational sub-states of lower states of same multiplicity may also happen via non-radiative internal conversion.

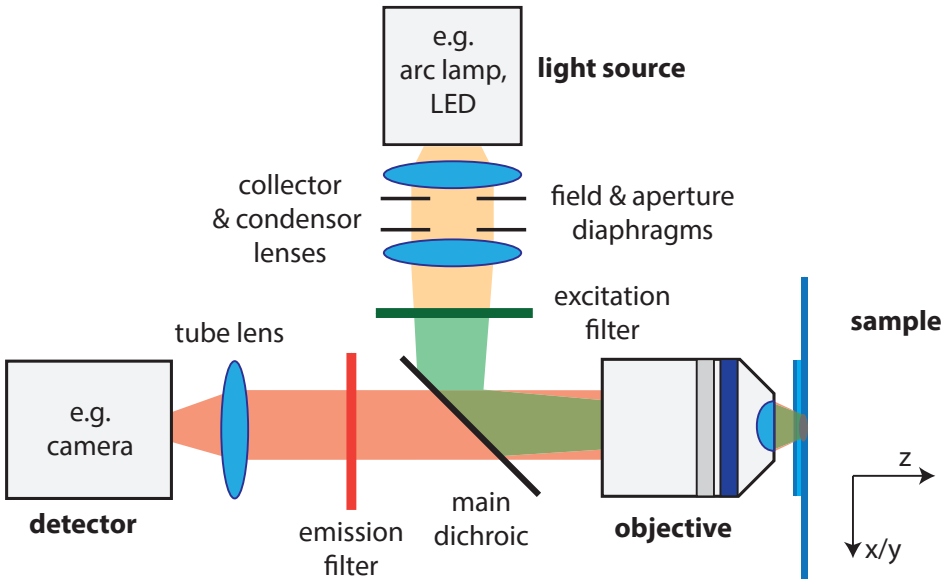


Figure 1.4: Simplified schematic of a basic fluorescence microscope, consisting of a single objective lens, a light source and lenses and apertures for Köhler illumination, a detection system (e.g. a camera and tube lens) as well as a main dichroic mirror and excitation and emission filters to separate the illumination and detection path.

The fluorescence properties of a fluorophore depend strongly on its molecular environment. An excited fluorophore might transfer its energy to other molecules which dissipate it in a non-radiative way. This effect is typically unwanted and called quenching. On the other hand, the molecular environment of a fluorophore can be adjusted to increase photostability through anti-bleaching agents.

The dependence of fluorescence on the environment also gives an opportunity for the establishment of fluorescent indicators of environmental factors such as calcium concentration, pH or the presence of another fluorophore in close proximity to which excitation is transferred via Förster resonance energy transfer (FRET) [15].

1.1.2 Super-resolution microscopy

A first, obvious desired property of a "good image" of a sample is high spatial resolution, i.e. the ability to distinguish detailed structures. However, due to physical properties, the resolution of an optical instrument is fundamentally limited to roughly half the wavelength of light used. This posed significant challenges to the study of molecular systems with typical sizes on the order of a few or a few tens of nanometers with light microscopy using visible light at wavelengths of a few hundred nanometers. In the last decades, several techniques have been developed to circumvent this limit that can be roughly classified into two groups: Extended resolution techniques, which extend the resolution of a microscope by a defined factor include, among others, structured illumination, image scanning microscopy and (practically infeasible due to low signal) confocal microscopy with an infinitely small pinhole. These extensions arise from purely physical considerations and typically do not place constraints on the labelling of the sample. On the other hand, some techniques, such as STED and localization microscopy, allow for fundamentally unlimited resolution, but require the use of specific fluorophores.

As all methods with fundamentally unlimited resolution are the result of an interplay of optics and the chemical labels in the sample, it is not easy to classify super-resolution microscopy solely into the domain of physics or chemistry. In this section, the physical basics of the diffraction limit, as well as three hallmark techniques for super-resolution that were employed in the work presented in this thesis: SIM, STED and SMLM, which each encompass unique microscope instrumentation and data analysis, will be discussed. Methods that rely more heavily on properties or labelling will be described in the following section on sample preparation. PAINT, while providing an avenue towards localization-based super-resolution by itself, can also be seen as a flexible framework of labelling to extract various molecular properties of the sample under study, and will therefore be discussed below. Likewise, expansion microscopy relies much more on the modification of the sample than on the actual microscope used to image it and will also be described below.

The diffraction limit

A first formal study into the fundamental limits of optical systems was done by Ernst Abbe, working for the Carl Zeiss microscope company that sought to place its manufacturing onto solid theoretical foundations, in 1873 [23]. In his studies, Abbe considered passing light through a simple line grid and collecting it with an objective. He could prove that only if the first diffraction maximum of the grid could be passed through the aperture of the optical system, an image of the grid can be generated. As shown by Abbe, the minimal grid spacing d (and consequently, since any object can be seen as a sum of spatial frequencies, the minimal distance between arbitrary structures), that can be resolved with a given objective is:

$$d = \frac{\lambda}{2NA} = \frac{\lambda}{2\sin(\alpha)n} \tag{1.1}$$

Here, λ is the wavelength of light, n is the refractive index of the medium and α is the maximum half-angle at which light is collected by the microscope. This assumes that a condensor with the same properties as the objective (or one objective for both illumination and detection) is used to illuminate the grid from all sides. Illuminating from just one direction, without a condensor, would result in a twice as large minimal distance. Since the resolving power of an objective is thus solely determined by the maximum angle of light it can capture in its designated medium, the denominator $NA = n\sin(\alpha)$, its *numerical aperture*, became one of the main indicators describing the optical quality of an objective and is extensively used to this day.

Later studies into the resolution limits of microscopy were done by Helmholtz and Lord Rayleigh [24], who studied the problem from the perspective of imaging single, light emitting points through a microscope. Due to diffraction, an infinitesimally small point emitter will not be imaged as a point, but as a smeared-out disk surrounded by concentric rings, the so-called point spread function (PSF). Rayleigh derived an analytic expression for the PSF intensity h at a given distance r from the center (in the focal plane) of a microscope with an objective, of given NA:

$$h(r) = I_0 \left(\frac{2J_1\left(\frac{2\pi r NA}{\lambda}\right)}{\frac{2\pi r NA}{\lambda}} \right)^2 \tag{1.2}$$

Here J_1 is the Bessel function of the first kind of order 1 and I_0 is the PSF intensity at the maximum. When imaging more complex samples, the resulting image I is the result of a convolution of the underlying structure s (i.e. the distribution of light-emitting molecules) with the PSF h:

$$I(x) = (s * h)(x) = \int_{-\infty}^{\infty} s(\tau)h(x-\tau)d\tau \tag{1.3}$$

Employing this view, a slightly more conservative limit for the minimal resolvable distance, the Rayleigh criterion, can be constructed, which states that two point-emitters can still be separated if the PSF intensity maximum of one coincides with the first intensity minimum of the other, and vice-versa (figure 1.5). This is the case at a distance $d_{Rayleigh} = \frac{0.61\lambda}{NA}$.

Another way of describing the resolution limit is to think about images in terms of their Fourier transforms (often abbreviated by the algorithm used to compute them, the Fast Fourier Transform, FFT): any signal, including images, can be seen as a sum of simple sine waves; the (complex-valued) Fourier transform determines for every possible frequency (and orientation in the >1D case) at which phase and amplitude the corresponding wave contributes to the whole signal. In the conventional, spatial view, imaging a structure through a diffraction-limited microscope corresponds to convolving the structure s with a PSF h. Due to the Fourier convolution theorem $\mathcal{F}(s * h) = \mathcal{F}(s)\mathcal{F}(h)$, we have an analogous view of this process in frequency space: The FFT of the underlying structure is multiplied by the object transfer function (OTF), the FFT of the PSF (figure 1.6). In a diffraction limited system, the OTF is nonzero only up to a finite radius around the origin (the frequency corresponding to the smallest resolvable distances). This Fourier-optics view is not just a mathematical abstraction, as the light in the back focal plane of an objective actually corresponds to the FFT of the structure in the image plane. Therefore, placing a small aperture in the BFP (or conjugated plane) results in a narrower OTF and thus reduced resolution, for example.

Structured Illumination

Various techniques exist that can expand upon the diffraction limit by a defined factor. For example, in confocal laser scanning microscopy (CLSM), lateral resolution can in theory be improved by a factor of $\sqrt{2}$, as the effective PSF is a product of the excitation and detections PSFs and thus has a smaller full width at half

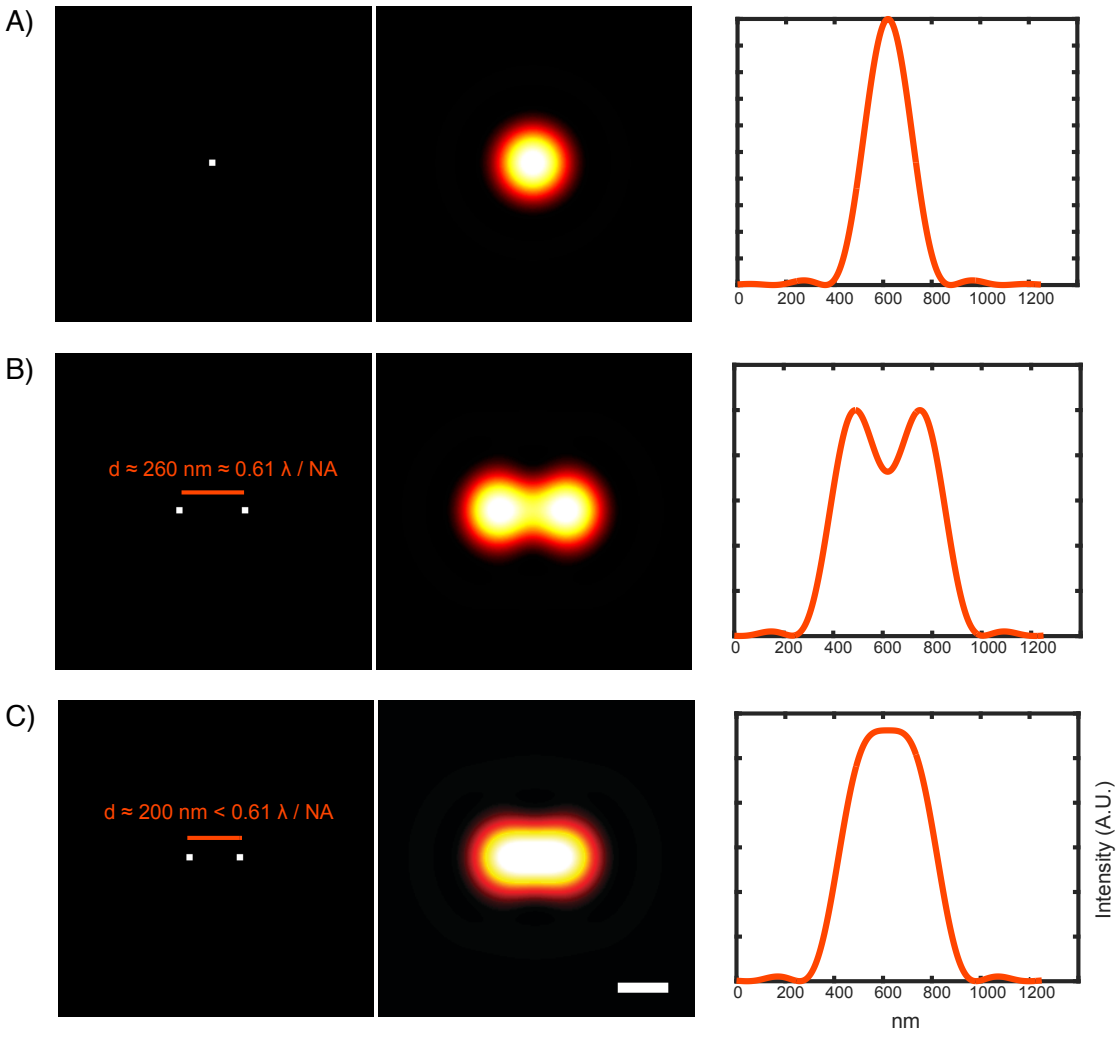


Figure 1.5: Simulated images of point-like light sources (left) and their convolution with the PSF of an objective with $NA = 1.4$ and $\lambda = 594nm$ (middle) as well as graphs of horizontal cuts through the images at the center of the simulated spots (right). PSFs were simulated according to equation 1.2. A) shows a single point-like light source, B) two point-like light sources resolvable according to the Rayleigh criterion and C) two point-like light sources closer than the diffraction limit. The scalebar equals 200 nm. Figure taken from [9].

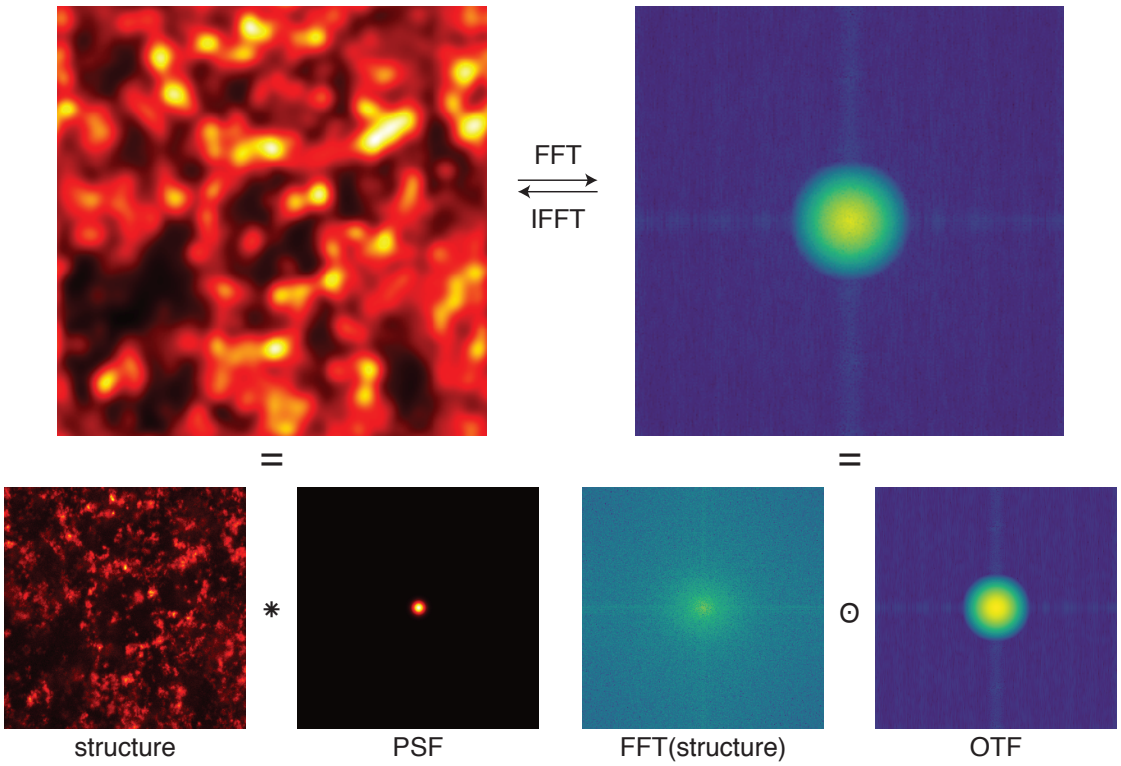


Figure 1.6: An image and its Fourier transform are equivalent views of the same object in ordinary or frequency space, related by forward or inverse Fourier transforms. Bottom left: In ordinary space, an image acquired by a microscope can be seen as the convolution of an underlying structure (for this illustration, a high-resolution STED image of EdU-labelled DNA is used) with the microscopes PSF (an idealized Gaussian in this case). Bottom right: In frequency space, the imaging process can be described as the elementwise product of the FFT of the underlying structure with the OTF.

maximum (FWHM). Likewise, in 2 photon microscopy (2PM), the 2P excitation depends on the square of the illumination light intensity, again yielding a $\sqrt{2}$ resolution improvement. In practice, however, these improvements have little effect, as resolution improvement in CLSM would require infinitely small pinholes and in 2PM, the resolution gain through the square illumination dependence is more than offset by the 2 times longer wavelengths used.

A technique that has in practice allowed a resolution increase by a factor up to two in every dimension is structured illumination (SIM). By taking multiple images with varying, non-uniform illumination patterns, high-resolution information that is *mixed into* the images can be computationally reconstructed. A commonly observed phenomenon based on the same principles that is often used to illustrate SIM is the Moiré effect: when overlaying two grid-like structures, intensity fringes of longer period than the original grids can be observed.

The process of generating images with increased resolution via SIM can be best understood in frequency space: If the underlying structure is not illuminated uniformly (i.e. it is multiplied point-wise by the illumination pattern), the image formation process takes the form of a convolution of the FFT of the structure with the FFT of the pattern, followed by multiplication by the OTF. It is the convolution with the pattern that can lead to high-frequency information (outside the support of the OTF) to be mixed into low-frequency parts that can pass the OTF. If multiple images of the same structure with different illumination patterns (e.g. grids with varying phase) are acquired, information that was moved into the central part of the FFT via convolution can be extracted and computationally moved back to its original position (figure 1.7).

Various implementations of the SIM principle exist, allowing xy-resolution doubling [25, 26], improved optical sectioning [27] or even resolution doubling in all dimensions by using a 3D illumination pattern [28]. Due to its capability to image

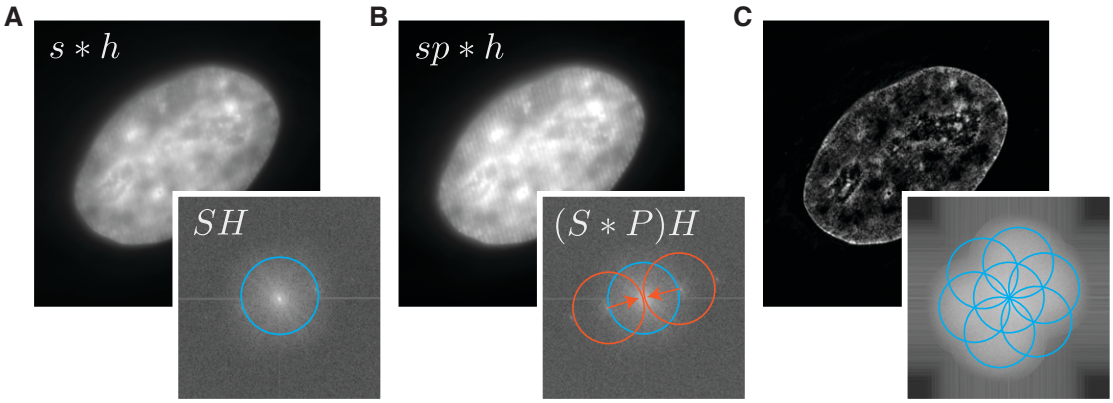


Figure 1.7: Principles of SIM, illustrated using images of a DAPI-stained HeLa nucleus imaged on a Deltavision OMX 3D-SIM microscope, courtesy of Andreas Maiser. (A) The diffraction limit corresponds to only the central, low-frequency parts of the FFT of the structure ($\mathcal{F}(s) = S$) being passed through an optical system, due to multiplication with the OTF ($\mathcal{F}(h) = H$) with limited support. (B) When non-uniform illumination ($p, \mathcal{F}(p) = P$) is used, higher-frequency components of the structure are mixed into the observable region via convolution. By taking several images with modified illumination patterns, e.g. phase-shifted grids, the individual components can be recovered and assigned to their original position in frequency space. (C) By acquiring images at several phases and grid orientations, a larger coverage of frequency space and therefore a super-resolved image can be reconstructed (reconstruction performed with the manufacturer's software, SoftWorX, slight asymmetries of the FFT due to misalignment of the instrument).

whole 3D volumes at improved resolution without requiring special fluorophores or excessive illumination intensities, SIM has become a workhorse technique for the study of subcellular and subnuclear structures [29, 30]. For conventional SIM, the resolution improvement is limited to a factor of two, as the illumination pattern itself is subject to the diffraction limit. By using saturable optical effects, like in STED (see below), an unlimited resolution improvement with saturated SIM (SSIM) can be implemented [31, 32].

STED

The theoretical description of Stimulated Emission Depletion (STED) microscopy by Hell and Wichmann in 1994 [33] provided a first avenue towards far-field microscopy without a fundamental resolution limit. The first practical implementations of the concept came around the turn of the millennium [34] and STED has since proven to be a reliable general-purpose super-resolution method in many studies. For his contributions to optical super-resolution microscopy by the development of STED, Stefan Hell was awarded the Nobel prize in chemistry in 2014 [35].

STED takes advantage of the phenomenon of stimulated emission: whereas in normal fluorescence, a molecule absorbs a photon and is pushed to an excited state and, after vibrational relaxation, eventually (typically after a few ns) falls back to the ground state by emitting a lower energy photon, stimulated emission happens when an excited molecule absorbs a second photon (with energy corresponding to a possible emission energy difference), which causes it to immediately fall back to the ground state, emitting two copies of the depleting photon in the process (figure 1.8). Stimulated emission is the process by which lasers (LASER: Light Amplification by Stimulated Emission of Radiation) generate highly coherent light.

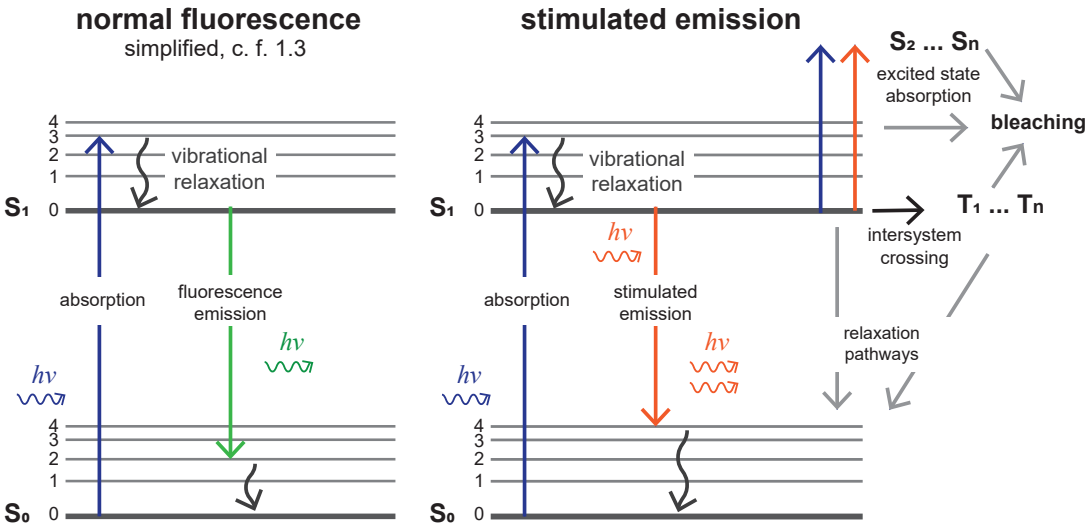


Figure 1.8: Energy transitions of a fluorophore during normal fluorescence and stimulated emission. Left: In normal fluorescence, the molecule is pushed into an excited state by absorption of a photon and will drop back into the ground state after vibrational relaxation, emitting a (typically longer wavelength) photon In the process. Right: In stimulated emission, an excited fluorophore absorbs a second photon with energy corresponding to an emission. This will cause the molecule to immediately drop back into the ground state, emitting two photons with exactly the same properties of the depleting photon. The depletion light can however also cause excitation into reactive higher states, in which photobleaching can occur due to interactions of the fluorophore with its environment [36].

When fluorophores are illuminated with both excitation and depletion light, the light they emit is predominantly in the form of stimulated emission, normal fluorescence emission is *depleted*. To make this effect usable for super-resolution, the depletion laser spot is shaped so that it will cause more depletion in some parts of the excited volume than others. For example, by using a ring (often referred to as a *doughnut* in STED literature) of depletion light surrounding an excitation light spot, only molecules in the very center exhibit normal fluorescence, whereas molecules on the side emit through stimulated emission. By scanning the combined excitation and depletion light across the sample (like in a conventional confocal laser scanning microscope) and only collecting normal fluorescence by filtering out the specific wavelength of stimulated emission, a super-resolution image can be acquired point-by-point (figure 1.9).

Since the depletion pattern is itself diffraction limited, it does not immediately provide unlimited resolution. To fully break the barrier, STED makes use of the fact that fluorescence depletion by stimulated emission is a non-linear, saturable process (the probability of depletion cannot go over 1, even if the depletion light intensity is further increased). By using more and more intense depletion light, the area in which depletion is not saturated (and in which normal emission can still occur) can be made arbitrarily small (figure 1.10). The resolution achievable by STED is thus dependent on the maximum depletion light intensity (I_{max}) and the fluorophore-specific intensity required for saturation (I_{sat}) which is reflected in the extended resolution formula (with the parameter α describing the steepness of the depletion light intensity gradient around the zero):

$$d \approx \frac{\lambda}{2NA\sqrt{1 + \frac{\alpha I_{max}}{I_{sat}}}} \tag{1.4}$$

The extremely high intensity depletion light is a major disadvantage of STED, since it places strong requirements on the photostability of the dyes used and

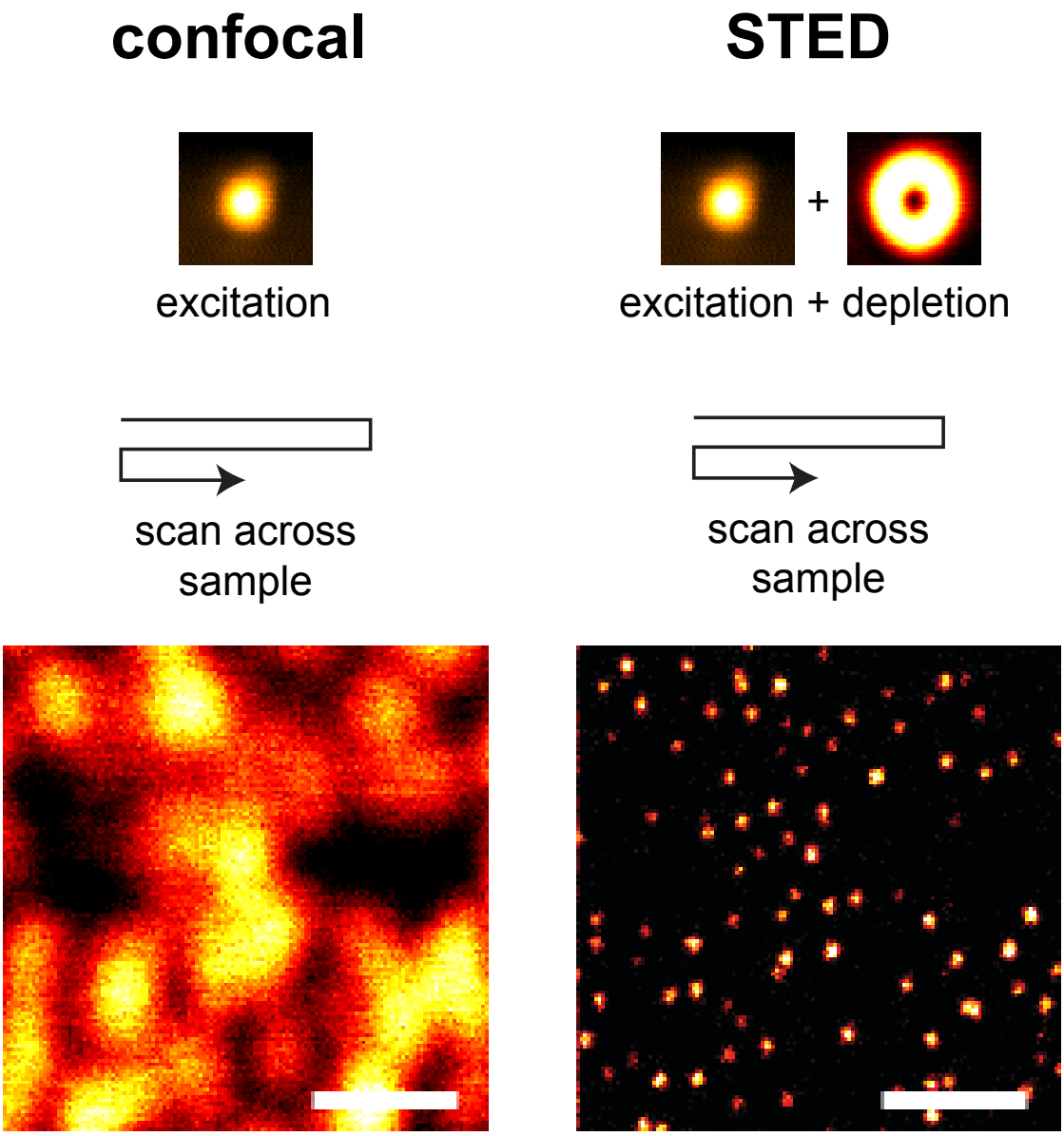


Figure 1.9: Comparison of image acquisition in a confocal laser scanning microscope and a STED microscope as well as sample images of 40 nm Crimson fluorescent beads. The scalebar equals 500 nm. Figure taken from [9].

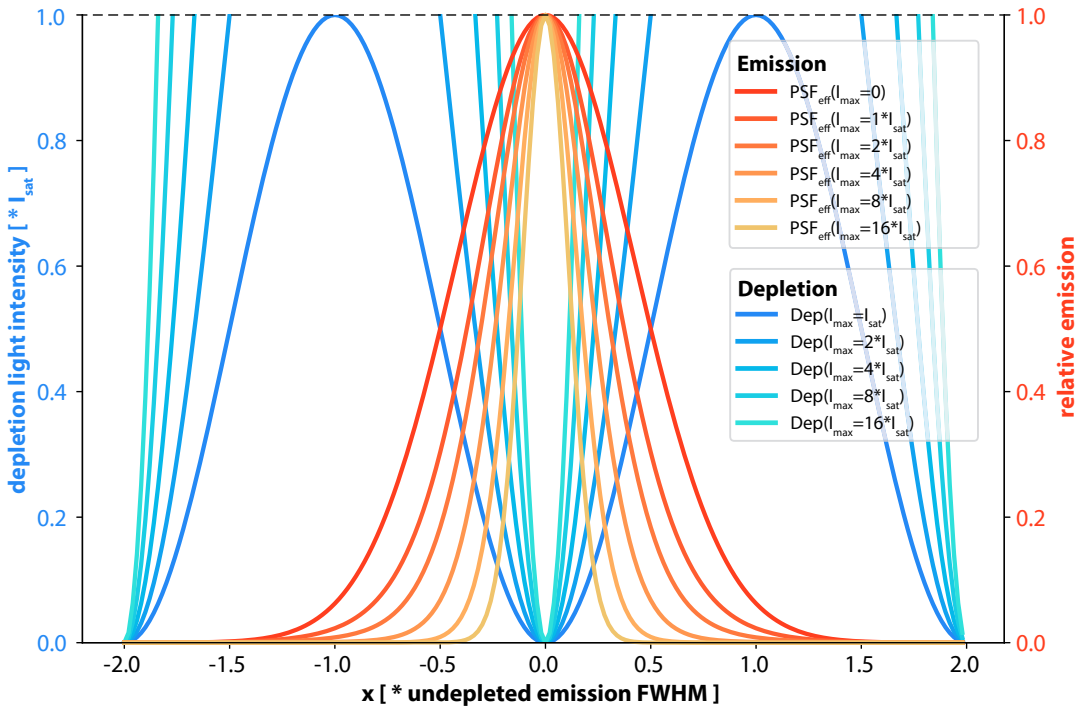


Figure 1.10: Effective STED PSF (red-yellow) as a function of the peak depletion intensity I_{max}, approximated according to [37], given corresponding sinusoidal depletion patterns (blue). Depletion pattern intensities greater than I_{sat} (saturated areas) are not shown. I_{sat} corresponds to the intensity at which $1 - e^{-1} \approx 63\%$ of fluorescence emission is depleted via stimulated emission.

places additional constraints on experiments. The problem can be alleviated by specialized acquisition schemes: for example, in the MINFIELD approach [38], the recorded image is smaller than the diameter of the depletion doughnut, and the fluorophores are never exposed to the maximum of the doughnut. Another possibility is to dynamically adjust dwell times at each location depending on whether fluorophores are present or not, preventing premature bleaching of fluorophores at succeeding positions [39–41].

Localization Microscopy

The fact that any point emitter imaged with a microscope will be smeared out into a blob by the instruments PSF limits the distance between two emitters that can be resolved. It does not limit, however, the precision with which a single emitter can be localized. One can think of the PSF as a probability distribution on the location of detected photons, given an emitter location. From basic statistic considerations, it follows that the mean location of the photons, i.e. the location of the emitter, can be localized with a standard error proportional to the inverse square root of the number of photons detected N (figure 1.11 A). Taking the discretized pixel nature of the acquired images and background noise into account, a more realistic estimate of localization precision is given in [42] as $\sigma_i = \sqrt{\frac{s_i^2}{N} + \frac{a^2/12}{N} + \frac{8\pi s_i^4 b^2}{a^2 N^2}}$, with s_i denoting the standard deviation of the PSF along axis i, a denoting the pixel size and b denoting the standard deviation of the background noise.

Initially, single molecule localization in micrographs was employed in particle tracking, e.g. [43]. A single fluorescent molecule imaged repeatedly over time will appear as a blob, but its center can be tracked with nanometer-scale accuracy. Particle tracking has been intensively employed in the study of the dynamics of motor proteins, for example. Furthermore, distances between two particles with distinct labels (e.g. different fluorophores) can be determined with similar high accuracy [44].

To actually generate an image and not just measure distances, each fluorophore in the sample would have to have a distinct label, e.g. color [45]. Since even on a state-of-the-art microscope it is tricky to distinguish more than a handful of different colors due to overlaps in the spectra of the fluorophores, the separation in single molecule localization microscopy (SMLM) comes through the alternative route of *separation in time*. By making sure that only a few labels are in an emitting state at a single timepoint, a small (non-overlapping) subset of them can be localized (figure 1.11 B). If there exists a way of then switching on a different subset of the labels, the process can be repeated for many frames, building a super-resolved image point-by-point.

The labels of choice for the real-world application of this principle are photoswitchable dyes [46], molecules that change their fluorescence properties upon illumination with light of a specific wavelength, e.g. photoactivatable dyes that are irreversibly switched from a dark to a bright state. By only activating a few of the fluorophores with a weak flash of activation light and then imaging them until they bleach and thus turn dark again and then repeating the process, the principle of SMLM can be implemented in practice.

First practical implementations of SMLM were independently described in 2006 by several research groups under the names photoactivated localization microscopy (PALM) [47], fluorescence photoactivation localization microscopy (fPALM) [48] or stochastic optical reconstruction microscopy (STORM) [49]. For work on the fundamental properties of photoswitchable molecules and a first application of localization of single photoswitchable molecules for super-resolution microscopy, William Moerner and Eric Betzig, respectively, were awarded the Nobel prize in Chemistry in 2014 [35].

Typically, SMLM is performed on time series of single xy-planes, often employing total internal reflection fluorescence (TIRF) illumination, in which the illumination

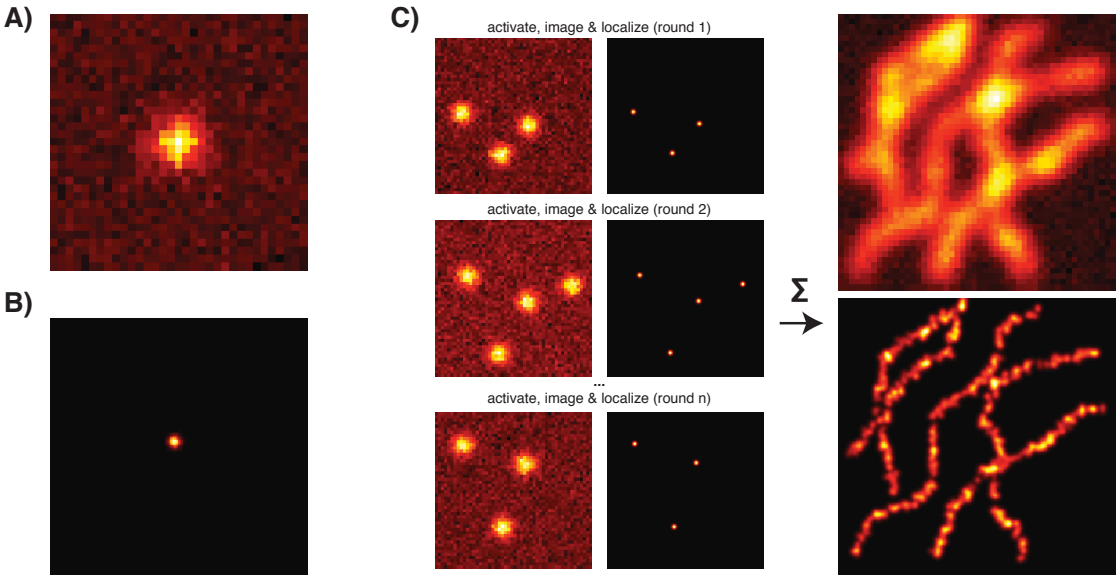


Figure 1.11: Principles of SMLM on simulated data A: A point emitter imaged through any microscope will be blurred by the instrument's PSF and the image will suffer from various types of noise. B: The center of the recorded PSF can be localized with high precision, however (distribution of centers of mass of 5000 simulated point emitters). C: When imaging a densely labelled structure with SMLM, subsets of the fluorophores are activated, imaged and localized in an iterative process (left). The resulting localizations can be combined to form an image with much higher resolution than the sum of the diffraction limited raw images (right).

light is reflected on the coverslip-sample border, and only a thin slice (of less than 200 nanometers thickness) above the coverslip is excited by an evanescent wave [50]. Since biological structures are inherently three-dimensional, methods to encode z information into the 2D images have been developed. The simplest way of extracting z information from the single planes is to deliberately introduce aberrations into the detection beam path that result in anisotropy of the PSF. For example, by inserting a cylindrical lens into the detection path, the recorded PSFs will be stretched in x if the emitter lies above the focal plane and in y if it lies below the focal plane, or vice-versa [51]. By using a pre-recorded calibration curve and calculating an x/y ratio by fitting a PSF function to the emitters, the z component can be recovered. There exist many different approaches of encoding z information via PSF engineering [52, 53], with various trade-offs between xy-precision, z precision and the depth-of-field in which three-dimensional localization can be performed. An alternative approach to extracting z information is to split the emission light into two paths with a slight path length difference and record it on two cameras (or two separate areas on one camera), so that the relative z location is encoded in the ratio of PSF sizes of one emitter in the two images [54].

Other super-resolution techniques

While SIM, STED and SMLM are the best known super-resolution microscopy techniques, there exists a multitude of other techniques that also promise extended or fully unlimited resolution, like, for example, image scanning microscopy (ISM, commercialized by Zeiss under the name AiryScan), in which point scanning is combined with array detection to achieve the resolution improvements of a confocal microscope with an infinitely small pinhole while still collecting enough light to preserve workable SNR [55]. A further family of techniques, consisting of SOFI [56] and SRRF [57], takes a time series of sequential images of the same structure

and reconstructs a super-resoled image from temporal fluctuations in the emissions of single fluorophores, which can be seen as a bulk version of SMLM.

A final development worth mentioning is MINFLUX [58], developed in the lab of STED inventor Stefan Hell, in which single molecule imaging is ingeniously combined with repeated point measurements with a doughnut-shaped excitation beam. In MINFLUX, accurate localization places the emitter in the zero of the doughnut and corresponds to a minimal number of fluorescence cycles by the fluorophore. Thus, localization precision no longer requires a maximum number of photons to be collected and the technique, though still awaiting widespread adaptation, promises single-nanometer resolution in biological samples [59].

1.1.3 Light-sheet microscopy

The developments of super-resolution fluorescence microcopy over the last two decades allowed for tremendous advances in imaging tiny subcellular structures. Another complementary but equally challenging task for microscopy is the capture of large, multicellular, three-dimensional objects such as tissues or organs and live organisms in their entirety. Many approaches have been developed to meet this challenge, for example confocal and two-photon microscopy [60], and their development often went hand in hand with computational advances, such as deconvolution [61].

While these techniques achieve *optical sectioning*, allowing imaging of large, fixed samples, they come with downsides that often limit the imaging of live organisms. Because illumination and detection are done by one objective, along a common axis, the whole sample is illuminated for every plane that is detected, which can lead to photobleaching or damage the (living) sample. Furthermore, many of these techniques work by scanning the sample point-by-point, which limits the acquisition speed and makes it hard to cope with moving samples.

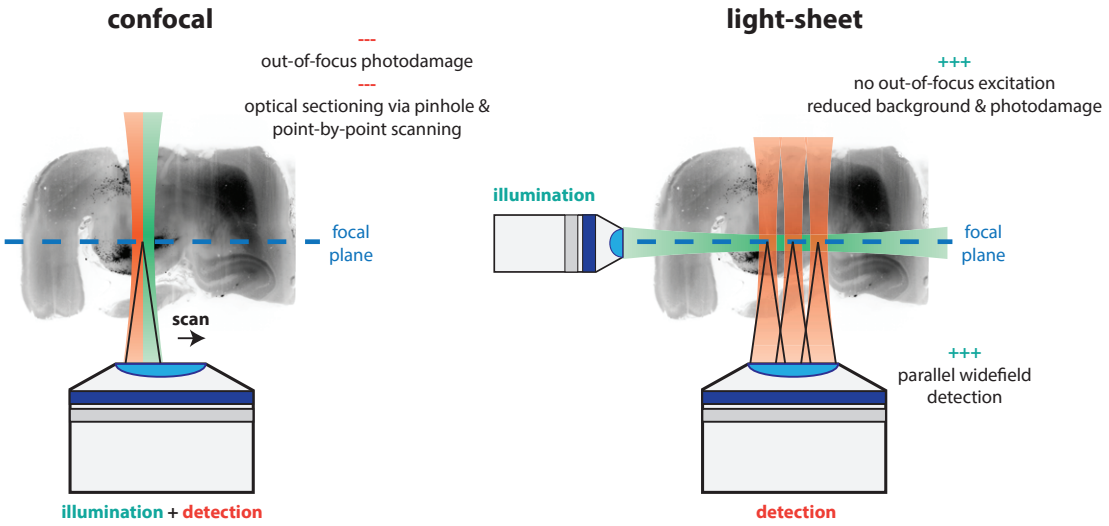


Figure 1.12: Principles of light-sheet fluorescence microscopy in comparison to conventional confocal microscopy. In conventional microscopes (left) illumination and detection happen through one objective lens along the same optical axis. This results in the illumination, and consequently in photodamage or background signal, behind and in front of the focal plane. In confocal microscopes, out-of-focus signal can be rejected via a pinhole, which however requires slow, point-by-point scanning. In LSFM (right) the sample is illuminated through a separate objective, oriented perpendicular to the main detection axis. In this way, only the focal plane can be illuminated, resulting in both reduced photodamage and reduced out-of-focus background (i.e. optical sectioning). LSFM does not require point-by-point scanning - an image of the entire focal plane can be captured with a widefield camera. Figure inspired by [62].

Light sheet fluorescence microscopy (LSFM) circumvents most of these problems in an elegant way: by illuminating the sample from the side using a second illumination objective at 90 degrees to the detection objective (figure 1.12). Through optical elements or by line scanning, the illumination can be focused into a thin sheet along the z-dimension, leading to illumination of only a thin plane in the sample that is aligned to the focal plane of the detection optics. That way, fluorophores outside the focal plane are not excited, limiting photodamage and phototoxicity outside the area of interest. Since the whole focal plane is illuminated at once and an image can quickly be recorded using a high-speed widefield camera, this also avoids the high peak intensities of a point-scanning confocal microscope [62–64].

Illumination through a perpendicular beam path is not a new development: In fact, it saw a famous early utilization under the name of *ultramicroscopy* by Zsigmondy and Siedentopf in their Nobel Prize-winning studies on colloid particles in glass in the early 20^{th} century [65]. It took until the end of the 20^{th} century for the principle of orthogonal illumination to see its earliest applications in biological imaging [66], and its potential finally reached widespread recognition under the name of selective plane illumination microscopy (SPIM) by Huisken and colleagues in 2004 [67].

Modern LSFM

A large variety of designs and optical configurations for light-sheet microscopes were described in the years following the initial SPIM paper, for example with upright [68], inverted [69] or horizontal beam path geometries, but also with illumination from multiple parallel directions [70, 71] and the capability to rotate the sample within the imaging chamber, e.g. [72]. Aside from just the layout of the optical axes, a major differentiating factor between the various designs is how the light-sheet is generated in the focal plane. Assuming that illumination happens

along the x axis in image space, the goal is to focus the illumination light in just the z direction, while keeping widefield illumination in the y direction. Similar to the trade-off between lateral resolution and depth-of-field in detection optics, the light-sheet generation can be optimized for a light-sheet that is thin center of the x axis but broadens quickly or a thick sheet that remains at relatively uniform thickness for a long distance. The simplest way of creating a light-sheet is to introduce a single cylindrical lens into the illumination path, which results in a thick, but relatively uniform sheet, ideal for imaging large (millimeter- to centimeter-sized) samples such as entire organs [73]. Alternatively, in the initial SPIM design, the illumination light is focused along one axis by a cylindrical lens but then passed through a higher-NA objective lens to form an orthogonal sheet with a thinner central section. Light-sheets can also be generated *virtually* by using just an objective to focus light along the illumination axis and then using a scanner to rapidly move the illumination beam in the y direction, resulting in a time-averaged light-sheet in the xy-plane [74].

To counteract the thickness vs. lateral length tradeoff, various approaches have been described. One example is the use of non-diffracting beams formed through interference, such as Bessel [75, 76] and Airy beams [77] or optical lattices [78]. Alternatively, one can introduce a quick focusing element, e.g. an electrically tunable lens (ETL) into the illumination path and quickly *scan* the center of the light-sheet along the x-axis, resulting in a more uniform thickness in the time-average [79].

Light-sheet microscopy is a relatively young field and the scientific community around it is still dominated by a tinkerer mentality of *building a microscope around the sample*. There exist designs specifically tailored for very large samples [73, 80] or detailed imaging of small structures [78], as well as single lens implementations of the light-sheet principle [81], for example. A unifying theme and challenge is

the large amount of image data that can be generated by these fast microscopes with large sample and live imaging compatibility, which pose new and specific challenges to downstream image analysis pipelines.

1.1.4 High-throughput and automated microscopy

Aside from, e.g., high spatial or temporal resolution, generating an accurate image of a biological system often also consists of acquiring a large dataset of individual images to capture distributions of features in a population. A simple example could be the distribution of the different cell cycle stages, which can be considered random when imaging a single cell and only becomes apparent when imaging large fields of many cells. Yet, the cell cycle distribution might be an essential readout when screening for the effects of antiproliferative drugs, for example [82]. Such screening studies, which are also common in the pharmaceutical industry, are often conducted using specialized high-content microscopes that are optimized for high throughput imaging of standardized multi-well plates [83].

While it is obvious that computational image processing can improve experimental results through downstream analysis or reconstruction of improved images from raw data, another way it can be employed in the experimental pipeline is during the image acquisition process itself. When studying rare phenotypes or the distribution of features at a fine spatial scale in a large population, a bulk high-throughput approach often becomes infeasible, which is why solutions for smart microscopy were developed that, for example, perform a pre-screening of the whole sample, e.g. multi-well plate, at a very coarse resolution and use image analysis to select candidate regions that are re-imaged at higher resolution [84–86].

Another application of smart microscopy comes from the fact that large samples typically introduce optical aberrations that might also vary in different regions of the sample and that cannot be accounted for by tweaking of the optical parameters

before image acquisition is started. A possibility to counteract this phenomenon is with adaptive optics, components like spatial light modulators (SLMs) or deformable mirrors that can adjust physical properties of the imaging system via electronic commands to adapt the microscope to provide optimal image quality throughout the experiment. For example, similar to applications in astronomy where optical parameters are automatically adjusted to spatially minimize the impulse response of a guide-star laser spot projected into the sodium layer of the Earth's atmosphere to account for aberrations introduced by the lower layers of the atmosphere, laser guide-stars can also be projected into a biological sample and serve as an optimization target for the adaptive optical elements of the system [87]. Another approach is to continuously adapt optical parameters to maximize a quality criterion derived from the detected images themselves [88].

It is my strong opinion that such smart microscopy approaches, where computational improvements are introduced at the earliest stages of an experiment, will continue to grow in importance in the coming years. Specifically, two implementations of the overview $\rightarrow$ candidate region detection $\rightarrow$ high-resolution imaging scheme on different microscopes were developed by me during the work leading up to this thesis, but are yet unfinished and unpublished.

1.2 Sample preparation

While the general importance of sample preparation and labelling was already discussed above, the focus of this section will be on three recent innovative techniques that were successfully employed in the papers presented below [21, 89]. For example, the fluorophore blinking necessary for SMLM can alternatively be implemented by transient binding of fluorophores to the labelled structure via DNA hybridization. This framework of PAINT can be used in a variety of ways, including spectral multiplexing. A further creative way of achieving increased spatial resolution is to physically enlarge the sample, as is done in expansion microscopy. In the other direction of spatial scale, very large samples often cause problems in imaging due to light scattering and attenuation, which can be counteracted by sample clearing techniques.

1.2.1 PAINT

In conventional SMLM, the number of photons that can be detected from a single emitter is limited by the inevitable eventual bleaching of the fluorophore. A strategy to circumvent this limitation is to associate each molecule that should be imaged not with a single fluorophore but with a handle that allows transient binding of a practically unlimited pool of fluorophores in the imaging buffer. In the technique of point accumulation for imaging in nanoscale topology (PAINT), this is implemented by attaching short DNA oligonucleotides to the molecule of interest, e.g. by labeling them with antibodies conjugated to the oligomer. These *docker strands* hybridize transiently with complementary *imager strands* with attached dye molecules in the buffer. By using short oligomers, the unstable binding kinetics lead to virtual *blinking* of the emitters, and localizations over a long series of frames can be combined to create final images with localization accuracies in the single-nanometer range.

The basic PAINT principle can be applied for a variety of tasks aside from very high resolution imaging [90], such as accurate estimation of molecular concentrations from fluorescence traces at a single location from the binding kinetics and associated blinking [91]. By labelling a variety of structures of interest with different docker strands and exchanging the complementary imager strands in the buffer in many rounds, PAINT can be used to enable many-fold color multiplexing [92], which is not only compatible with localization microscopy but many other super-resolution modalities [21]. Along the same lines, the principle can be used to sequentially image closely spaced loci on chromosomes, essentially *walking along the DNA* to capture the three-dimensional structure of chromatin point-by-point [93–95].

1.2.2 Clearing

Thick samples such as whole animal organs or embryos pose unique challenges to microscopic studies as light attenuation and light scattering at refractive index boundaries (e.g. membranes) lead to a deterioration of image quality as a function of imaging depth. While some microscopy modalities such as two-photon microscopy are less susceptible to these effects due to the long wavelengths employed and allow for imaging deep inside the brain of living mice [96], for example, the classical approach to imaging large structures was to physically cut them into thin slices and image the slices individually. Aside from distortions introduced by the cutting, that require laborious image registration after acquisition to reconstruct a three-dimensional volume, the sample preparation itself is also time- and work-intensive.

Clearing protocols [97] seek to enable the imaging of intact samples by making them optically transparent, which corresponds to equalizing the refractive index (RI) throughout the sample as much as possible. They therefore typically include steps of fixation, permeabilization and finally immersion of the sample in a defined

refractive index-matching medium. If there are pigment molecules in the sample, specialized chemical bleaching or decoloring agents can be used to reduce light attenuation. Likewise, lipids can be removed, e.g. by detergents, to reduce the amount of RI boundaries. Lipid removal is prone the disrupting the structural integrity of the sample, though, as membranes serve as an anchoring scaffold for many biomolecules.

A relatively novel clearing strategy consists of hydrogel-embedding-based protocols, such as CLARITY [98] and SWITCH [99] that seek to remove lipids from the sample in bulk while also retaining the structural integrity of the rest of it. To this end, hydrogel monomers are first crosslinked to the non-lipid structures in the sample and then the gel is made to polymerize *around the sample*. With the gel now replacing the lipids as a form of molecular scaffolding, the lipids can be removed with detergents or even electrophoresis before final incubation in a RI-matching medium (figure 1.13). A positive side-effect of such protocols is that the lipid removal leaves to sample very accessible to other molecules, for example antibodies, which often suffer from poor penetration depth in unmodified samples.

1.2.3 Expansion Microscopy (ExM)

Expansion microscopy (ExM) [100] is a perfect example of how out-of-the-box thinking can lead to surprising new solutions of longstanding problems. Instead of trying to improve the resolution of a microscope to allow the imaging of tiny samples, ExM seeks to physically enlarge the sample itself so that fine details can be imaged on a lower resolution microscope.

The basic workflow of ExM is very similar to sample clearing via hydrogel embedding, e.g. the CLARITY protocol. The sample structures of interest (e.g. proteins, RNA or other biomolecules) are crosslinked to hydrogel monomers and the gel is

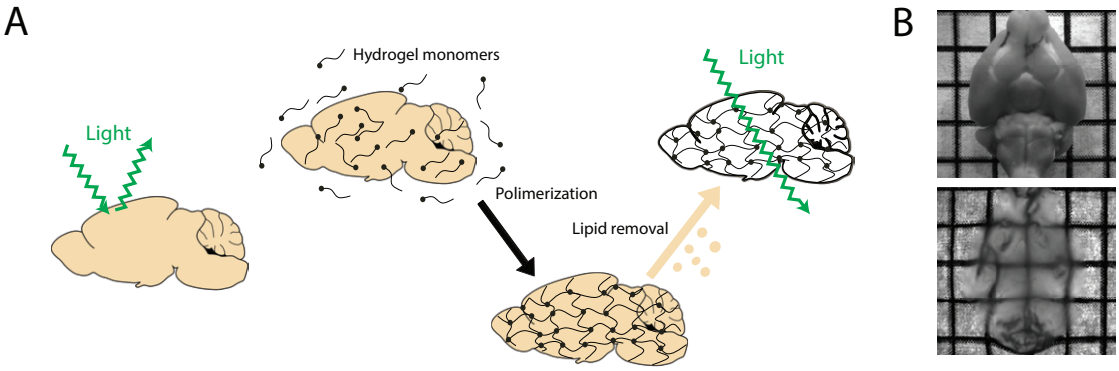


Figure 1.13: Tissue clearing using the CLARITY protocol. (A) Imaging deep within tissues is difficult mainly due to light scattering at refractive index boundaries (left). Schematic workflow of CLARITY: Hydrogel monomers are linked to the sample and polymerized around it. After passive or active electrophoretic removal of light-scattering lipids, imaging deep within the sample becomes possible. (B) CLARITY applied to and adult mouse brain: Photographs of a brain before (top) and after (bottom) clearing. Figure adapted from [89].

polymerized, followed by fragmentation of the sample (but not the hydrogel), typically via enzymatic digest. By employing special polymers that can be swollen by e.g. changing the osmolarity of the buffer, the sample fragments crosslinked to it can be displaced accordingly. Care needs to be taken in this step to ensure that all parts of the polymer expand by the same factor and that expansion happens equally in all directions, so-called *isotropic* expansion. A downside of ExM is that due to the harsh denaturing treatment involved, the protocols need to be adapted so that they leave specific structures of interest intact [101–103].

To measure absolute and not just relative lengths in an expanded sample, the expansion factor needs to be known. Expanded lengths can be converted into original sample lengths by dividing by the expansion factor, which can be estimated accurately from the chemical properties of the polymer mesh used. In addition to the resolution improvement, ExM also leaves the sample components more accessible, allowing for easier staining with antibodies deep in the sample, for example. This is another property that ExM shares with clearing protocols such

as CLARITY or SWITCH. By adapting the ExM protocol, the expansion process can also be compounded: an expanded sample can be expanded again for further resolution gain, a process called iterative expansion microscopy [104].

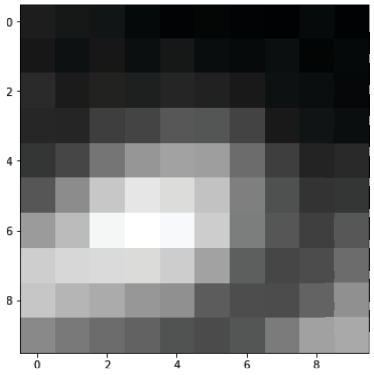
1.3 Computer Vision

Many modern microscopy modalities require processing of the raw data to form an image. This is often referred to as *reconstruction* of the image. For example, all super-resolution microscopy techniques rely on some form of computational reconstruction to some extent to form a final, high-resolution image. Even in STED, often termed a what-you-see-is-what-you-get (WYSIWYG) method, the recorded image, like in any point scanning microscope, cannot be observed by eye but only rendered on a computer after arranging the individual point measurements into a regular grid. The subfield of computer science concerned with processing images and extracting information from them is called *computer vision* [105, 106]. Digital image processing and analysis is an indispensable part of most microscopy-based science, from the stage of image acquisition to the final interpretation of the results (fig 1.1).

Every study presented in this thesis relies on image analysis to some extent, so this section will briefly cover the basics of image processing and a few selected examples that are of importance to the presented studies, before elaborating on the tasks of image alignment - the main problem solved by BigStitcher, the core paper of this thesis - and image segmentation, a task that repeatedly came up in the other projects in this thesis.

1.3.1 Basics of digital images

Digital images are a collection of numbers arranged into a regular grid or matrix (figure 1.14). The individual entries of the matrix are referred to as picture elements, or *pixels* [107]. For conventional color photographs, the pixel values are the color components, which in the simplest case form a red, green, blue 3-tuple (r,g,b), though different color models exist. Alternatively, one can think of a color image as a 3D array, with one axis containing the color components. The im-




```
[[344, 331, 328, 312, 303, 308, 303, 297, 313, 298],
 [333, 321, 334, 319, 331, 315, 314, 320, 305, 312],
 [360, 342, 350, 345, 353, 348, 338, 321, 317, 310],
 [355, 353, 391, 402, 433, 430, 399, 338, 326, 317],
 [377, 404, 490, 560, 584, 575, 474, 390, 352, 359],
 [431, 535, 666, 739, 714, 653, 508, 419, 375, 377],
 [569, 637, 776, 800, 783, 680, 507, 432, 392, 433],
 [683, 700, 709, 711, 679, 584, 446, 404, 413, 473],
 [663, 625, 601, 561, 543, 441, 415, 414, 456, 544],
 [529, 497, 472, 454, 423, 412, 430, 501, 582, 596]]
```

Figure 1.14: An image is just a collection of pixels, a matrix of light intensity measurements at distinct points in time and space. In color images, each pixel contains a list of measurements in each channel, e.g. red, green and blue components in conventional digital photographs.

ages encountered in biomedical image analysis often contain other axes such as z-focus position, time, or rotation (e.g. in multi-view LSFM or raw tomographic images) and can, in general, be considered n-dimensional arrays or rank-n tensors. It should also be noted that for microscopy images, in many cases, the color axis no longer represents vector components in a color space but results from multiple exposures taken with the same grayscale camera, but distinct filter sets to specifically detect certain wavelengths. Images can also be thought of as mappings of pixel coordinates to light intensity measurements $I : \mathbb{N}^n \to \mathbb{R}$. Pixel coordinates can be mapped to real (relative) locations in time and space through a transformation $T : \mathbb{N}^n \to \mathbb{R}^4$, which in the simplest case is just multiplication with pixel distances and inverse objective magnification and time intervals between acquisitions, but might be more complex if the camera is rotated or time intervals are not constant, for example. In real applications, the pixel values are stored in computer memory and need to be represented by bits. While low bit depth requires less storage, precaution has to be taken to not run into over- or underflow of the dynamic range, or numerical inaccuracies if integer representation instead of floating-point representation is used.

1.3.2 Overview of computer vision tasks and techniques

Since images are just a collection of numbers, there are obvious mathematical operations one can perform on them to build image analysis routines. The pixel values in an image or a subregion can be summed, averaged or other statistical measurements can be performed to achieve a low-dimensional summary. Regions can also be described by their shape parameters, like total area, circularity and so forth. Many steps in image analysis pipelines also rely on filters that calculate a number at each pixel location from its surroundings, which is typically implemented as a convolution $I * h$ with the filter kernel h, or the gradients of the image along its axes $\frac{\partial I}{\partial x_i}$.

While image processing can be done to manupulate images for purely aesthetic purposes, in biological image analysis the goal is to extract interpretable information from the input. Common tasks that occur again and again in computer vision include image alignment (or: registration) that seeks to combine multiple, partially overlapping images into one, image restoration - the transformation of real-world images to an estimate of the underlying structure without the contributions of random noise (or the optical system used to acquire the image in the case of deconvolution), the detection of objects and their description by geometric primitives, or segmentation of the images - the classification of all of the pixels into a defined set of labels.

1.3.3 Selected applications of computer vision in biology/microscopy

Principles of image analysis pipelines

While there is a staggering amount of possible processing steps in computer vision, producing enhanced images from input images, segmentation masks from images, lists of shape parameters from masks and so forth, their application to real biological data is done with the ultimate goal of extracting meaningful information

from raw data [108]. Like an experiment in a lab, an image analysis workflow will typically follow common steps: image pre-processing, such as denoising, flatfield-correction or fusion of tiled images into one, identification of biologically relevant objects, e.g. by segmentation or object detection, extraction of a low-dimensional description of the objects, and finally a statistical analysis of the resulting features to e.g. compare effects of mutations or drug treatments.

By keeping this structure of a pipeline made up of individual steps in mind, one can easily replace the individual processing steps if a new situation requires it, e.g. segmentation by a simple automatic threshold calculation that works on images of DNA-stained nuclei might be replaced by a machine-learning-based segmentation for more complex samples. If the building blocks adhere to a more or less strict contract, e.g. of a segmentation step producing a binary foreground-background image, they can be quickly replaced and put together in novel ways. Some software packages for (image) data analysis like KNIME [109] or CellProfiler [110, 111], but also novel developments in ImageJ/Fiji [112] try to incorporate this concept of analysis pipelines into their core user experience to make complicated workflows tractable to non-specialists.

Flatfield correction

One of the simplest examples of how computational image processing can be integrated into microscopy image acquisition and restoration is through the correction of uneven illumination and background signal. As the illumination light intensity is often highest in the center of the field-of-view and lower on the sides, vignetting artefacts that prevent accurate quantitative measurements can occur. Likewise, a digital camera might not register zero counts in the absence of light, leading to an offset that should also be corrected before intensity measurements are made in the image. If one knows the dark image, i.e. the image returned by the camera without any light exposure (I_d), and the bright image(I_b), i.e. the response to a

sample with uniform fluorophore density, one can easily correct for these effects by calculating:

$$C(x) = \frac{(R(x) - I_d(x))\overline{(I_b - I_d)}}{(I_b(x) - I_d(x))} \tag{1.5}$$

Here, R denotes the raw intensities and C the corrected intensities and $\overline{(I_b - I_d)}$ the average of $(I_b - I_d)$. While I_d and I_b can be experimentally determined by acquiring them in addition to the images of the sample, they can also be automatically estimated after the fact from multiple sample images acquired with the same instrument settings [113, 114].

Improvements of LSFM through computational means

The technical innovations of modern light-sheet microscopy, even in its earliest implementations [67], were complemented by image processing strategies to further emphasize the advantages of the novel technique. As the sample is often mounted in a way that it can be rotated around the axis perpendicular to the illumination & detection plane, it can easily be imaged from multiple sides. The individual image stacks can then be computationally combined into a single stack in a process called multi-view reconstruction (MVR) [115]. MVR expands the capabilities of the microscope in two ways: first, by combining opposing views, larger samples can be imaged even if light attenuation and scattering make it impossible to image them in their entirety from one side. Second, by combining views from perpendicular axes, the resolution of the reconstructed image stack can be made isotopic. As the lateral (xy) resolution of objectives is higher than the axial (z) resolution, acquiring a second image stack at a perpendicular angle to the first (in which the original z axis becomes one of the lateral axes) and fusing both stacks can produce results in which spatial resolution is equal along all spatial axes.

Light-sheet microscopy is also well suited for the imaging of very large samples due to its speed. While light-sheet designs for the explicit purpose of imaging large, cleared samples such as mouse brains exist, they often rely on large-FOV (field-of-view), low-NA objectives to capture the entirety of the specimen. These objectives lack the resolution to image intracellular details, however. To achieve high spatial resolution as well as large sample coverage, images can be acquired in a tiled fashion and later combined into one by a process called image stitching [89]. As light-sheet datasets can easily reach sizes of hundreds of gigabytes, they present distinct challenges to stitching software.

Quality control in (super-resolution) microscopy

In all super-resolution microscopy techniques that do not have fundamental resolution limits, such as STED and SMLM, the image quality achieved in practice depends not only on the physical characteristics of the microscope, but also on the sample itself, e.g. through the choice of fluorophores. But even in methods with deterministic resolution, such as SIM (or even conventional non-super-resolution methods), the resolution can be affected by improper alignment of the instrument, sample-dependent aberrations, low signal-to-noise ratio or reconstruction artifacts. To avoid arriving at false conclusions, it is therefore good scientific practice to do some quality control of the acquired image.

The simplest way of measuring resolution in an image is to manually look for small structures, usually points or filaments. One can plot the intensity along a cross-section of the object and in the resulting line profile determine the FWHM, which corresponds closely to the resolution according to the Abbe criterion (figure 1.15, A). For a more robust estimate, point-like objects can be detected automatically and a PSF function can be fitted to them to get an average estimate from multiple PSFs. If enough small structures are available distributed across the whole FOV, one can also detect non-uniformity in the resolution across the FOV.

A less biased method for resolution quantification, that has been applied in electron microscopy for a long time, is Fourier ring/shell correlation (FRC) [116]. FRC proceeds by taking two independent images of the same structure (or, in the case of SMLM, randomly splitting all localizations into two groups and rendering images from each) and calculating the (cross-)correlation of rings of pixels around the origin of the Fourier transform of both images. Close to the origin, the FFTs will contain highly correlated, low frequency information, but as the radius is increased and exceeds the resolution of the images, the two FFTs will only contain uncorrelated noise. Automatic thresholding methods exist to determine a cutoff frequency that corresponds to the highest usable resolution in the images. A recently proposed alternative to FRC that can work on single images but follows similar basic ideas is decorrelation analysis [117].

As the amount of digital (pre)processing that is applied to images is increased, methods for the automatic detection of reconstruction artifacts gain in importance. A straightforward method to identify reconstruction artifacts in super-resolution images derived from computational reconstruction is to compare an artificially blurred version of the reconstructed image to a diffraction-limited image (that can usually be generated easily from raw data). Areas with a large difference between the two images indicate the presence of reconstruction artifacts [118]. A similar approach can be used to ensure fidelity to the original data in deep learning-based super-resolution [119].

1.3.4 Image alignment

One of the most common tasks in computer vision is the alignment of multiple images. Applications that require image alignment include the stitching of panoramas from photographs or maps from satellite images, but also video compression [120]. Likewise, in microscopy, multiple images have to be aligned when imaging large histological slides or when compensating for optical effects such as chromatic

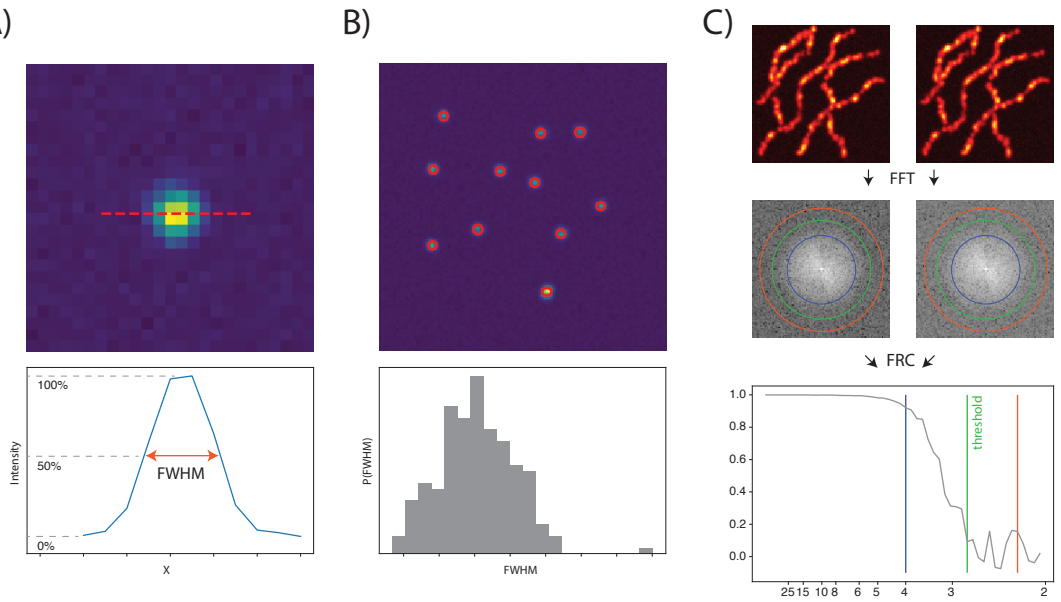


Figure 1.15: Methods to assess the resolution of microscope images. (A) The simplest way of estimating the resolution of a microscope is to acquire an image of an object smaller than the diffraction limit, which will be imaged as a blob corresponding to the microscope PSF. By plotting the intensity along a cross-section, the full width at half maximum (FWHM) and thus the minimal resolvable distance can be determined. (B) A more robust estimate (or distribution, bottom) of the FWHM can be obtained by automatically fitting a PSF model to a large number of point emitters. (C) In Fourier Ring Correlation, two independent images of the same sample are acquired, and the correlation of their FFTs at various distances from the origin is calculated. Due to the finite OTF, high-frequency information will not pass the optical system leaving only uncorrelated noise. The frequency at which the correlation drops below a pre-set threshold corresponds to the resolution of the system.

aberration (in which case the images of different spectral channels have to be aligned), or in general when trying to reconcile large samples with high resolution imaging (in which case the field-of-view is often limited). Furthermore, microscopy images often take the form of three-dimensional z-stacks. While this actually simplifies some alignment steps (e.g. it removes the need to separately account for three-dimensional camera movements), many standard algorithms for image alignment that have been developed for photographs have to be adapted to work in three dimensions.

The process of image alignment can be broken up into several steps: calculating pairwise transformations between overlapping images, finding an optimal global alignment that takes all pairwise transformations into account and saving or rendering the final composed result.

For the first step, the estimation of pairwise transformations T_{ij}, two main approaches can be distinguished: intensity-based alignment that seeks to minimize the difference of pixel values of two images, and interest-point based alignment that first detects corresponding keypoints (e.g. corners and bright or dark blobs) in each image and then fits a transformation mapping the locations to each other (figure 1.16).

1.3.5 Intensity-based image alignment

The first class of image alignment procedures are called intensity-based since they evaluate the intensities of all pixel pairs in the two images, with one of them transformed according to a transformation T with parameters t. For each image pair (I_1, I_2), the goal is to find parameters for the transformation t_{ij} that minimize some error/loss function L, calculated from the pixel intensities at x and $T(x,t)$:

$$t_{ij} = \arg\min_t \sum_x L\left(I_1\left(x\right), I_2\left(T(x,t)\right)\right) \tag{1.6}$$

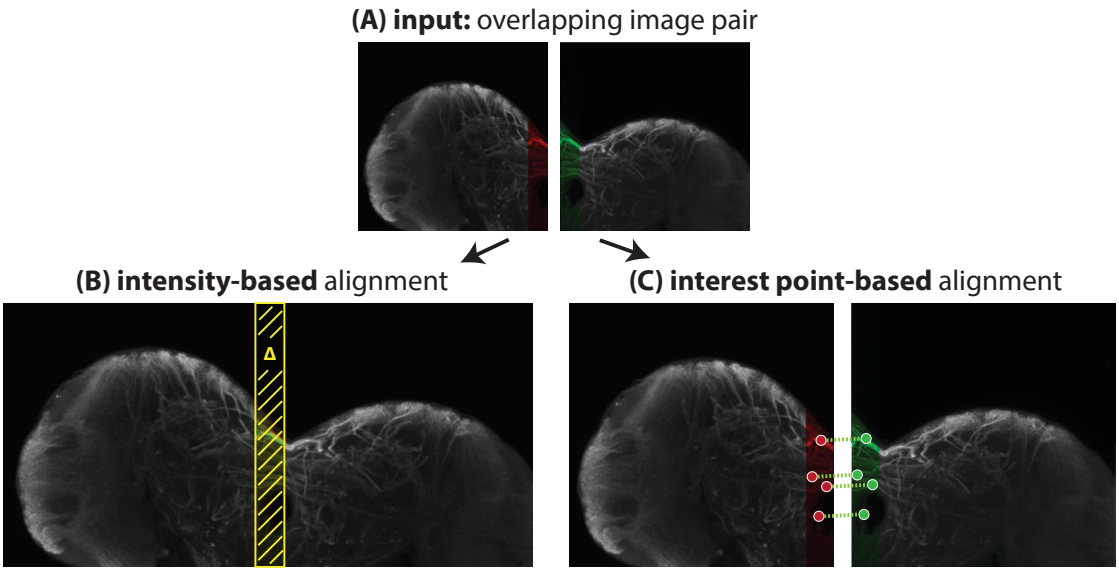


Figure 1.16: Strategies for pairwise image alignment. The input to the problem consists of two images of the same structure with nonzero overlap (A). Strategies for alignment include intensity-based methods, which seek to minimize some error metric calculated from all pixels in the overlapping region (after applying the transformation) (B), or interest point-based methods, which detect interest points in each image and then estimate a transformation from point correspondences (C). Example images from [121], two overlapping confocal images of the Drosophila central nervous system.

The resulting optimal parameters that describe how image I_1 should be transformed, so its pixel intensities have the least discrepancy with the corresponding intensities in I_2. An example for a basic error metric is the sum of squared differences (SSD):

$$SSD = \sum_x ||I_1(x), I_2(T(x,t)||_2 \tag{1.7}$$

Since the images only overlap 100% in the most trivial edge case, special care has to be taken of how to handle intensities at transformed coordinates that lie outside the image. Using a simple indicator function $w(x)$ that is 1 if the coordinates x are inside image I_2 and 0 otherwise, this gives the following windowed SSD:

$$SSD_w = \sum_x w(T(x,t))||I_1(x), I_2(T(x,t)||_2 \tag{1.8}$$

Calculating a difference between the two images with a given shift and N pixels overlap has computational cost of $\mathcal{O}(N)$, and since this has to be repeated for every possible parameter value, we end up with $\mathcal{O}(KN)$ operations for K possible parameter values. For example, the computational cost for checking all possible integer translations is $\mathcal{O}(N^2)$. Since this can quickly become prohibitively expensive, different strategies to reduce the computational cost of intensity-based alignment methods have been developed.

Hierarchical methods

For many computer vision tasks it is helpful to work with multiresolution *pyramids* of images. In the simplest case, one can construct a pyramid by repeatedly downsampling the image by a factor of 2. The image at level l of the pyramid is given by $I^l(x) = I(x * 2^l)$ (note that the original image should be smoothed prior to downsampling to prevent aliasing artifacts). For hierarchical image alignment,

one starts at the coarsest resolution l_{max}, checking all possible values of transformation parameters (sampled in an appropriately coarse grid). One then proceeds down the pyramid, refining the parameter estimate, but limiting the parameter search space to the vicinity of the parameters determined at the previous pyramid level, until level l_0 (full resolution images) is reached.

Fourier-based correlation

An alternative metric for the alignment quality is the *cross correlation* $I_1 \star I_2$ of the two images at a given shift t, which in this case is just the sum of the element-wise product of the images:

$$I_1 \star I_2(t) = \sum_x I_1(x) I_2(x+t) \tag{1.9}$$

Here, the transformation is assumed to be just a translation of I_1 by the shift vector t. An optimal alignment maximizes this cross correlation (or normalized versions that are less susceptible to very bright areas in the images). Like any other pixel-based objective, evaluation the cross correlation for every possible shift vector t is expensive, but similar to convolutions, cross-correlations can be calculated much faster by taking advantage of the Fourier correlation theorem (with $*$ denoting complex conjugation):

$$\mathcal{F}(I_1 \star I_2) = \mathcal{F}(I_1)\mathcal{F}(I_2)^* \tag{1.10}$$

Multiplying the two images elementwise in Fourier space and inverse transforming the result thus gives the correlation values at each possible shift. A modified version of this approach is phase correlation [121], in which the frequency space product is normalized elementwise. This has the advantage that if I_2 is just a shifted version of I_1 ($I_2(x) = I_1(x + t_{12})$ and by the Fourier shift theorem

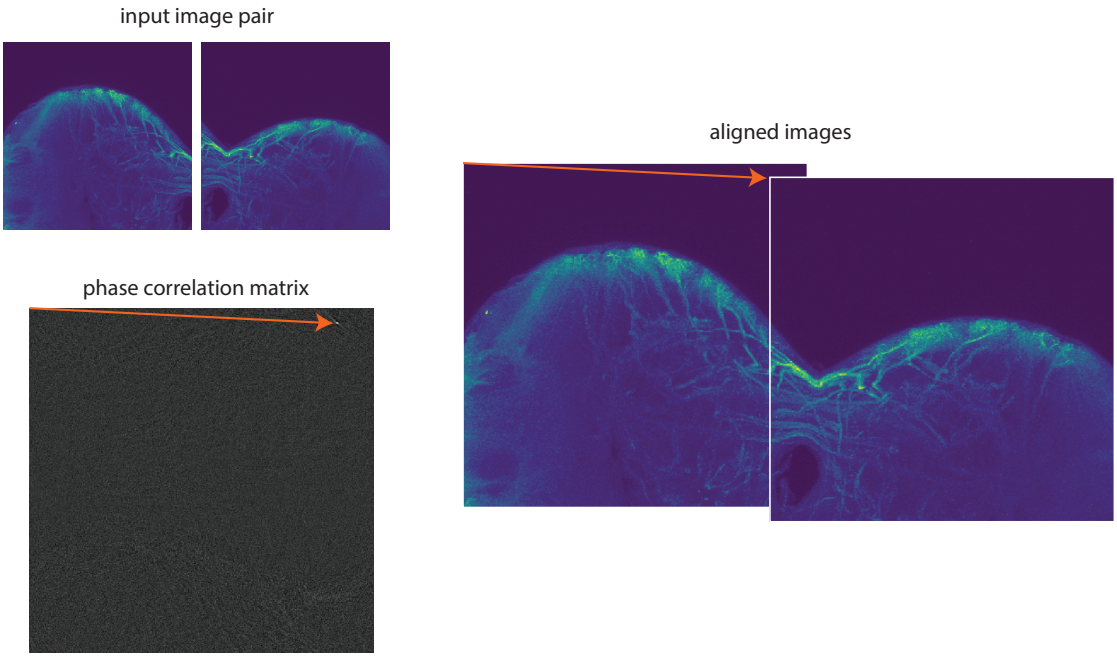


Figure 1.17: Image alignment with phase correlation. By multiplying and normalizing the FFTs of both images elementwise and inverse-transforming them, a phase correlation matrix (PCM) is calculated. Ideally, the PCM contains a single δ-peak at the position corresponding to the shift vector between the two images. Example images from [121].

$\mathcal{F}(I_2)(f) = \mathcal{F}(I_1)(f)e^{-2\pi i t_{12} f})$, the resulting phase correlation matrix (PCM) contains a single intensity peak at the location corresponding to the shift vector t_{12} (figure 1.17):

$$\begin{aligned}
\mathcal{F}(PCM(x)) &= \frac{\mathcal{I}_1(f)\left(\mathcal{I}_1(f)e^{-2\pi i t_{12} f}\right)^*}{|\mathcal{I}_1(f)\left(\mathcal{I}_1(f)e^{-2\pi i t_{12} f}\right)^*|} \\
&= \frac{\mathcal{I}_1(f)\mathcal{I}_1(f)^* e^{2\pi i t_{12} f}}{|\mathcal{I}_1(f)\mathcal{I}_1(f)^*|} \\
&= e^{2\pi i t_{12} f} \\
PCM(x) &= \delta(x + t_{12})
\end{aligned} \tag{1.11}$$

Here, $\mathcal{I}_1 = \mathcal{F}(I_1)$ is the Fourier transform of I_1. Since Fourier transforms can be computed in $\mathcal{O}(NlogN)$ time using FFT algorithms, the overall time complexity of Fourier-based image alignment is also $\mathcal{O}(NlogN)$. Since the Fourier shift and correlation theorems assume circular shifts of the images, there is some ambiguity to the results, which can be resolved by calculating the real space cross-correlation of the 2^n possible shifts or windowed versions of Fourier correlation. The method can also be extended to more complex transformations, e.g. it can determine scale and rotation transformations by working in log-polar coordinates [122].

Gradient-descent based methods

A third computationally effective strategy of optimizing the transformation parameters t consists of performing some sort of *gradient descent* on the parameters, removing the requirement to evaluate the error metric (e.g. SSD) for every possible transformation. Starting from some initial guess p_0, the parameters are iteratively updated by the local gradient of the error metric $\frac{\partial L}{\partial p}(p_i)$ until a local minimum is reached:

$$p_{i+1} = p_i + \epsilon * \frac{\partial L(p_i)}{\partial p} \tag{1.12}$$

Here, ϵ indicates the learning rate of the optimization algorithm. High learning rates might lead to faster convergence, but could also cause the optimization to overshoot and miss the minimum. Gradient descent is a very general method for any kind of optimization problem and various extensions of the basic strategy exist, for example the Lucas-Kanade method, which is specifically tuned for image alignment [123].

1.3.6 Interest point-based image alignment

The second approach to finding the pairwise transformations between overlapping images is by detecting corresponding interest points in both of them and aligning those interest points. Most commonly used transformations between two overlapping images can be estimated if a set of corresponding landmarks is known. For example, if the images are shifted by a translation, a single point pair is enough to align the two images, by simply shifting one image so that the two points coincide. Likewise, an affine transform (with augmented matrix Ab) between two images can be determined from at least $1+n$ n-dimensional point correspondences (x_i, x_i') by solving the linear system (the 2D case is shown for simplicity):

$$\begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 & y_1 & 1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_2 & y_2 & 1 \\ & & \vdots & & & \end{pmatrix} (a_{11}, a_{12}, b_1, a_{21}, a_{22}, b_2)^\top = (x_1', y_1', x_2', y_2', \dots) \quad (1.13)$$

While this provides a convenient framework for manual image alignment by picking interest points by hand and aligning them, high-throughput image analysis calls for automated methods. In general, automatic interest-point based alignment follows the following steps: landmark detection in both images, descriptor generation, descriptor matching, outlier removal and estimation of the transformation from matched inliers.

The first task in interest-point based registration is the detection of suitable landmarks in the images. Good classes of landmarks should be independent of the specific images of an experiment. Therefore, a sensible choice is to look for distinct low-level image features that may occur in any type of image. Examples of low-level features are corners and isolated bright or dark spots, that correspond

to maxima of the responses to single small filters or combinations of small sets of filters, such as the gradients of the images along the axes. Examples for this are the Harris corner detector which makes use of the structure tensor or the FAST corner detector. Another possibility is to filter the image with a Laplacian-of-Gaussian (LoG) kernel $h_{LoG} = \nabla^2\big(\frac{1}{\sqrt{(2\pi\sigma^2)^n}} e^{-\frac{||x||_2}{2\sigma^2}}\big)$ [124] or its approximation by a Difference-of-Gaussians (DoG), which give strong negative or positive response at blob-like structures of a predefined size and easily generalize to more than two dimensions.

Next, a descriptor of the interest point, a vector of defined length summarizing it and its surroundings, must be generated. An important property of a descriptor is invariance to the transformation one wants to estimate. For example, by calculating gradient orientations at a scale proportional to the detection σ_{max} in a pyramid of DoG responses with different σs and at an orientation relative to the dominant gradient direction, the SIFT [125] descriptor of an interest point will be similar in a rotated or scaled version of the image. An alternative that is also feasible in three dimensions is to describe an interest point by the distances to its nearest neighbors, which can be expressed in a local coordinate system that is also invariant to scale and rotation [115].

The next step is to match the descriptors in two images. This is typically done by comparing the distance of the feature vectors according to some metric, e.g. the Euclidean distance for real valued descriptors or Hamming distance for binary descriptors such as the ones generated in ORB [126]. To perform fast matching of the descriptors in one image to their nearest neighbors among the descriptors in the other image without having to do an expensive all-to-all comparison, spatial search data structures like k-d-Trees or approximate algorithms are often employed.

Simple descriptor matching will often produce some amount of erroneous correspondences, which is why the matching is often followed by a step of outlier removal. An

almost ubiquitous example for an outlier removal procedure in the context of image registration is the model-guided random sample consensus (RANSAC) algorithm. RANSAC works by randomly selecting a small subset of the correspondences, trying to fit a transformation to them and then expanding the correspondence set to other inlier correspondences that agree with the model. The process is repeated for a predefined number of iterations, and the largest inlier set discovered this way is retained at the end.

1.3.7 Global optimization

If only two images are to be aligned, the registration process is finished at this point. However, when aligning more than two images, one has to determine the transformations in a global coordinate system. It is convenient to view the images as forming a graph, with vertices corresponding to images and edges between overlapping images. One can just create a spanning tree of this graph and propagate the transformations from one fixed starting image. This alignment may not be optimal in the case of more than one contradictory pairwise transformation for one image. An improvement is to require that the spanning tree be minimal according to some quality criterion, e.g. the pairwise cross-correlation of the images [127], but this is still not globally optimal. The whole problem can also be formulated as a (overdetermined) linear system which can be solved in a least-squares optimal way via singular value decomposition [128, 129]. Alternatively, the problem can be solved via iterative re-application of pairwise transformations [130], allowing for easy introduction of weights for the pairwise transformations or iterative link removal during the optimization process.

1.3.8 Image composition and rendering

The final step after a global transformation T_i for each image I_i^{raw} has been determined is to fuse the individual images into one resulting composite image. In

the simplest case, this can be achieved by simply averaging the transformed images:

$$I^{fused} = \frac{\sum_i I_i^{raw}(T_i^{-1}(x))w_i(T_i^{-1}(x))}{\sum_i w_i(T_i^{-1}(x))} \tag{1.14}$$

Here w_i can be a simple windowing function, but also a more sophisticated weight for a given input pixel, e.g. to de-emphasize contributions from image borders or regions of low SNR. Of course, care has to be taken in the implementation of the average to set locations where none of the input images contribute to a predefined background value to prevent division by zero. The resulting composite image can either be saved or rendered on-the-fly. For very large datasets, for example LSFM data, on-the-fly rendering combined with hierarchical data representation in multiresolution pyramids, as it is done by BigDataViewer [131], is often desirable to allow quick inspection on conventional hardware (figure 1.18).

1.3.9 Image segmentation

Image alignment is an example of a task that typically occurs in the earlier steps of an experiment: it produces, for example, a new image providing a high-resolution view at a large sample constructed from many small fields-of-view or improved optical quality through fusion of many images. A later step in the experimental pipeline is to actually measure the properties of the biological system studied. In many cases, the system consists of distinct parts: specific tissues in a slice, single cells, organelles or other subcellular structures. Researchers are interested in the properties of those structures, which occupy only a subset of all pixels in an image, which leads to the tasks of object detection and image segmentation. Specifically, object detection refers to the extraction of a set of geometric primitives from the images, for example interest point localization as described above, detection of bounding boxes around non-point like objects or even more complex tasks such

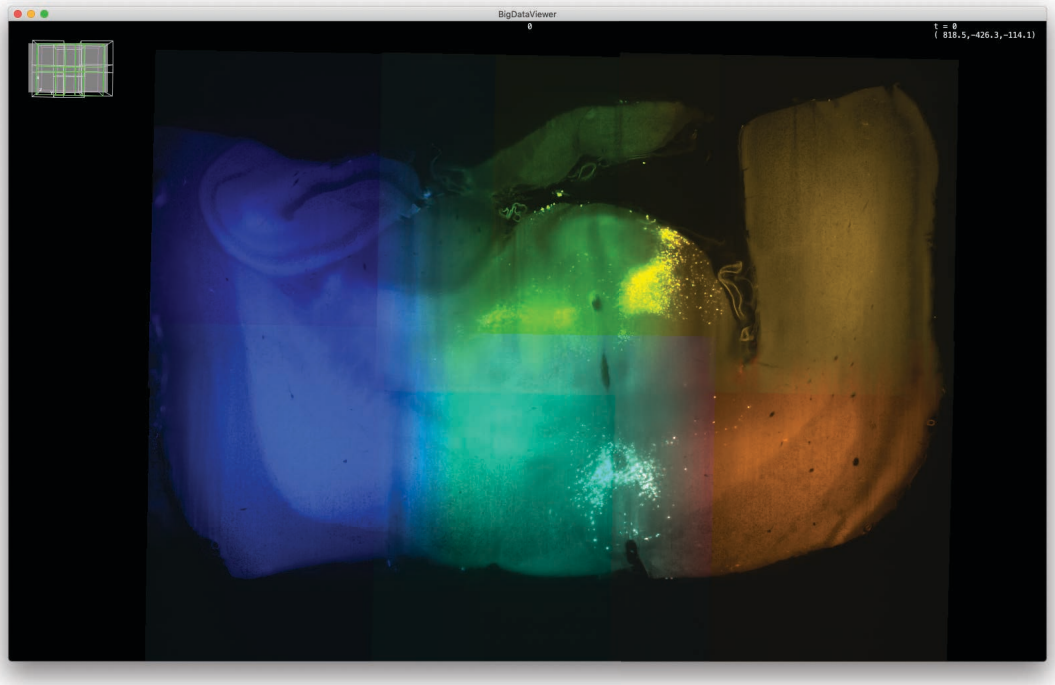


Figure 1.18: On-the-fly visualization of large multi-image datasets with BigDataViewer. The dataset is a multi-view LSFM recording of a cleared coronal slice through an adult mouse brain expressing H2B-GFP in BSX expressing neurons imaged at 2.5x magnification, consisting of 12 (6 shown) tiled and rotated image stacks, totalling at 166GB of data, as presented in [89]. The images were aligned using BigStitcher and the results can be inspected immediately without having to write a fused volume to disk. Individual images are shown in different false colors.

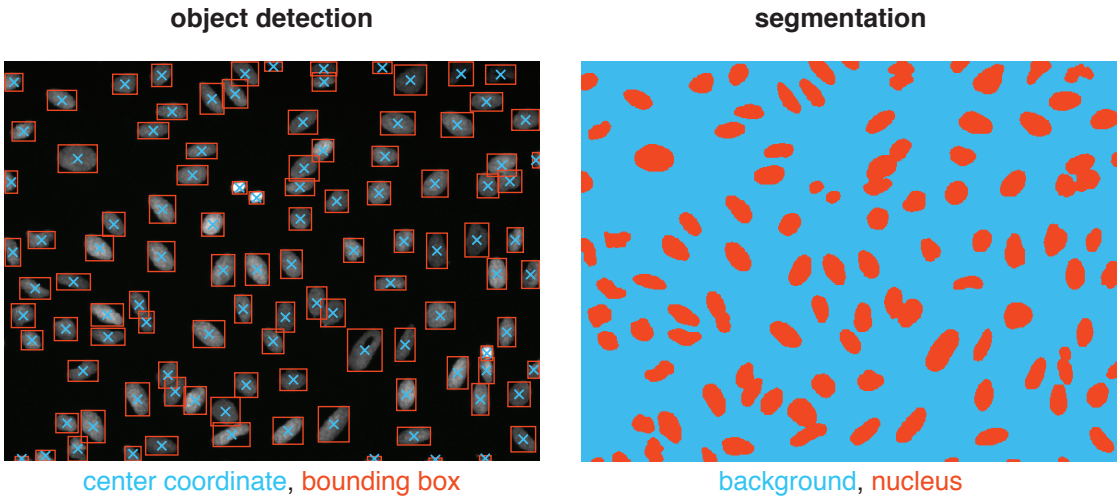


Figure 1.19: Principles of object detection and segmentation, illustrated on an image of DNA-stained nuclei taken from the Kaggle Data Science Bowl 2018 [133] training dataset (which contains manually labelled nuclei). The task in object detection (left) consists of extracting a set of geometric primitives describing instances of objects in the image, e.g. their center coordinate or an axis-aligned bounding box. In segmentation (right), each pixel of the input image is assigned a class identity (e.g. background, nucleus), so the result of the operation is another image.

as the extraction of a biometric face description from human portraits or pose descriptors (i.e. body part locations) from videos of animal tracking experiments [132]. Segmentation, on the other hand, refers to the task of assigning labels to every pixel (in the simplest case of binary segmentation the labels are just foreground and background). The result of segmentation is not a set of object descriptors, but rather another image (figure 1.19).

The most basic segmentation algorithm is to pick a threshold intensity value t and set every pixel brighter than t to foreground and the others to background. The threshold could be picked interactively by hand, but also through an automatic procedure, such as the minimization of intra-class variance in Otsu's method [134].

In this simple case, segmentation just gives a global labelling of the input image, but one is typically more interested in segmentation of all the instances of an object class (e.g. cell or nucleus). This problem is called instance segmentation, the subdivision of the mask image of a class into individual instances of objects. If the objects have round shapes and are only rarely touching, as is the case for DNA-stained nuclei, an instance segmentation can be achieved through simple morphological operations, e.g. the watershed transform, on the mask image. More powerful algorithms that provide instance segmentation from the start include Active Contours [135, 136], in which polygons are iteratively updated to minimize an energy function including shape constraints (calculated from the geometry of the polygons themselves, e.g. roundness) and intensity constraints (calculated from image or filter intensities at the locations of geometrical objects, e.g. response to an edge filter) and related methods of iterative region competition [137].

Specialized methods for many types of samples exist, but a general and powerful way to look at the problem is as a classification task on pixel features. To achieve robust performance on a variety of image datasets, looking at just the intensity of pixels is often not enough. A very general approach is to calculate a vector of features at each pixel, which can then be classified by machine learning (ML) approaches, either through unsupervised clustering or supervised classification using user-provided labels for some of the pixels in the dataset. Due to its importance in many cutting-edge computer vision applications, the principles of machine learning will be elaborated upon below.

1.3.10 Machine learning

While the field of machine learning is much broader than its applications in computer vision, it is that area that has seen some of the most spectacular applications of machine learning, and specifically *deep leaning*, in recent years. Machine learning, a subfield of computer science, aims to solve problems not by implementing

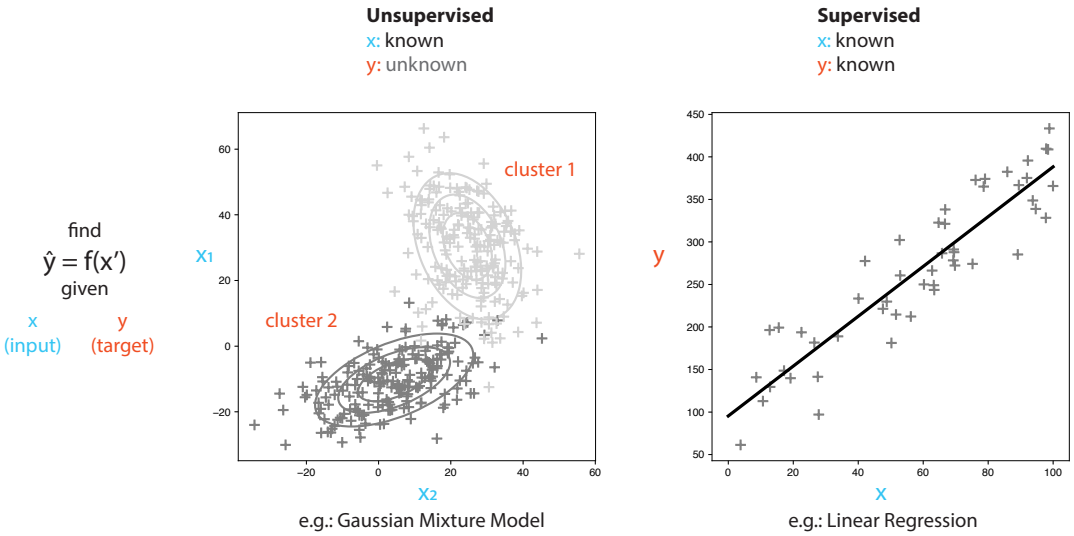


Figure 1.20: Principles of supervised and unsupervised machine learning. In supervised ML (right), the goal is to find parameters of a model to estimate a target $\hat{y}$ given an input x'. The parameters are set based on example (ground-truth) pairs of (x, y). In unsupervised learning (left), only the input x, but no explicit target y, is given for the example data. The goal of unsupervised methods is to automatically find patterns in the input features, e.g. to divide them into a number of clusters, or to use dimensionality reduction to find a compressed representation of the input for easy visualization and downstream processing.

step-by-step programs with manually written instructions, but by fitting flexible statistical models to data; a ML-based program thus *learns* from *experience* presented to it. While ML shares many concepts with statistical modelling, a field that has seen extensive study for over a hundred years, it differs in the aims it seeks to achieve. Whereas in statistical modeling, the fitted model is usually the end result of analysis and should ideally provide condensed and interpretable information about the matter of interest, ML typically employs similar methods to solve any kind of task that is hard to encode in a conventional algorithm, but for which example data can be provided by human operators. Methods and models used in this way do not necessarily need to be easily interpretable and can therefore be made much more complex.

More formally, the task solved in machine learning is to automatically find a procedure to produce an estimate $\hat{y}$ given a (new) input x'. While the general nature of the procedure is given, parameters are optimized given known inputs x, and often known outputs y, the so-called *ground-truth* data. Depending on whether y is known or not, we talk about supervised or unsupervised learning.

In unsupervised learning, only the input x, but no hard target y is known and the goal is often to produce a representation of x (or a new x') that is easier to handle or interpret, or an implicit target, e.g. assignment of the data points to similar subgroups. Typical tools of unsupervised learning include clustering methods that assign inputs into distinct groups (e.g. via K-Means or hierarchical clustering/Gaussian mixture models, figure 1.20) or dimensionality reduction methods such as PCA or t-SNE that seek to encode input vectors into lower dimensionality representations, that are more amenable to further automatic or manual analysis and visualization.

In supervised learning, pairs of (x_i, y_i) are provided and the goal is to find a procedure to estimate the target y from the input x. The ys might be values from a continuous space, in which case the task of estimating y is often referred to as *regression*, or form a discrete set of classes or labels, in which case the task of finding the most likely label or a probability for each label is referred to as *classification*. Simple examples of supervised learning are linear regression, which seeks to fit a line (or hyperplane for inputs of dimensionality >1) through a set of (x, y)-pairs, such that the square distance of y to $\hat{y}$ (the *residual*) is minimized: $\arg\min_{(\alpha,\beta)}(X\alpha + \beta - y)^\top(X\alpha + \beta - y)$ (with X being a design matrix whose rows are x_i and y the vector of all ys) or logistic regression in which the output of a linear transformation on x is fed into a logistic sigmoid function to produce a binary classification probability $P(class = 1) = \sigma(x\alpha + \beta)$.

Modern high-capacity models [138] for classification tasks include support vector machines (SVMs), in which classes are separated by hyperplanes of maximal margin (and nonlinear variants thereof in which the feature vectors are projected into a higher dimensional space through kernel functions) and decision tree models. Decision trees have the advantage of being easily interpretable for humans. On the other hand, one can construct very powerful classifiers by constructing an ensemble classifier from many decision trees, e.g. generated through Random Forest or boosting strategies.

Another very flexible type of machine learning model are artificial neural networks (ANNs), which have been studied for a long time, but saw a resurgence in recent years due to improvements in computational capacity and availability of training data (see below).

While it is tempting to use the most flexible and powerful (high capacity) models, that can approximate a difficult regression task or classification boundary arbitrarily well, extra care has to be taken to avoid overfitting: the model can learn to reproduce the target perfectly for xs in the ground-truth data by "learning the dataset by heart", but fail to generalize to new x'. It is therefore standard practice among machine learning practitioners to split the ground truth data into a training set and a test set - while the training set is used to fit a model, the test set is reserved for evaluation of how well the model generalizes to new data.

Image segmentation by machine learning

As motivated above, every pixel in an image can be described not only by its own intensity $I(x)$, but by the results of different filters F applied to the image $F(I)(x)$, somewhat analogous to the expansion of the feature space in kernel SVMs. Thus, each pixel has a feature vector $f_x = (F_1(I)(x), F_2(I)(x), \dots)^\top$ associated with it. The pixel features can either be used as-is to cluster the pixels into a set of classes

in an unsupervised manner or be assigned ground-truth labels and then be used to train a supervised learning model that is capable of assigning classes to the yet unlabeled pixels. Most tools for ML-based image segmentation [139, 140] follow the second route: the user has to manually draw labels onto some pixels in their image dataset, the tool will then calculate a set of feature maps (e.g. blurred versions of the input, edge or spot filters, such as a LoG with various σ or texture features such as the eigenvalues of the structure tensor). The features at each labelled pixel and the label itself are then used to train a model (with Random Forests being a popular choice) that can predict the class identity for other pixels.

While a generic set of low-level image features promises, and achieves, good performance on many segmentation tasks done in this manner, a step further is to make the features themselves a learnable part of the model. This is possible by using models based on ANNs, which underlie many state-of-the-art computer vision systems. The effectiveness of neural-network based models in image analysis tasks can be seen as the driving force behind the current resurgence of these models under the name of *deep learning* [141].

Deep learning

At the time of writing of this thesis, the terms deep learning and machine learning (as well as the more general principle of artificial intelligence, AI) are often used synonymously. However, more specifically, deep learning refers to machine learning performed with artificial neural networks. ANNs are models that propagate their input through an interconnected set (or network) of units called artificial neurons, due to their similarity to a coarse model of biological neurons.

Artificial neurons can also be seen as a flexible, general-purpose unit of computation. The output of an artificial neuron is defined as a nonlinear activation function f_a applied the weighted sum of the inputs plus an optional bias term:

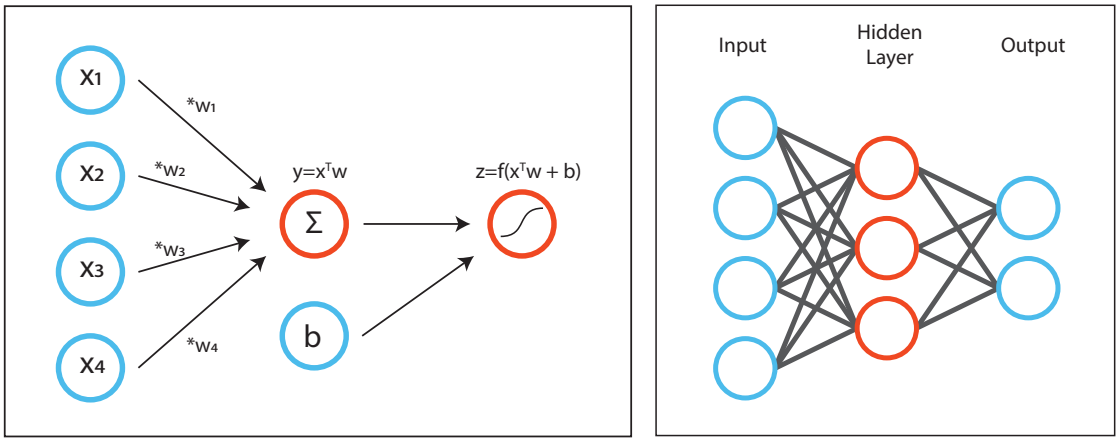


Figure 1.21: Building blocks of artificial neural networks. Left: A single artificial neuron computes a linear function of the inputs, i.e. a weighted sum plus a bias. A nonlinear activation function is applied to the intermediate result y to generate an output z. During training, the weights w_i of all neurons in a network are updated to approximate the desired ground-truth output. Right: By combining many artificial neurons and organizing them in layers, an artificial neural network (ANN) is constructed. The first (input) layer corresponds to the data x fed into the network, the activations of the last layer should correspond to the desired output y, intermediate layers are called hidden layers.

$y = f_a(a^\top w + b)$ (figure 1.21). Using multiple neurons with different weights and biases, a hidden layer whose output is $h = (y_1, y_2, \ldots, y_n)^\top$ can be formed. The outputs of a hidden layer n can be fed into the next layer $n + 1$ to give rise to a multilayer feedforward neural network, with the outputs of the last layer giving the targets $\hat{y}$ to be estimated. If an ANN consists of many layers (current architectures might have several dozens of layers), it is referred to as a deep neural network, hence the name deep learning. Multilayer networks also motivate the necessity of a nonlinear activation function f_a, since without that, all the intermediate linear transformations could be collapsed into one, making the network equivalent to simple linear regression.

The task of training a network consists of setting the weights w_i, b_i for every hidden unit so that the discrepancy of $\hat{y}$ and the ground truth targets y, according to some

task-specific loss function $L(y, \hat{y})$, is minimized. Since the network can be seen as a concatenation of many simple, differentiable functions, the gradient of L with respect to all the weights $\nabla_w L$ can be easily computed using a procedure called backpropagation. The weights can then be updated using simple gradient-based optimization [142].

When trying to process images with ANNs, one quickly runs into limitations as the inputs often consist of millions of pixels and it is not computationally feasible to calculate a weighted sum of all of them for every hidden unit. Furthermore, structures in images usually do not span the entire image and might occur at different locations in different instances of images of the same object. A way around these problems is to assign every hidden unit in a layer a position corresponding to the input image and not compute the weighted sum of all input pixels, but only of those in close proximity to it. Also, to get the same results for the same input structure at different locations, the hidden units are made to share the same weights. This corresponds to calculating a convolution with learnable kernels at every hidden layer (or, in practice, a predefined number of convolutions) and ANNs that follow the rules of weight sharing and local connectivity are called convolutional neural networks (CNNs, figure 1.22) [143].

Like in fully connected networks, multiple convolutional layers are stacked in CNNs. The outputs of them can be thought of as responses to a set of filters at the first hidden layer and filters applied to those responses at latter layers, with the kernels of the filters being learnable parameters. Other architectural features often encountered in CNNs include pooling layers that downsample their input, condensing it and expanding the field-of-perception, but also upsampling steps and convolutions with fractional stride, which can be used to synthetize an image from a low-dimensional representation. Depending on how the layers in a CNN are arranged, the final output could be just a class probability vector for classification of

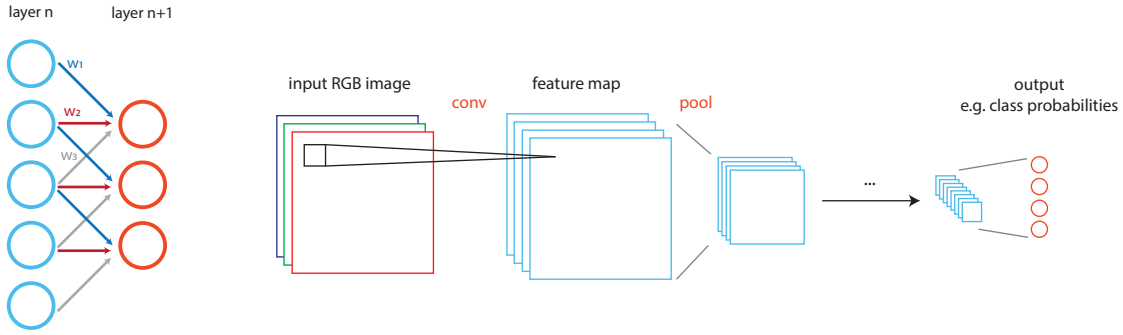


Figure 1.22: Principles of convolutional neural networks (CNNs). Left: CNNs implement the principles of parameter sharing and local connectivity: a neuron in layer n+1 is only connected to a subset of neurons in layer n; furthermore, the same weights (different colored arrows) are re-used (with shifted inputs) for all neurons in layer n+1. Right: When processing images, weighted sums of the neighbourhood with weight sharing correspond to convolutions that produce feature maps from input images (in practice, multiple convolution kernels are learned at each layer to produce multiple features maps). Convolutional layers are typically interleaved with pooling (downsampling) layers, and the same architectural building blocks are repeated for many layers, generating more abstract but semantically rich features-of-features as one goes deeper into the network. In the last layers of the network, images have been condensed into a large feature stack with low spatial dimension that can either directly be interpreted as an output or fed into a few fully connected layers to produce results such as class probabilities.

the image into distinct categories, another image (that is denoised, for example), an image of class probabilities at each pixel (and thus a segmentation of the input) or more complex, structured output that encodes, e.g. the presence and location of bounding boxes for object detection tasks. Aside from ubiquitous applications in industry, CNNs are used more and more in science and achieve unparalleled performance in biological tasks such as cell or nucleus segmentation [133, 144], cell type classification [145] or restoration of low SNR images [146] and computational super-resolution [119].

Chapter 2

Original Papers

2.1 BigStitcher: Reconstructing high-resolution image datasets of cleared and expanded samples

2.1.1 Main Paper

nature methods

BRIEF COMMUNICATION

https://doi.org/10.1038/s41592-019-0501-0

BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples

David Hörl[1,2,7], Fabio Rojas Rusak[3,7], Friedrich Preusser[2], Paul Tillberg[4], Nadine Randel[4], Raghav K. Chhetri[4], Albert Cardona[4,5], Philipp J. Keller[4], Hartmann Harz[1], Heinrich Leonhardt[1], Mathias Treier[3,6] and Stephan Preibisch[2*]

Light-sheet imaging of cleared and expanded samples creates terabyte-sized datasets that consist of many unaligned three-dimensional image tiles, which must be reconstructed before analysis. We developed the BigStitcher software to address this challenge. BigStitcher enables interactive visualization, fast and precise alignment, spatially resolved quality estimation, real-time fusion and deconvolution of dual-illumination, multitile, multiview datasets. The software also compensates for optical effects, thereby improving accuracy and enabling subsequent biological analysis.

Sample clearing[1] and expansion microscopy (ExM)[2] are powerful protocols that create large transparent volumes of whole tissues and organisms (Fig. 1a,b and Supplementary Notes 1–3; Methods). Using light-sheet microscopy[3–5] (Fig. 1c), these samples can be imaged with subcellular resolution in their entirety within a few hours[6]. These acquisitions have the potential to be powerful tools for whole-tissue and whole-organism studies because they preserve endogenous fluorescent proteins (Supplementary Fig. 1) and are compatible with most staining methods.

However, the raw data acquired with the microscope are not directly suitable for visualization and analysis. Many large overlapping three-dimensional (3D) image tiles are collected that amount to many terabytes in size and require image alignment (Fig. 1d–n). Owing to sample-induced refraction and scattering of the light sheet in the direction of illumination[7], 3D image tiles are typically acquired twice with alternating illumination from opposing directions[5] to achieve full coverage (Figs. 1d and 2, and Supplementary Fig. 2). Similarly, emitted light is distorted by the sample, effectively limiting the maximal imaging depth at which useful data can be collected (Fig. 1n and Supplementary Fig. 1). Additionally, overlapping images suffer from spherical and chromatic aberrations (Supplementary Figs. 3 and 4). For reconstruction, and to make these complex datasets easily accessible to biologists and computer scientists, we developed the BigStitcher software. It enables interactive visualization using BigDataViewer[8], fast and precise alignment, quality estimation, real-time fusion, deconvolution and alignment of multitile acquisitions taken from different physical orientations (so-called multitile 'views'), thereby effectively doubling the size of specimens that can be imaged (Fig. 1n), while further orthogonal views can render the resolution isotropic.

BigStitcher features a new user-friendly importer for a multitude of vendor-specific and custom formats that is based on BioFormats[9], and accesses image data through memory-cached virtual loading[10], which can optionally be combined with virtual flat-field correction (Supplementary Figs. 5 and 6, and Supplementary Notes 4 and 5). Performance is optimal when images are initially converted to a multiresolution, blocked and compressed format such as HDF5 (ref. [8]) enabling interactive visualization, processing and interaction with terabyte-sized image datasets.

Accurate reconstruction of these large complex datasets requires compensation for the different types of image and intensity transformations that are introduced by the acquisition process. We therefore developed an interactive stepwise process that compensates for all relevant transformations while providing spatially localized feedback on the quality of the acquired image data (Supplementary Figs. 7 and 8).

First, overlapping 3D image tiles are acquired to cover the entire sample for each acquisition angle. Although approximate tile locations are typically known, translation stages usually show substantial inaccuracies (Supplementary Fig. 9). To compute locations for every image tile, we developed an image stitching algorithm optimized for very large datasets that can deal with acquisitions arranged in non-regular grids (Fig. 3a) containing empty images and multiple independent samples (Supplementary Fig. 10). As acquisitions often consist of hundreds of gigabyte-sized image tiles, each containing very different information (Fig. 3), we initially compute each shift for all pairs of overlapping tiles (links) using the parameter-free phase correlation method[11–14] on downsampled images. It computes all possible shifts between two images, and intensity peaks in the resulting phase correlation matrix correspond to shifts with high correlation that we localize with subpixel precision (Supplementary Fig. 11 and Supplementary Note 6; Methods). Using simulations, we showed that our new pairwise stitching method achieved errors below 1 pixel while reducing computation times 100-fold (Supplementary Figs. 12–15 and Supplementary Note 7; Methods). As correlation-based approaches can fail for image pairs characterized by repetitive patterns, noise or low information content, computing final image tile locations requires global optimization, which sometimes needs to be combined with manual curation (Supplementary Fig. 16 and Supplementary Video 1). Our new global optimization method extends the concept of identifying tile positions by minimizing the distance between all links[12,15], which, in comparison to computing a minimum spanning tree[16], averages out normally distributed link errors (Supplementary Fig. 15) dur-

[1]Department of Biology II, Ludwig-Maximilians-Universität München, Munich, Germany. [2]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. [3]Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. [4]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. [5]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK. [6]Charité-Universitätsmedizin Berlin, Berlin, Germany. [7]These authors contributed equally: David Hörl, Fabio Rojas Rusak. *e-mail: stephan.preibisch@mdc-berlin.de

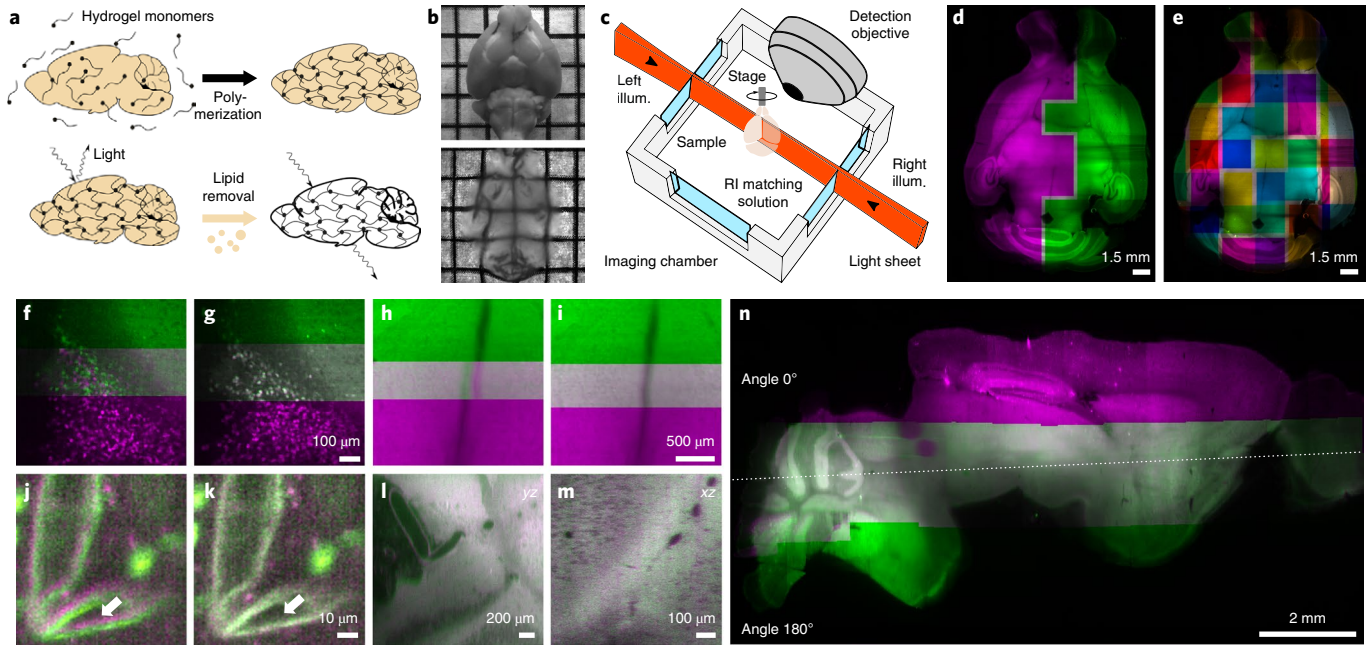


Fig. 1 | BigStitcher principles. a, Schematic of the CLARITY sample clearing process. **b**, Adult mouse brain before (top) and after (bottom) clearing. **c**, General layout of the type of light-sheet microscope used for most acquisitions[3,5]. RI, refractive index. **d**, A single slice through an entire adult mouse brain acquired with dual-sided illumination. The left illumination (illum.) is shown in pink and the right illumination is shown in green; image tiles indicate the illumination direction that was automatically selected for each tile. **e**, Overview of an entire section of an acquired adult mouse brain; different colors indicate individual image tiles (each 1920 × 1920 × 770 pixels). **f**–**i**, Illustration of the result of image stitching from a mouse brain expressing the histone 2B–eGFP lineage tracing marker in BSX-expressing neurons ($Bsx^{H2BeGFP}$) using phase correlation before (**f**,**h**) and after (**g**,**i**) stitching. **j**,**k**, The effect of ICP refinement on two different channels with sufficient autofluorescence visible in both channels. Arrows indicate the difference before (**j**) and after (**k**) refinement. **l**,**m**, Illustration of the high-quality multiview reconstruction for two overlapping multitile views at 0° (magenta) and 180° (green) for an axial versus axial (**l**) and lateral versus axial (**m**) view. **n**, One slice through an entire adult mouse brain (2.24 terabytes of raw data); both views are shown in axial orientation looking along the rotation axis of the microscope. The dotted line illustrates the middle of the section. In **b**,**d**–**n**, clearing and alignment (Methods) were performed on six independent samples with similar results (Supplementary Table 1).

ing optimization, as tiles are typically linked to many neighbors (Fig. 3a). Incorrect links are filtered by quality and by iteratively removing the link that disagrees most with the global optimization result[12] using a new compound metric. In current implementations, unconnected tiles (for example, empty images) and multiple independent objects in an acquisition are handled by ignoring them[12,13], or assuming regular two-dimensional (2D) translational grids[14]. Here we present a generic solution to this problem by introducing the concept of strong and weak links (Supplementary Fig. 10), which is independent of the original tile arrangement and not limited to translations. Strong links correspond to confirmed links, while weak links are derived from current transformations (for example, approximately known tile positions). Optimizing both link types in an acquisition yields accurate registration results within strongly linked regions and optimal alignments for weakly linked groups of tiles (Fig. 3a and Supplementary Fig. 10; Methods). However, correct tile placement (that is, solving the classical stitching problem) represents the first step and is usually not sufficient to properly align dual-illumination, multitile datasets (Fig. 2).

Second, because microscopy images suffer from spherical and chromatic aberrations that can be approximated reasonably well by affine transformations if distortions are in the range of a few pixels (Supplementary Figs. 3 and 4), we implemented a single-step interest-point-based alignment step that automatically extracts interest points and applies a variation of the iterative closest point algorithm (ICP)[17] on affine transformations. In combination with our new global optimization BigStitcher is able to compensate for small affine distortions that arise from spherical, and also chromatic aberrations if the same autofluorescent structures are visible across channels (Fig. 1i,j and Supplementary Figs. 3, 4 and 9; Methods).

Third, although samples are highly transparent (Fig. 1b), light scattering is an issue when imaging in tissues at depths on the order of centimeters. Although improved designs were recently proposed[18], dual-sided light-sheet illumination[5] (Fig. 1c,d) remains the most prominent method to double the sample size for which high-resolution image data can be collected laterally. Before alignment, we automatically suggest the best illumination direction for each tile by estimating image sharpness (Fig. 1d) using newly developed methods (Supplementary Fig. 2; Methods). Unexpectedly, we observed non-rigid image deformations between image tiles with different illumination direction (Figs. 1d and 2a–d). To understand how these transformations are created, we performed simulations of light propagation in tissue using ray tracing (Fig. 2e–h; Methods). These simulations showed that refractions within the illumination light path can lead to different parts of the sample being illuminated, and these parts can both lie in focus of the same detection objective owing to typical depths of field being in the range of several tens of micrometers (Supplementary Table 1). To compensate for these transformations, we implemented a virtual non-rigid alignment method based on identified corresponding interest points[19], as well as a piecewise ICP-based affine alignment based on virtual splitting of image tiles into smaller blocks (Fig. 2b–d,i–m; Methods). Depending on the magnitude of refraction, affine, split-affine or non-rigid alignment was the best choice for precise alignment (compare Fig. 2a–d and Fig. 2k–m), which is possible as long as the light-sheet remains within focus. However, once the light

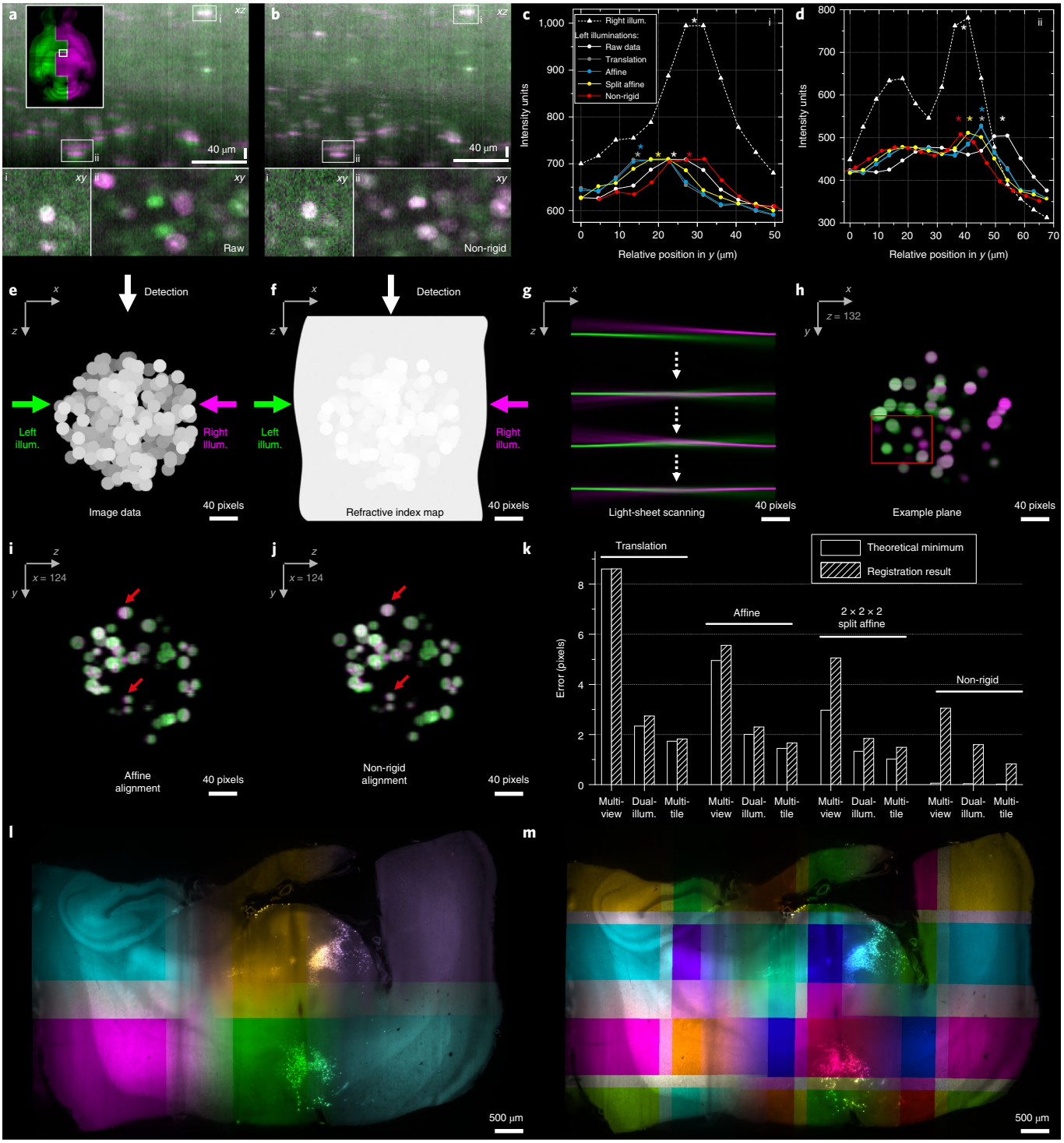


Fig. 2 | Optical aberrations, light simulations and alignment quantification. a, Expanded view of an area that shows misalignment between the two different illumination directions (left and right illuminations are shown in green and pink, respectively) of the original acquired images. An overview of the entire mouse brain is shown for orientation. Two selected areas (labeled i and ii) from the *xz* view are shown in the *xy* view below, highlighting the misalignment. **b**, The same views as in **a**, but corrected using non-rigid alignment; again, two are shown in the *xy* view below. **c**,**d**, Plots of cross-sections of the selected areas in **a** (ii plotted in **c**) and **b** (ii plotted in **d**) comparing different BigStitcher alignment methods. Asterisks mark the approximate centers of nuclei, which should ideally overlap between different illumination directions. **e**–**h**, Simulation of a spheroid-like object using ray tracing that recapitulates (red box) the non-rigid deformations indicated in **a**(ii). Supplementary Video 2 shows a corresponding animation. **e**, Simulated image data. **f**, Refractive index map. **g**, Illustration of the effects on the light sheet while scanning. **h**, An example slice of the simulated object. **i**,**j**, Comparison of the alignment quality for the affine and non-rigid alignments of the simulated data. **k**, Comparison of the best theoretically achievable registration quality from the manually selected corresponding points and a typical registration result when using different alignment modes in BigStitcher using the datasets shown in **l** and **m**. **l**,**m**, An as-small-as-possible 166-gigabyte multiview, dual-illumination, multitile dataset specifically acquired for error quantification in the original configuration (**l**) and after the virtual 2×2×2 split (**m**). In **a**–**m**, Non-rigid deformations of varying degree were observed for all dual-illumination datasets (Supplementary Table 1), a single simulated dataset was created and a single dataset for precise quantification was acquired. Different affine and non-rigid alignments were applied to those three datasets.

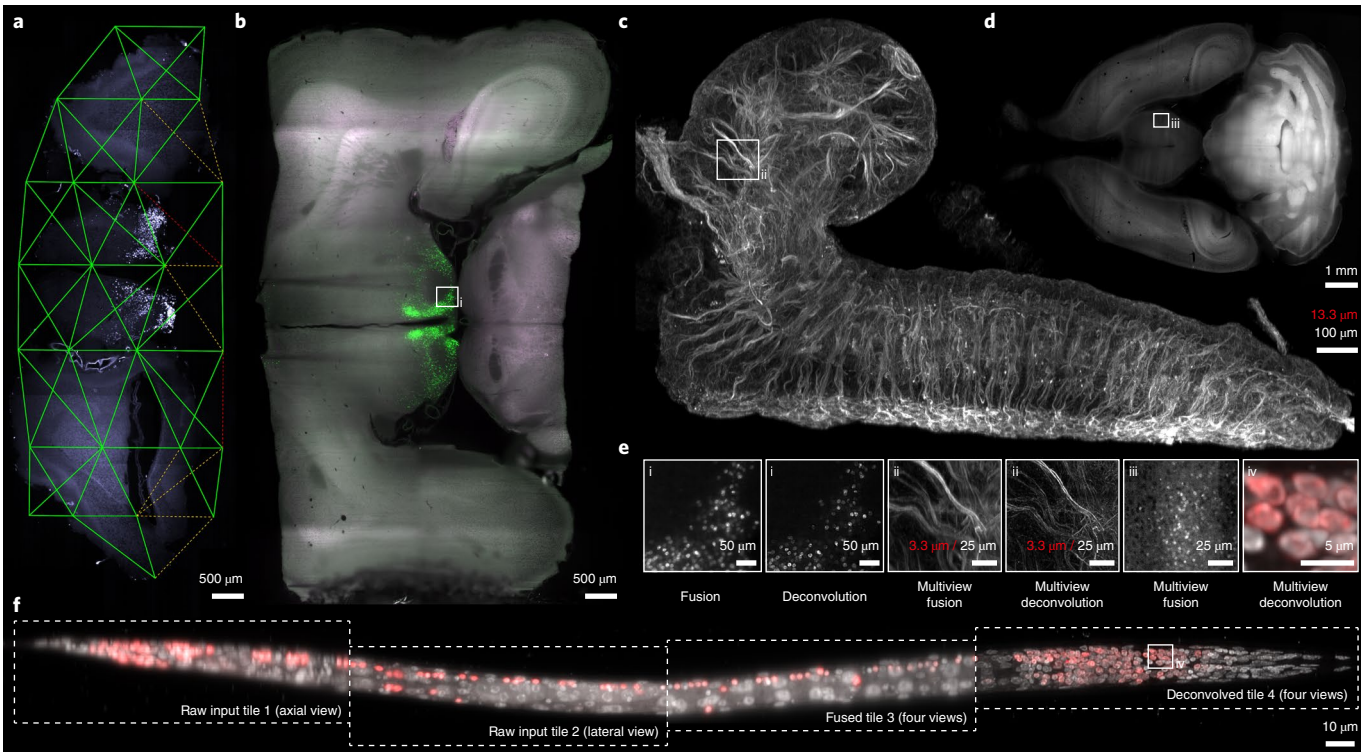


Fig. 3 | Reconstructed samples. a, One slice through an acquisition of an adult mouse Bsx^{H2BeGFP} coronal slice encompassing the hypothalamus (Supplementary Video 3). Green lines indicate strong links between overlapping image tiles, dotted orange lines indicate links rejected because of low correlation and red lines indicate links that were determined to be unconcise. **b**, One slice through an adult Bsx^{H2BeGFP} mouse brain. **c**, Maximum-intensity projection of the central part of the 7.5-fold expanded central nervous system from a *Drosophila* first instar larva with immunostaining for tubulin (Alexa Fluor 88) imaged with multitile IsoView light-sheet microscopy (Supplementary Videos 4 and 5), red indicates scale before expansion **d**, One slice through a whole multitile, multiview reconstructed adult Bsx$^{H2BeGFP/+}$ mouse brain (Supplementary Video 6). **e**, Expanded views of specific areas from **b**–**d** and **f** (labeled i–iv) illustrating (sub)cellular resolution and the advantage of (multiview) deconvolution over (multiview) fusion. **f**, A fixed *Caenorhabditis elegans* dauer larva acquired in four tiles with four views each; tagRFP is expressed in the nuclei of all neurons, which are co-stained with DAPI (Supplementary Video 7). Boxes indicate the quality of axial and lateral raw input data, multiview fusion and multiview deconvolution. In **a**–**f**, a total of eight different datasets were acquired for this publication (Supplementary Table 1), which were all reconstructed as described in the Methods. Additionally, the BigStitcher reconstruction pipeline has been applied to more than 50 samples in our lab (data not shown).

sheet is out of focus, blurred image data, which cannot be reconstructed using BigStitcher, are acquired. Such first-order defocusing can, however, be minimized by employing autofocusing during the acquisition process[20,21].

Finally, as emitted light is distorted by the sample, maximum imaging depth is limited. To overcome this problem, we acquired samples from opposing directions by rotation (Fig. 1c) or by simultaneous acquisition with two objectives[4]. We developed an optimized method for registration of large multitile views, where each view consists of a set of aligned image tiles from one physical orientation (Supplementary Note 8). This method robustly aligns large volumes using affine transformations, effectively doubling the imaging depth of any sample (Fig. 1n and Supplementary Table 2). Subsequently, applying ICP-based non-rigid, split-affine or affine registration allows precise multiview alignment that accounts for additional light refraction in the excitation light path. Using example data, we quantified theoretically possible and practically achievable registration performance (Fig. 2c,d,k, Supplementary Figs. 3, 4 and 17, and Supplementary Note 9), which illustrates that translations alone are not sufficient to achieve high-quality image reconstructions.

As image quality is not constant across the sample, it needs to be quantified to ensure that every part of the reconstructed dataset was acquired with high quality. However, manual inspection at the highest resolution for the entire sample is impossible owing to its size. Therefore, we developed relative Fourier ring correlation (rFRC), which is based on Fourier ring correlation (FRC)[22]. rFRC is able to automatically and rapidly estimate image quality throughout terabyte-sized light-sheet acquisitions while accounting for common scientific complementary metal oxide semiconductor (sCMOS) camera patterns (Supplementary Figs. 1, 7 and 8; Methods).

For downstream analysis, datasets can be fused or directly analyzed using BigDataViewer plugins. We implemented a new algorithm for real-time fusion by multithreaded processing of the currently visible plane in virtual images using blockwise multiresolution loading (Methods), which can optionally be performed with downsampling and on regions of the sample (Supplementary Fig. 18) while supporting brightness equalization. It enables fusion of terabyte-sized images on machines with little memory (Supplementary Fig. 19), while increased memory and compute power enable faster processing (Supplementary Table 2).

Deconvolution is an established method to increase contrast and resolution in light microscopy acquisitions, and required point spread functions (PSF) are typically estimated using fluorescent beads[23]. To handle multitile, multiview acquisitions, we extended deconvolution code[23] allowing BigStitcher to deconvolve selected regions and improve image quality (Fig. 3b–f and Supplementary Note 10). To estimate required PSFs in cleared samples, we developed a new protocol for embedding fluorescent beads in polymerization solution (Fig. 3b,e and Supplementary Note 11). Furthermore,

we combined ExM with IsoView light-sheet microscopy[4] allowing acquisition of multiview, multitile datasets of expanded tissues, which in turn enabled reconstruction of entire *Drosophila* larval nervous systems with a spatially isotropic subcellular resolution of ~100 nm (Fig. 3c,e). Image acquisition took 10 min as compared to the over 2 d that is required to image an expanded sample of similar volume using lattice light-sheet microscopy[24], although the acquisition was at lower resolution (Methods).

BigStitcher enables efficient and automatic processing of terabyte-sized datasets and addresses major unsolved issues such as easy import, management of large images, datasets acquired in a non-regular grid, globally optimal alignment of sparse datasets, illumination selection, rigid and non-rigid multiview alignment of multitile acquisitions, PSF extraction, quality estimation and interactive fusion. The aligned dataset and intermediate steps are interactively displayed, enabling the user to verify and interact with the alignment process to confirm and potentially guide proper alignment of complicated datasets (Supplementary Figs. 5, 16, 20 and 21). Automatic reconstruction of even large datasets can be achieved within tens of minutes, and BigStitcher outperforms existing software in terms of functionality, user-interaction and performance[12,13,16,24] (Supplementary Table 2; Methods). BigStitcher supports cleared samples (Fig. 3a,b,d), ExM samples (Fig. 3c,e and Supplementary Fig. 22), and standard 2D and 3D confocal and wide-field acquisitions, as well as tiled multiview light-sheet acquisitions (Fig. 3f). BigStitcher is open-source, implemented in ImgLib2 (ref. [10]) and provided as a Fiji[25] plugin with comprehensive documentation (https://imagej.net/BigStitcher). Most of its functionality is compatible with the ImageJ macro language (Methods) and can thus easily be automated. These properties make BigStitcher a powerful and scalable tool for the handling and reconstruction of tiled high-resolution image datasets acquired by new light microscopy technologies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41592-019-0501-0.

Received: 18 June 2018; Accepted: 25 June 2019;
Published online: 05 August 2019

References

1. Chung, K. et al. *Nature* **497**, 332–337 (2013).
2. Chen, F., Tillberg, P. W. & Boyden, E. S. *Science* **347**, 543–548 (2015).
3. Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J. & Stelzer, E. H. K. *Science* **305**, 1007–1009 (2004).
4. Chhetri, R. K. et al. *Nat. Methods* **12**, 1171–1178 (2015).
5. Huisken, J. & Stainier, D. Y. R. *Opt. Lett.* **32**, 2608–2610 (2007).
6. Tomer, R., Ye, L., Hsueh, B. & Deisseroth, K. *Nat. Protoc.* **9**, 1682–1697 (2014).
7. Richardson, D. S. & Lichtman, J. W. *Cell* **162**, 246–257 (2015).
8. Pietzsch, T., Saalfeld, S., Preibisch, S. & Tomancak, P. *Nat. Methods* **12**, 481–483 (2015).
9. Linkert, M. et al. *J. Cell Biol.* **189**, 777–782 (2010).
10. Pietzsch, T., Preibisch, S., Tomancák, P. & Saalfeld, S. *Bioinformatics* **28**, 3009–3011 (2012).
11. Kuglin, C. D. & Hines, D. C. in *Proc. IEEE International Conference on Cybernetics and Society* 163–165 (IEEE, 1975)
12. Preibisch, S., Saalfeld, S. & Tomancák, P. *Bioinformatics* **25**, 1463–1465 (2009).
13. Emmenlauer, M. et al. *J. Microsc.* **233**, 42–60 (2009).
14. Chalfoun, J. et al. *Sci. Rep.* **7**, 4988 (2017).
15. Saalfeld, S., Fetter, R., Cardona, A. & Tomancák, P. *Nat. Methods* **9**, 717–720 (2012).
16. Bria, A. & Ianello, G. *BMC Bioinformatics* **13**, 316 (2012).
17. Besl, P. J. & McKay, N. D. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 239–256 (1992).
18. Migliori, B. et al. *BMC Biol.* **16**, 57 (2018).
19. Schaefer, S., McPhail, T. & Warren, J. *ACM Trans. Graph.* **25**, 533–540 (2006).
20. Ryan, D. P. et al. *Nat. Commun.* **8**, 612 (2017).
21. Royer, L. A. et al. Adaptive light-sheet microscopy for long-term, high-resolution imaging in living organisms. *Nat. Biotechnol.* **34**, 1267–1278 (2016).
22. Nieuwenhuizen, R. P. J. et al. *Nat. Methods* **10**, 557–562 (2013).
23. Preibisch, S. et al. *Nat. Methods* **11**, 645–648 (2014).
24. Gao, R. et al. *Science* **363**, eaau8302 (2019).
25. Schindelin, J. et al. *Nat. Methods* **9**, 676–682 (2012).

Acknowledgements

We thank T. Pietzsch and S. Saalfeld for insightful discussions and BigDataViewer and ImgLib2 support; N. Vladimirov for very helpful microscopy discussions; C. Rueden for Fiji support and maintenance; N. Gompel for early-stage project discussions; and the *Caenorhabditis* Genetics Center at the University of Minnesota for providing *C. elegans* strains. S.P., F.P. and M.T. were funded by MDC Berlin; S.P. was supported by HFSP grant RGP0021/2018-102; F.P. was funded by a PhD fellowship from Studienstiftung des deutschen Volkes; F.R.R. and M.T. were funded by the Helmholtz Alliances ICEMED and AMPro; D.H., H.H. and H.L. were funded by the Deutsche Forschungsgemeinschaft (DFG, Nanosystems Initiative Munich), the NHGRI/NIH Center for Photogenomics (grant RM1 HG007743) and LMU Munich; and P.T., N.R., R.K.C., A.C. and P.J.K. were funded by HHMI Janelia.

Author contributions

S.P. conceived the idea in discussions with H.H., H.L. and M.T.; D.H. and S.P. developed the algorithms and implemented the software; F.R.R. performed all clearing experiments, reconstructions and benchmarks; F.P. imaged and reconstructed *C. elegans*; P.T. and N.R. performed ExM sample preparation; R.K.C. and P.J.K. developed the ExM-optimized IsoView microscope and imaged the sample; S.P. reconstructed the ExM sample; S.P., M.T., H.L., H.H., P.J.K. and A.C. supported and supervised the project; and S.P., D.H. and F.R.R. wrote the manuscript with input from the co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-019-0501-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.P.

Peer review information: Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Animals. For clearing, we used a previously generated $Bsx^{H2BeGFP}$ mouse line[26], in which exon 1 of the *Bsx* gene is replaced (starting at the ATG) with a sequence encoding a histone 2B–enhanced green fluorescent protein (eGFP) fusion protein. Brains from 10-week-old female $Bsx^{H2BeGFP/+}$ mice were used for tissue clearing and imaging. *C. elegans* dauer larvae expressing the tagRFP fluorescent protein fused to a nuclear localization sequence in all neurons[27] (strain AML32) were obtained by selecting dauer larvae in 1% SDS for 30 min (ref. [28]). Dauer larvae were fi ed with 4% paraformaldehyde for 30 min on ice, placed in 70% ethanol overnight at 4 °C and subsequently stained with DAPI. Experiments were conducted according to the institutional guidelines of the Max Delbrück Center for Molecular Medicine in the Helmholtz Association after approval from the Berlin State Office or Health and Social Aff irs (LAGeSo, Landesamt für Gesundheit und Soziales). *Drosophila* larvae used for ExM were obtained from the strain w;;attp2, carrying an empty attp2 landing site[29].

Clearing and expansion. Tissue clearing was performed using the CLARITY protocol[1] (Supplementary Note 1). Before imaging, the tissue samples were placed overnight in FocusClear for refractive index matching. For ExM, the nervous system of a first-instar *Drosophila* larva was extracted, fixed and stained with anti-tubulin antibodies (Supplementary Note 2). The stained sample was washed in 1× PBS and then processed using a modified ExM method to achieve 7.5-fold expansion in each dimension (Supplementary Note 2). In summary, the specimen was treated with acryloyl-X[30] as in standard ExM and embedded using a gel recipe modified from the original method[2]. The modified recipe uses a reduced cross-linker concentration to achieve greater expansion. After digestion with proteinase K, a new re-embedding step toughens up the gel, which would otherwise have poor mechanical properties.

Imaging. 3D images of cleared mouse brains placed in a FocusClear-filled imaging chamber were acquired using the Zeiss light-sheet Z.1 microscope. Fixed *C. elegans* dauer larvae were embedded in agarose that contains fluorescent beads, and imaged using the same microscope in a water-filled sample chamber. 3D images of the cleared and expanded central nervous system of a *Drosophila* first-instar larva were acquired using an IsoView light-sheet microscope[4] that has been modified for multitile acquisition. The ability of the IsoView microscope to rapidly record large specimens allowed us to image the entire 7.5-fold expanded nervous system ($2{,}040 \times 1{,}108 \times 1{,}201\ \mu m = 2.7\ mm^3$) in 10 min with an isotropic resolution of approximately 100 nm, acquiring two volumes that were oriented orthogonally to each other each with a spatial sampling of $55 \times 55 \times 110$ nm (unexpanded volume). In comparison, acquisition of a 4.09-fold expanded volume of half the size ($1{,}400 \times 2{,}700 \times 370\ \mu m = 1.4\ mm^3$) with two channels using lattice light-sheet microscopy[24] took 2.61 d; however, the image acquired using lattice light-sheet microscopy had a higher spatial sampling of $23.7 \times 23.7 \times 44.0$ nm (unexpanded volume). Details of the imaging strategies are described in Supplementary Note 3 and a summary of the most important acquisition parameters can be found in Supplementary Table 1.

Data processing pipeline. All data shown in this paper were processed using the BigStitcher Fiji[25] plugin. Zeiss CZI files and TIFF files exported by custom microscopes were imported using AutoLoader and subsequently converted to HDF5. For Zeiss CZI files, approximate tile positions and rotation angles were imported automatically; for other files, these parameters were specified by hand using BigStitcher tools (Supplementary Figs. 5 and 21). If not stated otherwise, reconstruction was performed using the following steps. For each tile, the best illumination was selected. Tiles were aligned using the phase correlation method using two-round global optimization followed by ICP[17] refinement on an affine model. Interest-point detection was performed for each multitile view. Either our extension of the fast descriptor-based rotation-invariant algorithm[31] or the descriptor-based translation-invariant algorithm[32], which was used after application of manual rotation, was used to register the interest points of each angle, and another round of ICP refinement was performed on all image tiles of the acquisition. Fused and deconvolved images were exported as TIFF files.

Pairwise stitching using Fourier-based phase correlation. We calculated pairwise translational shifts using our ImgLib2 (ref. [10]) implementation of the Fourier-based phase correlation algorithm[11]. The processing time was substantially reduced, while our simulations showed that, at the same time, registration errors below 1 pixel were achieved by computing the phase correlation on downsampled images and performing subpixel localization of the shift vector[33] (Supplementary Notes 6 and 7, and Supplementary Figs. 11–15).

Global optimization. To calculate the final transformations of each image tile we extended the concept of globally optimal registration by iterative minimization of square displacement of point correspondences[12,15,31,32] (Supplementary Note 12). Erroneous pairwise links that might not have been filtered out before global optimization (for example, wrong links caused by average cross-correlation, repetitive patterns or a low number of correspondences in the ICP refinement) lead to high registration errors after global optimization. This manifests in a large distance error, which is the difference between the individually computed distance between images (link) and the actual distance between images after global optimization. Iterative removal of the link with the highest distance error from the link graph and repetition of the global optimization lead to convergence to user-defined thresholds[12]. We extended this concept from Preibisch et al.[12] to affine transformations, introduced a new heuristic that additionally incorporates link quality, and implemented it in an extendable framework required for the two-round global optimization (Supplementary Note 12). If the dataset contains empty tiles or even consists of multiple disconnected objects, the final transformations are not propagated between them (Supplementary Fig. 10). We therefore developed a two-round global optimization that is capable of aligning independent connected components of the link graph using weak links defined by the current transformations of each tile (for example, approximate locations from metadata or manual alignments), which optimally preserves distances between neighboring objects (Supplementary Note 12 and Supplementary Fig. 10).

Our global optimization is agnostic to the nature of the point correspondences and transformation model, which allows us to use the same algorithm for translation-based alignment of tiled datasets using phase correlation, as well as affine registrations of multitile, multiview datasets based on ICP refinement or geometric descriptor matching.

Iterative closest point refinement. Although the phase-correlation-based image stitching produces relatively high-quality alignments, smaller errors can remain (Supplementary Figs. 13–15). Furthermore, this method is not able to correct for non-translational effects such as chromatic and spherical aberration or sample-induced light refraction (Fig. 2 and Supplementary Figs. 3, 4, 9 and 17). These effects can be better approximated using affine transformations. We therefore automatically detect interest points and run an ICP algorithm[17] for each overlapping pair of images, where the assignment of correspondences is limited by a distance threshold. We used the identified corresponding points of all pairwise links and computed a globally optimal affine transformation for each tile using our new global optimization algorithm. To avoid scaling of datasets, we regularized the affine transformation using a rigid transformation[15]. The resulting alignment usually improved the alignment quality and the same strategy could be applied to multichannel alignment if the same autofluorescent structures were visible in multiple channels (Supplementary Fig. 3). However, only small chromatic aberrations within a few pixels could be corrected by approximations based on affine or split-affine transformations (Supplementary Fig. 3). At the same time, non-rigid transformations can easily be unstable for this purpose as correspondences between different channels would have to be distributed over the entire image. Therefore, in those cases, it is better to use specialized software such as that described by Matsuda et al.[34] to correct for chromatic aberrations before importing data into BigStitcher.

Illumination selection. When imaging large samples using sequential dual-sided illumination[5], typically only illumination from one direction provided good image quality (Supplementary Fig. 2). We therefore implemented an illumination selection functionality in BigStitcher. We first combined all (selected) images by their illumination attribute and thereby grouped all images that shared other attributes besides illumination direction. In each of the resulting groups, we selected the best image. As quality criteria, we offered either the rFRC on full-resolution images or a fast approximation obtained using mean intensity or mean gradient magnitude at the lowest resolution level. While rFRC provided the highest distinctive power, both fast approximations were typically sufficient for robust estimation of the higher-quality illumination direction (Supplementary Fig. 2 and Fig. 1d). The image with the highest quality score was kept, while all other images were marked as missing in the dataset, which led to them being ignored in subsequent processing steps. However, before applying automatic illumination estimation, the user has the option to verify and potentially change the result. Optional resaving of the dataset after this step potentially decreases the storage requirement twofold.

Simulation of light propagation in tissue using ray tracing. We observed non-rigid deformations occurring in areas where image data from opposing light sheets were recorded by the same camera (Fig. 2a). While it is intuitive and clear from existing simulations of light propagation[35] that imaging of the same light sheet from opposing objectives can lead to non-rigid deformations caused by different aberrations in the detection light paths, we wanted to understand how changing just the light-sheet direction can introduce non-rigid deformations. These effects were clearly visible in large samples like entire mouse brains (Fig. 2a), which are still beyond the range of simulation using reasonable efforts. We therefore developed a simple ray-tracing-based method to simulate light propagation, which aimed to recapitulate these observations at a smaller and more manageable scale (Supplementary Note 13).

Although quite simple, this simulation recapitulated the effects observed in cleared images (Fig. 2a,h) and illustrated that different refraction of the illumination light sheets alone can lead to non-rigid deformations in the acquired image stacks as it lead to illumination of different contents of the sample in the z axis (Fig. 2g,h and Supplementary Video 2). As detection objectives with relatively

low numerical aperture (NA; and therefore large depth of field) are typically used for detection (Supplementary Table 1), both light sheets can still appear in focus despite being tens of micrometers away from each other. Therefore, these transformations need to be corrected for.

Non-rigid transformation. To be able to compensate for potentially strong aberrations introduced by light refraction due to the sample, we implemented a non-rigid alignment method. It is based on the concept of moving least squares[19] that has been shown to perform well in biomedical applications[15]. We implicitly regularized moving least squares using ICP[17] or random sample consensus[36] (Supplementary Note 14). Moving least squares requires computation of a transformation for each pixel, which is computationally expensive. We therefore implemented a virtual cached layer that only computes a transformation for every m^{th} pixel (m being the distance between pixels for which transfomations are computed) and linearly interpolates affine transformations for pixels in between. As BigDataViewer currently only supports affine transformations, we additionally implemented a multiresolution preview based on virtually fused non-rigid volumes that can be interactively displayed in an extra BigDataViewer window. BigStitcher also supports 'hybrid' fusion of non-rigid- and affine-transformed image tiles as non-rigid registration requires substantially increased computational effort (Supplementary Fig. 17).

Virtual reblocking. To allow piecewise affine transformations or a more refined illumination selection, we developed virtual reblocking of all 3D image stacks of an acquisition (Fig. 2l,m). The implementation distributes 3D blocks onto each input image stack using a defined overlap and thereby defines a new set of input image stacks for the acquisition. The new virtual image stacks are computed on the fly for all resolution levels from the original image data. Any ImgLoader is supported, including multiresolution image stacks. Optional resaving of the dataset as HDF5 or TIFF transforms the virtual image stacks into physical representations.

Quality estimation on the basis of Fourier ring correlation. To estimate image quality across entire terabyte-sized acquisitions, we developed an extension of the FRC[22] that is robust and insensitive to camera noise. $\mathrm{FRC}_{r_1,r_2}(f)$ constitutes the per-spatial-frequency (f) correlation between two independent realizations, r_1 and r_2, of the same signal. In localization-based super-resolution, point clouds are therefore typically split into two sets of independent pixels. Here we use consecutive slices instead and take advantage of the fact that they are nearly identical owing to the axial extent of the PSF. As a result, $\mathrm{FRC}_{r_1,r_2}(f)$ constitutes an entire correlation spectrum for each z plane, and we compute a single quality value $Q(z)$ by integration over all frequencies f

$$Q_{\mathrm{FRC}}(z) = \int_f \mathrm{FRC}_{z,z+1}(f)$$

A smoother symmetric result can be obtained by averaging the FRC spectra obtained using z planes above and below the measured plane

$$Q_{\mathrm{FRC}}(z) = \int_f \frac{\mathrm{FRC}_{z,z+1}(f) + \mathrm{FRC}_{z,z-1}(f)}{2}$$

For computing the 2D FRC, we adapted methods from the FRC ImageJ plugin (Supplementary Note 15). This results in a precise estimation of image quality, except if patterned noise (for example, sCMOS camera noise) is the dominant source of signal (Supplementary Fig. 1). To overcome this instability, we developed the rFRC, which subtracts a locally estimated scatterplot smoothing (loess)-smoothed[37] baseline FRC of z planes spaced by m slices that are beyond the axial extent of the PSF

$$Q_{\mathrm{rFRC}}(z) = \int_f \frac{\mathrm{FRC}_{z,z+1}(f) + \mathrm{FRC}_{z,z-1}(f)}{2} - \mathrm{loess}\left(\mathrm{FRC}_{z-m,z+m}(f)\right)$$

This effectively measures which additional frequencies the central planes z, $z+1$ and $z-1$, z have in common, as compared to the planes $z-m$, $z+m$ that are beyond the axial size of the PSF. The resulting values $Q_{\mathrm{rFRC}}(z)$ robustly measure image quality in the sample (Supplementary Figs. 1g, 7 and 8). As image content can change drastically within a slice, we support computation using a manually defined block-size (for example, 512×512) and with z stepping (for example, every 20 planes). To estimate the quality metric for entire acquisitions, we compute $Q_{\mathrm{rFRC}}(z)$ for defined points in each image stack. Over all input stacks, these measurements are held as sparse representations using ImgLib2 (ref. [10]) that can be rendered virtually and overlaid onto entire fused volumes (Supplementary Figs. 7 and 8, and Supplementary Video 8 and 9).

Virtual image fusion. A set of overlapping transformed image tiles are fused into one output image using a per-pixel weighted average that minimizes boundary artifacts and can increase contrast by incorporating entropy estimation[38] (Supplementary Note 16). To correct for unequal brightness and contrast in adjacent images, we optionally perform adjustment of the pixel intensities using a linear transformation for each image. An optimal brightness and contrast adjustment can be estimated using the same optimization framework used for image registration[39] (Supplementary Note 17). The memory requirements for fusion of large volumes can easily exceed the available random access memory (RAM) on a machine owing to the size of the output and the combined size of the input images. We, therefore, developed a framework for intensity transformations and coordinate transformations that is based on ImgLib2 (ref. [10]) and virtually fused all pixels of a defined bounding box using all input images and their associated weights. As the input images are provided through virtual image loading, the size of a virtually fused image is close to zero, irrespective of the size of input and output images. Ideally, input images are available in blocks so that affine transformations that slice the image stacks in arbitrary orientations do not require loading of the entire image[8]. The output image can now be rendered on a pixel-by-pixel basis with minimal memory requirements. Additional caching of the input image and the output images allows an efficient multithreaded fusion for optimally fast processing given the available memory. Therefore, more RAM will effectively speed up the fusion process (Supplementary Table 2 and Supplementary Fig. 17), but even machines with very low RAM are able to fuse terabyte-sized volumes (Supplementary Fig. 19). Fused images can be saved by choosing cached or virtual fusion and subsequently saving the ImageJ virtual stack using 'Save as image sequence…'.

Downsampling of the output can easily be incorporated by scaling the bounding box and preconcatenation of the downsampling transformation with each image transformation. If the input files are multiresolution, we automatically compute the optimal resolution level at which the input needs to be loaded. To optionally further reduce the image size of the fused image, the graphical user interface offers to conserve the original anisotropy between lateral and axial orientations of the acquired sample, which is a sensible choice if the dataset contains a single view or opposing (for example, 0° and 180°) multitile views.

Macro automation and headless operation. In addition to the graphical user interface, we offer standalone Fiji plugins for most of the individual steps, such as data import, illumination selection, pairwise shift calculation, link filtering, multiview alignment, global optimization, and image fusion and deconvolution. In macro mode results will not be displayed interactively but are instead saved to the XML project file or output files immediately. The individual steps can be recorded as ImageJ macros and easily combined into a script for headless batch processing[40].

Limitations of the framework and other software solutions. BigStitcher is designed for the reconstruction of large multitile, dual-illumination, multiview datasets, and supports affine, split-affine and non-rigid registrations to solve the alignment process for terabyte-sized image data. Several solutions based on image correlation support multitile-only data, such as Terastitcher[16], XUVTools[13] and ImageJ Stitching[11]; however, these solutions only support translation models, making them unsuitable for the types of datasets described above (Fig. 2 and Supplementary Figs. 3, 4, 9 and 17). A recent stitching solution developed by the Saalfeld lab that is also based on ImgLib2 (ref. [10]) can handle even larger datasets (up to hundreds of terabytes) and supports affine transformations based on local cross-correlation[24]. It is, however, also focused on multitile-only acquisitions, is designed to run on a cluster or in the cloud, does not support non-linear tile deformations, and has no user interface to access its functionality. Currently, BigStitcher scales well up to 1,000 large 3D image tiles per time point and image sizes on the order of tens of terabytes per time point. This is, however, mostly due to a limit in the rendering capacity of BigDataViewer. Future optimizations of BigDataViewer[8] and/or BigStitcher will further increase this limit. BigStitcher can correct for chromatic and spherical aberrations by approximation with affine transformations if errors are within a few pixels. For chromatic aberration correction, enough autofluorescent structures must be visible across channels (Supplementary Figs. 3, 4 and 9). Although BigStitcher can correct for geometric transformations introduced by the acquisition process (Fig. 2 and Supplementary Fig. 17), it is not possible to correct for images that are out of focus.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Small example datasets are available for download from the Open Science Foundation at https://osf.io/bufza/. Larger datasets are available on request. Additional datasets uploaded at a later stage will be linked from the documentation page which can be found at https://imagej.net/BigStitcher#Example_Datasets. Example datasets are explained in detail in Supplementary Note 18.

Code availability

All source code used in this publication (BigStitcher, phase correlation simulation and benchmarks, and the simulation of light propagation in tissue using ray tracing) is open-source and published under the GNU General Public License version 2. The latest stable releases used in this publication are provided as

Supplementary Software; current versions that include bugfixes and updates can be downloaded from GitHub (at https://github.com/PreibischLab/BigStitcher; https://github.com/PreibischLab/multiview-reconstruction; and https://github.com/PreibischLab/multiview-simulation; see Supplementary Notes 19 and 20 for further explanations). Details on how to use the software are described in Supplementary Note 21.

References

26. Sakkou, M. et al. *Cell Metab.* **5**, 450–463 (2007).
27. Nguyen, J. P., Linder, A. N., Plummer, G. S., Shaevitz, J. W. & Leifer, A. M. *PLoS Comput. Biol.* **13**, e1005517 (2017).
28. Karp, X. in *WormBook* (ed. The *C. elegans* Research Community, 2016); https://doi.org/10.1895/wormbook.1.180.1
29. Pfeiffer, D. B. et al. *Proc. Natl Acad. Sci. USA* **105**, 9715–9720 (2008).
30. Tillberg, P. W. et al. *Nat. Biotechnol.* **34**, 987–992 (2016).
31. Preibisch, S., Saalfeld, S., Schindelin, J. & Tomancák, P. *Nat. Methods* **7**, 418–419 (2010).
32. Smith, C. S. et al. *J. Cell Biol.* **209**, 609–619 (2015).
33. Lowe, D. G. *Int. J. Comput. Vis.* **60**, 91–110 (2004).
34. Matsuda, A., Schermelleh, L., Hirano, Y., Haraguchi, T. & Hiraoka, Y. *Sci. Rep.* **8**, 7583 (2018).
35. Weigert, M., Subramanian, K., Bundschuh, S. T., Myers, E. W. & Kreysing, M. *PLoS Comput. Biol.* **14**, e1006079 (2018).
36. Fischler, M. A. & Bolles, R. C. *Commun. ACM* **24**, 381–395 (1981).
37. Cleveland, W. S. *J. Am. Stat. Assoc.* **74**, 829–836 (1979).
38. Preibisch, S., Rohlfi g, T., Hasak, M. P. & Tomancák, P. in *Proc. of the International Society for Optics and Photonics, Medical Imaging* (eds. Reinhardt, J. M. & Pluim, J. P. W.) (SPIE, 2008).
39. Blasse, C. et al. *Bioinformatics* **33**, 2563–2569 (2017).
40. Schmied, C., Steinbach, P., Pietzsch, T., Preibisch, S. & Tomancák, P. *Bioinformatics* **32**, 1112–1114 (2016).

nature research

Corresponding author(s): Stephan Preibisch

Last updated by author(s): May 28, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Simulated data was created using the net.preibisch.simulation.SimulateMultiViewAberrations class in the multiview-simulation package (release version 0.2.2). Since it is a Maven artifact, the versions of all dependencies are clearly defined and the corresponding version can be built automatically from that source code state (https://github.com/PreibischLab/multiview-simulation/commit/b41b74cce9287f804b670d7de3396605446818a8).
Data analysis	The data was reconstructed using BigStitcher (release version 0.3.3). Since it is a Maven artifact, the versions of all dependencies (e.g. multiview-reconstruction) are clearly defined and the corresponding version can be built automatically from that source code state (https://github.com/PreibischLab/BigStitcher/commit/0d7f79a59ab15fb1805157ab72c5bc9802b02fbd).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Small example datasets are available online: https://osf.io/bufza/ (Open Science Foundation). Larger datasets are available on request. When we find ways to host larger datasets, they will be linked from the documentation page: https://imagej.net/BigStitcher#Example_Datasets

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size of 8 acquired samples was chosen to show a wide variety of samples and sample preparation that can be reconstructed using BigStitcher. Some samples are quite similar (clearing), some were chosen to highlight the generality (expanded, only fixed).
Data exclusions	No data was excluded.
Replication	In total 8 different samples (cleared, expanded, only fixed) were acquired and reconstructed for this publication. Furthermore, one sample was simulated and reconstructed. Additionally, the BigStitcher reconstruction pipeline has been applied to >50 samples in our lab (not part of the publication, most are prepared in collaboration on other projects), which highlights the generality of the method. The tissue sections shown in Fig. 3a + Suppl. Fig. 1 + Suppl. VIdeo 1 as well as Fig. 3b are very similar and highlight reproducibility.
Randomization	Randomization was used for a) the simulation and benchmarking of phase correlation (Supplementary Fig. 12-15), as well as b) for the raytracing (Fig. 2b-e). For a) random overlaps (uniformly distributed) were simulated 300 times for each condition, and the noise for the image generation process is based on Poisson noise. For b), illumination and detection rays in defined areas are sent randomly (uniformly distributed) into the scene.
Blinding	Blinding is not applicable for this study since there is no process that requires it.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	For clearing we used a previously generated BsxH2BeGFP mouse line, where the exon 1 of the bsx gene is replaced starting at the ATG with the coding sequence for histone2B eGFP. Brains from 10-week old female BsxH2BeGFP/+ mice were used for tissue clearing and imaging. C. elegans dauer larvae expressing tagRFP fused to a nuclear localizing sequence under the pan-neuronal rab-3 promotor in all neuron nuclei were obtained by selecting dauer larvae in 1% SDS for 30 minutes. Drosophila larva used for ExM were obtained from the strain w;;attp2, carrying an empty attp2 landing site.
Wild animals	No wild animals were used.
Field-collected samples	No field-collected samples were used.
Ethics oversight	Experiments were conducted according to institutional guidelines of the Max Delbrück Center for Molecular Medicine in the Helmholtz Association after approval from the Berlin State Office for Health and Social Affairs (LAGeSo, Landesamt für Gesundheit und Soziales, Berlin, Germany)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

2.1.2 Supplementary information

nature methods

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41592-019-0501-0

In the format provided by the authors and unedited.

BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples

David Hörl[1,2,7], Fabio Rojas Rusak[3,7], Friedrich Preusser[2], Paul Tillberg[4], Nadine Randel[4], Raghav K. Chhetri[4], Albert Cardona[4,5], Philipp J. Keller[4], Hartmann Harz[1], Heinrich Leonhardt[1], Mathias Treier[3,6] and Stephan Preibisch[2]*

[1]Department of Biology II, Ludwig-Maximilians-Universität München, Munich, Germany. [2]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. [3]Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany. [4]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. [5]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK. [6]Charité-Universitätsmedizin Berlin, Berlin, Germany. [7]These authors contributed equally: David Hörl, Fabio Rojas Rusak. *e-mail: stephan.preibisch@mdc-berlin.de

BigStitcher: Reconstructing high-resolution image datasets of cleared and expanded samples

Supplementary Figures, Tables, Notes

David Hörl[1,2,#], Fabio Rojas Rusak[3,#], Friedrich Preusser[2], Paul Tillberg[4], Nadine Randel[4], Raghav K. Chhetri[4], Albert Cardona[4,5], Philipp J. Keller[4], Hartmann Harz[1], Heinrich Leonhardt[1], Mathias Treier[3,6], Stephan Preibisch[2,*]

[1]Department of Biology II, Ludwig-Maximilians-Universität München, München, Germany
[2]Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany
[3]Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany
[4]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA
[5]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK
[6]Charité-Universitätsmedizin Berlin, Berlin, Germany
[#]D.H. and F.R.R. contributed equally

* Correspondence should be addressed to S.P. (stephan.preibisch@mdc-berlin.de)

Supplementary File	Title
Supplementary Figure 1	Quantification of fluorescence preservation in cleared tissue
Supplementary Figure 2	Quantification of automatic illumination selection
Supplementary Figure 3	Chromatic aberration correction
Supplementary Figure 4	Spherical aberration correction
Supplementary Figure 5	Manual alignment
Supplementary Figure 6	Flat-field correction
Supplementary Figure 7	Automatic quantification of image quality
Supplementary Figure 8	Quality estimation in whole-brain mouse acquisition
Supplementary Figure 9	Affine refinement via ICP
Supplementary Figure 10	Global optimization
Supplementary Figure 11	Pairwise registration by phase correlation
Supplementary Figure 12	Downsampling with different SNR
Supplementary Figure 13	Downsampling error statistics 1
Supplementary Figure 14	Downsampling error statistics 2
Supplementary Figure 15	Downsampling error statistics 3
Supplementary Figure 16	Interactive inspection and curation of pairwise links
Supplementary Figure 17	Quantification of image registration quality
Supplementary Figure 18	Bounding-box definition
Supplementary Figure 19	Virtual fusion of large image
Supplementary Figure 20	Interest point visualization
Supplementary Figure 21	Manual transformation of multi-view datasets
Supplementary Figure 22	Expansion microscopy reconstruction
Supplementary Figure 23	Principles of non-rigid alignment
Supplementary Table 1	Datasets used in this publication
Supplementary Table 2	Performance comparison
Supplementary Note 1	Sample clearing
Supplementary Note 2	Expansion microscopy

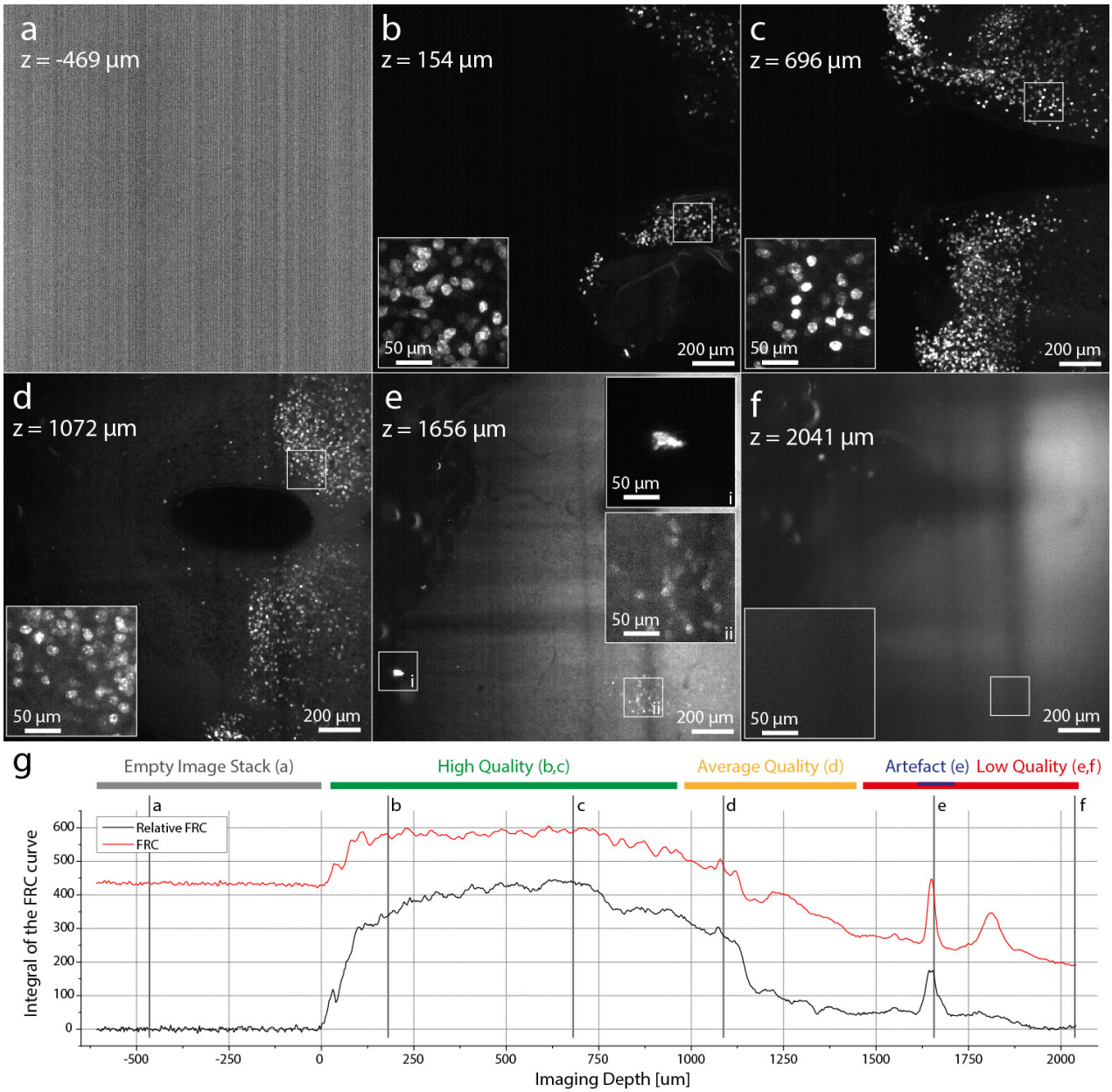
Continued on the following page

Supplementary File	Title
Supplementary Note 3	Imaging strategies
Supplementary Note 4	Data import
Supplementary Note 5	Flat-field correction
Supplementary Note 6	Pairwise shift calculation
Supplementary Note 7	Quantification of pairwise image stitching
Supplementary Note 8	MultiView Registration
Supplementary Note 9	Quantification of overall registration quality
Supplementary Note 10	Multiview deconvolution
Supplementary Note 11	PSF measurement and PSF extraction
Supplementary Note 12	Global optimization
Supplementary Note 13	Simulation of light propagation in tissue using raytracing
Supplementary Note 14	Non-rigid transformation
Supplementary Note 15	Quality estimation using relative Fourier Ring Correlation
Supplementary Note 16	Image fusion
Supplementary Note 17	Brightness and contrast adjustment
Supplementary Note 18	Example datasets for BigStitcher
Supplementary Note 19	Links to the current source codes
Supplementary Note 20	Bug reports and feature requests
Supplementary Note 21	BigStitcher user guide

Note: Supplementary Videos 1–9 are available for download on the journal homepage.

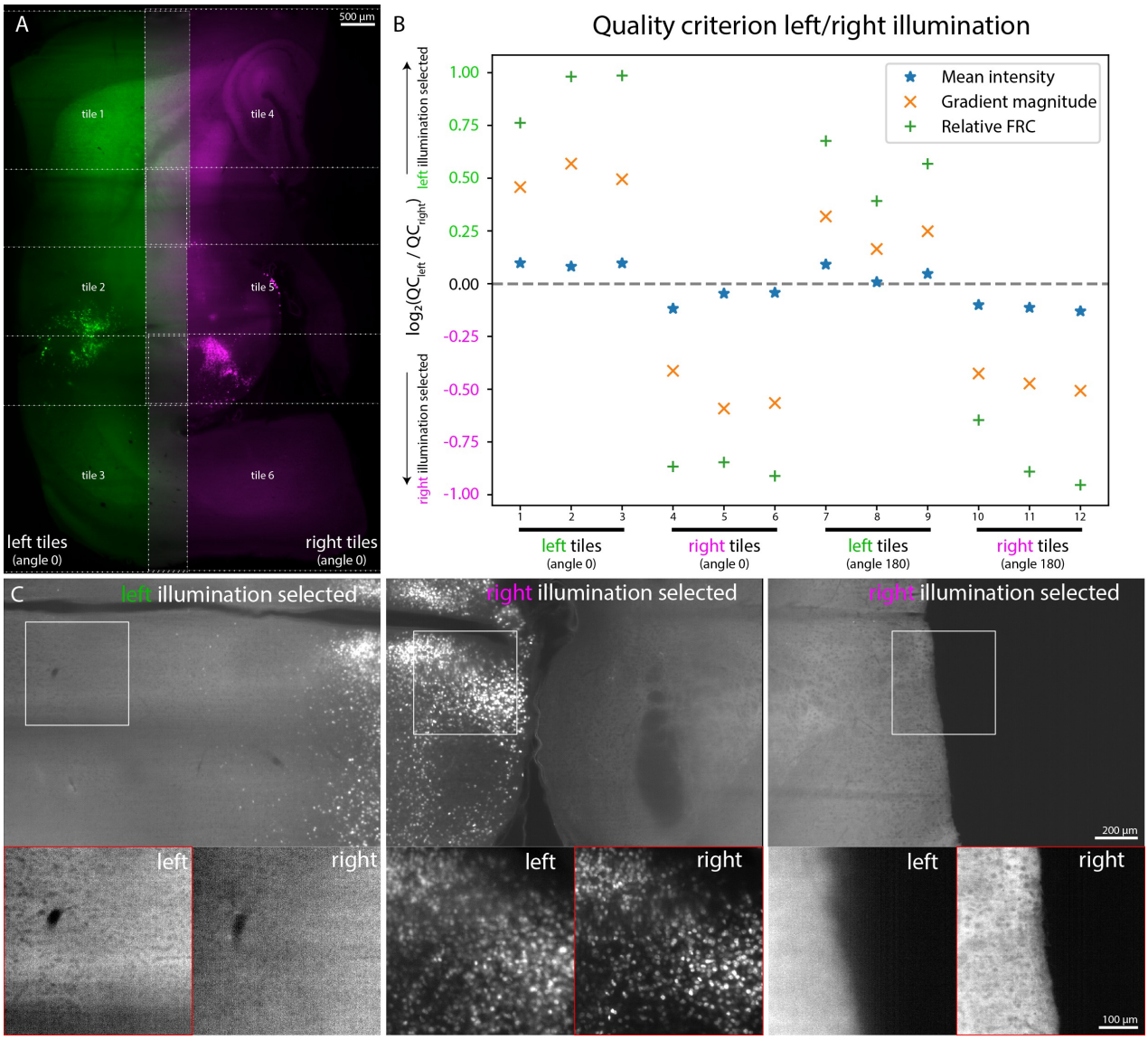
SUPPLEMENTARY FIGURES

SUPPLEMENTARY FIGURE 1: Quantification of fluorescence preservation in cleared tissue



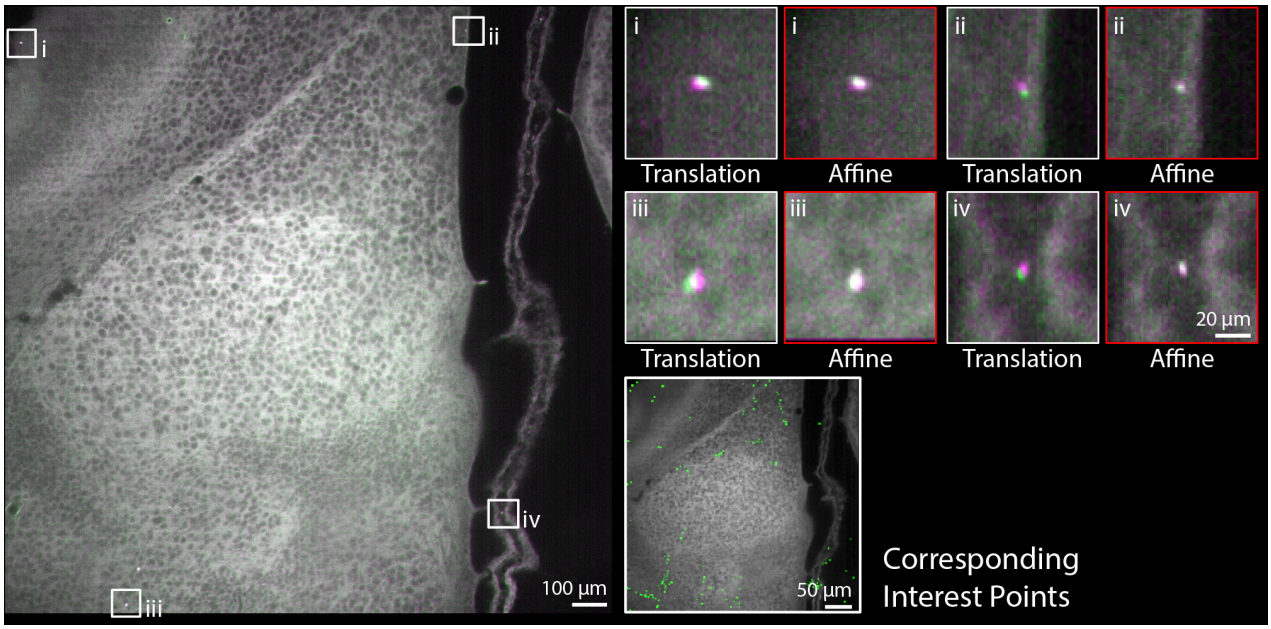
Supplementary Figure 1: *Quantification of fluorescence preservation in cleared tissue.* **(a-f)** Optical sections through an CLARITY-cleared adult mouse hypothalamus expressing H2B-GFP in all bsx neurons. Fluorescence is preserved throughout the clearing procedure. However, the signal is degrading with imaging depth and can typically be recorded up to 1 – 2 cm into the sample, depending on the tissue type and the quality of the clearing process limiting the size of the sample that can be acquired from a single orientation. Brightness and contrast was adjusted individually. **(g)** Quantification of image quality using (relative) Fourier Ring Correlation ([r]FRC, see **Online Methods**) in BigStitcher. Note that FRC produces high values for the camera patterns if no signal is present. The rFRC accurately measures image quality as illustrated by the position of the panels (a-f). As part of this publication similar experiments were performed 4× with comparable clearing results (**Fig. 1n, 3b, 3d**).

SUPPLEMENTARY FIGURE 2: Quantification of automatic illumination selection



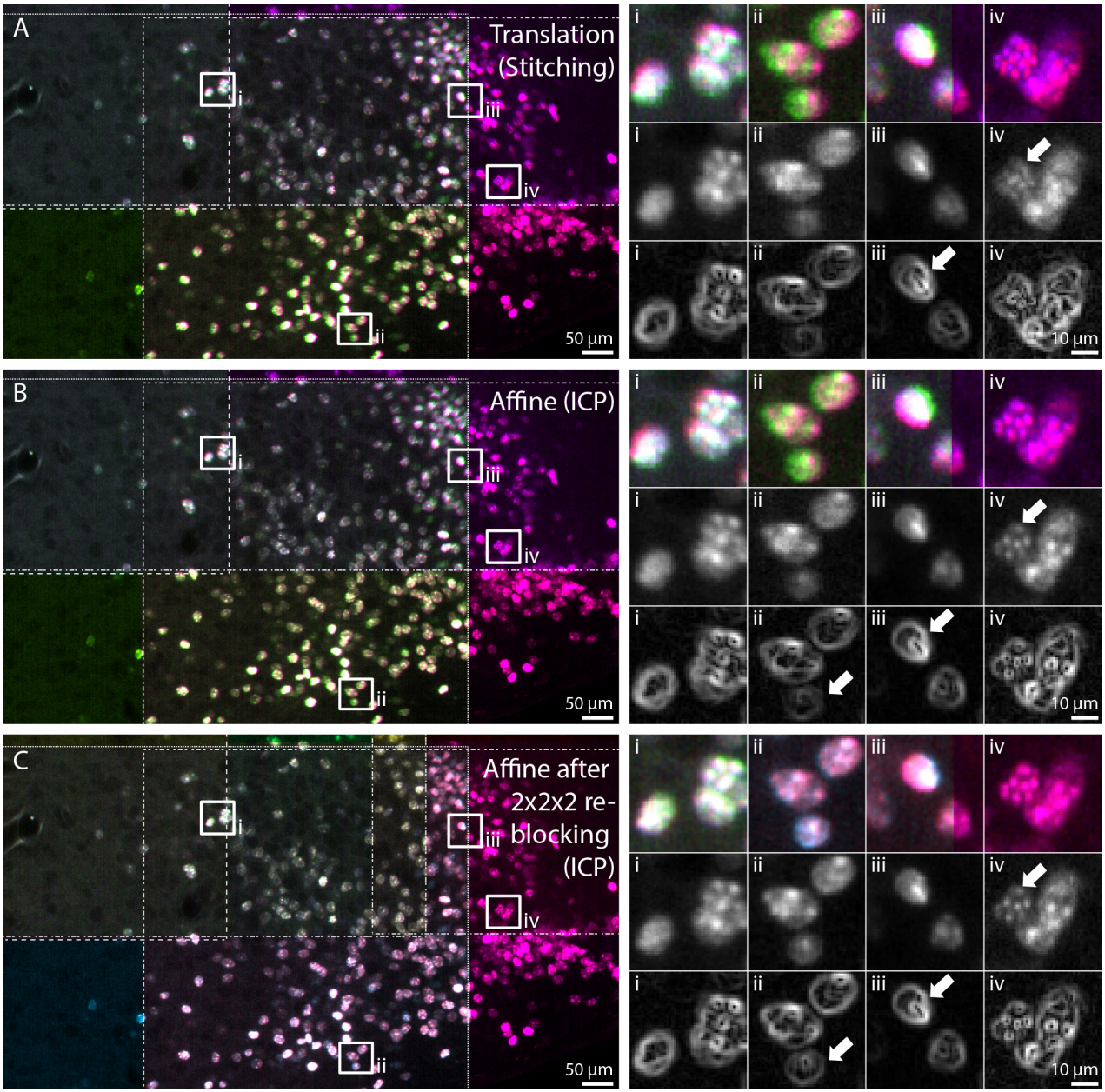
Supplementary Figure 2: *Automatic illumination selection.* **(A)** A small 166GB multi-view, dual-illumination, multi-tile dataset specifically acquired for verification purposes, here to quantify the properties of automatic illumination selection. Shown is a slice through the first of two angles, where six tiles and two illumination directions are highlighted. We manually confirmed that the left three tiles need to be assigned to left illumination, and the right three tiles to right illumination. **(B)** compares the distinction power of the methods *Mean intensity*, *Gradient magnitude*, and *Relative Fourier Ring Correlation* (see **Online Methods**) by their respective quality scores. All methods correctly predict the assignment, while the *Relative Fourier Ring Correlation* distinguishes the illumination directions best. Note that *Mean intensity* almost produces an error for tile 8 of the second angle (180 degrees) **(C)** Another example of best illumination for three consecutive tiles (left to right), selected based on *Mean intensity* for each tile. Close-ups shows the specified region for both illumination directions. **(A,C)** As part of this publication automatic illumination selection was performed on 4 datasets (see also **Fig. 1d, 3b, 3d**).

SUPPLEMENTARY FIGURE 3: Chromatic aberration correction



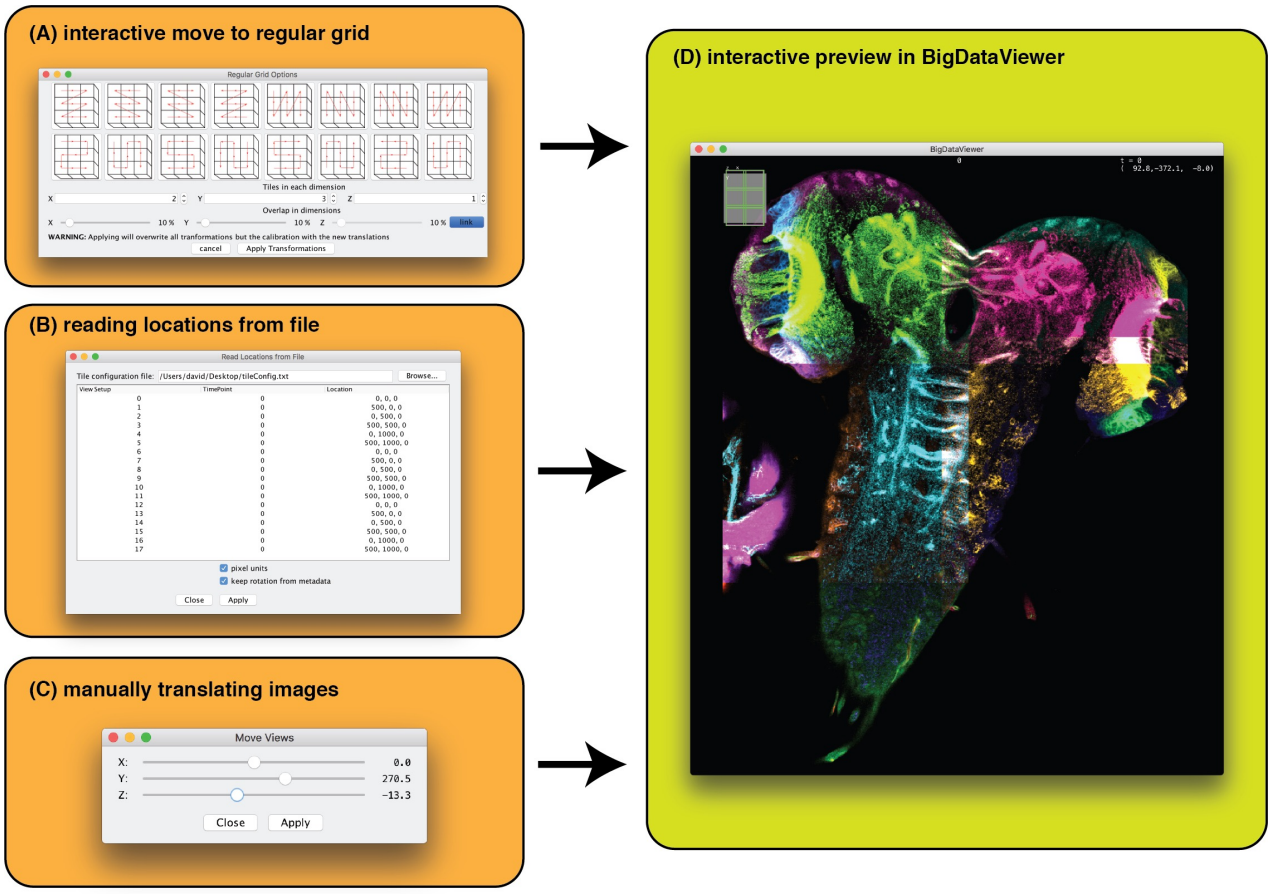
Supplementary Figure 3: *Chromatic aberration correction.* If sufficient autofluorescent signal is in common between channels the ICP refinement on an affine model can be used to approximately correct chromatic aberrations are within the range of a few pixels. Here, the 488 and 561 channels are shown in magenta and green, respectively. Zoom-ins (i) – (iv) illustrate the correction on one example image tile 1920×1920 pixels in size. In the bottom right the interest points (all points of the entire stack are shown for one slice) used for alignment are shown. Please note that for example the point in zoom-in (ii) was not used for alignment. If aberrations are significantly bigger than illustrated in this example or if not enough common autofluorescence between channels exist, images can be preprocessed with dedicated chromatic aberration software before import into BigStitcher (see Limitations section in **Online Methods**). Chromatic aberration correction was applied to all 26 image tiles of this dataset (**Supplementary Video 3**) as well as to all cleared samples that were acquired with 2 channels (**Supplementary Table 1**).

SUPPLEMENTARY FIGURE 4: Spherical aberration correction



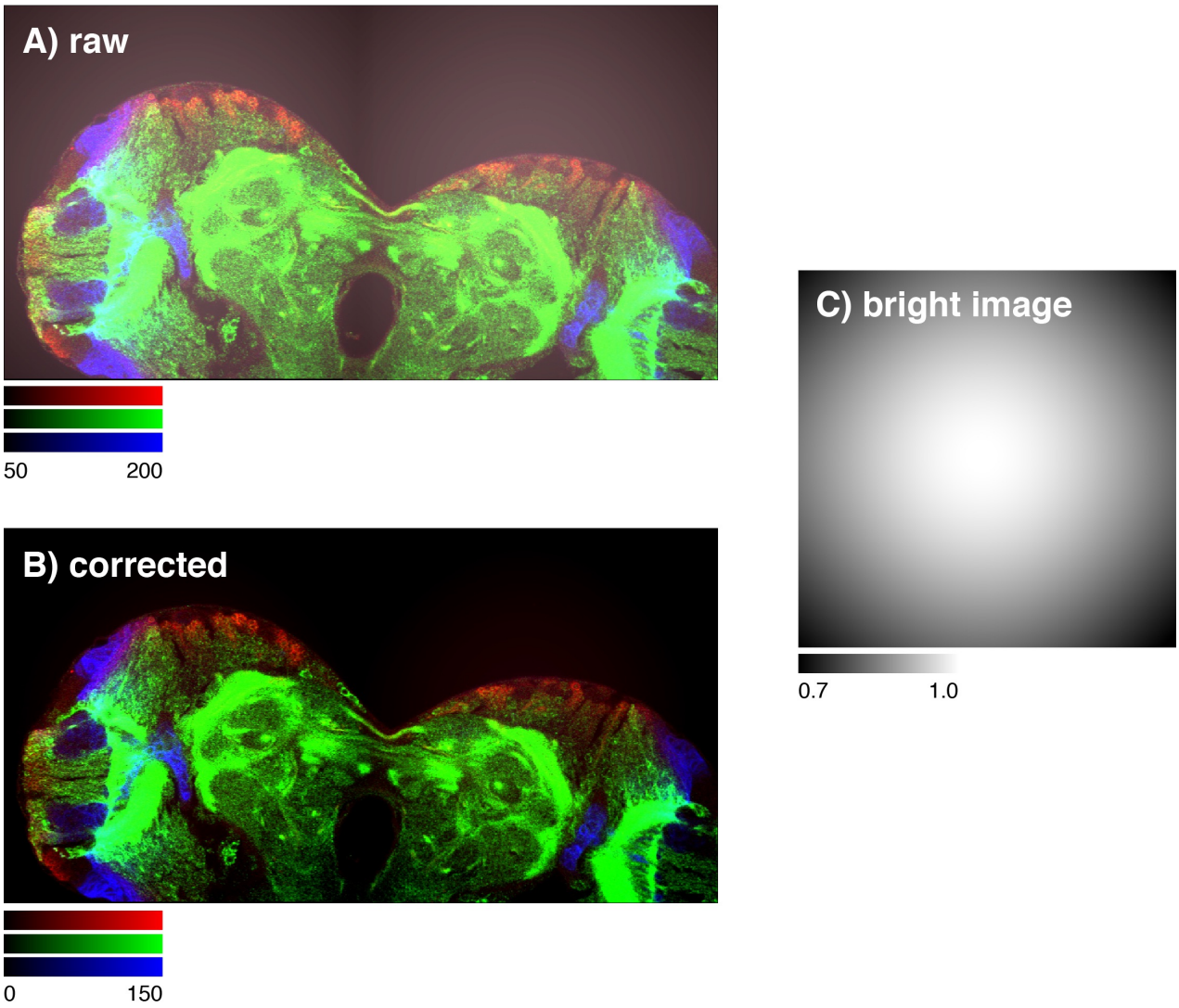
Supplementary Figure 4: *Spherical aberration correction.* **(A-C)** show the same area a cleared adult mouse hypothalamus expressing H2B-GFP in all bsx neurons where the corners of 4 image tiles of the same wavelength overlap. Zoom-ins (i)–(iv) show the alignment quality by overlaying different colors (1st row), after image fusing using blending (2nd row), and the sobel-filtering of the blending fusion (3rd row). **(A)** shows results for stitching, **(B)** when using affine refinement, and **(C)** when using affine refinement on re-blocked images. Note that affine, and split-affine improve the alignment quality. Blending-fusion can reduce artifacts as it reduces the contribution of pixels close to image borders. Arrows outline cases where artifacts persist after blending-fusion. E.g., the artifact visible in the fusion in (iii) stems from misalignments of the pink and red tile, since the green tile is almost invisible after fusion. **(A-C)** Spherical aberration correction using affine transformations achieving similar results was applied to all 26 tiles of the dataset, as well as the datasets shown in (**Fig. 1n, 3b-d**).

SUPPLEMENTARY FIGURE 5: Manual alignment



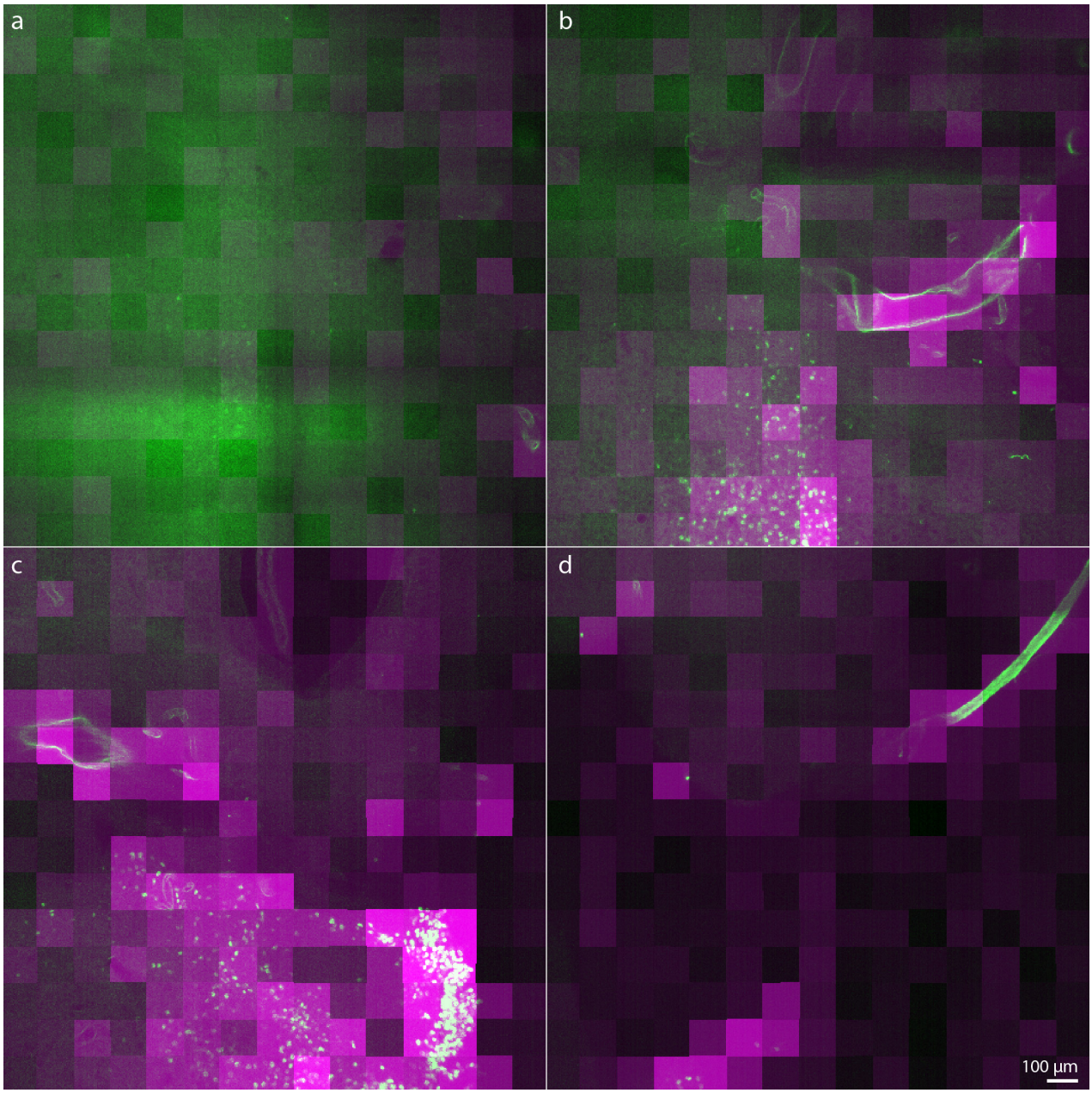
Supplementary Figure 5: *Interactive manual alignment of tiled images.* The BigStitcher GUI offers various ways of manually (pre-)aligning tiled images after import. **(A)** images can be moved to a regular grid with a given tile order and overlap. **(B)** image locations can also be read from a simple *tile configuration* text file. **(C)** selected image(s) can be moved along axes via sliders. **(D)** all changes will be displayed in the BigDataViewer window immediately **(D)** for quick verification.

SUPPLEMENTARY FIGURE 6: Flat-field correction



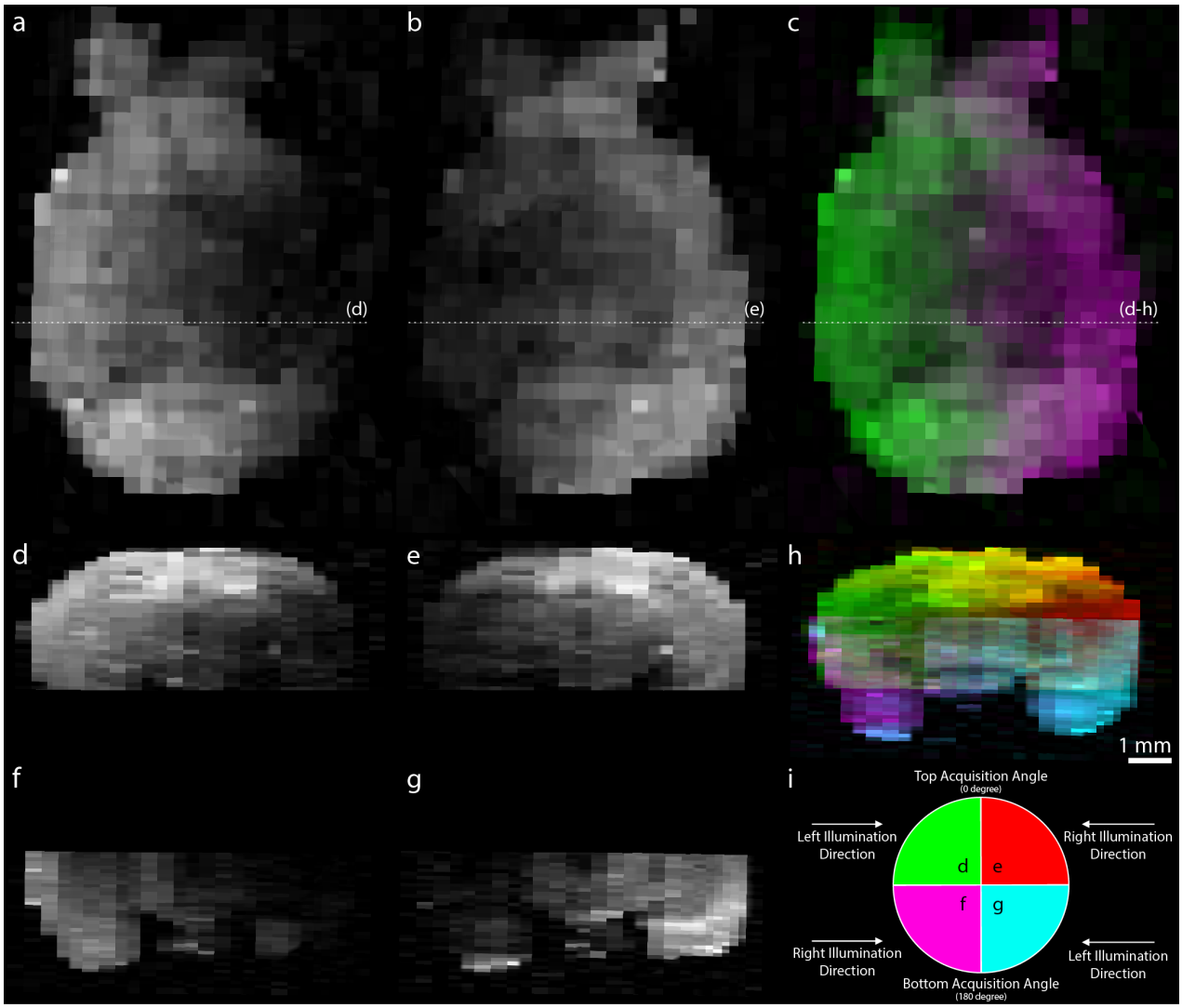
Supplementary Figure 6: *On-the-fly flat-field correction.* The BigStitcher offers correction for camera offsets, fixed pattern noise or uneven illumination. **(A)** Simulation of the effects of a constant background offset and Gaussian illumination/detection efficiency **(C)** on tiled images. By subtracting the *dark image* and modulating with the inverse relative intensity of the *bright image*, such artifacts can be corrected easily **(B)**. The correction is calculated virtually, with optional caching, to allow for immediate inspection of the results. **(A-C)** Flatfield correction as illustrated in this figure is a feature supported by BigStitcher, but has not been applied to any of the datasets shown in this publication.

SUPPLEMENTARY FIGURE 7: Automatic quantification of image quality



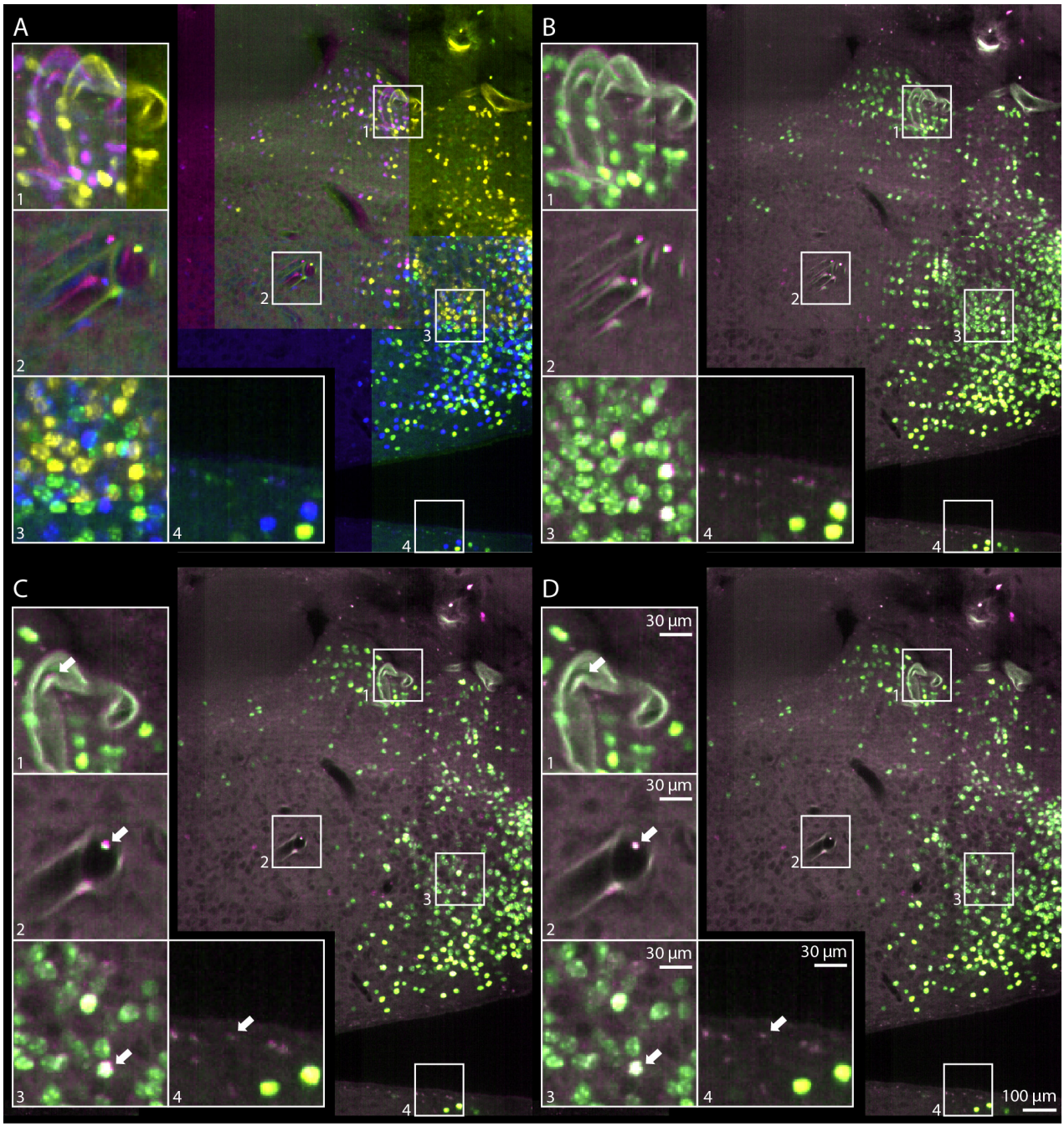
Supplementary Figure 7: *Automatic Quantification of Image Quality.* **(a-d)** Four different z-planes from a volume that overlays the results of the relative Fourier Ring Correlation (rFRC, see **Online Methods**) computed in 128×128 blocks using a spacing of 10 pixels (magenta) and cleared image data (same as **Supplementary Fig. 1**). The rFRC robustly detects areas with high image quality. Note that (a) is deepest inside the tissue and (d) is at the surface of the sample. See **Supplementary Video 8** for an animation of the entire stack. The rFRC was successfully applied to all cleared datasets in this publication (**Supplementary Table 1**), results are also shown in **Supplementary Fig. 1, 8** and **Supplementary Video 8,9**.

SUPPLEMENTARY FIGURE 8: Quality estimation in whole-brain mouse acquisition



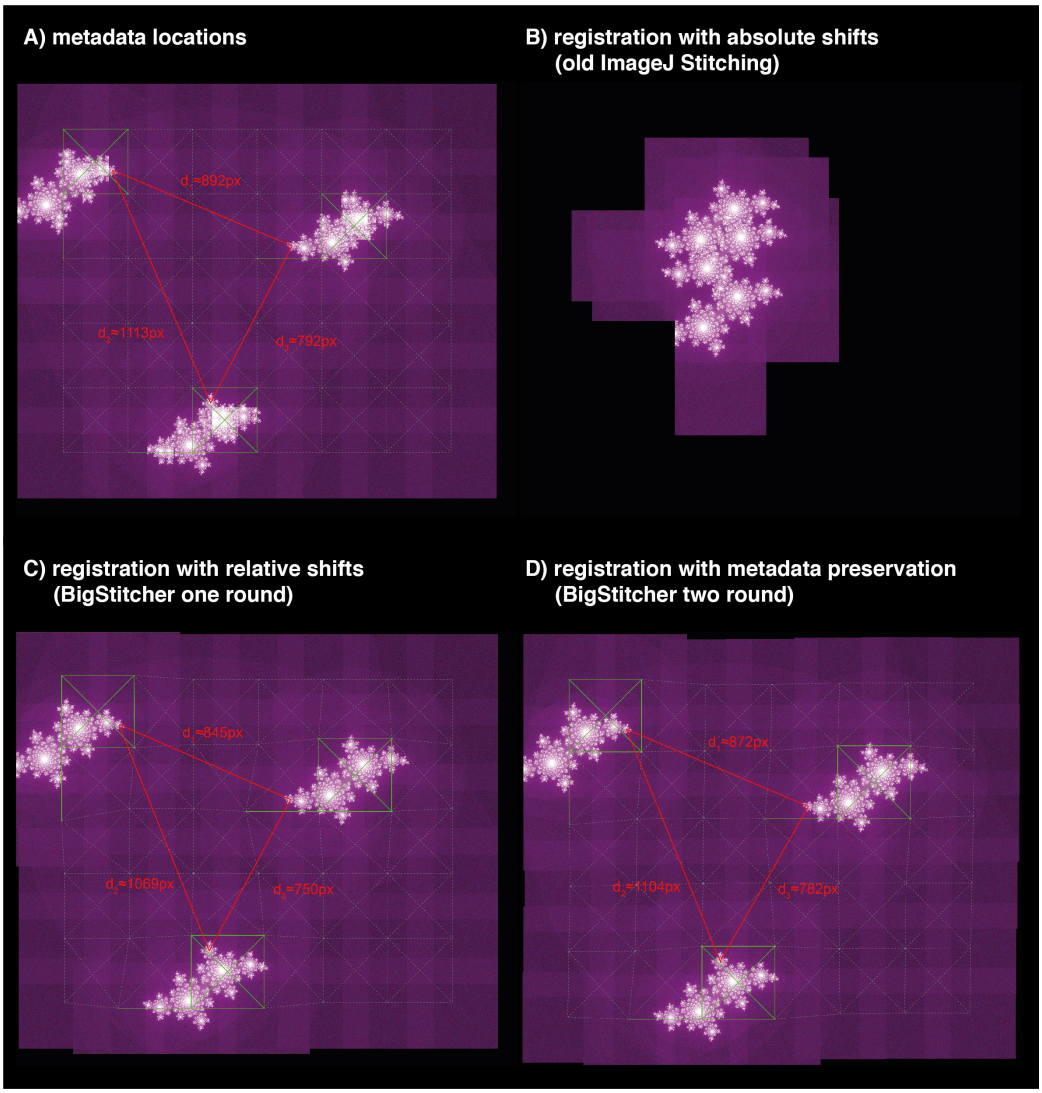
Supplementary Figure 8: *Quality Estimation in Whole-Brain Mouse Acquisition.* Application of our relative Fourier Ring Correlation (rFRC, see **Online Methods**) to the reconstruction of an entire adult mouse brain. The rFRC was computed in 512×512 blocks using a spacing of 256 pixels and transformed and rendered as the reconstructed volume (see Fig. 3d). **(a,b,c)** single slice through rFRC volume based on image data acquired with left illumination (a), right illumination (b), and both overlaid (c), dotted lines outline the orthogonal shown views in (d-h). **(d-h)** orthogonal views to (a-c) highlighting the contribution in image quality from different illumination directions and acquisition angles. **(i)** illustration of the color scheme used in (h) and the type of data displayed in (d-g). See **Supplementary Video 9** for an animation of the entire stack. The rFRC was successfully applied to all cleared datasets in this publication (**Supplementary Table 1**), results are also shown in **Supplementary Fig. 1, 7** and **Supplementary Video 8,9**.

SUPPLEMENTARY FIGURE 9: Affine refinement via ICP



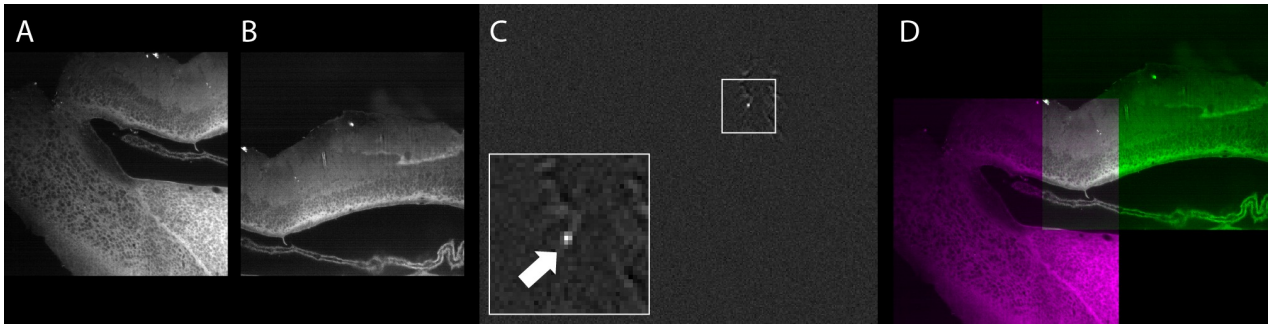
Supplementary Figure 9: *Illustration of different steps for multi-tile alignment* **(A)** Four randomly colored, overlapping image tiles show the typical error when using microscope metadata only. **(B)** Shows the same image tiles as (A), but without random color coding. **(C)** Quality of the registration after applying the phase-correlation based stitching with downsampling 4 and two-round global optimization. **(D)** Result after applying the automatic ICP refinement for tile alignment, spherical and chromatic aberration correction. **(A-D)** Insets highlight specific areas to better appreciate quality differences. **(A-D)** ICP-refinement using affine transformations achieving similar results was applied to all 26 tiles of the dataset, as well as the datasets shown in (**Fig. 1n, 3b-d**).

SUPPLEMENTARY FIGURE 10: Global optimization



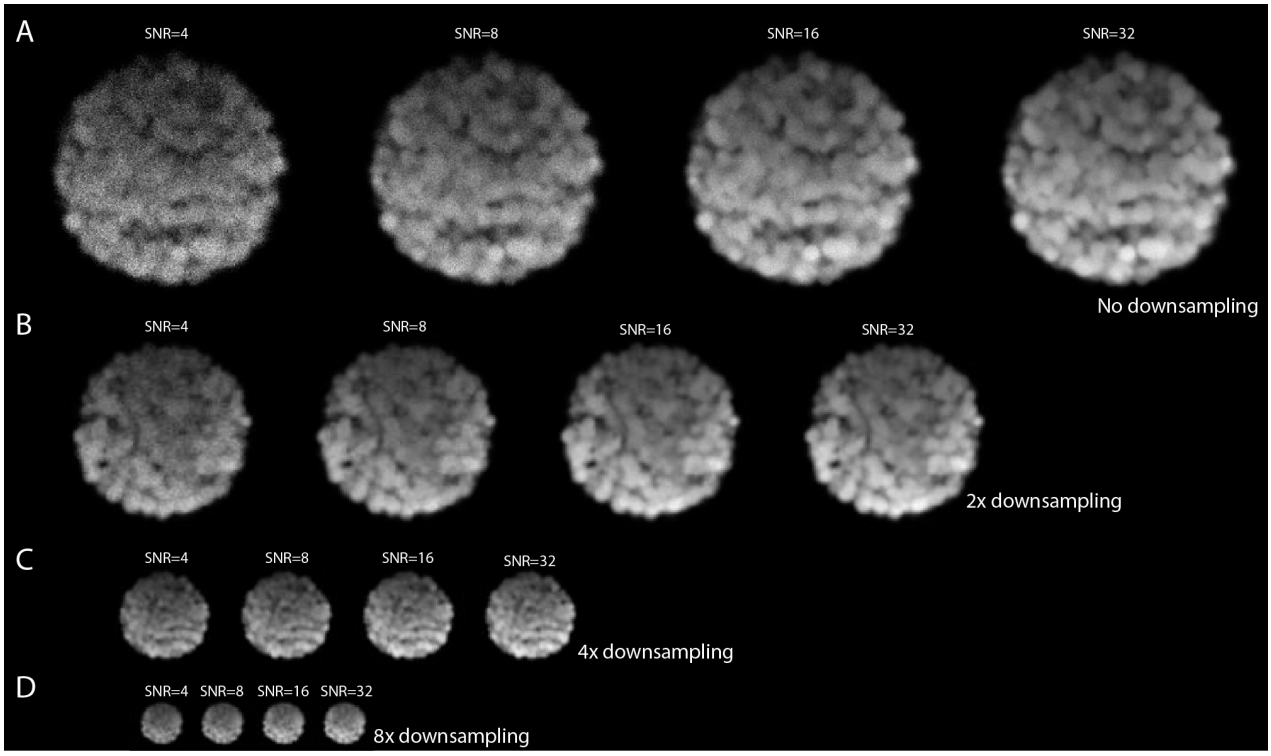
Supplementary Figure 10: *Global optimization of pairwise registration in sparse datasets connected by "empty tiles" (noise only)* . **(A)** Simulation of a tiled image dataset with sparse objects: tiled images of multiple translated Julia fractals moved to a grid according to approximate metadata (with too high overlap). Centers of images for which pairwise shifts can be determined via phase correlation are connected by green lines, whereas centers of neighboring tiles for which no meaningful shift can be calculated are linked by dashed grey lines. Manually measured distances between distinct points in the three fractals are shown in red. **(B)** performing global optimization with *absolute shifts* (as it is done BigStitcher's predecessor, the ImageJ Stitching plugin) will correctly align images within connected components of the link graph but place all fractals close to the origin. **(C)** by using *relative shifts*, BigStitcher will leave disconnected objects at their initial location while still aligning within connected components. **(D)** as registrations are not propagated between unconnected tiles, distances between neighboring objects might change. By running a second round of optimization to align connected components according to metadata shifts and applying the results to the in-component registrations, distances between neighboring objects are preserved as-good-as-possible. **(A-D)** Two-round global optimization as illustrated in this figure is a feature supported by BigStitcher, which has been applied to all datasets used in this publication. Especially the dataset shown in **Fig. 1d,e and Fig. 3d** profits from it since it contains empty tiles.

SUPPLEMENTARY FIGURE 11: Pairwise registration by phase correlation



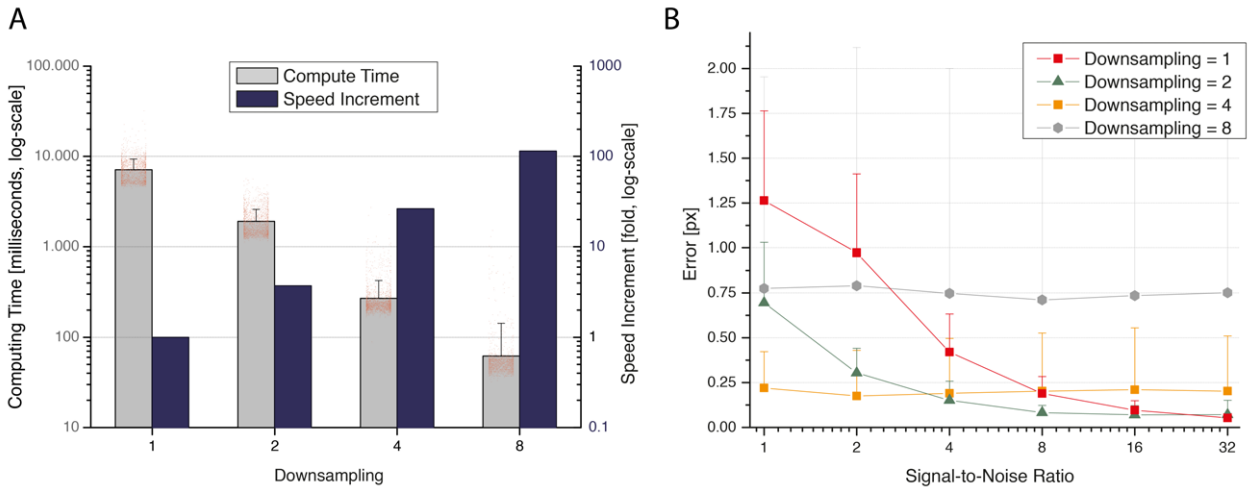
Supplementary Figure 11: *Pairwise registration by phase correlation*. **(A,B)** Central slices of image stacks from a tiled acquisition (non-regular tiling) of a cleared adult mouse hypothalamus. **(C)** Phase correlation matrix (PCM) calculated from the two images shows a single, distinct peak above nearly constant background. The peak location corresponds to the relative translation t of both tiles. **(D)** Central slice through the images aligned according to t, as displayed in interactively during the reconstruction process. **(A-D)** The pairwise registration using phase correlation was used as a first step in the alignment of all cleared and expanded samples used in this publication (**Supplementary Table 1**).

SUPPLEMENTARY FIGURE 12: Downsampling with different SNR



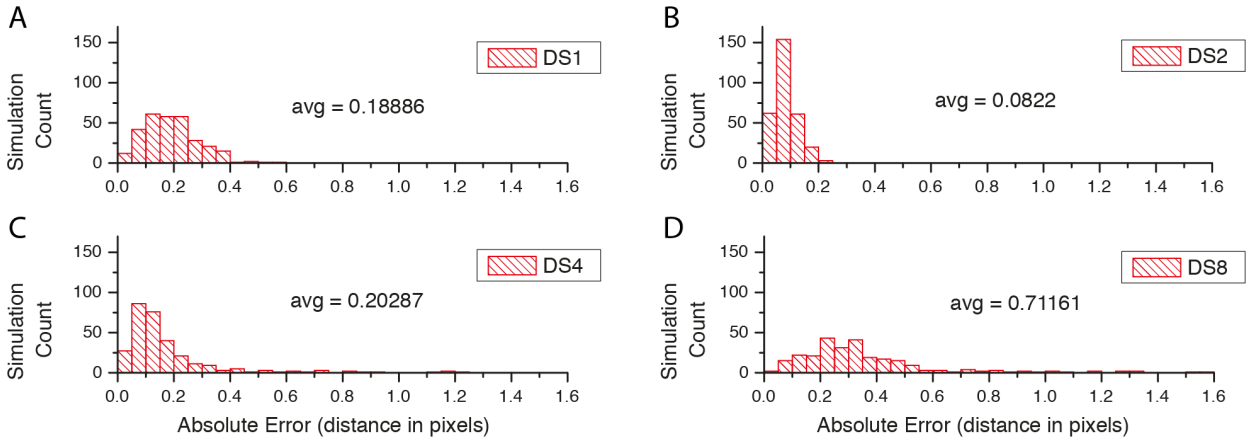
Supplementary Figure 12: *Effects of downsampling on simulated data with different SNR.* **(A)** Simulated image stacks of spheroid-like objects deteriorated by anisotropic sampling, light attenuation, convolution with an anisotropic PSF, and pixel intensity generation using Poisson processes to archive desired signal-to-noise-ratios (SNRs). A central slice through 3d volumes is shown. **(B,C,D)** Effects of downsampling on the simulated images. The effects of Poisson Shot Noise are gradually reduced by the blurring of increasing downsampling. **(A)** For quantification of the alignment quality using these simulations, 300 independent simulations were run for each combination of SNR and downsampling (see **Supp. Fig. 13 – 15**).

SUPPLEMENTARY FIGURE 13: Downsampling statistics 1



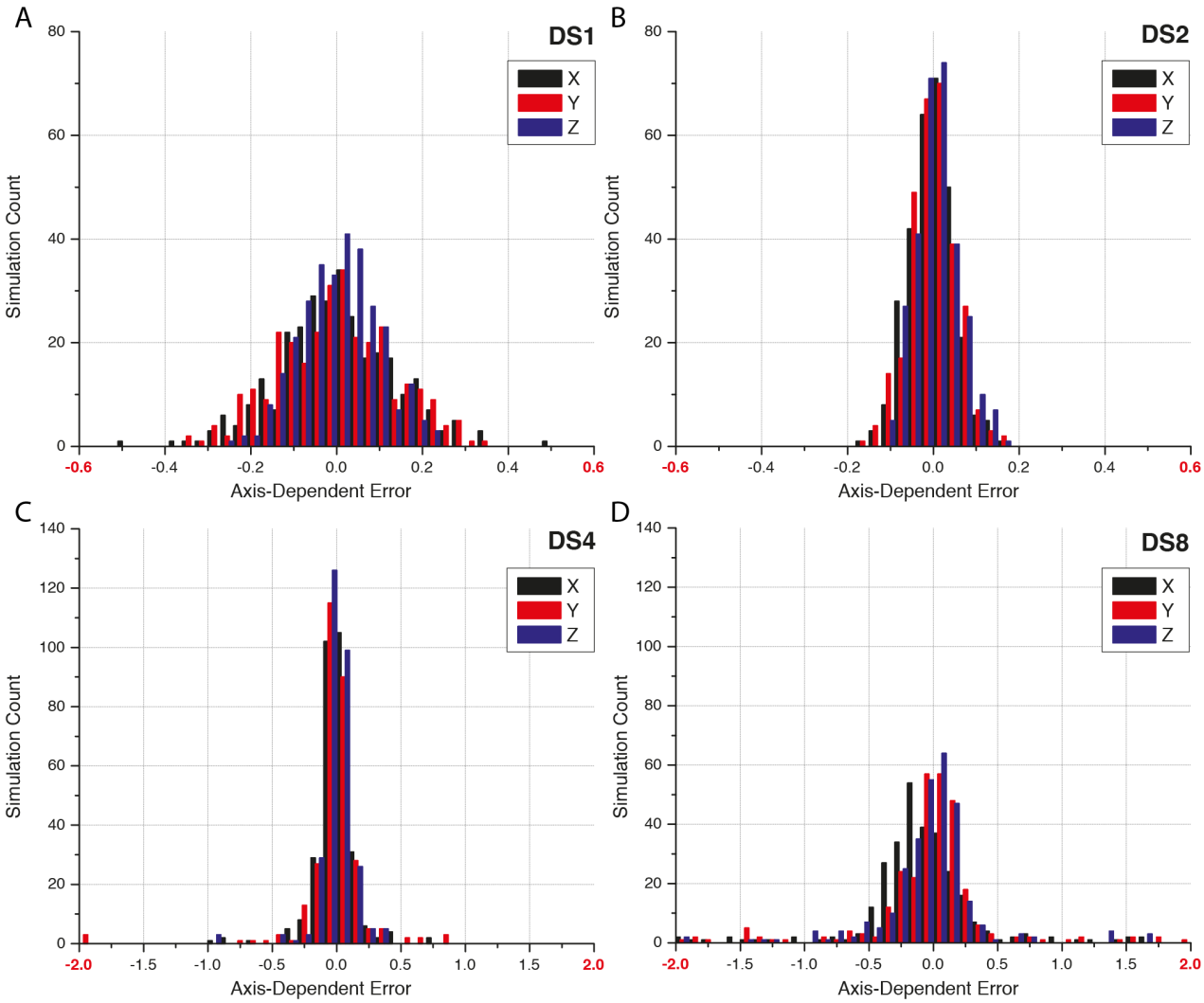
Supplementary Figure 13: *Processing times and overall errors.* **(A)** Processing times for sub-pixel precise identification of overlap between simulation spheroids. With increasing downsampling, the computation time drops significantly. Red dots show individual measurements, note the log-scale. Average (StDev) of computing time is 7122 (2224) msec, 1910 (681) msec, 271 (155) msec, and 62 (80) msec for downsampling 1, 2, 4 and 8, respectively. The speed increments are computed as the ratio of the average compute times, i.e. 1×, ∼4×, ∼26×, and ∼115×, respectively. Compute times were measured in a single thread on a Intel Xeon E5-2640 v4. **(B)** Average errors including their standard deviation for all combinations of SNR and downsampling. **(A,B)** All errors are in units of the input images (no downsampling). For each combination of SNR and downsampling 300 independent simulations were run to compute the values.

SUPPLEMENTARY FIGURE 14: Downsampling statistics 2



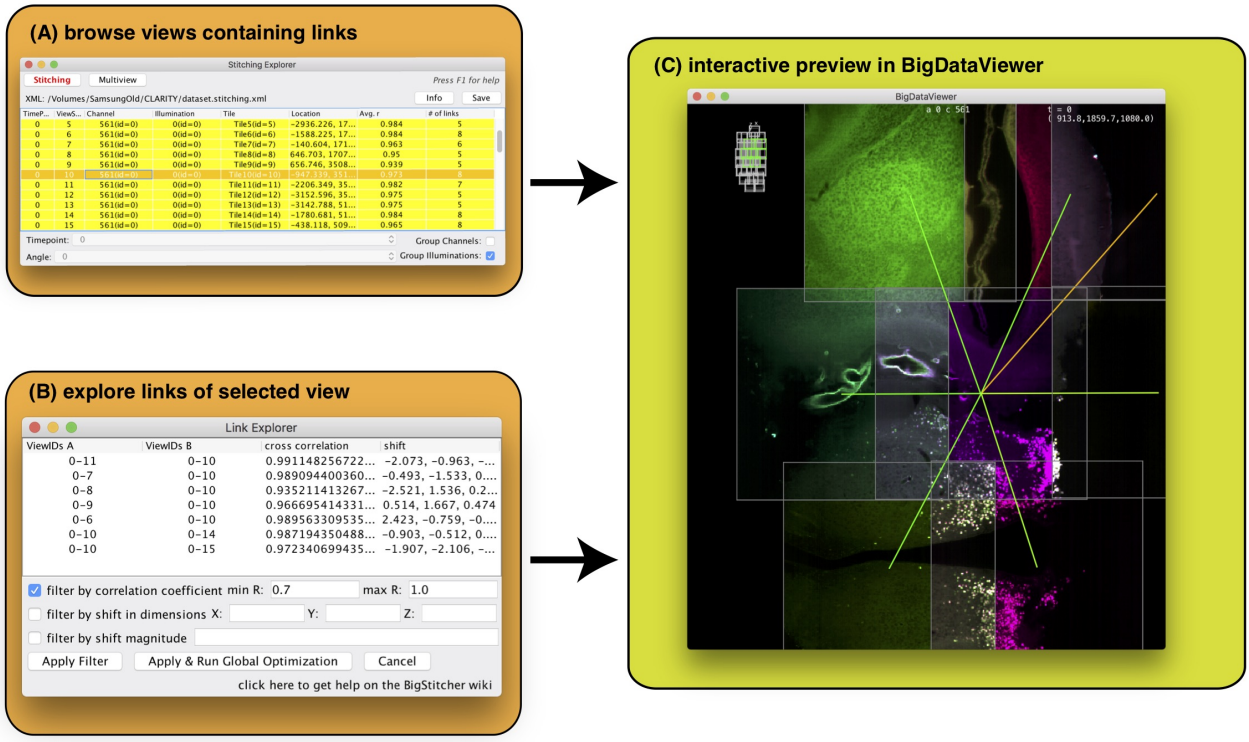
Supplementary Figure 14: *Errors for different downsamplings at SNR=8.* **(A-D)** Histograms showing the distributions of error of the simulations. Errors initially decrease due to the smoothing effect of the downsampling. All errors are in pixel units of the original resolution (DS1). Each histogram consists of 300 independent simulations.

SUPPLEMENTARY FIGURE 15: Downsampling statistics 3



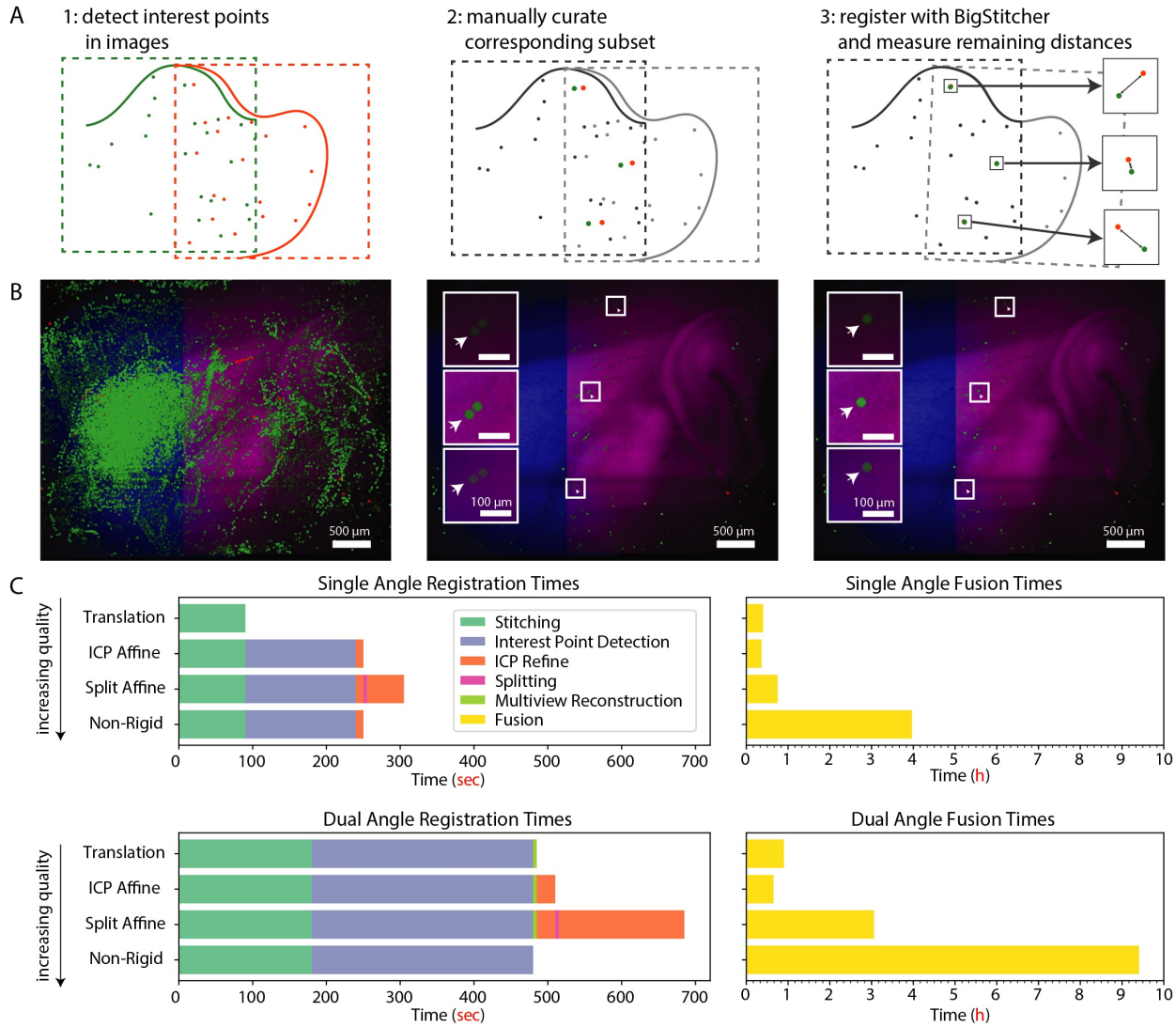
Supplementary Figure 15: *Absolute distance errors at SNR=8.* **(A-D)** Histograms showing the absolute distances between computed and known shift between two simulated spheroids, split by dimension. It illustrates a normal distribution of the error made during the pairwise phase correlation. All errors are in pixel units of the original resolution (DS1). Each histogram consists of 300 independent simulations.

SUPPLEMENTARY FIGURE 16: Interactive inspection and curation of pairwise links



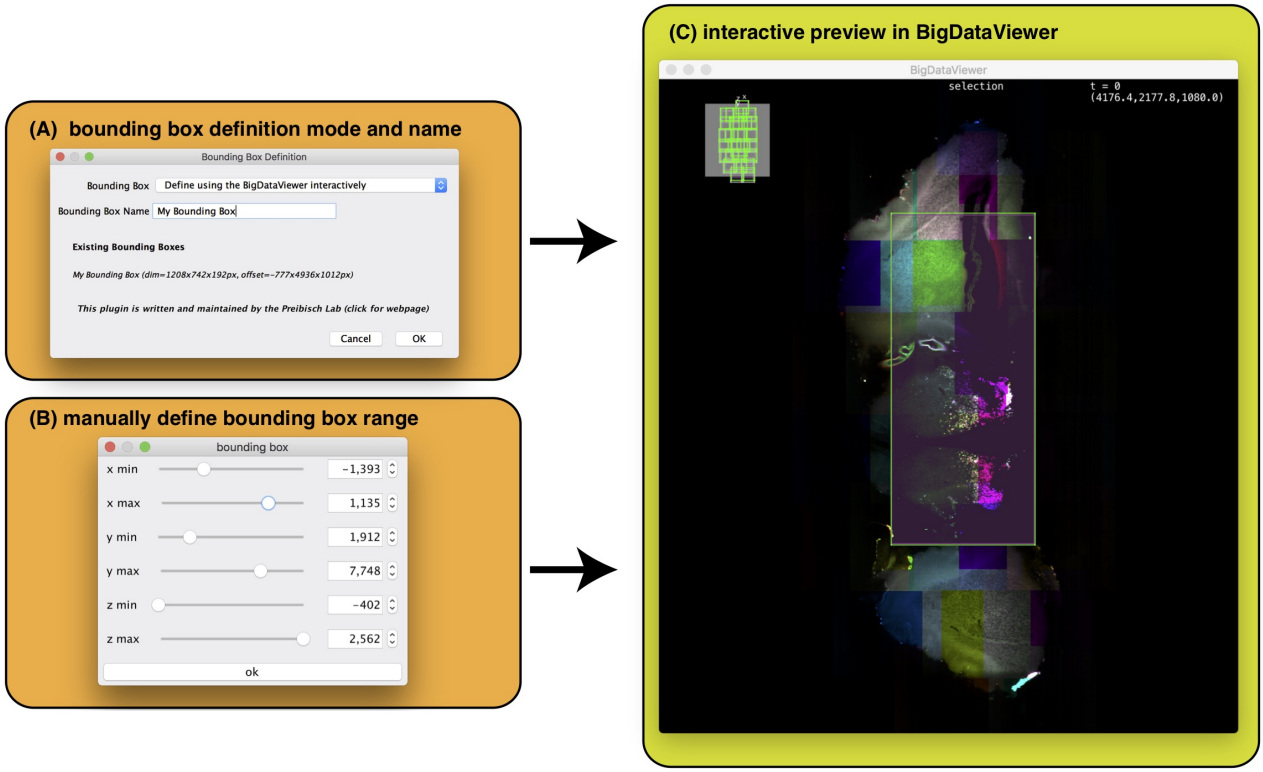
Supplementary Figure 16: *Interactive visualization of links in the link explorer.* The BigStitcher GUI offers to explore and modify calculated links between corresponding tiles in the link explorer menu. **(A)** tiles containing links are displayed in yellow and can be selected. **(B)** display corresponding tiles of the selected view. Single links can be removed manually or through available filtering options. **(C)** corresponding links of the selected view are displayed in real-time in the BigDataViewer.

SUPPLEMENTARY FIGURE 17: Quantification of image registration quality



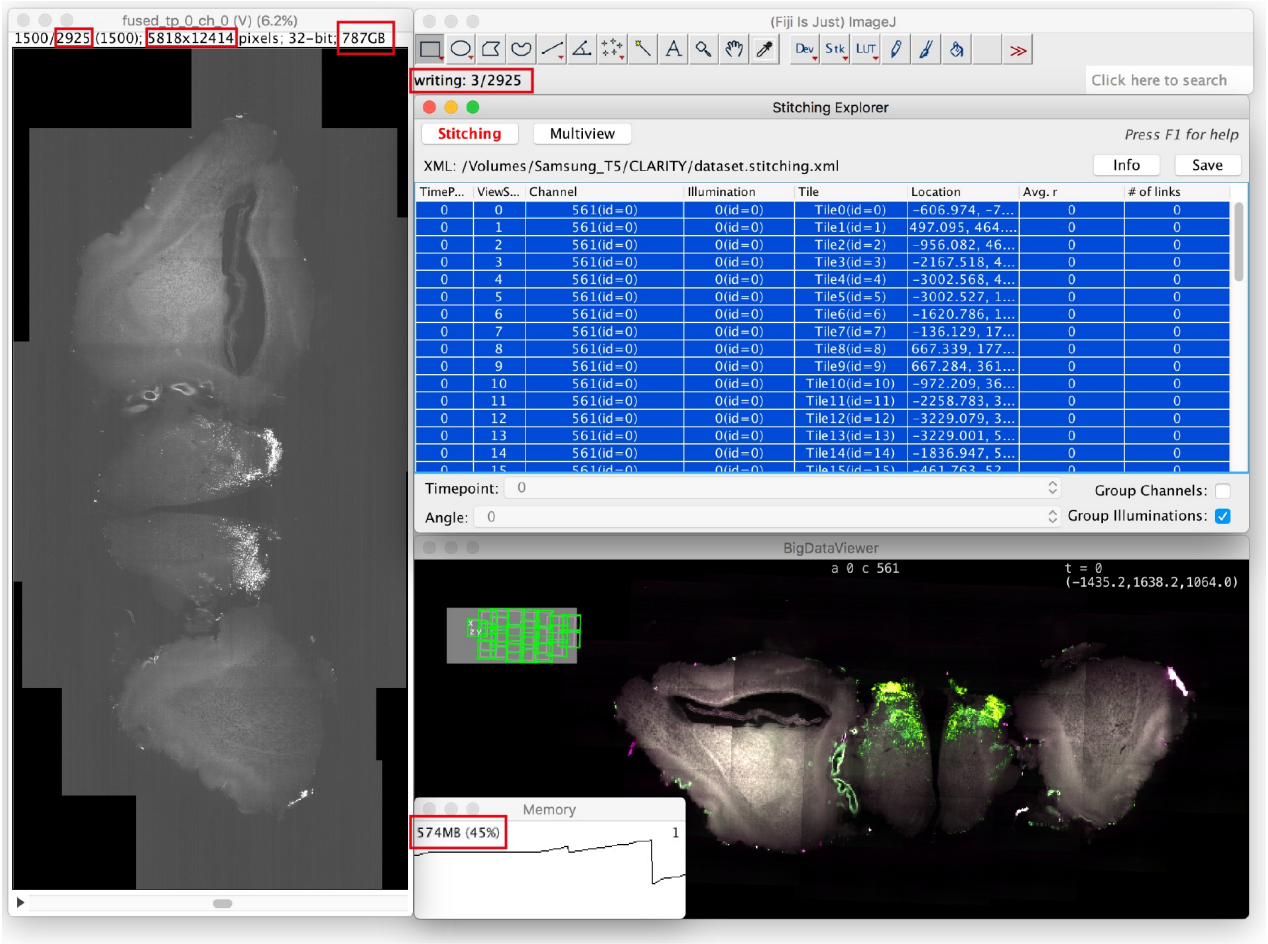
Supplementary Figure 17: *Quantification of Image Registration Quality.* A multi-view, dual-illumination, multi-tile dataset specifically acquired for verification purposes was used to quantify the registration error (see **Online Methods**, **Fig. 2k-m**, and **Supplementary Fig. 2**). **(A)**: Schematic description of the quantification process for registration accuracy. Interest points are first automatically detected in all images of the dataset (1). Of those, a subset of truly corresponding points interest points was selected for each pair of images (2). After registration with BigStitcher using various transformation models (translation, affine, split-affine, non-rigid), the remaining distance between the manually curated point pairs is used as a measure of registration error (3), actual errors are shown in **Fig. 2m**. **(B)** All interest points detected in two images of the example dataset overlaid on a slice view (left), manually selected corresponding points (middle, note the ”doubling” of the paired points as they are not yet aligned, arrows indicate examples) and the same points after registration (right). **(C)** Time required for registration (left) and fusion (right) of the dataset, for a single angle (top) or both angles (bottom). The single angle values are averages of both angles. Fusion was done at full resolution, preserving original data anisotropy. Multi-resolution pyramids of the images were computed beforehand. Processing was done on 2 Intel Xeon E5-2680v4 processors and 256GB RAM, data was loaded from SSDs in RAID0 configuration. This error quantification was performed only on this specifically acquired dataset.

SUPPLEMENTARY FIGURE 18: Bounding-box definition



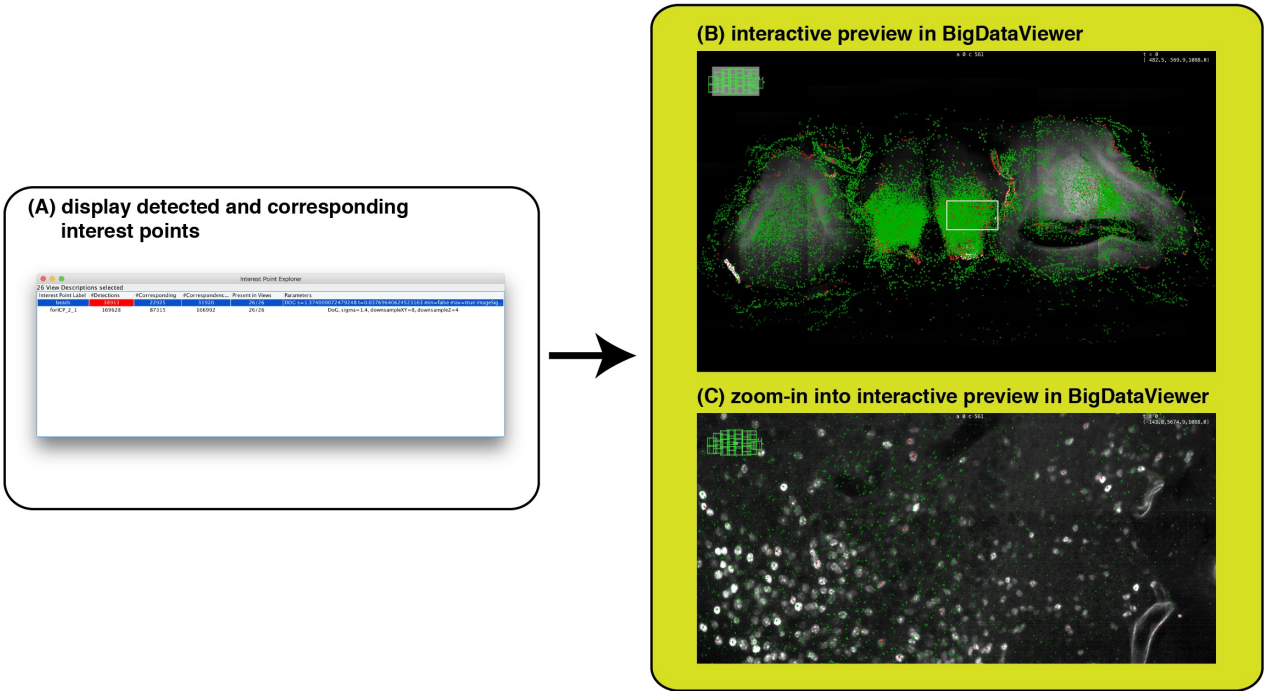
Supplementary Figure 18: *Interactive definition of bounding boxes.* The BigStitcher GUI offers the possibility of defining or modifying regions of interest via the creation of bounding boxes. **(A)** Choose the method used to define a new bounding box. In this case the interactive mode is selected. **(B)** manually define the bounding box range **(C)** Preview the size of the specified bounding box in the BigDataViewer in real-time.

SUPPLEMENTARY FIGURE 19: Virtual fusion of large Image



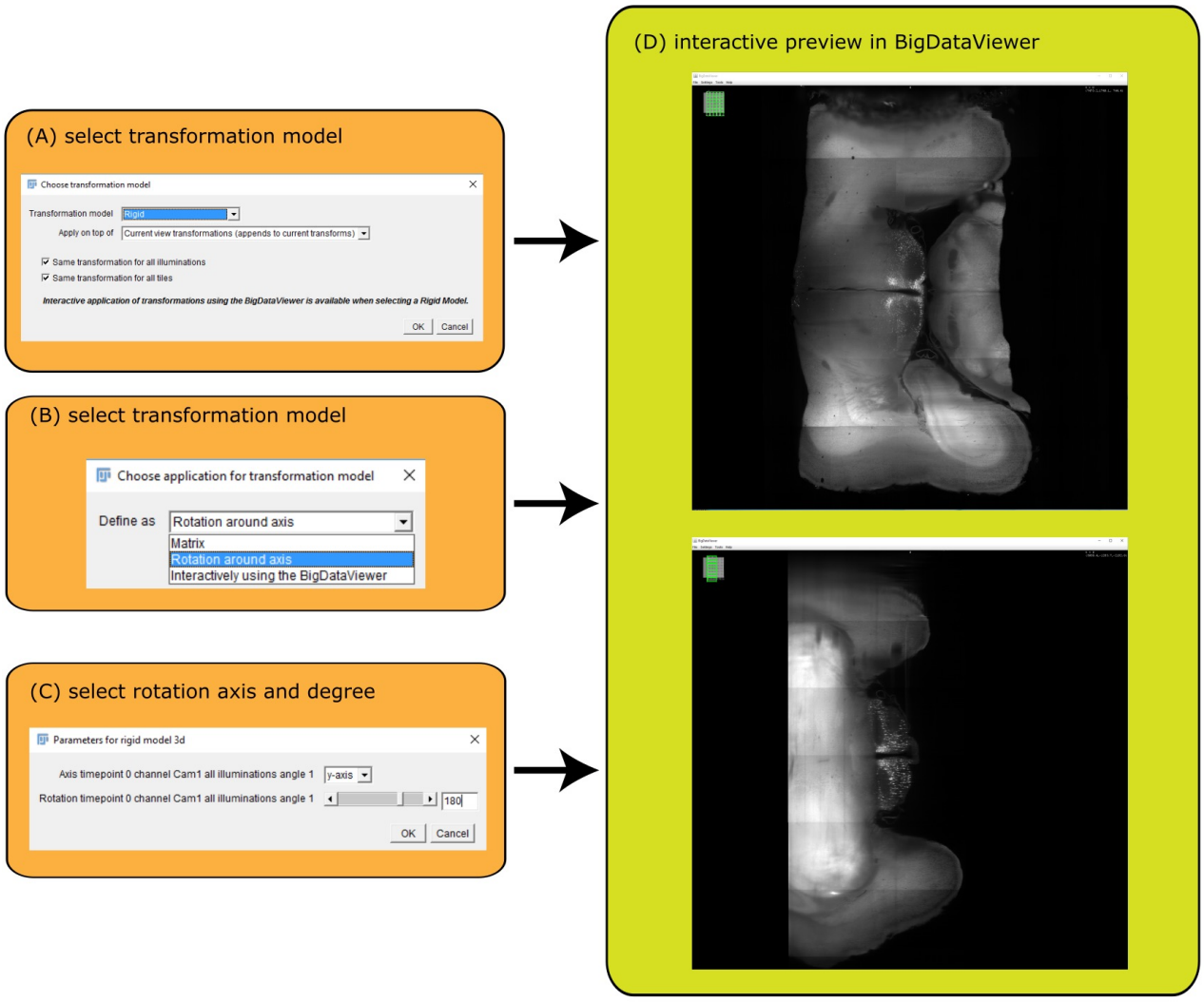
Supplementary Figure 19: *Virtual Fusion.* Screenshot of a Fiji instance running with **1.25GB of RAM** successfully fusing and saving a **787GB** volume 5818×12414×2925 pixels in size. This is achieved through **virtual fusion** combined with virtual, cached loading of blocked, multi-resolution input images. Red boxes highlight memory consumption, size, and progress. During the fusion process, the BigStitcher and BigDataViewer are interactively accessible.

SUPPLEMENTARY FIGURE 20: Interest point visualization



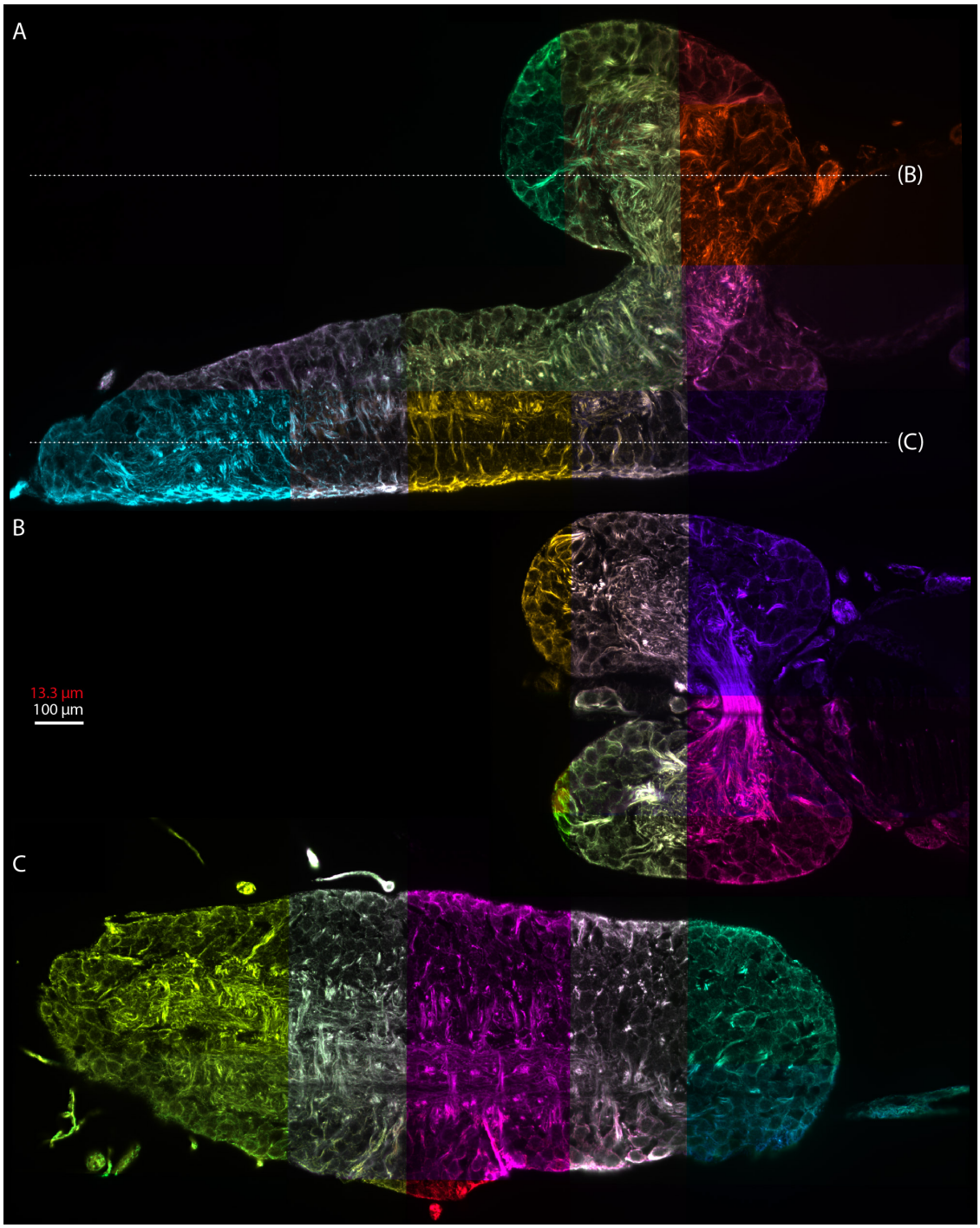
Supplementary Figure 20: *Interactive visualization of interest points.* The interest points explorer allows the visualization of interest points and corresponding interest points between views. **(A)** select desired interest points for visualization. **(B)** preview the interest points overlaid in the BigDataViewer. Red dots intersect with the current image plane, green dots are projections from different z-planes. The white box marks the zoom-in shown in (C). **(C)** Zoom-in into the region outlined in (B).

SUPPLEMENTARY FIGURE 21: Manual transformation of multi-view datasets



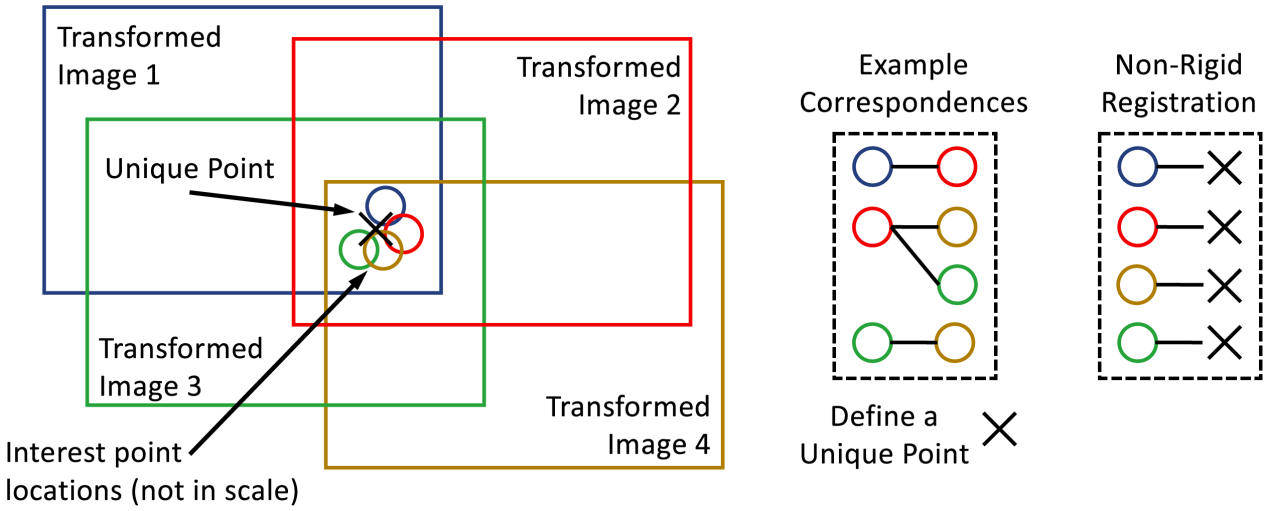
Supplementary Figure 21: *Interactive transformation of views.* Different transformation models can be applied to one or more views and simultaneously visualized in the BigDataViewer. **(A)** Choose transformation model grouping. **(B)** Further define the transformation model. In this case a rotation around the axis is selected. **(C)** Select rotation axis and angle. **(D)** Visualize rotation of the view in the BigDataViewer.

SUPPLEMENTARY FIGURE 22: Expansion microscopy reconstruction



Supplementary Figure 22: *Expansion microscopy stitching* **(A)** All tiles (randomly colored) of one view of the expanded Drosophila central nervous system. Dotted lines highlight orthogonal sections in (B) and (C). **(B,C)** Alignment of an orthogonal view showing at two different cut planes. Red scalebar takes expansion into account. **(A-C)** Expansion microscopy alignment using phase correlation followed by ICP refinement using affine transformations was performed only on this dataset.

SUPPLEMENTARY FIGURE 23: Principles of non-rigid alignment



Supplementary Figure 23: *Principle of Non-Rigid Alignment.* The non-rigid alignment applies a different affine transformation to each pixel of each transformed image. This continuous transformation space is defined by corresponding interest points between overlapping images. We therefore first identify all sets of all corresponding interest points that belong to each other as defined by pairwise correspondences (see Example Correspondences). Each individual set of correspondences across n images then define a $unique\ point$, of which typically hundreds to thousands per image exist. The left part of the figure illustrates a single $unique\ point$, which is defined as the average position of all corresponding interest points. Once all unique points are assigned to each correspondence, the non-rigid transformation can be individually computed for each transformed image.

SUPPLEMENTARY TABLE 1: Summary of all datasets used in this publication.

Dataset	Size	Microscope & Acquisition settings
Coronal slice from a adult mouse brain expressing H2B-eGFP under the neuronal BSX promoter (**Figure 3a + Supplement**)	1920×1920×1039 16 bit stacks 26 tiles from 1 angles (10% overlap) 2 channels, single illumination **0.36 TB total size**	Lightsheet Z.1 with EC Plan-Neofluar 5x/0.16 objective, Depth-of-field: $\sim$25μm 5.37μm LS thickness, 538.3647546μm Confocal parameter 0.915×0.915×2.57425409μm pixels 119.8ms exposure on PCO.edge camera lasers: 561 nm 100%, 488 nm 80% filters: EF1 BP 505-545, EF 2 BP 575-615
Whole adult mouse brain expressing H2B-eGFP in all BSX-expressing neurons (**Figure 1d-n + Figure 2a,f + Figure 3d,e + Supplement**)	1920×1920×770 16 bit stacks 56 tiles from 2 angles (10% overlap) 2 channels, dual illumination **2.2 TB total size**	Lightsheet Z.1 with EC Plan-Neofluar 5x/0.16 objective, Depth-of-field: $\sim$25μm 5.97μm LS thickness, 665.3906758999999μm Confocal parameter 0.915×0.915×4.929649351μm pixels 119.8ms exposure on PCO.edge camera lasers: 561 nm 40%, 488 nm 40% filters: EF1 BP 505-545, EF 2 BP 575-615
	1920×1920×645 16 bit stacks 63 tiles from 2 angles (10% overlap) 2 channels, dual illumination **2.1 TB total size**	Lightsheet Z.1 with EC Plan-Neofluar 5x/0.16 objective, Depth-of-field: $\sim$25μm 10.44μm LS thickness, 2034.834282μm Confocal parameter 0.915×0.915×6.203581395μm pixels 119.8ms exposure on PCO.edge camera lasers: 561 nm 40%, 488 nm 40% filters: EF1 BP 505-545, EF 2 BP 575-615
Coronal slice through an adult mouse brain expressing an H2B-eGFP lineage tracing marker in BSX-expressing neurons (**Figure 3b,e + Supplement**)	1920×1920×828 16 bit stacks 35 tiles from 2 angles (10% overlap) 2 channels, dual illumination **1.55 TB total size**	Lightsheet Z.1 with EC Plan-Neofluar 5x/0.16 objective, Depth-of-field: $\sim$25μm 10.44μm LS thickness, 2034.834282μm Confocal parameter 0.915×0.915×4.930096618μm pixels 119.8ms exposure on PCO.edge camera lasers: 561 nm 50%, 488 nm 50% filters: EF1 BP 505-545, EF 2 BP 575-615
	1920×1920×960 16 bit stacks 35 tiles from 2 angles (10% overlap) 2 channels, dual illumination **1.8 TB total size**	Lightsheet Z.1 with EC Plan-Neofluar 5x/0.16 objective, Depth-of-field: $\sim$25μm 10.44μm LS thickness, 2034.834282μm Confocal parameter 0.915×0.915×4.930916667μm pixels 119.8ms exposure on PCO.edge camera lasers: 561 nm 50%, 488 nm 50% filters: EF1 BP 505-545, EF 2 BP 575-615

Continued on next page

SUPPLEMENTARY TABLE 1: Summary of all datasets used in this publication.

Dataset	Size	Microscope & Acquisition settings
Coronal slice through an adult mouse brain expressing an H2B-eGFP lineage tracing marker in BSX-expressing neurons (zoomed-out) (**Figure 2k-m + Supplement**)	1920×1920×945 16 bit stacks 12 tiles from 2 angles (10% overlap) 1 channel, dual illumination **0.166 TB total size**	Lightsheet Z.1 with EC Plan-Neofluar 5x/0.16 objective (0.5× zoom) Depth-of-field: $\sim 35 \mu m$ $10.44 \mu m$ LS thickness, $2034.834282 \mu m$ Confocal parameter $1.83 \times 1.83 \times 4.93 \mu m$ pixels 119.8ms exposure on PCO.edge camera lasers: 488 nm 50% filters: EF1 BP 505-545
Whole *C. elegans* during dauer with all neuron nuclei expressing tagRFP in and co-stained with DAPI (**Figure 3e,f**)	750×1920×40 16 bit stacks 16 tiles from 4 angles (10% overlap) 2 channels, single illumination **0.003 (2,96GB) TB total size**	Lightsheet Z.1 with W Plan-Apochromat 20x/1.0 objective (2× zoom) Depth-of-field: $\sim 1 \mu m$ $1.82 \mu m$ LS thickness, $56.72238239 \mu m$ Confocal parameter $0.114 \times 0.114 \times 0.8035 \mu m$ pixels 59.9ms exposure on PCO.edge camera lasers: DAPI 505 nm 2%, DAPI 561 nm 0.2%, RFP 561 nm 5% filters: DAPI CAM BS Mirror, RFP CAM BS SBS LP 560, RFP EF 1 BP 575-615
Central nervous system of a Drosophila 1st instar larva with immunostaining for tubulin (**Figure 3c,e + Supplement**)	2048×2048×923 stacks 26 tiles from 2 angles (28% overlap) 1 channels, dual illumination **0.188 TB total size**	IsoView with SpecialOptics 16x/NA 0.71 objective Depth-of-field: $\sim 2 \mu m$ $4.92 \mu m$ LS thickness, $416.0 \mu m$ Confocal parameter $0.4125 \times 0.4125 \times 0.8125 \mu m$ pixels 20.1ms exposure on Orca Flash 4.0 v2 sCMOS camera lasers: 488 nm 10.6mW filters: BP488/10 EX, BP525/50 EM

Note: Depth-of-field was estimated based on the descriptions on this resource Depth of Field and Depth of Focus — MicroscopyU.

SUPPLEMENTARY TABLE 2: Performance comparison

Performance comparison						
Data Size	130 Mb / 63 Gb / 300 Gb	130 Mb / 63 Gb / 300 Gb	130 Mb / 63 Gb / 300 Gb	130 Mb / 63 Gb / 300 Gb	130 Mb / 63 Gb / 300 Gb	1.67 Tb
Software	**Illumination Selection**	**Stitching** (requires Illumination Selection)	**ICP refinement**	**Fusion** 8x down-sampling full resolution (requires Illumination Selection)	**Virtual Fusion** @ full resolution	**Multi-view Registration**
BigStitcher	n.a. / 9 s / 14 s	4 s / 1 min 28 s / 7 min 15 s	9 s / 57 s / 2 min 52 s	1 s / 9 s / 44 s 8 s / 1 h 19 min / 8 h 43 min	1 s / 15 s / 19 s	6 min 30 s
TeraStitcher	X	16 s / 34 min 43 s / 2 h 29 min	X	1 min 15 s / 14 min 24 s* / 1 h 4* 2 min 29 s / 26 min 41 s* / 2 h 35 min*	X	X
ImageJ Stitcher	X	10 s / 61 min / 5 h 35 min	X	X / X / X 16 s / 10 h 36 min* / 44 h 40 min*	X	X
Xuv Tools	X	2 s / X / X	X	X / X / X 9 s / X / X	X	X
Arivis	X	27 s / 47 h 15 min / 7 d 4 h*	X	X / X / X 2 s / 5 min* / 51 min*	X	X
BigStitcher benchmark for processing a terabyte-sized multi-view dataset						
Data Size	1.67 Tb					
Software	**Illumination Selection**	**Stitching**	**ICP refinement**	**Fusion** @ 4x downsampling	**Virtual Fusion** @ full res. (display / save)	**Multi-view Registration**
BigStitcher	59 s	1h 14 min	7 min 28 s	1 h 1 min	56 s / 23 h 50 min*	6 min 30 s

Supplementary Table 2 *Comparison of BigSticher features and performance with other available stitching programs for four different datasets with different sizes (130 Mb, 63 Gb, 300 Gb, and 1,67 Tb).* Benchmarks were performed on a HP Z840 workstation running Windows 10 Pro with two Intel Xenon E5-2667v4 CPUs and 512 GB memory. The benchmarks for the Arivis software were performed on a HIVE system running Windows Server 2012 R2 with two Intel Xenon E5-2640v3 CPUs and 256 GB memory. The latest stable version of each stitching program was used. BigStitcher datasets were stitched using 4$\times$ times (x,y) downsampling, and fusion was performed at stated downsampling levels. Correctness of the stitching could only be confirmed in the BigStitcher due to the flexibility of interactive inspection. Processing of multi-view, dual-illumination datasets as well as ICP refinement and virtual fusion is only possible in BigStitcher. Note that fusion in BigStitcher also performs intensity adjustment. All displayed values are averaged from three independent runs of each respective software except those marked with a black star. **Note that image fusion results are not comparable since all other applications fuse datasets using translation-only, which is a significantly simpler problem that cannot align the datasets sufficiently well while BigStitcher uses affine models. Therefore translation-only results on dual-illumination datasets are grayed and marked with a red star.** A cross indicates that the functionality is not supported by the software, 'n.a.' indicates that the dataset did not require this feature.

SUPPLEMENTARY NOTES

1. Sample clearing

Clearing of brain tissue was performed using the CLARITY protocol.[1] Mice were deeply anesthetized by intraperitoneal injection of 100 mg/kg Ketamine and 15 mg/kg Xylazine. Mice were exsanguinated by transcardial perfusion with 25 ml cold PBS followed by whole body perfusion with 25 ml cold monomer solution (4% v/v acrylamide, 4% w/v Paraformaldehyde (PFA), 0.25% w/v VA-044 in PBS). The brains were collected and fixed in monomer solution for 2 more days. Next, the whole brains were placed in fresh monomer solution and oxygen was removed from the tubes by vacuum and flushing the tube with nitrogen gas for 15 minutes. The samples were then polymerized by placing the tubes in a 37°C water bath under gentle shaking for 2 hours. Polymerized brains were placed in clearing solution (4% SDS in 200 mM Boric acid). Brains were incubated in clearing solution for 1 week at 37°C with daily solution change. Then, the brains were actively cleared using the X-Clarity setup from Logos Bioscience for 24 hours with a current of 1 A at 37°C. Cleared brains were washed twice overnight with 0.1% v/v Triton X-100 in PBS and once with PBS.

2. Expansion microscopy (ExM)

Expansion microscopy sample preparation

For Expansion Microscopy (ExM), the nervous system of a 1st instar Drosophila larva of was extracted, fixed in 4% PFA/1xPBS/0.1%Triton for 1 hour and washed 2x10 min in 1xPBS/0.1% Triton. Before each antibody usage, the nervous system and the antibodies were blocked in 5% goat serum/1xPBS/0.1% Triton for one hour. Following the blocking, the nervous system was incubated overnight at 4°C in 1:500 monoclonal Anti-α-Tubulin antibody produced in mouse (Sigma Aldrich T6199 1mg/ml). After 5x10 min washing (1xPBS/0.1% Triton), the secondary antibody 1:250 Anti-Mouse CF^{TM}488A antibody produced in goat (Sigma Aldrich AB4600387 2mg/ml) was added overnight at 4°C.

Detailed expansion microscopy protocol

Acryloyl-X, SE (6-((acryloyl)amino)hexanoic acid, succinimidyl ester; here abbreviated AcX; Thermo-Fisher) was resuspended in anhydrous DMSO at a concentration of 10 mg/mL, aliquoted and stored frozen in a desiccated environment. AcX stock solution was diluted in 1xPBS to a final concentration of 0.1 mg/mL AcX. Specimens were incubated in this 0.1mg/mL AcX solution for $>$6 h, at RT. Monomer solution (1xPBS, 1M NaCl, 1.84M sodium acrylate, 0.35M acrylamide, 3.2mM N,N'-methylenebisacrylamide) was mixed, frozen in aliquots, thawed fully, vortexed, and cooled to 4°C before use. Concentrated stocks of the initiator ammonium persulfate (APS, 10% w/w), accelerator tetramethylethylenediamine (TEMED, 10% v/w) and inhibitor 4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxyl (4-HT, 0.5% w/w) were prepared as concentrated stock solutions, which were frozen in aliquots and then fully thawed and vortexed before use. Initiator, accelerator and inhibitor stock solutions were added to the monomer solution at a ratio of 2uL each per 94uL monomer solution to produce complete monomer solution. Specimens were washed 2x15min in complete monomer solution, on ice with shaking. Specimens were transferred to 3D-printed gelation chambers sized 1cm x 1cm and 0.3mm deep, along with 30uL of complete monomer solution. Chamber was covered with a cover glass and transferred to a humidified 37°C incubator for 2hr for gelation and gel curing.

Proteinase K (New England Biolabs) was diluted 1:100 to 8 units/mL in digestion buffer (50 mM Tris (pH 8), 1 mM EDTA, 0.5% Triton X-100, 1 M NaCl) to produce proteinase solution. Gel was recovered from chamber and incubated fully immersed in proteinase solution overnight at RT, with shaking. The digested gel was next incubated in at least a 10-fold excess volume of monomer solution with accelerator and inhibitor (no initiator) 2x15min, followed by complete monomer solution 2x15min on ice, with shaking. (Initiator is omitted from the first two washes to prevent premature gelation.) During incubation, a glass slide and cover glass are coated with parafilm by laying parafilm with paper backing onto the glass surface (parafilm down) and scraping a razor blade

across the backing, then removing the backing. A 5mL syringe filled with silicone grease was used to apply four dabs (approx. 50uL each) of grease to the glass slide, at the corners of a rectangle slightly smaller than the cover glass. The gel was transferred to the coated slide and excess fluid was removed. The cover glass was placed over the gel, parafilm side down, contacting the dabs of grease. The cover glass was gently pressed down, squeezing the grease, until the coverglass contacted and sat flat across the gel. Excess complete monomer solution was backfilled into the resulting chamber to impede access of atmospheric oxygen to the gel. The completed chamber was moved to the 37°C humidified incubator for 2hr for gelation and curing.

The resulting doubly-gelled specimen was recovered from the chamber, and excess gel was trimmed away. The trimmed double gel was washed in excess volumes of doubly deionized water for 0.25-2 h to expand. This step was repeated 3-5 times in fresh water, until the size of the expanding sample plateaued.

3. Imaging strategies

3D images of cleared mouse brains were imaged using the Zeiss Lightsheet Z.1 microscope. Each sample was attached to the sample holder using a cyanoacrylate-based glue. The mounted sample was placed in the FocusClear pre-filled imaging chamber. Images were acquired using the EC Plan-NEOFLUAR 5×/NA 0.16 objective together with the LSFM 5×/NA 0.1 illumination objectives on a Zeiss Light-sheet Z.1. The data was acquired using dual side illumination and from different angles. Images were collected with two 1920×1920 pixel sCMOS cameras and stored in the Zeiss CZI file format.

Fixed C. elegans dauer larvae were embedded in 1.2% agarose containing fluorescent beads and imaged using the same microscope in a water-filled sample chamber. Imaging was performed using the 20×/NA 1.0 objective with additional 2× zoom.

3D images from a cleared and expanded central nervous system of a Drosophila 1st instar larva were acquired using an IsoView light-sheet microscope[2] that has been modified for multi-tile acquisition. To prepare the sample for imaging, excess gel surrounding the expanded sample was removed using a scalpel, leaving four flat and smooth gel surfaces for imaging. Some extra gel was left underneath the sample for mounting, and the sample was affixed to a cylindrical post using a cyanoacrylate-based glue. The mounted sample was placed in the imaging chamber filled with deionized water. Orthogonal views for each tile of the sample were acquired sequentially by switching the illumination and detection orders in IsoView. Images were acquired using SpecialOptics 16×/NA 0.71 objectives and recorded using full frames (2048×2048 pixels, pixel pitch of 0.4125 μm in sample space) of Orca Flash 4.0 v2 sCMOS cameras. The sample was held stationary during multi-view acquisition of each tile, and depth-sectioned images were acquired every 0.8125 μm by translating the detection piezos over a range of 750 μm. A tile for each view thus covered a field of 832 μm (X), 832 μm (Y), and 750 μm (Z). Automated tiling across the entire sample was achieved by moving the sample in predefined steps of 600 μm in X, Y, and Z until full coverage was achieved. Bi-directional light-sheet illumination was achieved using opposing SpecialOptics objectives and the illumination NA was chosen to be 0.0315 for a confocal parameter of approximately 416 μm. The light-sheets from opposing arms were shifted approximately by their Rayleigh length (208 μm) toward the illumination objectives so that each light-sheet provided uniform coverage of the respective half in the acquired image.

A summary of the most important acquisition parameters can be found in **Supplementary Table 1**.

4. Data import

Import of data

Microscopy acquisitions are saved in a multitude of vendor-specific formats, custom formats, and general formats such as TIFF stacks. We developed an extendable, user-friendly interface that automatically imports almost any format and extracts relevant metadata such as illumination directions, sample rotation, and approximate image positions using Bioformats.[3] The assignment of attributes to the image stacks in the raw data is usually automatic, or can be achieved with minimal interaction from the users. Therefore, the importer supports

interactive placement of image tiles using regular grids or text file-based definitions (**Supplementary Fig. 5**). BigStitcher accesses image data through memory-cached, virtual loading,[4] optionally combined with virtual flat-field correction (**Supplementary Fig. 6** and **Supplementary Note 5**). Performance is optimal when images are stored using a multiresolution, blocked, compressed format enabling interactive visualization, processing and interaction with terabyte-sized image datasets. The importer therefore suggests by default to resave single-block images (e.g. TIFF) into the BigDataViewer HDF5 format.[5] Alternatively, by making use of cached ImgLib2 data structures, we support virtual loading of image planes from the raw files including caching of already loaded planes.

Data import is described in detail on the BigStitcher Wiki `https://imagej.net/BigStitcher_Define_new_dataset`. We additionally added an example youtube video that illustrates how the most generic import from TIFF stacks works in BigStitcher: `https://youtu.be/aUofNP6V0lg`. In case direct import from a custom format fails, we therefore suggest to manually re-save data as TIFF stacks and subsequently importing them into BigStitcher. It is important to preserve the calibration of the image stacks in the process.

`SpimData` data format

We internally represent our image data and metadata using an extended version of the `SpimData` data format of BigDataViewer.[5] Each image stack is defined by a (`ViewSetup`, `TimePoint`)-combination. We extend the format by giving each `ViewSetup` the following *attributes*: `Channel` to represent color channels, `Illumination` to represent illumination directions, `Angle` to represent multi-view acquisition angles and finally `Tile`, representing (local) x,y points in a multipoint acquisition.

In addition to those attributes, we store detected interest points, bounding boxes (named sub-volumes in which we can *fuse* or deconvolve images), point spread functions for deconvolution and pairwise registrations (that have yet to be used in global optimization) for each (`ViewSetup`, `TimePoint`) *view*. For each image stack, we also store its *registration* (i.e. the transformation from pixel to world coordinates) as a list of affine transform matrices. The registration steps described below will typically prepend another transformation matrix to this list. Finally, the `SpimData` is associated with an `ImgLoader` object that can make image pixel data available as an ImgLib2[4] `RandomAccessibleInterval` given a (`ViewSetup`, `TimePoint`) *view id*.

The `SpimData` data structure can be saved as an XML *project file*, allowing users to manually edit it with any text editor. We automatically save previous versions of the project file to provide the user with the ability to un-do registration steps.

5. Flat-field correction

Flat-field correction is the process of correcting for image artifacts due to uneven illumination or detection efficiency or fixed-pattern noise. Aside from being visually unpleasing, especially in tiled acquisitions, these artifacts can also effect image registration and downstream quantitative image analyses. We therefore offer simple on-the-fly correction for a *dark image* (which might be nonzero due to e.g. camera offset) and a *bright image* (representing uneven illumination or detection efficiency across the field-of-view). We calculate corrected pixel intensities C from a raw image R and bright and dark images B and D as:

$$C_x = \frac{(R_x - D_{x'}) * \overline{(B - D)}}{(B_{x'} - D_{x'})} \tag{1}$$

The correction images can either have the same dimensionality as the raw images, in which case $x' = x$, or have lower dimensionality (e.g. when using 2D correction images on a 3D image stack), in which case $x' = (x_1 \dots x_n)$ with n being the dimensionality of the correction images. If a dark image is not provided by the user, we assume it to have constant intensity of 0 (corresponding to no background offset). Likewise, if no bright image is provided, we assume it to have constant intensity of 1 (uniform illumination and detection efficiency).

We implemented the flat-field correction as a wrapper around an `ImgLoader`, calculating corrected pixel intensity values on-the-fly (with optional caching) every time an image is loaded. That way, the corrected

images are available for all other processing steps such as intensity-based registration, interest point detection or image fusion, but it is still possible to activate or de-activate the correction or change bright or dark images after the initial flat-field correction. A separate (bright, dark)-correction image pair can be set for every image in the dataset by modifying the XML project file, while in the GUI we offer user-friendly assignment of correction images to every (channel, illumination direction)-pair.

6. Pairwise shift calculation

In BigStitcher, we currently offer three ways of calculating shifts between a pair of images: the Fourier-based *phase correlation* algorithm, the Gradient-descent-based *Lucas-Kanade* algorithm, both intensity-based methods, as well as interest point-based alignment.

Phase correlation

By default, we calculate pairwise translational shifts of two images I_1 and I_2 using phase correlation[6,7] using our new ImgLib2 implementation.[4] In noiseless images, the method produces a phase correlation matrix (PCM) Q containing a single δ-impulse at the location corresponding to the shift between the two images. Real images might contain multiple peaks (**Supplementary Fig. 11**) and we localize the n highest peaks in Q by detecting peaks with subpixel accuracy using a n-dimensional implementation of a quadratic fit.[8] Aside from allowing subpixel-accurate registration, we can use the precision obtained from the subpixel accuracy of the phase correlation to counteract the effects of downsampling (**Supplementary Fig. 12**), allowing us to achieve registration of similar quality to full-resolution with significant performance gains (**Supplementary Fig. 13–15**). Due to the periodic nature of the Fourier shift theorem, each peak in the PCM actually correspond to 2^n possible shifts in n dimensions. We therefore test each of these candidate shifts by calculating the cross-correlation between the images I_1 and I_2, optionally with interpolation in the case of sub-pixel shifts.[8] We choose the shift vector t corresponding to the highest cross correlation as the final result after applying downsampling correction, if necessary.

It is often necessary to not only align two single images but groups of images, e.g. all channels of a tile. We therefore implemented a flexible framework for the registration of grouped images (see below). The two images I_1 and I_2 can have arbitrary affine pre-registrations such as sample rotation, correction of axial scaling, or already performed registration steps. If pre-registrations of I_1 and I_2 are identical, or are only based on different translations or axis-aligned scalings, we run the phase correlation on (downsampled) raw input images, otherwise on virtually fused images (**Supplementary Note 16**).

Lucas-Kanade

In addition to the default phase correlation-based pairwise shift calculation, we offer registration via an ImgLib2 implementation of the *inverse compositional* formulation of the gradient descent-based Lucas-Kanade optical flow algorithm.[9] While the algorithm is applicable to a variety of transformation models, we currently stick to estimating a translation vector t. If the pairwise registration converges, we calculate the cross correlation of the overlapping portions of the images as a quality metric for the pairwise registration.

Intensity-based registration of grouped images

In many use cases, one might want to align not single images but groups of images, e.g. all channels of a tile, in the pairwise registration step. For this, we implemented a flexible framework for the registration of grouped images.

Each attribute of the images can be set to be an *axis of application*, an *axis of comparison* or an *axis of grouping*. The registration will proceed by first splitting the images by the application attributes, i.e. grouping all images that have the same value for these attributes. In each group, the images are then split by the comparison attributes and finally, the remaining image groups (that differ only in the grouping attributes) are combined into one image stack by either averaging all images for each grouping attribute or picking the image with a specific instance of the attribute.

In a typical application, the stitching of tiled datasets, we would, for example, start by *applying* the registration to all (`Angle`, `TimePoint`)-combinations individually, *comparing* by `Tiles` and finally *grouping* by `Illumination` and `Channel` for each tile, e.g. by averaging illumination directions and picking a specific channel.

Intensity-based registration of images with pre-registrations

The two images I_1 and I_2 can have arbitrary pre-registrations, i.e. pixel coordinates x_{px} are mapped to world coordinates x_w via the affine transforms $x_{w,I_1} = A_{I_1} x_{px,I_1} + b_{I_1}$ and $x_{w,I_2} = A_{I_2} x_{px,I_2} + b_{I_2}$. Depending on the values of A_{I_1} and A_{I_2}, we consider two cases: If they are equal, i.e. the pre-registrations differ only by a translation, we perform the shift calculation on the raw pixel data of the overlapping volume to get a shift vector t for I_2 in pixel coordinates. The transformation in world coordinates is then given by $R\left(\begin{smallmatrix} I & t \\ 0 \cdots 0 & 1 \end{smallmatrix}\right)R^{-1}$ with $R = \left(\begin{smallmatrix} A_{I_2} & b_{I_2} \\ 0 \cdots 0 & 1 \end{smallmatrix}\right)$. If the pre-registrations differ in more than just translation, we create virtually transformed images of the smallest rectangular bounding box enclosing the overlapping volume and use them as input to the registration. As the virtual input images are already in world coordinates in this case, the resulting transformation matrix for I_2 is simply $\left(\begin{smallmatrix} I & t \\ 0 \cdots 0 & 1 \end{smallmatrix}\right)$

Interest-point based

For interest-point based pairwise registration, we detect local extrema in either Difference-of-Gaussian or Difference-of-mean filtered images, optionally followed by subpixel refinement of the detections via a quadratic fit.[8] If we are registering a pair of image *groups*, the interest points of each image in the group are pooled, with optional replacement of point clusters within a user-defined radius by their center.

For each image, we apply the current (affine) registrations to the pixel-coordinate interest points and then determine *candidate point matches* via descriptor matching.[10,11] We then perform model-based outlier removal via the RANSAC algorithm,[12] yielding a set of *inlier point pairs*, $C_{inliers}$, and an optimal translation t for I_2, minimizing $\sum_{(ip_1,ip_2)\in C_{inliers}} ||ip_1 - ip_2 - t||^2$

7. Quantification of pairwise image stitching using downsampling

To assess the effect of downsampling on the pairwise stitching we use simulations of spheroid-like objects at different signal-to-noise ratios (SNRs) as ground truth. We create realistic images by mimicking image creation in light-sheet microscopy including optical sectioning, 3-fold anisotropy between xy and z, light attenuation, convolution, and pixel intensity generation using Poisson processes.[13] Importantly, pairs of overlapping images that we use for benchmarking the subpixel phase correlation method are created using different Poisson processes and are additionally rendered with half a pixel offset of the full resolution images to avoid nearly identical overlaps at high SNRs due to the simulation process (**Supplementary Fig. 12**).

We simulate 300 pairwise overlaps, each at SNRs ranging from 1 to 32, and lateral downsamplings ranging from 1× to 8×, where axial downsampling is matched as good as possible to achieve near-isotropic resolution as in the actual software. We illustrate that across SNRs downsampled images yield a constant registration quality, which even exceeds that of registration at full resolution for low SNRs. This is achieved through a combination of the smoothing effect during downsampling (**Supplementary Fig. 12**) and precise subpixel-localization (**Supplementary Fig. 13–15**). Due to the smoothing effect, registration quality therefore initially increases at 2-fold and 4-fold downsampling (**Supplementary Fig. 13–15**), while when using more downsampling, the loss of pixel resolution outweighs the effect of smoothing and hence the quality drops. Registrations with a constant quality of an average error of below one pixel can be computed at a fraction of the computing time compared to full resolution, typically 4 - 115 times faster. Existing outliers are filtered during global optimization and overall registration quality can further be improved during the ICP[14]refinement step.

Simulated data was created using the $net.preibisch.stitcher.headless.StitchingPairwise$ class in the BigStitcher package (release version 0.3.3). Since it is a Maven artifact, the versions of all dependencies are defined and the corresponding version can be built automatically from that source code state (https://github.com/PreibischLab/BigStitcher/commit/0d7f79a59ab15fb1805157ab72c5bc9802b02fbd).

8. MultiView Registration

For MultiView registration, i.e. alignment of image taken from different angles (or also time series stabilization), we first detect interest points in the individual images as described above (6). Images may be grouped (and are by default if we are, e.g. registering tiled acquisitions from multiple angles for which we already aligned the tiles via an intensity-based method) according to their attributes, by pooling their interest points and optionally merging clusters of interest points.

Pairwise point correspondences can either be established by geometric local descriptor matching, a modified version of the iterative closest point (ICP[14]) algorithm or by simply matching the center of mass of the point clouds of both images (note that in this case the registration will be constrained to be a translation). Using the link graph (V, C) and pairwise point correspondences P_{ij} established thus, we calculate the final registration by performing global optimization as described above (12), optionally with iterative link removal and a second round to preserve metadata.

Geometric Local Descriptor Matching

To identify corresponding interest points in between two point clouds, geometric local descriptor matching combined with random sample consensus (RANSAC[12]) has been proven to be a powerful technique.[10,11] The basic idea to express each interest point as a geometric constellation using its n (typically three) nearest neighboring interest points. The vector difference between two descriptors then describes how similar the local area of two points is. A geometric local descriptor (GLD) is assumed to be a correspondence candidate if it is at least m (typically one to ten) times more similar than the second most similar GLD.[8] True corresponding interest points between two point clouds are finally identified using RANSAC on a regularized affine transformation model. However, fast GLD matching using the rotation-invariant technique based on geometric hashing[10] requires relatively randomly distributed points to robustly identify correspondences, while the non-accelerated, redundant, translation-invariant counterpart[11] identifies correspondences reliably in non-rotated point clouds of only up to a few thousand points in reasonable time. Here, we developed a new matching procedure by extending both techniques to better suit the requirements when attempting to identify corresponding interest point in between point clouds of prior unknow size derived from imaged structures that are potentially rotated relative to each other.

Redundancy is a powerful mechanism for GLD matching. It uses additional nearest neighbors but excludes some of them sequentially during matching making it more robust to potentially mis-detected interest points.[11] We therefore extend the fast rotation-invariant technique based on geometric hashing[10] with the capability for redundancy. This significantly increases the chance of being able to align randomly oriented point clouds very fast, albeit at low inlier ratios (ratio of true correspondences to total number of correspondence candidates). Rotation invariance is not desired if both point clouds are known to be approximately in same orientation, for example if the rotation of the sample performed by the microscope was known and has been applied to the dataset. Checking for potential rotations simply increases the chance for wrong correspondence candidates. We therefore implemented a fast translation-invariant GLD based on geometric hashing that supports redundancy. All four versions of GLD are available in BigStitcher to enable robust multi-view alignment.

9. Quantification of overall registration quality

To quantify registration quality, we acquired an as-small-as-possible (166GB), cleared section of an adult mouse brain. It is imaged at lower magnification from two angles (0° and 180°) and in a 2×3 tile configuration with dual-sided illumination for each angle (**Fig. 2l,m**, **Supplementary Fig. 2,17** and **Supplementary Table 1**).

We identified a ground-truth set of corresponding interest points in directly adjacent images by manually selecting bright spots from a set of interest points that were automatically detected using Difference-of-Gaussian filtering and subpixel-accurate local maxima determination (Supplementary Fig. 17a,b). For each image pair, we selected between 19 and 52 corresponding points, in total 692.

We then registered the dataset in BigStitcher for tiled acquisition only, tiled acquisition across illumination directions, and for the multi-tile, dual illumination, multi-view case. This is achieved by grouping the images either by angle and illumination direction, just by angle, or not at all. For the single-view cases, we performed translational alignment by stitching the images using phase correlation. For an all-to-all registration with a translation model, the images of angle 2 were manually rotated by 180° and then all images were aligned using interest points by fast translation-invariant GLD matching followed by RANSAC[12] and global optimization using a translation model. All translation-model alignments were refined using ICP[14] as described above. The point correspondences determined during ICP were further used for non-rigid refinement.

For virtual re-blocking, each original image was split into 2×2×2 sub-blocks (with 120px overlap in xy and 100px overlap in z). After the re-blocking, 4–28 manually selected point correspondences remained between each set of directly adjacent

blocks. For each of the image groupings and registration models used, we calculated an average error of the manually selected point correspondences:

$$e_{avg} = \frac{1}{|I|} \sum_{i_1 \in I} \left(\frac{1}{|C(i_1)|} \sum_{i_2 \in C(i_1)} \left(\frac{1}{|PM(i_1, i_2)|} \sum_{(p_1, p_2) \in PM(i_1, i_2)} ||T^{i_1}(p_1) - T^{i_2}(p_2)||^2 \right) \right) \quad (2)$$

With I being the set of images, $C(i)$ the adjacent images of an image i (ignoring diagonal pairs for which no corresponding interest points were manually selected as well as pairs that are not in the same group, e.g. when grouping by angle and illumination direction, and pairs from the same original image in the virtually blocked dataset), $PM(i, j)$ the corresponding manually selected interest points of images i, j and T_i the transformation of image i.

To estimate the lowest theoretically achievable errors given a certain transformation, we use only the manually selected point correspondences to calculate a globally optimal registration (and optionally the non-rigid refinement thereof) of the images and then calculate the average error from the same point correspondences as described above. In the virtually blocked case, we also use manually selected point correspondences (826 in total) between adjacent blocks within the same original image for the registration (but ignore them for the final error calculation).

Despite relatively small aberrations in this sample as compared to entire mouse brains (compare with **Fig. 2b,c,d** and **Fig. 2k**) we illustrate that using only translation as transformation model is only reasonable for tiled acquisitions that do not include multiple illuminations or multiple acquisition angles, yet even there spherical aberrations persist that question the standard use of translation models in general (**Supplementary Fig. 4,9**). The alignment errors increase when aligning across illumination directions and greatly increase when aligning different acquisition angles. Importantly, please note that the alignment quality across different illumination directions is significantly reduced on larger samples when using only translation models (compare **Fig. 2b,c,d** and **Fig. 2k**). Using the affine, split-affine or non-rigid registration functionality, BigStitcher can sharply reduce the registration errors in large cleared and expanded samples. As a trade-off between speed and quality we usually choose affine or split-affine registrations.

10. Multiview deconvolution

In addition to real-time image fusion, we offer deconvolution of bounding-box-defined volumes using a multi-view formulation of the iterative Richardson-Lucy deconvolution algorithm[15,16] with Tikhonov regularization[17] and various optimizations.[13] The PSFs required for deconvolution can be extracted from interest points detected in the images (e.g. when subdiffraction fluorescent beads were incorporated with the sample, see section 11) or supplied as TIFF stacks with odd dimensions by the user. BigStitcher offers GPU acceleration of the deconvolution on CUDA-capable Nvidia GPUs.

To allow deconvolution of multi-tile views, we extended the original deconvolution[13] to be based on the virtual fusion. Thereby, any number of input image tiles are virtually fused and serve as one of input views for the multi-view deconvolution. Proper multi-view deconvolution of partly overlapping samples requires sophisticated weight normalization in between views,[13] which we implemented to be computed virtually. Since also the input views are also virtually loaded, the memory requirement of the deconvolution solely depends on the output image size and shows a significantly increased memory-efficiency. All virtual inputs and weights are additionally cached, ensuring highest-possible processing performance for systems with large amounts of RAM.

11. PSF measurement and PSF extraction

In light-sheet microscopy, measured PSFs often differ significantly from simulated ones due to variable precision of light-sheet alignment in every experiment. Therefore, light-sheet deconvolution usually relies on the extraction of PSFs from the actual experiment19,25. To be able to perform PSF extraction in cleared tissue we developed a new protocol. Estapor Fluorescent Microspheres (F-XC 030) were diluted 1:20000 in monomer solution containing bis-acrylamide (0,05% v/v bis-acrylamide, 4% v/v acrylamide, 4% w/v Paraformaldehyde (PFA), 0.25% w/v VA-044 in PBS). The monomer solution was polymerized under constant vacuum and shaking at 37°C for 2 hours. The formed hydrogel was incubated in FocusClear overnight and imaged using the Zeiss Lightsheet Z.1 microscope with the same experimental settings used to acquire previous samples. For C. elegans dauer imaging fixed larvae were embedded in 1.2% agarose together with Estapor Fluorescent Microspheres (F-Z 030), diluted 1:2000. For ExM data acquired on the IsoView microscope depth-sectioned images (0.4125 μm step) of fluorescent beads (200nm diameter) embedded in 0.6% low-melting-temperature agarose were imaged using the same experimental settings as for sample imaging. For all samples, PSFs were extracted by detecting interest points in the acquired bead images. Potential bead aggregates were excluded by manual removal on the maximum intensity projection using the BigStitcher module “Manage Interest Points > Remove Interactively”.

12. Global optimization

Estimation of globally optimal transformations

The pairwise registration step results in *links* between image (groups) V (note that since we do not use the actual image *content* in the global optimization, we will refer to the images by their integer *id* in this section: $V \subset \mathbb{N}$). The links can be either in the form of pairwise transformations T^p (such that coordinates x from two images V_i and V_j can be transformed according to $T^p_{ij}(x_j) = x_i$) or *point correspondences* PM from which such transformations can be estimated. The pairwise registrations thus form a *link graph* (V, C) with edges $C = \{(i,j) \in V \times V | T^p_{ij} \in T^p\}$ between image pairs for which we could determine pairwise transformations. Simply traversing a spanning tree of the link graph and propagating the pairwise transformations can lead to the compounding of pairwise registration errors, even if the traversal is done along a *minimal* spanning tree determined according to some quality metric q_{ij}, e.g. cross-correlation, of the pairwise registrations.

We thus make use of an algorithm for globally optimal registration by iterative minimization of square displacement of point correspondences[18,19] for reaching a reasonable consensus in this case. This point match-based framework allows for flexible grouping and fixing of images, is applicable to, among others, time series-, chromatic channel- or view-registration and can easily be adapted to incorporate the pairwise transformations from e.g. phase correlation. The algorithm is agnostic of the transformation model (e.g. translation, affine transform,...), with the only requirement being that the model parameters can be estimated by a least-squares fit from point correspondences.

We determine the globally optimal registrations R given the image (groups) V, pairwise links C, pairwise n-dimensional point matches PM with $PM_{ij} \subset \mathbb{R}^n \times \mathbb{R}^n$ and a set of fixed views $F \subseteq V$ by minimizing:

$$\underset{R \setminus \{R_i | V_i \in F\}}{\arg\min} \sum_{(i,j) \in C} \left(\sum_{(x_k, y_k) \in PM_{ij}} ||R_i(x_k) - R_j(y_k)||^2 \right) \tag{3}$$

Note that for all fixed views, the registration will be constrained to be the identity transformation I: $\forall V_i \in F : R_i = I$.

Global optimization given pairwise transformations

The intensity-based pairwise shift calculations do not directly give us the point correspondences we need for the global optimization step, instead the results are pairwise transformations T^p in the form of affine transform matrices. We can, however, easily construct point correspondences by taking a set of points and transforming them with the *inverse* transform (the only requirement being that the n-dimensional points do not all lie in a subspace of lower dimensionality of $\mathbb{R}^n$).

Using the 3-dimensional pairwise transformations T^p ($T^p_{ij}(x_j) = x_i$) between two image (groups) V_i and V_j given their existing registrations R^{meta}, we use the 8-point approximate bounding box of their overlapping region BB_{ij} to construct the point correspondences: $PM_{ij} = \{\left(bb_k, (T^p_{ij})^{-1}(bb_k)\right) | bb_k \in BB_{ij}\}$. We can then determine the globally optimal registrations R by performing the minimization described above (3).

Global optimization with iterative link dropping

Once the global optimization terminates due to convergence or exceeding of the maximum number of iterations, we can calculate the *error* of the individual images as the average displacement of all interest points in an image to their point matches:

$$e_i = \frac{\sum_{\{j:(i,j) \in C\}} \sum_{(x_k, y_k) \in PM_{ij}} ||R_i(x_k) - R_j(y_k)||}{\sum_{\{j:(i,j) \in C\}} |PM_{ij}|} \tag{4}$$

If the link graph (V, C^n) contains links with contradicting point correspondences, stopping after one round of global optimization might leave us with unsatisfying results. In the *iterative* version of the global optimization, we therefore check that both the average error of all images and the ratio of maximal and average error fall below a user-defined threshold. If these conditions are not yet met, we will proceed to iteratively remove disagreeing links from the link graph and repeat the global optimization. To do this, we first determine the link with the highest error by maximizing:

$$c_{worst} = \underset{(i,j)}{\arg\max} \max_{(x_k, y_k) \in PM_{ij}} \left((1 - q_{ij})^2 \sqrt{d_{ijk}} \log_{10} \left(\max \left(deg(i), \deg(j) \right) \right) \right) \tag{5}$$

with d_{ijk} denoting the distance of the k'th point match of the link (i,j), $d_{ijk} = ||R_i(x_k) - R_j(y_k)||$, $\deg(i)$ denoting the degree (number of neighbors) of an image V_i in the link graph and q_{ij} being a *quality metric* $\in (0,1)$ of the link, e.g. 0-truncated cross correlation. We then remove the worst link from the links ($C^{n+1} \leftarrow C^n \setminus c_{worst}$) and repeat the optimization step 3 with the new link graph (V, C^{n+1}). The whole process is repeated until the errors fall below a user-defined threshold (in the worst case, links will be dropped until we end up with *spanning trees* of the connected components in the link graph).

Two-round global optimization using metadata

If some cases, the link graph might contain multiple connected components, e.g. in datasets from screening applications, where the actual sample only occupies isolated "islands" and most images contain only background. In this case, we can only reliably determine pairwise transformations within the connected components and align images within the components in the global optimization step. We might, however, have reasonable registrations R^{meta} from metadata and wish to keep as closely as possible to those if we do not have *strong* links.

For this, we developed a *two-round* version of the global optimization. In the first round, we determine registrations R^{strong} as described above, using the graph of *strong* links, i.e. links that are backed by pairwise transformations. In the second round, we determine the connected components in the (V, C^{strong}) graph and a mapping $CC : \mathbb{N} \rightarrow \mathbb{N}$ from image (group) indices to connected component indices as well as *weak links* $C^{weak} = \{(i,j) \in V \times V | CC(i) \neq CC(j)\}$ between images in different components. We then determine transformations R^{cc} for each connected component not containing a fixed image by minimizing:

$$\underset{R^{cc} \setminus \{r_i^{cc} \in R^{cc} | CC_i \cap F \neq \emptyset\}}{\arg\min} \sum_{(i,j) \in C^{weak}} \sum_{bb_k \in BB_{ij}} ||R^{cc}_{CC(i)}\big(R_i^{strong}(bb_k)\big) - R^{cc}_{CC(j)}\big(R_j^{strong}(bb_k)\big)||^2 \tag{6}$$

Note that we use the corners bb_k of the bounding box BB_{ij} of the overlapping volume of two images V_i and V_j as the point correspondences. The overlap is determined according to the metadata transformations R^{meta} and we essentially try to "un-do" the registrations of the first round as well as possible (while keeping the registrations *within* the connected components). The final transformations R are the concatenation of the registrations within the connected components with the relative transformations of the connected components: $R_i \leftarrow R^{cc}_{CC(i)} R_i^{strong}$.

13. Simulation of light propagation in tissue using raytracing

To describe the scene we will simulate we use two phantom images of the same size that separately define the visible light image (corresponding to fluorescent probe distribution) and the refractive indices map (**Fig. 2e,f**). We deliberately embed the spheroid-like object of varying refractive index ($R_i = 1.1$–1.21) within a dense, invisible material with high refractive index ($R_i = 1.1$) surrounded by air ($R_i = 1.0$) to recapitulate significant aberrations in the illumination light path using a relatively small simulation volume of 289×289×289px. The object simulations are implemented in the multiview-simulations package.[13]

We virtually scan a concave lightsheet (diameter of 1 pixel in the center, and 3 pixels at the edge) in 1-pixel steps and alternating left and right illumination through the sample (**Fig. 2g**), simulating an entire volume for each lightsheet position and direction (**Fig. 2h** and **Supplementary Video 2**). Therefore, we send 200.000 individual rays originating from random positions within the concave lightsheet through the sample for each lightsheet simulation. The initial vector of each ray points approximately along the lightsheet illumination direction and moves in 1-pixel steps through the volume. After each move we locally compute the Eigenvector of the largest Eigenvalue using the refractive index map, which defines the normal vector of the refractive surface at the current, sub-pixel ray position. Using this estimated refraction surface, we compute the refraction angle using raytracing algebra,[20] update the ray vector accordingly, and add a Gaussian distribution with an intensity defined by the visible light image to the simulation volume. For simplicity we ignore total reflection since it is mostly caused by numerical instabilities. We confirmed correct refraction of rays based on our computation of local Eigenvectors in discrete pixel-images by comparing it to refraction of rays in the corresponding continuous, parametric description of the same scene (not shown).

The result of these simulations are 578 3d-volumes that recapitulate the principles of dual-sided lightsheet illumination (**Fig. 2g**). Inspired by classical raytracing, we perform a simplified detection simulation and therefore invert the ray path and only modulate signal intensity as a function of distance from the focal plane. Per camera pixel (289×289) we send 500 rays at random positions within each pixel into the scene that are refracted as described above. For detection, we use the same the same refractive index map, and the result of each respective lightsheet illumination simulation serves as image

data. However, since we assume an extremely high refractive index mismatch for illumination simulation to recapitulate the behavior in large samples, we assume a lower refractive index mismatch for the embedding material ($R_i = 1.01$) to acquire reasonably distorted images. The relative refractive index mismatch within the spheroid-like object is conserved ($R_i = 1.01$–1.11). We assume the focal point of the objective to lie in the center of the currently simulated lightsheet position. The light captured by each ray on its path through the sample is then computed as the sum of all light integrated when traveling through the sample, at each ray location gaussian-weighted ($\sigma = 3.5$) by the distance to the expected lightsheet position. The simulations were performed in parallel on the local compute cluster at the MDC.

Simulated data was created using the $net.preibisch.simulation.SimulateMultiViewAberrations$ class in the multiview-simulation package (release version 0.2.2). Since it is a Maven artifact, the versions of all dependencies are defined and the corresponding version can be built automatically from that source code state (https://github.com/PreibischLab/multiview-simulation/commit/b41b74cce9287f804b670d7de3396605446818a8).

14. Non-rigid transformation

The underlying principle of moving least squares[21] is to non-rigidly transform images using a set of corresponding points. Therefore, a local transformation is computed for each pixel using a distance-weighted fit of all corresponding points ensuring smoothness. In BigStitcher, corresponding points are a direct result of all interest point-based registration algorithms. To provide a sufficient amount of corresponding interest points, it is yet most useful to derive them using ICP.[14] Regularization is achieved on the registration side as corresponding interest points are identified on a regularized affine transformation model either using RANSAC[12] or ICP,[14] which both specify a maximum error. This ensures that corresponding points cannot diverge more than this specified error from the regularized affine transformation of each image tile. In combination with virtual re-blocking, this error can be limited to smaller regions than the acquired, physical tiles.

When computing local transformations for each image, it is necessary to ensure smoothness across n overlapping images by defining appropriate point correspondences. However, the registration identifies only pairwise correspondences in between pairs of images. From those, we therefore first identify all unique interest points across all images defined by all pairwise correspondences (**Supplementary Fig. 23**). The location of each unique point is then determined by averaging the locations of all contributing interest points after applying their respective affine transformations. Thereby, the non-rigid transformation only accounts for the remaining error after affine alignment. For each image, corresponding points required for moving least squares are then subsequently defined between the unique point and the corresponding interest point of the transformed image only.

15. Quality estimation using relative Fourier Ring Correlation

For computing the 2d-Fourier Ring Correlation[22] we adapted methods from the FRC ImageJ plugin[23] as outlined in the **Online Methods**.

16. Image fusion

We *fuse* multiple images by performing a weighted average of the raw images I^{raw} transformed by their registrations R. Each raw image I_i^{raw} has a set weight images W_i. For example, we allow the user to weigh the images with a cosine-shaped fade-out, de-emphasizing the artifact-prone border regions of the individual images, as well as by the approximate local entropy, to emphasize images with sharper structures in overlapping regions. Since the raw images will be evaluated at non-integer coordinates, we offer the choice between nearest-neighbor and linear interpolation. Downsampling can easily be achieved by prepending a scaling transformation to each of the registrations R. The intensity of the fused volume at a coordinate x is given by:

$$I^{fused}(x) = \frac{\sum_{I_i^{raw} \in I^{raw}} \left(I_i^{raw}(R_i^{-1}(x)) * \prod_{w_j \in W_i} w_j(R_i^{-1}(x)) \right)}{\sum_{I_i^{raw} \in I^{raw}} \left(\prod_{w_j \in W_i} w_j(R_i^{-1}(x)) \right)} \tag{7}$$

In practice, we evaluate I^{fused} only at integer coordinates of a user-defined *bounding box*. We implemented the image fusion to perform all calculations virtually on-the-fly, with caching of previously computed planes using imglib2-cache. This allows the quick inspection of fusion results as well as creation and planewise saving of images that might exceed the RAM available to the user.

17. Brightness and contrast adjustment

Even after correcting for fixed-pattern noise (5), differences in brightness and contrast between images, e.g. due to bleaching, might persist and be visible in the fused images. To correct for this, we estimate optimal linear transforms of pixel intensities in adjacent images[24] to achieve uniform brightness and contrast in the whole dataset. We minimize the intensity difference of all pixels in the overlapping volume O_{AB} of two images I_A, I_B (with corresponding coordinates (x_A, x_B) according to the current registrations):

$$\underset{\alpha,\beta}{\arg\min} \sum_{I_A \in I} \left(\sum_{I_B \in I \setminus I_A} \left(\sum_{(x_A, x_B) \in O_{AB}} \left(I_B(x_B) - \left[\alpha^{I_A} I_A(x_A) + \beta^{I_A} \right] \right)^2 \right) \right) \tag{8}$$

Since this is equal to one-dimensional point correspondence registration, we can make use of the same iterative optimization algorithm used for image registration (12). To reduce influence of noise and computational costs, we use (precomputed) downsampled versions of the images for the optimization. A problem with unconstrained optimization is the possibility of convergence to the trivial solution of setting all pixel intensities to zero. We therefore formulate the linear transform $I(x) * \alpha + \beta$ as a weighted average between a linear transform, an additive transform and the identity transform:

$$\alpha I(x) + \beta = \lambda_1 * (\alpha' I(x) + \beta_1) + \lambda_2 * (I(x) + \beta_2) + \lambda_3 * I(x) \tag{9}$$

with user-definable regularization parameters $\lambda_1, \lambda_2, \lambda_3 : \lambda_1 + \lambda_2 + \lambda_3 = 1$. By using nonzero λ_2, λ_3, we can constrain the optimization to not converge to the trivial solution.

The size of overlaps between image tiles can differ significantly. Therefore, intensity transformations supported by many overlapping pixels will implicitly have a higher weight, which can lead to the fact that visible intensity differences between tiles with little overlap persist. To compensate this effect we allow to balance overlaps by setting a maximal number of corresponding pixels. To ensure equal distribution of these corresponding pixels, we randomly remove pixels from the set of all pixels until the desired number is achieved.

18. Example datasets for BigStitcher

We prepared three different examples of different size and complexity for testing the BigStitcher. We suggest to run BigStitcher on these first before applying it to your dataset. This allows you to quickly test features in an environment where you can easily ask for advice on GitHub or the ImageJ Forum.

The data can be downloaded from the Open Science Foundation (a Nature recommended data repository `https://www.nature.com/sdata/policies/repositories`) at `https://osf.io/bufza/`.

2d multi-tile dataset (2.8 MB)

Maximum intensity projection of the nervous system of a Drosophila larva containing 6 tiles and 3 channels each. You can download the raw input at `http://preibischlab.mdc-berlin.de/BigStitcher/Grid_2d.zip` and a reconstructed BigStitcher project at `http://preibischlab.mdc-berlin.de/BigStitcher/Grid_2d_h5_aligned.zip`. In the reconstructed project, the images were imported into the BigStitcher using the AutoLoader (with immediate resaving as HDF5 and Movement to a regular 2-by-3 grid with 10% overlap between the tiles). We calculated pairwise shifts using phase correlation with default parameters, using the precomputed 2x2 downsampling and averaging the channels. We ignored links with correlation < 0.7 and calculated the final registration using the two-round global optimization with strict constraints.

3d multi-tile dataset (123 MB)

3d confocal scan of the nervous system of a Drosophila larva containing 6 tiles and 3 channels each, channels are distributed over different files. You can download the raw input at `http://preibischlab.mdc-berlin.de/BigStitcher/Grid_3d.zip` and the reconstructed project at `http://preibischlab.mdc-berlin.de/BigStitcher/Grid_3d_h5_aligned.zip`. In the reconstructed project, we ran the same import and reconstruction steps as for the 2d dataset and in addition performed affine refinement of the registration using IPC with default parameters and simple tile refinement to create the final reconstructed project.

We will add larger and more complex examples on the BigStitcher website `https://imagej.net/BigStitcher` and will also link videos of the alignment process from there.

19. Links to the current source codes

The BigStitcher is distributed over two projects. Both are licensed under the GPL(v2) and the source code is freely available on GitHub, at https://github.com/PreibischLab/BigStitcher and https://github.com/PreibischLab/multiview-reconstruction, respectively.

The CUDA code for accelerated interest point detection and devonvolution is available from https://github.com/StephanPreibisch/SeparableConvolutionCUDALib and https://github.com/StephanPreibisch/FourierConvolutionCUDALib, respectively.

The light simulation is a standalone software and part of the multiview-simulation package https://github.com/PreibischLab/multiview-simulation, the main class can be found here: https://github.com/PreibischLab/multiview-simulation/blob/master/src/main/java/net/preibisch/simulation/SimulateMultiViewAberrations.java. The license is also GPL(v2).

Newer versions will be hosted using GitHub, and release announcements will be done via Twitter (https://twitter.com/preibischs), on the GitHub page (https://github.com/PreibischLab/BigStitcher), and on the ImageJ wiki (http://imagej.net/BigStitcher). Releases are and will be provided to end users via the Fiji update mechanism.[25]

The following classes are the main classes for the respective codes:

BigStitcher: $net.preibisch.stitcher.plugin.BigStitcher$ (in BigStitcher)

Light simulation: $net.preibisch.simulation.SimulateMultiViewAberrations$ (in multiview-simulation)

Phase correlation simulation: $net.preibisch.stitcher.headless.StitchingPairwise$ (in BigStitcher)

20. Bug reports and feature requests

For bug reports and feature requests regarding BigStitcher please use the GitHub issue system available here: https://github.com/preibischLab/BigStitcher/issues.

21. BigStitcher user guide

The BigStitcher comes with extensive documentation that is hosted on the ImageJ wiki. The current version of the continuously updated user guide can be found at https://imagej.net/BigStitcher#Documentation.

REFERENCES

[1] Chung, K., Wallace, J., Kim, S.-Y., Kalyanasundaram, S., Andalman, A. S., Davidson, T. J., Mirzabekov, J. J., Zalocusky, K. A., Mattis, J., Denisin, A. K., Pak, S., Bernstein, H., Ramakrishnan, C., Grosenick, L., Gradinaru, V., and Deisseroth, K., "Structural and molecular interrogation of intact biological systems," *Nature* **497**(7449), 332–337 (2013).

[2] Chhetri, R. K., Amat, F., Wan, Y., Höckendorf, B., Lemon, W. C., and Keller, P. J., "Whole-animal functional and developmental imaging with isotropic spatial resolution," *Nature Methods* **12**, 1171–1178 (Oct. 2015).

[3] Linkert, M., Rueden, C. T., Allan, C., Burel, J.-M., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., MacDonald, D., et al., "Metadata matters: access to image data in the real world," *The Journal of Cell Biology* **189**(5), 777–782 (2010).

[4] Pietzsch, T., Preibisch, S., Tomančák, P., and Saalfeld, S., "ImgLib2—generic image processing in Java," *Bioinformatics* **28**(22), 3009–3011 (2012).

[5] Pietzsch, T., Saalfeld, S., Preibisch, S., and Tomancak, P., "BigDataViewer: visualization and processing for large image data sets," *Nature methods* **12**(6), 481 (2015).

[6] Preibisch, S., Saalfeld, S., and Tomancak, P., "Globally optimal stitching of tiled 3D microscopic image acquisitions," *Bioinformatics* **25**(11), 1463–1465 (2009).

[7] Kuglin, C., "The phase correlation image alignment method," *IEEE International Conference on Cybernetics and Society* (1975).

[8] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* **60**(2), 91–110 (2004).

[9] Baker, S. and Matthews, I., "Lucas-Kanade 20 years on: A unifying framework," *International Journal of Computer Vision* **56**(3), 221–255 (2004).

[10] Preibisch, S., Saalfeld, S., Schindelin, J., and Tomancak, P., "Software for bead-based registration of selective plane illumination microscopy data," *Nature Methods* **7**(6), 418 (2010).

[11] Smith, C. S., Preibisch, S., Joseph, A., Abrahamsson, S., Rieger, B., Myers, E., Singer, R. H., and Grunwald, D., "Nuclear accessibility of β-actin mRNA is measured by 3D single-molecule real-time tracking," *J Cell Biol* **209**(4), 609–619 (2015).
[12] Fischler, M. A. and Bolles, R. C., "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* **24**, 381–395 (June 1981).
[13] Preibisch, S., Amat, F., Stamataki, E., Sarov, M., Singer, R. H., Myers, E., and Tomancak, P., "Efficient Bayesian-based multiview deconvolution," *Nature Methods* **11**(6), 645 (2014).
[14] Besl, P. J. and McKay, N. D., "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**, 239–256 (Feb 1992).
[15] Richardson, W. H., "Bayesian-based iterative method of image restoration," *JOSA* **62**(1), 55–59 (1972).
[16] Lucy, L. B., "An iterative technique for the rectification of observed distributions," *The astronomical journal* **79**, 745 (1974).
[17] Tikhonov, A. N. and Arsenin, V. Y., [*Solutions of Ill-posed Problems*], W.H. Winston (1977).
[18] Saalfeld, S., Cardona, A., Hartenstein, V., and Tomančák, P., "As-rigid-as-possible mosaicking and serial section registration of large ssTEM datasets," *Bioinformatics* **26**(12), i57–i63 (2010).
[19] Saalfeld, S., Fetter, R., Cardona, A., and Tomancak, P., "Elastic volume reconstruction from series of ultra-thin microscopy sections," *Nature Methods* **9**(7), 717 (2012).
[20] De Greve, B., "Reflections and refractions in ray tracing." https://graphics.stanford.edu/courses/cs148-10-summer/docs/2006--degreve--reflection_refraction.pdf (2006). Accessed: 2018-11-30.
[21] Schaefer, S., McPhail, T., and Warren, J., "Image deformation using moving least squares," *ACM Trans. Graph.* **25**, 533–540 (July 2006).
[22] Nieuwenhuizen, R. P. J., Lidke, K. A., Bates, M., Puig, D. L., Grunwald, D., Stallinga, S., and Rieger, B., "Measuring image resolution in optical nanoscopy," *Nature Methods* **10**, 557–562 (Apr. 2013).
[23] Burri, O. and Herbert, A., "Fourier ring correlation plugin." https://imagej.net/Fourier_Ring_Correlation_Plugin (2018). Accessed: 2018-11-08.
[24] Blasse, C., Saalfeld, S., Etournay, R., Sagner, A., Eaton, S., and Myers, E. W., "PreMosa: extracting 2D surfaces from 3D microscopy mosaics," *Bioinformatics* **33**(16), 2563–2569 (2017).
[25] Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al., "Fiji: an open-source platform for biological-image analysis," *Nature Methods* **9**(7), 676 (2012).

2.2 Universal Super-Resolution Multiplexing by DNA Exchange

2.2.1 Main Paper

Super-Resolution Microscopy

International Edition: DOI: 10.1002/anie.201611729
German Edition: DOI: 10.1002/ange.201611729

Universal Super-Resolution Multiplexing by DNA Exchange

*Florian Schueder, Maximilian T. Strauss, David Hoerl, Joerg Schnitzbauer, Thomas Schlichthaerle, Sebastian Strauss, Peng Yin, Hartmann Harz, Heinrich Leonhardt, and Ralf Jungmann**

Abstract: *Super-resolution microscopy allows optical imaging below the classical diffraction limit of light with currently up to 20× higher spatial resolution. However, the detection of multiple targets (multiplexing) is still hard to implement and time-consuming to conduct. Here, we report a straightforward sequential multiplexing approach based on the fast exchange of DNA probes which enables efficient and rapid multiplexed target detection with common super-resolution techniques such as (d)STORM, STED, and SIM. We assay our approach using DNA origami nanostructures to quantitatively assess labeling, imaging, and washing efficiency. We furthermore demonstrate the applicability of our approach by imaging multiple protein targets in fixed cells.*

Super-resolution microscopy allows researchers to obtain images with currently up to 20 × higher spatial resolution than the classical diffraction limit.[1] Although current techniques are already starting to transform research in the life sciences,[2] most implementations are still limited to the observation of only a few molecular species in the same sample, so-called multiplexing. Exchange-PAINT,[3] a recent implementation of the PAINT[4] concept (points accumulation in nanoscale topography) and extension of DNA-PAINT,[5] enables multiplexed super-resolution imaging by using transient, programmable binding between dye-labeled imager strands and target-bound complementary d ocking strands during sequential imaging rounds. Although Exchange-PAINT allows spectrally unlimited multiplexing independent of different dye spectra (i.e. by using the same dye for each exchange round), imager strands are not fluorogenic, which firstly limits its applicability beyond total internal reflection (TIR) or oblique illumination away from the coverslip and secondly sets an upper limit for the achievable image speed. Recently, sequential labeling and imaging approaches have been devised for (d)STORM[6] ((direct) stochastic optical reconstruction microscopy), where a target is immunolabeled and imaged, followed by a fluorophore inactivation or quenching step.[7] This procedure is repeated sequentially for the acquisition of all remaining targets. Although these implementations allow spectrally unlimited multiplexing, the fluorophore quenching step followed by immunolabeling of the next target is time-intensive, which overall limits experimental throughput. Furthermore, relabeling and reimaging of targets from previous rounds is difficult to achieve. Recently, Exchange-PAINT was applied to STED[8] (stimulated emission depletion) microscopy.[9] To achieve this, the concentration of imager strands in Exchange-PAINT was increased to render most target strands o ccupied during image acquisition. While this allows for rapid probe exchange between sequential imaging rounds, it comes at the cost of potentially unoccupied target strands (as a result of the stochastic binding and unbinding of strands) and increased background fluorescence because of elevated concentrations of imager strands in solution, both ultimately limiting the achievable image resolution and quality.

To overcome limitations of current sequential multiplexing approaches and translate DNA-based multiplexing to super-resolution techniques such as (d)STORM, STED, or SIM, we here describe a universal implementation using exchangeable DNA probes. We devised a procedure (Figure 1) that allows us to efficiently attach, image, and detach dye-modified DNA strands (l abeling strands) to and from corresponding complementary handles coupled to different targets. To achieve this, we designed labeling strands that are optimized for stable binding during image acquisition but can still be efficiently removed from their targets using low-salinity washing buffer containing denaturing agents such as formamide. First, all target species (e.g. proteins P1 to P*n*) are labeled with orthogonal DNA strands (e.g. using DNA-conjugated antibodies for proteins) in a one-pot reaction (Figure 1 a). Then, buffer containing complementary labeling strands to targets P1 is introduced and DNA hybridization can occur (Figure 1 b). Next, the labeling buffer is exchanged by imaging buffer (optimized for dSTORM, STED, or SIM), which does not contain any unbound labeling strands, and image acquisition is performed (Figure 1 c). Subsequently, the imaging buffer is exchanged by low-salinity washing buffer containing 30 % of the denaturing agent formamide (for more details see the Supporting information for experimental details), thereby facilitating the dissociation of the labeling strands from their targets by virtually d ecreasing the DNA

[*] F. Schueder, M. T. Strauss, Dr. J. Schnitzbauer, T. Schlichthaerle, S. Strauss, Prof. R. Jungmann
Faculty of Physics and Center for Nanoscience, LMU Munich
Geschwister-Scholl-Platz 1, 80539 Munich (Germany)
and
Max Planck Institute of Biochemistry
Am Klopferspitz 18, 82152 Martinsried (Germany)
E-mail: jungmann@biochem.mpg.de
D. Hoerl, Dr. H. Harz, Prof. H. Leonhardt
Department of Biology II and Center for Nanoscience, LMU Munich
Grosshaderner Strasse 2, 82152 Martinsried (Germany)
Prof. P. Yin
Wyss Institute for Biologically Inspired Engineering and Department of Systems Biology, Harvard University
3 Blackfan Circle, Boston, MA 02115 (USA)

Supporting information for this article can be found under: http://dx.doi.org/10.1002/anie.201611729.

 © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

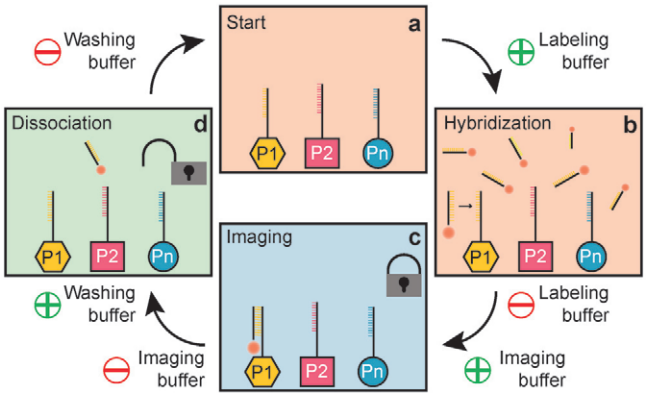


Figure 1. a) Targets 1–*n* are labeled with orthogonal approximately 12 nucleotide long DNA sequences P1–P *n*. b) Dye-modified la beling strands stably hybridize to complementary target strands P1. c) Acquisition is carried out in imaging buffer without free labeling strands in solution. d) Imaging buffer is replaced by denaturing washing buffer to facilitate the dissociation of labeling strands from targets P1. The labeling and washing procedure is repeated for all subsequent targets. Note that the labeling strands are coupled to the same spectral dye (e.g. Alexa 647) in each round, thus enabling spectrally unlimited multiplexing.

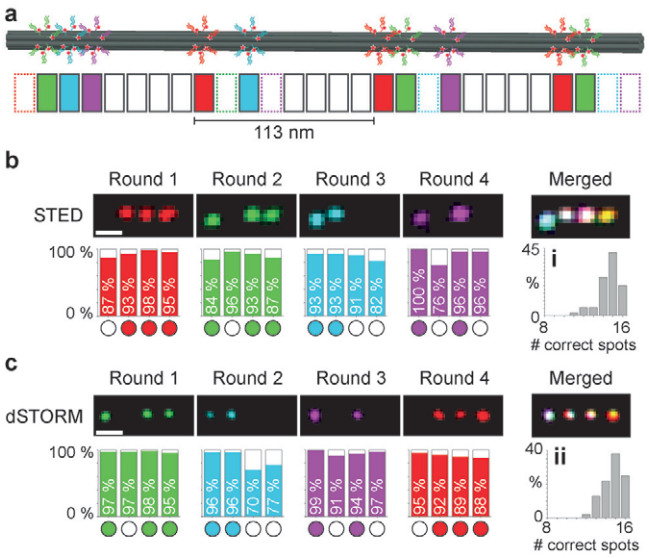


Figure 2. a) Illustration of the 6HB DNA origami b arcode. Four spots (with 6 binding sites each), spaced approximately 113 nm apart, can be d ecorated with up to four orthogonal target sequences each (colored in red, green, cyan, and magenta). b) Resulting super-resolution images of four rounds of Exchange-STED (top) with corresponding statistical analysis (bottom). Histograms for each round depict the percentage of correctly identified spots. (i) Statistical analysis showing the number of correct spots per structure in Exchange-STED (14.6 ± 0.7, mean ± standard deviation). c) Corresponding results for Exchange-dSTORM. (ii) Correct spots per structure in Exchange-dSTORM: 14.7 ± 0.4 (mean ± standard deviation). Scale bars: 200 nm.

melting temperature.[10] This washing procedure is usually performed for about 10 min, until all the labeling strands have dissociated. Finally, the washing buffer is replaced by hybridization buffer and the whole procedure is repeated for all the remaining target species. In the resulting multiplexed super-resolution micrograph, a unique pseudocolor is assigned to each imaging round (and, thus, each target). Most importantly, multiplexing is not limited by distinct spectral colors anymore, as the labeling strands for each exchange round carry the same spectral dye. The only limitation is the number of orthogonal DNA sequences (as in Exchange-PAINT), which could easily reach hundreds.

To demonstrate the feasibility of our approach, we used self-assembled DNA origami[11] (Figure 2). We designed so-called six-helix-bundle (6HB) structures[12] carrying four orthogonal single-stranded extensions on staple strands (which allows for four labeling and imaging rounds) at specific positions[13] (Figure 2a). For Exchange-STED and Exchange-dSTORM imaging, we arranged the sequences in four spots, approximately 113 nm apart (Figure 2a). For Exchange-SIM, we opted for a structure displaying three spots spaced about 168 nm apart (Figure S2). Each spot consists of six strands available for hybridization. Representative images of the respective imaging rounds are shown in Figure 2b,c for Exchange-STED and Exchange-dSTORM, respectively (see Figures S3 and S4 for expanded views). To assay the efficiency of our multiplexing approach, we interactively analyzed approximately 100 structures in the Exchange-STED and Exchange-dSTORM experiment. For quantification of correct versus incorrect spots in each labeling and imaging round, it is important to note that false negatives as well as false positives will lead to an er ror; however, these two fai lure modes have different root causes, and are thus important to distinguish. False positives occur when washing is inefficient, that is, labeling strands have not dissociated from their respective targets. False negatives occur when labeling or imaging is inefficient, that is, labeling strands have not hybridized to target strands or dyes are already bleached. The origami platform allows us to uncover both failure modes independently and thus reveals any potential bias of our approach. We analyzed each spot in each round separately to additionally assay for potential biases of different locations on the DNA origami structures. The results from our analysis show that on average about 91 % of spots are correct in the Exchange-STED experiments (Figure 2b) and 92 % in the Exchange-dSTORM experiments (Figure 2c).

To demonstrate that the order of Exchange rounds does not affect the experimental outcome, we varied the order for the dSTORM experiments. We found that, indeed, the outcome of the experiment is not affected by the order. We note, that in round 2 of the dSTORM experiment (Figure 2c), we do see a higher than expected number of false positives for spots 3 and 4 (70 % and 77 % correct, respectively). This potentially suggests insufficient washing between rounds 1 and 2. However, we also note that the expected number of correct spots was restored in round 3. To assay the influence of different washing and hybridization times, we performed additional experiments (Figure S5), where we first decreased the incubation time with the labeling strands from 10 min to 1 min (keeping the washing times constant). In a following experiment, we increased the washing time from 2 × 3 min to 3 × 10 min (keeping the incubation time of 10 min

 © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

constant). For the shorter probe incubation time, we detect a lower percentage of correctly labeled spots (true positives, see Figure S5). With longer washing times, we observe a similar performance as with our standard conditions. In conclusion, we note that our standard labeling and washing conditions (i.e. 10 min labeling, 2×3 min washing) should allow optimal results in exchange experiments. The statistical analysis of both Exchange-STED and -dSTORM experiments further shows that no positional dependency on the DNA origami structure was observed. There was also no bias towards false positives or negatives. Most importantly, there is also no bias towards later washing or labeling rounds, thus indicating that our approach is viable for more extensive multiplexing experiments (i.e beyond four rounds). Over four labeling and imaging rounds, we detected 14.6 ± 0.7 (mean $\pm$ standard deviation) correct spots in Exchange-STED (Figure 2b, (i)) and 14.7 ± 0.4 (mean $\pm$ standard deviation) correct spots in Exchange-dSTORM (Figure 2c, (ii)) from a total of 16 spots. Detailed experimental conditions and image processing specifics can be found in the Supporting Information.

Next, to translate our multiplexing concept from in vitro DNA origami structures to in situ labeling and imaging of protein targets in cells, we used primary and DNA-conjugated secondary antibodies against alpha-tubulin, LaminB, and TOM20. The respective secondary antibodies were coupled to three of our orthogonal target sequences. Hybridization, imaging, and washing steps were performed similarly to the in vitro studies on DNA origami. To demonstrate in situ imaging, we opted for dSTORM and STED as super-resolution methods (Figure 3), but the same procedure can be performed for SIM as well. The results for the respective three imaging rounds demonstrate the applicability of our labeling, washing, and imaging scheme to in situ cell samples. Relabeling and reimaging of targets from earlier imaging rounds is also possible with similar performance, thus highlighting the fact that labeling strands indeed dissociate, rather than being bleached and staying bound to their target strands (Figure S6).

In conclusion, we have devised a universal DNA-based multiplexed labeling and imaging technique that brings the advantages of DNA-PAINT and Exchange-PAINT imaging to super-resolution techniques such as dSTORM, STED, and SIM, while simultaneously overcoming some of the limitations of DNA-PAINT, that is, nonfluorogenic imager strands in solution and slower image acquisition. However, we also note that our presented multiplexing approach–as is the case for all sequential imaging techniques–i s limited to fixed cell applications and is not compatible with the imaging of live cells. Our concept has several advantages over previously reported sequential labeling and imaging approaches for multiplexed target detection: 1) Our approach is considerably faster than sequential immunolabeling[7,14] or DNA strand exchange cascades,[15] as immunolabeling of all target species is performed simultaneously and washing and labeling only takes about 20 min per round. Furthermore, the sample can remain on the microscope, thus no new registration is necessary. 2) Compared to Exchange-PAINT approaches,[9] no free i mager strands are present in the imaging buffer, as labeling strands stably hybridize to their targets, which furthermore ensures that these are constantly l abeled. This allows for optimized image-acquisition conditions for the respective super-resolution technique. 3) Targets can be relabeled and reimaged in subsequent rounds, which can provide r esistance to bleaching and increase image efficiency. Finally, by using DNA origami structures, we were able to assay the efficiency in labeling, imaging, and washing steps in a quantitative fashion.

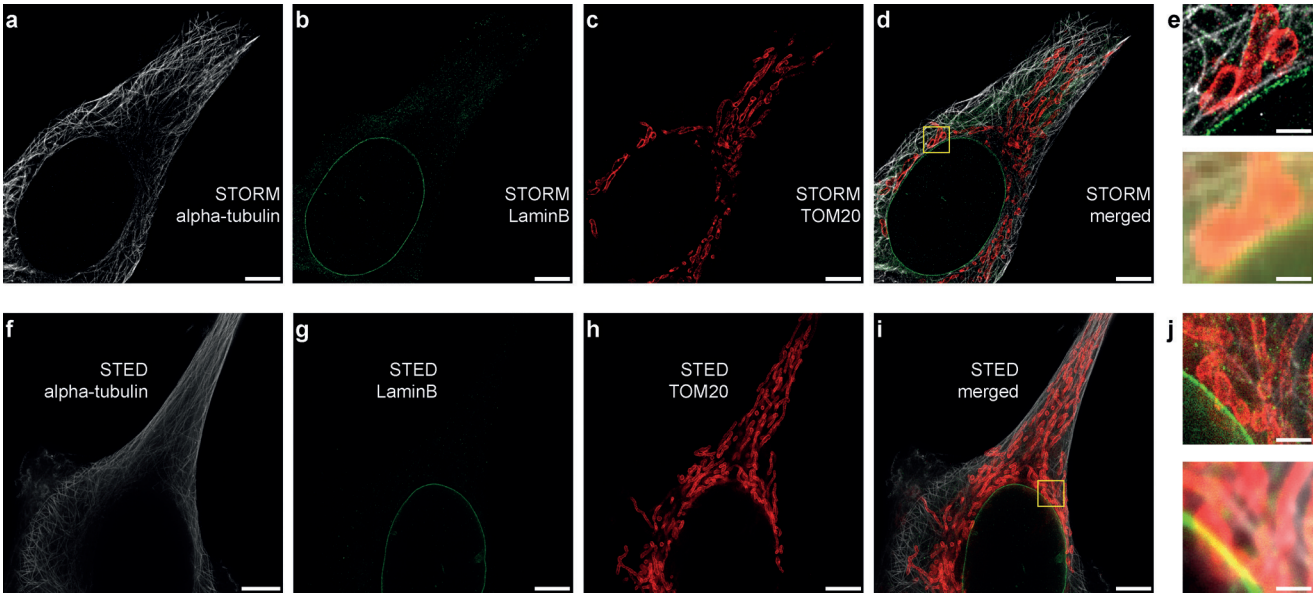


Figure 3. Three-round Exchange-dSTORM and Exchange-STED in situ. a) Alpha-tubulin is imaged in round 1. b) LaminB is imaged in round 2. c) TOM20 is imaged in round 3. d) Overlay of three-round Exchange-dSTORM. e) Zoom-in of the highlighted area from (d) with the corresponding diffraction-limited representation (bottom) demonstrating the increased spatial resolution in dSTORM. ff) Corresponding Exchange-STED results for the same protein targets. Scale bars: 5 µm (aď and fŕ), 1 µm (e, j).

 © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Acknowledgements

We thank J. B. Woehrstein and S. S. Agasti for helpful discussions and Y. Niyaz from Carl Zeiss Microscopy GmbH for the use of their SIM microscope and data acquisition support. This work was supported by the DFG through an Emmy Noether Fellowship (DFG JU 2957/1-1) and the SFB 1032 (Nanoagents for the spatiotemporal control of molecular and cellular reactions), the ERC through an ERC Starting Grant (MolMap, Grant agreement number 680241), the Max Planck Society, the Max Planck Foundation, the Center for Nanoscience (CeNS), and the Nanoinitiative Munich (NIM). M.T.S. acknowledges support from the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS). T.S. acknowledges support from the DFG through the Graduate School of Quantitative Biosciences Munich (QBM). P.Y. acknowledges support for from National Institute of Health (1-U01-MH106011-01 and 1R01EB018659-01).

Conflict of interest

Competing financial interest: A patent application has been filed. P.Y. and R.J. are cofounders of Ultivue, Inc., a startup company with interest in commercializing DNA-PAINT super-resolution technology.

Keywords: DNA nanotechnology ¥dSTORM ¥multiplexing ¥ SIM ¥STED

How to cite: *Angew. Chem. Int. Ed.* **2017**, *56*, 4052 2 4055
Angew. Chem. **2017**, *129*, 4111 2 4114

[1] S. W. Hell, S. J. Sahl, M. Bates, X. W. Zhuang, R. Heintzmann, M. J. Booth, J. Bewersdorf, G. Shtengel, H. Hess, P. Tinnefeld, A. Honigmann, S. Jakobs, I. Testa, L. Cognet, B. Lounis, H. Ewers, S. J. Davis, C. Eggeling, D. Klenerman, K. I. Willig, G. Vicidomini, M. Castello, A. Diaspro, T. Cordes, *J. Phys. D* **2015**, *48*, 443001.

[2] a) A. Szymborska, A. de Marco, N. Daigle, V. C. Cordes, J. A. Briggs, J. Ellenberg, *Science* **2013**, *341*, 655 2 658; b) K. Xu, G. Zhong, X. Zhuang, *Science* **2013**, *339*, 452 2 456; c) A. N. Boettiger, B. Bintu, J. R. Moffitt, S. Wang, B. J. Beliveau, G. Fudenberg, M. Imakaev, L. A. Mirny, C. T. Wu, X. Zhuang, *Nature* **2016**, *529*, 418 2 422.

[3] R. Jungmann, M. S. Avendano, J. B. Woehrstein, M. Dai, W. M. Shih, P. Yin, *Nat. Methods* **2014**, *11*, 313 2 318.

[4] A. Sharonov, R. M. Hochstrasser, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18911 2 18916.

[5] R. Jungmann, C. Steinhauer, M. Scheible, A. Kuzyk, P. Tinnefeld, F. C. Simmel, *Nano Lett.* **2010**, *10*, 4756 2 4761.

[6] a) M. J. Rust, M. Bates, X. Zhuang, *Nat. Methods* **2006**, *3*, 793 2 795; b) M. Heilemann, S. van de Linde, M. Schuttpelz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, M. Sauer, *Angew. Chem. Int. Ed.* **2008**, *47*, 6172 2 6176; *Angew. Chem.* **2008**, *120*, 6266 2 6271.

[7] a) J. Tam, G. A. Cordier, J. S. Borbely, A. Sandoval Alvarez, M. Lakadamyali, *PLoS One* **2014**, *9*, e101772; b) C. C. Valley, S. Liu, D. S. Lidke, K. A. Lidke, *PLoS One* **2015**, *10*, e0123941.

[8] S. W. Hell, J. Wichmann, *Opt. Lett.* **1994**, *19*, 780 2 782.

[9] S. Beater, P. Holzmeister, B. Lalkens, P. Tinnefeld, *Opt. Express* **2015**, *23*, 8630 2 8638.

[10] a) B. L. McConaughy, C. D. Laird, B. J. McCarthy, *Biochemistry* **1969**, *8*, 3289 2 3295; b) R. D. Blake, S. G. Delcourt, *Nucleic Acids Res.* **1996**, *24*, 2095 2 2103; c) R. Jungmann, T. Liedl, T. L. Sobey, W. Shih, F. C. Simmel, *J. Am. Chem. Soc.* **2008**, *130*, 10062 2 10063.

[11] P. W. Rothemund, *Nature* **2006**, *440*, 297 2 302.

[12] S. M. Douglas, J. J. Chou, W. M. Shih, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6644 2 6648.

[13] C. Lin, R. Jungmann, A. M. Leifer, C. Li, D. Levner, G. M. Church, W. M. Shih, P. Yin, *Nat. Chem.* **2012**, *4*, 832 2 839.

[14] M. J. Gerdes, C. J. Sevinsky, A. Sood, S. Adak, M. O. Bello, A. Bordwell, A. Can, A. Corwin, S. Dinn, R. J. Filkins, D. Hollman, V. Kamath, S. Kaanumalle, K. Kenny, M. Larsen, M. Lazare, Q. Li, C. Lowes, C. C. McCulloch, E. McDonough, M. C. Montalto, Z. Pang, J. Rittscher, A. Santamaria-Pang, B. D. Sarachan, M. L. Seel, A. Seppo, K. Shaikh, Y. Sui, J. Zhang, F. Ginty, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 11982 2 11987.

[15] a) D. Y. Duose, R. M. Schweller, W. N. Hittelman, M. R. Diehl, *Bioconjugate Chem.* **2010**, *21*, 2327 2 2331; b) R. M. Schweller, J. Zimak, D. Y. Duose, A. A. Qutub, W. N. Hittelman, M. R. Diehl, *Angew. Chem. Int. Ed.* **2012**, *51*, 9292 2 9296; *Angew. Chem.* **2012**, *124*, 9426 2 9430.

Manuscript received: December 1, 2016
Final Article published: March 3, 2017

 © 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

2.2.2 Supporting Information

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Universal Super-Resolution Multiplexing by DNA Exchange

*Florian Schueder, Maximilian T. Strauss, David Hoerl, Joerg Schnitzbauer, Thomas Schlichthaerle, Sebastian Strauss, Peng Yin, Hartmann Harz, Heinrich Leonhardt, and Ralf Jungmann**

anie_201611729_sm_miscellaneous_information.pdf
anie_201611729_sm_movie.avi

Supplementary Information

Materials. Unmodified, dye-labeled and biotinylated DNA oligonucleotides were purchased from MWG Eurofins. Streptavidin was purchased from Invitrogen (catalog number: S-888). Bovine serum albumin (BSA, catalog number: A4503-10G) and BSA-Biotin was obtained from Sigma-Aldrich (catalog number: A8549). Coverslips were purchased from Menzel-Gläser (Cover slips 24 × 60 mm, #1.5, catalog number: BBAD02400600#A*). Flow chambers were purchased from ibidi (Sticky-Slide VI $^{0.4}$ catalog number: 80608). M13mp18 scaffold was obtained from New England BioLabs (catalog number: N4040s). Freeze 'N Squeeze columns were ordered from Bio-Rad (catalog number: 7326165). Monoclonal antibodies against Alpha-tubulin (Thermo Scientific; catalog number: MA1-80017) were purchased from Thermo Scientific. Polyclonal antibodies against LaminB (Santa Cruz; catalog number: sc-6217) and TOM20 (Santa Cruz; catalog number: sc-11415) were ordered from Santa Cruz. The secondary antibodies Anti-Rat (catalog number: 712-005-150), Anti-Rabbit (catalog number: 711-005-152) and Anti-Goat (catalog number: 705-005-147) were purchased form Jackson ImmunoResarch. Cell imaging coverglass (catalog number: 0030 742.036) was purchased from Eppendorf. Formamide (catalog number: F9037-100ML), Protocatechuate 3,4-Dioxygenase from pseudomonas (PCD) (catalog number: P8279), 3,4-Dihydroxybenzoic acid (PCA) (catalog number: 37580-25G-F) and (+-)-6-Hydroxy-2,5,7,8-tetra-methylchromane-2-carboxzlic acid (Trolox) (catalog number: 238813-5G) were obtained from Sigma. 1M Tris pH 8.0 (catalog number: AM9856) was obtained from Ambion, Beta Mercaptoethanol (Catalog number: 63689-25ml) from Sigma, D+ Glucose(w/vol) (catalog number: 410955000) from Acros, Glucose Oxidase (catalog number: G7141-50KU) from Sigma, Glycerol (catalog number: G5516-25UN) from Sigma, Catalase (catalog number: C3155-50MG) from Sigma and H_2O (catalog number: 10977-035) was ordered from gibco.

The following buffers were used:

- Buffer A: 10 mM Tris-HCl, 100 mM NaCl, 0.05 % Tween 20, pH 7.5
- Buffer B: 5 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM EDTA, 0.05 % Tween 20, pH 8.0
- Buffer C: 1 × PBS, 500 mM NaCl, pH 8.0
- 1.2 × BME: Tris pH 8.0 50 mM, beta Mercaptoethanol 179 mM, MgCl 50 mM, Glucose 12.5 mM in H_2O
- 6 × GLOX: Glucose Oxidase 2.5 mg, Tris 50 mM, Glycerol 10 mM, Catalase 200 mg/ml in H_2O
- 40 × PCA: PCA solution consists of 154 mg PCA in 10 ml water adjusted to pH 9.0 with NaOH
- 100 × PCD: 9.3 mg PCD, 13.3 ml of buffer (50 % glycerol stock in 50 mM KCl, 1 mM EDTA and 100 mM Tris-HCl pH 8.0)
- 100 × Trolox: 100 mg Trolox, 430 µl 100 % Methanol, 345 µl 1 M NaOH in 3.2 ml H_2O

DNA origami self-assembly. The 6 helix bundle (6HB) DNA origami structures for the dSTORM, STED and SIM experiments, were formed in a one-pot reaction with a 40 µl total volume containing 10 nM scaffold strand (M13mp18), 100 nM folding staples and biotin handles, 1000 nM biotin anti-handles and 500 nM DNA-PAINT docking strands in folding buffer (1× TE buffer with 12 mM $MgCl_2$).
The solution was annealed using a thermal ramp cooling from 80 °C to 14 °C over the course of 15 h. After self-assembly, monomeric structures were purified by agarose gel electrophoresis (1.5 % agarose, 1× TBE, 10 mM $MgCl_2$, 1× SybrSafe) at 3 V/cm for 3 h. Gel bands were cut, crushed and filled into a Freeze 'N Squeeze column and spun for 5 min at 1 000× *g* at 4 °C.

Cell culture. Cells were grown in Falcon Tissue Culture Treated Flasks (catalog number: 353136) from Falcon. A mixture of 500 ml Dulbecco's Modified Eagle Medium (catalog number: 31966-021), 50 ml Fetal Bovine Serum (catalog number: 10500-064) and 5 ml Penicillin Streptomycin (catalog number: 15140-122) was used as growing media purchased from gibco. For passaging and washing the cells, 1× PBS pH 7.2 (catalog number: 20012-019) and 0.05 % Trypsin – EDTA (catalog number: 25300-054) were purchased from gibco. For the fixation paraformaldehyde and glutaraldehyde were obtained from Electron Microscopy Sciences. Quenching was done using Sodium Borohydride >97% (catalog number: 4051.1) from Roth. For permeabilization and blocking, Triton X 100 (catalog number: 6683.1) from Roth and Bovine Serum Albumin (catalog number: A4503-10G) were used.

Immunostaining of cells. HeLa cells were cultured with Dulbecco's Modified Eagle Medium supplemented with 10 % (v/v) heat inactivated FBS with 1 % (v/v) penicillin and streptomycin and incubated at 37 °C with 5 % CO_2. At approximately 30 % confluence, cells were seeded into Eppendorf chambered coverglass ~24 h before fixation. Microtubules, mitochondria and Lamin were immunostained using the following procedure: fixation in a mixture of 3 % paraformaldehyde and 0.1 % glutaraldehyde in PBS for 10 min; 3 × washing with PBS for 5 min; reduction with ~1 mg/ml $NaBH_4$ for 7 min; 3 × washing with PBS for 5 min; blocking and permeabilization with 3 % (w/v) BSA and 0.25 % (v/v) Triton X-100 in 1 × PBS for 2 h; staining overnight at 4 ºC with the primary antibodies against alpha-tubulin, TOM20 and LaminB (antibodies were diluted to 10 µg/ml in 5 % BSA); 3 × washing with PBS for 5 min each; and finally were stained for 1 h at RT with preassembled secondary antibody-DNA conjugates[1] against Rat-antibody, Rabbit-Antibody and Goat-Antibody (conjugates were diluted to 10 µg/ml in 5 % BSA); 3 × washing with PBS for 5 min each; post fixation in a mixture of 3 % paraformaldehyde and 0.1 % glutaraldehyde in PBS for 10 min; and 3 × washing with PBS for 5 min each.

Super-resolution setups. dSTORM. Fluorescence imaging was carried out on an inverted Nikon Eclipse Ti microscope (Nikon Instruments) with the Perfect Focus System, applying an objective-type TIRF configuration with an oil-immersion objective (CFI

Apo TIRF 100×, NA 1.49, Oil). For excitation and dark state transition of Alexa647 fluorophores, a 640 nm laser (150 mW nominal, Toptica iBeam Smart) was used. For facilitation of the transition back form the dark state to the ground state, a 405 nm laser (150 mW nominal, Toptica iBeam Smart) was used. The laser beams were passed through a cleanup filter (ZET405/488/561/640x, Chroma Technology, Bellows Falls, VT) and coupled into the microscope objective using a quad-band beam splitter (ZT405/488/561/640rpc, Chroma Technology). Fluorescence light was spectrally filtered with an emission filter (ZET405/488/561/640m-TRF, Chroma Technology) and imaged on an EMCCD camera (iXon Ultra 897 EMCCD, Andor Technology). Imaging was performed without additional magnification in the detection path and yielding a pixel size of 160 nm.
STED. Stimulated Emission Depletion images were acquired using a 3D STED microscope (Abberior Instruments, Göttingen, Germany). The system was equipped with 594 nm and 640 nm pulsed excitation lasers and a pulsed 775 nm depletion laser. The depletion pattern was generated via phase-modulation by a SLM (Abberior Instruments easy3D STED module). The objectives employed were a UPlanSApo 100x / 1.4 NA oil immersion objective and a UPlanSApo 60x / 1.2 NA water immersion objective (Olympus, Tokyo, Japan). Unless noted otherwise, images were acquired with the 100x objective. Image acquisition was controlled via the software ImSpector (Abberior Instruments and MPI for Biophysical Chemistry, Göttingen, Germany).
SIM. Structured Illumination images were acquired using a commercial Zeiss Elyra PS.1 (Carl Zeiss Microscopy, Germany) system. A 63x Plan-Apochromat 1.40 Oil objective and 37.5 mW @ 642 nm laser excitation power was used to acquire the images onto a PCO Edge sCMOS camera. SIM image acquisition and reconstruction was carried out using Zeiss ZEN software, according to the instructions from the manufacturer.

Sample preparation and image acquisition of DNA origami structures. dSTORM. First, the ibidi flow chamber was cleaned by rinsing 100 µl of isopropanol through the chamber and then washed 3 × with ultra-pure water. Then, 200 µl of biotin-labeled bovine albumin (1 mg/ml, dissolved in buffer A) was flown into the chamber and incubated for 5 min. The chamber was then washed using 100 µl of buffer A. 200 µl of streptavidin (0.5 mg/ml, dissolved in buffer A) was then flown into the chamber and allowed to bind for 5 min. After washing with 100 µl of buffer A and subsequently with 100 µl of buffer B, 50 µl of biotin-labeled DNA structures (~100 pM monomer concentration) in buffer B were finally flown into the chamber and incubated for 45 min. The chamber was washed using 100 µl of buffer B. Next, the sample was incubated with 100 nM Alexa647-modified labeling strands in buffer B for 10 min. Finally, the imaging buffer containing 1× BME and 1× GLOX was flown into the chamber. After image acquisition, the chamber was flushed with 30 % formamide in 1× PBS twice for 2 times ~ 3min. Then washing with 1× PBS was performed to exchange the washing buffer and subsequently buffer B was added. Afterwards the next Alexa647-modified labeling strands were introduced. Acquisition and washing steps were repeated until all targets were imaged. The CCD readout bandwidth was set to 17 MHz at 16 bit and 5.1 pre-amp gain. 100 electron multiplying (EM) gain was used. Imaging was performed using TIR illumination with an excitation intensity of ~1 kW/cm^2 at 640 nm and ~8 W/cm^2 at 405 nm.
Image acquisition parameters: 50 ms integration time, 10 000 frames per exchange round. RAW data was processed using a custom software package called "Picasso"[2] (www.jungmannlab.org, https://github.com/jungmannlab/picasso), employing single-molecule spot detection and standard maximum likelihood fitting routines[3]. For the *in vitro* DNA origami data, we detected on average 3 475 photons per localization and achieved an average NeNA[4] localization precision of ~9.8 nm, yielding a FWHM resolution of ~23 nm.
STED and SIM. First, the ibidi flow chamber was cleaned by rinsing 100 µl of isopropanol through the chamber and then washed 3 × with ultra-pure water. Then, 200 µl of biotin-labeled bovine albumin (1 mg/ml, dissolved in buffer A) was flown into the chamber and incubated for 5 min. The chamber was then washed using 100 µl of buffer A. 200 µl of streptavidin (0.5 mg/ml, dissolved in buffer A) was then flown through the chamber and allowed to bind for 5 min. After washing with 100 µl of buffer A and subsequently with 100 µl of buffer B, 50 µl of biotin-labeled DNA structures (~100 pM monomer concentration) in buffer B were finally flown into the chamber and incubated for 45 min. The chamber was washed using 100 µl of buffer B. The chamber was washed using 100 µl of buffer B. Then the sample was incubated with 100 nM Atto647N-modified labeling strand in buffer B for 10 min. After three-times 3 min of washing, the imaging buffer was introduced. Finally, the imaging buffer containing 1× Trolox, 1× PCA and 1× PCD (diluted in buffer B) was flown into the chamber. After image acquisition, the chamber was flushed with 30 % formamide in 1x PBS twice for ~3 min. Then washing with 1× PBS was performed to exchange the washing buffer and subsequently Buffer B was added. Afterwards the next Atto647N-modified labeling strand was introduced. Acquisition and washing steps were repeated until all targets were imaged. Images were acquired with 20 nm pixel steps and 10 µs pixel dwell time and 10-fold line accumulation. The average laser powers in the back focal plane were set to ~13 µW (640 nm excitation) and ~195 mW (775 nm depletion). The size of the confocal pinhole was set to 1 A.U. For the *in vitro* DNA origami data, we achieved an average localization precision of ~30 nm, yielding a FWHM STED resolution of ~68 nm.

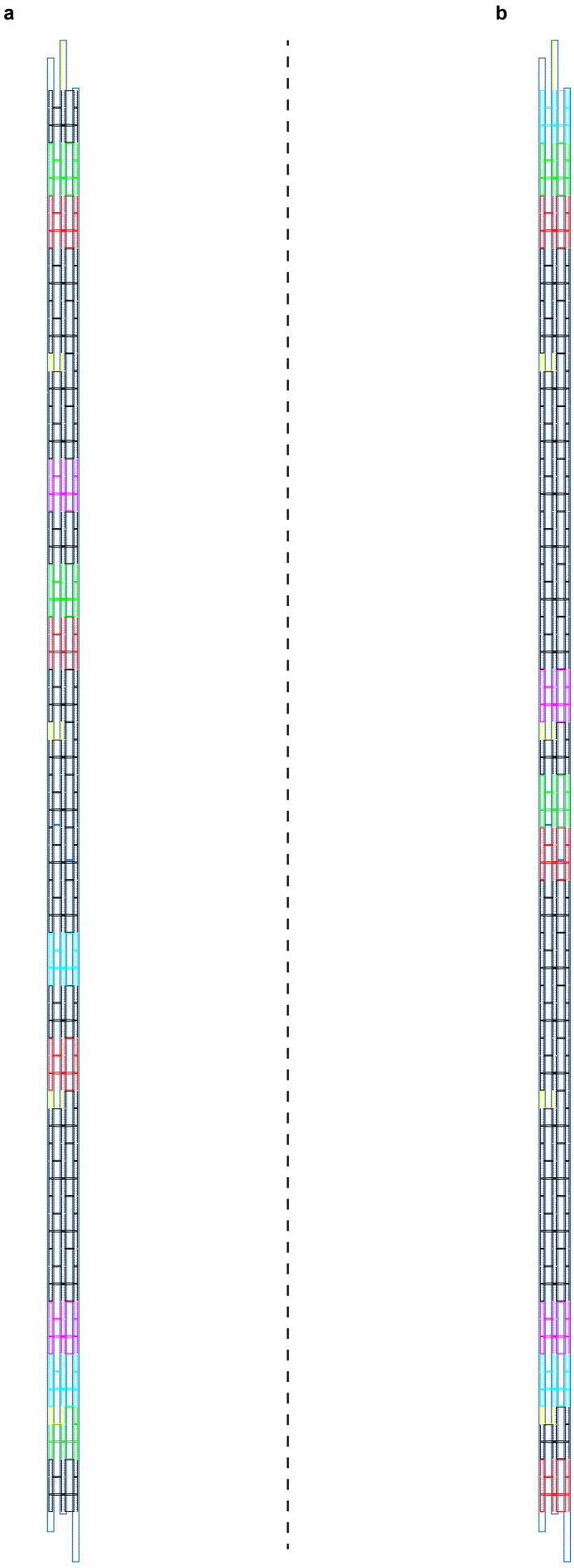
Sample preparation and image acquisition of cell samples. dSTORM. An Eppendorf chamber was adapted for fluid exchange. Images were acquired with an EMCCD readout bandwidth of 17 MHz at 14 bit, 5.1 pre-amp gain and 100 EM gain. Imaging was performed using oblique illumination. Sequential labeling, imaging and washing was performed analogous to the DNA origami *in vitro* experiments. The laser power densities at 640 nm were ~2 kW/cm^2 and 405 nm at ~6 W/cm^2.
Image acquisition parameters: 50 ms integration time, 50 000 frames per exchange round. RAW data was processed using a custom software package called "Picasso"[2] (www.jungmannlab.org, https://github.com/jungmannlab/picasso), employing single-molecule spot detection and standard maximum likelihood fitting routines[3]. For the *in situ* cell data, we detected on

average 5 400 photons per localization and achieved an average NeNA[4] localization precision of ~19 nm, yielding a FWHM resolution of ~44 nm.

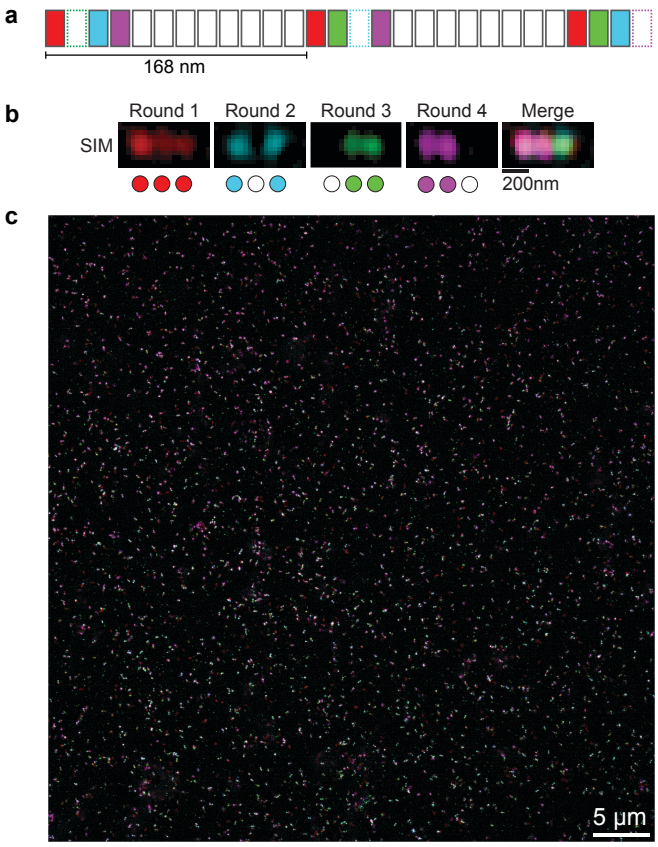
STED. Images were acquired with 20 nm pixel steps and 10 µs pixel dwell time and 3-fold line accumulation. The average laser powers in the back focal plane were set to ~13 µW (640 nm excitation) and ~190 mW (775 nm depletion). The size of the confocal pinhole was set to 1 A.U.

The STED-stack (Supplementary Video 1) was acquired with the 60x water immersion objective. The pixel steps were 40 nm in x and y and 400 nm in z, with a pixel dwell time of 15 µs and 2-fold line accumulation. The average laser powers in the back focal plane were set to ~13 µW (640 nm excitation) and ~195 mW (775 nm depletion). The size of the confocal pinhole was set to 1 A.U. 33 % of the depletion laser power was subjected to a circular phase mask to achieve axial resolution improvement in addition to the lateral super-resolution.

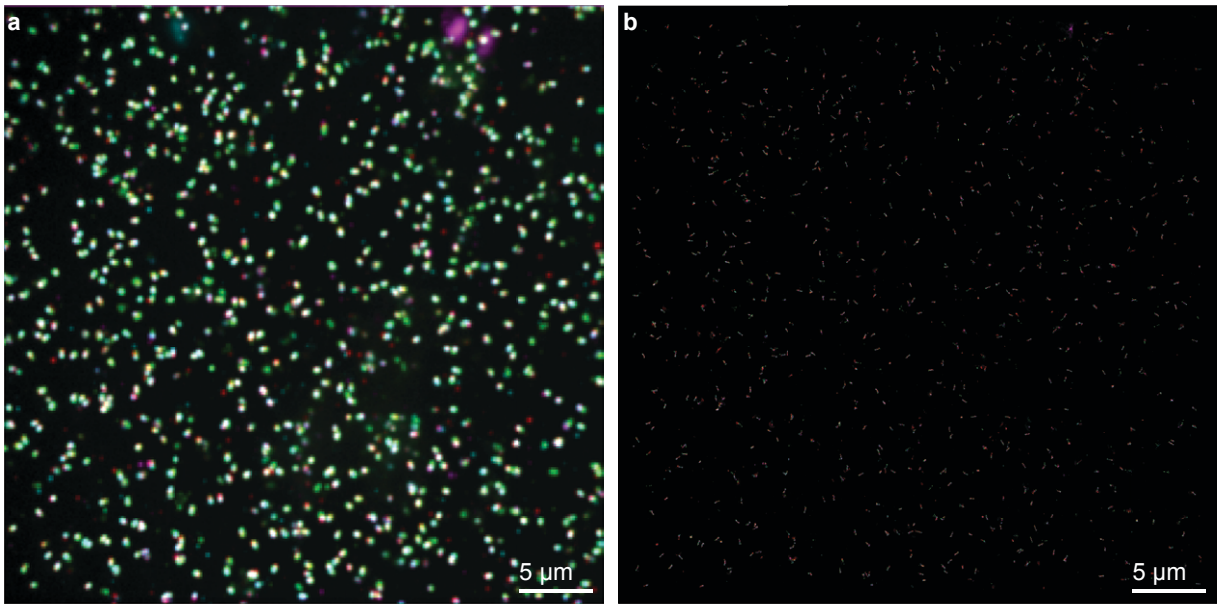
Supplementary Figures



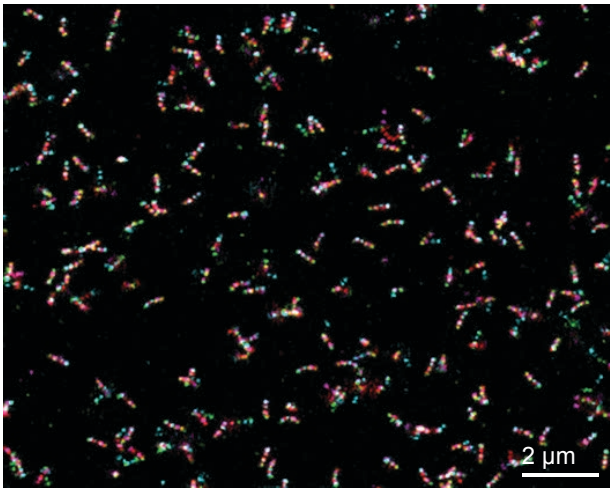
Supplementary Figure 1 | DNA origami designs. (**a**) 6HB origami for dSTORM and STED study. (**b**) 6HB origami for SIM study.



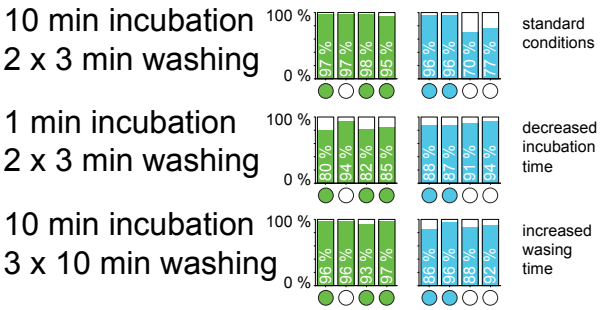
Supplementary Figure 2 | Exchange-SIM. (**a**) Schematic representation of 6HB DNA origami structure for Exchange-SIM demonstration. 3 spots are designed that can display up to 4 orthogonal DNA sequences for Exchange imaging. Spots are spaced ~168 nm apart. (**b**) Representative DNA origami structures for 4-round Exchange-SIM experiment. (**c**) 4-round Exchange-SIM overview.



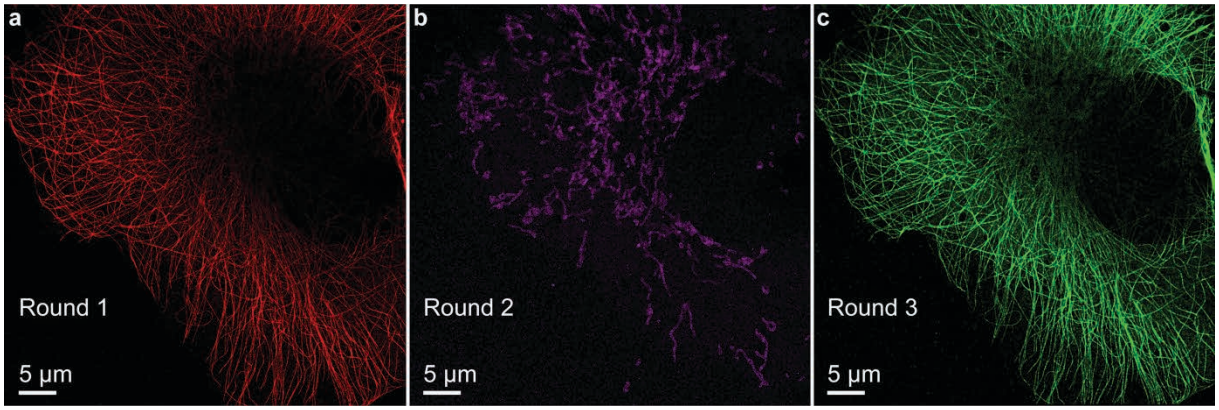
Supplementary Figure 3 | Exchange-dSTORM overview. (**a**) Diffraction-limited overview for 4-round Exchange-dSTORM. (**b**) Corresponding dSTORM super-resolution micrograph.



Supplementary Figure 4 | Exchange-STED overview. 4-color Exchange-STED micrograph.



Supplementary Figure 5 | Washing and hybridization assay. Results from different hybridization and washing times.

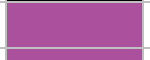
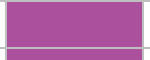
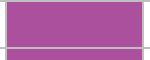
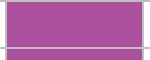


Supplementary Figure 6 | Exchange-dSTORM relabeling experiment. (**a**) Alpha-tubulin imaging in round 1. (**b**) TOM20 proteins are imaged in round 2 after probe exchange. Note that no microtubule structures are visible anymore. (**c**) Docking strands on anti-alpha-tubulin antibodies are relabeled and reimaged in round 3, highlighting the high labeling, erasing, and relabeling efficiency in our Exchange-dSTORM experiment.

Supplementary Table 1 | Strand sequences for 6HB DNA origami structure used for Exchange-STED and Exchange-dSTORM

Positon	Sequence	Color	Spot Number / Mod
0[293]4[280]	TGCTGAACCTCAAATAAAGCCAGAATGGGGAAGGTAAATATT		
0[335]4[322]	CGCTGAGAGCCAGCAGTAAGCGTCATACAGCGCCAAAGACAA		
0[377]4[364]	GAGGTGAGGCGGTCGGTAATAAGTTTTAGTTTATTTTGTCAC		
0[419]4[406]	CCATTAAAAATACCCCGTATAAACAGTTGGTGGCAACATATA		
0[461]4[448]	CTATTAGTCTTTAATATTCTGAAACATGATTACGCAGTATGT		
0[503]4[490]	GAAAGCGTAAGAATGAGAAGGATTAGGAATAACGGAATACCC		
0[545]4[532]	AAAGGGACATTCTGGATAAGTGCCGTCGAGCCGAACAAAGTT		
0[587]4[574]	TTATTTACATTGGCTAGGTGTATCACCGTAGCTATCTTACCG		
0[797]4[784]	AGGCCACCGAGTAAAAGTTTTGTCGTCTATTTTAAGAACTGG		
0[839]4[826]	AACGGTACGCCAGATATGGGATTTTGCTTGGTTTAATTTCAA		
0[881]4[868]	CGGGAGCTAAACAGGAGAATAGAAAGGACTGACGAGAAACAC		
0[923]4[910]	TTGACGAGCACGTATTTTCACGTTGAAAATCAACGTAACAAA		
0[965]4[952]	CCGCGCTTAATGCGCCTTTAATTGTATCATCAAGAGTAATCT		
0[1007]4[994]	CGAAAAACCGTCTATTCTTAAACAGCTTATGAACGGTGTACA		
0[1049]4[1036]	AAGAGTCCACTATTACAACCATCGCCCAAATCATAAGGGAAC		
0[1091]4[1078]	GAATAGCCCGAGATGCTTGCAGGGAGTTACCTGCTCCATGTT		
1[280]3[293]	AAAGAAATTGCGTATAATTTAGGCAGAGCCGACTTGCGGGAG		
1[322]3[335]	TACAGTAACAGTACAGGGCTTAATTGAGGCTATTTTGCACCC		
1[364]3[377]	TTCGCCTGATTGCTGTTATACAAATTCTACGCTAACGAGCGT		
1[406]3[419]	AGGCGAATTATTCAGGAATCATAATTACAAATAAACAGCCAT		
1[448]3[461]	TGAAACAAACATCACCGACCGTGTGATAGATTTTTTGTTTAA		
1[490]3[503]	TTTCATTTGAATTAAGTTAATTTCATCTAGAGAGAATAACAT		
1[532]3[545]	CAATATATGTGAGTCGCAAGACAAAGAAAATTAACTGAACAC		
1[574]3[587]	TATTAATTAATTTTTTGGGTTATATAACGCTAATATCAGAGA		
1[784]3[797]	GCTTCTGGTGCCGGTTGTTAAAATTCGCCTTTAAACAGTTCA		
1[826]3[839]	AGGCTGCGCAACTGAAGATTGTATAAGCTCAGGTCTTTACCC		
1[868]3[881]	TCGCTATTACGCCAATATGTACCCCGGTTTGCATCAAAAAGA		
1[910]3[923]	CGATTAAGTTGGGTAGAGAATCGATGAAATCGCGTTTTAATT		
1[952]3[965]	TGTAAAACGACGGCACAAAGGCTATCAGCAAACTCCAACAGG		
1[994]3[1007]	GACTCTAGAGGATCGATAAATTAATGCCCCTTTTGATAAGAG		
1[1036]3[1049]	ATGGTCATAGCTGTACAGTCAAATCACCATTGCTGAATATAA		
1[1078]3[1091]	AATTCCACACAACATAATGTGTAGGTAACAACTAAAGTACGG		
2[307]0[294]	ACAACGCCAACATGGATTTTCAGGTTTACTAAAGCATCACCT		
2[349]0[336]	CAACGCTCAACAGTCTTTTACATCGGGACTGCAACAGTGCCA		
2[391]0[378]	TTAGTATCATATGCTTGAATACCAAGTTCAGAAGATAAAACA		
2[475]0[462]	AATGGTTTGAAATAAGAAAACAAAATTAATATTTTTGAATGG		
2[517]0[504]	TTCAAATATATTTTCCTTTTTTAATGGAAACCCTTCTGACCT		
2[559]0[546]	GATGCAAATCCAATGAATAACCTTGCTTCACGACCAGTAATA		
2[601]0[588]	ACCTCCGGCTTAGGCCCTTAGAATCCTTTCGTCTGAAATGGA		
2[811]0[798]	TAAACGTTAATATTAAACCAGGCAAAGCTTTTATAATCAGTG		

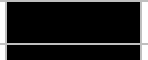
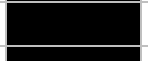
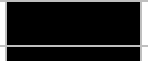
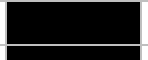
2[853]0[840]	GCCCCAAAAACAGGTTGGGAAGGGCGATGGATTTTAGACAGG	[redacted]	
2[895]0[882]	CTAGCATGTCAATCGCTGGCGAAAGGGGCGTTAGAATCAGAG	[redacted]	
2[937]0[924]	GTCTGGAGCAAACAAACGCCAGGGTTTTGTACTATGGTTGCT	[redacted]	
2[979]0[966]	TTTTTGAGAGATCTCAGTGCCAAGCTTGAACCACCACACCCG	[redacted]	
2[1063]0[1050]	GAGAAAGGCCGGAGTTCCTGTGTGAAATTTCCAGTTTGGAAC	[redacted]	
2[1105]0[1092]	ATGCAATGCCTGAGTACGAGCCGGAAGCCTTATAAATCAAAA	[redacted]	
3[294]5[307]	GTTTTGAAGCCTTAACCGATTGAGGGAGAAAGCGCAGTCTCT	[redacted]	
3[336]5[349]	AGCTACAATTTTATTCATATGGTTTACCATGGCTTTTGATGA	[redacted]	
3[378]5[391]	CTTTCCAGAGCCTAACACCACGGAATAAACGGGGTCAGTGCC	[redacted]	
3[420]5[433]	ATTATTTATCCCAAAATACATACATAAAAATGCCCCCTGCCT	[redacted]	
3[462]5[475]	CGTCAAAAATGAAAGATTAAGACTCCTTAAAGTATTAAGAGG	[redacted]	
3[504]5[517]	AAAAACAGGGAAGCGAGGAAACGCAATATTAGCGGGGTTTTG	[redacted]	
3[546]5[559]	CCTGAACAAAGTCAAAAAGTAAGCAGATAGAGGGTTGATATA	[redacted]	
3[588]5[601]	GATAACCCACAAGACAATGAAATAGCAATACTCAGGAGGTTT	[redacted]	
3[798]5[811]	GAAAACGAGAATGAAATTACCTTATGCGTTCCAGACGTTAGT	[redacted]	
3[840]5[853]	TGACTATTATAGTCAATTGGGCTTGAGAAAACAACTTTCAAC	[redacted]	
3[882]5[895]	TTAAGAGGAAGCCCGAATAAGGCTTGCCACAACTAAAGGAAT	[redacted]	
3[924]5[937]	CGAGCTTCAAAGCGTATTCATTACCCAAATCTCCAAAAAAAA	[redacted]	
3[966]5[979]	TCAGGATTAGAGAGGCTGGCTGACCTTCGGTTTATCAGCTTG	[redacted]	
3[1008]5[1021]	GTCATTTTTGCGGATTGAAAGAGGACAGGATACCGATAGTTG	[redacted]	
3[1050]5[1063]	TGCTGTAGCTCAACAGGCGCAGACGGTCCGCATAACCGATAT	[redacted]	
3[1092]5[1105]	TGTCTGGAAGTTTCTGTCGAAATCCGCGAAAGGCCGCTTTTG	[redacted]	
4[279]2[266]	GACGGAAATTATTCTTTTAGCGAACCTCGCATTTTCGAGCCA	[redacted]	
4[321]2[308]	AAGGGCGACATTCAAATCAAGATTAGTTAATCGCCATATTTA	[redacted]	
4[363]2[350]	AATCAATAGAAAATCCTGAATCTTACCATACCAGTATAAAGC	[redacted]	
4[405]2[392]	AAAGAAACGCAAAGATTTGCCAGTTACATAGAAAAAGCCTGT	[redacted]	
4[447]2[434]	TAGCAAACGTAGAATCCAAATAAGAAACAATAAGGCGTTAAA	[redacted]	
4[489]2[476]	AAAAGAACTGGCATATAGCAGCCTTTACTCTGACCTAAATTT	[redacted]	
4[531]2[518]	ACCAGAAGGAAACCGCATTAGACGGGAGCGCGAGAAAACTTT	[redacted]	
4[573]2[560]	AAGCCCTTTTTAAGGAGGGTAATTGAGCTATATGTAAATGCT	[redacted]	
4[783]2[770]	CTCATTATACCAGTATCCCCCTCAAATGATTAAATTTTTGTT	[redacted]	
4[825]2[812]	CTTTAATCATTGTGCCATAAATCAAAAAAAATATTTAAATTG	[redacted]	
4[867]2[854]	CAGAACGAGTAGTAAGAAGCAAAGCGGATGATAATCAGAAAA	[redacted]	
4[909]2[896]	GCTGCTCATTCAGTGAAAGACTTCAAATCGGTAATCGTAAAA	[redacted]	
4[951]2[938]	TGACAAGAACCGGAAACCAGACCGGAAGGTCATTGCCTGAGA	[redacted]	
4[993]2[980]	GACCAGGCGCATAGTACCTTTAATTGCTGGAGAGGGTAGCTA	[redacted]	
4[1035]2[1022]	CGAACTGACCAACTTGGCTTAGAGCTTAATCAATATGATATT	[redacted]	
4[1077]2[1064]	ACTTAGCCGGAACGATGTTTTAAATATGAGATTCAAAAGGGT	[redacted]	
5[266]1[279]	AAATAAATCCTCATTATCAAACCCTCAACGTAAAACAGAAAT	[redacted]	
5[308]1[321]	GAATTTACCGTTCCAGCAAATGAAAAATACGTCAGATGAATA	[redacted]	
5[350]1[363]	TACAGGAGTGTACTAGTATTAACACCGCGAAACAATAACGGA	[redacted]	
5[392]1[405]	TTGAGTAACAGTGCGAACGAACCACCAGACAAAATCGCGCAG	[redacted]	
5[434]1[447]	ATTTCGGAACCTATTGCGCGAACTGATAGCAAAAGAAGATGA	[redacted]	

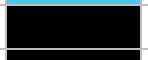
5[476]1[489]	CTGAGACTCCTCAAACGTGGCACAGACAATTACATTTAACAA		
5[518]1[531]	CTCAGTACCAGGCGGCCAACAGAGATAGAACAGTACATAAAT		
5[560]1[573]	AGTATAGCCCGGAAAGATTCACCAGTCACTGTAAATCGTCGC		
5[770]1[783]	AGCGTAACGATCTAAAGAGTCTGTCCATAGCTTTCCGGCACC		
5[812]1[825]	AAATGAATTTTCTGATCCTGAGAAGTGTGCCATTCGCCATTC		
5[854]1[867]	AGTTTCAGCGGAGTGAGGCCGATTAAAGCGGTGCGGGCCTCT		
5[896]1[909]	TGCGAATAATAATTTAACGTGCTTTCCTGATGTGCTGCAAGG		
5[938]1[951]	GGCTCCAAAAGGAGCCGCTACAGGGCGCCCCAGTCACGACGT		
5[980]1[993]	CTTTCGAGGTGAATTCAACGCTGCGCGTCATGCCTGCAGGTC		
5[1022]1[1035]	CGCCGACAATGACAAAAGAACGTGGACTTCGAATTCGTAATC		
5[1064]1[1077]	ATTCGGTCGCTGAGAGGGTTGAGTGTTGTGTTATCCGCTCAC		
0[251]4[238]	GTTGGCAAATCAACCAGACGATTGGCCTATCACCGTCACCGA		1
0[755]4[742]	GTTGTAGCAATACTTAGCATTCCACAGAGAAAAATCTACGTT		2
1[238]3[251]	AGAACCTACCATATTAAAGTACCGACAATTATCCGGTATTCT		1
1[742]3[755]	TCGGCCTCAGGAAGTTTAACCAATAGGACTGCGGAATCGTCA		2
2[265]0[252]	GTAATAAGAGAATACAAAATTATTTGCATCAATATCTGGTCA		1
2[769]0[756]	AAATCAGCTCATTTATCGCACTCCAGCCCACGCAAATTAACC		2
3[252]5[265]	AAGAACGCGAGGCGATTAAAGGTGAATTTGATATTCACAAAC		1
3[756]5[769]	TAAATATTCATTGACAGGACGTTGGGAACAGCCCTCATAGTT		2
4[237]2[224]	CTTGAGCCATTTGGCAGATATAGAAGGCAAGGTAAAGTAATT		1
4[741]2[728]	AATAAAACGAACTAGATAGCGTCCAATAACGCCATCAAAAAT		2
5[224]1[237]	AGGTTGAGGCAGGTAGTTGAAAGGAATTATAATGGAAGGGTT		1
5[728]1[741]	AACTACAACGCCTGTCTTTGATTAGTAAGGACGACGACAGTA		2
0[167]4[154]	TCAATAGATAATACCACCCTCAGAGCCAGTCACCAATGAAAC		1
0[671]4[658]	AGAACAATATTACCCCTCATTTTCAGGGCATTCAACTAATGC		2
0[1175]4[1162]	GCGGTCCACGCTGGACAGAGGCTTTGAGTGACCCCCAGCGAT		3
1[154]3[167]	ATCATATTCCTGATATCAACAATAGATATCATCGAGAACAAG		1
1[658]3[671]	TTGACCGTAATGGGAACAACCCGTCGGAAACCAAAATAGCGA		2
1[1162]3[1175]	GCGCTCACTGCCCGAATACTTTTGCGGGTAGATACATTTCGC		3
2[685]0[672]	AAATGTGAGCGAGTATAGGTCACGTTGGTGCTGGTAATATCC		2
2[1189]0[1176]	CATTATGACCCTGTCTTTCCAGTCGGGAGAGAGTTGCAGCAA		3
3[168]5[181]	CAAGCCGTTTTTATGCAAGGCCGGAAACCCACCCTCAGAGCC		1
3[672]5[685]	GAGGCTTTTGCAAAATTTAGGAATACCAATAGCAAGCCCAAT		2
3[1176]5[1189]	AAATGGTCAATAACAAAACACTCATCTTGACTAAAGACTTTT		3
4[153]2[140]	CATCGATAGCAGCAACCAAGTACCGCACAGTCCTGAACAAGA		1
4[657]2[644]	AGATACATAACGCCCAGACGACGATAAATTCTCCGTGGGATA		2
4[1161]2[1148]	TATACCAAGCGCGATTTAGTTTGACCATAGAAGCCTTTATTT		3
5[140]1[153]	ACCCTCAGAACCGCATTTGAGGATTTAGGGAGCGGAATTATC		1
5[644]1[657]	CTCAGAGCCACCACGCCAGCCATTGCAAAAACAAACGGCGGA		2
5[1148]1[1161]	GGGTAGCAACGGCTTTTGCCCCAGCAGGACATTAATTGCGTT		3
0[209]4[196]	AAAATATCTTTAGGCACCAGAGCCGCCGAAAATCACCAGTAG		1
0[713]4[700]	TGAGTAGAAGAACTCGTAACACTGAGTTTTACAGGTAGAAAG		2
0[1217]4[1204]	CTGATTGCCCTTCACCATTAAACGGGTATAAAACGAAAGAGG		3

1[196]3[209]	ATCCTGATTGTTTGACAATAAACAACATTCATTACCGCGCCC		1
1[700]3[713]	TCGTAACCGTGCATCTTCCTGTAGCCAGAGGGGGTAATAGTA		2
1[1204]3[1217]	GCTGCATTAATGAACATAAAGCTAAATCTTTCATTTGGGGCG		3
2[223]0[210]	CTGTCCAGACGACGGATTATACTTCTGAGAGGAAGGTTATCT		1
2[1231]0[1218]	ATAAAGCCTCAGAGTCGGCCAACGCGCGGAGACGGGCAACAG		3
3[210]5[223]	AATAGCAAGCAAATGAATTAGAGCCAGCCCAGCATTGACAGG		1
3[714]5[727]	AAATGTTTAGACTGACGGAACAACATTATCGTCACCAGTACA		2
3[1218]5[1231]	CGAGCTGAAAAGGTCGAAGGCACCAACCAAATACGTAATGCC		3
4[195]2[182]	CACCATTACCATTATTTCATCGTAGGAAGTTCAGCTAATGCA		1
4[699]2[686]	ATTCATCAGTTGAGAGAAGTTTTGCCAGCTTTCATCAACATT		2
4[1203]2[1190]	CAAAAGAATACACTCTGTTTAGCTATATGGTTGTACCAAAAA		3
5[182]1[195]	GCCACCAGAACCACAGCACTAACAACTAAATTCATCAATATA		1
5[686]1[699]	AGGAACCCATGTACCAAACTATCGGCCTTGTAGATGGGCGCA		2
5[1190]1[1203]	TCATGAGGAAGTTTCCGCCTGGCCCTGAAACCTGTCGTGCCA		3
0[125]4[112]	ACAAACAATTCGACCACCGGAACCGCCTGACAGAATCAAGTT		1
0[629]4[616]	ATGGAAATACCTACCAGAACCGCCACCCGGCATAGTAAGATA		2
0[1133]4[1120]	GATGGTGGTTCCGATCAGCAGCGAAAGAGAGATTTGTATCAT		3
1[112]3[125]	ATTTTGCGGAACAATCCTAATTTACGAGTCCTTATCATTCCA		1
1[616]3[629]	AGCTTAGATTAAGAATCATAGGTCTGAGAAGCAACACTATCA		2
1[1120]3[1133]	CTGGGGTGCCTAATAAATTTTTAGAACCTTGATTCCCAATTC		3
2[139]0[126]	AAAATAATATCCCAAGAAACCACCAGAAAAGTATTAGACTTT		1
2[643]0[630]	GTGAATTTATCAAACGCTGAGAAGAGTCCAGGAAAAACGCTC		2
2[1147]0[1134]	CAACGCAAGGATAAGAGTGAGCTAACTCCGAAAATCCTGTTT		3
3[126]5[139]	AGAACGGGTATTAACCGTAATCAGTAGCCCCTCAGAGCCGCC		1
3[630]5[643]	TAACCCTCGTTTACAAAAGGAATTACGATCAGAACCGCCACC		2
3[1134]5[1147]	TGCGAACGAGTAGAAACAAAGTACAACGCAGCATCGGAACGA		3
4[111]2[98]	TGCCTTTAGCGTCATAATCGGCTGTCTTCATGTAGAAACCAA		1
4[615]2[602]	ATAAGAGCAAGAAAATTGAGTTAAGCCCAGACTACCTTTTTA		2
4[1119]2[1106]	CGCCTGATAAATTGATTCCATATAACAGCTCATATATTTTAA		3
5[98]1[111]	GGAACCAGAGCCACAACTCGTATTAAATTGAGTAACATTATC		1
5[602]1[615]	AGTACCGCCACCCTATTTTGACGCTCAAGAAAACATAGCGAT		2
5[1106]1[1119]	CGGGATCGTCACCCAATCGGCAAAATCCATAAAGTGTAAAGC		3
2[181]0[168]	GAACGCGCCTGTTTTATCAGATGATGGCATAGATTAGAGCCG		Biotin
2[433]0[420]	TAAGAATAAACACCTTTCAATTACCTGAGCCCTAAAACATCG		Biotin
2[727]0[714]	AATTCGCGTCTGGCCTGCCAGTTTGAGGTAACATCACTTGCC		Biotin
2[1021]0[1008]	CAACCGTTCTAGCTCCCGGGTACCGAGCCCAACGTCAAAGGG		Biotin
2[1271]2[1232]	AAATCATACAGGCAAGGCAAAGAATTAGCAAAATTAAGCA		Biotin

Supplementary Table 2 | Strand sequences for 6HB DNA origami structure used for Exchange-SIM

Position	Sequence	Color	Spot Number / Mod
0[293]4[280]	TGCTGAACCTCAAATAAAGCCAGAATGGGGAAGGTAAATATT		
0[335]4[322]	CGCTGAGAGCCAGCAGTAAGCGTCATACAGCGCCAAAGACAA		
0[377]4[364]	GAGGTGAGGCGGTCGGTAATAAGTTTTAGTTTATTTTGTCAC		
0[419]4[406]	CCATTAAAAATACCCCGTATAAACAGTTGGTGGCAACATATA		
0[629]4[616]	ATGGAAATACCTACCAGAACCGCCACCCGGCATAGTAAGATA		
0[671]4[658]	AGAACAATATTACCCCTCATTTTCAGGGCATTCAACTAATGC		
0[713]4[700]	TGAGTAGAAGAACTCGTAACACTGAGTTTTACAGGTAGAAAG		
0[755]4[742]	GTTGTAGCAATACTTAGCATTCCACAGAGAAAAATCTACGTT		
0[965]4[952]	CCGCGCTTAATGCGCCTTTAATTGTATCATCAAGAGTAATCT		
0[1007]4[994]	CGAAAAACCGTCTATTCTTAAACAGCTTATGAACGGTGTACA		
0[1049]4[1036]	AAGAGTCCACTATTACAACCATCGCCCAAATCATAAGGGAAC		
0[1091]4[1078]	GAATAGCCCGAGATGCTTGCAGGGAGTTACCTGCTCCATGTT		
1[280]3[293]	AAAGAAATTGCGTATAATTTAGGCAGAGCCGACTTGCGGGAG		
1[322]3[335]	TACAGTAACAGTACAGGGCTTAATTGAGGCTATTTTGCACCC		
1[364]3[377]	TTCGCCTGATTGCTGTTATACAAATTCTACGCTAACGAGCGT		
1[406]3[419]	AGGCGAATTATTCAGGAATCATAATTACAAATAAACAGCCAT		
1[616]3[629]	AGCTTAGATTAAGAATCATAGGTCTGAGAAGCAACACTATCA		
1[658]3[671]	TTGACCGTAATGGGAACAACCCGTCGGAAACCAAAATAGCGA		
1[700]3[713]	TCGTAACCGTGCATCTTCCTGTAGCCAGAGGGGGTAATAGTA		
1[742]3[755]	TCGGCCTCAGGAAGTTTAACCAATAGGACTGCGGAATCGTCA		
1[952]3[965]	TGTAAAACGACGGCACAAAGGCTATCAGCAAACTCCAACAGG		
1[994]3[1007]	GACTCTAGAGGATCGATAAATTAATGCCCCTTTTGATAAGAG		
1[1036]3[1049]	ATGGTCATAGCTGTACAGTCAAATCACCATTGCTGAATATAA		
1[1078]3[1091]	AATTCCACACAACATAATGTGTAGGTAACAACTAAAGTACGG		
2[307]0[294]	ACAACGCCAACATGGATTTTCAGGTTTACTAAAGCATCACCT		
2[349]0[336]	CAACGCTCAACAGTCTTTTACATCGGGACTGCAACAGTGCCA		
2[391]0[378]	TTAGTATCATATGCTTGAATACCAAGTTCAGAAGATAAAACA		
2[643]0[630]	GTGAATTTATCAAACGCTGAGAAGAGTCCAGGAAAAACGCTC		
2[685]0[672]	AAATGTGAGCGAGTATAGGTCACGTTGGTGCTGGTAATATCC		
2[769]0[756]	AAATCAGCTCATTTATCGCACTCCAGCCCACGCAAATTAACC		
2[979]0[966]	TTTTTGAGAGATCTCAGTGCCAAGCTTGAACCACCACACCCG		
2[1063]0[1050]	GAGAAAGGCCGGAGTTCCTGTGTGAAATTTCCAGTTTGGAAC		
2[1105]0[1092]	ATGCAATGCCTGAGTACGAGCCGGAAGCCTTATAAATCAAAA		
3[294]5[307]	GTTTTGAAGCCTTAACCGATTGAGGGAGAAAGCGCAGTCTCT		
3[336]5[349]	AGCTACAATTTTATTCATATGGTTTACCATGGCTTTTGATGA		
3[378]5[391]	CTTTCCAGAGCCTAACACCACGGAATAAACGGGGTCAGTGCC		
3[420]5[433]	ATTATTTATCCCAAAATACATACATAAAAATGCCCCCTGCCT		
3[630]5[643]	TAACCCTCGTTTACAAAAGGAATTACGATCAGAACCGCCACC		
3[672]5[685]	GAGGCTTTTGCAAAATTTAGGAATACCAATAGCAAGCCCAAT		
3[714]5[727]	AAATGTTTAGACTGACGGAACAACATTATCGTCACCAGTACA		
3[756]5[769]	TAAATATTCATTGACAGGACGTTGGGAACAGCCCTCATAGTT		
3[966]5[979]	TCAGGATTAGAGAGGCTGGCTGACCTTCGGTTTATCAGCTTG		
3[1008]5[1021]	GTCATTTTTGCGGATTGAAAGAGGACAGGATACCGATAGTTG		

3[1050]5[1063]	TGCTGTAGCTCAACAGGCGCAGACGGTCCGCATAACCGATAT		
3[1092]5[1105]	TGTCTGGAAGTTTCTGTCGAAATCCGCGAAAGGCCGCTTTTG		
4[279]2[266]	GACGGAAATTATTCTTTTAGCGAACCTCGCATTTTCGAGCCA		
4[321]2[308]	AAGGGCGACATTCAAATCAAGATTAGTTAATCGCCATATTTA		
4[363]2[350]	AATCAATAGAAAATCCTGAATCTTACCATACCAGTATAAAGC		
4[405]2[392]	AAAGAAACGCAAAGATTTGCCAGTTACATAGAAAAAGCCTGT		
4[615]2[602]	ATAAGAGCAAGAAAATTGAGTTAAGCCCAGACTACCTTTTTA		
4[657]2[644]	AGATACATAACGCCCAGACGACGATAAATTCTCCGTGGGATA		
4[699]2[686]	ATTCATCAGTTGAGAGAAGTTTTGCCAGCTTTCATCAACATT		
4[741]2[728]	AATAAAACGAACTAGATAGCGTCCAATAACGCCATCAAAAAT		
4[951]2[938]	TGACAAGAACCGGAAACCAGACCGGAAGGTCATTGCCTGAGA		
4[993]2[980]	GACCAGGCGCATAGTACCTTTAATTGCTGGAGAGGGTAGCTA		
4[1035]2[1022]	CGAACTGACCAACTTGGCTTAGAGCTTAATCAATATGATATT		
4[1077]2[1064]	ACTTAGCCGGAACGATGTTTTAAATATGAGATTCAAAAGGGT		
5[266]1[279]	AAATAAATCCTCATTATCAAACCCTCAACGTAAAACAGAAAT		
5[308]1[321]	GAATTTACCGTTCCAGCAAATGAAAAATACGTCAGATGAATA		
5[350]1[363]	TACAGGAGTGTACTAGTATTAACACCGCGAAACAATAACGGA		
5[392]1[405]	TTGAGTAACAGTGCGAACGAACCACCAGACAAAATCGCGCAG		
5[602]1[615]	AGTACCGCCACCCTATTTTGACGCTCAAGAAAACATAGCGAT		
5[644]1[657]	CTCAGAGCCACCACGCCAGCCATTGCAAAAACAAACGGCGGA		
5[686]1[699]	AGGAACCCATGTACCAAACTATCGGCCTTGTAGATGGGCGCA		
5[728]1[741]	AACTACAACGCCTGTCTTTGATTAGTAAGGACGACGACAGTA		
5[938]1[951]	GGCTCCAAAAGGAGCCGCTACAGGGCGCCCCAGTCACGACGT		
5[980]1[993]	CTTTCGAGGTGAATTCAACGCTGCGCGTCATGCCTGCAGGTC		
5[1022]1[1035]	CGCCGACAATGACAAAAGAACGTGGACTTCGAATTCGTAATC		
5[1064]1[1077]	ATTCGGTCGCTGAGAGGGTTGAGTGTTGTGTTATCCGCTCAC		
0[251]4[238]	GTTGGCAAATCAACCAGACGATTGGCCTATCACCGTCACCGA		1
0[587]4[574]	TTATTTACATTGGCTAGGTGTATCACCGTAGCTATCTTACCG		2
0[923]4[910]	TTGACGAGCACGTATTTTCACGTTGAAAATCAACGTAACAAA		3
1[238]3[251]	AGAACCTACCATATTAAAGTACCGACAATTATCCGGTATTCT		1
1[574]3[587]	TATTAATTAATTTTTTGGGTTATATAACGCTAATATCAGAGA		2
1[910]3[923]	CGATTAAGTTGGGTAGAGAATCGATGAAATCGCGTTTTAATT		3
2[265]0[252]	GTAATAAGAGAATACAAAATTATTTGCATCAATATCTGGTCA		1
2[601]0[588]	ACCTCCGGCTTAGGCCCTTAGAATCCTTTCGTCTGAAATGGA		2
2[937]0[924]	GTCTGGAGCAAACAAACGCCAGGGTTTTGTACTATGGTTGCT		3
3[252]5[265]	AAGAACGCGAGGCGATTAAAGGTGAATTTGATATTCACAAAC		1
3[588]5[601]	GATAACCCACAAGACAATGAAATAGCAATACTCAGGAGGTTT		2
3[924]5[937]	CGAGCTTCAAAGCGTATTCATTACCCAAATCTCCAAAAAAAA		3
4[237]2[224]	CTTGAGCCATTTGGCAGATATAGAAGGCAAGGTAAAGTAATT		1
4[573]2[560]	AAGCCCTTTTTAAGGAGGGTAATTGAGCTATATGTAAATGCT		2
4[909]2[896]	GCTGCTCATTCAGTGAAAGACTTCAAATCGGTAATCGTAAAA		3
5[224]1[237]	AGGTTGAGGCAGGTAGTTGAAAGGAATTATAATGGAAGGGTT		1
5[560]1[573]	AGTATAGCCCGGAAAGATTCACCAGTCACTGTAAATCGTCGC		2
5[896]1[909]	TGCGAATAATAATTTAACGTGCTTTCCTGATGTGCTGCAAGG		3
0[167]4[154]	TCAATAGATAATACCACCCTCAGAGCCAGTCACCAATGAAAC		1

0[503]4[490]	GAAAGCGTAAGAATGAGAAGGATTAGGAATAACGGAATACCC		2
0[839]4[826]	AACGGTACGCCAGATATGGGATTTTGCTTGGTTTAATTTCAA		3
0[1175]4[1162]	GCGGTCCACGCTGGACAGAGGCTTTGAGTGACCCCCAGCGAT		4
1[154]3[167]	ATCATATTCCTGATATCAACAATAGATATCATCGAGAACAAG		1
1[490]3[503]	TTTCATTTGAATTAAGTTAATTTCATCTAGAGAGAATAACAT		2
1[826]3[839]	AGGCTGCGCAACTGAAGATTGTATAAGCTCAGGTCTTTACCC		3
1[1162]3[1175]	GCGCTCACTGCCCGAATACTTTTGCGGGTAGATACATTTCGC		4
2[517]0[504]	TTCAAATATATTTTCCTTTTTTAATGGAAACCCTTCTGACCT		2
2[853]0[840]	GCCCCAAAAACAGGTTGGGAAGGGCGATGGATTTTAGACAGG		3
2[1189]0[1176]	CATTATGACCCTGTCTTTCCAGTCGGGAGAGAGTTGCAGCAA		4
3[168]5[181]	CAAGCCGTTTTTATGCAAGGCCGGAAACCCACCCTCAGAGCC		1
3[504]5[517]	AAAAACAGGGAAGCGAGGAAACGCAATATTAGCGGGGTTTTG		2
3[840]5[853]	TGACTATTATAGTCAATTGGGCTTGAGAAAACAACTTTCAAC		3
3[1176]5[1189]	AAATGGTCAATAACAAAACACTCATCTTGACTAAAGACTTTT		4
4[153]2[140]	CATCGATAGCAGCAACCAAGTACCGCACAGTCCTGAACAAGA		1
4[489]2[476]	AAAAGAACTGGCATATAGCAGCCTTTACTCTGACCTAAATTT		2
4[825]2[812]	CTTTAATCATTGTGCCATAAATCAAAAAAAATATTTAAATTG		3
4[1161]2[1148]	TATACCAAGCGCGATTTAGTTTGACCATAGAAGCCTTTATTT		4
5[140]1[153]	ACCCTCAGAACCGCATTTGAGGATTTAGGGAGCGGAATTATC		1
5[476]1[489]	CTGAGACTCCTCAAACGTGGCACAGACAATTACATTTAACAA		2
5[812]1[825]	AAATGAATTTTCTGATCCTGAGAAGTGTGCCATTCGCCATTC		3
5[1148]1[1161]	GGGTAGCAACGGCTTTTGCCCCAGCAGGACATTAATTGCGTT		4
0[209]4[196]	AAAATATCTTTAGGCACCAGAGCCGCCGAAAATCACCAGTAG		1
0[545]4[532]	AAAGGGACATTCTGGATAAGTGCCGTCGAGCCGAACAAAGTT		2
0[881]4[868]	CGGGAGCTAAACAGGAGAATAGAAAGGACTGACGAGAAACAC		3
0[1217]4[1204]	CTGATTGCCCTTCACCATTAAACGGGTATAAAACGAAAGAGG		4
1[196]3[209]	ATCCTGATTGTTTGACAATAAACAACATTCATTACCGCGCCC		1
1[532]3[545]	CAATATATGTGAGTCGCAAGACAAAGAAAATTAACTGAACAC		2
1[868]3[881]	TCGCTATTACGCCAATATGTACCCCGGTTTGCATCAAAAAGA		3
1[1204]3[1217]	GCTGCATTAATGAACATAAAGCTAAATCTTTCATTTGGGGCG		4
2[223]0[210]	CTGTCCAGACGACGGATTATACTTCTGAGAGGAAGGTTATCT		1
2[559]0[546]	GATGCAAATCCAATGAATAACCTTGCTTCACGACCAGTAATA		2
2[895]0[882]	CTAGCATGTCAATCGCTGGCGAAAGGGGCGTTAGAATCAGAG		3
2[1231]0[1218]	ATAAAGCCTCAGAGTCGGCCAACGCGCGGAGACGGGCAACAG		4
3[210]5[223]	AATAGCAAGCAAATGAATTAGAGCCAGCCCAGCATTGACAGG		1
3[546]5[559]	CCTGAACAAAGTCAAAAAGTAAGCAGATAGAGGGTTGATATA		2
3[882]5[895]	TTAAGAGGAAGCCCGAATAAGGCTTGCCACAACTAAAGGAAT		3
3[1218]5[1231]	CGAGCTGAAAAGGTCGAAGGCACCAACCAAATACGTAATGCC		4
4[195]2[182]	CACCATTACCATTATTTCATCGTAGGAAGTTCAGCTAATGCA		1
4[531]2[518]	ACCAGAAGGAAACCGCATTAGACGGGAGCGCGAGAAAACTTT		2
4[867]2[854]	CAGAACGAGTAGTAAGAAGCAAAGCGGATGATAATCAGAAAA		3
4[1203]2[1190]	CAAAAGAATACACTCTGTTTAGCTATATGGTTGTACCAAAAA		4
5[182]1[195]	GCCACCAGAACCACAGCACTAACAACTAAATTCATCAATATA		1
5[518]1[531]	CTCAGTACCAGGCGGCCAACAGAGATAGAACAGTACATAAAT		2
5[854]1[867]	AGTTTCAGCGGAGTGAGGCCGATTAAAGCGGTGCGGGCCTCT		3

5[1190]1[1203]	TCATGAGGAAGTTTCCGCCTGGCCCTGAAACCTGTCGTGCCA		4
0[125]4[112]	ACAAACAATTCGACCACCGGAACCGCCTGACAGAATCAAGTT		1
0[461]4[448]	CTATTAGTCTTTAATATTCTGAAACATGATTACGCAGTATGT		2
0[797]4[784]	AGGCCACCGAGTAAAAGTTTTGTCGTCTATTTTAAGAACTGG		3
0[1133]4[1120]	GATGGTGGTTCCGATCAGCAGCGAAAGAGAGATTTGTATCAT		4
1[112]3[125]	ATTTTGCGGAACAATCCTAATTTACGAGTCCTTATCATTCCA		1
1[448]3[461]	TGAAACAAACATCACCGACCGTGTGATAGATTTTTTGTTTAA		2
1[784]3[797]	GCTTCTGGTGCCGGTTGTTAAAATTCGCCTTTAAACAGTTCA		3
1[1120]3[1133]	CTGGGGTGCCTAATAAATTTTTAGAACCTTGATTCCCAATTC		4
2[139]0[126]	AAAATAATATCCCAAGAAACCACCAGAAAAGTATTAGACTTT		1
2[475]0[462]	AATGGTTTGAAATAAGAAAACAAAATTAATATTTTTGAATGG		2
2[811]0[798]	TAAACGTTAATATTAAACCAGGCAAAGCTTTTATAATCAGTG		3
2[1147]0[1134]	CAACGCAAGGATAAGAGTGAGCTAACTCCGAAAATCCTGTTT		4
3[126]5[139]	AGAACGGGTATTAACCGTAATCAGTAGCCCCTCAGAGCCGCC		1
3[462]5[475]	CGTCAAAAATGAAAGATTAAGACTCCTTAAAGTATTAAGAGG		2
3[798]5[811]	GAAAACGAGAATGAAATTACCTTATGCGTTCCAGACGTTAGT		3
3[1134]5[1147]	TGCGAACGAGTAGAAACAAAGTACAACGCAGCATCGGAACGA		4
4[111]2[98]	TGCCTTTAGCGTCATAATCGGCTGTCTTCATGTAGAAACCAA		1
4[447]2[434]	TAGCAAACGTAGAATCCAAATAAGAAACAATAAGGCGTTAAA		2
4[783]2[770]	CTCATTATACCAGTATCCCCCTCAAATGATTAAATTTTTGTT		3
4[1119]2[1106]	CGCCTGATAAATTGATTCCATATAACAGCTCATATATTTTAA		4
5[98]1[111]	GGAACCAGAGCCACAACTCGTATTAAATTGAGTAACATTATC		1
5[434]1[447]	ATTTCGGAACCTATTGCGCGAACTGATAGCAAAAGAAGATGA		2
5[770]1[783]	AGCGTAACGATCTAAAGAGTCTGTCCATAGCTTTCCGGCACC		3
5[1106]1[1119]	CGGGATCGTCACCCAATCGGCAAAATCCATAAAGTGTAAAGC		4
2[181]0[168]	GAACGCGCCTGTTTTATCAGATGATGGCATAGATTAGAGCCG		Biotin
2[433]0[420]	TAAGAATAAACACCTTTCAATTACCTGAGCCCTAAAACATCG		Biotin
2[727]0[714]	AATTCGCGTCTGGCCTGCCAGTTTGAGGTAACATCACTTGCC		Biotin
2[1021]0[1008]	CAACCGTTCTAGCTCCCGGGTACCGAGCCCAACGTCAAAGGG		Biotin
2[1271]2[1232]	AAATCATACAGGCAAGGCAAAGAATTAGCAAAATTAAGCA		Biotin

Supplementary Table 3 | DNA Exchange extensions and labeling strand sequences

Name	Sequence	Corresponding "color" or antibody
S1 anti-handle	5′-GTGGTAGAGGAA-dye-3′	
S2 anti-handle	5′-GTTAGGAATGTTA-dye-3′	
S3 anti-handle	5′-TGGTGAGGGATT-dye-3′	
S4 anti-handle	5′-AGGTGGTAAGTT-dye-3′	
Biotin anti-handle	5′-GAATCGGTCACAGTACAACCG	
S1 handle	5′-TTCCTCTACCAC-3′	Anti-rat (alpha-tubulin)
S2 handle	5′-TAACATTCCTAAC-3′	Anti-goat (LaminB)
S3 handle	5′-AATCCCTCACCA-3′	Anti-rabbit (TOM20)
S4 handle	5′-AACTTACCACCT-3′	
Biotin handle	5′-Staple-TTCGGTTGTACTGTGACCGATTC-3′	

Supplementary Table 4 | Scaffold strand sequence M13mp18

TGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGGC
TATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGC
AACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCT
CTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAG
CTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGC
TATGACCATGATTACGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAACG
TCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGA
TCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGA
TCTTCCTGAGGCCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTGACCTATCCCATTACGGT
CAATCCGCCGTTTGTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAAT
TATTTTTGATGGCGTTCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAATGCGAATTTTAACAAAATATTAACGTTTACAATTTAAATA
TTTGCTTATACAATCTTCCTGTTTTTGGGGCTTTTCTGATTATCAACCGGGGTACATATGATTGACATGCTAGTTTTACGATTACCGTTCATCGAT
TCTCTTGTTTGCTCCAGACTCTCAGGCAATGACCTGATAGCCTTTGTAGATCTCTCAAAAATAGCTACCCTCTCCGGCATTAATTTATCAGCTAGA
ACGGTTGAATATCATATTGATGGTGATTTGACTGTCTCCGGCCTTTCTCACCCTTTTGAATCTTTACCTACACATTACTCAGGCATTGCATTTAAA
ATATATGAGGGTTCTAAAAATTTTTATCCTTGCGTTGAAATAAAGGCTTCTCCCGCAAAAGTATTACAGGGTCATAATGTTTTTGGTACAACCGAT
TTAGCTTTATGCTCTGAGGCTTTATTGCTTAATTTTGCTAATTCTTTGCCTTGCCTGTATGATTTATTGGATGTTAATGCTACTACTATTAGTAGA
ATTGATGCCACCTTTTCAGCTCGCGCCCCAAATGAAAATATAGCTAAACAGGTTATTGACCATTTGCGAAATGTATCTAATGGTCAAACTAAATCT
ACTCGTTCGCAGAATTGGGAATCAACTGTTATATGGAATGAAACTTCCAGACACCGTACTTTAGTTGCATATTTAAAACATGTTGAGCTACAGCAT
TATATTCAGCAATTAAGCTCTAAGCCATCCGCAAAAATGACCTCTTATCAAAAGGAGCAATTAAAGGTACTCTCTAATCCTGACCTGTTGGAGTTT
GCTTCCGGTCTGGTTCGCTTTGAAGCTCGAATTAAAACGCGATATTTGAAGTCTTTCGGGCTTCCTCTTAATCTTTTTGATGCAATCCGCTTTGCT
TCTGACTATAATAGTCAGGGTAAAGACCTGATTTTTGATTTATGGTCATTCTCGTTTTCTGAACTGTTTAAAGCATTTGAGGGGGATTCAATGAAT
ATTTATGACGATTCCGCAGTATTGGACGCTATCCAGTCTAAACATTTTACTATTACCCCCTCTGGCAAAACTTCTTTTGCAAAAGCCTCTCGCTAT
TTTGGTTTTTATCGTCGTCTGGTAAACGAGGGTTATGATAGTGTTGCTCTTACTATGCCTCGTAATTCCTTTTGGCGTTATGTATCTGCATTAGTT
GAATGTGGTATTCCTAAATCTCAACTGATGAATCTTTCTACCTGTAATAATGTTGTTCCGTTAGTTCGTTTTATTAACGTAGATTTTTCTTCCCAA
CGTCCTGACTGGTATAATGAGCCAGTTCTTAAAATCGCATAAGGTAATTCACAATGATTAAAGTTGAAATTAAACCATCTCAAGCCCAATTTACTA
CTCGTTCTGGTGTTTCTCGTCAGGGCAAGCCTTATTCACTGAATGAGCAGCTTTGTTACGTTGATTTGGGTAATGAATATCCGGTTCTTGTCAAGA
TTACTCTTGATGAAGGTCAGCCAGCCTATGCGCCTGGTCTGTACACCGTTCATCTGTCCTCTTTCAAAGTTGGTCAGTTCGGTTCCCTTATGATTG
ACCGTCTGCGCCTCGTTCCGGCTAAGTAACATGGAGCAGGTCGCGGATTTCGACACAATTTATCAGGCGATGATACAAATCTCCGTTGTACTTTGT
TTCGCGCTTGGTATAATCGCTGGGGGTCAAAGATGAGTGTTTTAGTGTATTCTTTTGCCTCTTTCGTTTTAGGTTGGTGCCTTCGTAGTGGCATTA
CGTATTTTACCCGTTTAATGGAAACTTCCTCATGAAAAAGTCTTTAGTCCTCAAAGCCTCTGTAGCCGTTGCTACCCTCGTTCCGATGCTGTCTTT
CGCTGCTGAGGGTGACGATCCCGCAAAAGCGGCCTTTAACTCCCTGCAAGCCTCAGCGACCGAATATATCGGTTATGCGTGGGCGATGGTTGTTGT
CATTGTCGGCGCAACTATCGGTATCAAGCTGTTTAAGAAATTCACCTCGAAAGCAAGCTGATAAACCGATACAATTAAAGGCTCCTTTTGGAGCCT
TTTTTTTGGAGATTTTCAACGTGAAAAAATTATTATTCGCAATTCCTTTAGTTGTTCCTTTCTATTCTCACTCCGCTGAAACTGTTGAAAGTTGTT
TAGCAAAATCCCATACAGAAAATTCATTTACTAACGTCTGGAAAGACGACAAAACTTTAGATCGTTACGCTAACTATGAGGGCTGTCTGTGGAATG
CTACAGGCGTTGTAGTTTGTACTGGTGACGAAACTCAGTGTTACGGTACATGGGTTCCTATTGGGCTTGCTATCCCTGAAAATGAGGGTGGTGGCT
CTGAGGGTGGCGGTTCTGAGGGTGGCGGTTCTGAGGGTGGCGGTACTAAACCTCCTGAGTACGGTGATACACCTATTCCGGGCTATACTTATATCA
ACCCTCTCGACGGCACTTATCCGCCTGGTACTGAGCAAAACCCCGCTAATCCTAATCCTTCTCTTGAGGAGTCTCAGCCTCTTAATACTTTCATGT
TTCAGAATAATAGGTTCCGAAATAGGCAGGGGGCATTAACTGTTTATACGGGCACTGTTACTCAAGGCACTGACCCCGTTAAAACTTATTACCAGT
ACACTCCTGTATCATCAAAAGCCATGTATGACGCTTACTGGAACGGTAAATTCAGAGACTGCGCTTTCCATTCTGGCTTTAATGAGGATTTATTTG
TTTGTGAATATCAAGGCCAATCGTCTGACCTGCCTCAACCTCCTGTCAATGCTGGCGGCGGCTCTGGTGGTGGTTCTGGTGGCGGCTCTGAGGGTG
GTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGCTCTGAGGGAGGCGGTTCCGGTGGTGGCTCTGGTTCCGGTGATTTTGATTATGAAAAGATGG
CAAACGCTAATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTACAGTCTGACGCTAAAGGCAAACTTGATTCTGTCGCTACTGATTACG
GTGCTGCTATCGATGGTTTCATTGGTGACGTTTCCGGCCTTGCTAATGGTAATGGTGCTACTGGTGATTTTGCTGGCTCTAATTCCCAAATGGCTC
AAGTCGGTGACGGTGATAATTCACCTTTAATGAATAATTTCCGTCAATATTTACCTTCCCTCCCTCAATCGGTTGAATGTCGCCCTTTTGTCTTTG
GCGCTGGTAAACCATATGAATTTTCTATTGATTGTGACAAAATAAACTTATTCCGTGGTGTCTTTGCGTTTCTTTTATATGTTGCCACCTTTATGT
ATGTATTTTCTACGTTTGCTAACATACTGCGTAATAAGGAGTCTTAATCATGCCAGTTCTTTTGGGTATTCCGTTATTATTGCGTTTCCTCGGTTT
CCTTCTGGTAACTTTGTTCGGCTATCTGCTTACTTTTCTTAAAAAGGGCTTCGGTAAGATAGCTATTGCTATTTCATTGTTTCTTGCTCTTATTAT
TGGGCTTAACTCAATTCTTGTGGGTTATCTCTCTGATATTAGCGCTCAATTACCCTCTGACTTTGTTCAGGGTGTTCAGTTAATTCTCCCGTCTAA
TGCGCTTCCCTGTTTTTATGTTATTCTCTCTGTAAAGGCTGCTATTTTCATTTTTGACGTTAAACAAAAAATCGTTTCTTATTTGGATTGGGATAA
ATAATATGGCTGTTTATTTTGTAACTGGCAAATTAGGCTCTGGAAAGACGCTCGTTAGCGTTGGTAAGATTCAGGATAAAATTGTAGCTGGGTGCA
AAATAGCAACTAATCTTGATTTAAGGCTTCAAAACCTCCCGCAAGTCGGGAGGTTCGCTAAAACGCCTCGCGTTCTTAGAATACCGGATAAGCCTT
CTATATCTGATTTGCTTGCTATTGGGCGCGGTAATGATTCCTACGATGAAAATAAAAACGGCTTGCTTGTTCTCGATGAGTGCGGTACTTGGTTTA
ATACCCGTTCTTGGAATGATAAGGAAAGACAGCCGATTATTGATTGGTTTCTACATGCTCGTAAATTAGGATGGGATATTATTTTTCTTGTTCAGG
ACTTATCTATTGTTGATAAACAGGCGCGTTCTGCATTAGCTGAACATGTTGTTTATTGTCGTCGTCTGGACAGAATTACTTTACCTTTTGTCGGTA
CTTTATATTCTCTTATTACTGGCTCGAAAATGCCTCTGCCTAAATTACATGTTGGCGTTGTTAAATATGGCGATTCTCAATTAAGCCCTACTGTTG
AGCGTTGGCTTTATACTGGTAAGAATTTGTATAACGCATATGATACTAAACAGGCTTTTTCTAGTAATTATGATTCCGGTGTTTATTCTTATTTAA
CGCCTTATTTATCACACGGTCGGTATTTCAAACCATTAAATTTAGGTCAGAAGATGAAATTAACTAAAATATATTTGAAAAAGTTTTCTCGCGTTC
TTTGTCTTGCGATTGGATTTGCATCAGCATTTACATATAGTTATATAACCCAACCTAAGCCGGAGGTTAAAAAGGTAGTCTCTCAGACCTATGATT
TTGATAAATTCACTATTGACTCTTCTCAGCGTCTTAATCTAAGCTATCGCTATGTTTTCAAGGATTCTAAGGGAAAATTAATTAATAGCGACGATT
TACAGAAGCAAGGTTATTCACTCACATATATTGATTTATGTACTGTTTCCATTAAAAAAGGTAATTCAAATGAAATTGTTAAATGTAATTAATTTTT
GTTTTCTTGATGTTTGTTTCATCATCTTCTTTTGCTCAGGTAATTGAAATGAATAATTCGCCTCTGCGCGATTTTGTAACTTGGTATTCAAAGCAA
TCAGGCGAATCCGTTATTGTTTCTCCCGATGTAAAAGGTACTGTTACTGTATATTCATCTGACGTTAAACCTGAAAATCTACGCAATTTCTTTATT
TCTGTTTTACGTGCAAATAATTTTGATATGGTAGGTTCTAACCCTTCCATTATTCAGAAGTATAATCCAAACAATCAGGATTATATTGATGAATTG
CCATCATCTGATAATCAGGAATATGATGATAATTCCGCTCCTTCTGGTGGTTTCTTTGTTCCGCAAAATGATAATGTTACTCAAACTTTTAAAATT
AATAACGTTCGGGCAAAGGATTTAATACGAGTTGTCGAATTGTTTGTAAAGTCTAATACTTCTAAATCCTCAAATGTATTATCTATTGACGGCTCT
AATCTATTAGTTGTTAGTGCTCCTAAAGATATTTTAGATAACCTTCCTCAATTCCTTTCAACTGTTGATTTGCCAACTGACCAGATATTGATTGAG
GGTTTGATATTTGAGGTTCAGCAAGGTGATGCTTTAGATTTTTCATTTGCTGCTGGCTCTCAGCGTGGCACTGTTGCAGGCGGTGTTAATACTGAC
CGCCTCACCTCTGTTTTATCTTCTGCTGGTGGTTCGTTCGGTATTTTTAATGGCGATGTTTTAGGGCTATCAGTTCGCGCATTAAAGACTAATAGC
CATTCAAAAATATTGTCTGTGCCACGTATTCTTACGCTTTCAGGTCAGAAGGGTTCTATCTCTGTTGGCCAGAATGTCCCTTTTATTACTGGTCGT
GTGACTGGTGAATCTGCCAATGTAAATAATCCATTTCAGACGATTGAGCGTCAAAATGTAGGTATTTCCATGAGCGTTTTTCCTGTTGCAATGGCT
GGCGGTAATATTGTTCTGGATATTACCAGCAAGGCCGATAGTTTGAGTTCTTCTACTCAGGCAAGTGATGTTATTACTAATCAAAGAAGTATTGCT

```
ACAACGGTTAATTTGCGTGATGGACAGACTCTTTTACTCGGTGGCCTCACTGATTATAAAAACACTTCTCAGGATTCTGGCGTACCGTTCCTGTCT
AAAATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTGATTCTAACGAGGAAAGCACGTTATACGTGCTCGTCAAAGCAACCATAGTACGCGCC
CTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTT
CCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGA
CCCCAAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCC
```

Supplementary References

[1] R. Jungmann, M. S. Avendano, M. Dai, J. B. Woehrstein, S. S. Agasti, Z. Feiger, A. Rodal, P. Yin, *Nat Methods* **2016**, *13*, 439-442.
[2] J. Schnitzbauer, M. T. Strauss, T. Schlichthaerle, F. Schueder, R. Jungmann, *Nat Protoc* **in review**.
[3] C. S. Smith, N. Joseph, B. Rieger, K. A. Lidke, *Nature Methods* **2010**, *7*, 373-U352.
[4] U. Endesfelder, S. Malkusch, F. Fricke, M. Heilemann, *Histochem Cell Biol* **2014**, *141*, 629-638.

2.3 Identification of the elementary structural units of the DNA damage response

2.3.1 Main Paper

nature COMMUNICATIONS

ARTICLE

Received 20 Dec 2016 | Accepted 26 Apr 2017 | Published 12 Jun 2017

DOI: 10.1038/ncomms15760 OPEN

Identification of the elementary structural units of the DNA damage response

Francesco Natale[1,*], Alexander Rapp[1,*], Wei Yu[1,†], Andreas Maiser[2], Hartmann Harz[2], Annina Scholl[1], Stephan Grulich[1], Tobias Anton[2], David Hörl[2], Wei Chen[3], Marco Durante[4,†], Gisela Taucher-Scholz[4], Heinrich Leonhardt[2] & M. Cristina Cardoso[1]

Histone H2AX phosphorylation is an early signalling event triggered by DNA double-strand breaks (DSBs). To elucidate the elementary units of phospho-H2AX-labelled chromatin, we integrate super-resolution microscopy of phospho-H2AX during DNA repair in human cells with genome-wide sequencing analyses. Here we identify phospho-H2AX chromatin domains in the nanometre range with median length of ~75 kb. Correlation analysis with over 60 genomic features shows a time-dependent euchromatin-to-heterochromatin repair trend. After X-ray or CRISPR-Cas9-mediated DSBs, phospho-H2AX-labelled heterochromatin exhibits DNA decondensation while retaining heterochromatic histone marks, indicating that chromatin structural and molecular determinants are uncoupled during repair. The phospho-H2AX nano-domains arrange into higher-order clustered structures of discontinuously phosphorylated chromatin, flanked by CTCF. CTCF knockdown impairs spreading of the phosphorylation throughout the 3D-looped nano-domains. Co-staining of phospho-H2AX with phospho-Ku70 and TUNEL reveals that clusters rather than nano-foci represent single DSBs. Hence, each chromatin loop is a nano-focus, whose clusters correspond to previously known phospho-H2AX foci.

[1] Department of Biology, Technische Universität Darmstadt, 64287 Darmstadt, Germany. [2] Department of Biology II, Center for Integrated Protein Science Munich (CIPSM), LMU Munich, 82152 Planegg-Martinsried, Germany. [3] Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany. [4] Department of Biophysics, GSI Helmholtzzentrum für Schwerionenforschung, 64291 Darmstadt, Germany. * These authors contributed equally to this work. † Present addresses: G5 Lymphocyte Development and Oncogenesis, Immunology Department, Pasteur Institute, 75724 Paris Cedex 15, France (W.Y.); Trento Institute for Fundamental Physics and Application (TIFPA-INFN), via Sommarive 14, 38123 Trento, Italy (M.D.). Correspondence and requests for materials should be addressed to M.C.C. (email: cardoso@bio.tu-darmstadt.de).

DNA double-strand breaks (DSBs) are the most harmful lesions induced by either endogenous (for example, replication) or exogenous (for example, ionizing radiation-IR) genotoxic stress, which may lead to chromosomal aberrations and tumorigenesis if not correctly repaired. To deal with DSBs, cells activate a rapid and hierarchically coordinated signalling cascade known as DNA damage response (DDR), leading to cell cycle arrest and allowing the DNA repair machinery to exert its function. One of the earliest events of DDR is the phosphatidylinositol-3-kinase-like-dependent phosphorylation of serine 139 of histone H2AX (γH2AX)[1], a histone H2A variant whose role at the interface of DNA repair, chromatin structure regulation and cell cycle checkpoint activation[2] is yet to be fully elucidated.

Detection of γH2AX has become the most widely used method for quantification of DSBs and their repair kinetics. Activated DDR, as scored by quantification of nuclear γH2AX focal structures, has been extensively described in both precancerous and cancer cells[3,4]. The majority of these studies were performed by conventional microscopy techniques, including confocal microscopy, and the structures resolved were in the micrometre or sub-micrometre range, with a predicted DNA content in the megabase-pair (Mbp) range. Indeed, γH2AX is proposed to spread up to several Mbps from the original lesion site, in higher eukaryotes[5]. The distribution of such histone modification is neither symmetrical around DSB sites nor uniform on chromatin, as assessed by chromatin immunoprecipitation (ChIP) studies conducted in mammals[6–8] and yeast[9,10]. Such uneven spreading may be accounted for by gene transcription[11], or cohesin complex binding[12], which antagonize γH2AX formation along the chromosomes.

An increasing body of evidence underlines the crucial role of genome topology and chromatin spatial organization in the regulation of biological processes[13]. Recent chromosome conformation capture studies have revealed the complexity of genome architecture, with large compartments in the Mbp range conserved across cell lineages and species[14,15], as well as smaller contact domains with a variable size in the range of a few hundreds of kilobase pairs (kb)[15]. This spatial organization can be dynamic and underlines cell-type-specific networks, possibly driving the expression of specific sets of genes[16] or organizing the replication process[17].

Nonetheless, the three-dimensional (3D) arrangement of γH2AX-decorated chromatin in the nuclear volume and its dynamic evolution during the DDR remains elusive. Here we investigate the DDR over time at nanometre resolution by employing super-resolution microscopy techniques on human cells exposed to X-ray radiation. By overcoming the optical diffraction limit, structured illumination microscopy (3D-SIM)[18] and stimulated-emission-depletion (STED)[19] fluorescence microscopy present high prospecting capacity, thus allowing us to dissect complex structures of γH2AX-decorated chromatin at nanometre resolution (~100 nm). Furthermore, the integration of the microscopy results with CRISPR-Cas-targeted DNA damage, RNAi of the key structural factor CCCTC-binding factor (CTCF), γH2AX ChIP-Seq(uencing) profiles during DDR, and more than 60 genomic features reveal temporal, functional and structural insights into the elementary chromatin units read by the DNA DSB repair machinery.

Results

Cellular system and experimental strategy validation. For our study, we employed HeLa cells, an established human cell line whose (epi)genome is extensively annotated in the context of the ENCODE project (genome.ucsc.edu/ENCODE/). To test the DDR, we assessed the formation of γH2AX before and after exposure to IR. We investigated the early (0.5 h), mid (3 h) and late (24 h) stages of DDR, which, according to earlier reports[20], represent 60–100%, 20–60% and less than 10% of the initial DSBs, respectively. Our confocal immunofluorescence analysis of γH2AX revealed that the show endogenous γH2AX signal. This is frequently observed in cancer cell lines and can be attributed to randomly produced DSBs at stalled and collapsed replication forks[21,22]. On exposure to IR, γH2AX followed the predicted repair kinetics, with nuclear γH2AX fluorescence intensity increasing, and then decreasing over time (Supplementary Fig. 1A). Similar kinetics was observed by western blot analysis (Supplementary Fig. 1B). Together, these methods revealed a four- to eightfold increase in γH2AX signal after IR. Overall, cells were able to activate a DDR and underwent cell cycle arrest, accumulating in S-phase (Supplementary Fig. 1C). No apoptosis was detected (Supplementary Fig. 1D), and 24 h post IR cells were viable, re-entered the cell cycle (Supplementary Fig. 1C) and proliferated, although at a lower rate compared with the mock-irradiated controls (Supplementary Fig. 1E).

To investigate γH2AX kinetics at high resolution, we recorded super-resolution image sets before and during DDR, and acquired γH2AX ChIP-Seq genome-wide data at matching time points (Fig. 1a). In all of our immuno-based approaches, we probed γH2AX-decorated chromatin with the same antibody, whose specificity was verified by slot blot analysis employing the γH2AX-immunizing peptide (Supplementary Fig. 1F). The reproducibility of the sequencing data was assessed and confirmed by comparing biological replicates (Supplementary Fig. 1G).

Super-resolution microscopy of γH2AX kinetics during DDR. To first address the effect of improved optical resolution, we compared the number of γH2AX foci from cells imaged by conventional confocal and 3D-SIM microscopy, and analysed in addition the pseudo-wide-field images re-computed from the same 3D-SIM images, before and after deconvolution (Fig. 1b). A detailed analysis workflow is in the 'Methods' section and summarized in Supplementary Fig. 1H. Compared with confocal images (Fig. 1c), we observed a fivefold increase in foci numbers in pseudo-wide-field images, with an additional twofold increase in deconvolved images (Fig. 1d). Despite employing IR doses that are challenging for conventional confocal microscopy (10 Gy X-ray), the enhanced optical resolution enabled us to resolve thousands of foci, increasing by about one order of magnitude the foci counts compared with the pseudo-wide-field, and about two orders of magnitude when comparing with confocal microscopy (Fig. 1e). Thus, it becomes obvious that a single focus identified by confocal microscopy can be further resolved by 3D-SIM into substructures (Fig. 1b, bottom panels, and Supplementary Fig. 2A), which we referred to as nano-foci. In addition, we controlled the imaging and reconstruction process of 3D-SIM by visual inspection of the reconstructed images in Fourier's space (Supplementary Fig. 2B). No reconstruction artifacts are visible as can be seen from the fast Fourier transformed images, which would contain regular stripe patterns otherwise.

Coherently, we observed a two- to fourfold decrease in the diameters of the segmented objects, when comparing 3D-SIM images with re-computed pseudo-wide-field images, with or without deconvolution, respectively (Supplementary Fig. 3A). Notably, in the 3D-SIM images, the nano-foci diameters were constant during the DDR (median lateral diameter: ~200 nm; Fig. 2a), indicating that we detected the smallest substructures of γH2AX-decorated chromatin at the limit defined by the foci segmentation process (eight voxels). To gauge the actual size of γH2AX nano-foci, we recorded γH2AX immunofluorescence

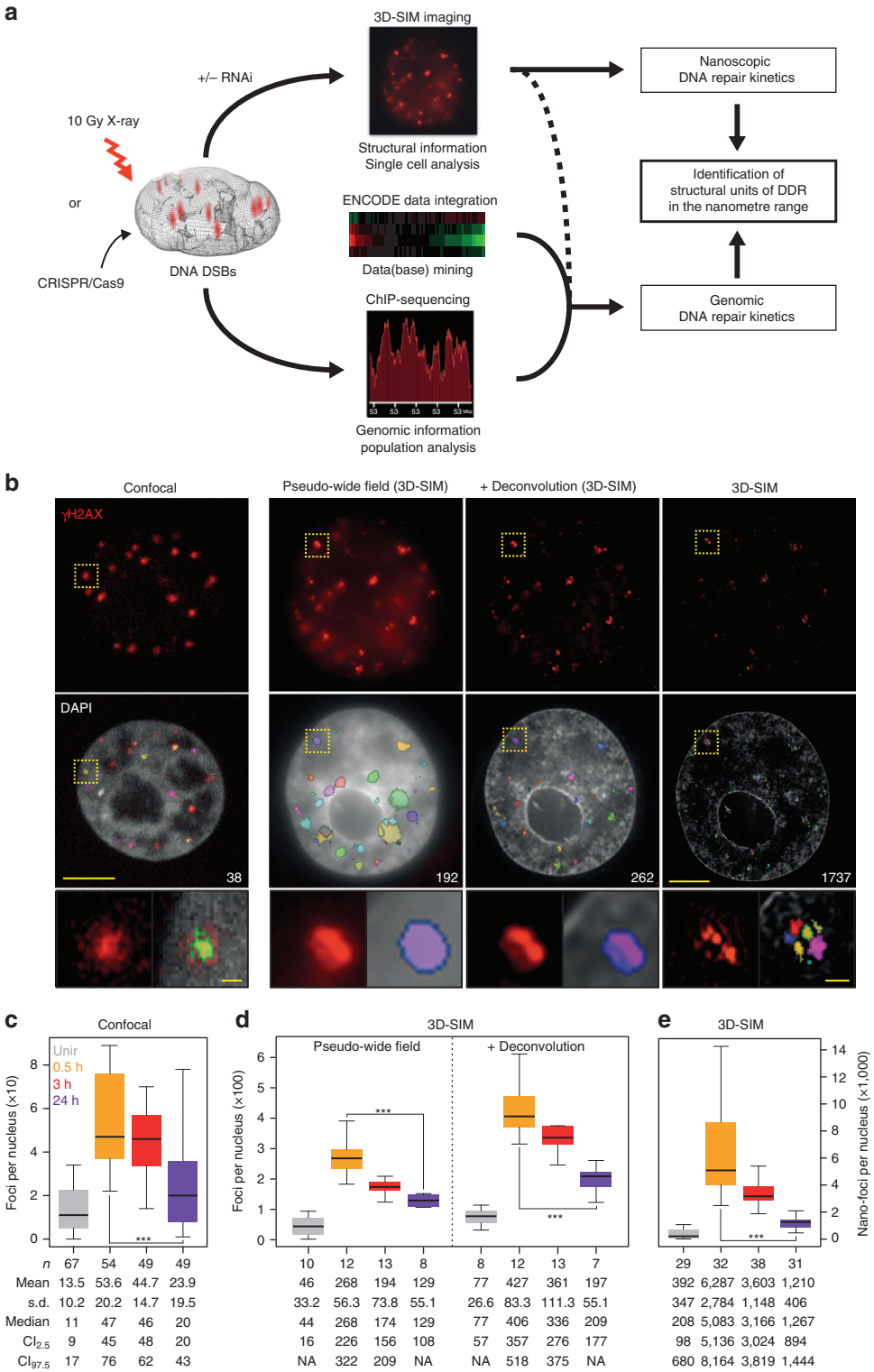


Figure 1 | Characterization of γH2AX foci at different resolution levels. (**a**) Schematics of the experimental approach. (**b**) Mid-nuclear sections of confocal microscopy (*z*: 200 nm) and 3D-SIM (*z*: 125 nm) representative images of cells, 24 h post IR. Only for 3D-SIM, the same exemplary cell is shown as re-computed pseudo-wide-field image before or after deconvolution as well as the original 3D-SIM output. The total number of detected foci (highlighted in colours) in the whole nuclear volume is shown in the DAPI panels. The lower panels show magnified views of the yellow dashed frame. Scale bars, 5 μm and 500 nm for main micrographs and magnified regions, respectively. γH2AX foci number distributions before and during DDR, from confocal images (**c**), 3D-SIM re-computed pseudo-wide-field of identical cell nuclei, before or after deconvolution (**d**) and original 3D-SIM images (**e**). *n*: total number of imaged cells from three independent experiments. All boxes and whiskers represent 25–75 percentiles and three times the IQD. The mean number of foci and corresponding s.d., the median as well as the 95% confidence intervals (CI) for the median are shown below each box. NA: not applicable. For **c**–**e**: one-way ANOVA with Dunnett's correction; $***P<10^{-3}$.

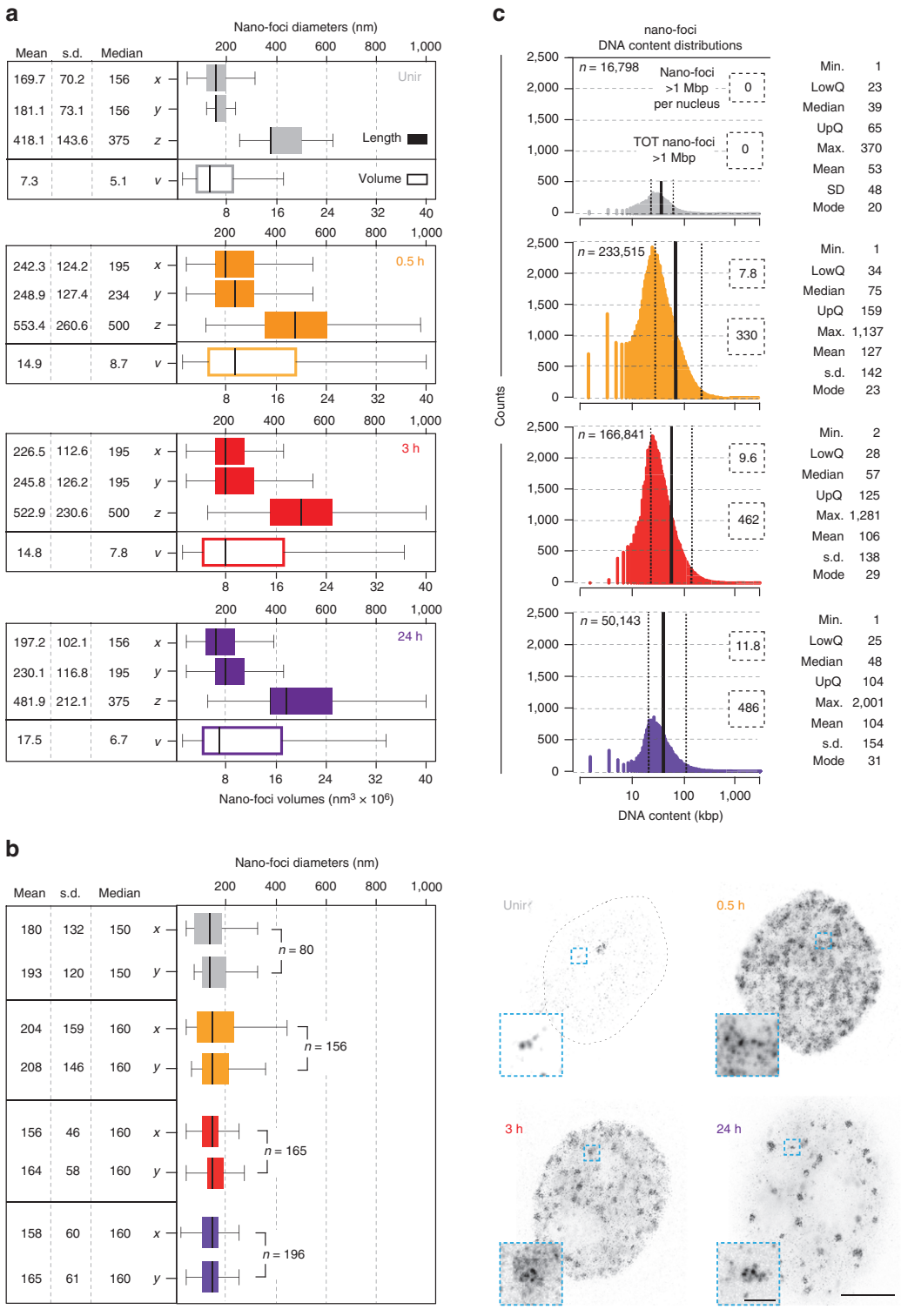


Figure 2 | Metrics of γH2AX nano-foci dimensions and DNA content. (**a**) Quantification of nano-foci diameters in the three dimensions (filled boxes, top) during DDR. From these three dimensions, the volumes were calculated (empty boxes, bottom). The difference between lateral and axial measurements is due to the lower resolution in the axial direction. Figures in nm or $nm^3 \times 10^6$ are shown. (**b**) STED microscopy of γH2AX immunofluorescence. (left) Quantification of lateral diameters of γH2AX nano-foci. Statistics and size scale are as in **a**. (right) Exemplary STED images of cells before and after IR are shown together with the magnified views of the light-blue boxes. Scale bars, 5 μm and 500 nm for main micrographs and magnified regions, respectively. (**c**) DNA content distributions of γH2AX nano-foci before and during DDR. Only in IR-exposed cells, we found nano-foci larger than 1 Mbp (dashed boxes), and their frequency never exceeded 1% (0.14%, 0.28%, 0.95% for 0.5 h, 3 h and 24 h, respectively). Kruskal–Wallis $\chi^2 = 18{,}503$, df = 3, $P < 2.2 \times 10^{-16}$. Statistics (in kb) are shown next to each distribution. All boxes and whiskers are as in Fig. 1. *n*: total number of measured nano-foci from all imaged cells in two independent experiments, for 3D-SIM (**a**,**c**) or STED (**b**).

images by STED microscopy. Compared with our 3D-SIM set-up, STED provided a twofold increase in optical resolution[18]. Yet, the measured lateral diameters (Fig. 2b) were statistically undistinguishable from those recorded by 3D-SIM under sham-irradiation conditions (unpaired two-tailed *t*-test: $P>0.05$). Upon irradiation, the mean lateral diameters imaged by STED were only ~20% smaller than those we measured by 3D-SIM imaging (unpaired two-tailed *t*-test: $P<10^{-3}$). These results validate our 3D-SIM measurements and indicate that γH2AX nano-foci are the chromatin elementary units of the cellular response to DSBs.

Next, to estimate the DNA content of nano-foci, we related the integrated 4,6-diamidino-2-phenylindole (DAPI) intensity of each γH2AX nano-focus to the total DNA content represented by the integrated whole nuclear DAPI intensity (Supplementary Fig. 3B). The resulting DNA fractions were first corrected for the total HeLa genome size (determined by spectral karyotyping, Supplementary Fig. 3C), and then further corrected for the cell cycle phase of each given cell (Supplementary Fig. 3D). Finally, values smaller than the 0.5th and bigger than 99.5th percentile were discarded to avoid artifactual biases. The resulting distributions are shown in Fig. 2c. Before exposure to IR, the interquartile distance (IQD) of the nano-foci DNA content was ~23–65 kb. On IR (0.5 h)—after γH2AX spreading—it increased to ~34–159 kb, with a median length of 75 kb (Fig. 2c and Supplementary Tables 1 and 2).

To provide another line of evidence supporting our 3D-SIM metrics, we produced γH2AX ChIP-Seq profiles under the same experimental conditions employed for the microscopic analysis. Next, we integrated the genomic data with the super-resolution microscopy data to establish a novel combined approach (described in detail in the Methods and Supplementary Fig. 4) and, thus, provide estimates of the γH2AX-decorated chromatin domain size. Overall, the resulting γH2AX genomic domains' size was in good agreement with that of 3D-SIM γH2AX nano-foci, although the former were ~30% smaller (IQD: 10–110 kb at 0.5 h). Because our approach only takes into account the *in cis* contribution to the size of the genomic domains, the difference between the latter and those measured by 3D-SIM can be attributed to inter-chromosomal contribution[23].

The DDR uncouples histone modifications and DNA compaction. To characterize the (epi)genetic composition of γH2AX-decorated chromatin during DDR, we related the ChIP-Seq γH2AX profiles to multiple genomic features, (Supplementary Table 3). First, we computed the density of such genomic features as well as the abundance of γH2AX in 10 kb genomic intervals. Next, we calculated the genome-wide Spearman's ρ correlation coefficient of each feature with γH2AX profiles before and during the DDR (Fig. 3a). The outcome of the analysis showed a strong correlation at early time post IR between γH2AX and euchromatic features such as GC content (Supplementary Fig. 5A; maximum Spearman's ρ: 0.81, $P<2.2\times10^{-16}$), DNase hypersensitivity sites, Regions of IncreaseD Gene Expression (RIDGEs), early replication timing and histone modifications associated with transcriptionally active chromatin state (for example, H3K36me3, H3K4me1/2/3 and H3K9ac). Heterochromatic features, such as AT content (Topo.CAT-YTA-RAK motif), lamin-binding sites, late replication timing, intensity of Giemsa shades and H3K9me3, were negatively correlated to γH2AX, instead. Notably, this trend was inverted at later times, with heterochromatic features correlating to residual γH2AX levels. An exemplary γH2AX profile on chromosome 21 is shown in Fig. 3b. Quantification of γH2AX levels, before and during DDR, in (anti-)RIDGEs, Giemsa shades as well as in H3K36me3- and H3K9me3-decorated chromatin domains is shown in Supplementary Fig. 5B–D and Supplementary Table 4.

To validate and extend these findings at the single-cell level, we recorded 3D-SIM images of γH2AX immunofluorescence combined with either H3K36me3 or H3K9me3 labelling (Fig. 4a). These two histone modifications recapitulate the results from Fig. 3a, with the former being mainly associated with actively transcribed genes[24], while the latter is abundant in heterochromatic (for example, pericentromeric regions) and transcriptionally silent regions[25]. We segmented γH2AX nano-foci as previously described and, in addition, we measured the H3K36me3 or H3K9me3 fluorescence intensity in the volume occupied by γH2AX nano-foci. In the latter, H3K36me3 signal was high at early time points, but not at 24 h post IR, as opposed to H3K9me3 signal, which was low at early time points but higher 24 h post IR (Fig. 4b). We observed similar results when measuring γH2AX fluorescence intensity in the volume of H3K36me3- and H3K9me3-decorated chromatin (Fig. 4c). Together, these findings recapitulate our genomic results, indicating that γH2AX nano-foci are mainly associated to an active chromatin state during the early and mid-stages of DDR, whereas the residual phosphorylation signal is enriched in heterochromatin at later times.

Based on these data, we expected an enrichment of γH2AX nano-foci in compact chromatin (that is, DAPI-dense structures) at later times. However, the mean DAPI content of γH2AX nano-foci remained unvaried over the time, and, if at all, was lower at 24 h (Fig. 4d). In fact, γH2AX nano-foci were located in close proximity to DAPI-dense structures, and the two seldom overlapped. To quantify this, we measured the maximum DAPI intensity in a 3D-region dilated by three voxels in all dimensions around each γH2AX nano-focus, which we referred to as 'shell' (Supplementary Fig. 5E). Shells always presented higher DAPI signal than the nano-foci (Fig. 4e). This is in agreement with previous observations, whereby γH2AX-decorated chromatin was excluded from DAPI-dense structures following DSB induction[26,27]. These findings prompted us to investigate the condensation state of H3K9me3-decorated chromatin after DNA damage induction. On IR, we observed a progressive decrease of DAPI intensity in H3K9me3-decorated chromatin, up to 24 h (Fig. 4f). Such decrease was not observed in H3K36me3-decorated chromatin. Together, this implies that heterochromatic regions underwent DNA decondensation, although they retained their histone marks. To independently validate this finding, we investigated γH2AX and H3K9me3 levels before and after the induction of CRISPR-Cas9-mediated DNA DSBs targeted at heterochromatic murine major satellite repetitive DNA elements, in C2C12 cells (Fig. 5a). These genomic regions are predominantly found at H3K9me3-rich chromatin and are the most condensed chromatin domains in the mouse genome (chromocentres). As early as the ectopically expressed Cas9 was active (>3 h), γH2AX was visible at H3K9me3-decorated chromatin (chromocentres) (Fig. 5b). Quantification of the H3K9me3 and γH2AX fluorescence intensity in the segmented chromocentres revealed that both signals co-localized (Fig. 5c). Next, we analysed the condensation state of Cas9-targeted chromocentres by means of dual-colour STED microscopy and DNA density measurements. On Cas9-mediated DSBs induction, chromocentres were dramatically decondensed (Fig. 5d,e). Remarkably, they retained the γH2AX mark, which was more abundant where the DNA signal was diminished (Fig. 5d). This observation is in agreement with our 3D-SIM data, whereby the γH2AX nano-foci present a partially decondensed state, with diminished DNA levels relative to their surroundings (Fig. 4f).

Taken together, these findings show that γH2AX nano-foci are chromatin units over represented in transcriptionally active

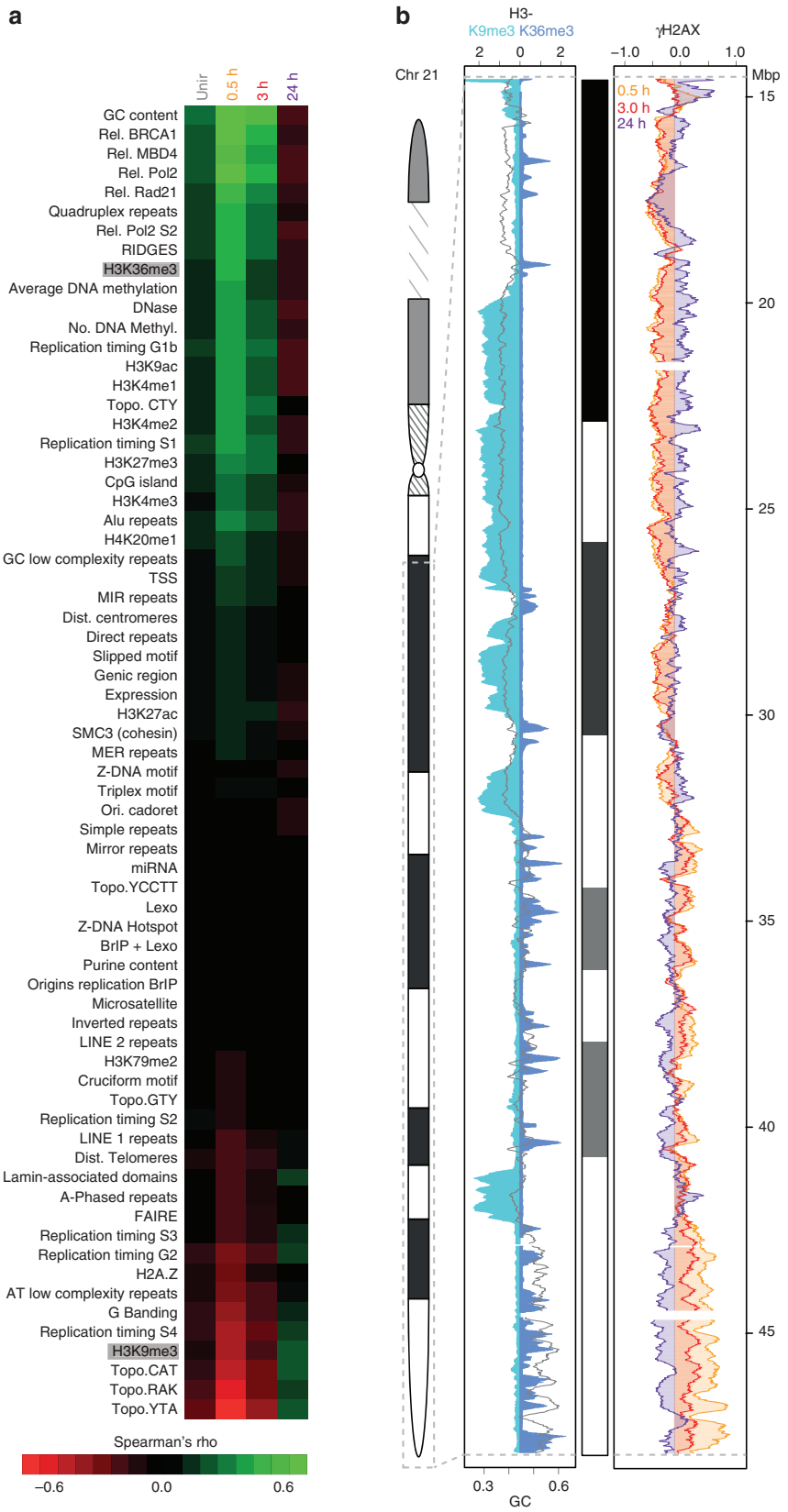


Figure 3 | Temporal correlation of γH2AX ChIP-Seq signal and genomic features. (**a**) Genome-wide correlation between ChIP-Seq γH2AX profiles and genomic features, before and during DDR. Spearman's ρ correlation coefficient is calculated between 10 kb-binned γH2AX profiles and the genomic features (Supplementary Table 3), and colour-coded from red (anti-correlation) to green (correlation). All genomic features are ordered decreasingly, according to the highest correlation value (γH2AX and GC, 0.5 h: 0.81). For all correlations: $P \ll 2.2 \times 10^{-16}$. (**b**) Exemplary ChIP-Seq γH2AX profile on chromosome 21. (left) H3K9me3, H3K36me3 and GC content (grey line); (right) γH2AX levels during DDR.

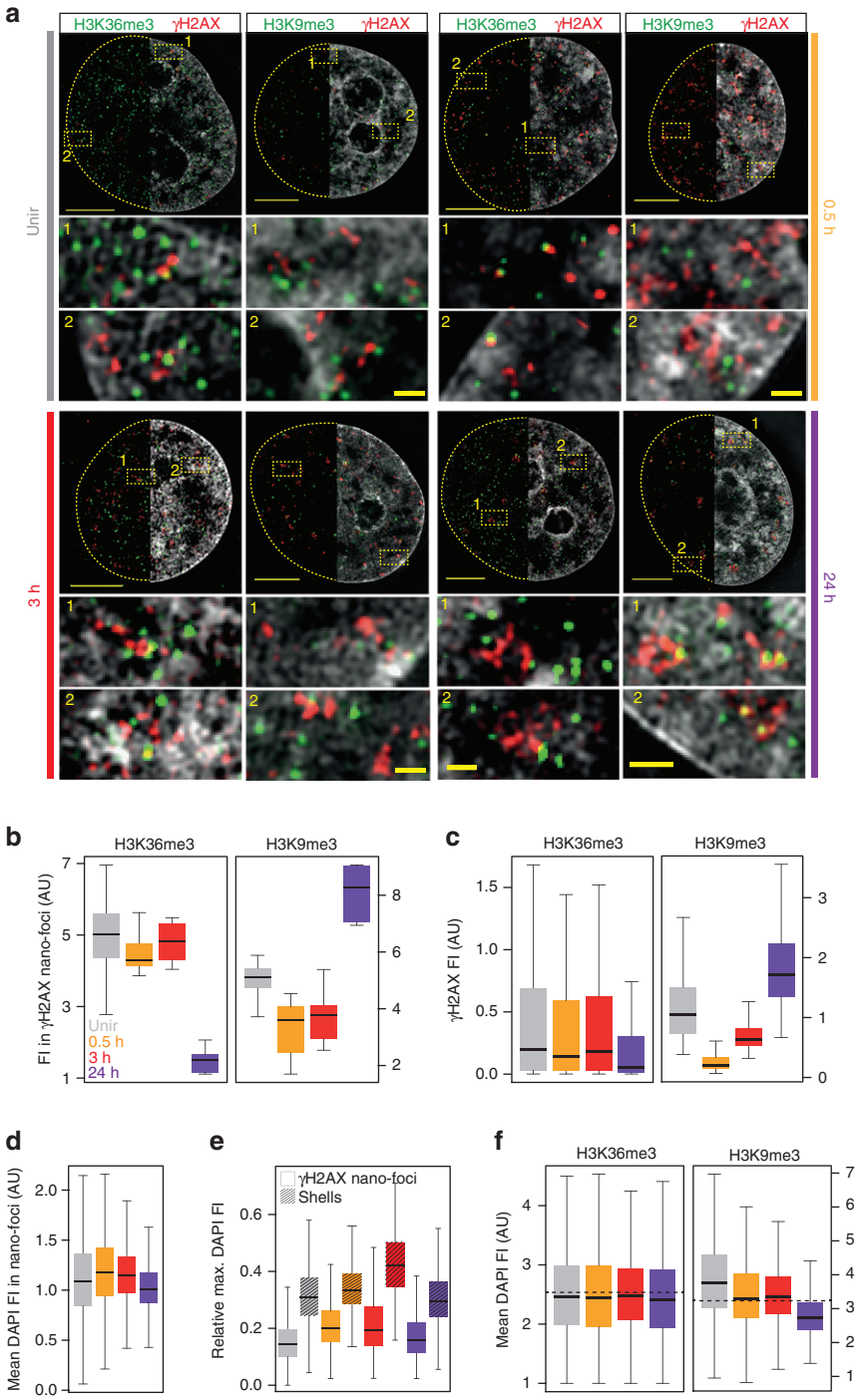


Figure 4 | 3D-SIM chromatin composition analysis of γH2AX nano-foci before and during DDR. (**a**) Exemplary 3D-SIM images of γH2AX (red) and H3K9me3/H3K36me3 (green) co-immunostaining before and after IR. Top panels: mid-nuclear sections showing γH2AX and histone marks with (right half) or without (left half) DAPI counterstaining. The dashed lines depict the nuclear contour. Bottom panels: magnification of the yellow dashed boxes with corresponding reference number. Scale bars, 5 μm and 500 nm for main micrographs and magnified regions, respectively. (**b**) Quantification of the H3K36me3 and H3K9me3 fluorescence intensities measured in γH2AX nano-foci volumes. Kruskal–Wallis $\chi^2 = 19.875$, df = 3, $P = 1.802 \times 10^{-4}$ and Kruskal–Wallis $\chi^2 = 24{,}451$, df = 3, $P = 2.011 \times 10^{-5}$. (**c**) Quantification of the γH2AX fluorescence intensity in H3K36me3- (Kruskal–Wallis $\chi^2 = 261{,}960$, df = 191,020, $P < 2.2 \times 10^{-16}$) and H3K9me3- (Kruskal–Wallis $\chi^2 = 246{,}300$, df = 232,750, $P < 2.2 \times 10^{-16}$) decorated chromatin. (**d**) Mean DAPI intensity in γH2AX nano-foci. Kruskal–Wallis $\chi^2 = 247{,}910$, df = 245,320, $P = 1.129 \times 10^{-4}$. (**e**) Quantification of maximum DAPI intensity in the volume occupied by γH2AX nano-foci (regular boxes) and shells (pattern), relative to the maximum integrated nuclear intensity. Shells represent 3D hollow structures surrounding γH2AX nano-foci (Supplementary Fig. 5E and 'Methods' section). Wilcoxon rank sum all $< 2.2 \times 10^{-16}$. (**f**) Mean DAPI fluorescence intensity in H3K36me3- or H3K9me3-decorated chromatin. Kruskal–Wallis $\chi^2 = 303{,}050$, df = 292,700, $P < 2.2 \times 10^{-16}$ and Kruskal–Wallis $\chi^2 = 25{,}500$, df = 25,002, $P = 0.01338$. Dotted lines: mean DAPI intensity measured over the whole analysed nuclei. All boxes and whiskers are as in Fig. 1. AU: arbitrary units. Results are from two independent experiments.

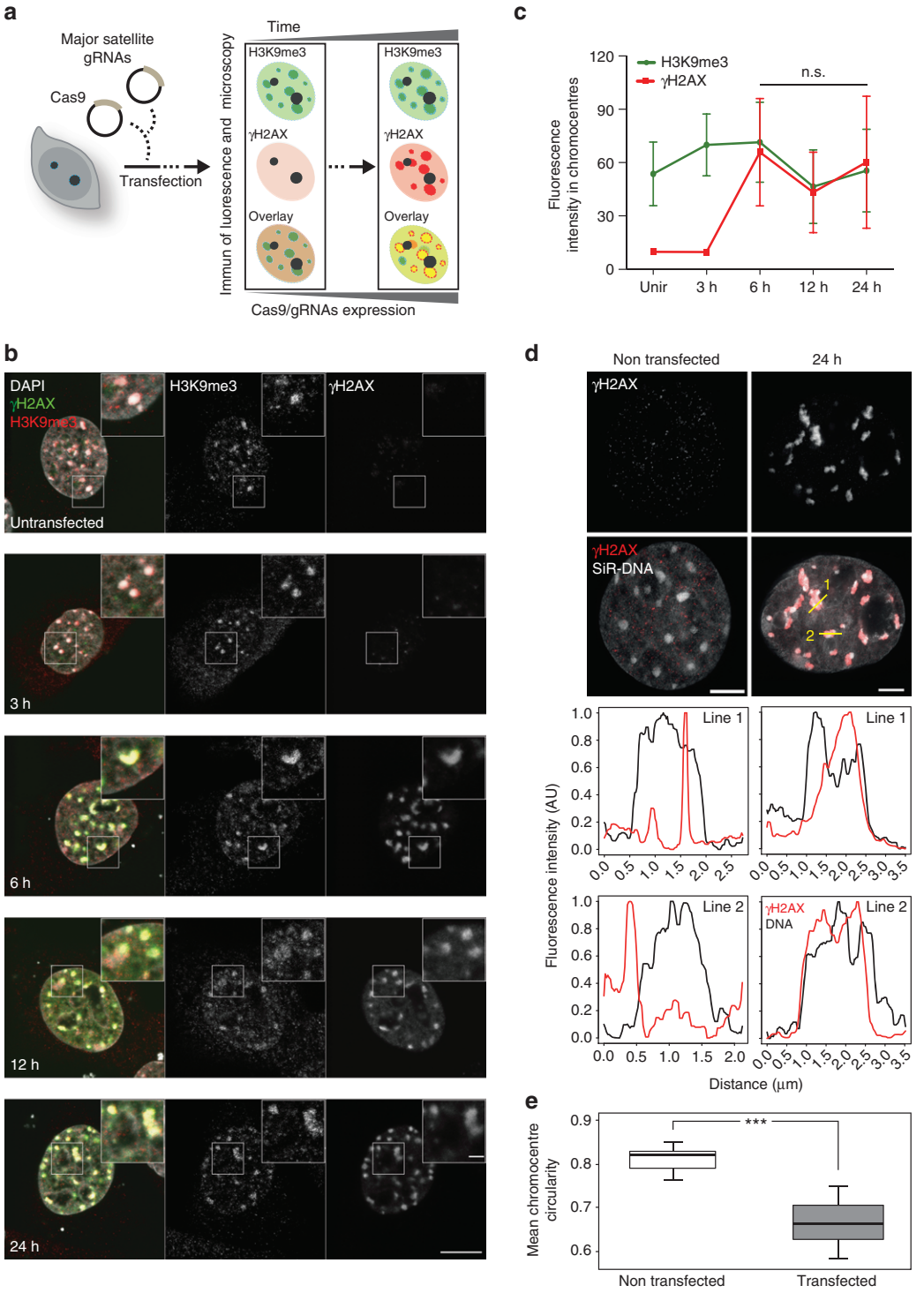


Figure 5 | Analysis of γH2AX and H3K9me3 levels at heterochromatin-targeted CRISPR-Cas9-mediated DSBs. (**a**) Schematics of the CRISPR-Cas9-mediated DSBs induction at murine major satellites DNA. C2C12 cells were transfected with Cas9 and major satellites gRNAs plasmids and fixed after the indicated times. (**b**) Representative immunofluorescence images of γH2AX and H3K9me3 in C2C12 cells. Scale bars, 10 μm and 2 μm for micrograph and inset, respectively. (**c**) Quantification of γH2AX and H3K9me3 fluorescence intensity from DAPI-segmented chromocentres. Mean and s.d. from (**b**) are shown. $n=5$ cells (2–19 chromocentres), 5 cells (6–15 chromocentres), 5 cells (8–19 chromocentres), 5 cells (13–19 chromocentres) and 4 cells (9–19 chromocentres), for untransfected, 3 h, 6 h, 12 h and 24 h time points, respectively. See image analysis in the 'Methods' section for details. (**d**) Representative STED immunofluorescence images of γH2AX and SiR-labelled DNA as indicated. Yellow lines: line profiles (shown below). For the latter, fluorescence intensities were normalized to the min–max range of values of each profile. Lines were smoothed by a 5-window running median. (**e**) Chromocentres decondensation after major satellite-targeted Cas9, assessed as mean chromocentre circularity in transfected ($n=9$) and untransfected ($n=10$) cells. For each cell, the circularity of chromocentres ($>100\,px^2$) within the nucleus was determined as described in the 'Methods' section, yielding shape information for 165 (transfected cells) and 148 (untransfected cells) chromocentres. Statistics: Wilcoxon rank sum test ($^{***}P<10^{-3}$). Scale bar, 2 μm.

regions early on exposure to IR. During the late stage of DDR, they mark heterochromatic regions whose DNA is in a locally decondensed state while keeping the characterizing histone marks (for example, H3K9me3). We propose that by retaining their histone mark, the chromatin identity of such domains is preserved. This not only indicates that the actual chromatin compaction state can be uncoupled from the histone modifications of a given chromatin domain, but also it suggests a modality to reestablish the original chromatin state, once DNA repair is accomplished.

γH2AX foci consist of spatially clustered γH2AX nano-foci. On exposure to IR, and as DDR progressed, γH2AX nano-foci were distributed throughout the nuclear volume, though they appeared to be spatially clustered (Fig. 6a and Supplementary Fig. 6A). To investigate such spatial clustering, we reconstructed the position of γH2AX nano-foci in the 3D nuclear space by collecting their 3D coordinates. Next, we measured the distances between the centroid of each nano-focus and all the other nano-foci in the nucleus. If the centroids of two objects were closer than a given cutoff distance, we assigned the corresponding nano-foci to the same cluster (Fig. 6b and Supplementary Fig. 1H and 'Methods' section). Based on the median lateral nano-focus radius of ~100 nm, we reasoned that two adjacent nano-foci would be spatially positioned so that their centroids would be at least 200 nm (2 × radius) away. Indeed, cutoff distances smaller than 300 nm resulted in poor clustering (Supplementary Fig. 6B). Similarly, distances bigger than 700 nm reduced the number of clusters at all time points, cancelling out differences over the time and, hence, impeding the analysis of the repair kinetics (Supplementary Fig. 6B). A cutoff distance of 500 nm (Fig. 6b) resulted in the highest number of clusters and a clear repair kinetics (Fig. 6c and Supplementary Table 5). Overall, the number of clusters was significantly higher than that of foci resolved by confocal microscopy, and comparable to the number of foci observed in pseudo-wide-field images (Fig. 1b,d). After IR, clusters were composed of a median number of four nano-foci (Fig. 6d), with the distributions remaining remarkably similar for all time points. This indicates that at times when the DSBs are repaired, the complete clusters, rather than single nano-foci, are removed en bloc. Coherently, clusters had an integrated median volume of about 0.05 μm^3 (Supplementary Fig. 6C), which decreased at later times. The average inter-centroid distance measured between all nano-foci belonging to a given cluster, the shortest path connecting all the centroids in a given cluster, and the inter-focal volume delimited by the 3D coordinates of the centroids of each nano-focus belonging to a cluster showed similar kinetics (Fig. 6b and Supplementary Fig. 6D–F). In all cases, these parameters increased after IR and then decreased, indicating that the nano-foci in each cluster were progressively closer to one another as the DDR progressed. One possible explanation is an active chromatin structure change bringing the clustered nano-foci in close proximity and, thus, facilitating the repair process of complex lesions at later times. However, the possibility that the clusters repaired at later times might correspond to a subset of damaged chromatin fibres whose location was in close spatial proximity already at earlier times is equally possible.

Finally, based on the previous nano-foci DNA content estimates, we calculated the DNA content of clusters by summing the DNA content of all γH2AX nano-foci belonging to a given cluster (Supplementary Tables 1 and 2). After IR, we observed broad-size distributions, with IQDs of about 197–938, 137–622 and 112–554 kb for 0.5 h, 3 h and 24 h time points, respectively (Fig. 6e). Overall, the cluster DNA content is in the (sub-)Mbp range, being directly relevant to genome regulation processes, as reported by genomic[14,15,17] or super-resolution microscopy[28] methods.

In view of these findings, and taking into account that the cutoff distance we applied for the cluster analysis is comparable in size to the γH2AX objects segmented in the pseudo-wide-field images (Supplementary Fig. 3A), we conclude that γH2AX foci, as previously identified by conventional microscopy techniques, correspond to spatially organized clusters, composed of several distinct nano-foci of phosphorylated H2AX in close spatial proximity whose pattern in the nucleus depends on the progression of DDR. While clusters are chromatin higher-order organization units in the half-a-megabase-pair size range, nano-foci are lower-order chromatin organization units whose size spans 40–160 kb.

γH2AX clusters contain single DNA DSBs. As previously reported, in higher eukaryotes[6–8], γH2AX is proposed to spread up to Mbps from the lesion site in a non-homogenous non-symmetrical fashion[11,12]. This implies that γH2AX may also be found reasonably far from the actual DNA break. Indeed, on severe localized DNA damage (for example, caused by accelerated charged particles), pan-nuclear H2AX phosphorylation is promptly induced by ATM and DNA-PK[29]. It is then obvious that not all γH2AX-decorated chromatin contains a DNA DSB in the immediate vicinity.

Based on the linear increase of γH2AX nano-foci numbers, we observed up to 10 Gy (Supplementary Fig. 6G), and on the assumption that 1 Gy X-ray induce 30–55 DSBs per diploid human genome[30–33], we estimated that 10 Gy X-ray would result in 470–860 DSBs in the ploidy-adjusted genome. Such numbers are conspicuously close to the number of γH2AX clusters we observed on IR (95% confidence interval of median cluster number at 0.5 h: 767–1,133; Fig. 6c and Supplementary Fig. 6H).

To directly estimate the number of DNA DSBs before and during the DDR, we recorded 3D-SIM super-resolution images of immunofluorescently labelled phospho-Ku70 proteins, which are directly associated to the broken ends, together with γH2AX. As shown in Fig. 7a,b, most of the phospho-Ku70 signal was surrounded by several γH2AX nano-foci. Remarkably, the number of phospho-Ku70 focal structures matched with good agreement that of our previously measured clusters (Fig. 7c). Also, the slopes of the linear regression lines computed while fitting the number of phospho-Ku70 and γH2AX nano-foci or clusters indicate that we measured ~3.4 γH2AX nano-foci per phospho-Ku70 focal structure, or in other words, that there are ~1.3 phospho-Ku70 focal structures per γH2AX cluster (Fig. 7d,e). We observed similar results by assessing the number of DNA DSBs by terminal deoxynucleotidyl transferase dUTP nick end-labelling (TUNEL). TUNEL signal was often surrounded by several γH2AX nano-foci (Fig. 7f,g) and the number of TUNEL focal structures recapitulates the DDR (Fig. 7h). Finally, we observed a robust agreement between the numbers of TUNEL focal structures and phospho-Ku70 (Fig. 7i) or γH2AX clusters (Fig. 7j). Together, these data demonstrate that γH2AX clusters are γH2AX-decorated multi-unit chromatin structures containing a single DNA DSB.

CTCF delimits phosphorylated H2AX chromatin domains. Altogether, the structural features we described about γH2AX clusters underpin the role of a structural organization factor in regulating their formation and kinetics. CTCF is involved in diverse cellular processes, including V(D)J recombination[34], regulation of transcription[35,36] and replication[17]. It mainly acts as a regulator of chromatin architecture[37,38] by forming and keeping chromatin loops, and the presence of CTCF-binding motif close to the boundaries of large looping chromatin domains

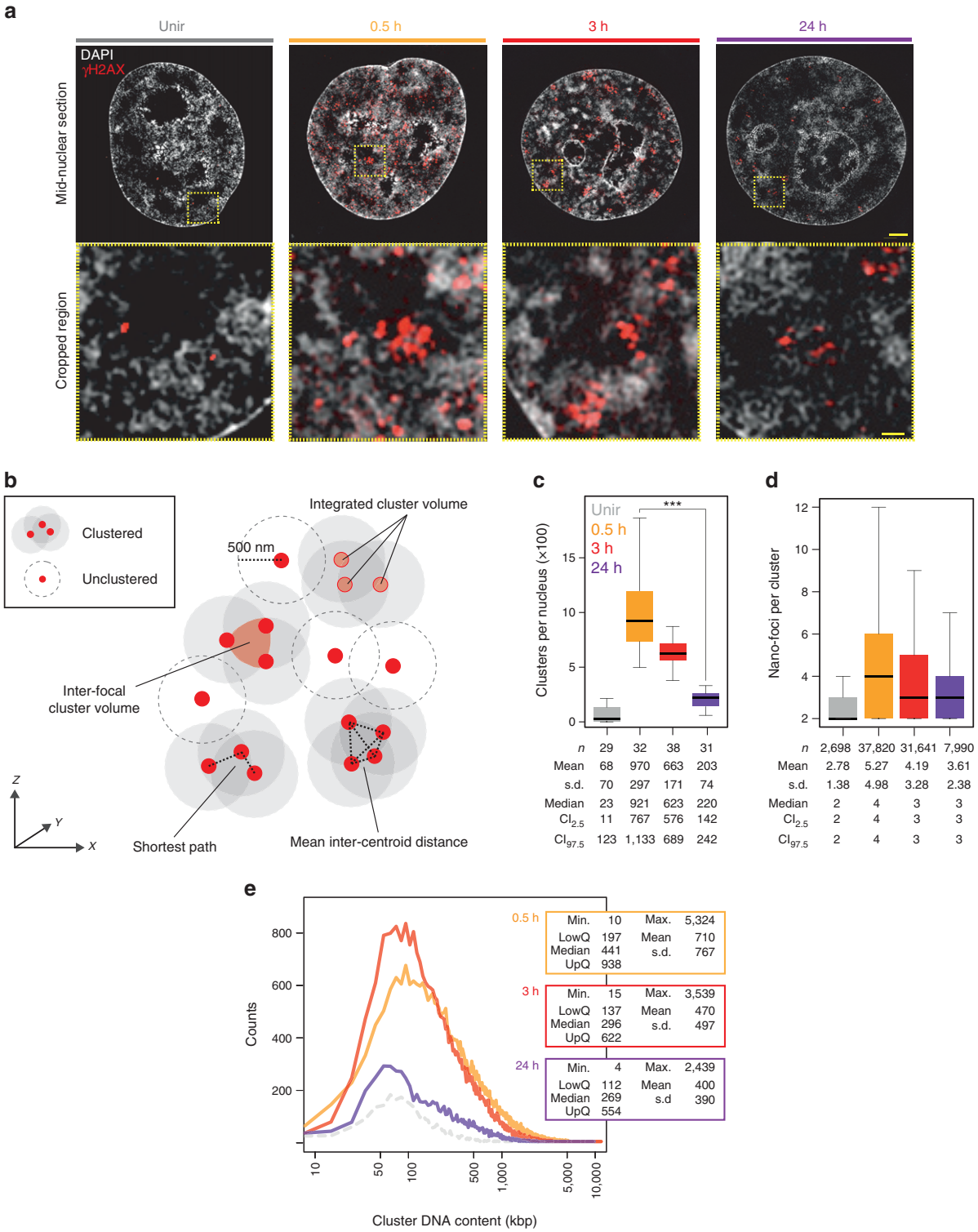


Figure 6 | Analysis of γH2AX nano-foci spatial clustering. (**a**) Exemplary 3D-SIM images of γH2AX immunofluorescence before and during DDR. Shown are the mid-nuclear section with DAPI and γH2AX signals, and magnified view from the yellow frame. Scale bars, 2 μm and 400 nm for main micrographs and magnified regions, respectively. (**b**) Schematics of γH2AX 3D-clusters analysis. All centroids (red dots) within a sphere defined by a given cutoff radius (500 nm in further analysis) are included in a cluster. For all nano-foci belonging to each given cluster, the sum of the volume of single nano-foci (integrated cluster volume), the volume delimited by the centroids (inter-focal volume), the shortest path connecting all centroids as well as the mean distance between centroids (mean inter-centroid distance) are computed (Supplementary Fig. 6C–F). (**c**) γH2AX 3D-clusters per nucleus. One-way ANOVA with Dunnett's correction; $^{***}P<10^{-3}$. (**d**) γH2AX nano-foci per 3D-clusters. Kruskal–Wallis $\chi^2=1{,}926.3$, df = 3, $P<2.2\times10^{-16}$. (**e**) DNA content distributions of γH2AX 3D-clusters during DDR. The DNA content of each nano-focus belonging to a given cluster is summed. The dashed line depicts the distribution of γH2AX 3D-clusters before IR. Kruskal–Wallis $\chi^2=5{,}964.1$, df = 3, $P<2.2\times10^{-16}$. Statistics are presented as in Fig. 2. All boxes and whiskers are as in Fig. 1. *n*: number of analysed cells (**c**) or 3D-clusters (**d**).

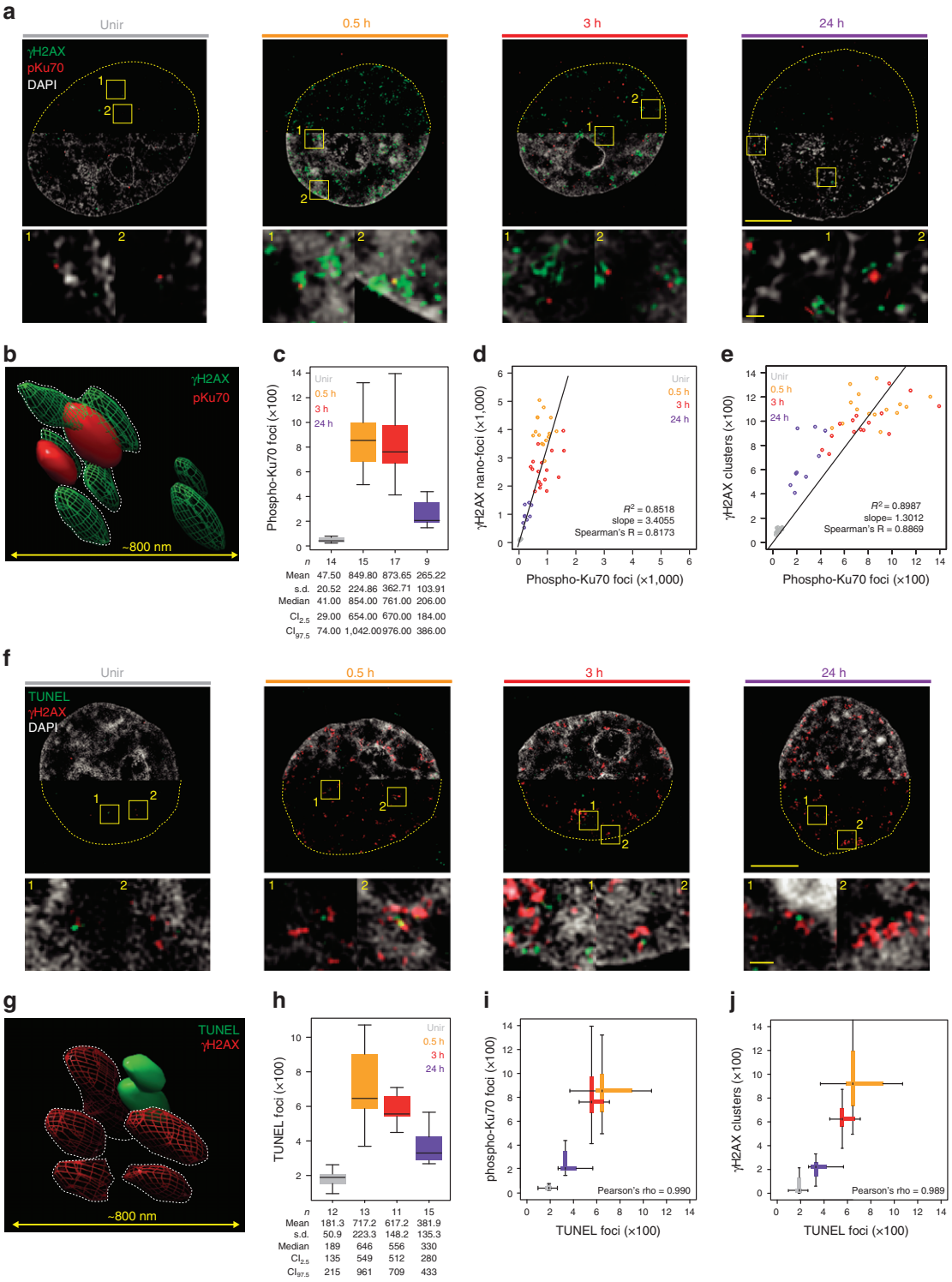
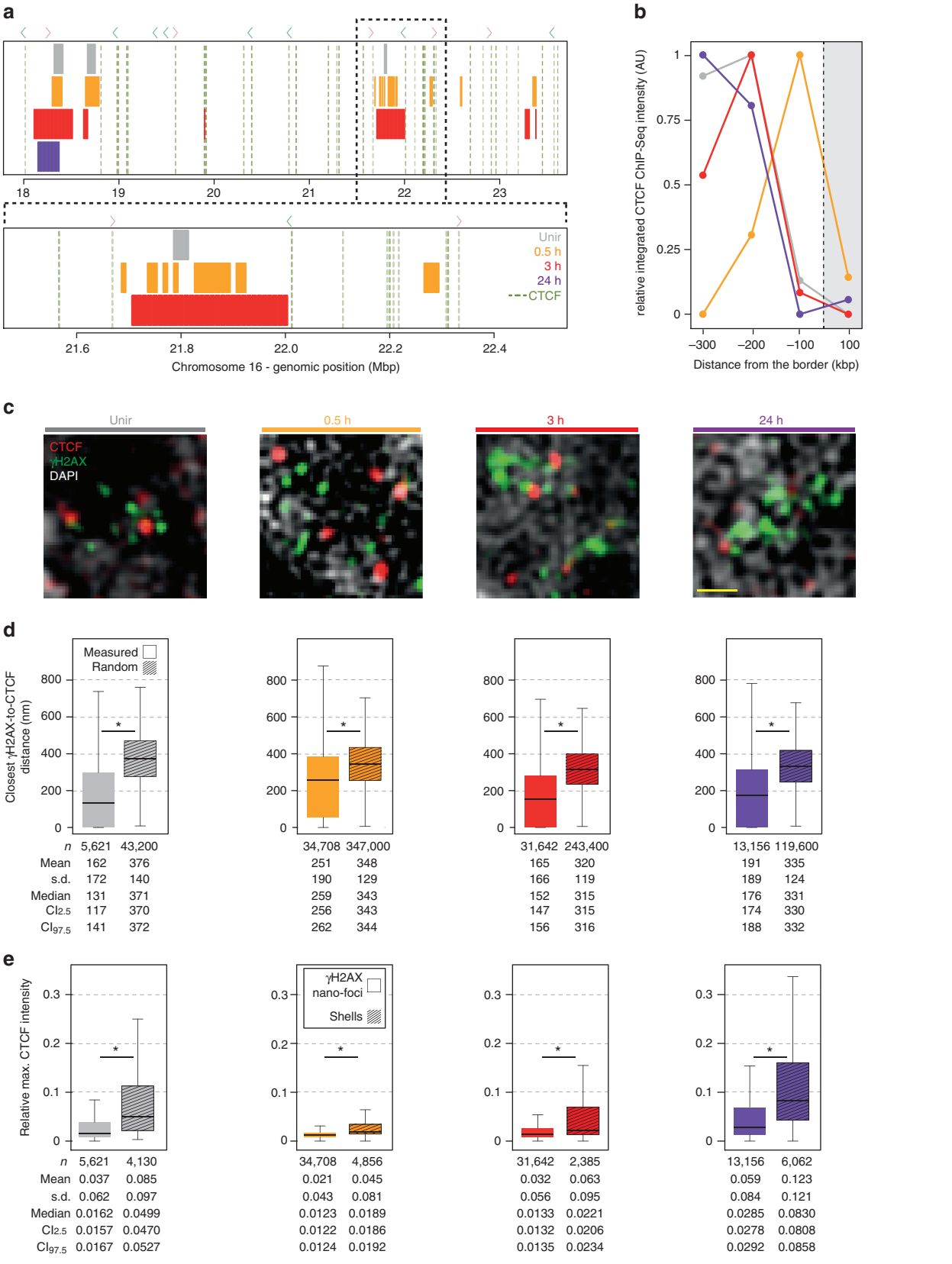


Figure 7 | Single phospho-Ku70- or TUNEL-labelled DNA DSBs are embedded in γH2AX clusters. (**a**) Exemplary 3D-SIM images of γH2AX and phospho-ku70 immunofluorescence before and during DDR. Shown are the mid-nuclear section (top) and enlarged views from the yellow frames (bottom). (**b**) 3D rendering of γH2AX and phospho-Ku70 immunostaining, 24 h after IR. (**c**) Phospho-Ku70 foci number distributions before and during DDR, from 3D-SIM images (one-way ANOVA with Dunnett's correction: $P<10^{-3}$). Scatter plots of phospho-Ku70 foci and γH2AX nano-foci (**d**) or γH2AX clusters (**e**). Each dot represents a single-cell nucleus. (**f**) Exemplary 3D-SIM images of γH2AX and TUNEL immunofluorescence before and during DDR. Shown are the mid-nuclear section (top) and enlarged views from the yellow frames (bottom). (**g**) 3D rendering of γH2AX and TUNEL immunostaining, 24 h after IR. (**h**) TUNEL foci number distributions before and during DDR, from 3D-SIM images (one-way ANOVA with Dunnett's correction: $P<10^{-3}$). Comparison between TUNEL and phospho-Ku70 (**i**) or γH2AX clusters (**j**) distributions, before and during DDR ($P<10^{-3}$). Scale bars, 5 μm and 500 nm for main micrographs and magnified regions, respectively. All boxes and whiskers are as in Fig. 1. *n*: number of analysed cells. Results are from two independent experiments.

has been recently confirmed by *in situ* Hi-C[15]. In view of these observations, and based on CTCF insulating properties, we next investigated the relationship between CTCF and γH2AX levels during DDR.

We identified the genomic location of putative CTCF-binding sites, based on a consensus motif modified from previous studies[15,39] (Supplementary Fig. 7A). The analysis resulted in 3,909 CTCF-binding sites, separated by a median intervening



distance of ~370 kb (IQD: 127–914 kb; Supplementary Fig. 7B). The orientation of CTCF motif had little to no impact on the measured distances (Supplementary Fig. 7B). This size range was comparable to that of γH2AX clusters rather than with that of single γH2AX nano-foci (Supplementary Fig. 7B), suggesting that individual clusters can be delimited by CTCF-binding sites. To validate such hypothesis at genomic level, we integrated our 3D-SIM-filtered γH2AX ChIP-Seq profiles (Supplementary Fig. 4) with publicly available HepG2 CTCF ChIP-Seq data. We identified ~140,000 CTCF genomic footprints, including CTCF occupancy levels ranging from very low to very highy. Due to the inherent nature of this ChIP-Seq data, it is unlikely that all those CTCF peaks would actually be present at the same time in a given cell. Therefore, we focused our analysis only on those CTCF genomic footprints whose occupancy score was maximum, assuming these sites would be conserved among different cell types. This reduced the number of CTCF footprints to 5,322. Remarkably, these sites were flanking most of the genomic γH2AX domains, before and during the DDR (Fig. 8a), yet the two signals seldom overlapped. In addition, CTCF ChIP-Seq signal intensity (that is, CTCF abundance) was higher upstream or downstream of the borders of each γH2AX genomic domain than that computed inside the domain (Fig. 8b), indicating that high-occupancy CTCF sites function as barriers for γH2AX spreading.

Next, we investigated the 3D-distribution of γH2AX and CTCF before and during DDR at single-cell level by 3D-SIM. On IR, CTCF foci were often in the immediate proximity of γH2AX nano-foci (Fig. 8c and Supplementary Fig. 7C,D). The majority (~75%) of the centroid-to-centroid distances between each γH2AX nano-focus and the closest CTCF focal structure were within 400 nm, and starting from 3 h post infrared, they all were below 200 nm (Fig. 8d). In all cases, the measured distances were smaller than distances between simulated random objects whose populations were comparable in numbers to those of CTCF and γH2AX nano-foci at each stage of DDR (Fig. 8d and Supplementary Fig. 7E). Because γH2AX nano-foci in our 3D-SIM images have a radius of ~100 nm, and CTCF focal structures showed comparable size, our results imply that the two objects would thus be in tight contact, with CTCF focal structures flanking γH2AX nano-foci. On exposure to IR, and based on the higher CTCF density in GC-rich regions, the expected γH2AX-to-CTCF distance should be equal to, if not shorter than, that we observed in the control sample (Fig. 8d, Unir, median: 131 nm). However, 0.5 h post IR, the median γH2AX-to-CTCF distance was two times longer (259 nm). Moreover, during the late stage of the DDR, the majority of DSBs were associated to heterochromatic regions (with lower GC content). In these regions, CTCF density is lower (compared with euchromatin) and the expected γH2AX-to-CTCF distance should be equal to, if not longer than, that we measured in a random distribution. Yet, the observed median γH2AX-to-CTCF distance was only half of that we obtained from a random distribution (Fig. 8d, 24 h measured: 176 nm; 24 h random: 331 nm). Such close spatial proximity was confirmed by the observation that CTCF signal was more abundant in the surroundings of γH2AX nano-foci (as measured in the previously described shells) rather than overlapping with them (Fig. 8e and Supplementary Fig. 7F).

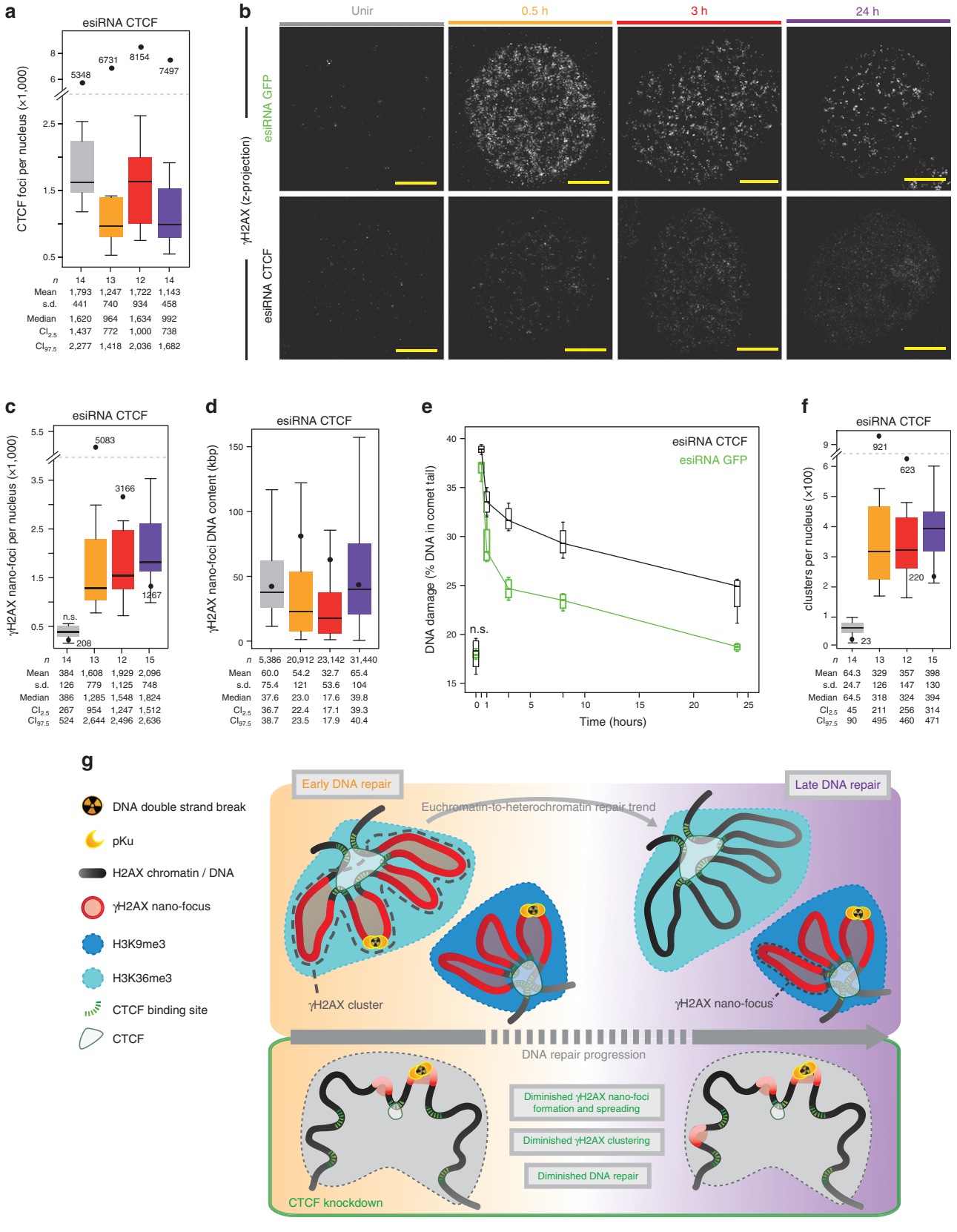
Taken together, our genomic and microscopy data strongly support that CTCF delimits γH2AX chromatin, and the two are in close spatial proximity.

CTCF is critical for spatial regulation of γH2AX chromatin. Finally, we investigated whether the perturbation of CTCF levels would affect the spatial distribution of γH2AX-decorated chromatin. While CTCF knockout is lethal, a number of studies have shown neither effects on the cellular and nuclear morphology, nor in the cell cycle progression up to 72 h post CTCF knockdown[40,41]. In our experimental system, esiRNA-mediated CTCF depletion to ~40% of the control protein levels (Supplementary Fig. 8A,B), resulted in a mild radiosensitization (~20%; Supplementary Fig. 8C) and a coherent decrease (70–85%) of CTCF foci in 3D-SIM micrographs, before and during DDR (Fig. 9a). Notably, CTCF depletion strongly impaired the formation of γH2AX nano-foci (Fig. 9b), which were smaller, diminished in numbers, and presented decreased volume and DNA content (Fig. 9c,d and Supplementary Fig. 8D,E). Only at 24 h post IR, the number of γH2AX nano-foci was comparable to that of the mock-knockdown samples, although with decreased fluorescence intensity, indicating a defect in the activation of the DDR. Indeed, CTCF-depleted cells showed a diminished DNA repair capability as assayed by comet single-cell analysis (Fig. 9e). Such defect was more prominent at the mid and late stages of DDR, suggesting that optimal CTCF levels are required to mount an efficient DDR. In this context, CTCF role in chromatin structural regulation may be crucial. Overall, the diminished γH2AX response resulted in a ~2.9-fold decrease in cluster formation (Fig. 9f). Remarkably, ATM and DNA-PKcs, the main signalling effectors involved in H2AX phosphorylation, were promptly activated on IR in both mock- and CTCF-depleted cells (Supplementary Fig. 9A,B), indicating that the presence of functional key factors of the DDR is necessary but not sufficient to trigger a proper response to DNA damage. In conclusion, we propose that CTCF, by preserving the 3D organization of the chromatin, is critical for the activation of an efficient DDR and, in such context, it functions as a regulator of the structural component of DDR.

Figure 8 | Genomic and microscopic analysis of CTCF spatial distribution in γH2AX-decorated chromatin. (**a**) Genomic localization of γH2AX ChIP-Seq domains (coloured bars) and CTCF genomic footprint (dashed green lines) in a representative region of chromosome 16. Dashed black line: magnification. Coloured arrowheads: orientation of CTCF-binding sites (red: forward; green: reverse). Details about γH2AX ChIP-Seq domains are in Supplementary Methods and Supplementary Fig. 4. ChIP-Seq CTCF profiles were retrieved from publicly available databases (UCSC Accession: Encode wgEH000080, wgEH000543, wgEH000401 and wgEH000470). (**b**) CTCF occupancy outside or inside γH2AX ChIP-Seq domains. The intensity of each CTCF peak in 100 kb bins upstream and downstream of the border of γH2AX ChIP-Seq domains (grey box) is summed and then presented as one-sided distribution. The bins range from ±300 to ±200, ±200 to ±100, ±100 to 0 and 0 to ±100 kb (inside the domain), with 0 being the border of each domain. AU: arbitrary unit. Genome-wide CTCF footprint localization relative to γH2AX ChIP-Seq domains' borders. For each domain, the distance in kb between its boundaries and the closest CTCF peak is measured and plotted as a bar (dashed lines). (**c**) Representative 3D-SIM images of immuno-stained γH2AX and CTCF before and during DDR. Scale bar, 500 nm. (**d**) Quantification of the closest centroid-to-centroid distance between CTCF and γH2AX nano-foci from 3D-SIM images. Measured (filled boxes) and simulated (patterned boxes) distances are shown. The latter were obtained from simulated random distributions of CTCF and γH2AX nano-foci (100 iterations). (**e**) Quantification of maximum CTCF intensity in γH2AX nano-foci and in surrounding shells. Maximum CTCF fluorescence in the segmented space normalized over the maximum CTCF fluorescence of the entire nucleus is plotted. All boxes and whiskers are as in Fig. 1. *n*: measured distances (**d**) or analysed shells (**e**) from two independent experiments. **d**,**e**: Mann–Whitney test: $P<10^{-3}$.

Discussion

In this study, the use of high prospecting super-resolution light microscopy technologies enabled us to identify the elementary structural units read by the DNA repair machinery, analysed as γH2AX focal structures following the exposure to IR. The γH2AX nano-foci we identified are two- to threefold smaller—with lateral diameters of ~200 nm—and contain ~10% of the conventionally estimated Mbp DNA content[42]. Similar γH2AX substructures sizes were recently measured after heavy ion irradiation[43], despite the highly



ionizing power charged particles possess, thus further supporting our findings.

Importantly, γH2AX nano-foci form clusters of approximately four chromatin units, and each cluster, rather than each of its structural components, contains one DSB, assessed by direct DNA end-labelling or by the presence of phospho-Ku70. This is supported by the good agreement between the predicted number of DSBs induced by the dose of IR employed in this work and the numbers of γH2AX clusters in control cells. γH2AX clusters are spatially distributed in the nuclear space according to a pattern that is dependent on the progression of DDR. Such pattern recapitulates the previously described repair kinetics, underlining an euchromatin-to-heterochromatin repair trend, which is likely dictated by the chromatin compaction state: chromatin regions that were already in an open state (for example, marked by H3K36me3) would be repaired earlier, while compact chromatin requires further structural remodelling before the DNA repair machinery could eventually exert its activity (Fig. 9g), For the latter, actual DNA decondensation, assessed as decrease of DAPI intensity, occurred while maintaining the main local histone modification (for example, H3K9me3), thus uncoupling DNA compaction from histone modifications. While chromatin relaxation seems to be dispensable for the DNA repair to occur at pericentromeric heterochromatin[44], we propose that the uncoupling of chromatin modifications and the actual chromatin decondensation is crucial to reestablish the original chromatin structure once DNA repair is accomplished.

In our 3D-SIM images, γH2AX clusters presented a discontinuous phosphorylation pattern, with γH2AX and CTCF showing mutually exclusive signals, although the two were in close spatial proximity. However, not all γH2AX nano-foci presented proximal CTCF foci. The latter likely consist of more than one CTCF molecule, and their detection may be influenced by a variety of factors, such as the CTCF-binding site density, differences in the binding affinity[45] of such sites and CTCF protein levels. It is tempting to speculate that the discontinuously phosphorylated pattern we observed is due to the presence of multiple CTCF molecules bound to their cognate consensus sequences but not resolvable by our imaging techniques. To discriminate between each individual chromatin loop bound by a pair of CTCF molecules, would demand single molecule sensitivity *in situ* 3D methods. Nonetheless, it is equally possible that other chromatin structure regulators (for example, cohesion complex[12]), histone turnover (for example, during DNA repair[46]) as well as biological processes such as transcription[11] antagonizing γH2AX formation and/or spreading along the chromosome contribute to the discontinuously phosphorylated pattern.

Finally, we show that CTCF has a critical role in the formation and spatial clustering of γH2AX nano-foci. CTCF-depleted cells present less γH2AX nano-foci, which are smaller and contain less DNA than those we observed in mock-treated cells. As a consequence, the DDR is delayed and the repair capability is diminished, despite the efficient activation of the main signalling effectors involved in H2AX phosphorylation (for example, DNA-PKcs or ATM). This indicates that a structural organization impairment—caused by CTCF depletion—results in a poor DDR. On CTCF depletion, the frequency of interactions of CTCF molecules with one another is decreased, leading to a diminished loop formation and a more sparse (that is, non-clustered) distribution of γH2AX nano-foci (Fig. 9b–g). Overall, this scenario emphasizes the need for a (dynamically) regulated 3D organization of the chromatin, whereby the 3D spatial proximity of chromatin loops could boost the local processivity of the committed kinases and assure an efficient DDR. In such context, because the CTCF-knocked-down cells display similar numbers of γH2AX nano-foci to the number of nano-foci cluster in control cells, we propose that in the absence of CTCF, spreading of γH2AX is impaired and, thus, this mark is restricted to the vicinity of the DSBs, that is, within one nano-focus (Fig. 9g).

In conclusion, our study demonstrates that the decreased levels of a single structural factor (CTCF), accounting for the (dynamic) stability of chromatin, *per se* dramatically hinder γH2AX spreading. While it is likely that additional factors (for example, DNA and histone methylation readers) contribute to this process, namely at heterochromatic regions, we propose that CTCF functions as a regulator of the structural component of DDR, preserving a crucial (dynamic) 3D organization of the chromatin and, thus, enabling an efficient DDR.

Methods

Cell culture and irradiation. Cervical carcinoma HeLa cells (ATCC No. CCL-2) cells were used throughout the study. A single exposure to 10 Gy X-ray was applied (250 kV, 16 mA, 2.5 Gy min^{-1} – GE Isovolt Titan) to induce DNA damage and trigger DDR. On exposure to IR, cells were incubated in a humidified environment, with 5% CO_2 at 37 °C as indicated. Sham-irradiated control cells were included. C2C12 (ATCC No CRL-1772) cells were used for CRISPR-Cas9 experiments. HeLa and C2C12 cells were cultured in DMEM (4.5 g l^{-1} glucose, Biochrom AG) supplemented with 10% and 20% fetal calf serum (Biochrom AG), respectively. All media were supplemented with 2 mM L-glutamine (Sigma), 100 U per ml penicillin and 100 μg ml^{-1} streptomycin (Sigma). All cell lines were tested for mycoplasma and found free of contamination (MycoAlert, Lonza).

Growth curve and cell cycle distribution. Cells were seeded 24 h before exposure to IR. After IR, cells were incubated for indicated times, before trypsinization and count with a coulter counter, in triplicates. The remaining cells were then fixed in 2% formaldehyde, permeabilized for 8 min with 0.5% Triton X-100 in PBS, stained with DAPI (1 μg ml^{-1}) and analysed at the flow cytometer Partec PAS III system (Partec) for cell cycle distribution. Data were analysed with FlowJo software (Tree Star, Inc.).

Apoptosis assay. To detect apoptosis, TUNEL assay was performed according to the manufacturer's instructions (Roche, #11684795910) and a minimum of 1,000 cells was scored by microscopy in two independent experiments.

Spectral karyotyping. Cells were treated with colcemid (0.1 μg ml^{-1}; Invitrogen, Darmstadt, Germany) 2 h before collecting to accumulate metaphase cells.

Figure 9 | CTCF depletion inhibits γH2AX nano-foci and cluster formation and diminishes the DNA repair capability. (**a**) Number of CTCF foci in esiRNA-depleted cells before and during DDR. Black dots: median number of CTCF foci in wild-type cells. (**b**) Impairment of γH2AX nano-foci and 3D-clusters formation during DDR as assessed by immunofluorescence of 3D-SIM images in CTCF-depleted cells. Scale bar, 5 μm. (**c**) γH2AX nano-foci number distributions before and after IR, in CTCF siRNA-treated cells. Black dots: median number of γH2AX nano-foci of untreated cells (from Fig. 1).NS: two-tailed *t*-test, $P>0.05$. (**d**) γH2AX nano-foci DNA content distributions before and after IR, in CTCF siRNA-treated cells. Black dots: median DNA content of γH2AX nano-foci of untreated cells (from Fig. 2). (**e**) DNA fragmentation measured by the neutral comet assay. Boxes represent the mean of medians from four replicates (two biological replicates in duplicate), each consisting of 60 comet measurements. NS: not significant (*t*-test, $P>0.05$). (**f**) γH2AX cluster distributions before and after IR, in CTCF siRNA-treated cells. Black dots: median number of γH2AX clusters in untreated cells (from Fig. 6). All boxes and whiskers are as in Fig. 1. Comparisons between time points (one-way ANOVA with Dunnett's correction) or between esiRNA-treated and wild-type cells (Wilcoxon/Mann–Whitney rank sum) are all statistically significant unless otherwise specified. (**g**) Model for cluster special arrangement during DDR, showing the time-dependent euchromatin-to-heterochromatin repair trend (top) and how γH2AX spreading is hampered by CTCF depletion with the concomitant loss of 3D-arrangement of chromatin loops (bottom).

Chromosome preparations were made according to standard procedures and hybridized with the 24XCyte Multicolor FISH Probe Kit (MetaSystems, Altlussheim, Germany). Metaphase spreads were examined with an Axio Imager Z1 microscope (Zeiss, Oberkochen, Germany) equipped with appropriate filter sets. At least 100 images of metaphases were taken, further processed using ISIS software (MetaSystems) and analysed to produce the karyotype.

CTCF knockdown. A number of 10^5 cells were transfected with 15 nM of a esiRNA pool (Sigma-Aldrich) using HiPerfect (Qiagen). The CTCF esiRNA is corresponding to the region 692–1195 of the human CTCF transcript (NM_006565.3). For mock treatments, cells were transfected using an esiRNA pool (Sigma-Aldrich) targeting the *GFP* gene. Cells were incubated 24–96 h post transfection and knockdown efficiency was monitored every 24 h.

Immunoblotting. Whole-cell extracts were prepared by freeze and thaw lysis (three cycles) in 600 mM NaCl, 20 mM Tris-HCl pH 7.8, 20% glycerol. After SDS–PAGE, proteins were transferred onto PVDF membrane in semi-dry conditions. The membrane was then blocked in 5% non-fat dry milk buffer and incubated with mouse anti-γH2AX (Clone JBW301, Upstate, 1:5,000). Immunoblots were stained with corresponding HRP-conjugated secondary antibodies (GE Healthcare, 1:20,000) and detected with the enhanced chemiluminescence detection system (Amersham Biosciences). Quantification was performed using ImageJ.

For the validation of antibody specificity and cross-reactivity, a dilution series of synthetic peptides (CKATQASQEY; Peptide Specialty Laboratories GmbH), with the underlined serine in either phosphorylated or non-phosphorylated form, was immobilized on a nitrocellulose membrane at the indicated concentrations and probed with anti-γH2AX and anti-H2AX as described above.

CTCF knockdown western blots were developed using a rabbit anti-CTCF (#D31H2, Cell Signaling, 1:700) and a mouse anti-actin (AC-40, Sigma-Aldrich, 1:1,000) and overnight incubation at 4 °C, followed by a direct immunofluorescence detection using anti-rabbit-IgG-Cy5 (#711-175-152, Jackson, 1:1,000) and an anti-mouse-IgG-Alexa488 (A11029, Invitrogen, 1:1,000). Images were recorded using a AI600 Imager (Amersham) and quantified using ImageJ.

Immunofluorescence. Cells were fixed in 3.7% formaldehyde and permeabilized in 0.5% Triton X-100 in PBS at room temperature (RT). The following primary antibodies were used: mouse anti-γH2AX (Clone JBW301, 1:500, Upstate), rabbit anti-H3K9me3 (#07–422, Upstate, 1:500), rabbit anti-H3K9me3 (#39161, Active Motif, 1:500), rabbit anti-H3K36me3 (ab9050, Abcam, 1:2,000); rabbit anti-phospho-Ku70 (pS5) (#ab61783, Abcam, 1:400); mouse anti-phospho-ATM (pS1981) (#MAB3806, Millipore, 1:100); rabbit anti-phospho-DNA-PKcs (pS2056) (#ab18192, Abcam, 1:100) and rabbit anti-CTCF (#2899, Cell Signaling, 1:900). For phospho-Ku70 detection cells were prefixed in 1% formaldehyde and then extracted with 0.7% Triton X-100 two times by 5 min[47] and subsequently fixed in 3.7% formaldehyde. Antibody incubation was performed at 4 °C over night in 1% BSA in PBS. For CLSM and 3D-SIM, signals were detected with goat anti-mouse-IgG-AlexaFluor 488, goat anti-rabbit-IgG-AlexaFluor 594 (1:800, Invitrogen), donkey anti-mouse-IgG-AlexaFluor 488 (A-21202, Thermo Fisher Scientific, 1:400), donkey anti-rabbit-IgG-AlexaFluor 594 (A-21207, Thermo Fisher Scientific, 1:400). For STED, γH2AX was detected with goat anti-mouse-IgG STAR 635P (#2-0002-007-5, Abberior, 1:100) or goat anti-mouse-IgG STAR 580 (#2-0002-005-1, Abberior, 1:100). DNA was counterstained with 36 nM DAPI (for 3D-SIM), 1 μM propidium iodide (confocal microscopy) or 2.5 μM SiR-DNA (Spirochrome), before cells were mounted with Vectashield antifade medium (Vectorlabs).

CRISPR-Cas9 targeting to heterochromatic major satellite DNA. Subconfluent C2C12 cells were transfected with Cas9 (pCMV-hCas9, Addgene ID: 41815) and major satellite gRNAs (U6-MaSgRNA) by means of Lipofectamine 3000 (Thermo Fisher Scientific) according to the manufacturer's instructions. Cells were then fixed in 3.7% formaldehyde for 10 min and immunofluorescence followed (as described above).

DNA DSB detection by TUNEL assay. Cells were grown and irradiated as described above. At the indicated time points, cells were fixed in 3.7% paraformaldehyde for 10 min. The fixation was quenched with 125 mM glycine in PBS for 10 min. Fixed cells where permeabilized in 0.5% Triton X-100 for 20 min, and equilibrated for 10 min in blunting buffer (100 mM Tris-HCl, 50 mM NaCl, 10 mM $MgCl_2$, 0.025% Triton X-100 and 5 mM DTT, pH 7.5). End repair was performed using 4 μl T4 polymerase (NEB: M0203S 3,000 units ml^{-1}) and 4 μl T4 polynucleotide kinase (NEB: M0201S 10,000 units ml^{-1}) in 82 μl blunting buffer, supplemented with 10 μl 1 mM dNTPs for 45 min. Slides were then equilibrated in TdT buffer for 10 min and the TUNEL reaction was performed according to the 'In Situ Cell Death Detection Kit' (Roche) with Fluorescein modified dUTPs, for 4 h at 37 °C according to the manufacturer's instructions. Following the TUNEL reaction, cells were blocked in 1% BSA in PBS for 20 min. γH2AX staining was performed as described above. Incorporated fluorescein-dUTPs were detected by a rabbit anti-FITC (CUSABIO, 1:500) and a anti-rabbit-IgG Alexa488 secondary antibody (Jackson ImmunoResearch, 1:800). All steps were conducted at RT, unless otherwise specified.

Comet assay. DNA repair kinetics in CTCF knockdown cells were measured using the neutral comet assay. In brief, CTCF was depleted as described above and 72 h post esiRNA transfection, the cells were exposed to 10 Gy X-ray. At the indicated time points, cells were trypsinized and 2×10^5 cells ml^{-1} were embedded in 0.8% low-melting point agarose (Sigma type VII). Lysis was performed for 4 h at 4 °C in lysis buffer (10 mM Tris, 150 mM NaCl, 1% N-lauryl-sarcosinate, 1% Triton X-100, 0.5% DMSO, pH 8.0) and electrophoresis was done in 1 × TBE at 4 °C (1 V cm^{-1}) for 25 min. Slides were then dehydrated in 70% ethanol and rehydrated in staining buffer (TBE supplemented with SybrGreen, 1:10,000) to stain the DNA[48]. Two biological replicates (in duplicates) were performed and 60 comets per slide were scored using Komet 4 (Kinetic Imaging Ltd.).

Microscopy. Confocal microscopy images were acquired using a Spinning Disk microscope (Perkin Elmer Vox1000) equipped with a ×60 NA 1.4 oil immersion lens (CFI Apochromat TIRF), with a pixel size of 120 nm or with a Leica TCS SP5 confocal microscope using a Plan Apo ×63 NA 1.4 oil immersion objective. Cells were recorded as z-stacks with a z-spacing of 0.2 μm.

Super-resolution microscopy images were acquired using a 3D structured illumination microscope (DeltaVision OMX V3, GE Healthcare) and a 2C STED 775 QUAD Scan microscope (Abberior Instruments). 3D-SIM was performed with a ×100 NA 1.4 objective lens with a pixel size of 39 nm and a z-spacing of 125 nm (ref. 18). STED was performed with a ×100 NA 1.4 Olympus UPlanSApo objective lens with a pixel size of 20 nm and excitation lasers of 488, 594 or 640 nm, and a 775 nm depletion laser.

High-content imaging was performed using the Operetta system (Perkin Elmer). Samples were imaged using a ×20 NA 0.45 air objective with three planes of 1 μm spacing, using the following filters: DAPI: excitation wavelength (ex): 360–400 nm, emission wavelength (em): 420–480 nm; Alexa488: ex: 460–490 nm, em: 500–550 nm; Alexa594: ex: 560–580 nm, em: 590–640 nm.

Image analysis. For confocal microscopy, the images were analysed in ImageJ using the nuclear staining as a mask to measure the total intensity of the γH2AX signal per nucleus. Foci were scored in 3D using Volocity (Perkin Elmer) by the following workflow: find objects (nucleus), threshold automatic, size minimum 500 μm^3; find foci: threshold 4,000 constant for pseudo-wide-field and 5,000 for deconvolved images, respectively. Minimum size: 0.05 μm^3, followed by 'separate touching objects' with a guide size of 0.5 μm^3. Different thresholds were applied, because pseudo-wide-field and deconvolved images are in different bit depth. All counts were double-checked by manual counting of randomly chosen samples by at least three experimenters.

For CRISPR-Cas9 experiments, confocal images of C2C12 cells were segmented into background, nuclei and chromocentres by pixel-wise classification via supervised machine learning (default Random Forest classifier and pixel features from the Trainable Weka Segmentation plugin in Fiji). The classifier was trained on manually labelled pixels of the DAPI channel in one image and then applied to all images. For each image, mean intensities in the H3K9me3 and γH2AX channels were determined for each chromocentre object (>100 px^2) within the largest object in the nuclear mask. To analyse DNA decondensation at repair sites in CRISPR-Cas9 experiments STED images of C2C12 cells were segmented into background, nuclei and chromocentres by pixel-wise classification as described above for confocal images. The classifier was trained on manually labelled pixels of the SiR-DNA channel in one image and then applied to all images (each image's pixel intensity range was mapped to the 8-bit range to account for differences in staining intensities). For each image, the circularity of chromocentre objects (>100 px^2) within the nucleus was determined. Three rounds of binary erosion with a 3 × 3 px-box followed by three rounds of binary dilation were applied to the segmentation results to smooth the borders of segmented objects.

3D-SIM images were exported from the DeltaVision software (softWoRx 6.0 Beta 19, Applied Precision) and converted to 16-bit images per channel. Foci counting was done using Volocity 6.3 (Perkin Elmer) or with the 3D foci picker plugin in ImageJ (imagej.nih.gov/ij/). Nearly identical results were obtained and the numbers from Volocity were used. In detail, the individual z-sections were imported and merged to a volume with the above-mentioned pixel sizes and z-spacing. First, the nucleus was identified by setting a manual threshold and a lower volume limit of 200 μm^3 followed by a 'Fill in Holes' step and two iterations of 'Dilate' and 'Close' to fill in all the DAPI weak volumes. The intensities and voxel coordinates of the whole nucleus were registered. Next, the γH2AX and H3K36me3 or H3K9me3 foci were identified with a lower threshold of 1,000 and a minimum object size of 0.001 μm^3. To separate close spaced objects, a final 'Separate Touching Objects' step with a nominal volume of 0.05 μm^3 was used. The foci identified were restricted to the previously defined nuclear volume to remove possible unspecific signals from outside of the nucleus.

3D-SIM pseudo-wide-field imaging: after sample acquisition, the pseudo-wide-field images were calculated using softWoRx 6.0 Beta 19 according to the following

workflow: the raw data from each 3D-SIM image z-stack was subdivided to isolate the first angle of acquisition. To this purpose, the maximum number of z-sections in each individual stack is divided by three. Then the projected five grid shifted section is averaged per z-position and colour channel. After that, the voxel dimensions are adjusted from 0.625 to 0.125 µm in the z-dimension by adjusting the file headers. The alignment of the new stack was done with the parameters used for 3D-SIM reconstruction. The following parameters of the softWoRx software were used: normalize intensity, use photosensor, correct bleaching, replace z-lines and smooth z-lines. To reverse the optical distortion in the images, the aligned 3D stack was deconvolved with the instrument-specific optical transfer function (OTF) with the following settings: 'enhanced ratio (aggressive)' and 'noise filtering medium'.

For CTCF distance analysis, the previously described protocol was extended as follows: CTCF domains detection was restricted to the nuclear volume, with an automated threshold and a minimum size of 0.001 μm^3. Then, the segmented γH2AX nano-foci were extended in all dimensions by three voxels ($117 \times 117 \times 375$ nm) and the resulting γH2AX nano-foci volume was subtracted to obtain the γH2AX foci shells. Finally, the Euclidian distances between each γH2AX nano-focus and the closest CTCF domain were measured. All identified foci with the corresponding 3D coordinates and intensities for all recorded channels were exported and post-processed in R[49]. ImageJ and UCSF chimera[50] were used for image visualization and 3D rendering, respectively. Simulations of CTCF and γH2AX distributions were run under R, using rgl and sphereplot packages. Hundred simulations of a sphere matching the average nuclear size of cells were run per time point. Every simulation contained objects whose numbers matched CTCF and γH2AX foci we recorded in 3D-SIM images.

For STED images, object dimensions (for example, diameters) were measured by manual object segmentation of randomly selected foci in ImageJ, using the analyse particle tool. For high-content images, analysis was performed using Harmony software (Perkin Elmer) with the following workflow: maximum projection of the planes, flatfield correction, find nuclei in DAPI channel, method M, splitting coefficient 0.1, general threshold 0.4 and guide size of 15 µm in diameter. Calculate intensity and morphology parameters for the nuclei. Discard nuclei touching the border, smaller than 100 μm^2 and larger than 350 μm^2. Filter nuclei for roundness >0.83 and with a 4 px Haralick contrast >0.8 and a DAPI signal CV of less than 30%. Measure the mean and integrated intensity for DAPI, γH2AX and CTCF in the selected nuclei areas.

ChIP. Cells were fixed with 1% formaldehyde for 10 min at RT and cross-link was quenched with 125 mM glycine (5 min at RT). Nuclei were isolated after mild lysis in hypotonic buffer (10 mM HEPES pH 8, 1.5 mM $MgCl_2$, 60 mM KCl) and 20 strokes in a tight dounce homogenizer. Chromatin was sheared in sonication buffer (0.5% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.1). Fragmentation of chromatin was carried out by ultrasound treatment (Bioruptor UCD200) so that fragments of 200–300 bp length were obtained. Chromatin from 1×10^6–2×10^6 cells was immunoprecipitated with anti-γH2AX (Clone JBW301, Upstate, 3 µg) antibody. Chromatin was then incubated ON at 4 °C with protein G-coated magnetic beads (ChIP-IT Express, Active Motif). The collected chromatin (ChIP sample) was then reverse-crosslinked in the presence of 200 mM NaCl at 65 °C for at least 5 h, followed by RNase A (50 µg ml^{-1}) treatment for 30 min at 37 °C and proteinase K (100 µg ml^{-1}) treatment for 3 h at 50 °C. DNA elution was carried out in 1% SDS, 100 mM $NaHCO_3$, in a rotary shaker at RT for 15 min. Pure DNA was isolated using the Qiagen PCR purification kit and 15–30 ng of size selected DNA fragments (Qubit fluorometric quantification) were used to produce ChIP-seq libraries (Illumina ChIP-Seq DNA sample Prep Kit). Input sample was essentially prepared following the same protocol, but the immunoprecipitation step was skipped.

Next-generation sequencing and data analyses. ChIP-Seq libraries were processed through a high-throughput sequencing pipeline (Illumina Genome Analyzer II). Reads were mapped to the human genome (University of California, Santa Cruz (UCSC) hg19 assembly, based on the National Center for Biotechnology Information (NCBI) build 37.1) by means of SOAP2 software[51], allowing up to two mismatches for each 36 bp read. All data sets were deposited in the Gene Expression Omnibus database (accession number: GSE60526). All γH2AX ChIP-Seq tracks were smoothed with a moving average of five intervals before further analysis. Genomic features and correlation analysis: all genomic features data were retrieved from publicly available databases (UCSC) (Supplementary Table 3). Most of the data were generated in HepG2 cells, but not all. Data that were originally generated in the hg18 assembly were transposed to hg19 using LiftOver (http://genome.ucsc.edu/cgi-bin/hgLiftOver). Reads per kilobase per million reads (RPKM)[52] were calculated for non-overlapping 10 kb genomic intervals for all sequence tracks. The features were further normalized to the corresponding genome-wide average and correlation with γH2AX tracks was performed (Spearman's ρ correlation coefficient with $P<2.2 \times 10^{-16}$ in all cases).

Statistical analysis. Overall, sample size was chosen so that groups (for example, time points) had comparable numbers (for example, number of imaged cells), whenever possible. High-content microscopy and next-generation sequencing provided large data sets ensuring statistical significance. All statistical analysis has been performed using R or GraphPad Prism. Briefly, in case data were normally distributed (Shapiro–Wilk test), ANOVA or Student's *t*-test were performed for groups or pairs, respectively. Else, Kruskal–Wallis or Wilcoxon/Mann–Whitney rank sum tests were used for groups or pairs, respectively.

Integration of 3D-SIM and ChIP-Seq data. To integrate the ChIP-Seq data with 3D-SIM information, we first generated 25 independent profiles by applying a smoothing factor to each γH2AX ChIP-Seq data set (Supplementary Fig. 4A). Such smoothing factor is a moving average ranging from 1 (no smoothing) to 25 genomic intervals (indicated as '1D', in Supplementary Fig. 4A). In parallel, we measured the volume fraction occupied by γH2AX nano-foci as well as their corresponding DNA content, before and during the DDR (Supplementary Fig. 4B). In response to ionizing radiation, we observed an increase of the mean γH2AX-occupied nuclear volume (from 0.21 ± 0.21% to 7.81 ± 3.19%), which recapitulated the DDR (the volume was reduced to 3.70 ± 1.39% and 0.66 ± 0.43%, at 3 h or 24 h post-ionizing radiation, respectively). Next, we applied the mean volume fractions (0.21%, 7.81%, 3.70% and 0.66% for unirradiated, 0.5 h, 3 h and 24 h, respectively) to filter the previously smoothed genomic γH2AX ChIP-Seq data so that only the 10 kb genomic intervals from the top percentiles of the read density distributions were retrieved (Supplementary Fig. 4C). For example, as for the unirradiated cells, we sampled the 99.79th percentile (top 100–0.21%) of the intervals, while for the 0.5 h time point, we sampled the 92.19th percentile (top 100–7.81%) of the total RPKM γH2AX ChIP-Seq distribution. A representative image of filtered ChIP-Seq profiles is shown in Supplementary Fig. 4D. By applying these imaging-based thresholds, we obtained a linear coverage of 4.7 Mbp, 159.0 Mbp, 92.3 Mbp and 21.8 Mbp, at unirradiated, 0.5 h, 3 h and 24 h time points, respectively (Supplementary Fig. 4E). Finally, we employed the numbers of 3D γH2AX nano-foci to match the numbers of 1D nano-domains as follows: first, the number of 3D γH2AX nano-foci before and after the DDR was scaled down to the haploid genome size to match the genomic data (ploidy correction factor: $\text{HeLa}_{\text{genome size}}/\text{haploid reference}_{\text{genome size}} = 3.12$); next, we chose the smoothing factor at which the number of γH2AX nano-foci and the number of retrieved genomic intervals matched best, at any given time point (Supplementary Fig. 4A, over-imposed crosses). All ChIP-Seq domains identified via such approach are referred to as '1D domains' and an estimate of the 1D domain size distribution is presented in Supplementary Fig. 4F.

Data availability. Next-generation sequencing results are available at GEO (https://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE60526. Other data that support the findings of this study are available from the corresponding author on reasonable request.

References

1. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273,** 5858–5868 (1998).
2. Stucki, M. & Jackson, S. P. GammaH2AX and MDC1: anchoring the DNA-damage-response machinery to broken chromosomes. *DNA Repair* **5,** 534–543 (2006).
3. Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434,** 864–870 (2005).
4. Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434,** 907–913 (2005).
5. Turinetto, V. & Giachino, C. Multiple facets of histone variant H2AX: a DNA double-strand-break marker with several biological functions. *Nucleic Acids Res.* **43,** 2489–2498 (2015).
6. Berkovich, E., Monnat, Jr R. J. & Kastan, M. B. Roles of ATM and NBS1 in chromatin structure modulation and DNA double-strand break repair. *Nat. Cell Biol.* **9,** 683–690 (2007).
7. Meier, A. *et al.* Spreading of mammalian DNA-damage response factors studied by ChIP-chip at damaged telomeres. *EMBO J.* **26,** 2707–2718 (2007).
8. Savic, V., Sanborn, K. B., Orange, J. S. & Bassing, C. H. Chipping away at gamma-H2AX foci. *Cell Cycle* **8,** 3285–3290 (2009).
9. Kim, J. A., Kruhlak, M., Dotiwala, F., Nussenzweig, A. & Haber, J. E. Heterochromatin is refractory to gamma-H2AX modification in yeast and mammals. *J. Cell Biol.* **178,** 209–218 (2007).
10. Shroff, R. *et al.* Distribution and dynamics of chromatin modification induced by a defined DNA double-strand break. *Curr. Biol.* **14,** 1703–1711 (2004).
11. Iacovoni, J. S. *et al.* High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. *EMBO J.* **29,** 1446–1457 (2010).
12. Caron, P. *et al.* Cohesin protects genes against gammaH2AX Induced by DNA double-strand breaks. *PLoS Genet.* **8,** e1002460 (2012).
13. Ong, C. T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15,** 234–246 (2014).
14. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).

15. Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159,** 1665–1680 (2014).
16. Gonzalez-Sandoval, A. *et al.* Perinuclear anchoring of H3K9-methylated chromatin stabilizes induced cell fate in *C. elegans* embryos. *Cell* **163,** 1333–1347 (2015).
17. Pope, B. D. *et al.* Topologically associating domains are stable units of replication-timing regulation. *Nature* **515,** 402–405 (2014).
18. Schermelleh, L., Heintzmann, R. & Leonhardt, H. A guide to super-resolution fluorescence microscopy. *J. Cell Biol.* **190,** 165–175 (2010).
19. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19,** 780–782 (1994).
20. Lobrich, M. *et al.* GammaH2AX foci analysis for monitoring DNA double-strand break repair: strengths, limitations and optimization. *Cell Cycle* **9,** 662–669 (2010).
21. Halazonetis, T. D., Gorgoulis, V. G. & Bartek, J. An oncogene-induced DNA damage model for cancer development. *Science* **319,** 1352–1355 (2008).
22. Tsantoulis, P. K. *et al.* Oncogene-induced replication stress preferentially targets common fragile sites in preneoplastic lesions. A genome-wide study. *Oncogene* **27,** 3256–3264 (2008).
23. Lee, C. S., Lee, K., Legube, G. & Haber, J. E. Dynamics of yeast histone H2A and H2B phosphorylation in response to a double-strand break. *Nat. Struct. Mol. Biol.* **21,** 103–109 (2014).
24. Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* **41,** 376–381 (2009).
25. Peters, A. H. *et al.* Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat. Genet.* **30,** 77–80 (2002).
26. Chiolo, I. *et al.* Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* **144,** 732–744 (2011).
27. Jakob, B. *et al.* DNA double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone H2AX phosphorylation and relocation to euchromatin. *Nucleic Acids Res.* **39,** 6489–6499 (2011).
28. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529,** 418–422 (2016).
29. Meyer, B. *et al.* Clustered DNA damage induces pan-nuclear H2AX phosphorylation mediated by ATM and DNA-PK. *Nucleic Acids Res.* **41,** 6109–6118 (2013).
30. Rothkamm, K., Kruger, I., Thompson, L. H. & Lobrich, M. Pathways of DNA double-strand break repair during the mammalian cell cycle. *Mol. Cell Biol.* **23,** 5706–5715 (2003).
31. Polo, S. E. & Jackson, S. P. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes Dev.* **25,** 409–433 (2011).
32. Newman, H. C., Prise, K. M., Folkard, M. & Michael, B. D. DNA double-strand break distributions in X-ray and alpha-particle irradiated V79 cells: evidence for non-random breakage. *Int. J. Radiat. Biol.* **71,** 347–363 (1997).
33. Lobrich, M., Rydberg, B. & Cooper, P. K. Repair of x-ray-induced DNA double-strand breaks in specific Not I restriction fragments in human fibroblasts: joining of correct and incorrect ends. *Proc. Natl Acad. Sci. USA* **92,** 12050–12054 (1995).
34. Chaumeil, J. & Skok, J. A. The role of CTCF in regulating V(D)J recombination. *Curr. Opin. Immunol.* **24,** 153–159 (2012).
35. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19,** 24–32 (2009).
36. Hou, C., Zhao, H., Tanimoto, K. & Dean, A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc. Natl Acad. Sci. USA* **105,** 20398–20403 (2008).
37. Fu, Y., Sinha, M., Peterson, C. L. & Weng, Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4,** e1000138 (2008).
38. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137,** 1194–1211 (2009).
39. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128,** 1231–1245 (2007).
40. Schmidt, D. *et al.* A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* **20,** 578–588 (2010).
41. Tark-Dame, M., Jerabek, H., Manders, E. M., Heermann, D. W. & van Driel, R. Depletion of the chromatin looping proteins CTCF and cohesin causes chromatin compaction: insight into chromatin folding by polymer modelling. *PLoS Comput. Biol.* **10,** e1003877 (2014).
42. Solovei, I., Thanisch, K. & Feodorova, Y. How to rule the nucleus: divide et impera. *Curr. Opin. Cell Biol.* **40,** 47–59 (2016).
43. Lopez Perez, R. *et al.* Superresolution light microscopy shows nanostructure of carbon ion radiation-induced DNA double-strand break repair foci. *FASEB J.* **30,** 2767–2776 (2016).
44. Tsouroula, K. *et al.* Temporal and spatial uncoupling of DNA double strand break repair pathways within mammalian heterochromatin. *Mol. Cell* **63,** 293–305 (2016).
45. Plasschaert, R. N. *et al.* CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation. *Nucleic Acids Res.* **42,** 774–789 (2014).
46. Ikura, T. *et al.* DNA damage-dependent acetylation and ubiquitination of H2AX enhances chromatin dynamics. *Mol. Cell Biol.* **27,** 7028–7040 (2007).
47. Anton, T., Bultmann, S., Leonhardt, H. & Markaki, Y. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus* **5,** 163–172 (2014).
48. Greinert, R. *et al.* UVA-induced DNA double-strand breaks result from the repair of clustered oxidative DNA damages. *Nucleic Acids Res.* **40,** 10263–10273 (2012).
49. R-Core-Team. R. A language and environment for statistical computing (2014).
50. Pettersen, E. F. *et al.* UCSF chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
51. Li, R. *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25,** 1966–1967 (2009).
52. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5,** 621–628 (2008).

Acknowledgements

We are indebted to Dr Diana Pignalosa and Dr Sylvia Ritter for SKY analysis. This work was supported by grants of the Deutsche Forschungsgemeinschaft (GRK1657/TP1B to M.C.C.; GRK1657/TP1C to A.R.; DFG CA198/8-1 and 2 to M.C.C.), the Bundesministerium für Bildung und Forschung Grants 02NUK017D and 02S8355 to M.C.C. and 02NUK036D to A.R. and the Beilsetein-Institute, NanoBiC collaboration.

Author contributions

F.N. and A.R. performed the ChIP-Seq experiments. A.M. and A.R. performed the 3D-SIM microscopy. A.S., H.H. and D.H. performed the STED microscopy. F.N., A.R. and W.Y. performed the bioinformatics. T.A. and D.H. performed the CRISPR-Cas9 experiments. M.C.C., F.N. and A.R. designed the project, analysed the data and wrote the manuscript. M.D., W.C., H.L. and G.T.-S. provided tools and expertise, and revised the manuscript.

Additional information

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

Competing interests: The authors declare no competing financial interests.

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

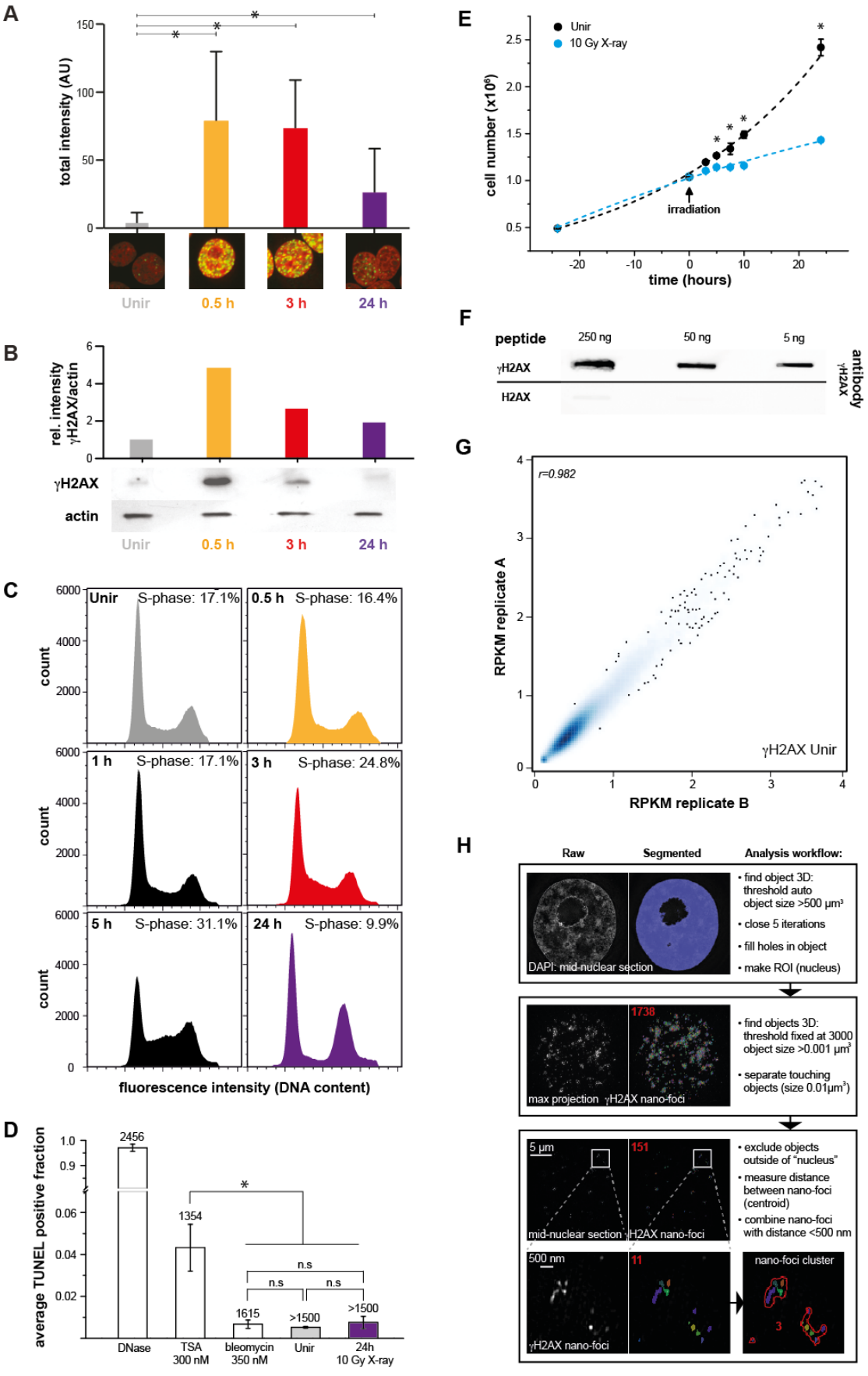
How to cite this article: Natale, F. *et al.* Identification of the elementary structural units of the DNA damage response. *Nat. Commun.* **8,** 15760 doi: 10.1038/ncomms15760 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

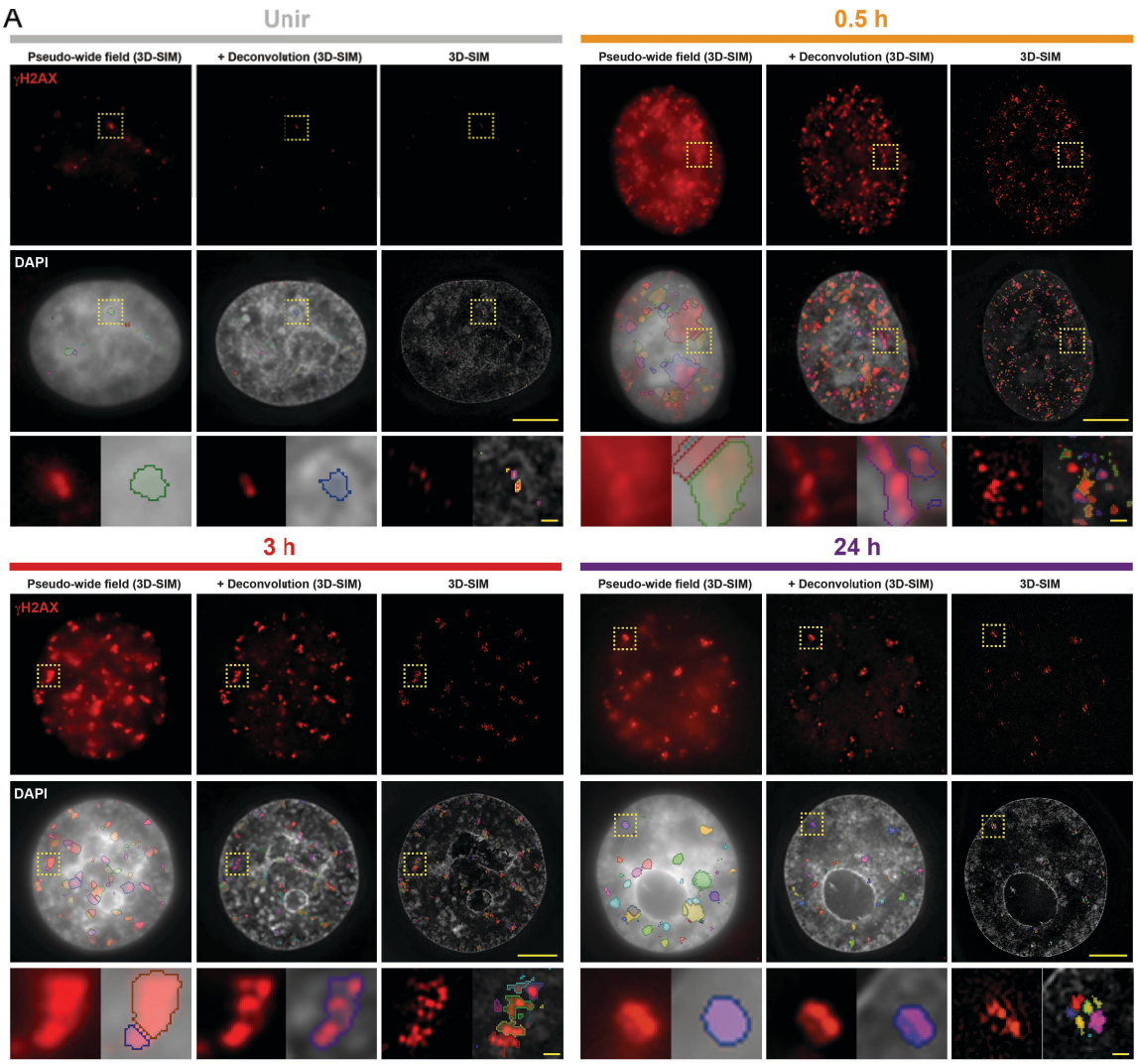
© The Author(s) 2017

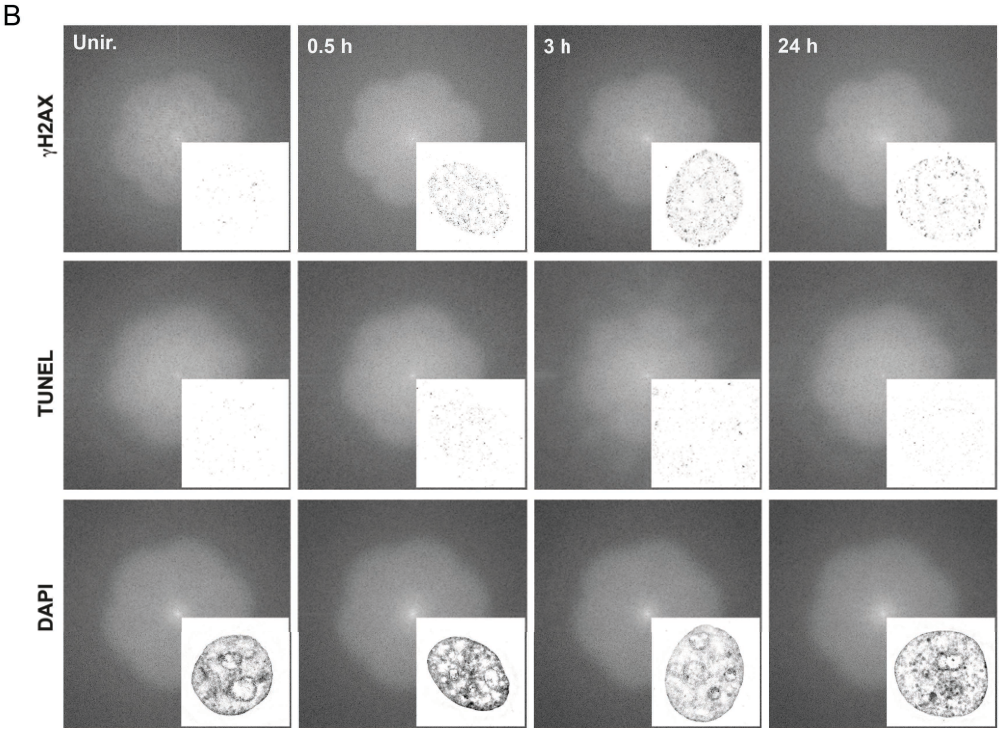
2.3.2 Supplementary information


A
total intensity (AU)
Unir
0.5 h
3 h
24 h
B
rel. intensity γH2AX/actin
γH2AX
actin
C
Unir S-phase: 17.1%
0.5 h S-phase: 16.4%
1 h S-phase: 17.1%
3 h S-phase: 24.8%
5 h S-phase: 31.1%
24 h S-phase: 9.9%
count
fluorescence intensity (DNA content)
D
average TUNEL positive fraction
2456
1354
1615
>1500
>1500
n.s
DNase
TSA 300 nM
bleomycin 350 nM
Unir
24h 10 Gy X-ray
E
Unir
10 Gy X-ray
cell number (x10⁶)
irradiation
time (hours)
F
peptide
250 ng
50 ng
5 ng
γH2AX
H2AX
antibody γH2AX
G
r=0.982
RPKM replicate A
RPKM replicate B
γH2AX Unir
H
Raw
Segmented
Analysis workflow:
DAPI: mid-nuclear section
• find object 3D: threshold auto object size >500 µm³
• close 5 iterations
• fill holes in object
• make ROI (nucleus)
1738
max projection γH2AX nano-foci
• find objects 3D: threshold fixed at 3000 object size >0.001 µm³
• separate touching objects (size 0.01µm³)
5 µm
151
mid-nuclear section γH2AX nano-foci
• exclude objects outside of "nucleus"
• measure distance between nano-foci (centroid)
• combine nano-foci with distance <500 nm
500 nm
11
γH2AX nano-foci
nano-foci cluster
3

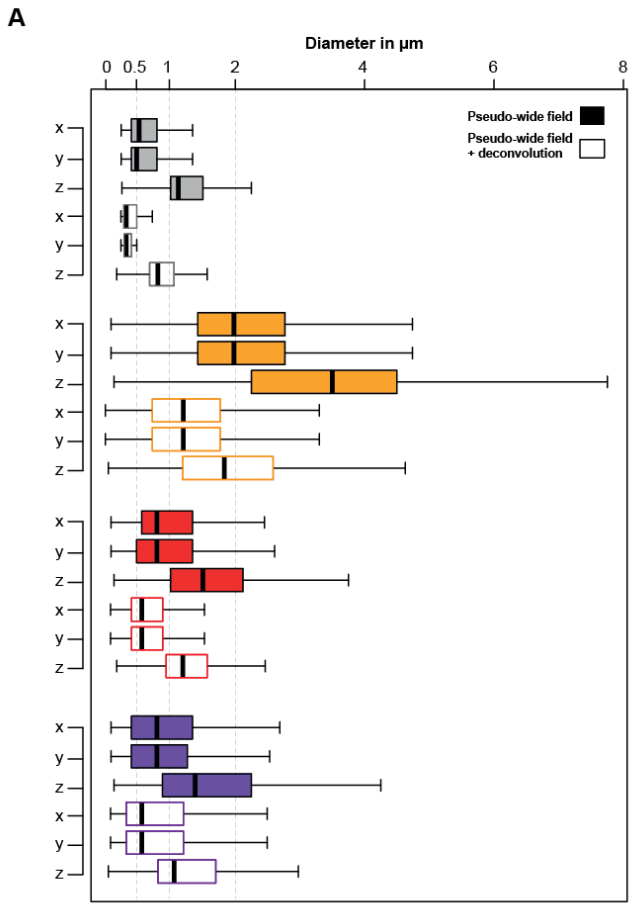
Supplementary Figure 1. Characterization and validation of cellular system and experimental strategy. To characterize the DDR as assessed by γH2AX formation, cells were exposed to 10 Gy X-ray and incubated as indicated. **(A)** Confocal microscopy immunofluorescence analysis of γH2AX before and after exposure to ionizing radiation (IR). Total γH2AX fluorescence intensity (Arbitrary Units) with exemplary micrographs matched to the corresponding time point is shown. In the micrograph: γH2AX (green); propidium iodide counterstained DNA (red). Results are mean and standard deviation from two independent experiments. *: significantly different from the mean of control unirradiated cells (one-way ANOVA, $p<10^{-4}$). **(B)** Immunoblot analysis of γH2AX (top blot) before and after exposure to IR. Loading control: β-actin (bottom blot). The ratios between γH2AX and β-actin chemiluminescence signal intensities is normalized to one for the unirradiated sample (Unir) and shown as bars in an exemplary barplot. **(C)** Cell cycle analysis by flow cytometry. After exposure to IR, cells underwent cell cycle arrest and accumulated in S-phase up to five hours post IR. After repair of DNA damage, cells progressed from the S-phase arrest into G2-phase (24h). Note the reduced population in S-phase 24h post IR. Fraction of cells in S-phase is indicated in each box. Two independent experiments were performed and ~25,000 cells per time point were analyzed. **(D)** Apoptosis analysis by TUNEL assay. Cells were treated with TSA or bleomycin at the indicated concentrations, or exposed to 10 Gy X-ray and incubated for 24h prior to analysis. "DNAse" (positive control) and unirradiated sample (negative control) are included. Results are mean and standard deviation from two independent experiments (n = 30; 15 imaged fields per condition per experiment). Total numbers of screened cells for each sample are indicated above each bar. The fraction of apoptotic cells never exceeded 1% in irradiated cells. *: two-tailed t-test, $p<10^{-4}$. n.s.: not significant. **(E)** Growth curve of cells before and after IR (cyan curve) as opposed to unirradiated control (black curve). Cells were seeded 24h before irradiation or mock-irradiation and cell number was assessed at indicated times. Note that after growth arrest, cells re-entered cell cycle and started proliferating again (24h post IR). Results

represent mean ± SEM from three independent growth curves, each performed in triplicate. *: two-tailed t-test, $p<10^{-2}$. **(F)** Slot blot analysis to test γH2AX antibody specificity. The γH2AX and H2AX peptides used for immunization were blotted at increasing indicated amount. The membrane was then probed with γH2AX antibody. Little (250 ng) to no cross-reactivity of anti-γH2AX antibody with H2AX peptide was observed. **(G)** ChIP-Seq reproducibility was assessed by comparing the RPKM values from two biological replicates. γH2AX ChIP and ChIP-Seq library preparation from two independent experiments are compared. γH2AX RPKM values were computed in 10 kbp genomic intervals, totalling 286,729 intervals. The two biological replicates show high linear correlation, with a Pearson's *r* of 0.982 ($p<2.2\times10^{-16}$). **(H)** Workflow of the image analysis protocol to quantify 3D-SIM data, including nuclear segmentation (top), γH2AX (nano-)foci segmentation (mid) and cluster analysis (bottom). A minimum segmentation unit of 2×2×2 voxels was allowed. An exemplary cell from the 24h time point is shown, together with the number of foci/clusters at all stages (in red). A detailed protocol of the microscopy analysis is in the "Image analysis" section in Methods.


A
Unir
0.5 h
3 h
24 h
Pseudo-wide field (3D-SIM)
+ Deconvolution (3D-SIM)
3D-SIM
γH2AX
DAPI


B
Unir.
0.5 h
3 h
24 h
γH2AX
TUNEL
DAPI

Supplementary Figure 2. (A) Segmentation of γH2AX nano-foci in 3D-SIM images before and during the DNA damage response. Mid-nuclear section (*z*: 125 nm) of representative images of cells before or after exposure to 10 Gy X-ray. The same cell is shown as re-computed pseudo-wide field image before or after deconvolution, as well as the original 3D-SIM output. The total number of segmented focal structures is presented in the middle panel, together with DAPI. The lower panels show magnified views of the yellow dashed frame. Scale bars: 5 µm and 500 nm for main micrographs and magnified regions, respectively. **(B)** 3D-SIM images represented in the Fourier's space. To avoid reconstruction artifacts, the images were controlled in Fourier space. Here, sample images from the γH2AX and TUNEL co-staining are presented with the Fast Fourier Transformed (FFT) images of mid nuclear sections together with the underlying images as insets. No reconstruction artifacts are visible in the information containing central rosettes.



C

SKY Results: Pseudo-chromosome length estimate

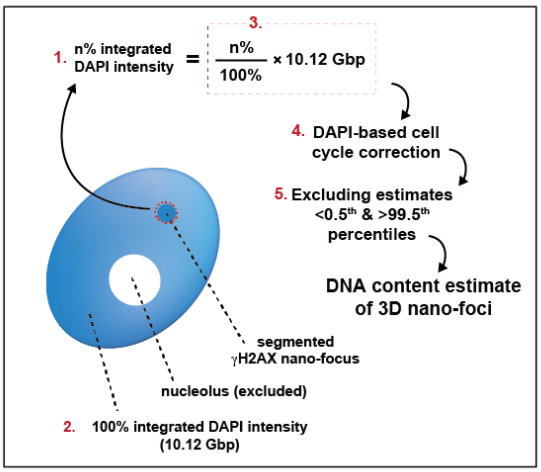
Chr	Frequency (haploid)	Human chromosome length (bp)	Frequency adjusted chromosome length (haploid; bp)
1	1.95	248,956,422	486,596,643
2	1.50	242,193,529	363,290,293
3	1.05	198,295,559	207,308,984
4	1.05	190,214,555	198,860,662
5	3.36	181,538,259	610,628,714
6	1.05	170,805,979	178,569,879
7	2.00	159,345,973	318,691,946
8	2.09	145,138,636	303,471,680
9	1.50	138,394,717	207,592,075
10	1.45	133,797,422	194,614,438
11	1.50	135,086,622	202,629,933
12	2.00	133,275,309	266,550,618
13	1.00	114,364,328	114,364,328
14	1.50	107,043,718	160,565,577
15	1.91	101,991,189	194,710,461
16	1.50	90,338,345	135,507,517
17	2.00	83,257,441	166,514,882
18	1.50	80,373,285	120,559,927
19	1.50	58,617,616	87,926,424
20	1.82	64,444,167	117,171,224
21	1.45	46,709,983	67,941,795
22	1.00	50,818,468	50,818,468
X	1.95	156,040,895	304,989,029
Y	0.00	57,227,415	0

Sum: 5,059,875,502

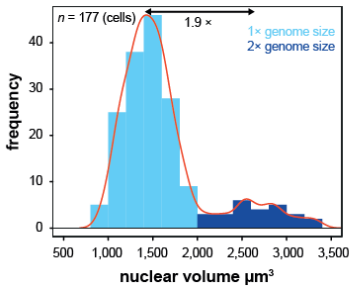
Summary Karyotype

N	46
Mean	73.9
Median	74
SD	2.678
SEM	0.395
Lower 95% CI of Mean	73.03
Upper 95% CI of Mean	74.62
Minimum	61
Maximum	81

B

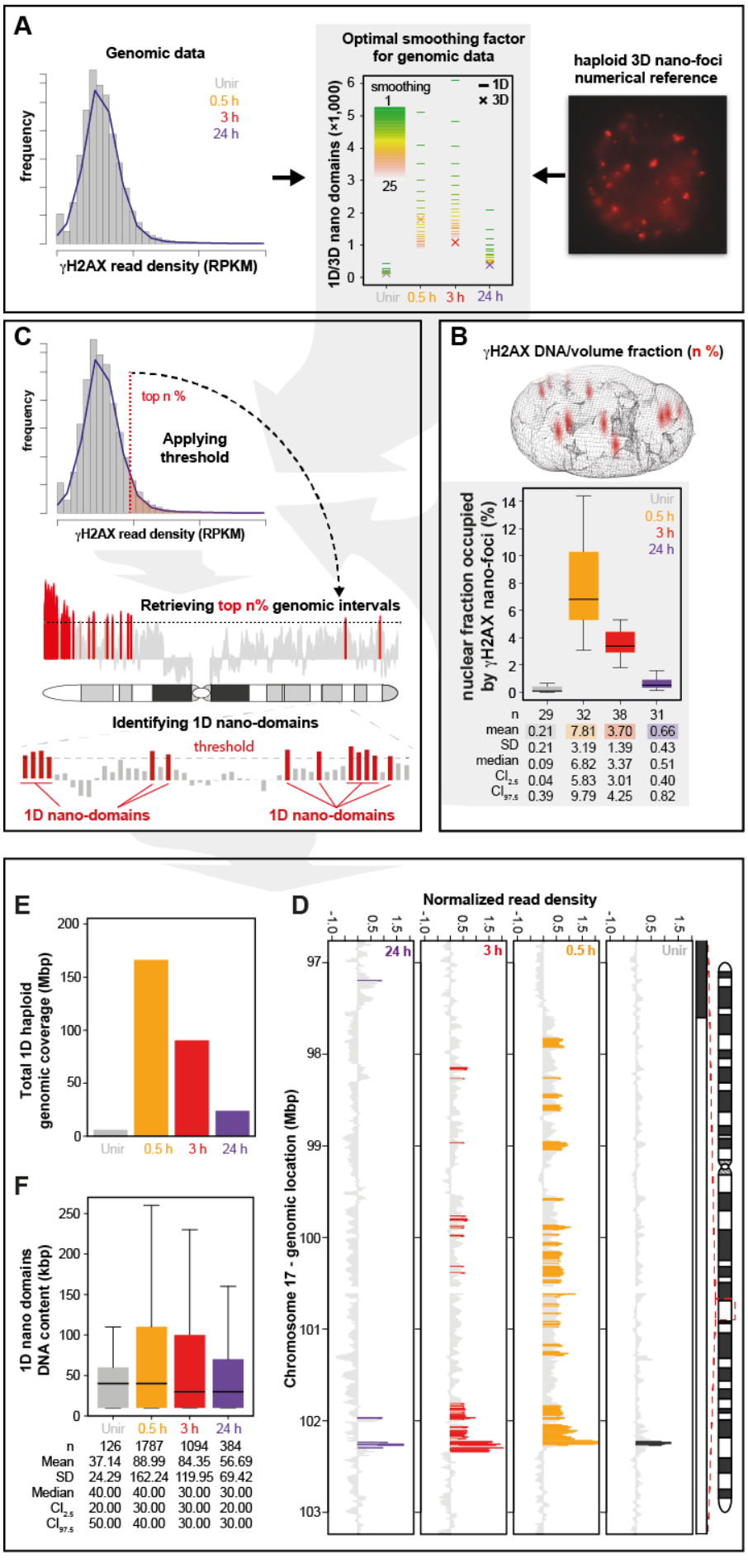


D

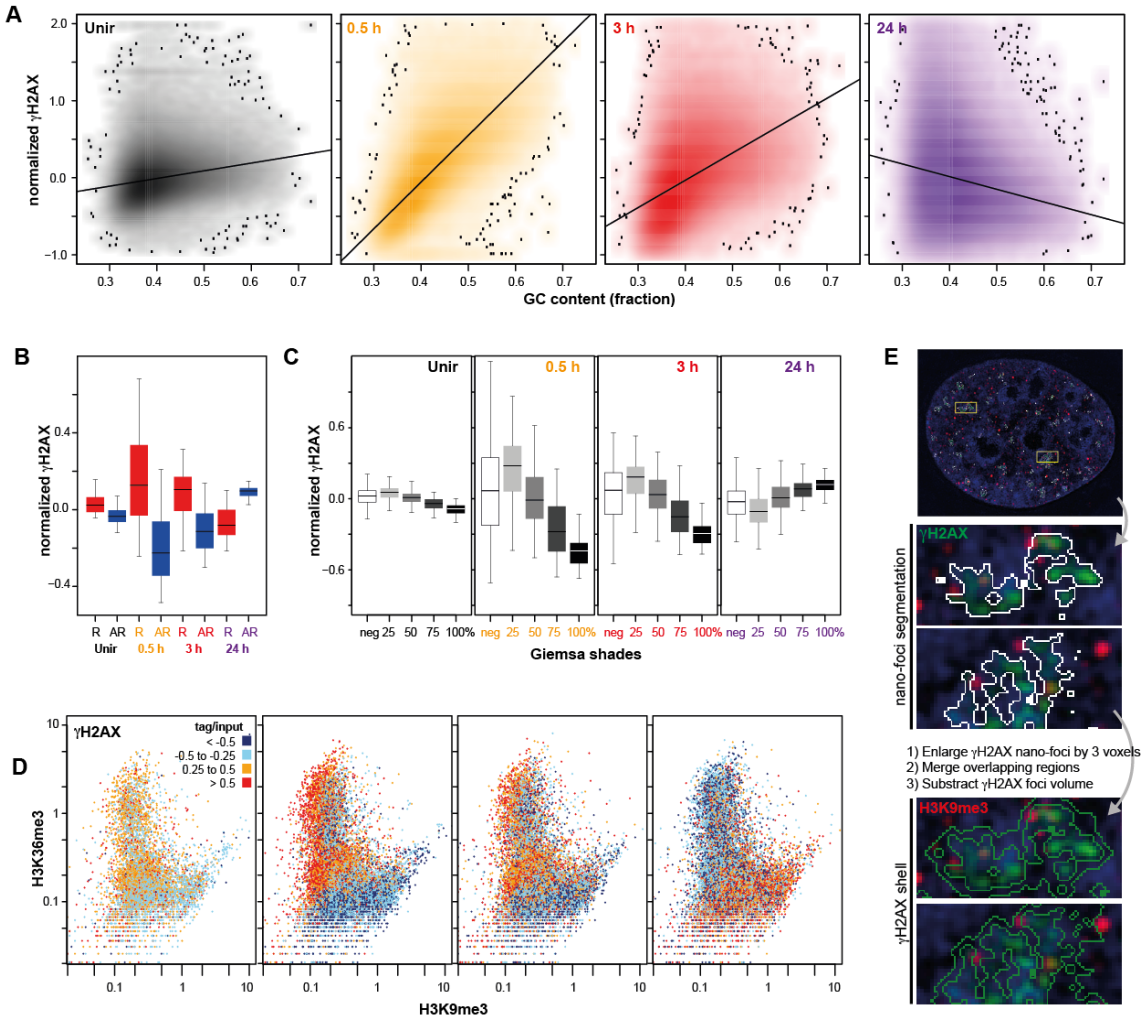


Supplementary Figure 3. Spectral karyotyping analysis of HeLa cells and DAPI-based cell cycle correction. **(A)** Quantification of the lateral and axial diameters of segmented objects in re-computed images, before and during DDR. The difference between lateral and axial measurements is due to the decreased resolution in the axial direction. **(B)** Schematic representation of the measurement of γH2AX nano-foci DNA content. The

whole procedure is summarized in bullet points as follows: i) the nucleus of a cell (excluding the nucleoli) and each γH2AX nano-focus are segmented; ii) the sum of all voxel in the segmented nucleus corresponds to the total integrated DAPI intensity (indicated as „2.“); iii) for each γH2AX nano-focus, the DAPI values of each voxel belonging to the segmented volume are summed (indicated as „1.“); iv) the resulting values are then normalized over the total integrated DAPI intensity (indicated as „3.“); v) this provides the fraction of total DAPI embedded in a single nano-focus, independent of the local DNA condensation state; vi) finally, to estimate the DNA content, the DNA fractions were corrected for the total genome size (determined by spectral karyotyping, panel C) and the cell cycle phase (panel D) **(C)** Relative haploid chromosome frequencies were combined with the human reference chromosome length to generate frequency-adjusted haploid pseudo-chromosomes. The total pseudo-haploid genome (5.06 Gbp) is the sum of all pseudo-chromosomes. A summary of all statistics from SKY is shown in the bottom box and reveals HeLa quasi-tetraploidy. **(D)** Distribution of the nuclear volume of all wild type cells analyzed during the DDR (*n* = 177). The distribution was arbitrarily split into two halves, and the corresponding “genome size” correction factor was used to adjust the nano-foci size (Fig. 2C). The major contribution to the nano-foci size is provided by the “1× genome” fraction.

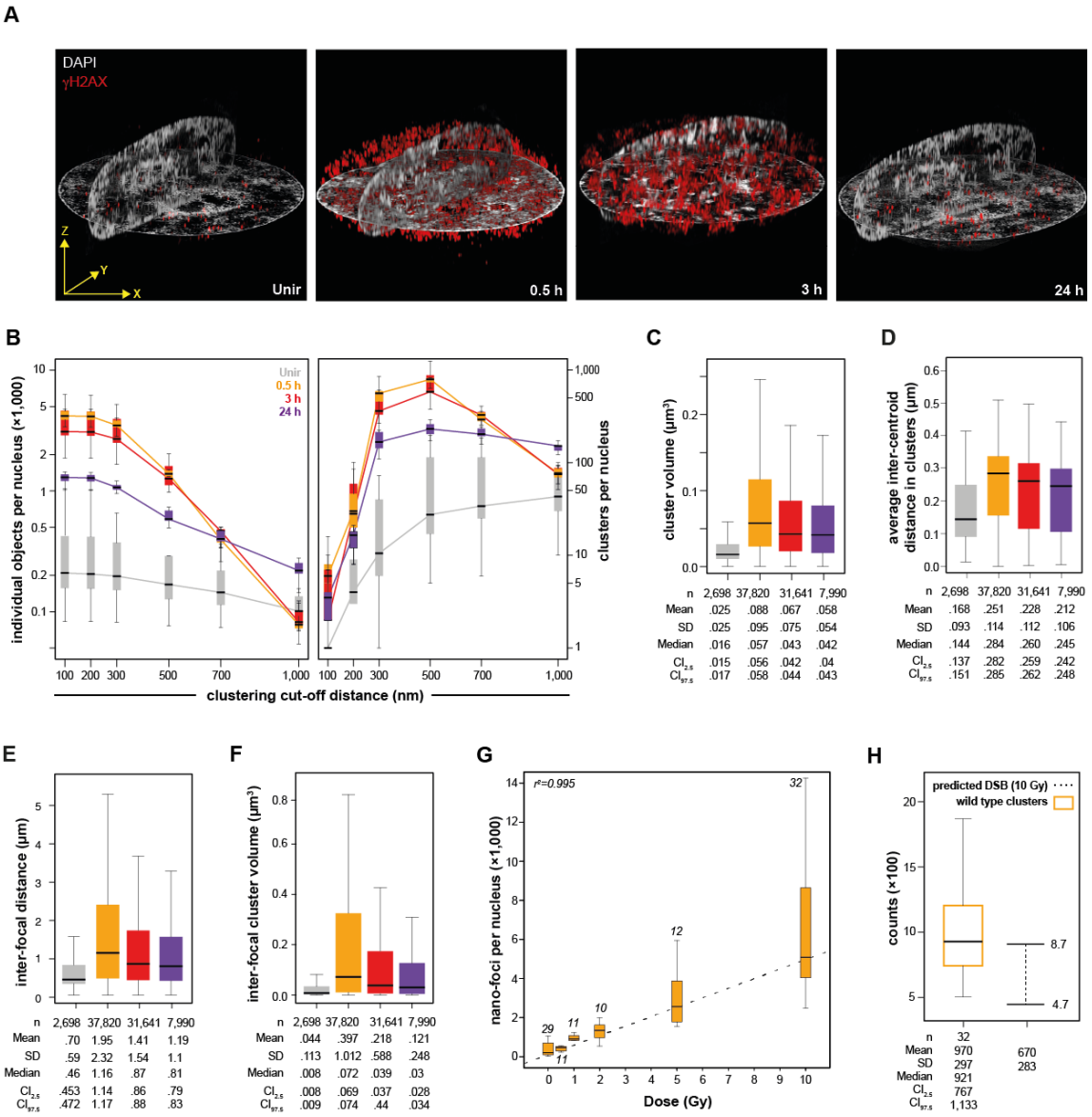

A
Genomic data
frequency
Unir
0.5 h
3 h
24 h
γH2AX read density (RPKM)
Optimal smoothing factor for genomic data
1D/3D nano domains (×1,000)
smoothing
1
25
1D
3D
Unir 0.5 h 3 h 24 h
haploid 3D nano-foci numerical reference
C
frequency
top n %
Applying threshold
γH2AX read density (RPKM)
Retrieving top n% genomic intervals
Identifying 1D nano-domains
threshold
1D nano-domains
1D nano-domains
B
γH2AX DNA/volume fraction (n %)
nuclear fraction occupied by γH2AX nano-foci (%)
Unir
0.5 h
3 h
24 h
n 29 32 38 31
mean 0.21 7.81 3.70 0.66
SD 0.21 3.19 1.39 0.43
median 0.09 6.82 3.37 0.51
CI2.5 0.04 5.83 3.01 0.40
CI97.5 0.39 9.79 4.25 0.82
E
Total 1D haploid genomic coverage (Mbp)
Unir 0.5 h 3 h 24 h
F
1D nano domains DNA content (kbp)
Unir 0.5 h 3 h 24 h
n 126 1787 1094 384
Mean 37.14 88.99 84.35 56.69
SD 24.29 162.24 119.95 69.42
Median 40.00 40.00 30.00 30.00
CI2.5 20.00 30.00 30.00 20.00
CI97.5 50.00 40.00 30.00 30.00
D
Normalized read density
24 h
3 h
0.5 h
Unir
Chromosome 17 - genomic location (Mbp)

Supplementary Figure 4. Integration of 3D-SIM and ChIP-Seq data to estimate the size of genomic γH2AX-decorated chromatin. (**A**) generation of 25 independent profiles by applying a smoothing factor (moving average) to each γH2AX ChIP-Seq dataset (middle panel, coloured lines). Such smoothing factor is a moving average ranging from 1 (no smoothing) to 25 genomic intervals (indicated as "1D"). Crosses (indicated as "3D") are the ploidy-corrected 3D-SIM γH2AX nano-foci. The smoothing factor is chosen according to the best fit between genomic and microscopy data. (**B**) Volume fraction occupied by γH2AX nano-foci as well as their corresponding DNA content, before and during the DNA damage response. (**C**) Filtering the previously smoothed genomic γH2AX ChIP-Seq data by applying the mean volume fractions measured in (B), so that only the 10 kbp genomic intervals from the top-percentiles of the read density distributions at matched time-points were retrieved. (**D**) Exemplary panel showing the filtered intervals from the underlying ChIP-Seq profiles. (**E**) Linear coverage of the filtered ChIP-Seq datasets. The total genomic coverage corresponds to the DNA content estimate we measured in the mean volume fraction from (B). (**F**) Estimate of the 1D domain size distribution.



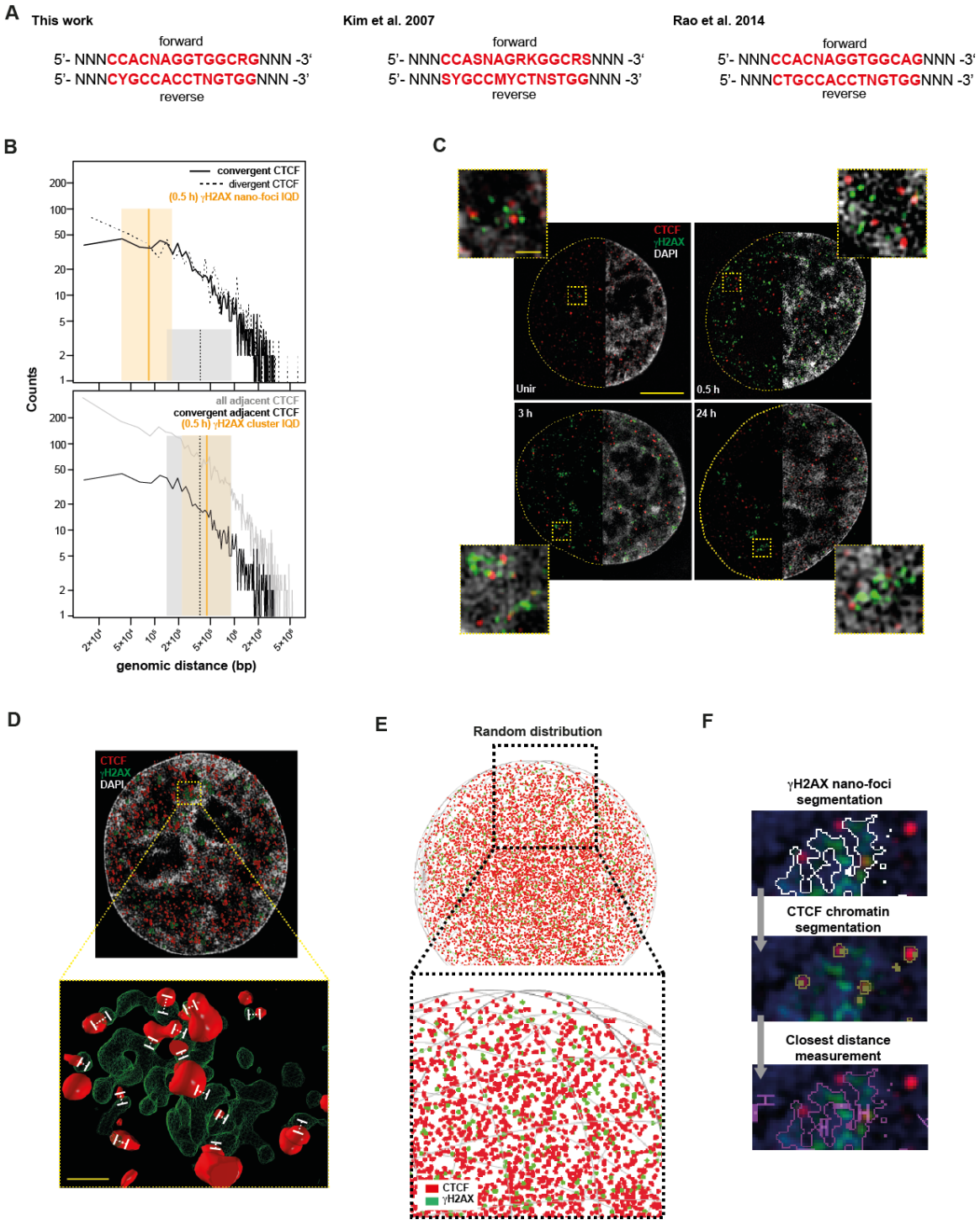
Supplementary Figure 5. Genomic and microscopic analysis of γH2AX-decorated chromatin. **(A)** Genome-wide correlation between γH2AX levels and GC content before and after IR. Normalized levels were calculated as follows: [($\gamma H2AX_{interval}$ RPKM/$input_{interval}$ RPKM)-($\gamma H2AX_{average}$ RPKM/$input_{average}$ RPKM)], where "interval" is a 10 kbp genomic interval and "average" is the genome-wide RPKM average value of all intervals in each corresponding dataset. Data are presented as density scatter plots of normalized γH2AX levels as a function of GC content. The early (0.5h, orange) mid- (3h, red) and late (24h, purple) stages of DDR as well as the sham-irradiated levels (Unir, grey) are shown. Black line: linear regression. Positive correlation with increasing GC content was observed before and up to 3h post IR. At 24h, the tendency was inverted, as indicated by the negative slope of the regression line. **(B)** γH2AX levels in

(anti-)RIDGEs before and after IR. Normalized levels are presented as [(γH2AX $_{(anti-)RIDGE}$ RPKM/input$_{(anti-)RIDGE}$ RPKM) - (γH2AX$_{average}$ RPKM/input$_{average}$ RPKM)] where "(anti-)RIDGE" is the total genomic coverage for all RIDGEs or anti-RIDGEs and "average" is the genome-wide RPKM average value of all genomic intervals. Upon IR, γH2AX is enriched in RIDGEs, whereas at later times the trend is inverted. Wilcoxon rank sum rest; $p<10^{-5}$. **(C)** γH2AX levels in Giemsa-shaded band ideograms before and after IR. Normalized levels are presented as [(γH2AX$_{band\ type}$ RPKM/input$_{band\ type}$ RPKM) - (γH2AX$_{average}$ RPKM/input$_{average}$ RPKM)] where "band type" is the total genomic coverage for each band and "average" is the genome-wide RPKM average value of all genomic intervals. Upon IR, γH2AX is enriched in Giemsa light bands (negative and 25%) whereas at later times the trend is inverted (75-100%). Kruskal-Wallis test and p-values in Supplementary Table 4. **(D)** Genome-wide γH2AX levels before and after IR. Each dot in the scatterplot represents a 10 kbp genomic interval whose coordinates correspond to H3K9me3 (x-axis) and H3K36me3 (y-axis) levels. The relative γH2AX enrichment in each genomic interval is presented as a heat-map, increasing from blue to red. It is to be noted that, upon IR, γH2AX is enriched in H3K36me3-rich/H3K9me3-poor compartments. Conversely, at later times, residual γH2AX signal is mainly found in H3K36me3-poor/H3K9me3-rich compartments. **(E)** γH2AX nano-foci (green) are segmented as described in Methods. The resulting volume units are then enlarged by three voxels in the three dimensions. All overlapping regions are merged to form a distinct volume unit. Finally, the volume of the original γH2AX nano-foci is subtracted to generate γH2AX shells. Fluorescence intensity of other probed features (e.g. H3K9me3, red) are then measured in the shells. The enlarged panels correspond to regions defined by the yellow frames. All boxes and whiskers represent 25-75 percentiles and three times the interquartile distance.



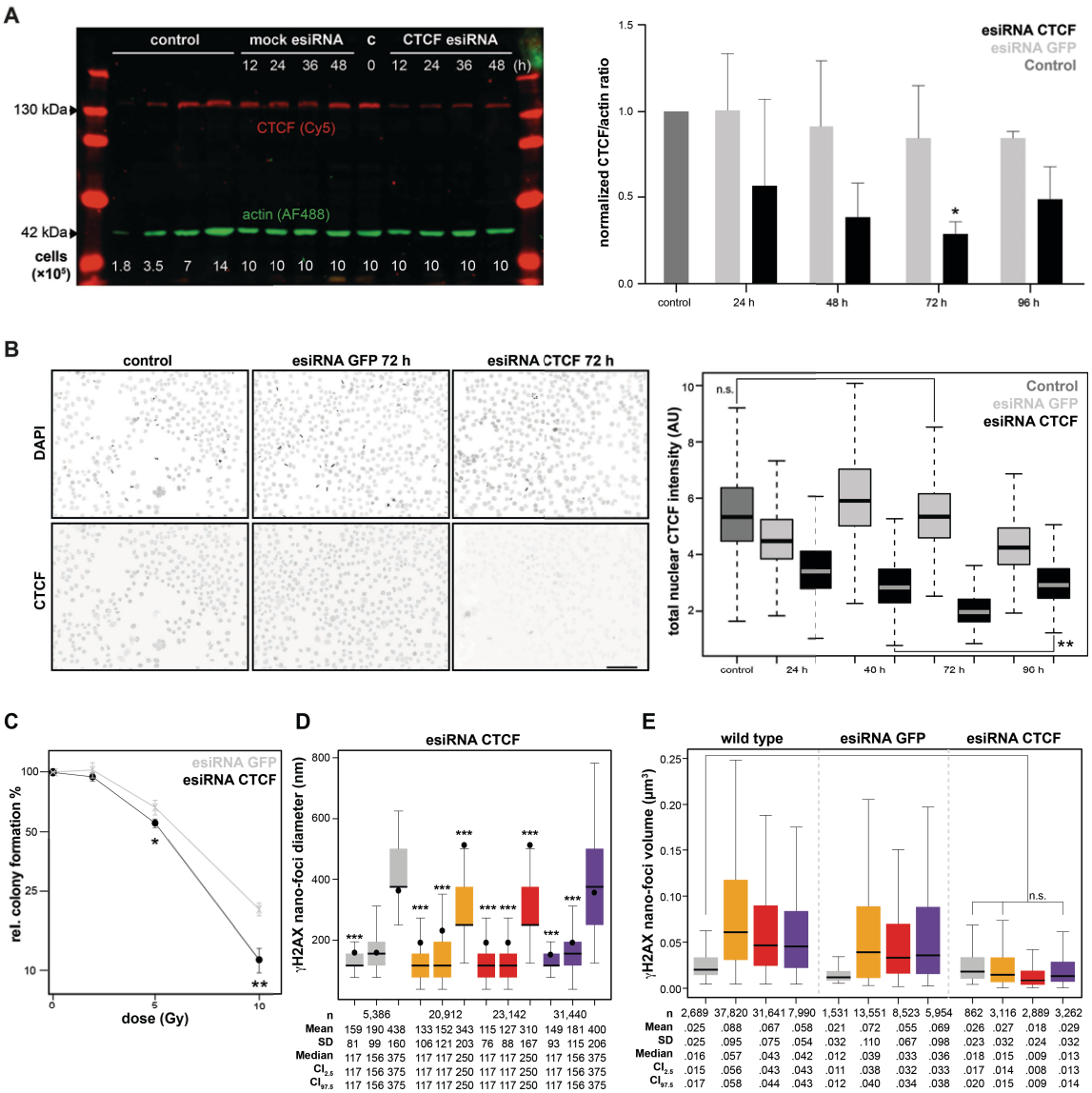
Supplementary Figure 6. Validation of γH2AX nano-foci and nano-foci clusters in cells. **(A)** Exemplary 3D-SIM images of γH2AX (red) before and during DDR showing a 3D representation of γH2AX nano-foci with DAPI channels in *xy* and *yz* mid-nuclear cross-sections. **(B)** Effect of cut-off distance between nano-foci for the cluster analysis. 10 cells per time point were analyzed for the effect of the clustering threshold distance (from 100 to 1,000 nm) and the resulting distributions are presented as boxplots. (left) Sum of 3D-clusters plus individual non-clustered nano-foci. (right) Total number of 3D-clusters. 500 nm was the cut-off distance resulting in both the highest number of clusters and clear repair kinetics. Solid lines connect the medians of each distribution.

(C) γH2AX 3D cluster integrated volume distributions. The volume of each nano-focus included in a cluster is summed. Kruskal-Wallis chi-squared = 2,941.4, df = 3, $p<2.2\times10^{-16}$. **(D)** Distribution of the average inter-centroid distances measured between each nano-focus belonging to a given cluster. Kruskal-Wallis chi-squared = 1,889.3, df = 3, $p<2.2\times10^{-16}$. **(E)** Distributions of the shortest paths connecting the centroids of all nano-foci belonging to a 3D cluster Kruskal-Wallis chi-squared = 2,223.7, df = 3, $p<2.2\times10^{-16}$. **(F)** Inter-focal 3D-clusters volume distributions, presented as the volume delimited by the centroids of each nano-focus belonging to a 3D-clusters Kruskal-Wallis chi-squared=2,217.5, df=3, $p<2.2\times10^{-16}$. **(G)** Dose-curve showing linear increase of γH2AX nano-foci. Cells were irradiated with 0.5, 1, 2, 5 and 10 Gy X-ray and incubated 0.5h before fixation. γH2AX immunofluorescence was followed by nano-foci quantification on 3D-SIM images. The number of imaged cells per dose is shown in italic. Dashed line: linear regression calculated over the median of each distribution, after subtracting the median number of nano-foci from unirradiated cells. Estimated nano-foci per Gy: 495, after background subtraction. Kruskal-Wallis chi-squared = 88.028, df = 5, $p<2.2\times10^{-16}$. All boxes and whiskers are 25^{th}-75^{th} percentile and three times the interquartile distance, respectively. n: number of analyzed 3D clusters. **(H)** Comparison between the numbers of γH2AX clusters, 0.5h post IR and the predicted number of DSBs induced by 10 Gy X-ray.



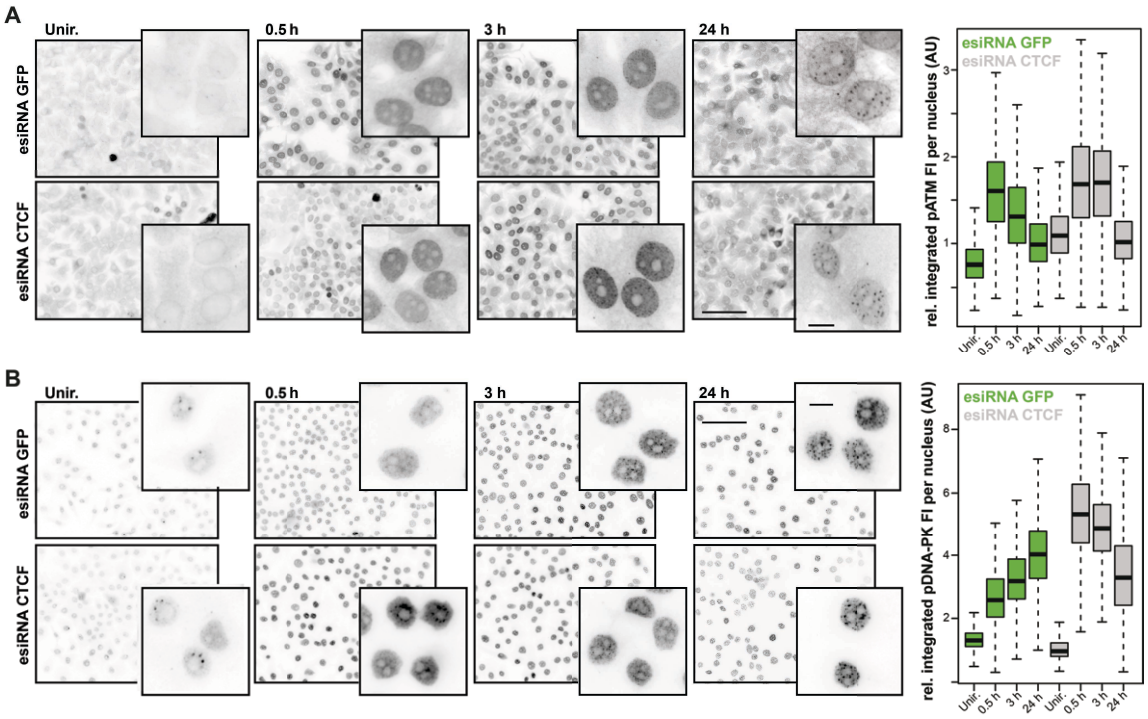
Supplementary Figure 7. Spatial localization of CTCF with respect of γH2AX-decorated chromatin. (A) CTCF consensus motifs used in this work and from previous works. **(B)** Size comparison between CTCF-delimited chromatin segments and γH2AX 3D-nano-foci and clusters. Because the CTCF motif is not a palindrome, two adjacent motifs can have four possible orientations. Recent findings indicate that, when a

chromatin loop is formed, the CTCF sites are facing one another in a convergent fashion in almost all cases. The distances between all adjacent CTCF genomic sites independent of their orientation (grey curve) as well as those between two adjacent convergent (black solid curves) or divergent (dashed curves) CTCF genomic sites are shown. The IQD of the "convergent adjacent" distribution (grey box; dashed line: median) is compared to that of γH2AX 3D clusters (left; orange box; line: median) or γH2AX nano-foci size (right; orange box; line: median). Little to no difference is observed when comparing distances between convergent adjacent CTCF sites and distances between random or divergent orientation (IQD: 150-987 kbp). **(C)** 3D-SIM images of immuno-stained γH2AX and CTCF before and during DDR. The DAPI channel represents the mid-nuclear section. The dotted curved line delimits the nuclear contour. Panels on the right are enlarged views of a representative region (yellow dashed lines). **(D)** Three-dimensional rendering of γH2AX (green) and CTCF (red) immunostaining in a mid nuclear section, 24h after IR. The enlarged region represents γH2AX foci clusters surrounded by CTCF. White dashed lines: exemplary measurements. **(E)** Graphical representation of simulated γH2AX and CTCF foci 0.5h post IR in a sphere of volume comparable to that of a cell nucleus. The number of γH2AX and CTCF foci used in the simulation are matched to the number of foci detected in 3D-SIM images of each time-point. Specifically, 5,348, 6,731, 8,154, 7,497 CTCF and 374, 4,357, 4,065 and 1,200 γH2AX nano-foci were used for unirradiated, 0.5h, 3h and 24h time points, respectively. **(F)** Shell segmentation and analysis workflow for the measurement of CTCF proximity to γH2AX foci: i) γH2AX foci are segmented; ii) CTCF foci are segmented; iii) the closest Euclidian distance between the centroids of γH2AX and CTCF foci is measured.



Supplementary Figure 8. CTCF knock-down via RNAi. (A) Left, representative immunoblot of CTCF protein in the absence or presence of CTCF-esiRNA; lanes 1 and 15: protein ladder; lanes 2 to 5: loading control with increasing amount from left to right (the cell number is indicated below); lanes 6 to 9: mock (GFP) esiRNA, quadruplicate; lane 10: untransfected control; lanes 11 to 14: CTCF esiRNA, quadruplicate. Right, quantification of CTCF protein levels, relative to the untransfected control. 72h after incubation with CTCF esiRNA, we observed the maximum depletion, with CTCF protein levels being about 40%, compared to the corresponding mock-treated sample (ANOVA with Dunnett's correction; *: p<0.05, relative to control). **(B)** High-content immunofluorescence microscopy of cells in which CTCF protein was knocked-down via

esiRNA. Wide field images of DNA (up) and CTCF (down) immunofluorescence in the absence or presence of CTCF-esiRNA for the indicated times are shown. GFP esiRNA (middle panels) was used as mock transfection control. Scale bar: 100 µm. Right, the boxes are the distributions of total nuclear CTCF fluorescence intensity from at least 4,500 cells at the indicated times post esiRNA treatment. Kruskal-Wallis test with Dunn's multiple comparison correction. **: $p<10^{-2}$; n.s.: not significant; all other pairs are significantly different with a $p<10^{-3}$. **(C)** Diminished CTCF levels increase radiosensitivity. Colony formation assay was performed after exposing cells to the indicated X-ray doses. Values are mean and standard deviation from two independent assays. For each experiment, 3 and 6 technical replicates, for unirradiated and irradiated samples at the indicated doses, respectively, were analyzed. Two-tailed t-test with $p<0.05$ (*) or 0.01 (**). **(D)** Quantification of γH2AX nano-foci diameters in CTCF siRNA-treated cells before and after IR. Black dots: median length of γH2AX nano-foci diameters in untreated cells (from Fig. 2A). ***: Wilcoxon rank-sum test, $p<<10^{-3}$. **(E)** γH2AX 3D-clusters integrated volume distributions (as in Supplementary Fig. 6C) in untreated (control), mock- and CTCF-depleted cells. n.s.: Kruskal-Wallis test with Dunn's multiple comparison, with $p>0.05$.



Supplementary Figure 9. High content immunofluorescence microscopy of phospho-ATM **(A)** or phospho-DNA-PKcs **(B)** before or during DDR, in CTCF-depleted cells. Briefly, cells were exposed to 10 Gy IR, incubated for the indicated times and then fixed. High content immunofluorescence microscopy and analysis were then performed with an Operetta System. Scale bar: 100 µm, inset scale bar: 10 µm. Results are from two independent experiments, with >5,000 individual cells per condition per time-point analyzed. Kruskal-Wallis test with Dunn's multiple comparison correction; all pairs are significantly different with a $p<10^{-3}$. All boxes and whiskers represent 25-75 percentiles and three times the interquartile distance.

Supplementary Table 1

DNA content of γH2AX nano-foci and clusters

3D nano-foci (kbp)	Min.	LowQ	Med.	Mean	SD	UpQ	Max
Unirradiated (*n* = 16,798)	0.8	23.8	38.9	53.6	48.1	64.9	369.2
0.5 h (*n* = 233,515)	0.7	31.7	69.7	119.9	142.5	148.5	1,100.0
3 h (*n* = 166,841)	1.9	27.9	57.8	106.6	138.3	125.9	1,281.0
24 h (*n* = 50,143)	0.7	17.8	32.7	80.8	174.7	67.7	2,008.0
3D clusters (kbp)	**Min.**	**LowQ**	**Med.**	**Mean**	**SD**	**UpQ**	**Max**
Unirradiated (*n* = 2,698)	21.1	81.7	132.1	190.7	173.8	234.6	1,624
0.5 h (*n* = 37,820)	20.7	195.3	424.2	687.2	782.6	884.6	10,224
3 h (*n* = 31,641)	21.7	141.5	300.7	478.9	547.8	622.1	12,251
24 h (*n* = 7,990)	20.9	87.7	189.1	286.9	299.8	377.3	3,746

Supplementary Table 2

Ploidy-corrected DNA content of γH2AX nano-foci and clusters

3D nano-foci (kbp) ploidy-corrected	Min.	LowQ	Med.	Mean	SD	UpQ	Max
Unirradiated (*n* = 16,798)	0.7	23.4	38.9	53.2	48.1	64.9	369.2
0.5 h (*n* = 233,515)	0.7	33.7	74.6	126.9	149.9	159.1	1,137.0
3 h (*n* = 166,841)	1.9	27.9	57.4	106.2	138.2	125.2	1,281.0
24 h (*n* = 50,143)	0.7	25.1	48.1	104.2	180.2	104.4	2,008.0
3D clusters (kbp)	**Min.**	**LowQ**	**Med.**	**Mean**	**SD**	**UpQ**	**Max**
Unirradiated (*n* = 2,698)	2.1	74.2	123.3	178.1	154.2	223.1	1,035.8
0.5 h (*n* = 37,820)	10.2	197.3	440.8	710.2	767.3	937.7	5,324.3
3 h (*n* = 31,641)	15.1	136.8	296.1	469.8	469.7	622.4	3,539.2
24 h (*n* = 7,990)	4.3	111.9	269.2	400.3	389.8	553.9	2,439.3

n: number of nano-foci (top) or clusters (bottom); Min., Max.: minimum and maximum value in the distribution; LowQ, UpQ: 25th and 75th percentiles of the distribution; Med.: median; SD: standard deviation.

Supplementary Table 3

Overview of genomic features used

Feature	Cell Type	Type of Data	Data Source / Reference
General Features			
G-banding	Human	% Shading	UCSC Genome Brower
Distance to the telomere	Hg19	Distance in bp	UCSC Genome Brower
Distance to the centromere	Hg19	Distance in bp	UCSC Genome Brower
Purine percent	Hg19	Percentage	In-house calculation
GC content	Hg19	Percentage	In-house calculation
DNase	HepG2	DNase-seq	GSM816662
FAIRE	HepG2	FAIRE-seq	GSM864354
CpG island	Hg19	Count	UCSC Genome Brower
Transcription			
miRNA	Human	Count	miRBase[1]
TSS	Hg19	Distance in bp	UCSC Genome Brower
Expression	HepG2	Micro array	GSM646144-5[2]
Rel. Pol2	HepG2	Chip-Seq	GSM822284
Rel. Pol2_S2	HepG2	Chip-Seq	GSM935543
RIDGES	Human	Coordinates	http://r2.amc.nl
Genic region	Hg19	Count	UCSC Genome Brower
DNA Methylation			
Average DNA Methylation	HepG2	Micro array	GSM999338
Number of DNA methylation sites (No. DNA Methyl.)	HepG2	Micro array count	GSM999338
Relative MBD4 abundance (Rel. MBD4)	HepG2	ChIP-seq	GSM1010740
Histones and Histone Modifications			
H2A.Z	HepG2	Chip-Seq	GSM733774[3]
H3K4me1	HepG2	Chip-Seq	GSM798321[3]
H3K36me3	HepG2	Chip-Seq	GSM733685[3]
H3K9me3	HepG2	Chip-Seq	GSM1003519[3]
H3K79me2	HepG2	Chip-Seq	GSM733641[3]
H3K27ac	HepG2	Chip-Seq	[3]
H3K27me3	HepG2	Chip-Seq	[3]
H3K4me2	HepG2	Chip-Seq	[3]
H3K4me3	HepG2	Chip-Seq	[3]
H3K9ac	HepG2	Chip-Seq	[3]
H4K20me1	HepG2	Chip-Seq	[3]
DNA Sequence Elements			
Alu repeats	Human	Count	RepeatMasker[4]
MIR repeats	Human	Count	RepeatMasker[4]
LINE1 repeats	Human	Count	RepeatMasker[4]
LINE2 repeats	Human	Count	RepeatMasker[4]
MER repeats	Human	Count	RepeatMasker[4]
AT Low Complexity repeats	Human	Count	RepeatMasker[4]
GC Low Complexity repeats	Human	Count	RepeatMasker[4]
Simple repeats	Human	Count	RepeatMasker[4]
G-Quadruplex Forming repeats (Quadruplex repeats)	Human	Count	RepeatMasker[4]
Z-DNA Motif	Human	Count	[5]
Z-DNA hotspot	Human	Count	[5]
Inverted repeats	Human	Count	[5]
Cruciform Motif	Human	Count	[5]
Direct repeats	Human	Count	[5]
Slipped Motif	Human	Count	[5]

Mirror repeats	Human	Count	[5]
Triplex Motif	Human	Count	[5]
A-Phased repeats	Human	Count	[5]
Microsatellite	Human	Count	RepeatMasker[4]
DNA Replication			
Replication timing S1	GM12801	RepliSeq	GSM923440[6]
Replication timing S2	GM12801	RepliSeq	GSM923440[6]
Replication timing S3	GM12801	RepliSeq	GSM923440[6]
Replication timing S4	GM12801	RepliSeq	GSM923440[6]
Replication timing G1b	GM12801	RepliSeq	GSM923440[6]
Replication timing G2	GM12801	RepliSeq	GSM923440[6]
Origins of replication by lambda exonuclease digestion (Origin Replication Lexo)	HeLa	Genomic array	[7]
Origins of replication by anti-bromodeoxyuridine IP (Origin Replication BrIP)	HeLa	Genomic array	[7]
Origins of replication by common anti-bromodeoxyuridine IP and lambda exonuclease digestion (Lexo + BrIP)	HeLa	Genomic array	[7]
Origins of replication (Ori. Cadoret)	HeLa	Genomic array	[8]
Topoisomerase motif (Topo.CAT)	Hg19	Density	[9]
Topoisomerase motif (Topo.CTY)	Hg19	Density	[9]
Topoisomerase motif (Topo.GTY)	Hg19	Density	[9]
Topoisomerase motif (Topo.RAK)	Hg19	Density	[9]
Topoisomerase motif (Topo.YCCTT)	Hg19	Density	[9]
Topoisomerase motif (Topo.YTA)	Hg19	Density	[9]
DNA Binding Factors			
SMC3 (cohesin)	HepG2	ChIP-seq	GSM935542
Lamina Associated Domain	Tig3ET	Coverage	[10]
Rel.BRCA1	HepG2	ChIP-seq	[3]
Rel.Rad21	HepG2	ChIP-seq	[3]

Supplementary Table 4

Summary of statistical analyses related to figure S5

Figure	Sample	Test	p-value	Comment
S5B	Unir	Wilcoxon rank sum	2.91e-6	RIDGEs (xx) vs. anti-RIDGEs (xx)
S5B	0.5 h	Wilcoxon rank sum	2.98e-10	RIDGEs (xx) vs. anti-RIDGEs (xx)
S5B	3 h	Wilcoxon rank sum	2.95e-9	RIDGEs (xx) vs. anti-RIDGEs (xx)
S5B	24 h	Wilcoxon rank sum	< 2.2e-16	RIDGEs (xx) vs. anti-RIDGEs (xx)
S5C	Unir “0 vs 25”	Kruskal-Wallis	< 5e-2	Giemsa bands group comparison
S5C	Unir “0 vs 50”	Kruskal-Wallis	n.s.	
S5C	Unir “0 vs 75”	Kruskal-Wallis	< 1e-3	
S5C	Unir “0 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	Unir “25 vs 50”	Kruskal-Wallis	< 1e-2	
S5C	Unir “25 vs 75”	Kruskal-Wallis	< 1e-3	
S5C	Unir “25 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	Unir “50 vs 75”	Kruskal-Wallis	< 1e-3	
S5C	Unir “50 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	Unir “75 vs 100”	Kruskal-Wallis	< 5e-2	
S5C	0.5 h “0 vs 25”	Kruskal-Wallis	< 1e-2	
S5C	0.5 h “0 vs 50”	Kruskal-Wallis	n.s.	
S5C	0.5 h “0 vs 75”	Kruskal-Wallis	< 1e-3	
S5C	0.5 h “0 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	0.5 h “25 vs 50”	Kruskal-Wallis	< 1e-2	
S5C	0.5 h “25 vs 75”	Kruskal-Wallis	< 1e-3	
S5C	0.5 h “25 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	0.5 h “50 vs 75”	Kruskal-Wallis	< 1e-3	
S5C	0.5 h “50 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	0.5 h “75 vs 100”	Kruskal-Wallis	< 1e-2	
S5C	3 h “0 vs 25”	Kruskal-Wallis	< 1e-3	
S5C	3 h “0 vs 50”	Kruskal-Wallis	n.s.	
S5C	3 h “0 vs 75”	Kruskal-Wallis	< 1e-3	
S5C	3 h “0 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	3 h “25 vs 50”	Kruskal-Wallis	< 1e-2	
S5C	3 h “25 vs 75”	Kruskal-Wallis	< 1e-3	
S5C	3 h “25 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	3 h “50 vs 75”	Kruskal-Wallis	< 1e-3	
S5C	3 h “50 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	3 h “75 vs 100”	Kruskal-Wallis	< 1e-2	
S5C	24 h “0 vs 25”	Kruskal-Wallis	< 1e-2	
S5C	24 h “0 vs 50”	Kruskal-Wallis	n.s.	
S5C	24 h “0 vs 75”	Kruskal-Wallis	< 1e-3	
S5C	24 h “0 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	24 h “25 vs 50”	Kruskal-Wallis	< 1e-3	
S5C	24 h “25 vs 75”	Kruskal-Wallis	< 1e-3	
S5C	24 h “25 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	24 h “50 vs 75”	Kruskal-Wallis	< 1e-2	
S5C	24 h “50 vs 100”	Kruskal-Wallis	< 1e-3	
S5C	24 h “75 vs 100”	Kruskal-Wallis	n.s.	

Supplementary Table 5

Summary of γH2AX (nano-)foci and cluster numbers

	γH2AX foci		γH2AX nano-foci							γH2AX nano-foci clusters		
	Confocal microscopy		pseudo-wide field		deconvolved pseudo-wide field		3D-SIM		% DDR*	3D-SIM cluster		% DDR*
Unir	13±10	(11)	46±33	(44)	75±26	(77.5)	392±347	(208)	(4.1)	68±70	(23)	(2.5)
0.5 h	53±20	(47)	268±56	(268)	427±83	(406)	6,287±2,785	(5,083.5)	(100)	970±297	(920.5)	(100)
3 h	44±14	(46)	194±73	(174)	361±111	(336)	3,603±1,148	(3,166.5)	(62.3)	663±171	(623)	(67.7)
24 h	23±19	(20)	128±55	(129)	197±45	(209)	1,210±406	(1,267)	(24.9)	203±74	(220)	(23.9)

Indicated are the mean number of γH2AX (nano-)foci ± SD as well as the median (in brackets).
*: assessed as percentage of γH2AX nano-foci or clusters relative to the median value from 0.5h (100%). Note that percentages are comparable between nano-foci and clusters, indicating that the cut-off distance from Supplementary Figure 6B did not impede the analysis of DDR.

Supplementary References

1. Griffiths-Jones, S., Saini, H.K., van Dongen, S. & Enright, A.J. miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**, D154-8 (2008).

2. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-9 (2011).

3. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

4. Smit, A.F.A., Hubley, R. , Green, P. RepeatMasker Open 3.0. (1996-2010).

5. Cer, R.Z. et al. Non-B DB: a database of predicted non-B DNA-forming motifs in mammalian genomes. *Nucleic Acids Res* **39**, D383-91 (2011).

6. Hansen, R.S. et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci U S A* **107**, 139-44 (2010).

7. Karnani, N., Taylor, C.M., Malhotra, A. & Dutta, A. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol Biol Cell* **21**, 393-404 (2010).

8. Cadoret, J.C. et al. Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc Natl Acad Sci U S A* **105**, 15837-42 (2008).

9. Arlt, M.F. & Glover, T.W. Inhibition of topoisomerase I prevents chromosome breakage at common fragile sites. *DNA Repair (Amst)* **9**, 678-89 (2010).

10. Guelen, L. et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948-51 (2008).

2.4 A Simple and Sensitive High-Content Assay for the Characterization of Antiproliferative Therapeutic Antibodies

2.4.1 Main Paper

Check for updates

Technical Note

A Simple and Sensitive High-Content Assay for the Characterization of Antiproliferative Therapeutic Antibodies

SLAS Discovery
2017, Vol. 22(3) 309–315
© 2016 Society for Laboratory Automation and Screening
DOI: 10.1177/1087057116677821
journals.sagepub.com/home/jbx
SAGE

Andreas Stengl[1], David Hörl[1], Heinrich Leonhardt[1], and Jonas Helma[1]

Abstract

Monoclonal antibodies (mAbs) have become a central class of therapeutic agents in particular as antiproliferative compounds. Their often complex modes of action require sensitive assays during early, functional characterization. Current cell-based proliferation assays often detect metabolites that are indicative of metabolic activity but do not directly account for cell proliferation. Measuring DNA replication by incorporation of base analogues such as 5-bromo-2′-deoxyuridine (BrdU) fills this analytical gap but was previously restricted to bulk effect characterization in enzyme-linked immunosorbent assay formats. Here, we describe a cell-based assay format for the characterization of antiproliferative mAbs regarding potency and mode of action in a single experiment. The assay makes use of single cell–based high-content-analysis (HCA) for the reliable quantification of replicating cells and DNA content via 5-ethynyl-2′-deoxyuridine (EdU) and 4′,6-diamidino-2-phenylindole (DAPI), respectively, as sensitive measures of antiproliferative mAb activity. We used trastuzumab, an antiproliferative therapeutic antibody interfering with HER2 cell surface receptor-mediated growth signal transduction, and HER2-overexpressing cell lines BT474 and SKBR3 to demonstrate up to 10-fold signal-to-background (S/B) ratios for treated versus untreated cells and a shift in cell cycle profiles indicating antibody-induced cell cycle arrest. The assay is simple, cost-effective, and sensitive, providing a cell-based format for preclinical characterization of therapeutic mAbs.

Keywords

therapeutic antibodies, cell-based assays, high-content screening, EdU, proliferation

Introduction

Biological drugs such as therapeutic antibodies are in the process of replacing chemical compounds as the major class of future medicines. Therapeutic antibodies are often characterized by complex modes of action, such as inhibition of cell proliferation, induction of apoptosis, and targeted immune recruitment. Moreover, antibody drug conjugates (ADCs) that combine chemotherapeutic cytotoxicity with antibody-mediated tumor specificity even increase the diversity of potential modes of action.[1] Thus, the functional characterization during early drug development requires sensitive cell-based high-throughput assays that address this complexity and measure multiple cellular parameters.[2] One of the major modes of action of therapeutic antibodies is based on inhibition of target cell growth by, for example, blocking growth signaling pathways in cancer cells.[3] For assessing the antiproliferative potency of such candidates, several methods have been described.[4] A simple approach to quantify the number of cells that survive treatment consists of automated cell counting.[5] However, a significant proportion of remaining cells is likely to have entered apoptosis or cell cycle arrest, leading to an overestimation of the proliferating cell population. A more precise approximation of proliferation can be achieved by detecting metabolic activity in viable cells and thus excluding apoptotic cells. Compounds such as 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) are converted to a colored product by NAD(P)H-dependent cellular oxidoreductases, providing a quantifiable measure for metabolic activity.[6] An alternative approach to assess viability is the detection of intracellular adenosine triphosphate (ATP), which is maintained only at high levels in metabolically active cells and declines rapidly upon cell death or apoptosis. The release of intracellular ATP and its

[1]Department of Biology II, LMU Munich, Planegg-Martinsried, Germany

Received Aug 16, 2016, and in revised form Sep 27, 2016. Accepted for publication Oct 13, 2016.

Supplementary material is available online with this article.

Corresponding Author:
[1]Department of Biology II, LMU Munich, Grosshadernerstrasse 2, 82152 Planegg-Martinsried, Germany.
Email: helma@biologie.uni-münchen.de

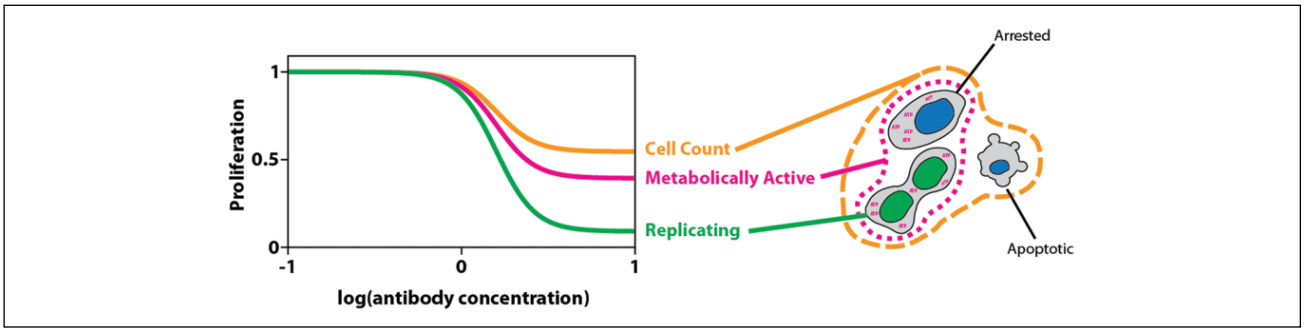


Figure 1. Addressing proliferation at different layers. Antiproliferative antibodies interfere with a cell's ability to replicate. Directly, detecting replicating cells (green) allows for the largest separation between maximal and minimal number of affected cells. Indirectly, restrained DNA replication also reduces the amount of metabolically active cells and the total number of cells remaining after treatment. However, the detection of metabolically active cells (magenta) includes arrested cells, resulting in an overestimation of proliferating cells. This effect is even more drastic when further generalizing the detection to all remaining cells (orange), which also includes apoptotic cells.

detection via ATP-dependent luciferase activity is widely used in proliferation assays.[7]

However, cells that have undergone cell cycle arrest are still metabolically active and consequently not distinguishable from proliferating cells by above-described assays. A major characteristic of proliferating cells is the replication of DNA during S phase. Thus, the incorporation of nucleotide analogues such as 5-bromo-2′-deoxyuridine (BrdU) into chromosomal DNA during replication allows for the distinction between proliferating and arrested cells. BrdU can be detected by antibodies and thus may be implemented with highly sensitive enzyme-linked immunosorbent assay (ELISA)–based multiwell assays.[8] It has been shown that a wider separation between signals from treated and untreated samples (signal-to-background [S/B] ratio) can be achieved with BrdU incorporation compared with assays detecting metabolic activity.[9] 5-Ethynyl-2′-deoxyuridine (EdU), an alternative nucleotide analogue, enables a simpler, milder, and more efficient detection via copper-catalyzed azide alkyne cycloaddition (CuAAC) of fluorescent dyes, such as 6-FAM–azide. The use of EdU coupled to fluorescent dyes simplifies the assay procedure and in addition improves compatibility with other nuclear stains such as 4′,6-diamidino-2-phenylindole (DAPI), thus representing the method of choice for sensitive microscopy-based detection of proliferation.

Accurate distinction between proliferating and nonproliferating cells improves the sensitivity of an antiproliferative potency assay (**Fig. 1**). Changing the mode of signal detection, on one hand, can further improve sensitivity but also provide additional information about the antiproliferative effect. Plate reader–based readouts are commonly used in screening experiments to validate lead candidates and produce statistically relevant data. Commonly used colorimetric multiwell proliferation assays are restricted to single-course parameters such as mean metabolic activity per well. To better understand the mode of action underlying an antiproliferative effect, cellular or subcellular information on signal localization and intensity may prove useful, which is usually not accessible with plate reader systems. Fluorescence microscopy is the method of choice to gain information about single cells with a variety of microscopic high-content screening (HCS) platforms developed in recent years that allow for automated image acquisition and analysis in a high-throughput manner.[10]

In the present study, we describe a simple and sensitive microscopic high-content assay for the quantification and characterization of the antiproliferative potency of therapeutic antibodies. The quantification of replicating cells, via EdU incorporation, as a measure for proliferation allows for most sensitive distinction between proliferating and nonproliferating cells. In addition to quantifying the antiproliferative potency of a monoclonal antibody (mAb), the mode of action can be investigated in the course of the same experiment. For example, potential induction of cell cycle arrest can be studied by cell cycle profiling based on nuclear DNA content quantification.

Materials and Methods

Cell Lines and Cell Culture

Antibodies were produced in FreeStyle HEK 293-F cells (Thermo Fisher Scientific, Waltham, MA, USA) cultured in FreeStyle 293 Expression Medium and maintained at cell densities from 3×10^5 to 3×10^6 cells/mL in a shaker flask at 37 °C, 5% CO_2, shaking at 120 rpm.

HER2 overexpression cell lines BT474 (ATCC HTB20) and SKBR3 (ATCC HTB30) and a control cell line with neglectable HER2 expression levels (1000-fold less than SKBR3), MDA-MB-468 (ATCC HTB-132), were cultured in Dulbecco's modified Eagle's medium (DMEM)/F12 + Gibco Glutamax-I (Thermo Fisher Scientific, Waltham,

MA, USA) supplemented with 10% fetal calf serum (FCS) at 37 °C, 5% CO_2.

Protein Expression and Purification

Trastuzumab was expressed in FreeStyle HEK 293-F cells as described previously from the pVITRO1-trastuzumab-IgG1/κ vector (Addgene plasmid 61883; Addgene, Cambridge, MA, USA).[11]

Antibody purification from cleared and sterile filtered cell culture supernatants was performed with an Äkta purifier system equipped with a 1-mL HiTrap Protein A HP column (GE Healthcare, Piscataway, NJ). The system was operated with a constant flow rate of 1 mL/min. After sample application, the column was washed with 10 column volumes (CVs) of wash buffer (20 mM phosphate buffer, 150 mM NaCl, pH 7.3). Bound antibody was eluted with a one-step pH decrease to 3.0 (10 mM Na-citrate buffer, pH 3.0). Eluted fractions of size 0.2 mL or 0.5 mL were collected followed by immediate neutralization of the pH with one-third volume 1 M Tris HCl, pH 8.0. Peak fractions were pooled and concentrated using an Amicon Ultra 4-mL Centrifugal Filter NMWL 10 kDa (Merck Millipore, Billerica, MA, USA) and stored at 4 °C or snap frozen in liquid nitrogen and transferred to −80 °C for long-term storage.

Antibody Treatment, EdU Incorporation, and Nuclear Staining

In total, 1×10^4 cells were seeded in each well of a 96-well optical cell culture plate supplemented with 100 µL culture media. To ensure proper attachment, cells were incubated for 4 h prior to addition of antibody. The 1:3 serial dilutions of trastuzumab in culture media were performed at threefold the desired final concentration, ranging from 50 nM to 0 nM. Then, 50 µL of each dilution was added in triplicates to individual wells. Cells were incubated with antibody for 4 days followed by the addition of EdU to a final concentration of 10 µM. To guarantee labeling of all proliferating cells, EdU treatment was done for 20 h followed by fixation of cells in phosphate-buffered saline (PBS) + 4% paraformaldehyde (PFA), permeabilization in PBS + 0.5% Triton X-100, and blocking of the well surface with PBT (PBS, 2% BSA, and 0.02% Tween 20). EdU was labeled via CuAAC by the addition of 30 µL of staining reagent (4 mM $CuSO_4$, 20 µM 6-FAM–azide, 50 µM Na-ascorbate in 100 mM Tris/HCl, pH 7.0) per well and incubated for 30 min at room temperature. Remaining unconjugated dye was removed by washing three times with 100 µL PBST (PBS + 0.02% Tween 20). Then, 100 ng/mL DAPI in PBST was added for 10 min at room temperature to counterstain nuclear DNA, followed by three washing steps with PBST and one additional wash with ddH_2O.

Image Acquisition and Data Analysis

Images were acquired with an Operetta High-Content Imaging system (PerkinElmer, Waltham, MA, USA) equipped with a 40× high NA objective. The 380/40-nm excitation and 410- to 480-nm emission filters were used to image DAPI, and the 475/30-nm excitation and 500- to 550-nm emission filters were used to image 6-FAM–EdU.

DAPI images were used to segment and count the total number of nuclei for each well, representing the total cell count. Each antibody concentration was tested in technical triplicates. Total cell counts of triplicates were averaged and normalized to the cell count of an untreated control (c(trastuzumab) = 0). Averaged and normalized cell counts were plotted against $\log_{10}$-transformed antibody concentrations. Fitting a nonlinear four-parametric model equation $y = \min + \frac{\max - \min}{1 + \left(\frac{IC_{50}}{x}\right)\text{Hill slope}}$ to the normalized cell counts y and antibody concentration x yielded inhibition curves with the descriptive parameters IC_{50} (concentration of half-maximal inhibition) and Hill slope.

Relative nucleic DNA quantities were obtained by calculating total DAPI intensities of segmented nuclei. Absolute DAPI intensities per nucleus were subgrouped by binning and plotted as a probability histogram to analyze probability distributions. A 10-parameter model function $p_x = P(x) = G_1(x) + G_2(x) + S(x)$, comprising the sum of two Gaussian terms $G_1(x) = a_{G1} * exp(-\frac{1}{2}\frac{(x-\mu_{G1})^2}{\sigma_{G1}^2})$ and $G_2(x) = a_{G2} * exp(-\frac{1}{2}\frac{(x-\mu_{G2})^2}{\sigma_{G2}^2})$ representing G1 and G2/M cell cycle phases, and a constant term with Gaussian fadeout,

$$S(x) = \begin{cases} h * exp(-\frac{1}{2}\frac{(x-x_{lower})^2}{\sigma_S^2}) & x < x_{lower} \\ h * exp(-\frac{1}{2}\frac{(x-x_{upper})^2}{\sigma_S^2}) & x > x_{upper} \\ h & else \end{cases},$$

modeling S phase, were fitted to the DAPI intensity probability densities p_x and histogram bin centers x to model the DNA content distribution throughout the cell cycle. The function was fitted by globally minimizing the squared error via simulated annealing using the GenSA package in R. By integrating over the respective term of the derived fit equation representing the G1, S, or G2/M phase, the relative proportion of each phase of the whole cell population was calculated—for example,

$$P(\mathrm{G1}) = \frac{\int_{-\infty}^{\infty} G_1(x)}{\int_{-\infty}^{\infty} P(x)}.$$

Based on 6-FAM–EdU signal, nuclei were classified as proliferating or nonproliferating. Data averaging, normalization, and curve fitting were done in a similar manner as described above for total cell counts.

All image processing was performed with the Harmony software (PerkinElmer); data analysis and curve fitting were done in MATLAB and R (2016, https://www.R-project.org). The R script used for the estimation of cell cycle distributions from DAPI intensity distributions is available at https://github.com/hoerldavid/CellCycleFit.

Results and Discussion

In the field of biologics, therapeutic antibodies have emerged as an especially promising drug format over the past years.[2] A role model for this class of drugs is trastuzumab, which binds the extracellular domain of the HER2 cell surface receptor. In a subset of breast cancers, the growth factor receptor HER2 is overexpressed and mediates increased proliferation.[12] Trastuzumab counteracts this accelerated growth by reducing HER2-mediated signaling and therefore acting as an antiproliferative drug on HER2-overexpressing cells.[13] To assess the antiproliferative potency of a therapeutic antibody, cells are subjected to a range of antibody concentrations. Higher antibody concentrations are expected to lead to lower numbers of viable cells and an even more pronounced decrease in proliferating cells (**Fig. 1**).

In the described assay, HER2-overexpressing cells (BT474 and SKBR3) and control cells (MDA-MB-468) were supplemented with EdU after 4 days of trastuzumab treatment. The proliferating fraction of the cell population incorporates EdU molecules into newly synthesized DNA during S phase. Surviving cells are stained with DAPI, whereas the incorporated EdU is labeled by CuAAC-mediated coupling of the fluorescent dye 6-FAM–azide. Imaging of stained cells on an Operetta system facilitates the detection and segmentation of nuclei, DNA content analysis using the DAPI signal, and definition of the proliferation status according to the EdU signal. Testing multiple antibodies over a range of concentrations is conveniently done in a multiwell tissue culture plate, which is compatible with the Operetta HCS imaging system. With this setup, an inhibition curve with 10 data points as technical triplicates can easily be generated for two individual antibodies in a 96-well format. Quantification of counted nuclei and detected proliferating cells can readily be done with the built-in software package of the Operetta system (Harmony), whereas statistical analysis and curve fitting are conveniently handled with respective MATLAB toolboxes. Besides the quantification of total cell counts and proliferating cells, the relative intensities of the DAPI and/or EdU signal per nucleus provide additional information with regard to cell cycle phase distributions.

Cell Survival and Cell Cycle Progression

Treatment of HER2-overexpressing cell lines with trastuzumab leads to a reduction in cell growth, but BT474 cells have been reported to be more susceptible than SKBR3 cells.[14] After 4 days of treatment, fluorescence microscopy of DAPI-stained nuclei indicates a clear reduction in cells with increasing concentrations of trastuzumab for BT474 (**Fig. 2A**) as well as SKBR3 cells. Next, we performed high-content image analysis by nuclei segmentation and subsequent quantification of surviving cells as a function of antibody concentration. By fitting a four-parametric nonlinear model to the obtained data points, we calculated inhibition curves. These fits revealed a decrease in total cell number with increasing antibody concentration and S/B ratios lower than 3 for BT474 (**Fig. 2B**) and SKBR3 (**Fig. 2C**). The maximal induction of cell death is 64% with a concentration of half maximal inhibition (IC_{50}) of 1.8 nM for BT474 cells and 65% with an IC_{50} value of 1.9 nM for SKBR3 cells. The low S/B values can be explained by the specific mode of action mediated by trastuzumab, decelerating cell proliferation rather than actively promoting cell death.[14] Therefore, cells that have already passed G1 phase will further progress in cell cycle. With BT474 and SKBR3 cells exhibiting long doubling times (2–3 days), S/B ratios greater than 4 (two doublings) are not to be expected in the time course of the assay, which holds also true for other assays merely detecting survival or viability.[9] Moreover, a very low Hill slope could be observed for SKBR3 cells compared with BT474, which is linked to the lower susceptibility of SKBR3 to trastuzumab.[9,14] Consistently, an unsusceptible cell line (MDA-MB-468) showed no difference in the number of viable cells between treated and untreated conditions (**Fig. 2B**,**C**). These results indicate that exclusively measuring cell survival is limiting the S/B ratio of proliferation assays, since arrested cells, which are still metabolically active, cannot be distinguished from proliferating cells.

High-content image analysis of DAPI-stained nuclei allows not only segmentation and quantification of nuclei but also the measurement of relative nuclear DNA contents. Since the amount of chromosomal DNA doubles through S phase from G1 to G2 phase, the absolute DAPI signal per nucleus can be used to analyze changes in cell cycle distributions. In this line, we generated frequency histograms of the absolute DAPI intensity per nucleus (**Fig. 3A** and **Suppl. Fig. S1**). Fitting a three-term model function to the data allowed us to determine the proportion of cells within each cell cycle phase (**Fig. 3B**). SKBR3 cells exhibited a clear change in cell cycle

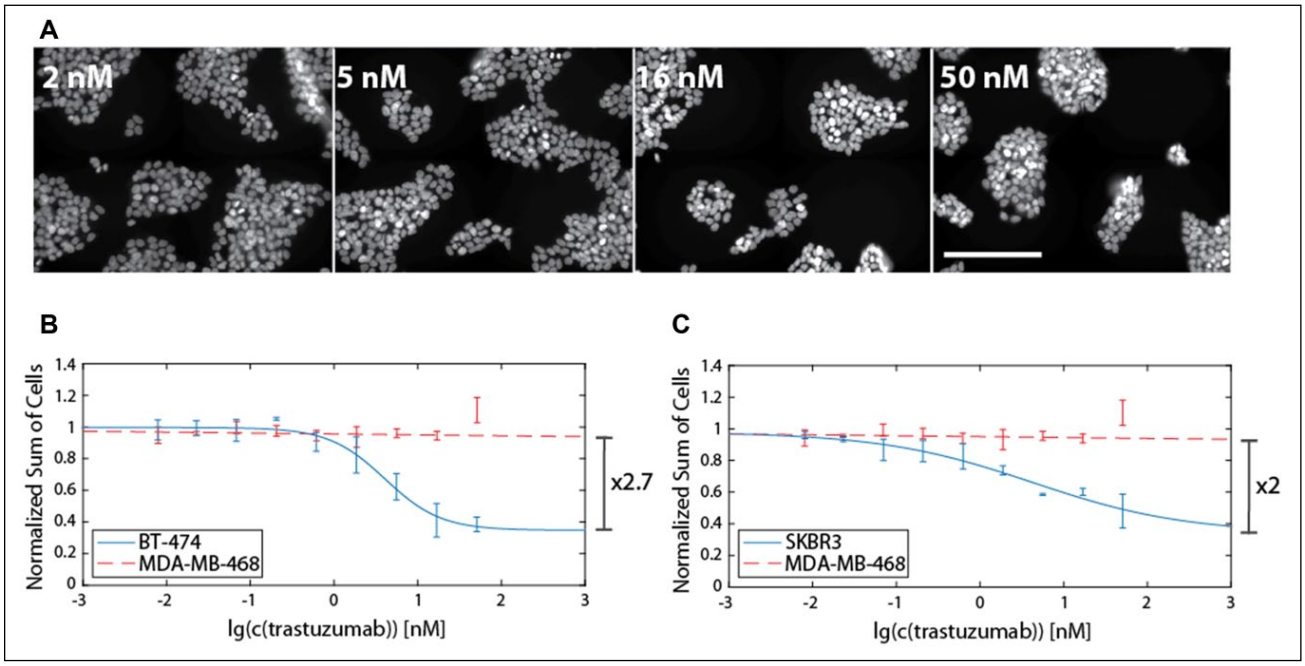


Figure 2. Quantification of antiproliferative potency by counting nuclei of surviving cells. 4′,6-Diamidino-2-phenylindole (DAPI)–stained nuclei were imaged with an Operetta high-content screening (HCS) system. Representative images of BT474 cells for four different antibody concentrations are shown in (**A**). Scale bar represents 100 µm. The observed decrease in surviving cells was quantified from technical triplicates for nine individual antibody concentrations (0.008–50 nM) and an untreated control. Averaged triplicates normalized to untreated control were plotted against $\log_{10}$-transformed trastuzumab concentrations for BT474 (**B**) and SKBR3 (**C**) and fitted to a four-parametric inhibition curve model equation (solid lines). Proliferation of a negative control cell line, MDA-MB-468, was unaffected by trastuzumab treatment (dashed line). The maximal difference in the number of surviving cells was 2.7-fold for BT474 as well as for SKBR3 cells.

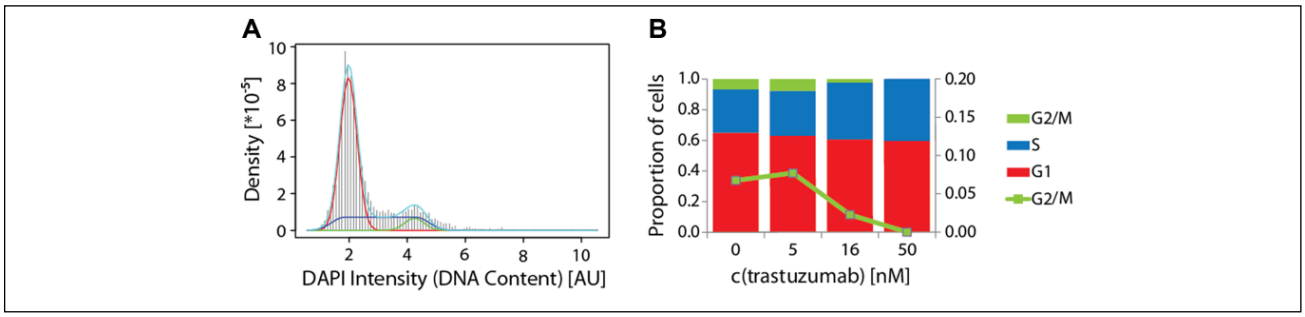


Figure 3. Shift in cell cycle distribution of trastuzumab-treated SKBR3 cells. Nuclear 4′,6-diamidino-2-phenylindole (DAPI) intensities were analyzed to categorize cells into cell cycle phases according to their relative DNA content. Probability density histograms of DAPI intensities were used to fit a model equation to the observed distribution. An exemplary histogram for c(trastuzumab) = 16 nM is given in (**A**) with the fitted curve in cyan and respective cell cycle phase terms in red (G1), blue (S), and green (G2/M). Integration over the individual terms yields the proportion of cells in each cell cycle phase treated with different trastuzumab concentrations (**B**). High concentrations of trastuzumab lead to a reduction in the G2/M phase proportion, indicating cell cycle arrest.

profiles upon trastuzumab treatment. The quantification of these data shows a decrease in the G2 phase population with increasing antibody concentration, which suggests an arrest in either G1 or S phase. This is consistent with the proposed G1 arrest induced by trastuzumab.[15]

Cell cycle profiles are an additional readout of the described assay and provide supplementary information about the mode of action of an antiproliferative antibody. Investigation of potency and mode of action in a single experiment was facilitated by increasing resolution to the single-cell level combined with high-throughput sample and data handling implemented in HCS systems. Cell cycle analysis of the less susceptible SKBR3 cell line showed that we are able to analyze an antibody's mode of action even if the overall

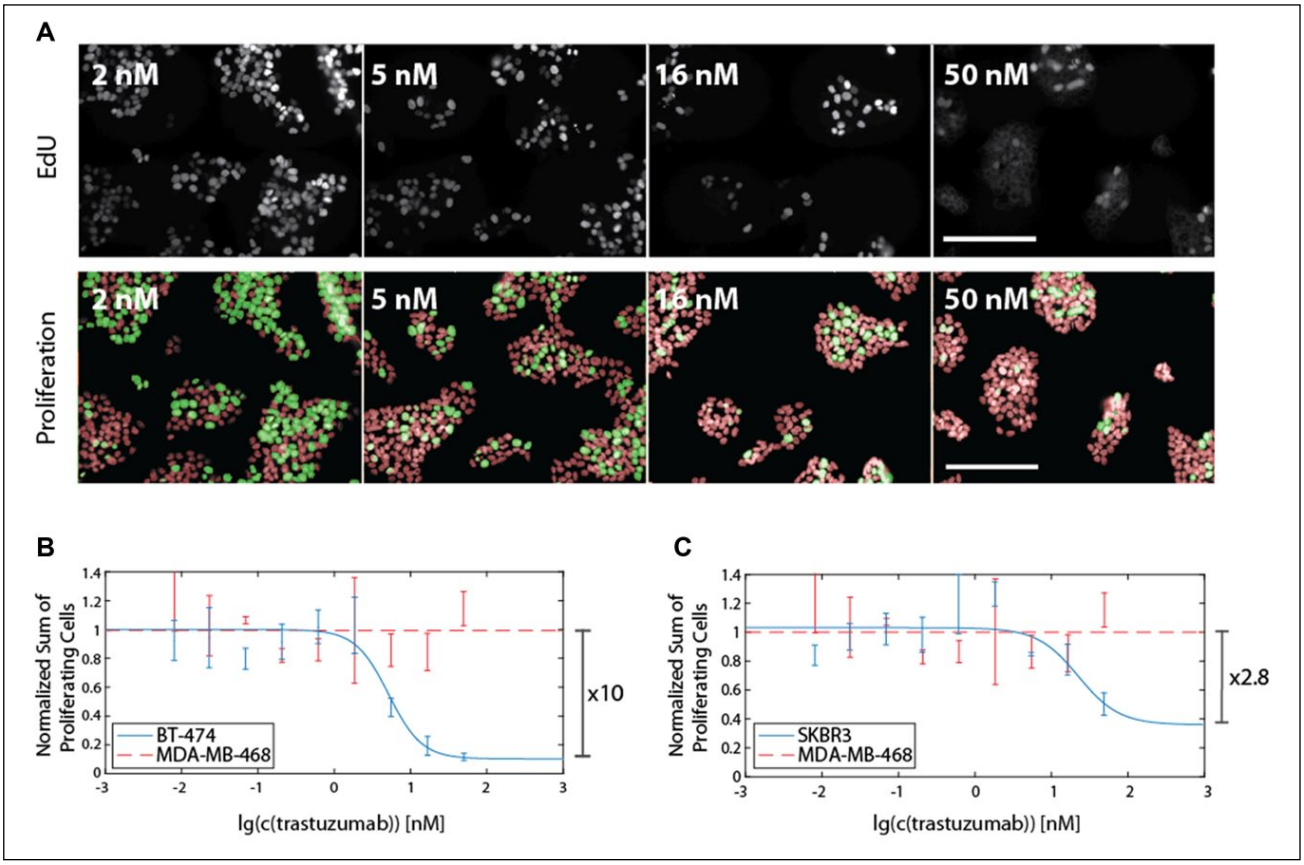


Figure 4. Improving assay sensitivity by detecting proliferating cells via 5-ethynyl-2′-deoxyuridine (EdU) incorporation. EdU, incorporated into chromosomal DNA during replication, was labeled by copper-catalyzed azide alkyne cycloaddition (CuAAC) with 6-FAM and imaged with an Operetta high-content screening (HCS) system. Representative images of BT474 cells are shown in (**A**). Scale bar represents 100 µm. Segmented nuclei from **Figure 2A** were classified as proliferating (green) or nonproliferating (red) based on EdU signal presence. It is clearly visible that only a small fraction of all surviving cells is still proliferating at high antibody concentrations. Results of quantification of proliferating cells and data fitting similar to data in **Figure 2** are shown for BT474 cells (**B**) and SKBR3 cells (**C**). The signal to background (S/B) ratio could be greatly improved for BT474 cells from 2.7 to 10 compared with surviving cell quantification (**Fig. 2**). SKBR3 cells exhibit an S/B ratio of 2.8, which is comparable to the value derived from 4′,6-diamidino-2-phenylindole (DAPI)–based quantification of surviving cells (2.7).

antiproliferative effect is weak. Nevertheless, it is also desirable to detect this weak proliferation inhibition with greater resolution. To address this need, we chose EdU incorporation for sensitive detection of proliferating cells.

Increased Assay Sensitivity via Quantification of EdU Incorporating Cells

Since DNA replication is a major characteristic of proliferation, we decided to use EdU incorporation as a marker for proliferating cells. Labeling EdU with a fluorescent dye allowed the distinction between proliferating and nonproliferating cells by fluorescence microscopy. Automated quantification of EdU-positive cells increased the S/B ratio to 10 for treated versus untreated BT474 cells (**Fig. 4B**). A concentration of half maximal inhibition (IC_{50}) of 4.9 nM was obtained from the fitted inhibition curve, whereas the maximal induction of proliferation inhibition was 90%. For SKBR3 cells, we observed a maximal induction of proliferation inhibition of 64% and IC_{50} of 3.9 nM. To ensure that the detected inhibition of proliferation was due to trastuzumab-mediated effects, we subjected a control cell line, MDA-MB-468, to the same treatment. As expected, we could not observe any difference in the proliferating fraction upon addition of trastuzumab (**Fig. 4B**,**C**). We could show that EdU incorporation-based detection of proliferating cells by microscopy greatly increases the S/B ratio compared with detecting surviving cells and improves the inhibition curve parameters such as Hill slope in the case of SKBR3 (**Fig. 4C**). A 10-fold change in proliferation has recently also been demonstrated with a DELFIA-BrdU–based assay.[9] However, the assay described in the present article uses the more sensitive and mild EdU staining method, provides the possibility for multiplexed readout of

various parameters, and increases the assay resolution by the detection of single cells instead of averaging over a bulk population.

In summary, we could show that EdU-based labeling of proliferating cells with subsequent automated imaging and analysis combined with DAPI-based cell cycle profiling is a simple and sensitive way for parallel investigation of anti-proliferative potency and mode of action of therapeutic antibodies.

Acknowledgments

We thank Dr. Shane Miersch for providing cell lines and advice on assay setup.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a grant from the Priority Program SPP1623 of the Deutsche Forschungsgemeinschaft by H. L. A. S. was trained and supported by the graduate school GRK1721 of the Deutsche Forschungsgemeinschaft as an associate member.

References

1. Schumacher, D.; Hackenberger, C. P.; Leonhardt, H.; et al. Current Status: Site-Specific Antibody Drug Conjugates. *J. Clin. Immunol*. **2016**, *36*(Suppl 1), 100–107.
2. Shi, S. Biologics: An Update and Challenge of Their Pharmacokinetics. *Curr. Drug Metab*. **2014**, *15*, 271–290.
3. Crombet-Ramos, T.; Rak, J.; Perez, R.; et al. Antiproliferative, Antiangiogenic and Proapoptotic Activity of h-R3: A Humanized Anti-EGFR Antibody. *Int. J. Cancer* **2002**, *101*, 567–575.
4. Vega-Avila, E.; Pugsley, M. K. An Overview of Colorimetric Assay Methods Used to Assess Survival or Proliferation of Mammalian Cells. *Proc. West Pharmacol. Soc*. **2011**, *54*, 10–14.
5. Dehlinger, D.; Suer, L.; Elsheikh, M.; et al. Dye Free Automated Cell Counting and Analysis. *Biotechnol. Bioeng*. **2013**, *110*, 838–847.
6. Mosmann, T. Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays. *J. Immunol. Methods* **1983**, *65*, 55–63.
7. Crouch, S. P.; Kozlowski, R.; Slater, K. J.; et al. The Use of ATP Bioluminescence as a Measure of Cell Proliferation and Cytotoxicity. *J. Immunol. Methods* **1993**, *160*, 81–88.
8. Porstmann, T.; Ternynck, T.; Avrameas, S. Quantitation of 5-Bromo-2-Deoxyuridine Incorporation into DNA: An Enzyme Immunoassay for the Assessment of the Lymphoid Cell Proliferative Response. *J. Immunol. Methods* **1985**, *82*, 169–179.
9. Lu, X.; Bergelson, S. Development of a Sensitive Potency Assay to Measure the Anti-Proliferation Effect of an Anti-HER2 Antibody. *J. Immunol. Methods* **2014**, *415*, 80–85.
10. Fraietta, I.; Gasparri, F. The Development of High-Content Screening (HCS) Technology and Its Importance to Drug Discovery. *Expert Opin. Drug Discov*. **2016**, *11*, 501–514.
11. Dodev, T. S.; Karagiannis, P.; Gilbert, A. E.; et al. A Tool Kit for Rapid Cloning and Expression of Recombinant Antibodies. *Sci. Rep*. **2014**, *4*, 5885.
12. Browne, B. C.; O'Brien, N.; Duffy, M. J.; et al. HER-2 Signaling and Inhibition in Breast Cancer. *Curr. Cancer Drug Targets* **2009**, *9*, 419–438.
13. Vu, T.; Claret, F. X. Trastuzumab: Updated Mechanisms of Action and Resistance in Breast Cancer. *Front. Oncol*. **2012**, *2*, 62.
14. Brockhoff, G.; Heckel, B.; Schmidt-Bruecken, E.; et al. Differential Impact of Cetuximab, Pertuzumab and Trastuzumab on BT474 and SK-BR-3 Breast Cancer Cell Proliferation. *Cell Prolif*. **2007**, *40*, 488–507.
15. Lane, H. A.; Motoyama, A. B.; Beuvink, I.; et al. Modulation of p27/Cdk2 Complex Formation through 4D5-Mediated Inhibition of HER2 Receptor Signaling. *Ann. Oncol*. **2001**, *12*(Suppl 1), S21–S22.

2.4.2 Supplementary Material

Supplementary Information

A Simple and Sensitive High Content Assay for the Characterization of Anti-Proliferative Therapeutic Antibodies

Andreas Stengl[1], David Hörl[1], Heinrich Leonhardt[1], Jonas Helma[1]*

[1] Department of Biology II, LMU Munich, Grosshadernerstrasse 2, 82152 Planegg-Martinsried, Germany

* Corresponding author: helma@biologie.uni-münchen.de

Keywords: therapeutic antibodies, cell-based, high-content assay, EdU, proliferation

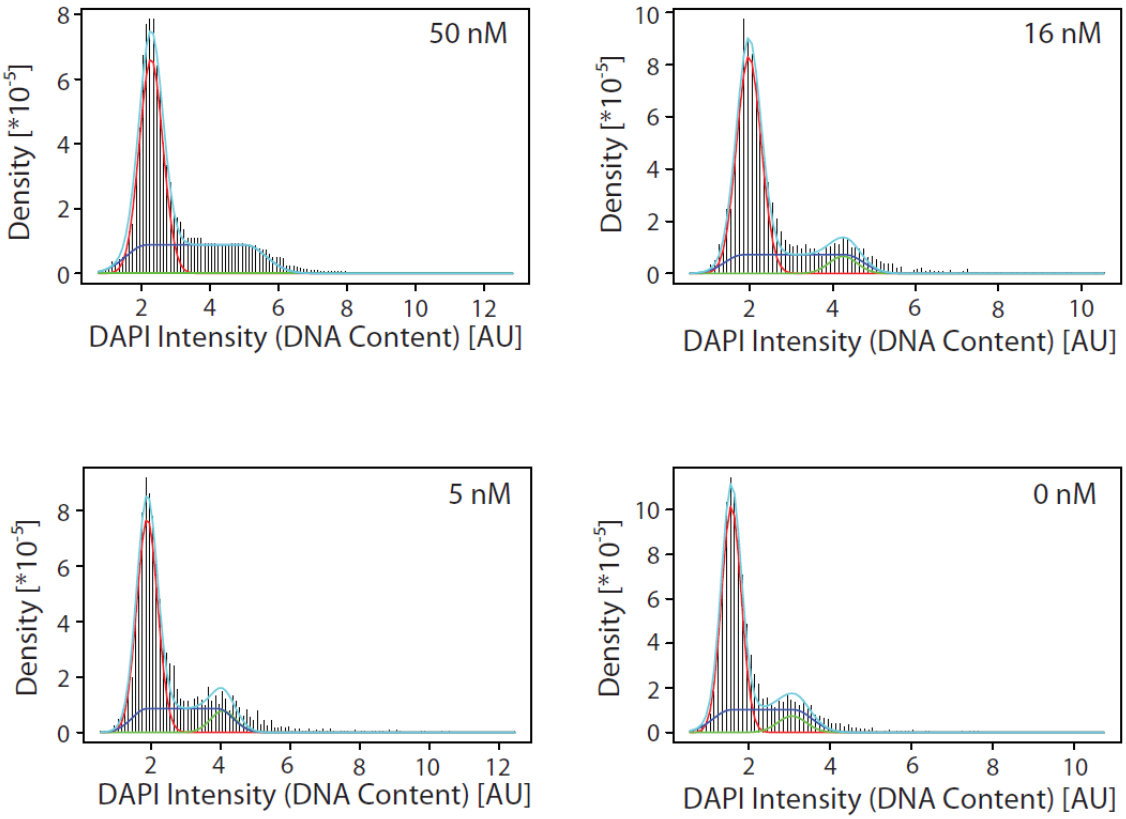


Figure S1 Histograms of total nuclear DAPI intensities and fitted cell cycle phase curves. Total DAPI intensities per nucleus were calculated for SKBR3 cells and plotted as probability density histograms at four different trastuzumab concentrations. A prominent peak for G1 phase was observed, a plateau representing S phase and a smaller peak at approximately two times the DNA content of G1 phase cells, consisting of cells in G2/M phase. Solid line graphs represent either the G1 phase term (red), S phase term (blue) or G2/M phase term (green) of the resulting fit (cyan). Integration over the individual terms yielded relative quantities for each cell cycle phase (see Figure 2).

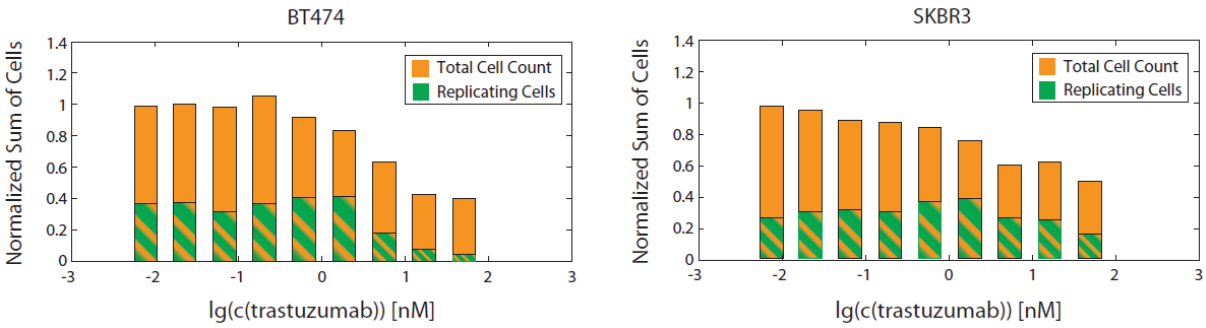


Figure S2 Quantification of proliferating cells improves assay sensitivity compared to counting total surviving cells. Total cell count (orange) and proliferating cell fraction (green) are depicted for different trastuzumab concentrations. With increasing antibody concentration the number of cells surviving treatment is decreased for BT474 as well as for SKBR3 cells. Even more pronounced is the decrease in proliferating cells of the surviving cell population illustrating the increased sensitivity demonstrated with the described assay. These data reflect the assay schematic given in Figure 1.

2.5 DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination

2.5.1 Main Paper

Open
ORIGINAL ARTICLE

Cell Research (2015) 25:911-929.
www.nature.com/cr

DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination

Weihua Qin[1, 3, 4, *], Patricia Wolf[1, 3, 4, *], Nanli Liu[1, 3, 4], Stephanie Link[1, 3, 4], Martha Smets[1, 3, 4], Federica La Mastra[1, 3, 5], Ignasi Forné[2, 3], Garwin Pichler[1, 3, 6], David Hörl[1, 3, 4], Karin Fellinger[1, 3, 7], Fabio Spada[1, 3, 8], Ianmarco Spacek[5], Axel Imhof[2, 3], Hartmann Harz[1, 3, 4], Heinrich Leonhardt[1, 3, 4]

[1]*Department of Biology II, Ludwig Maximilians University Munich, Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany;* [2]*Adolf-Butenandt Institute, Ludwig Maximilians University Munich, Schillerstr. 44, 80336 Munich, Germany;* [3]*Center for Integrated Protein Science Munich (CIPSM),* [4]*Nanosystems Initiative Munich (NIM),* [5]*Department of Theoretical and Applied Sciences, University of Insubria, Via Manara 7, 21052 Busto Arsizio (VA), Italy*

DNMT1 is recruited by PCNA and UHRF1 to maintain DNA methylation after replication. UHRF1 recognizes hemimethylated DNA substrates via the SRA domain, but also repressive H3K9me3 histone marks with its TTD. With systematic mutagenesis and functional assays, we could show that chromatin binding further involved UHRF1 PHD binding to unmodified H3R2. These complementation assays clearly demonstrated that the ubiquitin ligase activity of the UHRF1 RING domain is required for maintenance DNA methylation. Mass spectrometry of UHRF1-deficient cells revealed H3K18 as a novel ubiquitination target of UHRF1 in mammalian cells. With bioinformatics and mutational analyses, we identified a ubiquitin interacting motif (UIM) in the N-terminal regulatory domain of DNMT1 that binds to ubiquitinated H3 tails and is essential for DNA methylation *in vivo*. H3 ubiquitination and subsequent DNA methylation required UHRF1 PHD binding to H3R2. These results show the manifold regulatory mechanisms controlling DNMT1 activity that require the reading and writing of epigenetic marks by UHRF1 and illustrate the multifaceted interplay between DNA and histone modifications. The identification and functional characterization of the DNMT1 UIM suggests a novel regulatory principle and we speculate that histone H2AK119 ubiquitination might also lead to UIM-dependent recruitment of DNMT1 and DNA methylation beyond classic maintenance.

Keywords:UHRF1; histone ubiquitination; DNMT1; DNA methylation
Cell Research( ) **25**:9 1 2 :0 cr 2 published online 2 ne

Introduction

Epigenetic mechanisms including DNA and histone modifications are crucial for the regulation of gene expression during development. DNA methylation occurs at the C5 position of cytosine residues, mostly within cytosine-guanine dinucleotides (CpG), and is involved in imprinting, X-chromosome inactivation, stable transcriptional repression, genome stability and tumorigenesis [ ]. DNA methylation patterns are established by the *de novo* methyltransferases DNMT3A and DNMT3B during gametogenesis and early development, and are propagated by the maintenance methyltransferase DNMT1 after DNA replication in somatic cells.

DNMT1 comprises a regulatory N-terminal domain (NTD), which covers two thirds of the molecule, and a C-terminal catalytic domain (CD), which contains all essential motifs of active C5 DNA methyltransferases. The NTD controls the subcellular distribution of DNMT1 during the cell cycle and its enzymatic activity. A subdomain in the NTD was initially described as a targeting sequence (TS) as it was found to mediate the associa-

The set work is the scientific responsibility of the work
Correspondence: Heinrich Leonhardt
E-mail: heinrich.leonhardt@lmu.de
Tel: + Fax +

[6]Current address: Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany

[7]Current address: Intervet International GmbH, Unterschleissheim, Germany

[8]Current address: Department of Chemistry, Ludwig Maximilians University Munich, Germany

Received 21 November 2014; revised 27 March 2015; accepted 7 May ; published online 2 ne

tion of DNMT1 with late replicating pericentromeric heterochromatin [2]. Subsequent studies defined a distinct proliferating cell nuclear antigen (PCNA) binding domain (PBD) responsible for the interaction with the replication machinery [3]. The subnuclear localization of DNMT1 undergoes characteristic changes throughout the cell cycle reflecting PBD-mediated PCNA binding during S phase and TS domain-mediated heterochromatin association during late S and G2 phase [4, 5]. The association of DNMT1 with the replication machinery enhances methylation efficiency, but is not strictly required for postreplicative maintenance DNA methylation [6, 7]. In contrast, the TS domain was found to be required for DNMT1 enzymatic activity [8, 9]. However, the molecular mechanism of TS domain function in the regulation of maintenanceDNAmethyltransferase is unknown.

Besides its role in replication-independent heterochromatin binding, the TS domain mediates DNMT1 homodimerization [9] and autoinhibition [10, 11]. A recent crystal structure shows that the TS domain inserts into the DNA binding pocket of the CD, indicating a role of intramolecular interactions in the regulation of DNMT1 activity [10, 11]. Moreover, the TS domain interacts with the SET- and RING-associated (SRA) domain of ubiquitin like, containing PHD and RING finger domains 1 (UHRF1) [12–14]. In contrast to UHRF2, the interaction of UHRF1 with DNMT1 was found to beS phase-dependent[15].

UHRF1, also known as NP95 (mouse) or ICBP90 (human), has been reported as a crucial cofactor for maintenance DNA methylation. Mice lacking UHRF1 show a similar phenotype as *Dnmt1* null (*Dnmt1*$^{-/-}$) mice that manifests in genomic DNA hypomethylation and developmental arrest at embryonic day 9 [16–18]. The SRA domain of UHRF1 preferentially binds to hemimethylated DNA resulting from semiconservative DNA replication and is, therefore, thought to play an important role in loading DNMT1 onto newly synthesized DNA substrates [16, 17, 19–22]. The heterochromatin association of UHRF1 is also mediated by the tandem Tudor domain (TTD), which forms an aromatic cage for specific binding of histone H3 tails containing a trimethylated lysine 9 (H3K9me3) residue [23–25]. The plant homeodomain (PHD) was reported to act in combination with the TTD to read the H3K9me3 mark [26] and to contribute to large-scale reorganization of pericentromeric heterochromatin [27]. In addition, UHRF1 harbors a really interesting new gene (RING) domain endowed with ubiquitin E3 ligase activity *in vitro*, which is required for growth regulation of tumor cells [28–30]. The ubiquitination state and stability of DNMT1 is controlled by UHRF1 and the ubiquitin-specific protease USP7 [29, 30]. UHRF1 overexpression leads to DNA hypermethylation by the destabilization and delocalization of DNMT1 [31]. Besides its role in marking DNMT1 for proteasomal degradation, UHRF1 also exerts its ubiquitin E3 ligase activity on histone substrates[28–30].

A recent study describes replication-dependent H3K23 ubiquitination by UHRF1 in *Xenopus* extracts [32]. Knockdown and rescue experiments in HeLa cells showed that SRA domain-mediated DNA binding as well as RING domain-dependent E3 ubiquitin ligase activity of UHRF1 are required for H3 ubiquitination. Expression of the SRA and RING domain mutants in *Uhrf1*$^{-/-}$ mouse cells could neither restore DNMT1 replication targeting nor DNA methylation levels. A deletion of large parts of the DNMT1 TS domain abolished binding to ubiquitinated H3K23 *in vitro*, but effects on enzymatic activity were not investigated. In particular, the structure and function of the rather large TS domain with its multiple roles and interactions remain to be clarified.

In this study, we elucidate the complex interplay between UHRF1 and DNMT1. While we could confirm the general role of UHRF1 in recruiting DNMT1 to substrate sites by direct interaction, we found that DNMT1 targeting and activities are essentially controlled by specific binding to histone tails ubiquitinated by UHRF1. We generated defined mutations in different UHRF1 domains that retained SRA domain-mediated binding to hemimethylated DNA substrate sites, TTD-mediated recognition of H3K9me3 and binding of DNMT1, but did not allow maintenance DNA methylation. We could show that binding to unmodified H3R2 via the PHD and ubiquitination of H3K18 via the RING domain are required for UHRF1 to mediate maintenance DNA methylation. In turn, we identified a ubiquitin interacting motif (UIM) in the TS domain of DNMT1 that reads this ubiquitin mark and is strictly required for maintenance DNA methylation *in vivo*. These results show the manifold regulatory mechanisms controlling DNMT1 activity and illustrate the multifaceted interplay between DNA and histone modifications.

Results

The interaction of DNMT1 with UHRF1 is required for maintenance DNA methylation

To test whether the interaction of DNMT1 with UHRF1 is indeed required for maintenance DNA methylation, we generated stable cell lines based on *Dnmt1*$^{-/-}$ ESCs expressing green fluorescent protein (GFP) fusions of either DNMT1 wild-type (GFP-DNMT1 wt) or a truncated TS domain deletion mutant (GFP-DNMT1 Δ458-500) that is defective in binding to UHRF1 (Figure

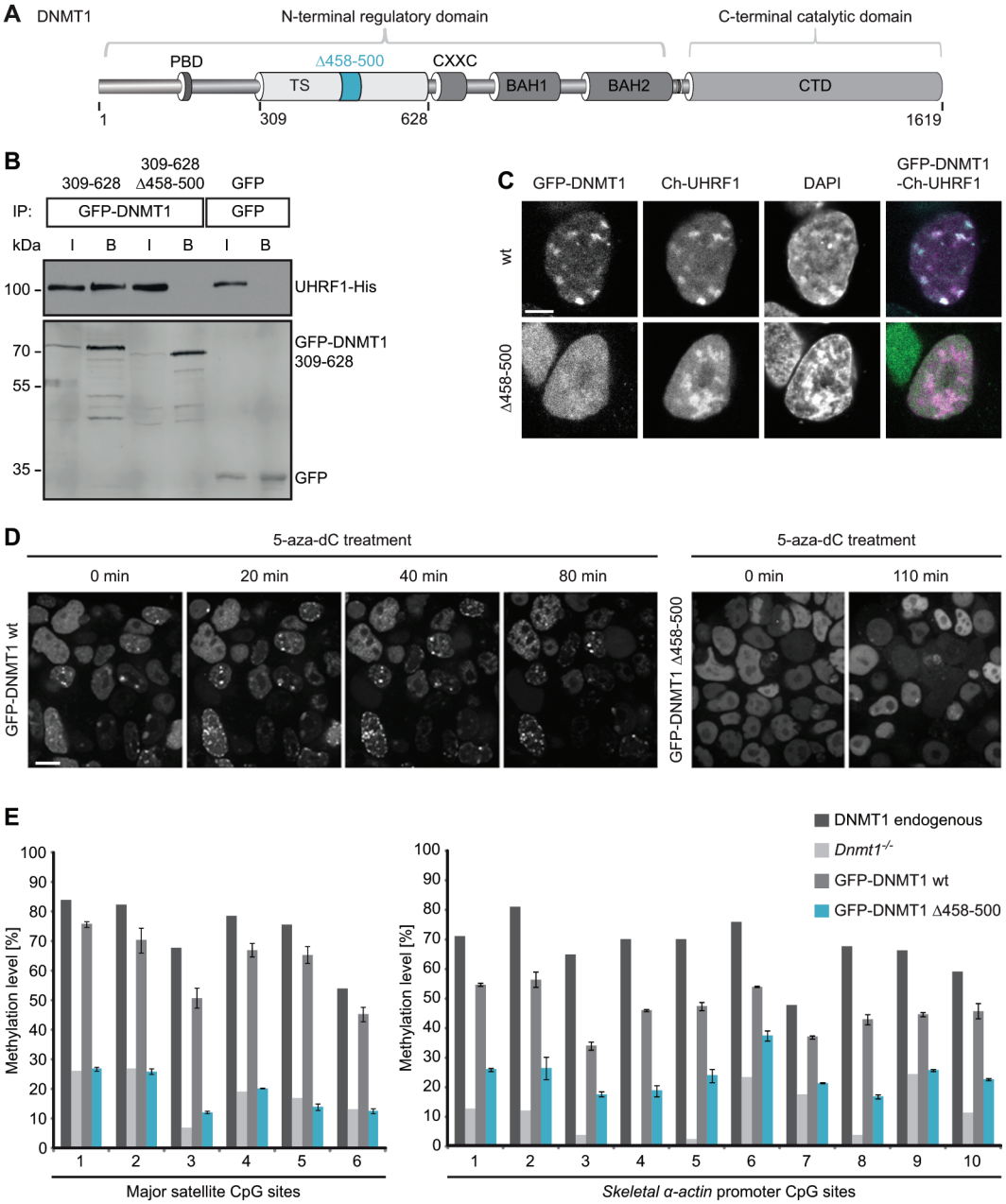


Figure 1 The DNMT1 TS domain is required for UHRF1 interaction, heterochromatin targeting and maintenance DNA methylation. **(A)** Schematic outline of DNMT1 domains and the TS domain deletion (Δ458-500). DNMT1 comprises a large N-terminal domain (NTD) harboring the PCNA binding domain (PBD), the targeting sequence (TS) domain and two bromo adjacent homology (BAH) domains. The active catalytic center of DNMT1 resides within its C-terminal domain (CD). **(B)** Co-immunoprecipitation of UHRF1-His and the GFP-DNMT1 TS domain (309-628) wild-type (wt) or Δ458-500 constructs. Both constructs were co-expressed in HEK 293T cells and after immunoprecipitation of GFP fusions, bound proteins were detected by western blot with an anti-UHRF1 and an anti-GFP antibody. GFP was used as negative control. I, input; B, bound. **(C)** Confocal mid sections of fixed ESCs stably expressing GFP-DNMT1 wt or Δ458-500 mutant constructs. Ch-UHRF1 was transiently co-expressed to illustrate heterochromatic regions, DAPI was used for counterstaining. Scale bar, 5 μm. **(D)** Covalent complex formation of GFP-DNMT1 wt and GFP-DNMT1 Δ458-500 mutant were analyzed by an *in vivo* trapping assay. Confocal mid-sections of ESCs stably expressing GFP-DNMT1 wt and deletion mutant constructs before and after treatment with the mechanism-based inhibitor 5-aza-dC are displayed. Scale bar, 10 μm. **(E)** Local DNA methylation analyses at the major satellite repeats and the *skeletal α-actin* promoter. CpG methylation levels of mouse *Dnmt1*$^{-/-}$ ESCs stably expressing GFP-DNMT1 wt or GFP-DNMT1 Δ458-500 mutant constructs were analyzed by bisulfite treatment of genomic DNA, PCR amplification and direct pyrosequencing. The methylation level of the J1 wt cell line (endogenous DNMT1) and untransfected *Dnmt1*$^{-/-}$ cells are shown for comparison. Mean values ± SD from two different clones were calculated, respectively.

A and B). The deleted region was determined by a sequence alignment of TS domains from higher eukaryotes and a conserved core region of the domain was chosen for mutational analysis (Supplementary information, Figure S1A). In contrast to GFP-DNMT1 wt, GFP-DNMT1 Δ458-500 did not co-localize with cherry (Ch)-UHRF1 and showed a dispersed distribution in the nucleus (Figure 1C), suggesting that the interaction with UHRF1 is essential for subnuclear localization of DNMT1.

Next, we investigated the role of UHRF1 interaction for the catalytic function of DNMT1. Notably, GFP-DNMT1 Δ458-500 that did not interact with UHRF1 was able to fully methylate hemimethylated DNA substrates *in vitro* (Supplementary information, Figure S1C). To test the DNA methylation activity of the deletion mutant *in vivo*, we made use of a trapping assay. In this assay, the cytosine analogue 5-aza-2′-deoxycytidine (5-aza-dC) forms an irreversible covalent complex with the methyltransferase at the C6 position of the cytosine residue when incorporated into DNA during replication thereby trapping the enzyme at DNA replication foci. Trapped DNMT1 fractions increase over time and allow monitoring the activity-dependent accumulation of DNMT1 at its target sites [33]. In ESCs stably expressing GFP-DNMT1 wt, foci of immobilized protein emerged already within 20 min (Figure 1D, left panel). In contrast, GFP-DNMT1 Δ458-500 was not enriched at replication foci even after 100 min, indicating that the deletion mutant is unable to methylate newly replicated DNA in living cells (Figure 1D, right panel). To pursue this idea, we further analyzed site-specific DNA methylation levels of stable GFP-DNMT1 wt and Δ458-500 ESC lines (Supplementary information, Figure S1B). GFP-DNMT1 could restore local DNA methylation at the major satellite repeats in *Dnmt1*$^{-/-}$ ESCs leading to an average methylation level of 62% that is comparable to the level of the wt cell line expressing the endogenous protein (74%, Figure 1E, left panel). In contrast, the DNMT1 mutant deficient in UHRF1 binding was unable to reestablish local DNA methylation patterns resulting in decreased levels at the major satellite repeats (average 19%) similar to the *Dnmt1*$^{-/-}$ control cell line (average 8%). Consistently, a similar defect of GFP-DNMT1 Δ458-500 in DNA methylation activity was observed at the single-copy sequence of the *skeletal α-actin* promoter (Figure 1E, right panel). Furthermore, similar results were obtained from DNA methylation analyses at the minor satellite repeats and the *Dnmt1o* promoter confirming that stable expression of GFP-DNMT1 Δ458-500 could not restore DNA methylation in a *Dnmt1*$^{-/-}$ cell line (Supplementary information, Figure S1D).

In summary, we provide strong evidence that the GFP-DNMT1 Δ458-500 mutant deficient in UHRF1 binding, even though able to methylate DNA substrates *in vitro*, cannot restore DNA methylation patterns in *Dnmt1*$^{-/-}$ ESCs. These findings suggest that the interaction of DNMT1 with UHRF1 is required to maintain DNA methylation *in vivo*.

The PHD and RING domain of UHRF1 are essential for maintenance DNA methylation

Cooperative binding of the UHRF1 TTD to di- and trimethylated histone H3K9 and of the SRA domain to hemimethylated DNA was described as a prerequisite for targeting DNMT1 to its substrate and for subsequent DNA methylation [34]. Given the regulatory impact of these two domains, we were interested in how the PHD and RING domain of UHRF1 may functionally contribute to maintenance DNA methylation by DNMT1. To this end, we introduced point mutations in the PHD and RING domain (UHRF1-GFP H346G and UHRF1-GFP H730A, respectively) that are expected to prevent coordination of zinc ions by zinc-finger motifs (Figure 2A and Supplementary information, Figure S2A). Consequently, the mutation in the RING domain significantly reduced the E3 ubiquitin ligase activity of UHRF1 *in vivo* (Supplementary information, Figure S2C and S2D). Notably, the preference of UHRF1-GFP for hemimethylated DNA was not impaired by the PHD and RING domain mutations (Supplementary information, Figure S2B).

First, we tested whether the point mutations in the PHD and RING domain influence the interaction of UHRF1 with DNMT1. UHRF1-GFP wt as well as UHRF1-GFP H346G and UHRF1-GFP H730A still co-precipitated with red fluorescent protein (RFP)-DNMT1, indicating that the mutations did not affect the interaction with DNMT1 directly (Figure 2B). In addition, the unaltered interactions were confirmed by a fluorescent three-hybrid assay [35, 36]. In this assay, UHRF1-GFP fusion constructs were used as baits by tethering them to a *lac* operator (*lac*O) array present in baby hamster kidney (BHK) cells that simultaneously expressed RFP-DNMT1 as a prey. Accumulation of RFP-DNMT1 at the *lac*O spot enriched for UHRF1-GFP wt, UHRF1-GFP H346G or UHRF1-GFP H730A clearly demonstrates that the mutant proteins were still able to interact with DNMT1 *in vivo* (Figure 2C).

In order to perform functional studies on the PHD and RING domain mutants, we stably expressed GFP-tagged UHRF1 wt, UHRF1 H346G or UHRF1 H730A in *Uhrf1*$^{-/-}$ ESCs. Similar to wt, also UHRF1-GFP H346G and UHRF1-GFP H730A showed focal enrichment at heterochromatin (Figure 2D, first panel and Supplementary information, Figure S2E). Thus, the mutations

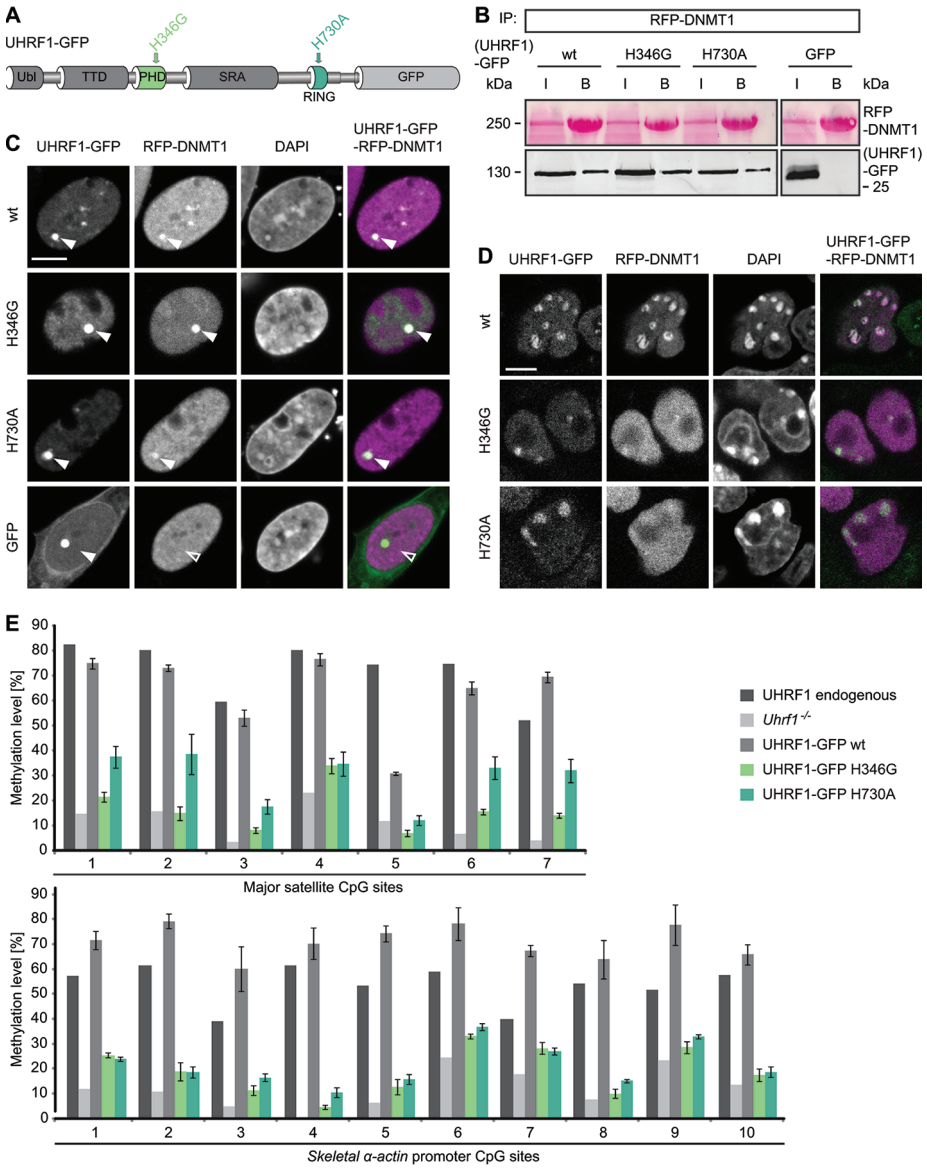


Figure 2 Mutations in the PHD and RING domain of UHRF1 affect DNMT1 targeting and maintenance DNA methylation, but not the interaction with DNMT1. **(A)** Schematic outline of the multidomain protein UHRF1. UHRF1 harbors a ubiquitin-like (Ubl) domain, a plant homeodomain (PHD) and a tandem Tudor domain (TTD) followed by a SET and RING-associated (SRA) domain and a really interesting new gene (RING) domain. UHRF1-GFP expression constructs carrying point mutations in the PHD (H346G) and RING domain (H730A) are illustrated. **(B)** Co-immunoprecipitation of UHRF1-GFP wt or PHD and RING domain mutants co-expressed with RFP-DNMT1 in HEK 293T cells. RFP-DNMT1 was immunoprecipitated using the RFP-Trap and bound UHRF1-GFP was detected by western blot with an anti-GFP antibody. GFP was used as negative control. Immunoprecipitated RFP-DNMT1 is shown by Ponceau staining. I, input; B. bound. **(C)** Fluorescence three-hybrid assay for visualization of the interaction RFP-DNMT1 with UHRF1-GFP wt or PHD and RING domain mutants. Displayed are confocal mid sections of BHK cells carrying a stably integrated Lac-operator array that were triple transfected with LacI fused to the GFP-binder, UHRF1-GFP constructs and RFP-DNMT1. DAPI was used for chromatin counterstaining. Closed arrows indicate the co-localization of both proteins at the *lac*O spot, open arrows indicate no co-localization. GFP was used as negative control. Scale bar, 5 μm. **(D)** Confocal mid sections of fixed *Uhrf1*$^{-/-}$ ESCs stably expressing UHRF1-GFP wt or PHD and RING domain mutant constructs. RFP-DNMT1 was transiently co-expressed and DNA was counterstained by DAPI. Scale bar, 5 μm. **(E)** Local DNA methylation analyses at major satellite repeats and the *skeletal* α*-actin* promoter. CpG site methylation levels of mouse E14 *Uhrf1*$^{-/-}$ ESCs stably expressing UHRF1-GFP wt or PHD and RING domain mutant constructs were analyzed by bisulfite treatment of genomic DNA, PCR amplification and direct pyrosequencing. The methylation level of E14 wt ESCs (endogenous UHRF1) and untransfected E14 *Uhrf1*$^{-/-}$ cells are shown for comparison. Mean values ± SD from two different clones were calculated, respectively.

do not affect localization of UHRF1. In contrast to its chromatin association in the UHRF1-GFP wt cell line, transiently co-expressed RFP-DNMT1 did not co-localize with UHRF1-GFP H6G and UHRF1-GFP H6A, but showed a dispersed distribution in the nucleus (Figure 2D, second panel). This observation is consistent with the result of staining for endogenous DNMT1 (Supplementary information, Figure S2E). Only in the UHRF1-GFP wt cell line, endogenous DNMT1 was enriched at S phase-specific replication sites, whereas it was diffusely distributed in the nucleus of the mutant cell lines pointing towards a defective DNMT1 targeting mechanism. To examine if DNMT1 methylation activity depends on the PHD and RING domain of UHRF1, we performed site-specific methylation analyses at heterochromatic regions. Consistent with defects in targeting DNMT1 to replication sites, DNA methylation levels at the major satellite repeats and the *skeletal α-actin* promoter revealed that both UHRF1-GFP H6G and UHRF1-GFP H6A were not able to mediate DNA remethylation by DNMT1 in *Uhrf1*$^{-/-}$ ESCs in contrast to UHRF1-GFP wt (Figure 2E). Especially at the major satellite repeats, the average DNA methylation in the PHD mutant cell lines remained nearly unchanged (16%) from the *Uhrf1*$^{-/-}$ control cell line (11%). Also, the average methylation levels in the RING domain mutant cell lines (29%) did not reach the wt DNA methylation level (62%) at the major satellite repeats. Similar results were obtained for the minor satellite repeats and the *Dnmt1o* promoter (Supplementary information, Figure S3A). Consistent with this site-specific DNA hypomethylation, the stable UHRF1 mutant cell lines also showed decreased global DNA methylation levels as compared with the wt (Supplementary information, Figure S3B and S3C). Partial rescue of global DNA methylation in the RING domain mutant cell lines could be due to residual E3 ubiquitin ligase activity of UHRF1-GFP H6A (Figure 3B, Supplementary information, Figure S2C and S2D).

To exclude the possibility that DNA hypomethylation might result from lower expression of the PHD and RING domain mutant (Figure 3A), we performed a transient rescue assay in *Uhrf1*$^{-/-}$ ESCs. Even though expression levels of the mutant constructs exceeded those of the UHRF1-GFP wt, the PHD and RING domain mutants could not mediate remethylation at the major satellite repeats (Supplementary information, Figure S3D and S3E) arguing for functional rather than expressional defects.

In summary, the PHD and RING domain mutants, although not affecting UHRF1 heterochromatin localization or the direct interaction with DNMT1, cannot mediate either targeting of DNMT1 to replication foci nor maintenance DNA methylation. These findings suggest that these UHRF1 domains contribute to the recruitment of DNMT1 by independent mechanisms.

The PHD and RING domain of UHRF1 are required for ubiquitination of histone H3

Histone H3 has been reported as a UHRF1-dependent ubiquitination target in *Xenopus* egg extracts [32], providing a potential mechanism for the recruitment of DNMT1 to chromatin. Thus, we set out to investigate whether H3 ubiquitination required PHD-mediated histone binding and RING domain-mediated ubiquitin E3 ligase activity of UHRF1 in mammalian cells. To this end, we extracted histones from wt or *Uhrf1*$^{-/-}$ ESCs and detected modified H3. As expected, histone H3 was less ubiquitinated in the absence of UHRF1 (Figure 3A and 3B), indicating that UHRF1 serves as a ubiquitin E3 ligase for H3 in mammalian cells. We also found that ubiquitination levels of histone H3 in *Uhrf1*$^{-/-}$ ESCs stably expressing the RING domain mutant UHRF1-GFP H6A were not rescued to the level of wt cells. Surprisingly, the PHD mutant UHRF1-GFP H6G also could not restore ubiquitination of histone H3 (Figure 3A and 3B).

Since the PHD has been reported to bind to unmodified H3R2 [26, 37-39], we investigated the role of this histone residue in H3 ubiquitination by mutational analyses. Compared with GFP-H3 wt, ubiquitination of a GFP-H3 R2A mutant expressed in human embryonic kidney (HEK) 293T cells was clearly reduced (Figure 3C) pointing towards an important role of the R2 residue for UHRF1-dependent H3 ubiquitination.

To further test the histone binding properties of the PHD mutant *in vitro*, we performed a peptide pull-down assay with wt or PHD and RING domain mutant UHRF1-GFP using H3 peptides with an unmodified, trimethylated or acetylated K9 residue. The mutation in the RING domain did not alter the histone binding of UHRF1-GFP showing a preference for unmodified and K9 trimethylated H3 peptides similar to the wt protein (Figure 3D). The mutation in the PHD, however, decreased the binding to both, the unmodified and the K9 trimethylated peptide. We further examined the histone binding preferences of UHRF1-GFP with an *in vitro* histone tail binding assay. The results revealed the binding of UHRF1-GFP to unmodified but not R2 dimethylated H3 histone tails (Supplementary information, Figure S4A, consistent with prior K_d measurements [39]. As the PHD of UHRF1 has been shown to bind unmethylated H3R2 residues and to contribute to the K9 methylated H3 histone binding of the TTD [26, 37, 39], we propose that PHD-dependent histone binding is required for UHRF1-mediated ubiquitination of histone H3

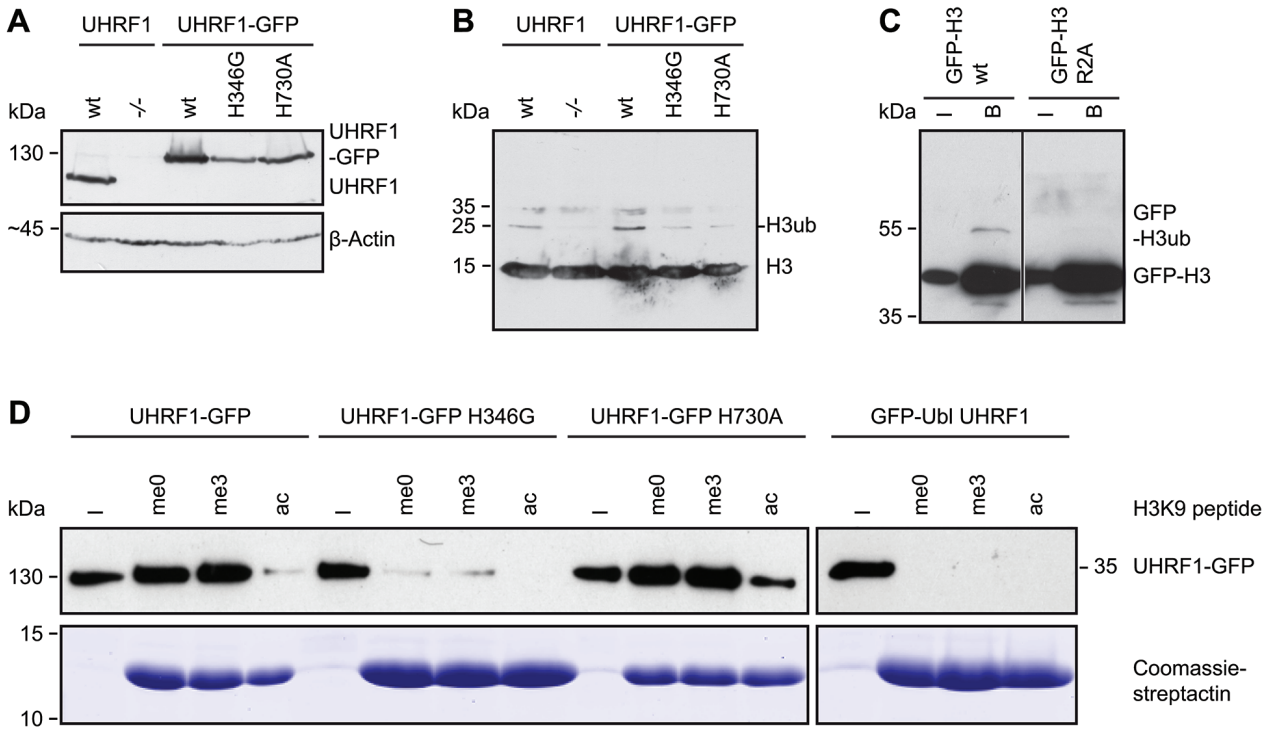


Figure 3 Histone H3 ubiquitination requires the UHRF1 PHD and RING domain. **(A)** Western blot analyses of endogenous UHRF1 or stably expressed UHRF1-GFP wt or H346G and H730A mutants in E14 *Uhrf1*$^{-/-}$ ESCs with an anti-UHRF1 antibody. Equal loading is shown by an anti-β-Actin antibody. **(B)** Analyses of H3 ubiquitination from acid extracted histones derived from the different cell lines in (A). A specific anti-H3 antibody was used for detection. **(C)** Ubiquitination of GFP-H3 in dependence on R2. GFP-H3 wt and the arginine to alanine mutant (R2A) were co-expressed with UHRF1-His in HEK 293T cells, respectively, and after immunoprecipitation with the GFP-Trap, the bound fraction was detected by western blotting with a specific anti-H3 antibody. I, input; B, bound. **(D)** *In vitro* peptide pull-down assay of UHRF1-GFP wt or the PHD and RING domain mutants from crude cells extracts of HEK 293T cells using H3 peptides (amino acid 1-20) that were either unmodified (me0), K9 trimethylated (me3) or K9 acetylated (ac) and functionalized on streptactin beads. The GFP-Ubl domain of UHRF1 was used as negative control, Coomassie-stained streptactin is shown as loading control. I, input.

UHRF1 ubiquitinates histone H3 on K18 in mammalian cells

Using *Xenopus* extracts immunodepleted for DNMT1, H3 was shown to be ubiquitinated at the K23 residue [32]. To map ubiquitination sites on histone H3 tails in mammalian cells, we performed mass spectrometry using human and mouse cells. In contrast to the results from *Xenopus* extracts, the K18 residue of histone H3 was identified as novel ubiquitination site in mouse ESCs, while the K23 residue was unmodified or acetylated (Figure 4A and 4B). Relative quantification of H3 peptides containing ubiquitinated K18 and an unmodified or acetylated K23 residue showed a reduction of K18 ubiquitination in ESCs lacking UHRF1 (Figure 4C and 4D). Similarly, immunoprecipitation of GFP-UHRF1 from HEK 293T cells and subsequent mass spectrometry also revealed ubiquitination at K18 but not at K23 (Supplementary information, Figure S5A). Comparison of ubiquitination levels of overexpressed GFP-H3 carrying an R2A, K18A or K23A mutations suggests that in this constellation K23 could also be modified (Supplementary information, Figure S5B). Interestingly, the GFP-H3 R2A construct showed reduced ubiquitination levels indicating that the R2 residue plays a role in efficient H3 ubiquitination.

DNMT1 harbors a UIM that mediates binding to ubiquitinated H3 and is essential for DNA methylation activity in vivo

To unravel how H3 ubiquitination may contribute to maintenance DNA methylation we screened DNMT1 for potential binding motifs. With bioinformatics analyses, we identified a ubiquitin interacting motif (UIM) in the N-terminal regulatory domain of DNMT1. This motif is located in a region spanning from amino acid 98 to 9 of mouse DNMT1 and shows striking similarity to UIMs of known ubiquitin interacting proteins (Figure 5A). Comparison of the ubiquitin binding properties between GFP-DNMT1 wt and mutants either lacking

the UIM (Δ356-404) or containing substitutions of the relevant and conserved amino acids in the motif to alanine (D381A-E382A-S392A, D381A-E382A-M385A-S392A-D395A, Figure 5A and Supplementary information, Figure S6A) showed a defect in the association with ubiquitinated histone H3 and ubiquitinated H2AK119 (Figure 5B, 5C and Supplementary information, Figure S6B-S6D). To further elucidate UIM-dependent ubiquitinated histone binding, we quantified modified H3 18-26 peptides bound by GFP-DNMT1 wt or the UIM mutants by mass spectrometry. Whereas H3 histone peptides ubiquitinated at K18 and acetylated or unmodified at K23 co-immunoprecipitated with GFP-DNMT1 wt, only little to no ubiquitinated peptide signals were detected for the UIM mutants (Figure 6A, 6B). GFP-DNMT1 Δ458-500 defective in UHRF1 interaction (Figure 5B) showed reduced (Figure 6B) or undetectable (Figure 5B, 5C) binding to ubiquitinated H3 and H2A. This deletion located in a TS domain region C-terminal of the UIM might affect the integrity and functionality of the motif responsible for ubiquitin binding. Therefore, we cannot rule out that apart from disrupted UHRF1 binding also defects in the association with ubiquitinated histones contributed to the observed changes in subnuclear distribution and protein function of GFP-DNMT1 Δ458-500 (Figure 1C-1E).

Besides a decreased binding to ubiquitinated H3, the TS domain point and deletion mutants exhibited an increased binding to H3 or core histones compared with GFP-DNMT1 wt (Figure 5B). Therefore, specific binding of DNMT1 to ubiquitinated H3 via its UIM might prevent the enzyme from stable chromatin association and thereby facilitate DNA methylation.

To clarify the functional role of the UIM in maintenance DNA methylation *in vivo*, we performed a functional complementation assay in *Dnmt1*$^{-/-}$ ESC lines transiently expressing GFP-DNMT1 wt, GFP-DNMT1 Δ356-404, GFP-DNMT1 D381A-E382A-S392A or GFP-DNMT1 D381A-E382A-M385A-S392A-D395A. Local DNA methylation analyses at the major satellite repeats and the *skeletal α-actin* promoter showed that the UIM mutants were not able to reestablish DNA methylation patterns (Figure 6C). GFP-DNMT1 wt restored DNA methylation at the major satellite repeats to 84%. By comparison, the UIM deletion and point mutants were not able to rescue resulting in low average methylation levels of 20% to 23% comparable to untransfected *Dnmt1*$^{-/-}$ ESCs (15%). Similar results were also observed at the minor satellite repeats and the *Dnmt1o* promoter (Supplementary information, Figure S7A).

Given that the GFP-DNMT1 TS UIM deletion and point mutants were able to interact with Ch-UHRF1 (Supplementary information, Figure S7B), we were in-

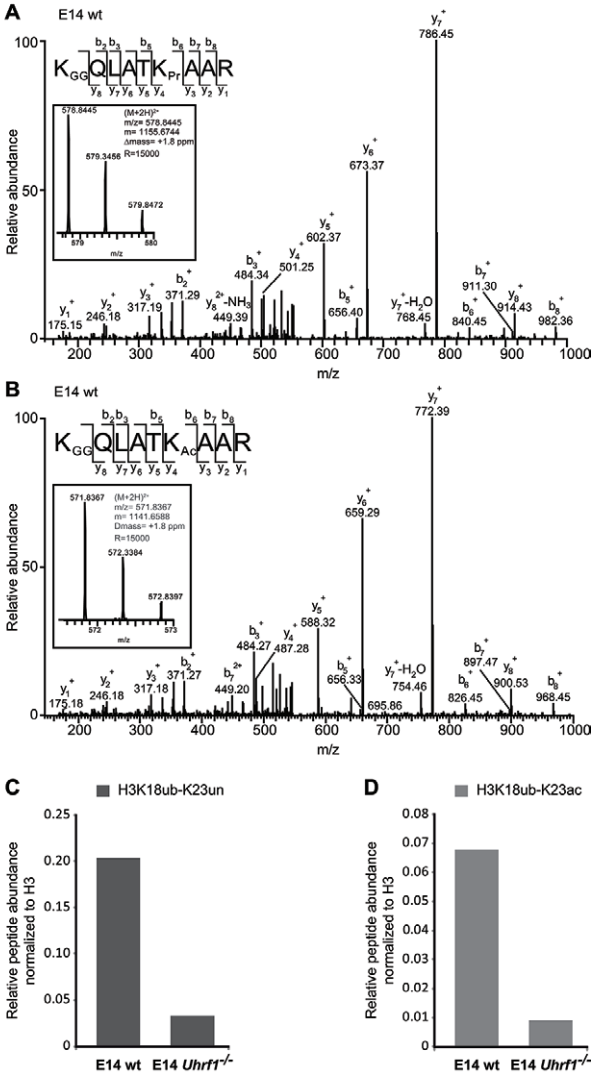


Figure 4 UHRF1 ubiquitinates histone H3 at K18 in mammalian cells. **(A)** Identification of H3 18-26 peptides carrying ubiquitination (GG) at K18 and no modification (Pr) at K23 by LC-MS/MS. MS2 fragmentation spectrum of the precursor ion is shown in the inset. An almost complete series of b and full y product ions generated by CID fragmentation were detectable providing a high confidence in its correct identification and localization of the ubiquitin modification. Inset: mass, charge and measurement error determination of the H3 18-26 peptides K18GGK23Pr in the E14 wt sample. Displayed is the isotopic distribution of the H3 peptide from which the mass to charge ratio (*m*/*z*), the charge (2+) and the monoisotopic mass value (*m*) were derived. Δ*m*: difference between the expected and the measured masses; *R*: resolution of the MS measurement. **(B)** Identification of H3 18-26 peptides carrying ubiquitination (GG) at K18 and acetylation (Ac) at K23 by LC-MS/MS as in (A). **(C**, **D)** Quantification of H3 18-26 peptides carrying ubiquitination (ub) at K18 and an unmodified (un) or acetylated (ac) K23 residue from E14 wt and E14 *Uhrf1*$^{-/-}$ samples. Extracted ion chromatograms of the ions corresponding to the peptides of interest were used for the quantification. The signals were normalized against the total amount of analyzed H3 proteins.

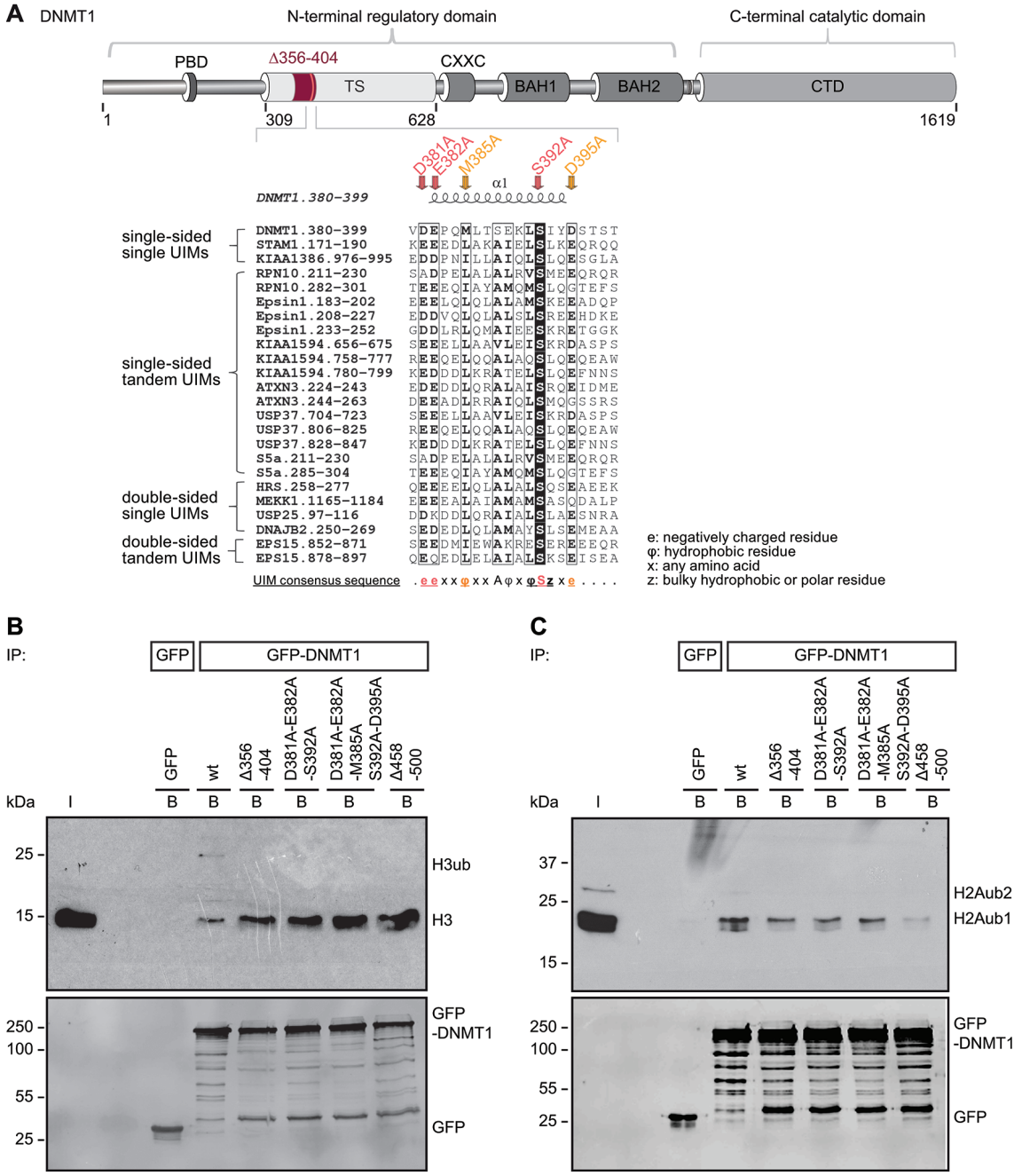


Figure 5 The TS domain of DNMT1 harbors a ubiquitin interacting motif (UIM) that is essential for binding to ubiquitinated H3 and H2A. **(A)** Schematic outline of the UIM in the TS domain of DNMT1 and indication of the UIM deletion (Δ356-404) and the point mutations (D381A-E382A-S392A and D381A-E382A-M385A-S392A-D395A). A peptide sequence of DNMT1 encompassing amino acid 380-399 was aligned with peptide sequences of proteins previously known to contain UIMs. Identical amino acids are highlighted in black, highly similar amino acids are framed in black. The secondary structure of the DNMT1 region (pdb: 3EPZ [10]) harboring the UIM is displayed on top of the sequence alignment generated using ESPript [78]. The consensus sequence for single-sided UIMs [58] is shown below. The UIMs were found by scanning the protein primary sequences against a collection of motifs in ExPASy Prosite. Putative subgroups of UIMs are indicated on the left. **(B)** Ubiquitinated histone H3 binding assay. After extraction of histones from HEK 293T cells, the extracts were incubated with GFP-DNMT1 wt or mutants immobilized on the GFP-Trap and the bound fractions were analyzed by western blotting with specific anti-H3 and anti-GFP antibodies. GFP was used as negative control. I, input; B, bound. **(C)** Ubiquitinated histone H2A binding assay as in (B). Bound fractions were analyzed by western blotting with specific anti-H2AK119ub and anti-GFP antibodies. Analyses of the anti-H2AK119ub antibody specificity and of peptides isolated from the corresponding band are shown in Supplementary information, Figure S6C and S6D. H2Aub1, monoubiquitinated H2A; H2Aub2, diubiquitinated H2A.

terested in how the UIM in DNMT1 has an influence on the subnuclear localization of the protein. Immunostaining of replicating DNA with a specific anti-PCNA antibody indicated that GFP-DNMT1 wt was enriched at S phase-specific replication foci, while GFP-DNMT1 Δ356-404, GFP-DNMT1 D381A-E382A-S392A and GFP-DNMT1 D381A-E382A-M385A-S392A-D395A showed only weak association with the PCNA-stained replication sites especially in late S phase (Supplementary information, Figure S8). To analyze the UIM-dependent enrichment of DNMT1 at late-replicating heterochromatin, we quantified mean fluorescence intensities at chromocenters compared with the nucleoplasmic region (Figure 7A). In late S phase ES and mouse embryonic fibroblasts (MEF) cells, GFP-DNMT1 wt localized at chromocenters, whereas the UIM mutations abolished heterochromatin enrichment (Figure 7B and 7C). These results clearly demonstrate the key role of the UIM in DNMT1 targeting via ubiquitinated histone H3 binding and for maintenance DNA methylation in mammalian cells.

Discussion

DNA methylation is an important epigenetic modification regulating gene expression in development and disease. A key question is how methylation marks are set, maintained and removed. According to previous models, DNA methylation marks are set by the *de novo* methyltransferases DNMT3A and DNMT3B during development and maintained by the maintenance DNA methyltransferase DNMT1 that specifically recognizes and modifies hemimethylated DNA substrates. However, the preference of DNMT1 for hemimethylated DNA measured *in vitro* [40-43] is not sufficient to explain efficient maintenance of DNA methylation patterns over many cell division cycles *in vivo*. The interaction of DNMT1 with the replication protein PCNA was shown to enhance maintenance DNA methylation by a factor of two, but not to be essential [6, 7]. In contrast, the interacting factor UHRF1 recruiting and allosterically activating DNMT1 is essential for DNA methylation [14, 16, 17, 44]. In this study, we have now dissected the distinct role of different UHRF1 and DNMT1 domains in directing DNA methylation.

In line with previous studies, we show that, albeit being weak, the TS domain-mediated interaction of DNMT1 with the SRA domain of UHRF1 is required for targeting and function of DNMT1 *in vivo*. Accordingly, truncated DNMT1 (Δ458-500) deficient in UHRF1 binding showed weaker association with chromocenters in late S phase mouse fibroblasts [4] and failed to maintain DNA methylation in ESC (Figure 1).

Heterochromatin binding of UHRF1 is mediated by the TTD, PHD and SRA domain and defects in any of these three domains lead to decreased DNA methylation by DNMT1 [34, 45, 46]. Accordingly, it was postulated that UHRF1 reads and binds repressive histone marks and hemimethylated DNA and via direct protein-protein interaction recruits DNMT1 for maintenance DNA methylation.

Defects of a RING domain mutant (C713A, C515A and C716A) in restoring ubiquitinated H3 in HeLa cells after knockdown of human DNMT1 and UHRF1 have previously been reported [32]. We found that the RING domain, though not directly involved in UHRF1 chromatin binding or interaction with DNMT1, is indispensable for DNA methylation by DNMT1. Remarkably, a UHRF1 RING domain mutant (H730A) with diminished ubiquitin E3 ligase activity (Supplementary information, Figure S2C and S2D) that could still bind DNMT1 (Figure 2B), hemimethylated DNA and K9 trimethylated H3 peptides *in vitro* (Figure 3D and Supplementary information, Figure S2B) and chromocenters *in vivo* (Supplementary information, Figure S2E), nonetheless failed in recruiting DNMT1 to replication sites (Figure 2D and Supplementary information, Figure S2F). These findings suggest that DNMT1 recruitment to replication forks is not based on direct interaction with UHRF1, but on the catalytic activity of the RING domain. Previously, the RING domain of UHRF1 has been reported to have an autoubiquitination activity [28] and, in addition, to ubiquitinate DNMT1 [29, 30] and histone substrates [24, 25]. A recent study describes that ubiquitination of H3 by UHRF1 provides docking sites for DNMT1 on chromatin and thus couples maintenance DNA methylation and replication [32]. While we could confirm the essential role of UHRF1, we obtained new insights into the complex functional interplay of UHRF1 and DNMT1 domains.

First, in contrast to ubiquitination at K23 in *Xenopus* egg extracts [32], our mass spectrometry results identified H3K18 as ubiquitination target of UHRF1 in mammalian cells (Figure 4A, 4B and Supplementary information, Figure S5A). By mutational analysis in HEK 293T cells, we found that in absence of K18, the mutated GFP-tagged H3 might be ubiquitinated at K23 (Supplementary information, Figure S5B). However, by semiquantitative analysis of endogenous ubiquitinated H3 peptides in wt versus *Uhrf1*$^{-/-}$ mouse ESCs using mass spectrometry, we clearly show the specificity of K18 ubiquitination by UHRF1 and its reduction by UHRF1 depletion (Figure 4C, 4D). Second, in the previous study, a deletion of 100 amino acids within the DNMT1 TS domain (Δ325-425) caused a loss of histone binding *in vitro* [32]. The TS domain is, however, involved in multiple interactions and required for proper folding, stability and activity of DNMT1. The incomplete structural information indicates

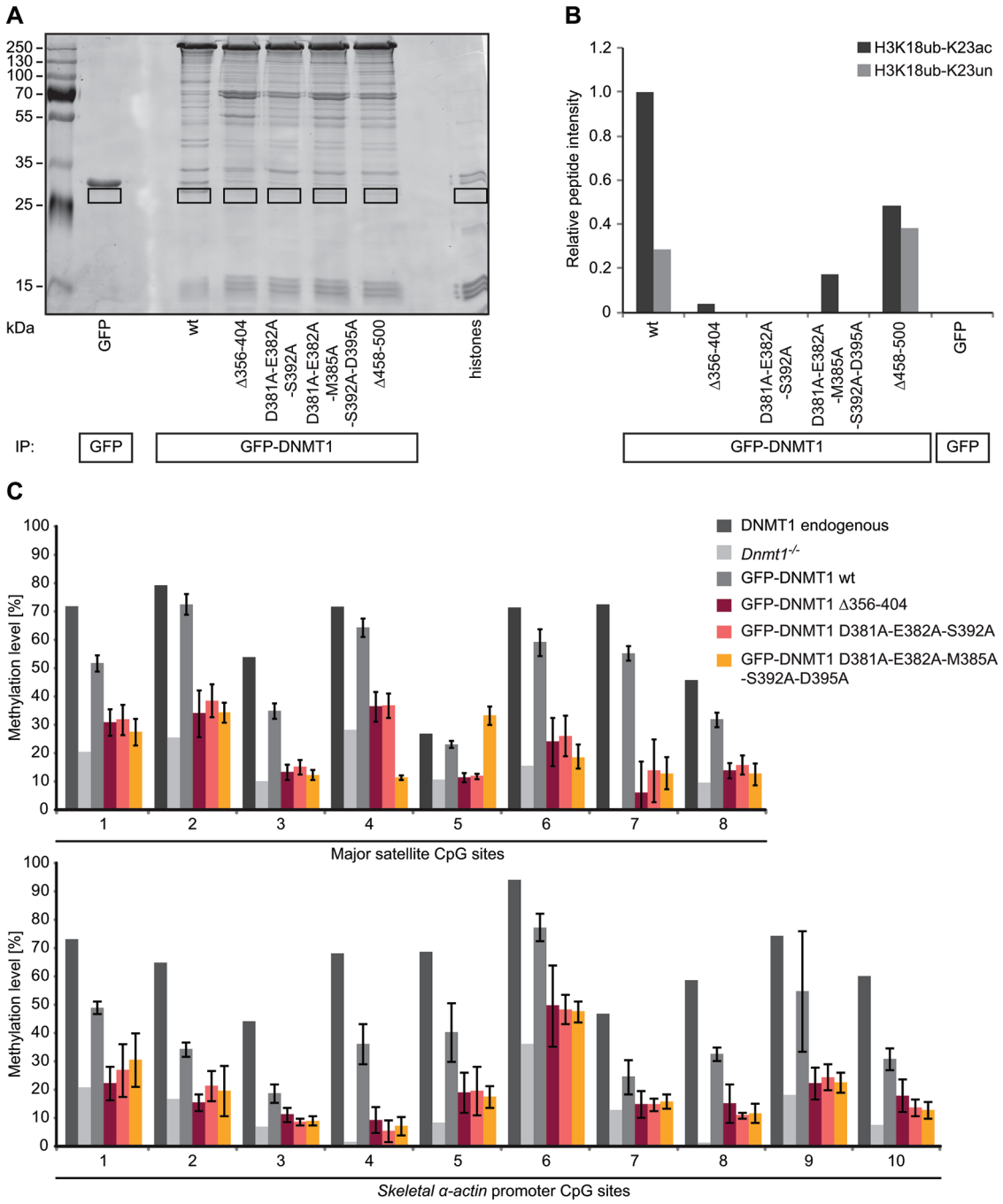


Figure 6 The DNMT1 UIM is required for ubiquitinated H3K18 binding and for DNA methylation. **(A)** Ubiquitinated histone binding experiments using GFP-DNMT1 wt or UIM mutants as well as the Δ458-500 mutant deficient in binding to UHRF1. Equal amounts of GFP fusions were immobilized on the GFP-Trap and incubated with acid extracted histones. Bound proteins were visualized by Coomassie staining and the fractions highlighted by black rectangles were analyzed by mass spectrometry. GFP was used as negative control. **(B)** Quantification of H3 18-26 peptides carrying ubiquitination (ub) at K18 and an acetylated (ac) or unmodified (un) K23 residue from histone binding experiment shown in (A). Extracted ion chromatograms of the ions corresponding to the peptides of interest were used for quantification (H3K18ubK23ac: m/z = 571.8353 ± 10 ppm; H3K18ubK23un: m/z = 578.8441 ± 10 ppm). **(C)** Local DNA methylation analyses of J1 *Dnmt1*$^{-/-}$ ESCs expressing GFP-DNMT1 wt or Δ356-404 and UIM point mutants. CpG methylation levels at the major satellite repeats and the *skeletal α-actin* promoter were analyzed by bisulfite treatment of genomic DNA, PCR amplification and direct pyrosequencing. Methylation levels of untransfected J1 *Dnmt1*$^{-/-}$ cells are shown for comparison. Mean values ± SD from three to four biological replicates were calculated, respectively.

different TS domain conformations and a role in autoinhibition of the CD, but does not provide any further mechanistic insights [10, 11, 47]. With bioinformatics and mutational analyses, we identified a conserved UIM located in the TS domain of DNMT1 (amino acids 381-395) that mediates the recognition of ubiquitinated H3 *in vitro* (Figure 5, 6A, 6B and Supplementary information, Figure S6B). Localization and activity analyses with specific mutants *in vivo* clearly indicated that the UIM is required for DNMT1 subnuclear distribution and maintenance DNA methylation (Figures 6C, 7 and Supplementary information, Figure S7A and S8).

Last, we could show that besides hemimethylated DNA binding by the SRA domain [32], UHRF1 PHD binding to H3R2 is also required for H3 ubiquitination and subsequent DNA methylation (Figure 2E and 3B). Therefore, we propose that cooperative chromatin binding of the TTD, the PHD and the SRA domain constitutes a prerequisite for H3K18 ubiquitination. These ubiquitinated histone tails are recognized by the UIM and thus mediate DNMT1 chromatin binding. Thereby, UHRF1 acts as a reader and writer of histone marks and via recruitment of DNMT1 dynamically links DNA and histone modification pathways. Based on these results, we propose a ubiquitination-dependent chromatin targeting mechanism for DNMT1 that is essential for maintenance DNA methylation after replication (Figure 8A). The identification and functional characterization of a UIM in DNMT1 not only changes our view of maintenance DNA methylation, but also opens new perspectives for the involvement of DNMT1 in other repressive epigenetic pathways (Figure 8B).

Besides association with ubiquitinated H3, we found that DNMT1 also binds ubiquitinated H2AK119 (Figure 5C and Supplementary information, Figure S6C, S6D). Consistently, DNMT1 was recently detected among proteins binding to H2A ubiquitinated at K118 in *Drosophila*, corresponding to K119 in mammals [48]. H2AK119 ubiquitination is catalyzed by RING1A/1B, two components of the Polycomb repressive complex 1 (PRC1), and plays an important role in regulating gene expression [49]. Similar to UHRF1-dependent H3 ubiquitination, H2A ubiquitination by RING1A/1B might also contribute to DNA methylation. We speculate that UIM-mediated binding of DNMT1 to ubiquitinated H2AK119 might direct DNMT1 to un- or hemimethylated sites dependent on PRC1 ubiquitination activity (Figure 8B, left half).

PRC1-dependent H2A ubiquitination further leads to PRC2 recruitment and subsequent H3K27 methylation [50]. Enhancer of Zeste homolog 2 (EZH2), a component of PRC2, writes methylated H3K27 and interacts with DNMTs. This interaction was shown to be required for DNA methylation of EZH2 target promoters [51]. DNMT1 depletion in differentiated cells affects H2A ubiquitination-dependent PRC2 recruitment at pericentromeric heterochromatin [52]. Thus, UIM binding to ubiquitinated H2A is likely DNA replication independent and DNMT1 might function as adaptor protein mediating PRC2 recruitment and repressive Polycomb domain formation.

Besides recruiting DNMT1 to specific sites on chromatin, the UIM could also play a role in the allosteric activation of the enzyme. The UIM is located within the TS domain of DNMT1 that had been shown to bind the CD and thereby inhibit catalytic activity [10, 11]. It is tempting to speculate that competitive UIM binding to ubiquitinated histone tails displaces the TS domain from the DNA binding pocket and abolishes autoinhibition of DNMT1.

Given the emerging role of ubiquitination in DNA methylation, it is interesting to notice that ubiquitination is a highly dynamic post-translational modification that can be reversed by ubiquitin-specific proteases (USPs). The UHRF1-DNMT1 complex has been reported to contain USP7 that deubiquitinates and stabilizes DNMT1 [29, 30]. Thus, USP7 might in addition modulate the ubiquitination status of histone H3 and thereby regulate DNMT1 association with chromatin. An alternative pathway controlling DNMT1 chromatin association could involve the recently described chromatin acetylation of H3K18 and K23 [53, 54]. Acetylated H3K18 is enriched at the transcriptional start sites of active and poised genes [55]. Thus, H3K18 acetylation might counteract ubiquitination and thereby prevent binding and silencing of active genes by DNMT1. The dynamic interplay of ubiquitination and acetylation of H3K18 likely controls DNMT1 chromatin binding and thereby directs methylation activity. Studies of UHRF1 and DNMT1 complex composition in different cell cycle phases and cell types should provide further insights into the fine-tuning of DNMT1 activity *in vivo*.

Given the complex role of the large TS domain on the one hand and the scarce structural and mechanistic data on the other hand, our identification of a well defined UIM provides a concrete basis for functional insights. Ubiquitin binding proteins with defined UIMs have been described in various cellular processes like, e.g., sorting of ubiquitinated membrane proteins for lysosomal degradation. The crystal structure of the signal transducing adaptor molecule 1 (STAM1) [56] suggests that three central amino acids in the UIM, L176, A179 and S183 form a hydrophobic interface for ubiquitin binding [57]. Similar to the UIM in STAM1, the UIM in DNMT1 also harbors a conserved hydrophobic amino acid M385 and S392 flanked by negatively charged amino acids (D381, E382 and D395), which we found to be essential in n analyses (Figures 5, 6 and 7). Different from other UIMs,

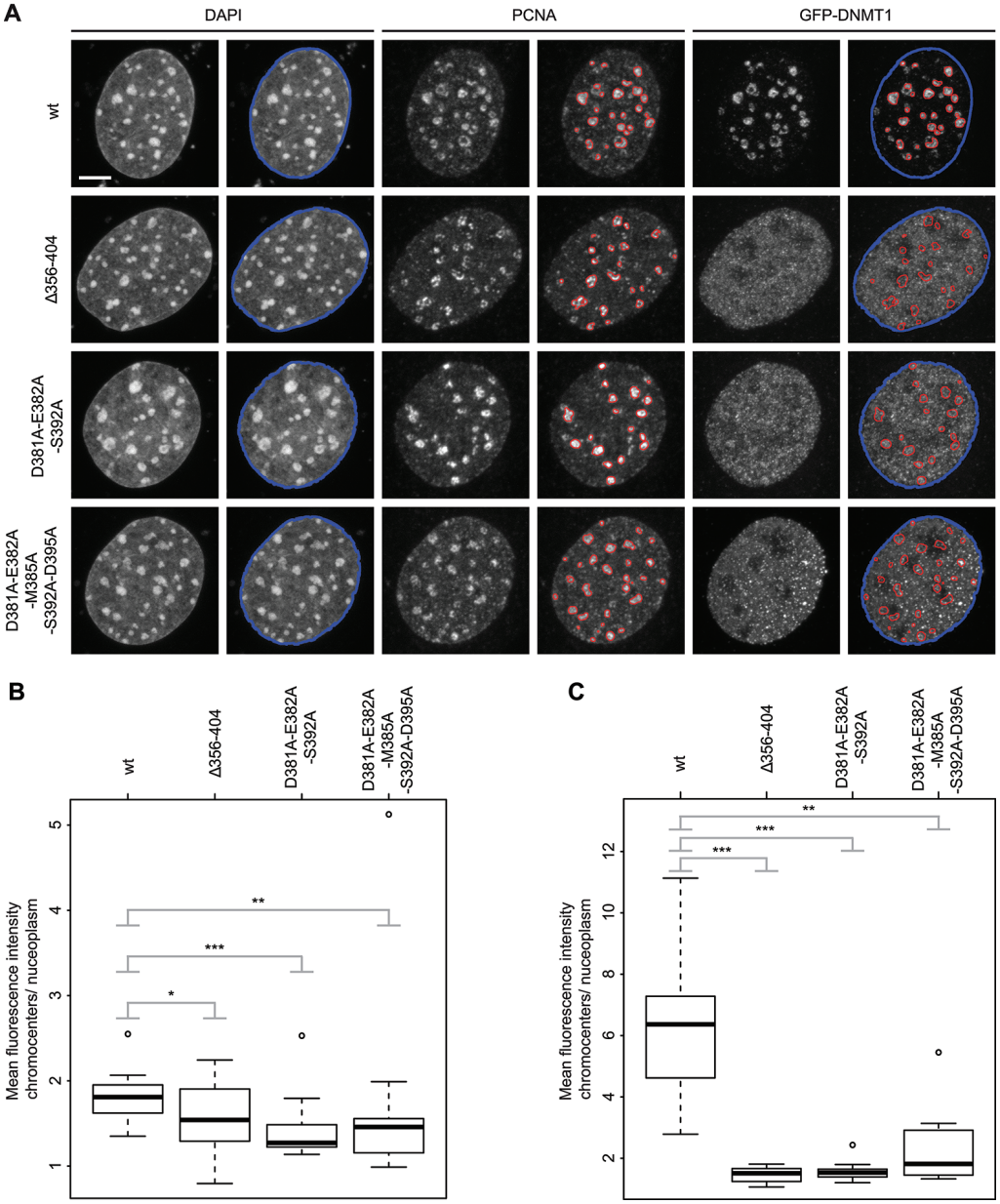


Figure 7 GFP-DNMT1 UIM mutants show a decreased association with PCNA-stained replication sites in late S phase compared with the wt. **(A)** Maximum intensity projections of MEF cells transiently expressing GFP-DNMT1 wt or UIM mutants. Replicating DNA was stained with a specific anti-PCNA antibody and chromatin was counterstained with DAPI. Replication foci masks (red) match the enrichment of GFP-DNMT1 wt in late S phase, whereas the UIM mutants do not show a focal enrichment. Segmentations were generated in an automated fashion using a machine learning algorithm (WEKA). The nuclear mask outlined in blue was based on the DAPI staining, whereas the replication foci masks outlined in red were based on the PCNA staining. Both masks were superimposed on the GFP channels. The GFP-DNMT1 signal inside the red masks (chromocenters) relative to the remainder of the nucleus (nucleoplasm) was quantified. Scale bar, 5 μm. **(B)** Quantification of chromocenter association of GFP-DNMT1 wt or UIM mutants in late S phase J1 *Dnmt1*$^{-/-}$ ESCs. The ratio of the mean GFP fluorescence intensity at chromocenters over the mean intensity in the nucleoplasm is shown in the box plot from 15 (wt), 16 (Δ356-404), 12 (D381A-E382A-S392A) or 18 (D381A-E382A-M385A-S392A-D395A) cells. The results were further analyzed in R using a Wilcoxon test and considered as statistical significant for $P < 0.05$ (*) and $P < 0.01$ (**) or highly significant for $P < 0.001$ (***). The following P values were calculated: Δ356-404: $P = 0.049$, D381A-E382A-S392A: $P = 0.0016$ and D381A-E382A-M385A-S392A-D395A: $P = 0.0056$. **(C)** Quantification of chromocenter association of GFP-DNMT1 wt or UIM mutants in late S phase MEF cells as in (B). Eleven (wt), 12 (Δ356-404, D381A-E382A-S392A) or 10 (D381A-E382A-M385A-S392A-D395A) cells were analyzed. The following P values were calculated in R using a Wilcoxon test: Δ356-404: $P = 0.00000148$, D381A-E382A-S392A: $P = 0.00000148$ and D381A-E382A-M385A-S392A-D395A: $P = 0.0012$.

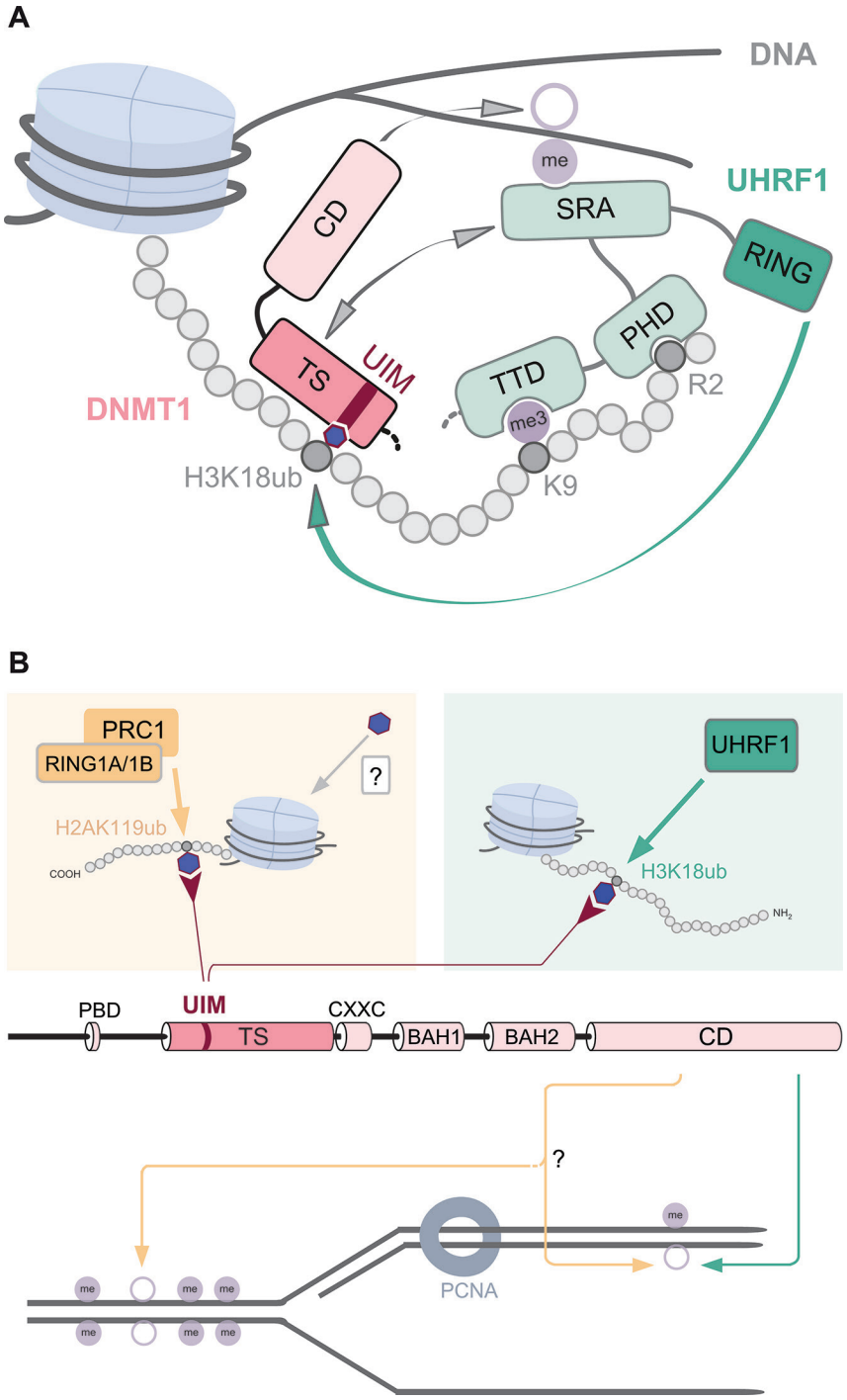


Figure 8 Overview of interactions and modifications controlling DNMT1 activity. (**A**) UHRF1 is enriched at H3 tails as a result of the PHD-mediated binding to H3R2, the TTD-mediated binding to methylated H3K9 and recognition of hemimethylated CpG sites via the SRA domain. By interaction of the SRA domain with the TS domain, DNMT1 is directly recruited to its target sites. UHRF1 chromatin binding via its TTD, PHD and SRA domain is a prerequisite for subsequent H3 ubiquitination by the RING domain. The UIM of DNMT1 binds to H3 tails ubiquitinated at K18 by UHRF1 and is essential for DNMT1 targeting and DNA methylation *in vivo*. (**B**) The previously described direct interaction of DNMT1 with UHRF1 and PCNA is not sufficient for maintenance DNA methylation. Besides the UHRF1-dependent H3K18 ubiquitination recruiting DNMT1 via its UIM for maintenance DNA methylation (right half), we propose an alternative pathway that involves H2AK119 ubiquitination by RING1A/1B of PRC1 (left part). The identification of the DNMT1 UIM now opens the possibility that ubiquitination of histone tail residues by ubiquitin E3 ligases might constitute alternative pathways for DNA methylation by DNMT1 CD beyond classic maintenance. Blue hexagons represent a ubiquitin moiety.

the central conserved Arg residue is not present in DNMT1 (Supplementary information, Figure S6A). Based on sequence alignments and structural information, UIMs can be subdivided in single-sided single UIMs, as in STAM1, and in single-sided tandem UIMs, as in the proteasome subunit S5a [58] (Figure 7A). The tandem UIMs in S5a provide a model for the recognition of polyubiquitin chains [59]. In contrast, a double-sided single UIM in the hepatocyte growth factor-regulated tyrosine kinase substrate (HRS) allows for efficient binding of multiple monoubiquitinated receptors in the processing of endosomal protein sorting [58]. Comparison with these known UIMs suggests that the motif in DNMT1 belongs to the group of single-sided single UIMs, which would be compatible with their recognition of single ubiquitinated substrates.

In summary, the functional analysis of UHRF1 domains and the identification of a UIM in DNMT1 challenge traditional views of maintenance DNA methylation as a simple copying mechanism. Instead, DNA methylation by DNMT1 requires reading of H3R2, H3K9me3 and hemimethylated DNA by UHRF1 and subsequent ubiquitination of H3K18 by its RING domain thereby integrating signals from different epigenetic pathways. These multiple layers controlling DNMT1 activity suggest that overall methylation densities in chromatin domains are maintained rather than specific methylation patterns precisely copied. The functional characterization of the UIM further raises the possibility that other ubiquitin E3 ligases like RING1A/1B of PRC1 might direct DNMT1 activity to repressive chromatin domains beyond implem aintenance.

Materials and Methods

Expression constructs and antibodies

Fusion constructs were generated using enhanced GFP, monomeric RFP or monomeric Ch. The expression constructs for GFP, RFP-DNMT1, GFP-DNMT1 wt, GFP-DNMT1 Δ458-500, GFP-DNMT1 309-628 (GFP-TS) and UHRF1-His have been described previously [9, 29, 33, 60, 61]. GFP-TS Δ458-500, GFP-DNMT1 Δ356-404 and GFP-DNMT1 point mutant (D381A-E382A-S392A and D381A-E382A-M385A-S392A-D395A) expression constructs as well as UHRF1-GFP H346G and H730A were derived from the corresponding wt constructs by overlap extension PCR [62]. The GFP-UHRF1 single-domain construct for the ubiquitin-like domain (Ubl) was generated by PCR using the corresponding wt full-length construct. Ch UHRF1 and GFP-UHRF1 expression constructs have been described previously [29]. Expression constructs for GFP-H3 R2A, K8A, K23A as well as K18A-K23A were obtained by overlap extension PCR on the corresponding wt constructs. The construct for LacI-GBP has been reported before [36, 64, 65]. All constructs were verified by DNA sequencing (MWG Biotech).

For immunofluorescence staining of heterochromatin, a mouse anti-H3K9me3 and an anti-H4K20me3 antibody were used (Active Motif). Endogenous DNMT1 was stained with the rat monoclonal antibody 5A10 [4] and PCNA with the rat monoclonal antibody 16D10 [66]. As secondary antibodies an anti-mouse Alexa Fluor 594 and anti-rat Alexa Fluor 647 antibody were applied, respectively (Invitrogen).

For detection of GFP fusion proteins by western blot, a mouse anti-GFP (Roche) or a rat anti-GFP (Chromotek) antibody was used. RFP or Ch fusion proteins were detected by the rat anti-red antibody 5F8. UHRF1 was visualized by a rabbit anti-UHRF1 antibody [29] and HA-biotinylated mouse monoclonal anti-HA antibody 12CA5. Equal loading of cell lysates was assessed by a mouse anti-β-Actin antibody (Sigma-Aldrich). The rabbit anti-H3 antibody was purchased from Abcam and the anti-H2AK119ub from New England Biolabs. Depending on the expected intensity of the signals, secondary antibodies either conjugated to horseradish peroxidase (anti-rabbit (Biorad), anti-rat and anti-mouse (Dianova)) or conjugated to fluorescent dyes (anti-mouse and anti-rat Alexa Fluor 647 as well as anti-rat Alexa Fluor 488 (Invitrogen)) were applied. For detection of HRP-conjugated antibodies, an ECL Plus reagent (GE Healthcare, Thermo Scientific) was used.

Cell culture, transfection and immunofluorescence staining

HEK 293T and BHK cells were cultured in DMEM supplemented with 10% fetal calf serum and 50 μg/ml gentamycine (PAA). MEF cells were cultured in DMEM supplemented with 15% fetal calf serum, 0.1 mM β-mercaptoethanol (Invitrogen), 2 mM l-glutamine, 1× MEM non-essential amino acids, 100 U/ml penicillin and 100 g/ml streptomycin (PAA). ESCs including J1 wt, *Dnmt1*$^{-/-}$, E14 wt and *Uhrf1*$^{-/-}$ were cultured without feeder cells in gelatinized flasks as described [33]. Culture medium was supplemented with 1 000 U/ml recombinant leukemia inhibitory factor (Millipore). The *Dnmt1*$^{-/-}$ ESCs used in this study are homozygous for the c allele [63]. Mouse E14 wt and *Uhrf1*$^{-/-}$ cells have been reported before [6]. Mouse ESCs and MEF cells were transfected with FuGENE HD (Roche), Lipofectamine® 2 000 or 3 000 reagent (Invitrogen) according to the manufacturer's instructions. HEK 293T cells and BHK cells were transfected using polyethylenimine as transfection reagent (Sigma) according to the manufacturer's instructions. Cell fixation and microscopy were carried as described [33].

Generation of stable ESC lines and DNA methylation analyses

For typing eight hours after expression of GFP-tagged constructs in *Dnmt1*$^{-/-}$ or *Uhrf1*$^{-/-}$ ESCs, GFP-positive mouse ESCs were separated using a fluorescence-activated cell sorting (FACS) Aria II instrument (Becton Dickinson). Stably expressing cells were expanded in selection medium containing 10 μg/ml blasticidin (GFP-DNMT1 wt and GFP-DNMT1 Δ458-500) or 500 ng/ml puromycin (UHRF1-GFP wt, H346G and H730A) and GFP-positive cells were FACS sorted a second time. From the remove, the UHRF1-GFP wt, H346G and H730A cell lines were single-cell sorted. Single clones of GFP-DNMT1 Δ458-500 and corresponding wt [29] were picked manually. For all cell lines, clones with low expression levels were enriched for further analyses. The level and accuracy of the expressed GFP fusion constructs were checked by western blot analyses (Figure 7A and Supplementary information, Figures S1B and S3D). For functional analyses of GFP-DNMT1 wt and GFP-DNMT1 UIM mutants (Δ356-404, D381A-E382A-S392A and D381A-E382A-M385A-S392A-D395A) as well as UHRF1-

GFP wt or UHRF1-GFP point mutants (H346G and H730A) by transient rescue assay, 48 h after expression of these proteins in *Dnmt1*$^{-/-}$ or *Uhrf1*$^{-/-}$ ESCs, respectively, GFP-positive cells were collected with FACS. Genomic DNA isolation, bisulfite conversion and PCR conditions were described before [69, 70]. Primer sets used for amplification of minor satellites, major satellites, *skeletal α-actin* and the *Dnmt1o* promoter are listed in Supplementary information, Table S1. All PCR products were analyzed by pyrosequencing (Varionostic), which results in a quantitative data set for individual CpG sites[70].

Co-immunoprecipitation and western blotting

For co-immunoprecipitation assays, the GFP and RFP, Chor His fusion constructs were co-expressed in HEK 293T cells and protein extracts were normalized to the same GFP or RFP concentration prior to co-immunoprecipitation with the GFP-Trap or RFP-Trap (Chromotek). Bound fractions were first detected by fluorescence intensity measurements and second by western blot analyses.

Acid extraction and TCA precipitation of histones

Histones were isolated by acid extraction as reported previously [71]. In brief, 10^7 mouse ESCs or HEK 293T cells were treated in hypotonic lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT and 1× Protease Inhibitor, 2 mM PMSF) for 30 min and centrifuged at 10 000 *g* at 4 °C to get the intact nuclei. After washing steps, nuclei were resuspended in 0.4 N H_2SO_4 and incubated on a rotator at 4 °C overnight. After centrifugation, histones in the supernatant were transferred into a fresh reaction tube and precipitated using 33% trichloroacetic acid (TCA). After washing twice with cold acetone, histones were dissolved in H_2O. Histone concentrations were measured using the Pierce™ 660 nm protein assay kit (Thermo Scientific).

Ubiquitinated histone binding experiment

For ubiquitinated histone binding experiment, acid extracted histones from HEK 293T cells were used. GFP-DNMT1 and its mutants were immobilized on the GFP-Trap (Chromotek) and incubated with equal amounts of acid extracted histones for 30 min at 4 °C. After washing steps, the bound fractions were analyzed by western blot.

Due to unspecific binding of histones to the eppendorf tubes, we used eppendorf tubes with low binding affinity during mass spectrometry sample preparation.

Immunoprecipitation of ubiquitinated GFP-H3

GFP-H3 wt and R2A mutant constructs were co-expressed in HEK 293T cells with UHRF1-His and harvested after treatment with 2 mM N-ethyl maleimide (NEM, AppliChem) for 5 min. Lysates were prepared by firstly isolating nuclei in hypotonic buffer (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34% glycerol, 0.1 mM EDTA, 0.1 mM DTT, 1× protease inhibitor, 2 mM PMSF, 0.6% NP-40, 0.25 mg/ml NEM) and second lysis of the nuclei in hypertonic buffer (20 mM HEPES pH 7.9, 150 mM KCl, 1.5 mM $MgCl_2$, 10% glycerol, 0.1 mM EDTA, 1 mM DTT, 1× protease inhibitor, 2 mM PMSF, 1 mg/ml DnaseI (AppliChem), 0.25 mg/ml NEM). Prior to immunoprecipitation, the GFP concentration was equalized using lysates from UHRF1-His transfected HEK 293T cells for dilution. After immunoprecipitation of GFP-H3 with the GFP-Trap (Chromotek) and washing (20 mM HEPES pH 7.9, 150 mM KCl, 0.2% glycerol, 0.2% Triton X-100, the bound fraction was analyzed by western blot.

For semiquantitative analysis of the GFP-H3 wt or K18A, K23A, K18A-K23A and R2A ubiquitination, the GFP fusion constructs were co-expressed with HA-ubiquitin in HEK 293T cells and 2 days after transfection, the cells were harvested as described above and underwent the reported sequential IP [72].

F3H assay and trapping assay

The F3H assay was performed as described previously [56]. In the trapping assay, mouse ESCs stably expressing GFP-DNMT1 wt or Δ458-500 were cultured in Ibidi chambers and incubated with 10 μM of the cytosine analogue 5-aza-2′-deoxycytidine (Sigma). Images were acquired with a UltraVIEW VoX spinning disc microscope (PerkinElmer) assembled to an Axio Observer D1 inverted stand (Zeiss) and using a 63×/1.4 NA Plan-Apochromat oil immersion objective.

In vitro peptide pull-down assay

The peptide pull-down assay from nuclear cell extracts of HEK 293T cells expressing UHRF1-GFP fusion constructs was performed as described [72] with the following modifications. C-terminally biotinylated histone peptides were purchased from PSL and are listed in Supplementary information, Table S2. Streptactin beads (IBA) were used for the immobilization of biotinylated peptides in binding buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT). After the binding reaction, beads were washed four times with wash buffer (20 mM HEPES pH 7.9, 0.2% glycerol, 2 mM EDTA, 150 mM KCl, 0.2% Triton X-100). Bound fractions were eluted by boiling in 2× Laemmli sample buffer and analyzed by western blot.

Mass spectrometry

The gel was stained with Coomassie and H3 bands were manually excised, propionylated and digested with trypsin as described before [73] with minor modifications. For peptide extraction, gel slices were incubated twice with 50 μl of 50% acetonitrile/5% TFA and twice more with 50 μl of acetonitrile. The resulting liquid containing the digested peptides was totally evaporated, redissolved with 15 μl of 0.1% formic acid and stored at −20 °C until further processing.

Tryptic peptides were injected (5 μl) in an Ultimate 3 000 HPLC system (LC Packings Dionex). Samples were desalted online in a C18 microbiotrap (300 μm id × 5 mm, packed with C18 PepMap™, 5 μm, 100 Å by LC Packings), and peptides were separated with a gradient from 5% to 60% acetonitrile in 0.1% formic acid over 40 min at 300 nl/min on a C18 analytical column (75 μm id × 15 cm, packed with C18 PepMap™, 3 μm, 100 Å by LC Packings).

The effluent from the HPLC was directly electrosprayed into a linear trap quadrupole-Orbitrap XL mass spectrometer (Thermo Fisher Scientific). The MS instrument was operated in data-dependent mode. Survey full-scan MS spectra (from m/z 300–2 000) were acquired in the Orbitrap with resolution R = 60 000 at m/z 400 (after accumulation to a "target value" of 500 000 in the linear ion trap). The six most intense peptide ions with charge states between two and four were sequentially isolated to a target value of 10 000 and fragmented by collision-induced dissociation and recorded in the linear ion trap. For all measurements with the Orbitrap detector, three lock-mass ions were used for internal calibration [74].

Typical MS conditions were spray voltage, 5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 200 °C; normalized collision-induced dissociation energy 35%; activation $q = 0.25$; and activation time = 30 ms.

Mascot 2.3.02 was used for protein identification with the following settings: Database: Swissprot 57.7; Taxonomy: *Homo sapiens* (human); MS tolerance: 10 ppm; MS/MS tolerance: 0.6 Da; peptide FDR: 0.01; protein FDR: 0.01; minimum peptide length 5; and variable modifications: propionyl (K, N-term), GlyGly (K).

Quantification of modified H3 18-26 peptides was based on the intensities of the MS1 peaks. The spectra depicted in Figure 4A and 4B were used to determine the exact masses ($m/z \pm 10$ ppm) and used as a reference for further quantification.

Quantitative analysis of DNMT1 subnuclear localization

During late S phase, DNMT1 is enriched in replication foci at chromocenters. In order to quantify the subnuclear distribution of GFP-DNMT1 wt and defined UIM mutants the following procedure was used: confocal z-stacks (100 µm interval) were acquired with identical scan settings in three color channels to visualize replication foci (anti-PCNA staining, 594 nm excitation), DNMT1 localization (GFP-DNMT1 fusions enhanced with GFP-booster (Chromotek), 488 nm excitation) and DNA counterstaining (DAPI, 405 nm excitation). For each color channel, maximum intensity projections were calculated and only GFP-expressing cells were analyzed. Segmentation of replication foci or whole nuclei was performed with the Weka segmentation plugin [75] in Fiji [76]. Training of the classifier was finalized until the result matched the visual impression (Figure 7A). Due to variations in ESC samples, replication foci were segmented using different classifiers for wt or the different UIM mutants. In contrast, for all somatic cells, one classifier was sufficient to segment replication foci. Whole nuclei were segmented by a classifier based on the DAPI signal. After Weka segmentation, the resulting binary masks were filtered using the particle analyzer of Fiji with a circularity value ≥ 0.25. To select for cells in late S phase, only replication foci ≥ 150 pixel where further analyzed in the ESC samples. In MEF cell samples, only late S phase cells were imaged and analyzed without applying size exclusion for replication foci. Nuclear masks (size ≥ 3 000 pixel) were used to quantify the total amount of GFP fusion protein in a single nucleus. Nuclei were further subsegmented by replication foci masks. For each nucleus, the ratio between the mean GFP signals in replication foci relative to the mean GFP signal outside the foci was calculated. Raw data were corrected for background signals by subtracting the modal grey value. Ratios from all nuclei expressing GFP-DNMT1 wt or UIM mutants were visualized as boxplots. Numerical calculations and statistical analysis were performed with R [77].

Statistical analysis

Results were expressed as mean values ± SD or as mean values ± SEM from the number of biological replicates indicated in the corresponding figure legend.

Acknowledgments

We thank Stefan Jentsch (Max Planck Institute of Biochemistry, Germany) for providing the HA-ubiquitin construct and Peter Becker (Adolf Butenandt Institute, Germany) for the 601 DNA construct. We are grateful to the following colleagues for providing ESCs and somatic cells: Masahiro Muto and Harikesh Kohe for mouse E14 wt and *Uhrf1*$^{-/-}$ ESCs; En Li and T. Chen for mouse J1 wt and *Dnmt1*$^{-/-}$; Thomas Jenuwein for MEF cells; and L. David Spector for providing BHK cells containing a lacoperator repeat array. We thank E.M. Baur (Ludwig Maximilians University, Germany) for technical help with the GFP-TS UIM point mutant plasmid constructs. This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG, SFB646 A7 to HL and Z03 to AI), the Nanosystem Initiative Munich (NIM to HL) and the Epigenomics Flagship Project (EPIGEN-CNR -IT to IMB). KF and GP were supported by the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS)PWN Lander Sarefellow sof the Graduate School Life Science Munich (LSM). MS is a fellow of the Integrated Research Training Group (IRTG) of the SFB1064. NL and WQ were also supported by the China Scholarship Council (CSC).

References

1 Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002; **16**:6-21.
2 Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 1992; **71**:865-873.
3 Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* 1997; **277**:1996-2000.
4 Schneider K, Fuchs C, Dobay A, *et al.* Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. *Nucleic Acids Res* 2013; **41**:4860-4876.
5 Easwaran HP, Schermelleh L, Leonhardt H, Cardoso MC. Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep* 2004; **5**:1181-1186.
6 Schermelleh L, Haemmer A, Spada F, *et al.* Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res* 2007; **35**:4301-4312.
7 Spada F, Haemmer A, Kuch D, *et al.* DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J Cell Biol* 2007; **176**:565-571.
8 Margot JB, Aguirre-Arteta AM, Di Giacco BV, *et al.* Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. *J Mol Biol* 2000; **297**:293-300.
9 Fellinger K, Rothbauer U, Felle M, Langst G, Leonhardt H. Dimerization of DNA methyltransferase 1 is mediated by its regulatory domain. *J Cell Biochem* 2009; **106**:521-528.
10 Syeda F, Fagan RL, Wean M, *et al.* The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *J Biol Chem* 2011; **286**:15344-15351.
11 Takeshita K, Suetake I, Yamashita E, *et al.* Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proc Natl Acad Sci USA* 2011; **108**:9055-9059.
12 Achour M, Jacq X, Ronde P, *et al.* The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene*

2008; **27**:2187-2197.
13 Felle M, Joppien S, Nemeth A, *et al.* The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. *Nucleic Acids Res* 2011; **39**:8355-8365.
14 Berkyurek AC, Suetake I, Arita K, *et al.* The DNA Methyltransferase Dnmt1 directly interacts with the SET and RING finger associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. *J Biol Chem* 2013; **289**:379-386.
15 Zhang J, Gao Q, Li P, *et al.* S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. *Cell Res* 2011; **21**:1723-1739.
16 Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 2007; **317**:1760-1764.
17 Sharif J, Muto M, Takebayashi S, *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 2007; **450**:908-912.
18 Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992; **69**:915-926.
19 Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* 2008; **455**:818-821.
20 Avvakumov GV, Walker JR, Xue S, *et al.* Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* 2008; **455**:822-825.
21 Qian C, Li S, Jakoncic J, Zeng L, Walsh MJ, Zhou MM. Structure and hemimethylated CpG binding of the SRA domain from human UHRF1. *J Biol Chem* 2008; **283**:34490-34494.
22 Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, Leonhardt H. The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res* 2010; **38**:1796-1804.
23 Cheng J, Yang Y, Fang J, *et al.* Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. *J Biol Chem* 2013; **288**:1329-1339.
24 Citterio E, Papait R, Nicassio F, *et al.* Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol Cell Biol* 2004; **24**:2526-2535.
25 Karagianni P, Amazit L, Qin J, Wong J. ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Mol Cell Biol* 2008; **28**:705-717.
26 Xie S, Jakoncic J, Qian C. UHRF1 double tudor domain and the adjacent PHD finger act together to recognize K9me3-containing histone H3 tail. *J Mol Biol* 2012; **415**:318-328.
27 Papait R, Pistore C, Grazini U, *et al.* The PHD domain of Np95 (mUHRF1) is involved in large-scale reorganization of pericentromeric heterochromatin. *Mol Biol Cell* 2008; **19**:3554-3563.
28 Jenkins Y, Markovtsov V, Lang W, *et al.* Critical role of the ubiquitin ligase activity of UHRF1, a nuclear RING finger protein, in tumor cell growth. *Mol Biol Cell* 2005; **16**:5621-5629.
29 Qin W, Leonhardt H, Spada F. Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J Cell Biochem* 2011; **112**:439-444.
30 Du Z, Song J, Wang Y, *et al.* DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. *Sci Signal* 2010; **3**:ra80.
31 Mudbhary R, Hoshida Y, Chernyavskaya Y, *et al.* UHRF1 overexpression drives DNA hypomethylation and hepatocellular carcinoma. *Cancer Cell* 2014; **25**:196-209.
32 Nishiyama A, Yamaguchi L, Sharif J, *et al.* Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* 2013; **502**:249-253.
33 Schermelleh L, Spada F, Easwaran HP, *et al.* Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat Methods* 2005; **2**:751-756.
34 Liu X, Gao Q, Li P, *et al.* UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat Commun* 2013; **4**:1563.
35 LDambacher S, Deng W, Hahn M, *et al.* CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. *Nucleus* 2012; **3**:101-110.
36 Zolghadr K, Mortusewicz O, Rothbauer U, *et al.* A fluorescent two-hybrid assay for direct visualization of protein interactions in living cells. *Mol Cell Proteomics* 2008; **7**:2279-2287.
37 Hu L, Li Z, Wang P, Lin Y, Xu Y. Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. *Cell Res* 2011; **21**:1374-1378.
38 Rajakumara E, Wang Z, Ma H, *et al.* PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Mol Cell* 2011; **43**:275-284.
39 Wang C, Shen J, Yang Z *et al.* Structural basis for site-specific reading of unmodified R2 of histone H3 tail by UHRF1 PHD finger. *Cell Res* 2011; **21**:1379-1382.
40 Frauer C, Leonhardt H. A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucleic Acids Res* 2009; **37**:e22.
41 Bestor TH, Ingram VM. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc Natl Acad Sci USA* 1983; **80**:5559-5563.
42 Yoder JA, Soman NS, Verdine GL, Bestor TH. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J Mol Biol* 1997; **270**:385-395.
43 Jeltsch A. On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics* 2006; **1**:63-66.
44 Bashtrykov P, Jankevicius G, Jurkowska RZ, Ragozin S, Jeltsch A. The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. *J Biol Chem* 2014; **289**:4106-4115.
45 Rothbart SB, Dickson BM, Ong MS, *et al.* Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. *Genes Dev* 2013; **27**:1288-1298.
46 Rothbart SB, Krajewski K, Nady N, *et al.* Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat Struct Mol Biol* 2012; **19**:1155-1160.
47 Song J, Rechkoblit O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in main-

tenance DNA methylation. *Science* 2011; **331**:1036-1040.
48 Kalb R, Latwiel S, Baymaz HI, *et al.* Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* 2014; **21**:569-571.
49 Leeb M, Wutz A. RING1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J Cell Biol* 2007; **178**:219-229.
50 Blackledge NP, Farcas AM, Kondo T, *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* 2014; **157**:1445-1459.
51 Vire E, Brenner C, Deplus R, *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 2006; **439**:871-874.
52 Cooper S, Dienstbier M, Hassan R, *et al.* Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell reports* 2014; **7**:1456-1470.
53 Kurdistani SK, Tavazoie S, Grunstein M. Mapping global histone acetylation patterns to gene expression. *Cell* 2004; **117**:721-733.
54 Tsai WW, Wang Z, Yiu TT, *et al.* TRIM24 links a non-canonical histone signature to breast cancer. *Nature* 2010; **468**:927-932.
55 Wang Z, Zang C, Rosenfeld JA, *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 2008; **40**:897-903.
56 Raiborg C, Stenmark H. The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature* 2009; **458**:445-452.
57 Lim J, Son WS, Park JK, Kim EE, Lee BJ, Ahn HC. Solution structure of UIM and interaction of tandem ubiquitin binding domains in STAM1 with ubiquitin. *Biochem Biophys Res Commun* 2011; **405**:24-30.
58 Hirano S, Kawasaki M, Ura H, *et al.* Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. *Nat Struct Mol Biol* 2006; **13**:272-277.
59 Wang Q, Young P, Walters KJ. Structure of S5a bound to monoubiquitin provides a model for polyubiquitin recognition. *J Mol Biol* 2005; **348**:727-739.
60 Frauer C, Rottach A, Meilinger D, *et al.* Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS One* 2011; **6**:e16627.
61 Meilinger D, Fellinger K, Bultmann S, *et al.* Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep* 2009; **10**:1259-1264.
62 Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 1989; **77**:51-59.
63 Pichler G, Wolf P, Schmidt CS, *et al.* Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem* 2011; **112**:2585-2593.
64 Rothbauer U, Zolghadr K, Tillib S, *et al.* Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat Methods* 2006; **3**:887-889.
65 Herce HD, Deng W, Helma J, Leonhardt H, Cardoso MC. Visualization and targeted disruption of protein interactions in living cells. *Nat Commun* 2013; **4**:2660.
66 Rottach A, Kremmer E, Nowak D, *et al.* Generation and characterization of a rat monoclonal antibody specific for PCNA. *Hybridoma (Larchmt)* 2008; **27**:91-98.
67 Rottach A, Kremmer E, Nowak D, Leonhardt H, Cardoso MC. Generation and characterization of a rat monoclonal antibody specific for multiple red fluorescent proteins. *Hybridoma (Larchmt)* 2008; **27**:337-343.
68 Lei H, Oh SP, Okano M, *et al.* De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 1996; **122**:3195-3205.
69 Tucker KL, Beard C, Dausmann J, *et al.* Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev* 1996; **10**:1008-1020.
70 Tost J, Gut IG. DNA methylation analysis by pyrosequencing. *Nat Protoc* 2007; **2**:2265-2275.
71 Shechter D, Dormann HL, Allis CD, Hake SB. Extraction, purification and analysis of histones. *Nat Protoc* 2007; **2**:1445-1457.
72 Wysocka J. Identifying novel proteins recognizing histone modifications using peptide pull-down assay. *Methods* 2006; **40**:339-343.
73 Villar-Garea A, Israel L, Imhof A. Analysis of histone modifications by mass spectrometry. *Curr Protoc Protein Sci* 2008; Chapter 14:Unit 14.10.
74 Olsen JV, de Godoy LM, Li G, *et al.* Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* 2005; **4**:2010-2021.
75 Mark Hall EF, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* 2009; **11**:10-18.
76 Schindelin J, Arganda-Carreras I, Frise E, *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012; **9**:676-682.
77 R Core Team (2014). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
78 Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 2014; **42**:W320-W324.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)

This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0

2.5.2 Supplementary information

Supplementary information, Data S1

Materials and methods

Antibodies

For detection of (GFP-)DNMT1 by western blot, a mouse anti-DNMT1 antibody (pATH52 [1, 2]) was used. Equal loading was confirmed by immunoblotting with a specific anti-Lamin B1 (Abcam) or anti-β-Actin (Sigma) antibody. The rabbit anti-H2A antibody was purchased from Millipore.

In vitro DNA methylation assay

For analyses of *in vitro* DNA methylation activity, GFP-DNMT1 was purified by immunoprecipitation from HEK 293T extracts. The concentration of GFP-DNMT1 in the bound fractions was measured by fluorescent read out. In order to get enough unmodified DNA templates, the 601 DNA sequences were amplified with the primers: TGCATGTATTGAACAG (forward) and TGCACAGGATGTATATATC (reverse). 3 µg of GFP-DNMT1 were incubated with 88 ng of DNA template in methylation buffer containing 160 µM SAM and 100 ng/µl BSA at 37°C for 3 hours. After inactivation of the reaction at 65°C for 30 min, the DNA was isolated with a Nucleospin PCR cleaning kit (Macherey-Nagel) and bisulfite treated with EZ DNA Methylation-Gold Kit (Zymo research). Primer sequences for the 601 DNA were TGTATGTATTGAATAG (forward primer) and TACACAAAATATATATATC (reverse primer). For amplification we used Qiagen Hot Start Polymerase in 1x Qiagen Hot Start Polymerase buffer supplemented with 0.2 mM dNTPs, 0.2 µM forward primer, 0.2 µM reverse primer, 1.3 mM Betaine (Sigma) and 60 mM Tetramethylammonium-chloride (TMAC, Sigma). Pyrosequencing reactions were carried out by Varionostic GmbH (Ulm).

Preparation of hemimethylated DNA substrates

To prepare the hemimethylated DNA, an efficient method for long heteroduplex DNA was used as described [3]. One pair of modified PCR primers were synthesized, which are labeled with phosphate at 5'-end, 5'-phosphorylated-TGCATGTATTGAACAG-3' and 5'-phosphorylated-TGCACAGGATGTATATATC-3'. To get single and upper strand DNA, the DNA was amplified with the reverse primer labeled with phosphate at the 5'-end, following a lambda-nuclease digestion (NEB). The same procedure is required for making lower strand DNA. To prepare the methylated lower strand DNA, one more step, *in vitro* methylation by bacterial methyltransferase M.SssI (NEB), is required before treatment with the lambda-nuclease. In the end, equal amounts of upper and lower strand DNA were mixed and incubated at 95°C for 5 min, followed by annealing. To get rid of contamination from double strand DNA after lambda-nuclease treatment, the hydroxyapatite chromatography was carried out. Hydroxyapatite column (Sigma) was packed according to the manufacturer's instructions and the single stranded DNA was eluted by elution buffer containing 150 mM sodium phosphate.

In vitro DNA binding assays

In vitro DNA binding assays were performed as described previously [4]. Briefly, two double stranded DNA oligonucleotides labeled with different ATTO fluorophores were used as substrates in direct competition. DNA oligonucleotide substrates with identical sequence contained an unmodified or hemimethylated cytosine at a single, central CpG site (UMB: unmethylated binding substrate, ATTO550; HMB: hemimethylated binding substrates, ATTO647N; Supplementary Table S3). GFP fusion proteins were expressed in HEK 293T cells and immunoprecipitated using the GFP-Trap (Chromotek). Immobilized UHRF1-GFP wt and mutants were washed three times before incubation with DNA substrates at a final concentration of 160 nM each. After removal of unbound substrates, protein amounts (GFP fluorescence) and bound DNA were measured with an Infinite M1000 plate reader (Tecan).

In vivo autoubiquitination assay

The *in vivo* autoubiquitination assay of UHRF1-GFP was performed as described before [5]. The resulting ubiquitination levels were detected with a specific mouse monoclonal anti-HA antibody (12CA5) and quantified using Image J and a statistical Student's t-test analysis. Equal amounts of (UHRF1-)GFP in the bound fraction were verified by immunoblotting with a specific anti-GFP antibody (Chromotek).

Slot blot analysis

To quantify global DNA methylation levels, the Bio-Rad slot blot system was used according to the manufacturer's instruction. Prior to loading on a Nitrocellulose membrane (Amersham), genomic DNA was denatured in 6x SSC buffer for 10 min at 95°C and incubated for 10 min on ice. The membrane was crosslinked, blocked with 5% milk and immunostained with specific rabbit anti-ssDNA (Eurogentec) and mouse anti-5mC (IBL) antibodies. Quantification was performed using the ImageJ gel analysis tool.

In vitro histone tail peptide binding assay

The *in vitro* histone tail peptide binding assay was performed as described before [6] with the following modification. GFP fusion proteins were equalized to a GFP concentration of 130 nM prior to immunoprecipitation with the GFP-Trap. The TAMRA-labeled H3 peptides used in this assay are listed in Supplementary Table S2.

Ubiquitinated histone H3 binding experiment

For ubiquitinated H3 binding experiment, HEK 293T cells were incubated with 2 mM N-Ethylmaleimide (NEM, AppliChem) for 10 min before harvesting and were treated in hypotonic buffer (10 mM Tris-HCl pH 8, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT and 1x Protease Inhibitor, 2 mM PMSF) for 10 min on ice to isolate the intact nuclei. Nuclei resuspended in MNase digestion buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $CaCl_2$, 0.1% NP-40, 1x Protease Inhibitor (Serva), 2 mM PMSF) were digested with 40 U/ml MNase at 37°C for 5 min to get mononucleosomes.

GFP-DNMT1 and its mutants were immobilized on the GFP-Trap and incubated with equal amount of mononucleosomes for 2 hours. After washing steps, the bound fractions were analyzed by western blot.

Supplementary references

1 Bestor TH. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J* 1992; **11**:2611-2617.

2 Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992; **69**:915-926.

3 Thomas E, Pingoud A, Friedhoff P. An efficient method for the preparation of long heteroduplex DNA as substrate for mismatch repair by the Escherichia coli MutHLS system. *Biol Chem* 2002; **383**:1459-1462.

4 Frauer C, Leonhardt H. A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucleic Acids Res* 2009; **37**:e22.

5 Qin W, Leonhardt H, Spada F. Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J Cell Biochem* 2011; **112**:439-444.

6 Pichler G, Wolf P, Schmidt CS et al. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem* 2011; **112**:2585-2593.

7 Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 2014; **42**:W320-324.

8 Hirano S, Kawasaki M, Ura H et al. Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. *Nat Struct Mol Biol* 2006; **13**:272-277.

Chapter 3

Discussion

Applied methodological work in biological studies, such as microscopy and image analysis, necessarily takes a second place after the new findings that are discovered. Nevertheless, it is my hope to have helped my co-authors arrive at conclusions that they would not have reached so easily otherwise. I worked on the development of new methods as well, providing tools to drive scientific investigation in the years to come. On the following pages, I will provide a quick discussion of the studies included in this thesis and my contributions to each of them and end with some general thoughts on the current and future role of computational tools in microscopy-based science.

3.1 Localization of the contributions in the overall experimental workflow

The work pursued during the work on this thesis spans a wide range in the generalized experimental workflow shown in the introduction (figure 1.1), both in the stage of the experiment at which I contributed, but also on the experimentation-computation range (figure 3.1). It can be grouped into two categories, the studies

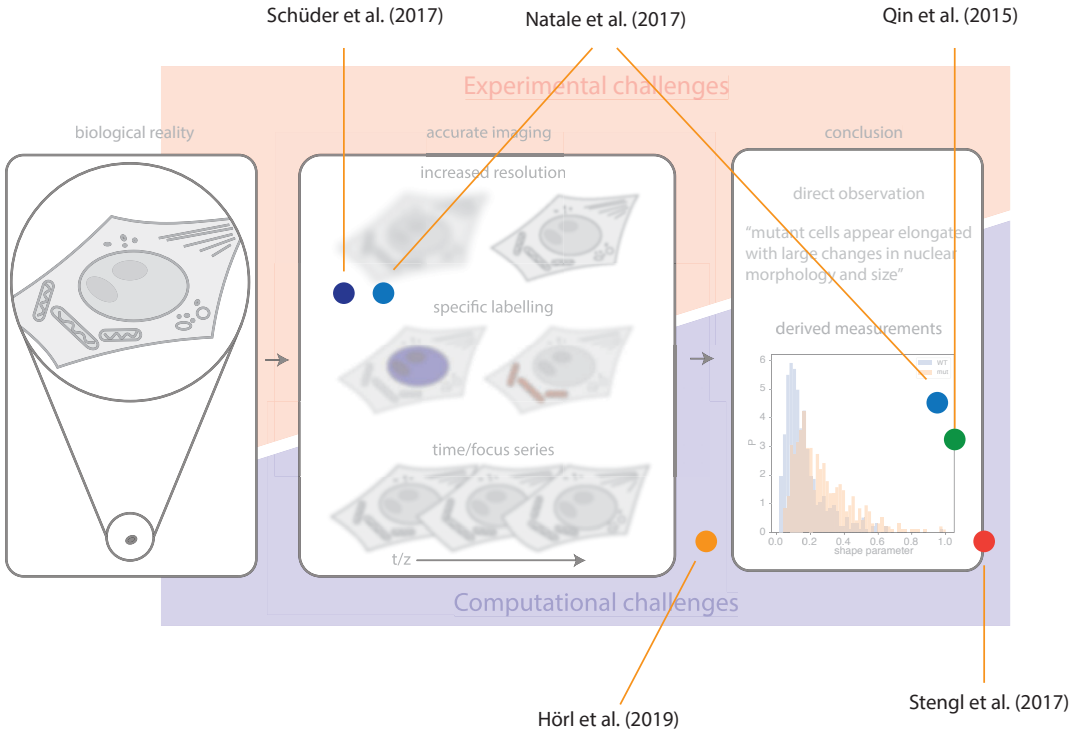


Figure 3.1: Location of the work presented in this thesis in the overall workflow of imaging-based experiments.

in which I contributed to the discovery of novel epigenetic and cell biological insights via microscopy and image analysis [147, 148], helping colleagues to study complex phenomena in nuclear cell biology, and those in which I contributed to methods development [21, 82, 89], benchmarking new labelling strategies and writing tools for quick screening of therapeutic antibodies for cancer treatment and reusable tools for the processing of terabyte-sized image datasets. On the following pages, I will discuss biological studies and methodological developments separately, with a dedicated section for the central work presented in this thesis, BigStitcher [89]

3.2 Novel insights into epigenetics and cell biology

Two of the studies presented in this thesis produced novel biological findings, specifically concerning the regulation of the DNA methyltransferase 1 (DNMT1) as well as spatial distribution of DNA damage response upon exposure to radiation, two areas that have been of central interest in the work of the Leonhardt lab and the collaborating Cardoso lab for a long time. Going into great depth here would overstate my contributions that, while forming an essential part of the whole, constitute single steps towards the final result. Therefore, the overall conclusions will be summarized briefly, while elaborating upon those results derived directly from imaging experiments with direct contribution from me.

3.2.1 DNMT1

The canonical role of DNMT1 is that of a maintenance DNA methyltransferase, transferring a methyl residue to cytosines at hemi-methylated cytosine-guanine dinucleotide (CpG) sites directly after replication. There are a variety of regulatory mechanisms controlling the activity of DNMT1 in vivo, through its N-terminal regulatory part, like association with the replication machinery through a proliferating cell nuclear antigen (PCNA)-binding domain or binding via a target sequence (TS) to ubiquitin-like containing PHD and RING finger domains 1 (UHRF1), which itself harbors complex regulatory (binding to a variety of chromatin features, like e.g. hemi-methylated CpG) and enzymatic (e.g. Ubiquitin-ligase) activity. In the study presented above [148], we could add another piece to the regulatory puzzle by discovering and characterizing an Ubiquitin interacting motif (UIM) within the TS that allows DNMT1 to bind to H3 tails ubiquitinated by UHRF1.

My contribution to the study consisted in quantifying the subnuclear localization of DNMT1 mutants, especially their association with late replication sites. To this end, confocal stacks of S-phase nuclei of fixed cells, both mouse embryonal fibrob-

lasts (MEFs) and J1 embryonal stem cells (ESCs), transiently expressing DNMT1 mutants as GFP-fusions were acquired by one of the main authors of the study, Dr. Patricia Wolf. The whole nuclei were labelled with DAPI and replication sites via immunolabelling with an anti-PCNA antibody. For the segmentation of nuclei as well as replication sites, we turned to machine learning-based methods, specifically the user-friendly Trainable Weka Segmentation (TWS) plug-in in ImageJ/Fiji [139]. This constituted a large step forward in the sophistication of segmentation techniques used in the lab, which was previously carried out by simple thresholding, sometimes with manual threshold selection. A workflow was set up in which the main author would interactively generate segmentation models in TWS, followed by simple morphological operations to remove small objects and identify connected components and automated intensity measurements from the masked areas. Used in this way, ML can be seen as a mere automation aid, propagating expert knowledge from a few labelled examples to the rest of the dataset. However, this is not necessarily a bad thing: for one, almost every state-of-the-art specialized AI approach shares this property, except the expert contribution is hidden in large datasets and usually not exposed in production ML applications outside of science. Furthermore, the direct observation of segmentation results provided by interactive tools such as TWS almost certainly entails manual quality control by the researcher doing the training, a step that might easily be missed in fully automated pipelines relying, for example, on simple automated thresholding that just spit out a collection of numbers at the end, never showing intermediate segmentations to the user. The (in my opinion) great importance of interactivity will be discussed further below. From the nuclear and replication masks generated with our semi-automatic pipeline, we could determine average fluorescence intensity in the spots, corrected for nuclear background and show significant differences between mutants in a downstream statistical analysis.

Using this relatively simple analysis pipeline in both J1 mouse embryonal stem cells and MEFs, we could in both cell types show significantly reduced association with late replicating heterochromatin of a Dnmt1-UIM deletion mutant as well as point mutants of conserved amino acids within the UIM, complementing other lines of evidence produced by colleagues in the study.

3.2.2 DNA damage response

The goal of the study presented in [147] was a comprehensive and quantitative description of the dynamics and spatial distribution of DNA damage response after double strand breaks (DSBs) introduced via X-ray irradiation or enzymatic activity. An early cellular response to DSBs is the accumulation of H2A histone family member X phosphorylated on serine 139 (γH2AX) at damage sites. In our study, we systematically quantified the spatial distribution of γH2AX sites and their interaction with other chromatin features by super-resolution fluorescence microscopy and biochemical assays.

One contribution to the project consisted of verifying that the structures visible in SIM micrographs indeed correspond to single repair sites and not just aggregates with non-resolvable substructures. To this end, we applied the higher-resolution microscopy technique of STED. Based on STED images acquired by me and comparing them to SIM images acquired by co-authors, we measured the size of γH2AX foci in HeLa cells. We could show that the sizes do not change significantly in unirradiated cells and that they are significantly smaller during DNA damage response, though only by a small factor less than the expected resolution gain of STED versus SIM, indicating that we have indeed imaged elementary units of DNA replication (apart from potential nanoscale fine structures).

The second contribution made to the study was in the form of confocal and STED imaging of heterochromatin after the introduction of DSBs by a clustered reg-

ular interspaced short palindromic repeats (CRISPR)/Cas9 system, followed by image analysis similar to the pipeline in [148]. In short, Cas9 was used in concert with a major satellite gRNA to introduce DSB in heterochromatic regions of murine myoblast C2C12 cells. Cells were fixed at set times after transfection and imaged using confocal or STED microscopy. The goal was twofold: first, to measure γH2AX accumulation at heterochromatic sites and second, to quantify a de-condensation, visible as fraying of heterochromatic chromocenters. Using a joint nucleus-chromocenter segmentation based on ML-based pixel classification (similar to the approach used in [148], with the same included manual quality control), we could indeed show γH2AX accumulation at heterochromatic sites (immunolabelled for the mark H2K9me3) in confocal images and a de-condensation of chromocenters, measured through the proxy of circularity ($4\pi\frac{area}{perimeter^2}$) in STED images.

In the study presented above, we could show, among others, that γH2AX forms nano-domain of distinct sizes that are flanked by CTCF and form clusters around single DSBs and a temporal trend of DNA damage response from euchromatin to heterochromatin. Interestingly, while heterochromatin undergoing repair seems to undergo de-condensation, it retains repressive features, such as the heterochromatin mark histone 3 trimethylated on lysine 9 (H3K9me3), which constitutes a de-coupling of the three-dimensional chromatin organization and linear identifying features along the chromosomes. Our findings rely on two types of super-resolution microscopy, state-of-the-art image analysis and biochemical assays, and the synthesis of multiple approaches allowed us to arrive at our conclusions.

3.3 Methods development

The other studies presented above [21, 82, 89], while not concerned with new discoveries themselves, consisted of the development of new methods that can serve

as a stepping stone for novel discoveries to come. In two of them, my contribution can again be seen as one small contribution to others' ideas. Like above, I will summarize these projects briefly before elaborating upon my own contributions in more detail.

3.3.1 Super-Resolution multiplexing by DNA exchange

Acquiring a complete image of a biological system can often entail measuring the spatial distribution of a large number of distinct biomolecules in one sample. However, simultaneous labeling with different fluorophores quickly approaches a limit, as it becomes harder and harder to distinguish the channels due to spectral overlap. Methods that promise the ability to perform unlimited multiplexing thus typically rely on sequential imaging of different targets in a sample with a small set of detection labels, e.g. fluorophores. Strategies for sequential labelling and imaging include antibody labeling in many rounds, with dissociation or bleaching of the fluorophores introduced in the previous round [149], or Exchange-PAINT [92], in which transiently binding, fluorophore-labelled DNA oligomers are introduced into the buffer and exchanged to sequentially image targets labelled with complimentary oligomers. The approach presented in our paper [21] is based on a similar strategy of reversible DNA hybridization, though with separate labeling, imaging and label removal (washing) phases using longer oligomers that are removed by denaturing agents and not stochastic dissociation. This sequential labeling via relatively stable DNA hybridization provides for an easy exchange of fluorophores, but also circumvents the background signal present in approaches based on transient binding. The applicability of the strategy for SIM, STED and (d)STORM super-resolution imaging was demonstrated in the study.

As part of the study, I was responsible for testing the suitability of the method for STED microscopy, imaging both DNA origami and cells in which up to four proteins were labelled with docking strand-conjugated antibodies, working closely

with the first author, Florian Schüder. We could demonstrate that the method, having previously been demonstrated with STORM imaging, could indeed be used with STED imaging as well, requiring minimal modifications (essentially just slightly different imaging buffers and imager strands labelled with the STED-compatible dye ATTO647N).

While the work contributed to DNA Exchange-based multiplexing can be seen as just a simple proof-of-principle on yet another microscope, it is also of special interest to STED imaging, as it opens up a way around several of the key limitations of STED microscopy. Due to the high intensity depletion light used, STED suffers from high bleaching rates and thus only a relatively small set of fluorescent dyes can be employed effectively with the technique. This severely limits the compatibility with three-dimensional imaging of large samples (as fluorophores in deeper sections will often be bleached prematurely by the imaging of preceding sections) and of many targets in parallel (due to the limited selection of fluorophores, with many of the most stable ones in the same red part of the spectrum).

Our strategy circumvents the limitation of STED to a few performant dyes via the core functionality of spectral multiplexing, but it can also be used to image large samples even if bleaching occurs. While our protocol exchanges the labelling buffer for an imager strand-free imaging buffer and thus does not supply a replenishable label pool during imaging, it could easily be adapted to incorporate multiple rounds of exchange with the same imager strand, each time replacing bleached fluorophores. Care needs to be taken to not cause mechanical movement of the sample during buffer exchange, e.g. by touching the chamber with pipettes, a problem that can be remedied with mechanized microfluidics systems [150] or computational alignment of the images (see section 1.3.4).

3.3.2 Therapeutic antibody characterization

Drug development often entails high-content analysis to screen a multitude of potential therapeutic agents, which is usually carried out using automated screening instruments such as flow cytometers or automated microscopes. The study presented above [82] was focused on the development of a robust assay to quantify the antiproliferative potential of therapeutics based on monoclonal antibodies that were site-specifically modified with a cytotoxic agent. For antiproliferative agents, simple flow cytometry, with its highly reduced readout, is often unable to distinguish between apoptotic, mitotically arrested and proliferating cells, which is why we turned to an image-based assay based on high-content microscopy.

Working with pre-segmented nuclei from a Perkin-Elmer Operetta high content screening (HCS) platform, we developed two models that could be fit to those already very derivative measurements. The first consisted of a statistical model for cell cycle distribution that was fit to the DAPI intensities in the segmented nuclei using global optimization via simulated annealing. From the probability distribution fitted to measurements from wells with varying antibody concentration, we could easily derive a trend towards G2-phase depletion after exposure to higher antibody concentrations. The second line of evidence in our assay consists of the fit of sigmoid inhibition curves to the (drug concentration, normalized cell count)-pairs. By basing the cell count on just a DNA stain, we would again run into problems differentiating proliferating, arrested and apoptotic cells, a problem we remedied by supplementing the medium with EdU, a base analog that gets incorporated into the chromatin during replication and which can be labelled by a click reaction after fixation, and basing our cell count on EdU-positive (i.e. replicating) cells.

While we benchmarked the performance of our assay using the established antiproliferative antibody Trastuzumab, the rationale behind the development was

to create a flexible tool to characterize newly developed antibody-drug conjugates (ADCs) developed by the biotechnology startup Tubulis, a spinoff of the Leonhardt lab. Indeed, the tool already saw application in the R&D of Tubulis [151], being used to evaluate the antiproliferative potency of ADCs generated from Trastuzumab and Brentuximab via novel conjugation strategies developed by the company.

3.4 BigStitcher

The rest of the discussion will focus on a first-author study of mine, dealing with the development of the image alignment tool BigStitcher [89]. A large portion of my time over the last four years went into the development of BigStitcher, a well-received [81, 152, 153], user-friendly tool for alignment, pre-processing and management of very large, multidimensional image datasets, as are produced by LSFM, for example. It has already seen widespread adaptation in a handful of studies and newly developed methods, taking an important place between raw data acquisition and downstream analysis.

3.4.1 Unique challenges of large LSFM data met by BigStitcher

While modern microscopy modalities such as light-sheet microscopy allow for the easy acquisition of very large many-dimensional datasets, these data are often not immediately amenable to analysis due to their large size and dimensional complexity. Dedicated tools for data management and preprocessing are therefore essential to make data tractable for scientists wanting to perform downstream computational analyses on them. In the BigStitcher software, a tool for the management, alignment and visualization of such datasets, implemented as an open source ImageJ/Fiji [154] plug-in, my co-authors and I tried to provide a user-friendly way of handling these datasets.

A first major issue that a tool for a task as general as image alignment has to deal with if it is to gain widespread adoption is the fact that users with a variety of different microscopes will try to use it. Unfortunately, even though attempts at standardization exist [155], many microscopes produce images in proprietary, vendor-specific formats. In the open ImageJ ecosystem, handling of the various data formats is unified in the BioFormats library [155], but the resulting parts of datasets might still differ greatly in their dimensionality, from individual planes to whole multi-image datasets stored in one big array. A first service that BigStitcher provides is the (partially automated, figure 3.2) combination of a set of input files into a common representation of three dimensional stacks indexed by properties such as time, spectral channel, acquisition angle, acquisition position or illumination direction, using BioFormats to avoid having to needlessly re-save data to an intermediate format. To achieve performant processing of the data, we turn to algorithms that work on downsampled versions of the images again and again. A lot of time is lost just loading the data before downsampling, however. It is for that reason that we offer (and advertise) the option to calculate multi-resolution pyramids of the images once and save all levels to open, general-purpose hierarchical data formats such as HDF5 [131] and N5 [156]. That way, if a downsampled version of an image is required later in the pipeline, we can just open the precomputed layer from the pyramid, avoiding serious input-output (IO) bottlenecks.

After a dataset has been properly defined, the main functionality of BigStitcher lies in image alignment. While this is a common and largely solved problem in computer vision, the large size of the datasets and their three-dimensional structure place specific demands on the alignment tools, which we met by a variety of optimizations in BigStitcher. Two main problems arising from the large image sizes are increased computation times, but also the inability to fit all required raw images and intermediate results, such as FFTs or filter responses, into the main memory (RAM) of the computer used (a problem that is compounded when multi-

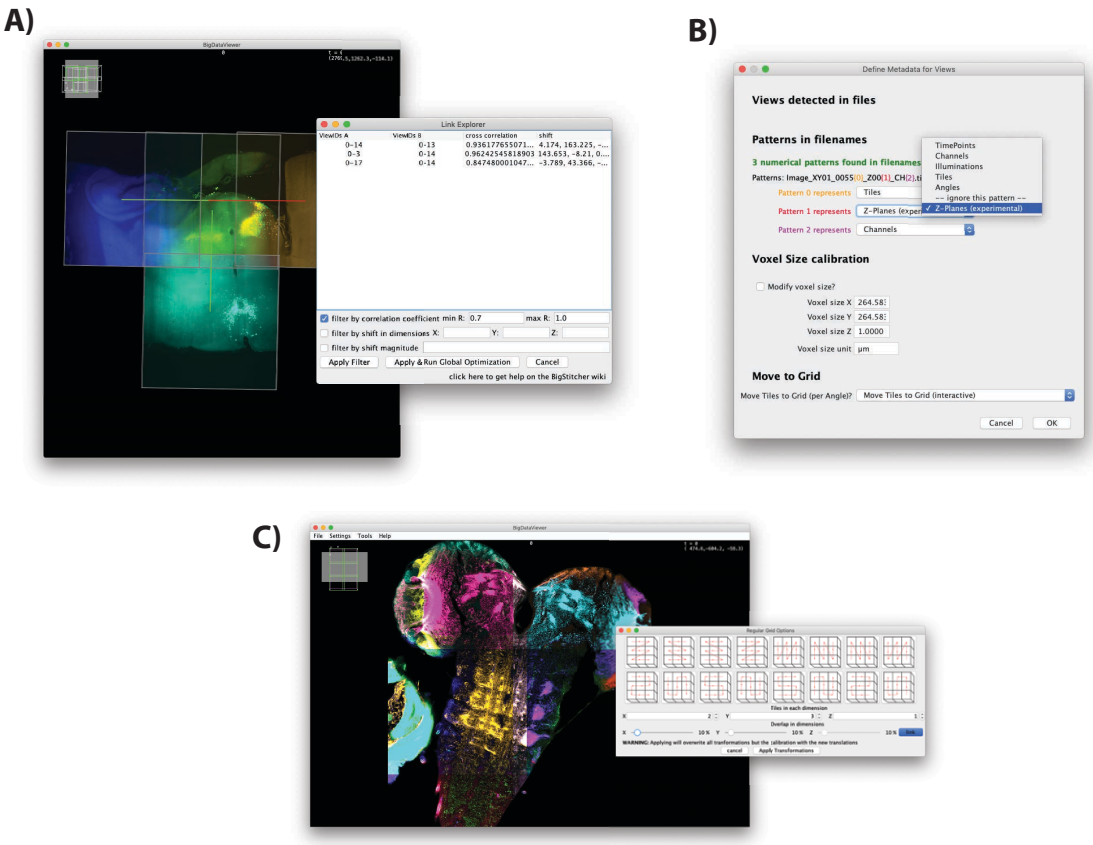


Figure 3.2: Interactivity and examples of manual data curation in BigStitcher. (A) After calculating pairwise transformations, they can be previewed on-the-fly in BigDataViewer and erroneous results can be filtered based on a set of criteria like a minimum correlation coefficient or removed manually. (B) When loading a dataset consisting of multiple files into BigStitcher, the software tries to assign attributes to the images automatically if metadata are given or requires only minimal guidance from the user, e.g. by assigning automatically detected patterns in the filenames to image attributes. (C) As a first step in the alignment of a dataset, images can be moved to a regular grid to produce a rough pre-alignment. Like in many other steps of BigStitcher, an interactive preview of the process is shown.

ple transformations are estimated in parallel to reduce compute time). We tackled both problems by employing a multi-resolution data representation throughout BigStitcher and working with the lowest possible resolution whenever possible, in a semi-hierarchical way. While this might lead to a decrease in registration quality at first sight, we could show that by clever utilization of subpixel-accuracy algorithms, such as subpixel localization of the shift peak in Fourier-based alignment or subpixel interest point detection in interest point-based alignment, we can achieve comparable quality to full-resolution methods while retaining up to several hundredfold speedups. BigStitcher is also a superset of the already existing Multiview Reconstruction tool [115], allowing for the alignment of images acquired from multiple angles or time series via interest-point based alignment using special descriptors developed for three dimensional problems. Stitching of tiled acquisitions and MVR are done in one common transformation framework (as both are just special cases of image alignment), a generic evolution of previous tools that focus on just some special cases.

Very large samples might entail regions of low image quality that are difficult to align with fully automated procedures. We therefore took great care to ensure global consistency of the alignment throughout the stitching process: by employing not only a single transformation mapping each individual image to the final volume, but a list of incremental transformations for each image, we can base the alignment on preexisting information, e.g. microscope stage metadata. During the global optimization of pairwise transformations, inconsistencies are detected and removed automatically, one pairwise link at a time, to keep as much information as possible compared to a minimal spanning tree approach that discards most of the pairwise links from the outset [127]. Furthermore, by using a two-step optimization process, we can find an optimal alignment of all images for which pairwise transformations are known while keeping as-close-as-possible to preexisting information like metadata. Via this strategy, BigStitcher can generate a consistent alignment

even when the sample only consists of ”islands” of meaningful structures, or refine parts of the transformations with higher-resolution calculations while keeping rough estimates from a previous round of fast registration using a high downsampling factor. This is a very generic approach to a regularized global optimization of transformations that works with arbitrary locations of the individual images and does not require images to be acquired in a regular grid [157].

While simple aberrations such as chromatic aberrations or spherical aberrations can be accounted for by registration with an affine model, another problem we address are optical aberrations due to inhomogeneous refractive index throughout large samples. These sample dependent distortions call for more general registration models, so-called non-rigid alignment. A simple way to achieve a finer grained local alignment is to virtually split all images into smaller parts and align them piecewise using an affine model. A further option is to calculate a separate transformation at each point by using a locally weighted average of interest point correspondences in the neighborhood, e.g. using the moving least squares approach. We implemented both strategies to achieve non-rigid alignment in BigStitcher and could show that they indeed improve registration results using manually curated landmarks in a real dataset of LSFM images of a murine brain slice.

While the precise alignment of many-image datasets is the core functionality of our tool, we added several further functions to improve the final images. As mentioned above, uneven illumination is one of the simplest optical phenomena that can be corrected with computational means. In terabyte-sized datasets, this might entail saving a corrected copy of equal size, further increasing the storage requirements of the data. We therefore implemented virtual flatfield-correction that performs the correction on-the-fly as images are loaded, and in full compatibility with the multiresolution data structures used throughout the tool. On a similar note, light-sheet microscopy of large sample often benefits from parallel illumination from

two sides. If images with each illumination direction are acquired separately, the quality of the image pairs shows a reciprocal relation as one moves laterally along the common illumination axis. An easy way of improving the final image quality is to pick the higher quality image at each location and only use that for the further steps. Quantifying the quality of an image is essential in many tasks, such as autofocusing and adaptive optics, but also for quality control of experimental results, and a variety of metrics to measure image quality exist. In BigStitcher, we support fast selection of the better illumination direction by comparison of the mean intensities of the two images or the mean gradient magnitudes, but also via a more powerful quality criterion derived from Fourier Ring Correlation (FRC). We also allow for the use of FRC to generate a quality image, to give the user immediate feedback on potential artifacts or low SNR regions in their images. Besides normal FRC, we developed an optimized relative FRC (rFRC) that is less sensitive to fixed-pattern noise that is common in (s)CMOS cameras. Finally, a very general way of improving image quality is deconvolution: estimating the physical structures in an image without the contribution of the microscopes PSF. To this end, we support multi-view deconvolution using the iterative Richardson-Lucy algorithm in BigStitcher [158]. All calculations are done in a virtualized fashion, making this memory and computation-intensive process usable for very large datasets.

3.5 The importance of interactivity

It might seem slightly confusing to the reader that, after showcasing methods of state-of-the-art imaging and automatization of the analysis of images produced by them, I should dedicate space to emphasize the importance of manual steps in image analysis. Yet, I still believe that for progress to be made, the end result of scientific experiments should be novel insights understandable to (human)

researchers, either directly or, in the case of highly complex systemic results, in eventual follow-up studies.

The importance of human expert knowledge becomes more obvious when one considers that many state-of-the art methods for a variety of tasks rely on machine learning, especially deep learning. It is easy to forget that the high performance of these "automatic" methods stems not only from the very flexible models used, but from their pairing with vast amounts of ground-truth data curated by human experts. Deep learning has seen its most successful applications in tasks for which large, standardized data sets are available, like, for example, the ImageNet dataset [159] for photographic image classification [160, 161], COCO [162] for object detection [163] or multilingual parliamentary records, as they are kept by the EU or Canada [164], that are used in the training of machine translation systems.

The cases in which machine learning has been used to derive novel insights, even though they can be constructed as thought experiments [165], is surprisingly small and narrow ML is mostly employed to replace manual steps in any kind of data processing pipeline. Until true general artificial intelligences are developed, a task that has eluded AI researchers for decades, deep learning can be seen as a framework of automatically extracting correlations from data and generating black-box representations, achieving a similar goal as rule-based AI approaches such as expert systems, though by different means. That is not to say that ML enables no methodological advances at all, as ingenious schemes for ground-truth data generation can be used and transferred to contexts in which the data are not available in comprehensive form at all [146, 166]. This is especially true in biological research, where available datasets are small and clever ways of expanding them to meet the requirements of deep learning has and will surely prove fruitful. Examples for this approach include the generation of (input, target)-pairs from separate experiments, virtual expansion of datasets via data augmentation, a task that it-

self can be tackled with deep generative models, or dedicated few-shot learning techniques.

As there are no avenues towards the advent of general AI yet and even shallow AI models that provide understandable feedback on their inner workings, though seeing tremendous research interest, are in their infancy, human expert intervention is still state-of-the-art for most applications. The challenge is to find a middle ground between fully automated pipelines that might produce sub-par results and mostly manual analyses that provide little time savings, which I tried to do in my work, e.g. by incorporating a manual inspection step into the segmentation parts of the otherwise automated image analysis pipelines used in [147, 148].

Along the same lines, interactivity can be seen as one of the main advantages of BigStitcher in comparison to alternative tools for image alignment [127, 157]. By closely integrating BigStitcher with the multidimensional image visualization provided by BigDataViewer, we can show a step-by-step progression of the alignment pipeline to the user (figure 3.2), allowing them to preview in real time e.g. pairwise shifts, detected interest points, selection of the best illumination direction, flatfield correction or rFRC quality throughout the dataset. Performing an alignment of terabyte sized datasets can take hours, even with the numerous optimizations we introduced and even with the regularizations we perform to make the process as-robust-as-possible, errors in individual steps are still possible, e.g. erroneous pairwise shifts throwing off global alignment or artifacts interfering with interest point-based alignment. If the whole procedure were implemented as a black box, the user would just be presented with an unsatisfactory result after hours of compute time. By providing visual feedback throughout the process, most errors become obvious immediately and the user has the ability to manually correct them at once, e.g. by re-doing single steps with different parameters. Especially in experiments on large cleared samples, such as entire organs, which often take

weeks to prepare and require costly reagents, discarding the whole experiment would constitute a major setback and the possibility for manual curation is vital. On the other hand, we did not sacrifice power for usability - BigStitcher can be automated with macros and run in batch mode for routine experiments for which effective parameters are known.

3.6 Outlook & other projects pursued during PhD

As a tool for the preprocessing of large LSFM datasets, BigStitcher is already seeing wide adaptation, as it allows for hassle-free manual inspection and preprocessing of the data. Nonetheless, an important next step is the development of actual automatic or semi-automatic analysis tools for such large datasets. A step in that direction is the development of large data annotation tools such as MaMuT [167]. What should follow is the development of actual analysis tools capable of handling these large datasets, which will likely rely heavily on parallel processing of subsets of the data, most likely not on single machines, but on dedicated compute clusters. Yet, to allow analysis of big datasets without (too) specialized infrastructure, the development of multi-resolution analysis workflows (similar to the routines we employ during image alignment and preprocessing) seems an avenue worth pursuing.

The finished publications presented in the thesis are (unfortunately) not a complete overview of all the work I did during my PhD. For example, a significant amount of time went into the development of frameworks for smart microscopy on two different commercially available systems, the Abberior Instruments Expert Line STED as well as the Nikon Ti platform. Using our STED setup and the possibility to automate measurements via a Python interface, I, with the help of students [168], implemented a two-step scanning scheme consisting of fast confocal overview imaging followed by the detection of structures of interest that are then

imaged at STED super-resolution in a second step. The framework was written in a very generic way, offering the possibility to quickly implement more complex acquisition pipelines. We have employed the system to capture population statistics of chromatin conformation in single cells by combining pairwise distance measurements in a high-throughput fashion with novel oligomer-pool-based FISH protocols developed by collaborators [1]. At the time of writing, the automation pipeline, FISH protocol and results collected by their use form the basis of multiple manuscripts with varying degree of finalization.

A similar smart microscopy platform was implemented for colleagues in the group of Nicolas Gompel to allow high-throughput, high-resolution imaging of Drosophila wings, under the working title WingScanner. Again, the bulk of automation code consists of high-level Python routines, working in concert with the commercial microscope control software (NIS Elements) by wrapping selected parts of its macro interface in a Pythonic way. As wings are a complex object to detect, we eventually moved towards deep learning-based object detection, employing a client-server architecture to perform expensive computations on a GPU-equipped server while not interfering with the microscope control workstation. As high-resolution wing images often span more than one FOV, BigStitcher was also included into the framework to perform alignment and fusion of the tiled images automatically. More than one manuscript based on results acquired on the WingScanner system are currently in preparation in the Gompel lab [2]

Finally, I supervised students in developing new deep learning model architectures for the task of image augmentation to facilitate limited ground-truth ML-based analyses [169, 170], employing generative adversarial networks (GANs) [171, 172] to generate a potentially infinite amount of synthetic images from a few real examples. While the early benchmarks of this line of investigation (e.g. perfor-

[1] Tobias Ragoczy, PhD, personal communication

[2] Prof. Dr. Nicolas Gompel, personal communication

mance of a downstream segmentation model) show only mediocre improvements, we nonetheless believe that strategies of incorporating ML into the low-data situations encountered in science are of great interest and will continue our efforts in this direction. It is my opinion that pure ML research tends to focus too narrowly on a few standard applications (and standard datasets) [173] and there is enormous potential in finding clever ways of integration of these techniques into applied scientific workflows, that inevitably entail novel, non-standard results and data structure.

On that note, I am convinced that method developers should not aim to replace researchers with ever smarter instrumentation and software, but complement and aid them in discovering new humanly tractable insights into the molecular basis of life. It is my hope that I have contributed a little to this goal with the work presented in this thesis and I am looking forward to doing so in the future.

Appendix A

Bibliography

[1] Erwin Schrödinger. *What Is Life? The Physical Aspect of the Living Cell; with Mind and Matter & Autobiographical Sketches*. Canto Classics. Cambridge ; New York: Cambridge University Press, 1992. 184 pp. ISBN: 978-1-107-60466-7.

[2] D. S. Johnson et al. "Genome-Wide Mapping of in Vivo Protein-DNA Interactions". In: *Science* 316.5830 (June 8, 2007), pp. 1497–1502. DOI: 10.1126/science.1141319.

[3] E. Lieberman-Aiden et al. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome". In: *Science* 326.5950 (Oct. 9, 2009), pp. 289–293. DOI: 10.1126/science.1181369.

[4] Joshua A. Weinstein, Aviv Regev, and Feng Zhang. "DNA Microscopy: Optics-Free Spatio-Genetic Imaging by a Stand-Alone Chemical Reaction". In: *Cell* 178.1 (June 2019), 229–241.e16. DOI: 10.1016/j.cell.2019.05.019.

[5] Timo Zimmermann. "Spectral Imaging and Linear Unmixing in Light Microscopy". In: *Microscopy Techniques*. Ed. by Jens Rietdorf. Vol. 95. Berlin, Heidelberg: Springer Berlin Heidelberg, May 27, 2005, pp. 245–265. ISBN: 978-3-540-23698-6 978-3-540-31545-2. DOI: 10.1007/b102216.

[6] H. Gest. "The Discovery of Microorganisms by Robert Hooke and Antoni van Leeuwenhoek, Fellows of The Royal Society". In: *Notes and Records of the Royal*

Society of London 58.2 (May 22, 2004), pp. 187–201. DOI: 10.1098/rsnr.2004.0055.

[7] Project Gutenberg. *The Fundamentals of Bacteriology, by Charles Bradfield Morrey.* URL: https://www.gutenberg.org/files/43227/43227-h/43227-h.htm (visited on 01/20/2020).

[8] Wikipedia, The Free Encyclopedia. *Cell Theory.* URL: https://en.wikipedia.org/wiki/Cell_theory (visited on 01/20/2020).

[9] David Hörl. "Imaging the Nucleus Using STED Microscopy". Master thesis. Munich: LMU, 2015.

[10] Richard N. Day and Michael W. Davidson. "The Fluorescent Protein Palette: Tools for Cellular Imaging". In: *Chemical Society Reviews* 38.10 (2009), p. 2887. DOI: 10.1039/b901966a.

[11] T. Cremer and M. Cremer. "Chromosome Territories". In: *Cold Spring Harbor Perspectives in Biology* 2.3 (Mar. 1, 2010), a003889–a003889. DOI: 10.1101/cshperspect.a003889.

[12] Andrea M Femino et al. "Visualization of Single RNA Transcripts in Situ". In: *Science* 280.5363 (Apr. 24, 1998), pp. 585–590. DOI: 10.1126/science.280.5363.585.

[13] Sara H Rouhanifard et al. "ClampFISH Detects Individual Nucleic Acid Molecules Using Click Chemistry–Based Amplification". In: *Nature Biotechnology* 37.1 (Jan. 2019), pp. 84–89. DOI: 10.1038/nbt.4286.

[14] Osamu Shimomura, Frank H. Johnson, and Yo Saiga. "Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea". In: *Journal of Cellular and Comparative Physiology* 59.3 (June 1962), pp. 223–239. DOI: 10.1002/jcp.1030590302.

[15] Roger Y. Tsien. "The Green Fluorescent Protein". In: *Annual Review of Biochemistry* 67.1 (June 1998), pp. 509–544. DOI: 10.1146/annurev.biochem.67.1.509.

[16] NobelPrize.org. *The Nobel Prize in Chemistry 2008 (Press Release)*. 2008. URL: https://www.nobelprize.org/prizes/chemistry/2008/press-release/ (visited on 01/22/2020).

[17] Tobias Anton, Heinrich Leonhardt, and Yolanda Markaki. "Visualization of Genomic Loci in Living Cells with a Fluorescent CRISPR/Cas9 System". In: *The Nuclear Envelope*. Ed. by Sue Shackleton, Philippe Collas, and Eric C. Schirmer. Vol. 1411. New York, NY: Springer New York, 2016, pp. 407–417. ISBN: 978-1-4939-3528-4 978-1-4939-3530-7. DOI: 10.1007/978-1-4939-3530-7_25.

[18] Hanhui Ma et al. "CRISPR-Sirius: RNA Scaffolds for Signal Amplification in Genome Imaging". In: *Nature Methods* 15.11 (Nov. 2018), pp. 928–931. DOI: 10.1038/s41592-018-0174-0.

[19] Antje Keppler et al. "Labeling of Fusion Proteins of O6-Alkylguanine-DNA Alkyltransferase with Small Molecules in Vivo and in Vitro". In: *Methods* 32.4 (Apr. 2004), pp. 437–444. DOI: 10.1016/j.ymeth.2003.10.007.

[20] Georgyi V. Los et al. "HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis". In: *ACS Chemical Biology* 3.6 (June 2008), pp. 373–382. DOI: 10.1021/cb800025k.

[21] Florian Schueder et al. "Universal Super-Resolution Multiplexing by DNA Exchange". In: *Angewandte Chemie International Edition* 56.14 (Mar. 27, 2017), pp. 4052–4055. DOI: 10.1002/anie.201611729.

[22] Jean-Karim Hériché, Stephanie Alexander, and Jan Ellenberg. "Integrating Imaging and Omics: Computational Methods and Challenges". In: *Annual Review of Biomedical Data Science* 2.1 (July 20, 2019), pp. 175–197. DOI: 10.1146/annurev-biodatasci-080917-013328.

[23] E. Abbe. "Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung". In: *Archiv für Mikroskopische Anatomie* 9.1 (Dec. 1873), pp. 413–468. DOI: 10.1007/BF02956173.

[24] Rayleigh. "XV. *On the Theory of Optical Images, with Special Reference to the Microscope*". In: *The London, Edinburgh, and Dublin Philosophical Maga-*

zine and Journal of Science 42.255 (Aug. 1896), pp. 167–195. DOI: 10.1080/14786449608620902.

[25] M. G. L. Gustafsson. "Surpassing the Lateral Resolution Limit by a Factor of Two Using Structured Illumination Microscopy". In: *Journal of Microscopy* 198.2 (May 2000), pp. 82–87. DOI: 10.1046/j.1365-2818.2000.00710.x.

[26] Rainer Heintzmann and Christoph G. Cremer. "Laterally Modulated Excitation Microscopy: Improvement of Resolution by Using a Diffraction Grating". In: BiOS Europe '98. Ed. by Irving J. Bigio et al. Stockholm, Sweden, Jan. 19, 1999, pp. 185–196. DOI: 10.1117/12.336833.

[27] M. A. A. Neil, R. Juškaitis, and T. Wilson. "Method of Obtaining Optical Sectioning by Using Structured Light in a Conventional Microscope". In: *Optics Letters* 22.24 (Dec. 15, 1997), p. 1905. DOI: 10.1364/OL.22.001905.

[28] Mats G.L. Gustafsson et al. "Three-Dimensional Resolution Doubling in Wide-Field Fluorescence Microscopy by Structured Illumination". In: *Biophysical Journal* 94.12 (June 2008), pp. 4957–4970. DOI: 10.1529/biophysj.107.120345.

[29] L. Schermelleh et al. "Subdiffraction Multicolor Imaging of the Nuclear Periphery with 3D Structured Illumination Microscopy". In: *Science* 320.5881 (June 6, 2008), pp. 1332–1336. DOI: 10.1126/science.1156947.

[30] Iain Williamson et al. "*Shh* and ZRS Enhancer Colocalisation Is Specific to the Zone of Polarising Activity". In: *Development* 143.16 (Aug. 15, 2016), pp. 2994–3001. DOI: 10.1242/dev.139188.

[31] M. G. L. Gustafsson. "Nonlinear Structured-Illumination Microscopy: Wide-Field Fluorescence Imaging with Theoretically Unlimited Resolution". In: *Proceedings of the National Academy of Sciences* 102.37 (Sept. 13, 2005), pp. 13081–13086. DOI: 10.1073/pnas.0406877102.

[32] Rainer Heintzmann, Thomas M. Jovin, and Christoph Cremer. "Saturated Patterned Excitation Microscopy—a Concept for Optical Resolution Improvement". In: *Journal of the Optical Society of America A* 19.8 (Aug. 1, 2002), p. 1599. DOI: 10.1364/JOSAA.19.001599.

[33] Stefan W. Hell and Jan Wichmann. “Breaking the Diffraction Resolution Limit by Stimulated Emission: Stimulated-Emission-Depletion Fluorescence Microscopy”. In: *Optics Letters* 19.11 (June 1, 1994), p. 780. DOI: 10.1364/OL.19.000780.

[34] Thomas A. Klar and Stefan W. Hell. “Subdiffraction Resolution in Far-Field Fluorescence Microscopy”. In: *Optics Letters* 24.14 (July 15, 1999), p. 954. DOI: 10.1364/OL.24.000954.

[35] NobelPrize.org. *The Nobel Prize in Chemistry 2014 (Press Release)*. 2014. URL: https://www.nobelprize.org/prizes/chemistry/2014/press-release/ (visited on 01/22/2020).

[36] Joanna Oracz et al. “Photobleaching in STED Nanoscopy and Its Dependence on the Photon Flux Applied for Reversible Silencing of the Fluorophore”. In: *Scientific Reports* 7.1 (Dec. 2017), p. 11354. DOI: 10.1038/s41598-017-09902-x.

[37] Benjamin Harke et al. “Resolution Scaling in STED Microscopy”. In: *Optics Express* 16.6 (Mar. 17, 2008), p. 4154. DOI: 10.1364/OE.16.004154.

[38] Fabian Göttfert et al. “Strong Signal Increase in STED Fluorescence Microscopy by Imaging Regions of Subdiffraction Extent”. In: *Proceedings of the National Academy of Sciences* 114.9 (Feb. 28, 2017), pp. 2125–2130. DOI: 10.1073/pnas.1621495114.

[39] Jörn Heine et al. “Adaptive-Illumination STED Nanoscopy”. In: *Proceedings of the National Academy of Sciences* 114.37 (Sept. 12, 2017), pp. 9797–9802. DOI: 10.1073/pnas.1708304114.

[40] R A Hoebe et al. “Controlled Light-Exposure Microscopy Reduces Photobleaching and Phototoxicity in Fluorescence Live-Cell Imaging”. In: *Nature Biotechnology* 25.2 (Feb. 2007), pp. 249–253. DOI: 10.1038/nbt1278.

[41] Thorsten Staudt et al. “Far-Field Optical Nanoscopy with Reduced Number of State Transition Cycles”. In: *Optics Express* 19.6 (Mar. 14, 2011), p. 5644. DOI: 10.1364/OE.19.005644.

[42] Russell E. Thompson, Daniel R. Larson, and Watt W. Webb. "Precise Nanometer Localization Analysis for Individual Fluorescent Probes". In: *Biophysical Journal* 82.5 (May 2002), pp. 2775–2783. DOI: 10.1016/S0006-3495(02)75618-X.

[43] A. Yildiz. "Myosin V Walks Hand-Over-Hand: Single Fluorophore Imaging with 1.5-Nm Localization". In: *Science* 300.5628 (June 27, 2003), pp. 2061–2065. DOI: 10.1126/science.1084398.

[44] Stefan Niekamp et al. "Nanometer-Accuracy Distance Measurements between Fluorophores at the Single-Molecule Level". In: *Proceedings of the National Academy of Sciences* 116.10 (Mar. 5, 2019), pp. 4275–4284. DOI: 10.1073/pnas.1815826116.

[45] Christoph Cremer et al. "Principles of Spectral Precision Distance Confocal Microscopy for the Analysis of Molecular Nuclear Structure". In: *Handbook of computer vision and applications* 3 (1999), pp. 839–857.

[46] G. H. Patterson. "A Photoactivatable GFP for Selective Photolabeling of Proteins and Cells". In: *Science* 297.5588 (Sept. 13, 2002), pp. 1873–1877. DOI: 10.1126/science.1074952.

[47] E. Betzig et al. "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution". In: *Science* 313.5793 (Sept. 15, 2006), pp. 1642–1645. DOI: 10.1126/science.1127344.

[48] Samuel T. Hess, Thanu P.K. Girirajan, and Michael D. Mason. "Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy". In: *Biophysical Journal* 91.11 (Dec. 2006), pp. 4258–4272. DOI: 10.1529/biophysj.106.091116.

[49] Michael J Rust, Mark Bates, and Xiaowei Zhuang. "Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM)". In: *Nature Methods* 3.10 (Oct. 2006), pp. 793–796. DOI: 10.1038/nmeth929.

[50] N.L. Thompson, T.P. Burghardt, and D. Axelrod. "Measuring Surface Dynamics of Biomolecules by Total Internal Reflection Fluorescence with Photobleaching Recovery or Correlation Spectroscopy". In: *Biophysical Journal* 33.3 (Mar. 1981), pp. 435–454. DOI: 10.1016/S0006-3495(81)84905-3.

[51] Bo Huang et al. “Whole-Cell 3D STORM Reveals Interactions between Cellular Structures with Nanometer-Scale Resolution”. In: *Nature Methods* 5.12 (Dec. 2008), pp. 1047–1052. DOI: 10.1038/nmeth.1274.

[52] Andrey Aristov et al. “ZOLA-3D Allows Flexible 3D Localization Microscopy over an Adjustable Axial Range”. In: *Nature Communications* 9.1 (Dec. 2018), p. 2409. DOI: 10.1038/s41467-018-04709-4.

[53] S. R. P. Pavani et al. “Three-Dimensional, Single-Molecule Fluorescence Imaging beyond the Diffraction Limit by Using a Double-Helix Point Spread Function”. In: *Proceedings of the National Academy of Sciences* 106.9 (Mar. 3, 2009), pp. 2995–2999. DOI: 10.1073/pnas.0900245106.

[54] Manuel F Juette et al. “Three-Dimensional Sub–100 Nm Resolution Fluorescence Microscopy of Thick Samples”. In: *Nature Methods* 5.6 (June 2008), pp. 527–529. DOI: 10.1038/nmeth.1211.

[55] Ingo Gregor and Jörg Enderlein. “Image Scanning Microscopy”. In: *Current Opinion in Chemical Biology* 51 (Aug. 2019), pp. 74–83. DOI: 10.1016/j.cbpa.2019.05.011.

[56] T. Dertinger et al. “Fast, Background-Free, 3D Super-Resolution Optical Fluctuation Imaging (SOFI)”. In: *Proceedings of the National Academy of Sciences* 106.52 (Dec. 29, 2009), pp. 22287–22292. DOI: 10.1073/pnas.0907866106.

[57] Siân Culley et al. “SRRF: Universal Live-Cell Super-Resolution Microscopy”. In: *The International Journal of Biochemistry & Cell Biology* 101 (Aug. 2018), pp. 74–79. DOI: 10.1016/j.biocel.2018.05.014.

[58] Francisco Balzarotti et al. “Nanometer Resolution Imaging and Tracking of Fluorescent Molecules with Minimal Photon Fluxes”. In: *Science* 355.6325 (Feb. 10, 2017), pp. 606–612. DOI: 10.1126/science.aak9913.

[59] Klaus C. Gwosch et al. “MINFLUX Nanoscopy Delivers 3D Multicolor Nanometer Resolution in Cells”. In: *Nature Methods* (Jan. 13, 2020). DOI: 10.1038/s41592-019-0688-0.

[60] James B. Pawley, ed. *Handbook of Biological Confocal Microscopy*. 3rd ed. New York, NY: Springer, 2006. 985 pp. ISBN: 978-0-387-25921-5.

[61] Daniel Sage et al. "DeconvolutionLab2: An Open-Source Software for Deconvolution Microscopy". In: *Methods* 115 (Feb. 2017), pp. 28–41. DOI: 10.1016/j.ymeth.2016.12.015.

[62] J. Huisken and D. Stainier. "Selective Plane Illumination Microscopy Techniques in Developmental Biology". In: *Development* 136.12 (June 15, 2009), pp. 1963–1975. DOI: 10.1242/dev.022426.

[63] Elizabeth M.C. Hillman et al. "Light-Sheet Microscopy in Neuroscience". In: *Annual Review of Neuroscience* 42.1 (July 8, 2019), pp. 295–313. DOI: 10.1146/annurev-neuro-070918-050357.

[64] Yinan Wan, Katie McDole, and Philipp J. Keller. "Light-Sheet Microscopy and Its Potential for Understanding Developmental Processes". In: *Annual Review of Cell and Developmental Biology* 35.1 (Oct. 7, 2019), annurev-cellbio-100818–125311. DOI: 10.1146/annurev-cellbio-100818-125311.

[65] H. Siedentopf and R. Zsigmondy. "Uber Sichtbarmachung und Größenbestimmung ultramikoskopischer Teilchen, mit besonderer Anwendung auf Goldrubingläser". In: *Annalen der Physik* 315.1 (1902), pp. 1–39. DOI: 10.1002/andp.19023150102.

[66] A. H. Voie, D. H. Burns, and F. A. Spelman. "Orthogonal-Plane Fluorescence Optical Sectioning: Three-Dimensional Imaging of Macroscopic Biological Specimens". In: *Journal of Microscopy* 170.3 (June 1993), pp. 229–236. DOI: 10.1111/j.1365-2818.1993.tb03346.x.

[67] J. Huisken. "Optical Sectioning Deep Inside Live Embryos by Selective Plane Illumination Microscopy". In: *Science* 305.5686 (Aug. 13, 2004), pp. 1007–1009. DOI: 10.1126/science.1100035.

[68] Y. Wu et al. "Spatially Isotropic Four-Dimensional Imaging with Dual-View Plane Illumination Microscopy". In: *Nature Biotechnology* 31.11 (Nov. 2013), pp. 1032–1038. DOI: 10.1038/nbt.2713.

[69] Petr Strnad et al. “Inverted Light-Sheet Microscope for Imaging Mouse Pre-Implantation Development”. In: *Nature Methods* 13.2 (Feb. 2016), pp. 139–142. DOI: 10.1038/nmeth.3690.

[70] Raghav K Chhetri et al. “Whole-Animal Functional and Developmental Imaging with Isotropic Spatial Resolution”. In: *Nature Methods* 12.12 (Dec. 2015), pp. 1171–1178. DOI: 10.1038/nmeth.3632.

[71] Raju Tomer et al. “Quantitative High-Speed Imaging of Entire Developing Embryos with Simultaneous Multiview Light-Sheet Microscopy”. In: *Nature Methods* 9.7 (July 2012), pp. 755–763. DOI: 10.1038/nmeth.2062.

[72] Peter G Pitrone et al. “OpenSPIM: An Open-Access Light-Sheet Microscopy Platform”. In: *Nature Methods* 10.7 (July 2013), pp. 598–599. DOI: 10.1038/nmeth.2507.

[73] Hans-Ulrich Dodt et al. “Ultramicroscopy: Three-Dimensional Visualization of Neuronal Networks in the Whole Mouse Brain”. In: *Nature Methods* 4.4 (Apr. 2007), pp. 331–336. DOI: 10.1038/nmeth1036.

[74] P. J. Keller et al. “Reconstruction of Zebrafish Early Embryonic Development by Scanned Light Sheet Microscopy”. In: *Science* 322.5904 (Nov. 14, 2008), pp. 1065–1069. DOI: 10.1126/science.1162493.

[75] Florian O. Fahrbach and Alexander Rohrbach. “A Line Scanned Light-Sheet Microscope with Phase Shaped Self-Reconstructing Beams”. In: *Optics Express* 18.23 (Nov. 8, 2010), p. 24229. DOI: 10.1364/OE.18.024229.

[76] Thomas A Planchon et al. “Rapid Three-Dimensional Isotropic Imaging of Living Cells Using Bessel Beam Plane Illumination”. In: *Nature Methods* 8.5 (May 2011), pp. 417–423. DOI: 10.1038/nmeth.1586.

[77] Tom Vettenburg et al. “Light-Sheet Microscopy Using an Airy Beam”. In: *Nature Methods* 11.5 (May 2014), pp. 541–544. DOI: 10.1038/nmeth.2922.

[78] Bi-Chang Chen et al. “Lattice Light-Sheet Microscopy: Imaging Molecules to Embryos at High Spatiotemporal Resolution”. In: *Science* 346.6208 (Oct. 24, 2014), p. 1257998. DOI: 10.1126/science.1257998.

[79] Kevin M. Dean et al. “Deconvolution-Free Subcellular Imaging with Axially Swept Light Sheet Microscopy”. In: *Biophysical Journal* 108.12 (June 2015), pp. 2807–2815. DOI: 10.1016/j.bpj.2015.05.013.

[80] Fabian F. Voigt et al. “The mesoSPIM Initiative: Open-Source Light-Sheet Microscopes for Imaging Cleared Tissue”. In: *Nature Methods* (Sept. 16, 2019). DOI: 10.1038/s41592-019-0554-0.

[81] Venkatakaushik Voleti et al. “Real-Time Volumetric Microscopy of in Vivo Dynamics and Large-Scale Samples with SCAPE 2.0”. In: *Nature Methods* 16.10 (Oct. 2019), pp. 1054–1062. DOI: 10.1038/s41592-019-0579-4.

[82] Andreas Stengl et al. “A Simple and Sensitive High-Content Assay for the Characterization of Antiproliferative Therapeutic Antibodies”. In: *SLAS DISCOVERY: Advancing Life Sciences R&D* 22.3 (Mar. 2017), pp. 309–315. DOI: 10.1177/1087057116677821.

[83] Rainer Pepperkok and Jan Ellenberg. “High-Throughput Fluorescence Microscopy for Systems Biology”. In: *Nature Reviews Molecular Cell Biology* 7.9 (Sept. 2006), pp. 690–696. DOI: 10.1038/nrm1979.

[84] Christian Conrad et al. “Micropilot: Automation of Fluorescence Microscopy–Based Imaging for Systems Biology”. In: *Nature Methods* 8.3 (Mar. 2011), pp. 246–249. DOI: 10.1038/nmeth.1558.

[85] Arthur Edelstein et al. “Computer Control of Microscopes Using µManager”. In: *Current Protocols in Molecular Biology* 92.1 (Oct. 2010). DOI: 10.1002/0471142727.mb1420s92.

[86] Christian Tischer et al. “Adaptive Fluorescence Microscopy by Online Feedback Image Analysis”. In: *Methods in Cell Biology*. Vol. 123. Elsevier, 2014, pp. 489–503. ISBN: 978-0-12-420138-5. DOI: 10.1016/B978-0-12-420138-5.00026-4.

[87] Tsung-Li Liu et al. “Observing the Cell in Its Native State: Imaging Subcellular Dynamics in Multicellular Organisms”. In: *Science* 360.6386 (Apr. 20, 2018), eaaq1392. DOI: 10.1126/science.aaq1392.

[88] Loïc A Royer et al. “Adaptive Light-Sheet Microscopy for Long-Term, High-Resolution Imaging in Living Organisms”. In: *Nature Biotechnology* 34.12 (Dec. 2016), pp. 1267–1278. DOI: 10.1038/nbt.3708.

[89] David Hörl et al. “BigStitcher: Reconstructing High-Resolution Image Datasets of Cleared and Expanded Samples”. In: *Nature Methods* (Aug. 5, 2019). DOI: 10.1038/s41592-019-0501-0.

[90] Joerg Schnitzbauer et al. “Super-Resolution Microscopy with DNA-PAINT”. In: *Nature Protocols* 12.6 (June 2017), pp. 1198–1228. DOI: 10.1038/nprot.2017.024.

[91] Ralf Jungmann et al. “Quantitative Super-Resolution Imaging with qPAINT”. In: *Nature Methods* 13.5 (May 2016), pp. 439–442. DOI: 10.1038/nmeth.3804.

[92] Ralf Jungmann et al. “Multiplexed 3D Cellular Super-Resolution Imaging with DNA-PAINT and Exchange-PAINT”. In: *Nature Methods* 11.3 (Mar. 2014), pp. 313–318. DOI: 10.1038/nmeth.2835.

[93] Brian J. Beliveau et al. “Single-Molecule Super-Resolution Imaging of Chromosomes and in Situ Haplotype Visualization Using Oligopaint FISH Probes”. In: *Nature Communications* 6.1 (Nov. 2015), p. 7147. DOI: 10.1038/ncomms8147.

[94] Bogdan Bintu et al. “Super-Resolution Chromatin Tracing Reveals Domains and Cooperative Interactions in Single Cells”. In: *Science* 362.6413 (Oct. 26, 2018), eaau1783. DOI: 10.1126/science.aau1783.

[95] Alistair N. Boettiger et al. “Super-Resolution Imaging Reveals Distinct Chromatin Folding for Different Epigenetic States”. In: *Nature* 529.7586 (Jan. 2016), pp. 418–422. DOI: 10.1038/nature16496.

[96] Karel Svoboda and Ryohei Yasuda. “Principles of Two-Photon Excitation Microscopy and Its Applications to Neuroscience”. In: *Neuron* 50.6 (June 2006), pp. 823–839. DOI: 10.1016/j.neuron.2006.05.019.

[97] Kazuki Tainaka et al. “Chemical Principles in Tissue Clearing and Staining Protocols for Whole-Body Cell Profiling”. In: *Annual Review of Cell and Developmen-*

tal Biology 32.1 (Oct. 6, 2016), pp. 713–741. DOI: 10.1146/annurev-cellbio-111315-125001.

[98] Kwanghun Chung et al. "Structural and Molecular Interrogation of Intact Biological Systems". In: *Nature* 497.7449 (May 2013), pp. 332–337. DOI: 10.1038/nature12107.

[99] Evan Murray et al. "Simple, Scalable Proteomic Imaging for High-Dimensional Profiling of Intact Systems". In: *Cell* 163.6 (Dec. 2015), pp. 1500–1514. DOI: 10.1016/j.cell.2015.11.025.

[100] F. Chen, P. W. Tillberg, and E. S. Boyden. "Expansion Microscopy". In: *Science* 347.6221 (Jan. 30, 2015), pp. 543–548. DOI: 10.1126/science.1260088.

[101] Fei Chen et al. "Nanoscale Imaging of RNA with Expansion Microscopy". In: *Nature Methods* 13.8 (Aug. 2016), pp. 679–684. DOI: 10.1038/nmeth.3899.

[102] Emmanouil D. Karagiannis et al. *Expansion Microscopy of Lipid Membranes*. preprint. Bioengineering, Nov. 4, 2019. DOI: 10.1101/829903. URL: http://biorxiv.org/lookup/doi/10.1101/829903 (visited on 01/14/2020).

[103] Paul W Tillberg et al. "Protein-Retention Expansion Microscopy of Cells and Tissues Labeled Using Standard Fluorescent Proteins and Antibodies". In: *Nature Biotechnology* 34.9 (Sept. 2016), pp. 987–992. DOI: 10.1038/nbt.3625.

[104] Jae-Byum Chang et al. "Iterative Expansion Microscopy". In: *Nature Methods* 14.6 (June 2017), pp. 593–599. DOI: 10.1038/nmeth.4261.

[105] Simon J. D. Prince. *Computer Vision: Models, Learning, and Inference*. Cambridge: Cambridge University Press, 2012. ISBN: 978-0-511-99650-4. DOI: 10.1017/CBO9780511996504.

[106] Richard Szeliski. *Computer Vision*. Texts in Computer Science. London: Springer London, 2011. ISBN: 978-1-84882-934-3 978-1-84882-935-0. DOI: 10.1007/978-1-84882-935-0.

[107] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. New York, NY: Pearson, 2018. 1168 pp. ISBN: 978-0-13-335672-4.

[108] Ignacio Arganda-Carreras and Philippe Andrey. “Designing Image Analysis Pipelines in Light Microscopy: A Rational Approach”. In: *Light Microscopy*. Ed. by Yolanda Markaki and Hartmann Harz. Vol. 1563. New York, NY: Springer New York, 2017, pp. 185–207. ISBN: 978-1-4939-6808-4 978-1-4939-6810-7. DOI: 10.1007/978-1-4939-6810-7_13.

[109] Christian Dietz and Michael R. Berthold. “KNIME for Open-Source Bioimage Analysis: A Tutorial”. In: *Focus on Bio-Image Informatics*. Ed. by Winnok H. De Vos, Sebastian Munck, and Jean-Pierre Timmermans. Cham: Springer International Publishing, 2016, pp. 179–197. ISBN: 978-3-319-28549-8. DOI: 10.1007/978-3-319-28549-8_7.

[110] Anne E Carpenter et al. “CellProfiler: Image Analysis Software for Identifying and Quantifying Cell Phenotypes”. In: *Genome Biology* 7.10 (2006), R100. DOI: 10.1186/gb-2006-7-10-r100.

[111] Claire McQuin et al. “CellProfiler 3.0: Next-Generation Image Processing for Biology”. In: *PLOS Biology* 16.7 (July 3, 2018). Ed. by Tom Misteli, e2005970. DOI: 10.1371/journal.pbio.2005970.

[112] Curtis T. Rueden et al. “ImageJ2: ImageJ for the next Generation of Scientific Image Data”. In: *BMC Bioinformatics* 18.1 (Dec. 2017), p. 529. DOI: 10.1186/s12859-017-1934-z.

[113] Tingying Peng et al. “A BaSiC Tool for Background and Shading Correction of Optical Microscopy Images”. In: *Nature Communications* 8.1 (Aug. 2017), p. 14836. DOI: 10.1038/ncomms14836.

[114] Kevin Smith et al. “CIDRE: An Illumination-Correction Method for Optical Microscopy”. In: *Nature Methods* 12.5 (May 2015), pp. 404–406. DOI: 10.1038/nmeth.3323.

[115] Stephan Preibisch et al. “Software for Bead-Based Registration of Selective Plane Illumination Microscopy Data”. In: *Nature Methods* 7.6 (June 2010), pp. 418–419. DOI: 10.1038/nmeth0610-418.

[116] R. Nieuwenhuizen et al. “Measuring Image Resolution in Optical Nanoscopy”. In: *Nature Methods* 10.6 (June 2013), pp. 557–562. DOI: 10.1038/nmeth.2448.

[117] A. Descloux, K. S. Grußmayer, and A. Radenovic. “Parameter-Free Image Resolution Estimation Based on Decorrelation Analysis”. In: *Nature Methods* (Aug. 26, 2019). DOI: 10.1038/s41592-019-0515-7.

[118] Siân Culley et al. “Quantitative Mapping and Minimization of Super-Resolution Optical Imaging Artifacts”. In: *Nature Methods* 15.4 (Apr. 2018), pp. 263–266. DOI: 10.1038/nmeth.4605.

[119] Wei Ouyang et al. “Deep Learning Massively Accelerates Super-Resolution Localization Microscopy”. In: *Nature Biotechnology* 36.5 (May 2018), pp. 460–468. DOI: 10.1038/nbt.4106.

[120] Richard Szeliski. “Image Alignment and Stitching: A Tutorial”. In: *Foundations and Trends® in Computer Graphics and Vision* 2.1 (2006), pp. 1–104. DOI: 10.1561/0600000009.

[121] S. Preibisch, S. Saalfeld, and P. Tomancak. “Globally Optimal Stitching of Tiled 3D Microscopic Image Acquisitions”. In: *Bioinformatics* 25.11 (June 1, 2009), pp. 1463–1465. DOI: 10.1093/bioinformatics/btp184.

[122] B.S. Reddy and B.N. Chatterji. “An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration”. In: *IEEE Transactions on Image Processing* 5.8 (Aug./1996), pp. 1266–1271. DOI: 10.1109/83.506761.

[123] Simon Baker and Iain Matthews. “Lucas-Kanade 20 Years On: A Unifying Framework”. In: *International Journal of Computer Vision* 56.3 (Feb. 2004), pp. 221–255. DOI: 10.1023/B:VISI.0000011205.11775.fd.

[124] Tony Lindeberg. “Feature Detection with Automatic Scale Selection”. In: *International journal of computer vision* 30.2 (1998), pp. 79–116.

[125] David G. Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. In: *International Journal of Computer Vision* 60.2 (Nov. 2004), pp. 91–110. DOI: 10.1023/B:VISI.0000029664.99615.94.

[126] Ethan Rublee et al. “ORB: An Efficient Alternative to SIFT or SURF”. In: *2011 International Conference on Computer Vision*. 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE, Nov. 2011, pp. 2564–2571. ISBN: 978-1-4577-1102-2 978-1-4577-1101-5 978-1-4577-1100-8. DOI: 10.1109/ICCV.2011.6126544.

[127] Alessandro Bria and Giulio Iannello. “TeraStitcher - A Tool for Fast Automatic 3D-Stitching of Teravoxel-Sized Microscopy Images”. In: *BMC Bioinformatics* 13.1 (Dec. 2012), p. 316. DOI: 10.1186/1471-2105-13-316.

[128] Khaled Khairy, Gennady Denisov, and Stephan Saalfeld. “Joint Deformable Registration of Large EM Image Volumes: A Matrix Solver Approach”. In: (Apr. 26, 2018). arXiv: 1804.10019 [cs].

[129] Zhihao Zheng et al. “A Complete Electron Microscopy Volume of the Brain of Adult Drosophila Melanogaster”. In: *Cell* 174.3 (July 2018), 730–743.e22. DOI: 10.1016/j.cell.2018.06.019.

[130] S. Saalfeld et al. “As-Rigid-as-Possible Mosaicking and Serial Section Registration of Large ssTEM Datasets”. In: *Bioinformatics* 26.12 (June 15, 2010), pp. i57–i63. DOI: 10.1093/bioinformatics/btq219.

[131] Tobias Pietzsch et al. “BigDataViewer: Visualization and Processing for Large Image Data Sets”. In: *Nature Methods* 12.6 (June 2015), pp. 481–483. DOI: 10.1038/nmeth.3392.

[132] Alexander Mathis et al. “DeepLabCut: Markerless Pose Estimation of User-Defined Body Parts with Deep Learning”. In: *Nature Neuroscience* 21.9 (Sept. 2018), pp. 1281–1289. DOI: 10.1038/s41593-018-0209-y.

[133] Juan C. Caicedo et al. “Nucleus Segmentation across Imaging Experiments: The 2018 Data Science Bowl”. In: *Nature Methods* (Oct. 21, 2019). DOI: 10.1038/s41592-019-0612-7.

[134] Nobuyuki Otsu. “A Threshold Selection Method from Gray-Level Histograms”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (Jan. 1979), pp. 62–66. DOI: 10.1109/TSMC.1979.4310076.

[135] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. "Snakes: Active Contour Models". In: *International Journal of Computer Vision* 1.4 (Jan. 1988), pp. 321–331. DOI: 10.1007/BF00133570.

[136] Virginie Uhlmann and Michael Unser. "Tip-Seeking Active Contours for Bioimage Segmentation". In: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI 2015). Brooklyn, NY, USA: IEEE, Apr. 2015, pp. 544–547. ISBN: 978-1-4799-2374-8. DOI: 10.1109/ISBI.2015.7163931.

[137] J. Cardinale, G. Paul, and I. F. Sbalzarini. "Discrete Region Competition for Unknown Numbers of Connected Regions". In: *IEEE Transactions on Image Processing* 21.8 (Aug. 2012), pp. 3531–3545. DOI: 10.1109/TIP.2012.2192129.

[138] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The Elements of Statistical Learning*. Vol. 1. 10. Springer series in statistics New York, 2001.

[139] Ignacio Arganda-Carreras et al. "Trainable Weka Segmentation: A Machine Learning Tool for Microscopy Pixel Classification". In: *Bioinformatics* 33.15 (Aug. 1, 2017). Ed. by Robert Murphy, pp. 2424–2426. DOI: 10.1093/bioinformatics/btx180.

[140] Stuart Berg et al. "Ilastik: Interactive Machine Learning for (Bio)Image Analysis". In: *Nature Methods* (Sept. 30, 2019). DOI: 10.1038/s41592-019-0582-9.

[141] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep Learning". In: *Nature* 521.7553 (May 2015), pp. 436–444. DOI: 10.1038/nature14539.

[142] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[143] Yann LeCun et al. "Gradient-Based Learning Applied to Document Recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[144] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab et

al. Vol. 9351. Cham: Springer International Publishing, 2015, pp. 234–241. ISBN: 978-3-319-24573-7 978-3-319-24574-4. DOI: 10.1007/978-3-319-24574-4_28.

[145] Christian Matek et al. “Human-Level Recognition of Blast Cells in Acute Myeloid Leukaemia with Convolutional Neural Networks”. In: *Nature Machine Intelligence* 1.11 (Nov. 2019), pp. 538–544. DOI: 10.1038/s42256-019-0101-9.

[146] Martin Weigert et al. “Content-Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy”. In: *Nature Methods* 15.12 (Dec. 2018), pp. 1090–1097. DOI: 10.1038/s41592-018-0216-7.

[147] Francesco Natale et al. “Identification of the Elementary Structural Units of the DNA Damage Response”. In: *Nature Communications* 8.1 (Aug. 2017), p. 15760. DOI: 10.1038/ncomms15760.

[148] Weihua Qin et al. “DNA Methylation Requires a DNMT1 Ubiquitin Interacting Motif (UIM) and Histone Ubiquitination”. In: *Cell Research* 25.8 (Aug. 2015), pp. 911–929. DOI: 10.1038/cr.2015.72.

[149] Johnny Tam et al. “Cross-Talk-Free Multi-Color STORM Imaging Using a Single Fluorophore”. In: *PLoS ONE* 9.7 (July 7, 2014). Ed. by Katharina Gaus, e101772. DOI: 10.1371/journal.pone.0101772.

[150] Pedro Almada et al. “Automating Multimodal Microscopy with NanoJ-Fluidics”. In: *Nature Communications* 10.1 (Dec. 2019), p. 1223. DOI: 10.1038/s41467-019-09231-9.

[151] Marc-André Kasper et al. “Ethynylphosphonamidates for the Rapid and Cysteine-Selective Generation of Efficacious Antibody-Drug Conjugates”. In: *Angewandte Chemie International Edition* 58.34 (Aug. 19, 2019), pp. 11631–11636. DOI: 10.1002/anie.201904193.

[152] Tonmoy Chakraborty et al. “Light-Sheet Microscopy of Cleared Tissues with Isotropic, Subcellular Resolution”. In: *Nature Methods* 16.11 (Nov. 2019), pp. 1109–1113. DOI: 10.1038/s41592-019-0615-4.

[153] Manish Kumar and Yevgenia Kozorovitskiy. "Tilt-Invariant Scanned Oblique Plane Illumination Microscopy for Large-Scale Volumetric Imaging". In: *Optics Letters* 44.7 (Apr. 1, 2019), p. 1706. DOI: 10.1364/OL.44.001706.

[154] Johannes Schindelin et al. "Fiji: An Open-Source Platform for Biological-Image Analysis". In: *Nature Methods* 9.7 (July 2012), pp. 676–682. DOI: 10.1038/nmeth.2019.

[155] Melissa Linkert et al. "Metadata Matters: Access to Image Data in the Real World". In: *The Journal of Cell Biology* 189.5 (May 31, 2010), pp. 777–782. DOI: 10.1083/jcb.201004104.

[156] Ruixuan Gao et al. "Cortical Column and Whole-Brain Imaging with Molecular Contrast and Nanoscale Resolution". In: *Science* 363.6424 (Jan. 18, 2019), eaau8302. DOI: 10.1126/science.aau8302.

[157] Joe Chalfoun et al. "MIST: Accurate and Scalable Microscopy Image Stitching Tool with Stage Modeling and Error Minimization". In: *Scientific Reports* 7.1 (Dec. 2017), p. 4988. DOI: 10.1038/s41598-017-04567-y.

[158] Stephan Preibisch et al. "Efficient Bayesian-Based Multiview Deconvolution". In: *Nature Methods* 11.6 (June 2014), pp. 645–648. DOI: 10.1038/nmeth.2929.

[159] Jia Deng et al. "ImageNet: A Large-Scale Hierarchical Image Database". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops). Miami, FL: IEEE, June 2009, pp. 248–255. ISBN: 978-1-4244-3992-8. DOI: 10.1109/CVPR.2009.5206848.

[160] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, June 2016, pp. 770–778. ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.90.

[161] Christian Szegedy et al. "Going Deeper with Convolutions". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015 IEEE Confer-

ence on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, June 2015, pp. 1–9. ISBN: 978-1-4673-6964-0. DOI: 10.1109/CVPR.2015.7298594.

[162] Tsung-Yi Lin et al. "Microsoft COCO: Common Objects in Context". In: *Computer Vision – ECCV 2014*. Ed. by David Fleet et al. Vol. 8693. Cham: Springer International Publishing, 2014, pp. 740–755. ISBN: 978-3-319-10601-4 978-3-319-10602-1. DOI: 10.1007/978-3-319-10602-1_48.

[163] Shaoqing Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes et al. Curran Associates, Inc., 2015, pp. 91–99.

[164] Philipp Koehn. "Europarl: A parallel corpus for statistical machine translation". In: *MT summit*. Vol. 5. Citeseer. 2005, pp. 79–86.

[165] Chinmay Belthangady and Loic A. Royer. "Applications, Promises, and Pitfalls of Deep Learning for Fluorescence Image Reconstruction". In: *Nature Methods* 16.12 (Dec. 2019), pp. 1215–1225. DOI: 10.1038/s41592-019-0458-z.

[166] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. "Noise2Void - Learning Denoising from Single Noisy Images". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019.

[167] Carsten Wolff et al. "Multi-View Light-Sheet Imaging and Tracking with the MaMuT Software Reveals the Cell Lineage of a Direct Developing Arthropod Limb". In: *eLife* 7 (Mar. 29, 2018), e34410. DOI: 10.7554/eLife.34410.

[168] Pascal Bawidamann. "Automation of a STED Microscope to Enable Super-Resolution Screening". Bachelor thesis. Munich: LMU, 2016.

[169] David Bunk et al. "Facilitating Limited Ground-Truth Biomedical Image Analysis by Deep Learning-Based Image Augmentation". Talk and Poster, 3rd NEUBIAS Conference (Luxembourg). Feb. 8, 2019.

[170] David Bunk. "Facilitating Limited Ground-Truth Biomedical Image Analysis by Deep Learning-Based Image Augmentation". Master thesis. Munich: LMU, 2019.

[171] Ian Goodfellow et al. “Generative Adversarial Nets”. In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., 2014, pp. 2672–2680.

[172] Tero Karras et al. “Progressive Growing of GANs for Improved Quality, Stability, and Variation”. In: (Oct. 27, 2017). arXiv: 1710.10196 [cs, stat].

[173] Kaiming He, Ross Girshick, and Piotr Dollar. “Rethinking ImageNet Pre-Training”. In: *The IEEE International Conference on Computer Vision (ICCV)*. Oct. 2019.

Appendix B

Abbreviations

1D 1-dimensional

2D 2-dimensional

2PM 2 photon microscopy

3D 3-dimensional

ADC antibody-drug conjugate

AI artificial intelligence

ANN artificial neural network

BFP back focal plane

BSX brain-specific homeobox

ChIP chromatin immunoprecipitation

CLSM confocal laser scanning microscopy

(s)CMOS (scientific) complementary metal-oxide-semiconductor

CNN convolutional neural network

COCO Common Objects in Context

CpG cytosine-guanine dinucleotide

CRISPR clustered regular interspaced short palindromic repeats

DAPI 4',6-diamidino-2-phenylindole

DIC differential interference contrast

DNA deoxyribonucleic acid

DNMT1 DNA methyltransferase 1

DoG difference of Gaussians

DSB double strand break

EdU 5-ethynyl-2'-deoxyuridine

EM electron microscopy

ESC embryonal stem cell

ETL electrically tuneable lens

EU European Union

ExM expansion microscopy

FFT fast Fourier transform

Fiji Fiji Is Just ImageJ

FISH fluorescence in-situ hybridization

FOV field-of-view

fPALM fluorescence photoactivation localization microscopy

FRC Fourier ring correlation

FRET Förster resonance energy transfer

FWHM full width at half maximum

GAN generative adversarial network

GFP green fluorescent protein

γ**H2AX** H2A histone family member X phosphorylated on serine 139

GPU graphics processing unit

H3 histone 3

H3K9me3 histone 3 trimethylated on lysine 9

HCS high content screening

HDF5 Hierarchical Data Format 5

H&E Hematoxylin & Eosin

HeLa Henrietta Lacks cervical cancer cell line

IO input-output

ISM image scanning microscopy

LASER light amplification by stimulated emission of radiation

LED light emitting diode

LoG Laplacian of Gaussian

LSFM light-sheet fluorescence microscopy

MaMuT Massive Multiview Tracker

MEF mouse embryonal fibroblast

ML machine learning

MVR multi-view reconstruction

N5 Not HDF5

NA numerical aperture

nD n-dimensional

OTF object transfer function

PAINT point accumulation in nanoscale topology

PALM photoactivated localization microscopy

PCA principal component analysis

PCM phase correlation matrix

PCNA proliferating cell nuclear antigen

PSF point spread function

RAM random access memory

RANSAC random sample consensus

RI refractive index

RNA ribonucleic acid

R&D research and development

Seq eequencing

SIM structured illumination microscopy

SLM spatial light modulator

SMLM single molecule localization microscopy

SNR signal-to-noise ratio

SPIM selective plane illumination microscopy

SSD sum of squared differences

SSIM saturated SIM

STED Stimulated Emission Depletion

STORM stochastic optical reconstruction microscopy

SVM support vector machine

TIRF total internal reflection fluorescence

TS target sequence

t-SNE t-distributed stochastic neighbour embedding

TWS Trainable WEKA Segmentation

UHRF1 ubiquitin-like containing PHD and RING finger domains 1

UIM ubiquitin interacting motif

WYSIWYG what-you-see-is-what-you-get

Appendix C

List of Publications

David Hörl[1], Fabio Rojas Rusak[1], Friedrich Preusser, Paul Tillberg, Nadine Randel, Raghav K. Chhetri, Albert Cardona, Philipp J. Keller, Hartmann Harz, Heinrich Leonhardt, Mathias Treier, and Stephan Preibisch. "BigStitcher: Reconstructing High-Resolution Image Datasets of Cleared and Expanded Samples". In: *Nature Methods* (Aug. 5, 2019). DOI: 10.1038/s41592-019-0501-0

Florian Schueder, Maximilian T. Strauss, **David Hoerl**, Joerg Schnitzbauer, Thomas Schlichthaerle, Sebastian Strauss, Peng Yin, Hartmann Harz, Heinrich Leonhardt, and Ralf Jungmann. "Universal Super-Resolution Multiplexing by DNA Exchange". In: *Angewandte Chemie International Edition* 56.14 (Mar. 27, 2017). DOI: 10. 1002/anie.201611729

[1] equal contribution

Francesco Natale, Alexander Rapp, Wei Yu, Andreas Maiser, Hartmann Harz, Annina Scholl, Stephan Grulich, Tobias Anton, **David Hörl**, Wei Chen, Marco Durante, Gisela Taucher-Scholz, Heinrich Leonhardt, and M. Cristina Cardoso. "Identification of the Elementary Structural Units of the DNA Damage Response". In: *Nature Communications* 8.1 (Aug. 2017). DOI: 10.1038/ncomms15760

Andreas Stengl, **David Hörl**, Heinrich Leonhardt, and Jonas Helma. "A Simple and Sensitive High-Content Assay for the Characterization of Antiproliferative Therapeutic Antibodies". In: *SLAS DISCOVERY: Advancing Life Sciences R&D* 22.3 (Mar. 2017). DOI: 10.1177/1087057116677821

Weihua Qin, Patricia Wolf, Nan Liu, Stephanie Link, Martha Smets, Federica La Mastra, Ignasi Forné, Garwin Pichler, **David Hörl**, Karin Fellinger, Fabio Spada, Ian Marc Bonapace, Axel Imhof, Hartmann Harz, and Heinrich Leonhardt. "DNA Methylation Requires a DNMT1 Ubiquitin Interacting Motif (UIM) and Histone Ubiquitination". In: *Cell Research* 25.8 (Aug. 2015). DOI: 10.1038/cr.2015.72

Appendix D

Declaration of Contributions

- **BigStitcher: Reconstructing High-Resolution Image Datasets of Cleared and Expanded Samples**
  I wrote the BigStitcher software (source code available at: `https://github.com/PreibischLab/BigStitcher` and `https://github.com/PreibischLab/multiview-reconstruction`) together with the last author, Dr. Stephan Preibisch, to deal with very large image datasets as produced by the co-first author, Dr. Fabio Rojas Rusak, and other co-authors. Fabio and I performed benchmarks of registration accuracy and runtime. Together with Fabio and Stephan, I wrote the manuscript and supplement and prepared all figures with input from the other authors.

Prof. Dr. Heinrich Leonhardt _____

Dr. Fabio Rojas Rusak _____

David Hörl _____

- **Universal Super-Resolution Multiplexing by DNA Exchange**
  I performed STED microscopy of DNA origami as well as HeLa cells, employing our DNA exchange multiplexing scheme and optimized imaging conditions together with the first author, Florian Schüder. I wrote a description of the microscope and imaging conditions for the manuscript together with Dr. Hartmann Harz and helped prepare figures showcasing the STED results. I proofread the entire manuscript.

- **Identification of the Elementary Structural Units of the DNA Damage Response**
  I performed STED microscopy of γH2AX-stained cells at various times after exposure to ionizing radiation. I wrote code to segment nuclei and chromocenters in confocal as well as STED images and performed statistical analysis of H3K9me3 and γH2AX signal in the chromocenters (for confocal images) as well as chromocenter shape (for STED images), working closely with Dr. Tobias Anton. I wrote paragraphs describing methods and helped prepare figures for our results. I proofread the entire manuscript.

Prof. Dr. Heinrich Leonhardt _____

David Hörl _____

- **A Simple and Sensitive High-Content Assay for the Characterization of Antiproliferative Therapeutic Antibodies**

  Working closely with the first author, Dr. Andreas Stengl, I helped prepare MATLAB scripts for fitting inhibition curves to cell counts at various drug concentrations, as well as an R script to fit a cell cycle stage distribution to DAPI intensity measurements (available at: `https://github.com/hoerldavid/CellCycleFit`). I wrote a description of our fitting procedures for the manuscript. I proofread to entire manuscript.

- **DNA Methylation Requires a DNMT1 Ubiquitin Interacting Motif (UIM) and Histone Ubiquitination**

  I wrote code to segment nuclei as well as replication sites in images of fluorescently labelled cells in a semi-automated fashion, together with Dr. Hartmann Harz and with close feedback from the co-first author Dr. Patricia Wolf. Based on segmentation masks, I performed statistical analysis of the amount of GFP-tagged DNMT1 at replication sites for multiple DNMT1 mutants. I wrote a description of our image analysis pipeline with Dr. Hartmann Harz and helped prepare figures from our results. I proofread the entire manuscript.

Prof. Dr. Heinrich Leonhardt _____

David Hörl _____

Appendix E

Acknowledgements

At the time of writing this thesis, I have spent about 24 years ”in school” - roughly five sixths of my life. While I hope that my persistence in this environment is at least partially due to some talent, the support of many people and some luck along the way definitely played their part as well. Trying to list everyone would certainly result in some omissions due to my forgetfulness, so I wish to use the following lines to thank some of the people who were most influential and helpful in my journey into science:

Looking back, I have to admit that I was extremely lucky to have been born into a family that fostered my creativity from my early childhood days and continued to be supportive of me ever since. Of all the people to thank, I foremost want to mention **my parents Christine and Simon Hörl**, who have helped me become who I am today and who have always encouraged me along the way. I also wish to thank the rest of my family - my aunts and uncles, cousins and my grandparents who still show great interest in my work at over 90 years of age - and all the friends I made in school and university!

Having developed a strong technical interest in high school, I initially went on to study computer science but quickly found the perspective of a "conventional" software development job afterwards somewhat lacking. Thus, somewhat naïvely, I decided to pursue an early childhood dream of becoming a natural scientist and enrolled in biology as a second field. Still, it took some time for me to find a direction that held my interest. It was a chance meeting with **Dr. Hartmann Harz** in a practical course that brought me into the lab of **Prof. Dr. Heinrich Leonhardt**, where I have spent the last years as a graduate student. I think I can now say with confidence that I have found a field - microscopy and computational image analysis - that allows me to use my diverse education and also has managed to hold my interest and continues to do so. Heinrich's lab is a true melting pot of ideas filled with enthusiastic people working on a variety of topics and I have yet to be bored in this environment. Representing all my colleagues - both scientists and non-scientists - who made this lab a great place to work, I want to thank my officemates over the time, **Dr. Sebastian Bultmann**, **Katharina Brandstetter**, **Dr. Heinrich Flaswinkel**, **Dr. Patricia Wolf**, **Dr. Katharina Thanisch**, **Dr. Svenja Rühland** and **Dr. Andreas Stengl**. A part of the work in this thesis stemmed from collaborations within the lab and I am deeply grateful to Heinrich, Andreas, Patricia, **Dr. Jonas Helma-Smets** and **Dr. Weihua Qin** for including me in their projects. I also want to take the time to thank Hartmann again for being my everyday supervisor - he continues to be a great inspiration to me!

Again, through a chance meeting, I got to know **Dr. Stephan Preibisch**, who became my de-facto second advisor in all things computational. He introduced me into the international community of software developers centered around the open source biomedical image analysis platform Fiji and I could learn a lot about cutting-edge computer vision from him and the people I met through him. The collaboration with Stephan resulted in the central project of this thesis, BigStitcher,

of which I am co-first author together with **Dr. Fabio Rojas Rusak**. I wish to thank Stephan, Fabio and all the other colleagues who made this great project work!

Two other studies presented in this thesis stemmed from collaborations with the labs of **Prof. Dr. Ralf Jungmann** and **Prof. Dr. M. Cristina Cardoso** - I wish to thank the first authors of the studies, **Florian Schüder**, **Dr. Francesco Natale** and **Dr. Alexander Rapp** for making me a part of their projects. Some large collaboration projects I participated in over the last years were not yet finished at the time of writing of this thesis, but I nonetheless wish to thank colleagues at the Altius institute in Seattle, especially **Tobias Ragoczy, Ph.D**, and in the labs of **Prof. Dr. Nicolas Gompel** and **Prof. Dr. Christof Osman** at LMU. I'm looking forward to bringing our projects to successful conclusions in the near future!

Finally, as I finished my own education, I gradually had the chance and honor to supervise talented students over the years. Representing all the students I helped supervise in seminars, practical courses, research internships and final theses, I want to thank **Pascal Bawidamann**, **David Bunk** and **Miguel Guirao**, the latter two following my path as graduate students in Heinrich's lab. I found that I greatly enjoy explaining complex scientific and technical matters and don't consider teaching a second-class aspect of working at a university, but of equal importance to research.

On that note, I hope I was able to present the work in this thesis in an interesting way that did not bore you, the reader, and wish to thank you for taking the time to read it!