



Contents lists available at ScienceDirect

## Computational and Structural Biotechnology Journal

journal homepage: [www.elsevier.com/locate/csbj](http://www.elsevier.com/locate/csbj)

## Research article

## Advanced image-free analysis of the nano-organization of chromatin and other biomolecules by Single Molecule Localization Microscopy (SMLM)



Jonas Weidner<sup>a</sup>, Charlotte Neitzel<sup>a</sup>, Martin Gote<sup>a,1</sup>, Jeanette Deck<sup>a</sup>, Kim Küntzelmann<sup>a</sup>, Götz Pilarczyk<sup>a</sup>, Martin Falk<sup>a,b,\*</sup>, Michael Hausmann<sup>a</sup>

<sup>a</sup> Kirchhoff-Institute for Physics, Heidelberg University, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany<sup>b</sup> Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 00 Brno, Czech Republic

## ARTICLE INFO

## Article history:

Received 30 November 2022

Received in revised form 8 March 2023

Accepted 8 March 2023

Available online 9 March 2023

## Keywords:

Single molecule localization microscopy (SMLM)

Ripley distance frequency histograms

Persistent homology

Persistent image

Principal component analysis

Application of mathematical analysis tools to chromatin organization and DNA repair processes

## ABSTRACT

The cell as a system of many components, governed by the laws of physics and chemistry drives molecular functions having an impact on the spatial organization of these systems and vice versa. Since the relationship between structure and function is an almost universal rule not only in biology, appropriate methods are required to parameterize the relationship between the structure and function of biomolecules and their networks, the mechanisms of the processes in which they are involved, and the mechanisms of regulation of these processes. Single molecule localization microscopy (SMLM), which we focus on here, offers a significant advantage for the quantitative parametrization of molecular organization: it provides matrices of coordinates of fluorescently labeled biomolecules that can be directly subjected to advanced mathematical analytical procedures without the need for laborious and sometimes misleading image processing. Here, we propose mathematical tools for comprehensive quantitative computer data analysis of SMLM point patterns that include Ripley distance frequency analysis, persistent homology analysis, persistent ‘imaging’, principal component analysis and co-localization analysis. The application of these methods is explained using artificial datasets simulating different, potentially possible and interpretatively important situations. Illustrative analyses of real complex biological SMLM data are presented to emphasize the applicability of the proposed algorithms. This manuscript demonstrated the extraction of features and parameters quantifying the influence of chromatin (re)organization on genome function, offering a novel approach to study chromatin architecture at the nanoscale. However, the ability to adapt the proposed algorithms to analyze essentially any molecular organizations, e.g., membrane receptors or protein trafficking in the cytosol, offers broad flexibility of use.

© 2023 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Although molecular and systems biology have significantly contributed to the understanding of cell signaling pathways at the molecular level, we still know little about how the cell regulates its

function and response to physiological and stress signals in such a chemically complex environment [1–4]. Increasing evidence shows that cells are dynamic systems in which the self-evolving architecture of molecular complexes and their networks at the meso- and nano-level governs the interactions of these complex systems (e.g., chromatin) with other molecules (e.g., DNA repair proteins) and hence their function [4–6]. Function, in turn, influences the architecture of molecular complexes, creating a self-regulating feedback loop. The biophysical principle of cell self-regulation can thus be aptly described by the axiom “form follows function” (and vice versa) taken from architecture.

Methods of super-resolution fluorescence light microscopy [7–11] or electron microscopy [12,13] are very well suited for biological investigations of the structural aspects of biomolecule

**Abbreviations:** SMLM, single molecule localization microscopy; IRIF, ionizing radiation induced foci; PCA, principal component analysis; NN, nearest neighbor; NHEJ, non-homologous end joining; HR, homologous recombination; DSB, DNA double-strand break; LET, linear energy transfer

\* Corresponding author at: Kirchhoff-Institute for Physics, Heidelberg University, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany.

E-mail address: [falk@ibp.cz](mailto:falk@ibp.cz) (M. Falk).

<sup>1</sup> Present address: Animal and Human Health Engineering, KU Leuven, Kasteelpark Arenberg 30, 3001 Leuven, Belgium.

organizations on the nanoscale [14]. However, these techniques usually result in a super-resolved image, from which all structural parameters must be extracted through challenging, time-consuming and sometimes misleading image processing ([15–17], etc.). In contrast to these microscopic techniques, Single Molecule Localization Microscopy (SMLM) [18–21] directly generates a coordinate matrix of fluorescent labeled molecules from which information about a biological system can be obtained straightly by advanced mathematical processing, without a need for image analysis. Pointillistic images can secondarily be generated from the mathematically recalculated data matrices to visualize selected features of the studied system. However, the fundamental advantage of SMLM over other superresolution methods can only be exploited if appropriate mathematical procedures are available for image-free analysis. Hence, a fundamental task is to create new procedures that would enable the identification of important parameters of individual molecular structures and networks that influence the behavior of cellular systems as a whole, their fate and development.

In the current manuscript, we propose new ideas and an appropriate set of advanced mathematical algorithms for the comprehensive analysis of SMLM data, specifically the geometric and topological organization of molecules and molecular complexes forming functional structures in cells. The algorithms provide comprehensive information on the distance frequency distributions, signal densities, and mutual interactions between the same or different (differently labeled) biomolecules within and outside specific structures. A vector representation framework is then introduced based on topological properties of the measured objects as an advanced tool for a quantitative comparison of the mentioned parameters and topological similarity between these objects in different samples (such as different cell types, cell conditions, treatments, development ways of the system in time (e.g., since treatment), etc.). The extracted features can subsequently suggest a lot about the biological function of the studied molecular complexes and their networks.

Currently applied inter-point distance, density-based and tessellation methods for SMLM cluster analysis were overviewed in [22,23]. The point-to-point distance method based on Ripley's K function and the pair correlation function were described in [24]. In the present manuscript, these methods have been extended to analyze the exact shape of clusters and shift the scope towards inter-cluster and global structures. The mesh techniques proposed by [23] have been replaced with the multiscale method of persistent homology. As follows from [25], the basic distance- and density-based methods are currently in use, while graph-based models such as persistent homology are increasingly spreading, as well as machine learning [26–28]. We show here how these approaches can be combined with persistent homology used to extract the underlying features and Principal Component Analysis (PCA) to reduce the data to relevant dimensions.

Methods to extract mesh like structures can be used for SMLM in the form of Voronoi tessellation [29,30] and Delaunay triangulation [31,32]. For a Voronoi vertices diagram, the 2D plane is divided into tiles around each point, where the edge is defined by the equidistant bisector to the next point (see for instance [33]). The Delaunay triangulation splits the plane into triangles with SMLM points as vertices (see for instance [34]). The main problem with these tessellation methods is that they are limited to a short scale, as information is only gathered to the next neighboring point. Therefore, the mesh information obtained in this way correlates strongly with a simple density map. To overcome this issue, persistent homology can be applied with multiple accessible length scales [35,36]. A common problem of tessellation, but sometimes also of persistent homology, might be a lack of unequivocal interpretability along large complex structures, such as the entire nucleus. Therefore, an important requirement is to further simplify the features using PCA

[37], as we propose in this manuscript. Reducing the dimensionality of the various local features extracted by mesh techniques then enables biological interpretation of main structures that occur with natural variability typical for a biological system.

We first explain the functions of the proposed algorithms and the interpretation of the results using suitable simulations covering a wide range of possible biological situations. Next, we demonstrate the broad applicability of the algorithms and their power on the challenging analyses of real biological objects. Chromatin has emerged as an ideal example for this purpose due to its structural and organizational complexity. Chromatin is a highly organized multimolecular network that undergoes dynamic rearrangements related to its functions and cells' response to complex stimuli [4,6,38–41]. For instance, oscillations of chromatin condensation were observed in connection with regulation of genome expression [42,43]. Significant alterations of chromatin architecture also occur after induction of double-strand breaks (DSB) into the DNA molecule by DNA-damaging agents, such as ionizing radiation, and during subsequent DNA repair [34,44]. In addition to changes at the level of the chromatin network as a whole [34,45], specific rearrangements take place at individual DSB sites [38,44,46–49]. This affects the accessibility of chromatin to specific proteins and RNAs [50–52], governs the formation of IIRFs (ionizing radiation induced foci), and presumably also the choice of the specific repair mechanism at the sites of individual DSBs [53–55]; etc. (see Section 2.2). Chromatin architecture thus undergoes characteristic rearrangements with specific dynamics as the cell moves towards a particular fate [4,5,42] and regulates both the local activity of chromatin and the functioning of the entire chromatin network [56]. The current view increasingly leans towards the view that chromatin architecture is significantly involved in the control of all DNA functions, including the transcription, replication, and repair [57]. An important question of current (radio)biology, for example, is how chromatin architecture affects the response of cells to radiation.

Another example that illustratively demonstrates the fundamental importance of architecture-based regulations is the self-formation of lipid membranes with embedded clusters of receptor proteins, driven by minimum entropy of the system [58,59]. Cell activities overexpressing receptors [60], ligand binding to receptors [61], or cell irradiation—all these and other processes provoke changes in the receptor cluster formation and induce incorporation of the receptor molecules into the cytosol followed by molecular trafficking [61–63].

Our results obtained using the proposed algorithms and presented in the current manuscript thus illustrate not only the innovation and wide applicability of these algorithms, but also the importance of physical processes in the control of cell life. We expect that the described analysis of molecular systems at the nanoscale may represent a new way out of the apparent impasse of current systems biology and provide generally valid information about the principles of physics-based (de)regulation of key cellular processes.

## 2. Results

### 2.1. SMLM data acquisition

The principle of SMLM, as well as methods for detection of biomolecules using fluorescently labeled specific antibodies (proteins) or oligonucleotides (DNA, RNA), have been described in detail in our previous and other publications [64,65]. For SMLM data acquisition, a time sequence of a few thousand image frames is recorded for a given specimen to accumulate a sufficient number of randomly occurring blinking events of labeled molecules. Given some background due to auto-fluorescence and non-specifically bound fluorophores, the position of the labeled molecules is determined by computational comparison of one image frame with the next using a

suitable fitting algorithm for the point spread function [66]. The signal intensity, the x and y positions, the errors of the x and y positions, the standard deviation of the x and y position, the number of photoelectrons and the frame number in which the signal was first found are then stored in a data matrix of all individual blinks. These coordinate matrices can then be subjected to quantitative mathematical algorithms for system feature analyses, for instance as we suggest in our previous papers [38,67,68] and below. Illustrative images visualizing selected analysis results can also be generated.

In Section 2.2, we present the partly known [68], but important for understanding of the entire analysis strategy, analysis of Ripley's distance histograms (pairwise distances between all signals) and nearest neighbor distances. These analyses provide first deeper insights into the organization of individual biomolecules within cellular/cell nuclear systems based on a priori knowledge of the simulation. The results are interpreted with respect to, among other things, (i) the formation and architecture of ionizing radiation induced repair clusters at DSB sites in different cell types and chromatin environments, (ii) the ability to distinguish different types of clusters (for greater clarity demonstrated on PML bodies vs. PML/RAR microspeckles instead of repair clusters), (iii) the arrangement of ionizing radiation induced repair clusters in the cell nucleus under specific irradiation conditions (exposure to different types of ionizing radiation), and (iv) the changes in mutual spatial relationships between specific chromatin elements (Alu consensus sequence regions and heterochromatin domains) in non-irradiated and irradiated cells.

In the next step (Section 2.3), the time course of changes in the organization of chromatin and individual ionizing radiation-induced repair clusters during DNA double-strand break (DSB) repair is analyzed using novel techniques based on persistent homology, persistent imaging, and principal component analysis (PCA). A real analysis of the spatial distribution of Alu regions dispersed in the DNA [64,67] revealed a characteristic shift along the first two principal components extracted from persistent imaging, providing evidence of changing chromatin organization during DSB repair, which are inaccessible with traditional methods.

## 2.2. Distances based analysis

### 2.2.1. Theoretical background to distance histograms

Initial information on the intracellular distribution of fluorochrome-labeled single molecules of interest (further referred to as points or signals) is gained from distances calculated for all signal pairs (Euclidean distance histograms). This can be done by using the `cldist` algorithm given in the Python package `scipy` [69,70]. Alternatively, only the shortest or few shortest distances can be evaluated [68,71]. By ordering these distances (referred to as the 'next neighbor distances') for each signal position, we can learn about the signal density in the immediate surroundings of each signal. Using the DBscan algorithm (Density-Based Spatial Clustering of Applications with Noise) [72], information can be obtained as to whether fluorophores occur in clusters or are distributed rather randomly. Another important capability of cluster analysis is the ability to analyze mutual colocalizations (i.e., interactions) between different molecular components, e.g., within DSB repair clusters or between these clusters and chromatin, etc.

Histograms of mutual distance frequencies provide a basic insight into a given point cloud as they reveal abnormalities to the random background. Such abnormalities can be interpreted in terms of biological organization features and, in turn, biological functions. To understand how various abnormalities can be interpreted, it must first be clarified what a random uniform distribution of points in such a histogram looks like. For a random uniform distribution of points, the frequency of distances linearly increases with their size. The area  $A$  (equation 1) of the considered ring with a bin width  $d$ , is

proportional to the distance  $r$ . Then, as the radius  $r$  increases, the area of the ring  $A$  grows linearly, and thus the probability of occurrence (number) of signals inside this area.

$$\begin{aligned} A &= \pi(r + d)^2 - \pi r^2 = (\pi r^2 + 2\pi dr + \pi d^2) - \pi r^2 \\ &= 2\pi dr + \pi d^2 \approx 2\pi dr \end{aligned} \quad (1)$$

In the case of a closed biological system, a Gaussian distribution of points is often observed inside, manifested as a Rayleigh distance distribution. Mathematically (Eq. 2) if  $X$  and  $Y$  represent two variable signal coordinates, normally distributed around zero with standard deviation  $\sigma$ , then the square root of the quadratic sum  $R$  is Rayleigh distributed.

$$R = \sqrt{X^2 + Y^2} \Rightarrow f_R(x) = \frac{x}{\sigma} \exp\left(\frac{-x^2}{2\sigma^2}\right) \quad (2)$$

$R$  therefore represents the distance of a two-dimensional Gaussian distribution. Unlike the similar Ripley's K function [71], where all distances within a certain radius are considered, in our case only points inside a given ring (one distance histogram bin) are used. This was implemented to prevent averaging cumulative effects. An illustrative example of the application of Ripley's K function in biology is given elsewhere [73].

Depending on the biological problem, various deviations from the simple Gaussian point distribution, derived from the freely jointed chain model [74], can be considered. More realistic models have been proposed that take excluded volume interaction into account [75]. This is leading to the accumulation of larger distances because the inner volume is already full. Such properties can be modeled (see Section 5.4 in [76]).

A linear curve of a uniform point distribution provides the basis for further analysis, with any deviation from this linear trend indicating a specific distribution of signals. It should be noted that its slope and smoothness depend on the density of the point cloud and the differences in smoothness can be explained by the law of large numbers.

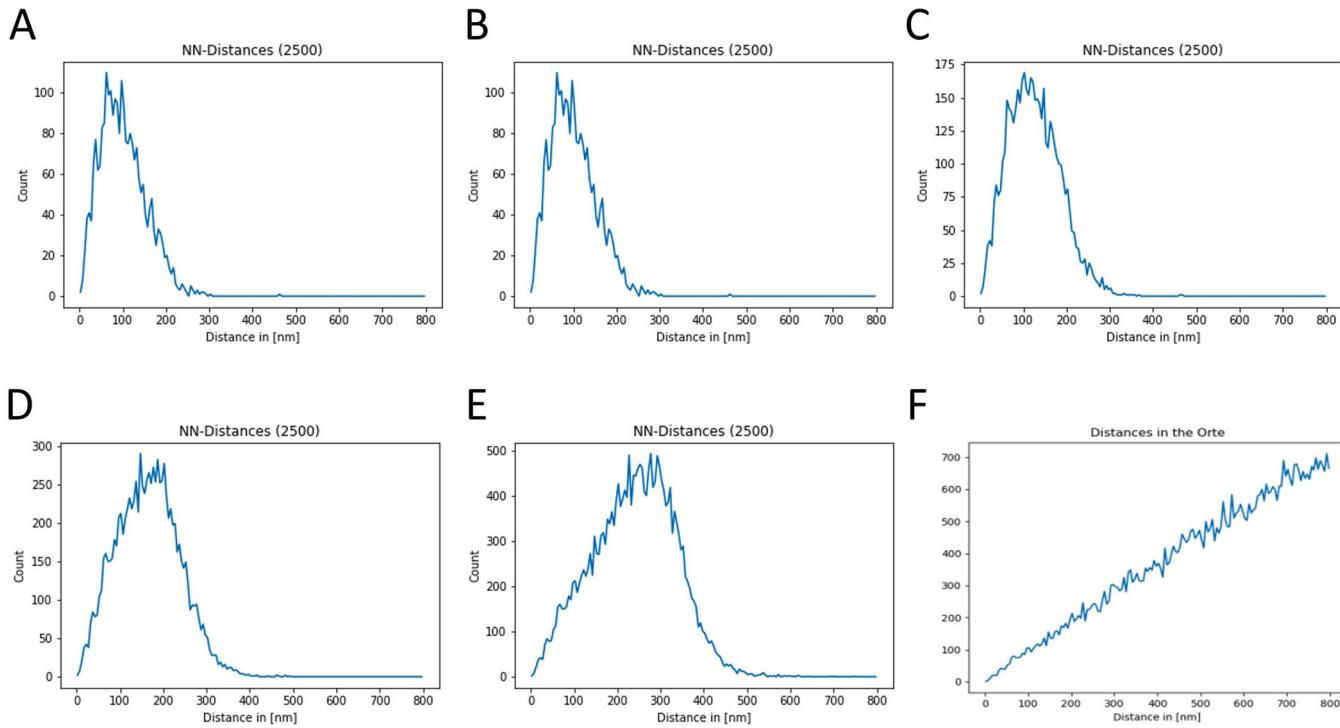
In addition to the all-to-all distance histogram, only nearest neighbor (NN) distances are considered in Fig. 1. These distributions provide a picture of the immediate surroundings of particularly dense regions, as these regions dominate the statistic. By a reasonable selection of allowed neighbors the distribution can be reduced to the part of interest. In other words, different local properties can be analyzed using the width and general shape of the peak depending on the allowed number of next neighbors.

### 2.2.2. Examples of biological clusters and cluster analysis

A phenomenon often occurring in (cell) biology is the formation of local accumulations of molecules, further referred to as clusters. As detailed later, molecular clusters, their topological architecture and cellular distribution are fundamental for self-organization of biological systems. Cluster analysis is thus critically important for understanding the functioning and regulation of biological systems.

Molecular clusters can be, for example, receptors on the cell membrane. A change in the topology of these clusters can then indicate their physiological or pathological activity [60,61,63,77]. Cyttoplasmic organelles [78], e.g., mitochondria or lysosomes, and nuclear bodies, e.g., PML or Cajal bodies, are further representatives of biomolecular clusters. The expression of the PML/RAR $\alpha$  oncoprotein in acute promyelocytic leukemia (APL) cells causes the disintegration of PML bodies into the form of so-called PML/RAR $\alpha$  microspeckles, which can be reversed by treatment. Proposed procedures of cluster analysis (see Fig. 2) can thus be advantageously demonstrated on the process of APL pathogenesis.

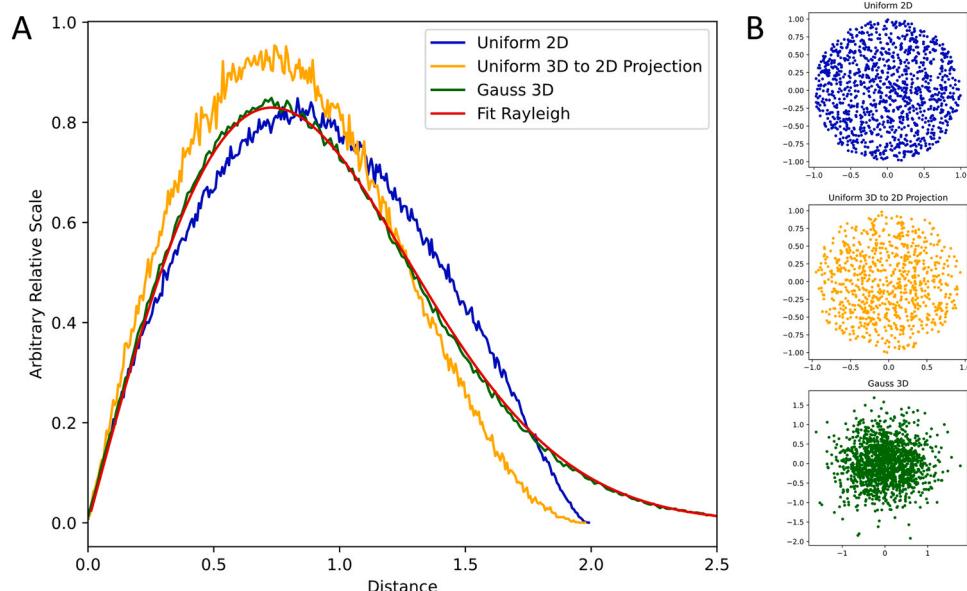
Among many other possible examples, DSB repair (IRIF) clusters, are used in this manuscript due to the high dynamics of their architecture and molecular composition. The induction of double



**Fig. 1.** Next neighbor distance frequency curves of 2500 uniformly randomly distributed points are shown. (A) only the first next neighbor; (B) two next neighbors; (C) three next neighbors; (D) five next neighbors; (E) 10 next neighbors; (F) all neighbors. The curve, linear for the all-to-all distance histogram, now shows a peak in the short distance region, as more distant points are not taken into account. As the number of neighbors increases, the left edge of the peak approaches a linear function, characteristic of the histogram with all distances included (F). The right edge drops off quickly because far points are not involved. Because of the underlying uniform distribution, the nearest neighbor distance histograms contain nearly a complete subset of small distances. This is punctuated by the strict cutoff at the right edge of the peak. A homogeneously filled point cloud with infinite density would result in a vertical cutoff. The further the measured point cloud is from this random distribution, the greater the deviation from the displayed triangular shape. Especially for clustering and other short-range interactions, nearest neighbor analysis becomes very valuable, as described below.

strand breaks (DSBs) in the DNA molecule initiates extensive trafficking of many different proteins to the damage sites [68], where they bind in an orchestrated manner first to free DNA ends and subsequently to nascent structural protein platforms. This leads to gradual formation of sophisticated repair complexes, manifesting in the nucleus as IRIFs [68,79–81]. Importantly, IRIFs contain proteins common to all repair mechanisms as well as proteins exclusively

involved only in one of the repair pathways (non-homologous end joining, NHEJ, or homologous recombination, HR). The interaction between chromatin structure, type of ionizing radiation ([50,82,83]; reviewed in [84]), cell cycle phase [85] and cell type [50,53,86] has been shown to influence the formation and internal organization of IRIFs at micro- and nanoscale resolution. Various radioprotectants and radiosensitizers (e.g., amifostine [87] and metal nanoparticles



**Fig. 2.** Three types of distance distributions (A) of underlying point clouds (B) are compared. A homogeneous circle in 2D (blue), a homogeneously filled sphere in 3D (orange) and a Gaussian distribution (green) that is fitted with a Rayleigh distribution (red).

[88], respectively) can also directly or indirectly affect DSB repair and the structure of IRIFs and chromatin [85,89]. Importantly, at nanoscale resolution, IRIFs can be divided into (sub)clusters [80,90]. It is thus hypothesized that the chromatin architecture at the DSB site influences the architecture of IRIFs, which in turn controls loading of additional specific proteins to particular DSB sites, and thus the repair mechanism selected at these sites [55]. In addition to questions concerning the basic principles of biological regulation, as in the case of the aforementioned hypothesis, the proposed procedures of geometric and topological analysis can offer answers to an unexpected number of specific questions related to radiobiology and cancer biology, as we show below. For example, since different types of ionizing radiation damage chromatin in different ways, knowledge of the geometry and topology of IRIF point clusters at different chromatin locations, and their changes over time after irradiation, appear to be promising indicators of cell fate after exposure to specific types of ionizing radiation in radiotherapy or space research [3,38,53–55,68,80,85,89,91,92].

Distance and cluster analyses also offer another possibility of evaluating SMLM data – tracking the “colocalization” of different types of molecules within clusters, based on the analysis of distances between the signals of different fluorochromes [93]. Interactions between labeled molecules can be quantified much more reliably than based on “color overlay” at standard optical microscopy [94].

**2.2.2.1. Internal cluster form.** As already stated, the clusters represent local biological events like repair protein accumulations at double strand breaks. The number of clusters, mean cluster size and the proportion of signals which are detected as clustered can be obtained. The underlying random uniform pattern distribution can be caused either by measurement noise of the microscopy procedure or due to imperfect sample preparation (unbound or non-specifically bound probe) [67]. Importantly, a biological ‘background’ can also occur, meaning the existence of a pool of (correctly labeled) free proteins that is not locally accumulated in clusters but is dispersed throughout the nucleus.

Another important aspect of clusters is their internal arrangement. To simulate this phenomenon, a comparison of distance diagrams resulting from different underlying point cloud distributions was conducted. Three different types of signal distributions are considered: a Gaussian distribution,<sup>2</sup> a 2D uniform circular distribution and a 3D uniform spherical distribution projected to 2D. It is shown that the overall shape of the resulting distance distributions is similar, independent of the exact local point distribution (Fig. 2).

A Gaussian distribution occurs when a central event leads to accumulation or diffusion towards or away from the event site. This can be seen, for instance, during DNA repair, where repair proteins gather around the site of damage and later leave it to allow for further repair steps. In contrast, a uniform distribution occurs when a certain area is covered evenly with the targeting fluorophores. The point distribution also depends on the 2D or 3D nature of the analyzed specimen. The SMLM specimens can be thought of as 2D on a large scale but 2D projections of 3D data on a small scale. Both cases are compared in Fig. 2. As given by Eq. 2, the Gaussian distribution is described by the Rayleigh fit. As can be seen from Fig. 2, all considered point clouds generally lead to a similar distribution. This allows the following analysis without paying too much attention to the exact internal distribution of the clusters.

Clusters within randomly distributed points result in frequency curves with characteristic peaks at short distances, as explained in Fig. 3. More detailed information about signal clusters can be

deduced from the next-neighbor (NN) distributions, as explained earlier (Section 2.2.2.1). In Fig. 4, the next-neighbor distributions are presented for 100 clusters in a point pattern of 2500 points. A characteristic peak can be observed that changes its position and shape with the number of neighbors considered. Most of the nearest neighbors lie within a short distance, indicating clustering of points. In comparison to the histograms of distances among all detected signals, the next-neighbor analysis provides information about the short distance interaction of fluorophores within clusters. The exact properties we are interested in are highly dependent on the biological question; however, the analysis is generally applicable to situations where different forms and shapes of clusters are compared.

Based on the point density, the clusters can be automatically masked by DBscan [72], which allows separate analyses for points within and outside clusters (Fig. 5). In this way, the distribution of the points within clusters can be compared to the various theoretical distributions (simulated in Fig. 2), providing conclusions on the character of the internal cluster organization. Cluster masking and decoupling analysis procedures are necessary to address many questions, because signals within and outside clusters have different biological meanings and the internal arrangement of clusters reflects their biological functions. In addition, unwanted background noise can be removed. When analyzing the background, attention must be paid to the fact that the underlying distribution does not contain the entire background, because holes remain in the original cluster locations. This problem can only be neglected for distributions with few clusters.

In comparison to the next neighbor analysis, the clusters here are determined by the density (DBscan). This approach is therefore particularly useful in cases where it is possible to clearly separate clusters from the surrounding environment. For samples that are characterized by a smooth transition between clusters and the surroundings, the next neighbor analysis can provide better insights.

Experimental application of the algorithms described above for complex signal distribution analysis, including cluster analysis with cluster and background signal separation, is demonstrated on real SMLM data for PML bodies in acute promyelocytic leukemia (APL) pathogenesis. This experimental system nicely illustrates the advantages of the proposed methodology, as separate analysis of the distribution of signals forming PML clusters, PML/RAR $\alpha$  (micro) clusters and the background, is essential for studying the mechanism of APL development (Fig. 6).

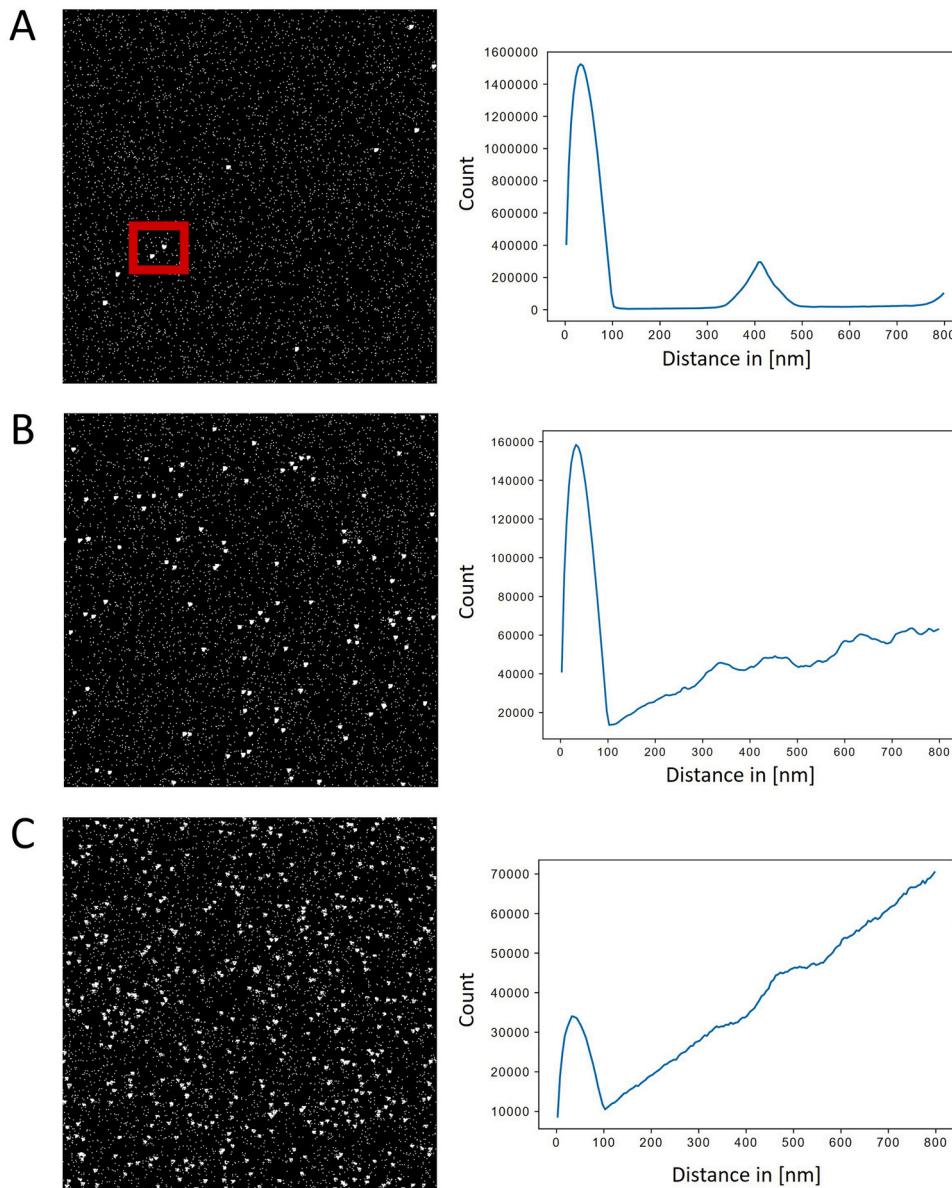
**2.2.2.2. Cluster formation within tracks.** Another important capability of our proposed procedures is to analyze the mutual distribution of clusters (Fig. 7). In contrast to Fig. 5, where the internal organization of the clusters (distributions of signals within clusters) as well as the distribution of signals within the background was of interest, the analysis in Fig. 7 focuses on the mutual organization of the clusters. The distance diagram in Fig. 7B, which includes all distances between signals within clusters and in the background, reveals global features of the signal distribution, i.e., both the formation of clusters and their non-random arrangement in a linear track.

In contrast to Fig. 5, where clusters are randomly distributed in the cell nucleus, a linear arrangement of clusters in Fig. 7 leads to formation of a wide second peak due to frequent occurrence of distances 400–600 nm between clusters. An illustrative biological example of a random and non-random (linear) cluster distribution is the distribution of repair protein (IRIF) clusters in cells irradiated with photon (simulation in Fig. 5) and densely ionizing particle radiation (Fig. 7), respectively. The parameter independence of this analysis is a great advantage compared to, e.g., cluster analysis.

### 2.2.3. Multi-color channel (multi-point cloud) analysis

A frequently asked question of fundamental importance in biology is whether two different (or more) biomolecules interact

<sup>2</sup> For the Gaussian, a 2D distribution and a 3D to 2D projection are equivalent. So only one simulation was conducted. 1500 events are used for each simulation.

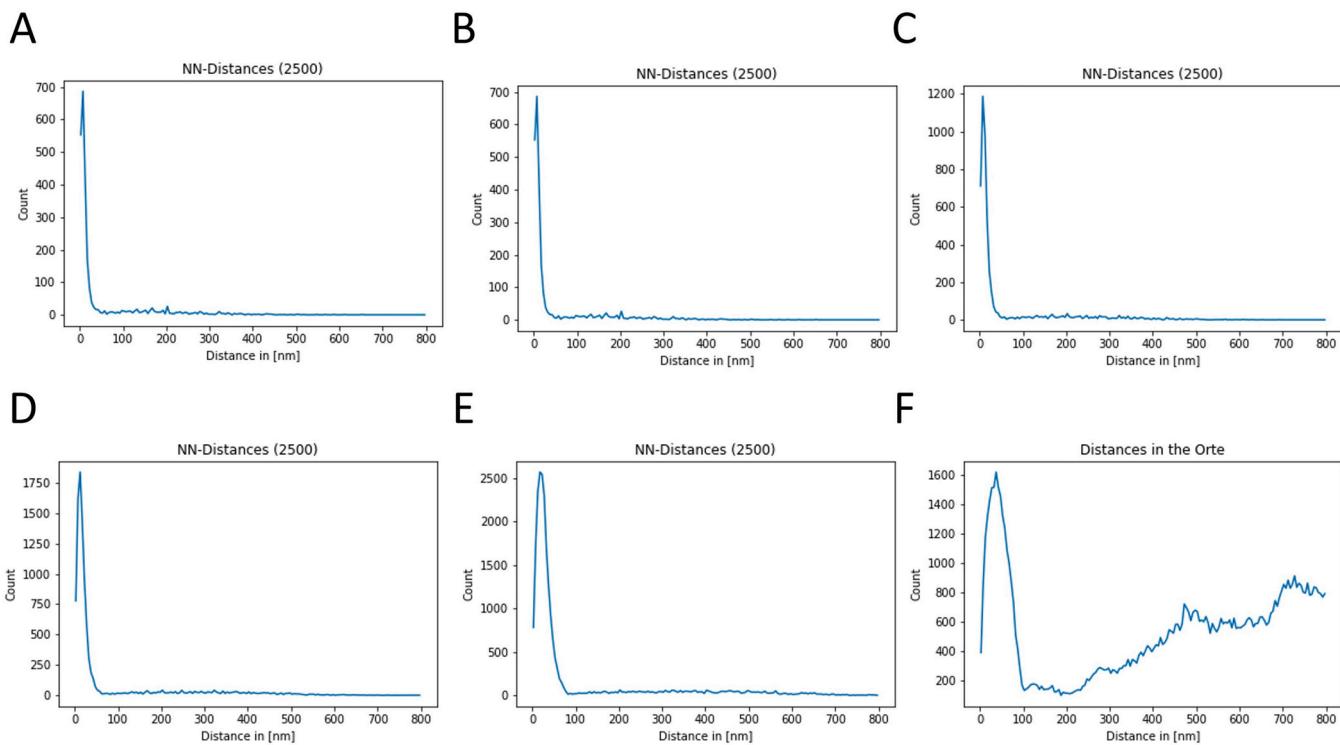


**Fig. 3.** Clusters (white spots) within a cloud of 25,000 points per  $10,000 \times 10,000 \text{ nm}^2$ . The curves indicate the clusters by a characteristic peak at small distances. The peaks at larger distances on top of the linearly increasing line (random point pattern) refer to inter-cluster distances. (A) 10 clusters; (B) 100 clusters, (C) 500 clusters. The area per cluster, the number of signals within the random uniform background distribution, and the number of signals within all clusters are all constant. This means that the number of signals per cluster decreases with an increasing number of clusters. As the number of signals per cluster decreases (6A→C), the peak height shrinks by the same factor. The width of the peak determines the cluster size. Because all clusters are simulated with the same size, the width is the same. Since the number of background signals is kept constant in the given simulation, the slope of the linear part (distances larger than 100 nm) is also constant. Only the presence of other clusters cause disturbance of the linear function. For example, a second peak is visible in the case where only a low number of clusters (A, 10 in this simulation) is present. This is due to the fact, that two clusters have a distance of 400 nm (red). With a growing number of clusters, the linear part (200–800 nm) smoothens based on the law of large numbers.

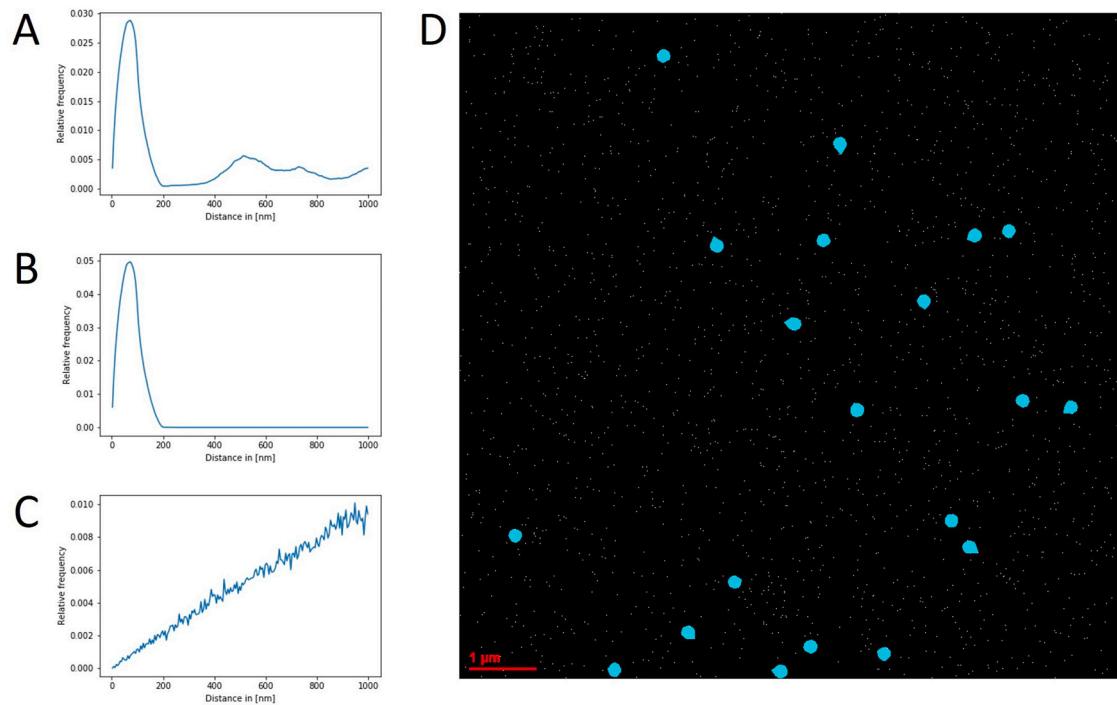
with each other, in the language of microscopic data, whether signals from differently labeled biomolecules ‘colocalize’ with each other. Colocalization analysis can be used to investigate the interaction of multiple proteins or other biomolecules, or even their specific parts (e.g., specific DNA sequences). In addition to the detection of local interactions of (bio)molecules, it is also possible to analyze their dynamics and the arrangement of molecules in forming molecular complexes. The matrices of signal coordinates can thus be interpreted in terms of the function of various biomolecules and the mechanism of the studied processes. An illustrative example of two-channel analysis of colocalization (i.e., interaction between clusters of biomolecules) can be the formation of IRIFs at DNA double-strand break sites. Another example is shown in Fig. 8,

where the spatial relationship between heterochromatin and clusters of Alu repeats is studied.

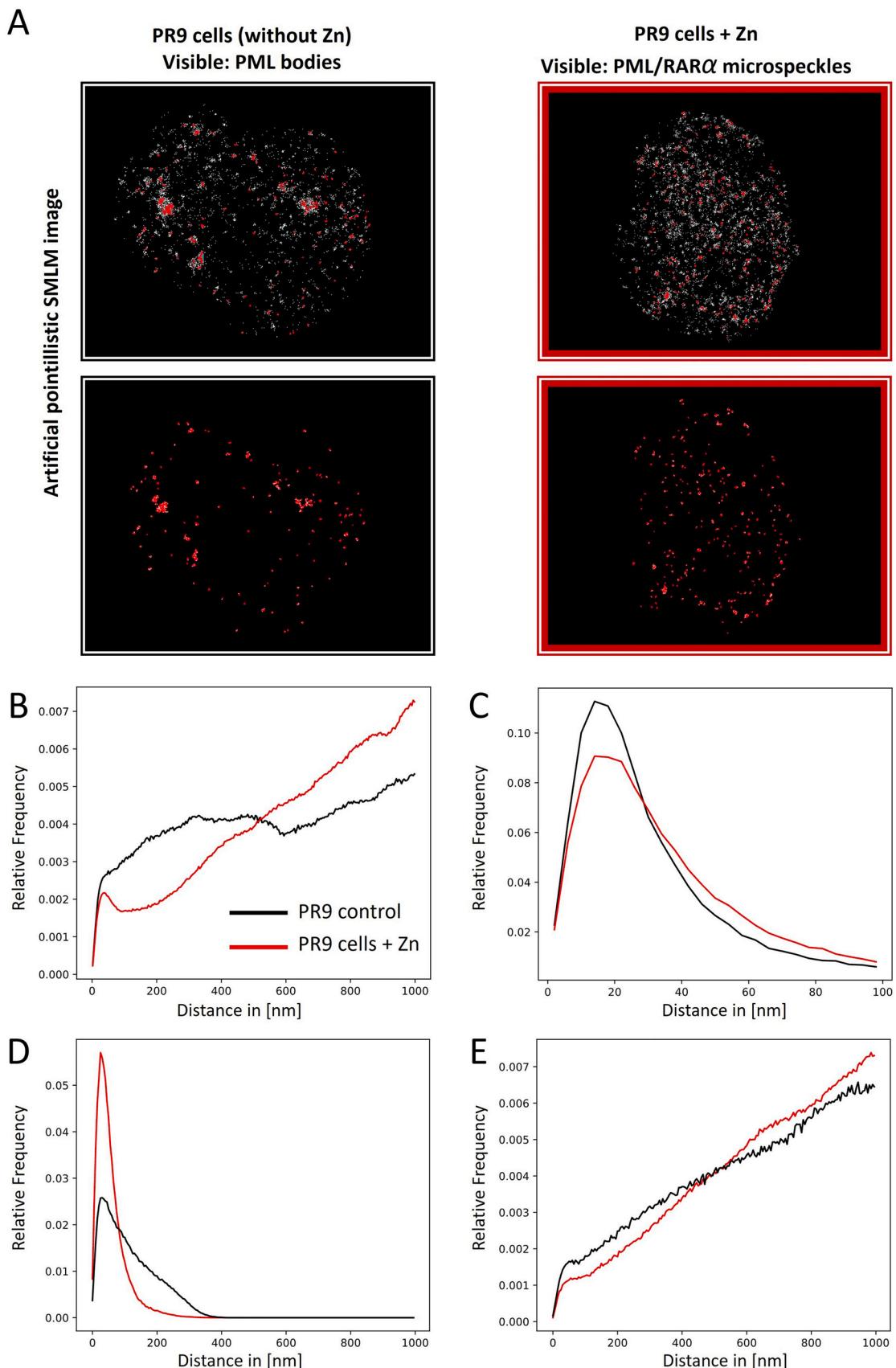
Ortho-matrices of two different and differently labeled molecules (e.g., proteins X + Y, protein X + DNA sequence, DNA sequences X + Y, etc.) can be evaluated separately for each of the color channels (as already described) or in an inter-matrix manner (Fig. 8). In the latter case, the distances of one point in the first channel from all points in the second channel are determined and frequency curves of these distances are generated for all points in the first channel. For this purpose, the individual coordinates of each point of the first color channel are transferred to the second channel and the Euclidian distances are determined, as already described for the single-channel analysis.



**Fig. 4.** Next neighbor distance frequency curves of 2500 points with 100 clusters integrated. (A) only the first next neighbor; (B) two next neighbors; (C) three next neighbors; (D) five next neighbors; (E) 10 next neighbors; (F) all neighbors (for comparison see Fig. 3 with a random uniform point distribution).

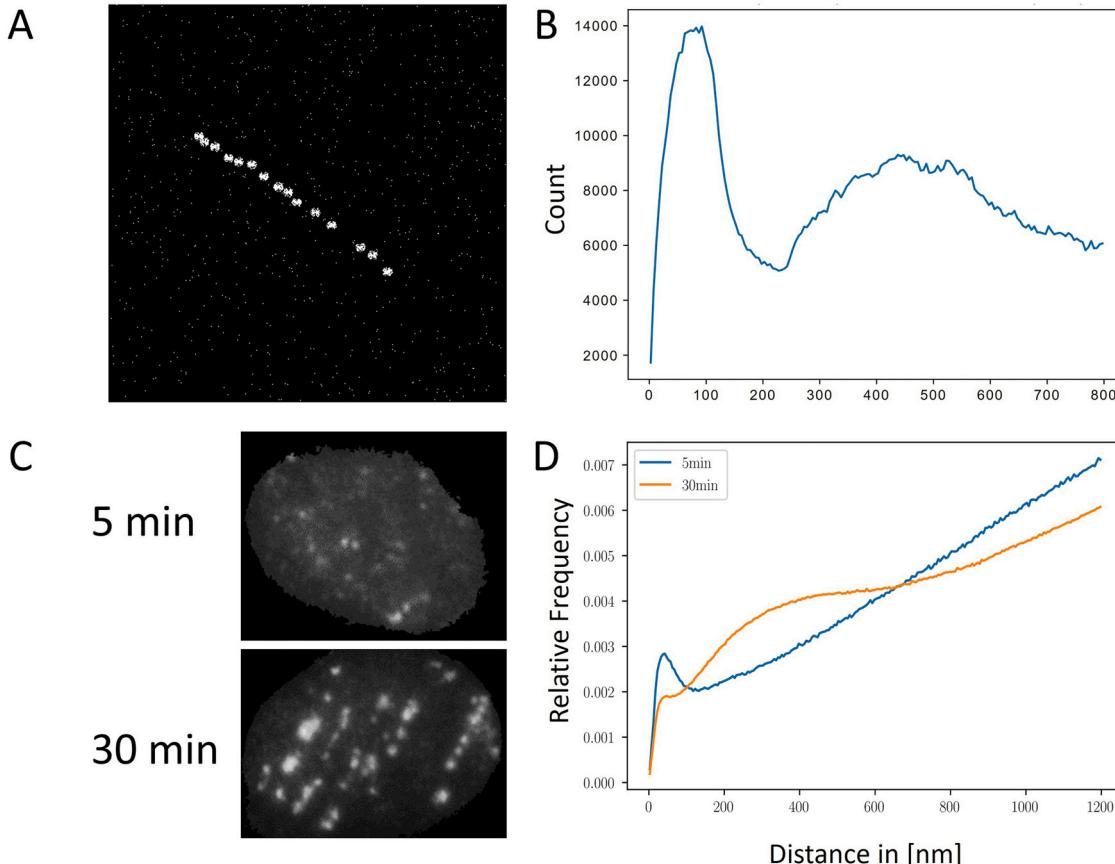


**Fig. 5.** Masking of clusters in 10,000 signals and 20 simulated clusters with background. The upper curves show all distances (A), the middle the distances inside each cluster (B), and the lower the distances outside clusters (C). A visualization is shown on the right (D). In (A), the second and third peaks of the distribution are caused by nonrandom mutual distances between clusters. These peaks are removed after the filtration of signals outside clusters in (B) or clustered signals in (C). Hence in (C), a linear growth can be observed that represents the uniform background. It is obvious that plot A allows the evaluation of the overall character of the distribution of the given molecule, i.e., the involvement of this molecule in different cellular (sub)networks. Cluster masking then allows to study these (sub)networks separately (B vs. C). When comparing the plots of A, B and C, it should be borne in mind that relative values are displayed and the scales of the vertical axes of these plots therefore differ significantly.



(caption on next page)

**Fig. 6.** Application of the comprehensive SMLM signal distribution analysis to real experimental system based on U937 PR9 cell, using the software procedures described in the main text, including the cluster analysis with cluster and background signal separation. U937 PR9 cells have an incorporated PML/RAR $\alpha$  fusion gene, which is regulated by a zinc-inducible promoter. In the absence of zinc ions in the culture medium, PML/RAR $\alpha$  expression does not occur – the cells behave like healthy promyelocytes and contain undamaged PML bodies. After adding Zn $^{2+}$  ions to the culture medium, the fusion oncogene PML/RAR $\alpha$  is expressed and causes disintegration of PML bodies into many so-called PML/RAR microspeckles – the cells now mimic acute promyelocytic leukemia (APL) cells. PR9 cells incubated with 100 M Zn $^{2+}$  for 6 h and expressing PML/RAR microspeckles (red curves) are compared to PR9 control cells with normal PML bodies (black curves). Detected cluster signals are shown for the control and treated cells (A). In the first row, cluster signals (red) and non-cluster signals (gray) are combined while in the second row only cluster signals are depicted. The distance distribution (B) and Next-Neighbor analysis (C) was conducted. The distance histogram (B) can be split in a diagram of points inside a cluster (D) and points outside clusters (E).

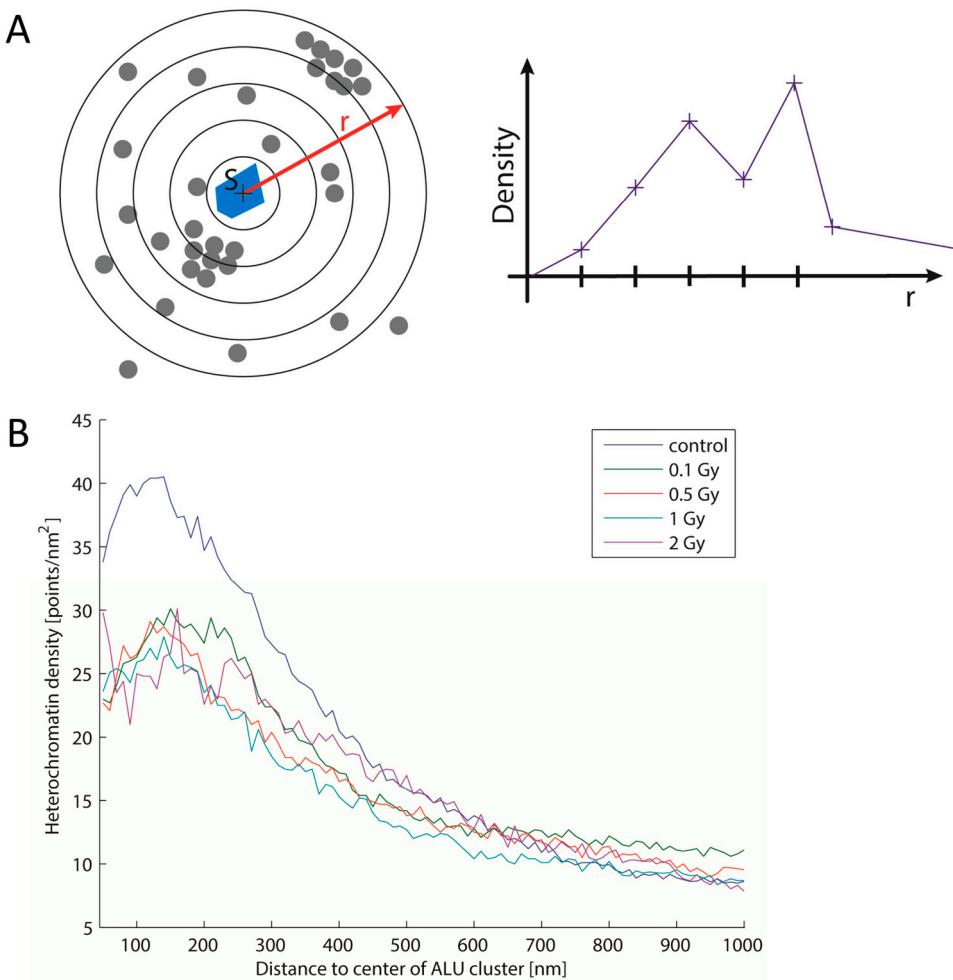


**Fig. 7.** Simulation of a rectilinear cluster track (A) and the corresponding distribution of all signal-to-signal distances. Fifteen clusters were identified and masked along the track within 10,000 signals. Panels C and D show an example of real data analysis of IRIF focus spatial arrangement. (C) Overview wide-field immunofluorescence microscopy images of 53BP1 foci in normal human dermal fibroblasts (NHDF) irradiated with densely ionizing  $^{15}\text{N}$  ions (13.1 MeV/n, 181.4 keV/ $\mu\text{m}$ , 1.25 Gy, 2.1 particles per nucleus on average) and fixed at 5 min or 30 min post-irradiation. (D) The all-to-all distance distributions of SMLM signals acquired for nuclei shown in (C). The curve for 5 min after irradiation shows a random distribution because only some 53BP1(C) clusters have evolved so far. At 30 min post-irradiation, 53BP1 clusters are already fully developed and organized in several (~6) parallel linear tracks (C); this is evidenced by the presence of two peaks—a sharp peak at a distance of ~50 nm, which reveals the formation of clusters, and a second broad peak at a distance of about 200–600 nm—which points to a non-random (rectilinear) arrangement of clusters. The decrease of the first peak (< 100 nm) in the curve for 30 min post-irradiation compared to the curve for 5 min post-irradiation is at first sight paradoxical due to the presence of more clusters in the 30 min PI. However, this decrease is relative and is due to a change in the spatial arrangement of the clusters (a second peak occurs at 30 min after irradiation). The biological reasons for the different distribution of 53BP1 clusters at 5 and 30 min after irradiation are as follows: The formation of DSB repair complexes (IRIF) in irradiated cells occurs with specific repair kinetics for each repair protein. In the case of 53BP1, this means that only a few foci are formed along each particle track 5 min after irradiation. This gives the impression of a random distribution of foci (clusters). In contrast, at 30 min post-irradiation, 'all' 53BP1 foci have already formed and their rectilinear distribution along the particle tracks is thus evident.

### 2.3. Topological data analysis

The axiom that structure determines function and vice versa can be considered as a universal rule not only in biology. Therefore, knowledge of the architecture of cellular structures, such as the DSB repair foci (IRIFs), and the ability to compare their topology in different cells, chromatin domains, post-irradiation times, doses and types of radiation, irradiation conditions, etc., is very important. In the case of classical microscopy, the analysis of cellular structures can be somewhat distorted by deconvolution and image analysis. Here, the advantage of SMLM becomes fully apparent, as we make use of the spatial coordinate matrices of individual biomolecules without the need for image analysis.

Topological features are those that remain invariant under continuous deformation. In particular, the invariance of rotation and bending (scale invariance) is of interest for the analysis of point cloud data inside the nucleus. Using this method, it is possible to find prominent parameters describing the underlying biological structure regardless of cell orientation and slight deformation induced by specimen preparation. The rotational invariance along the horizontal plane corresponds to the arbitrary orientation of the microscope slide (i.e., the single cell) on the microscope stage. In the local perspective, we do not measure only the nuclear section, but the projection from 3D to 2D. Thus, in addition to the irrelevance of the horizontal orientation of the section, we can also exploit the rotational invariance of the 3D orientation of the nucleus. This property is especially important for a substantial number of signals



**Fig. 8.** Two-channel analysis of heterochromatin distribution around clusters of Alu repeats [76]. (A) The distance distribution of heterochromatin (gray) around the center of Alu clusters (S, blue) is described. The density histogram (right) is equivalent to the distance histogram for a single (one color) matrix but it contains the distances between two different signal cohorts (i.e., Alu cluster centers and heterochromatin signals). (B) Mean distance distributions are shown for different cell exposures to ionizing radiation. About 200 Alu clusters were detected in each cell, with 30–40 cells were analyzed per radiation dose. It is clear from the graph that irradiation affects the density of heterochromatin around the Alu clusters (0–350 nm), as there is a clear difference between irradiated and non-irradiated cells. A dose-independent relaxation of the heterochromatic regions around Alu clusters can be observed. (Images originally published under CC BY license in Krufczik et al., 2017 [76]).

(> 10,000), as more points allow us to analyze structures in the magnitude of the Z-resolution limit. This leads us to consider the local point cloud as a projection rather than as slice. For a less dense sample, the distances between points are large compared to the Z-resolution. Thus, the whole sample can be seen as a thin disc. In this case, the point cloud data can be seen locally as a projection from 3D to 2D and globally as a slice.

### 2.3.1. Persistent homology

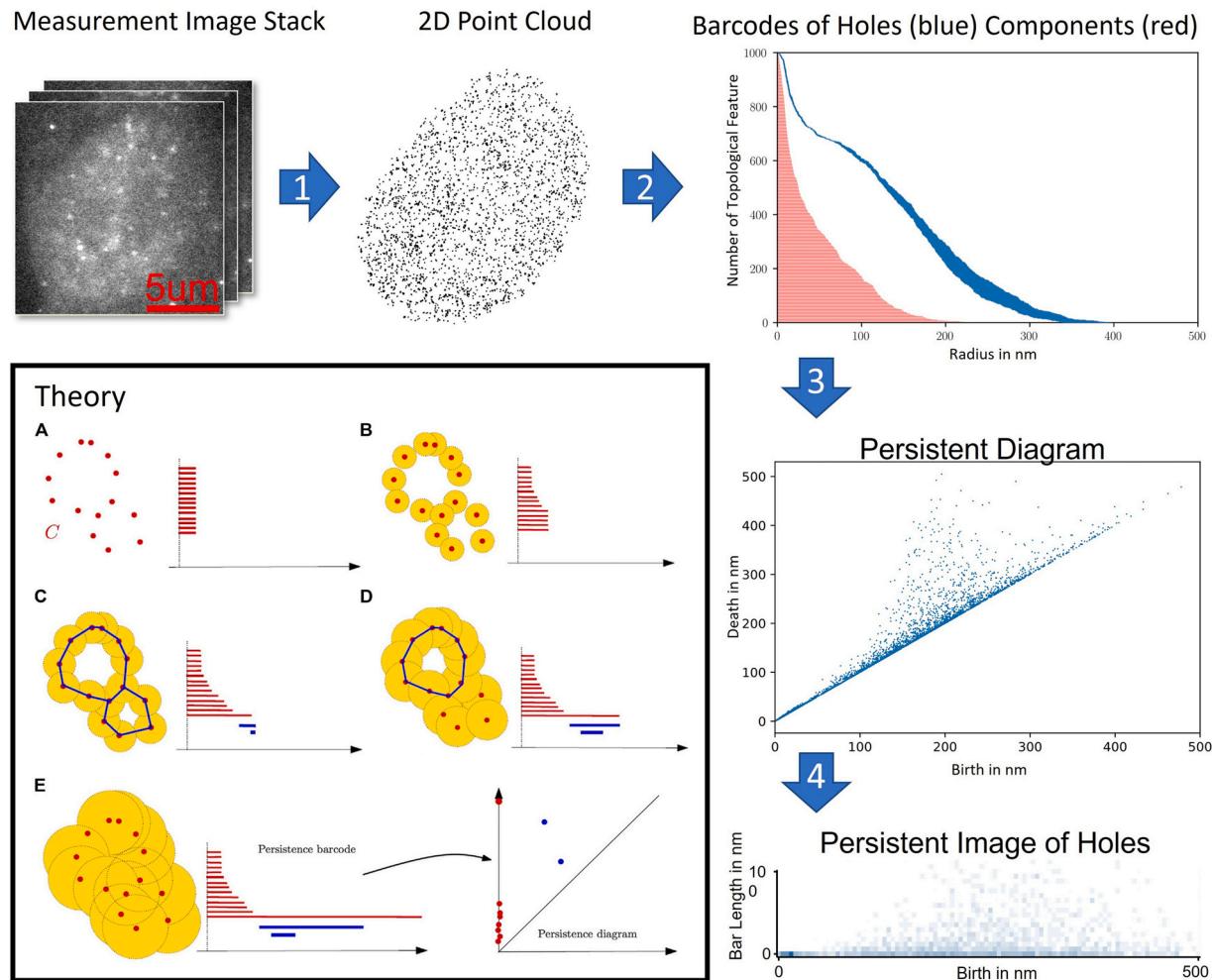
The persistent homology can be used to detect topological features [35,36]. This method aims to extract significant structures from point cloud data generated by SMLM. The extracted features can be represented in the form of a bar chart. Two key features are investigated as shown in Fig. 9. The first feature is called ‘component’, or ‘dimension 0 feature’ and describes the local neighborhood of each point. The second feature type is the ‘hole’, or ‘dimension 1 feature’. Unlike components, holes describe the empty spaces between points. The description of components and holes as dimensions were chosen according to the Betti number (for details see [36]).

To extract topological features of studied objects, multiple graphs are gradually generated out of the point cloud for multiple radii  $\alpha$ . The distance matrix as a single input can only be used in the

Vietoris–Rips complex for a general metric space. Additional positional information is needed for the closely related Čech complex applied in the following. The Čech complex for a given radius  $\alpha$  is defined as the set of simplices  $[x_0, \dots, x_k]$  such that the  $k+1$  closed balls  $B(x_i, \alpha)$  have a non-empty intersection. By using this set of simplices one can extract the components and holes for a given radius. Finally, the barcode representation as shown in Fig. 9 is generated by sweeping over multiple radii and collecting the holes and components. In a nutshell, this means that the point cloud is investigated on different scales and the appearance of features is tracked for the respective scale. Then the overall topological structure can be analyzed based on the scale over which certain features ‘live’. A good explanation of the theoretical background was given by [95] and an illustrative video was produced by [96].

### 2.3.2. SMLM-specific normalization for persistent homology

The point density is important for the result of persistent homology. Therefore, it is necessary to normalize measurements that differ significantly in the number of detected signals. However, if data is normalized, significant information may be lost. For this reason, only the necessary filtering is performed. Large variation in counts and densities would lead to meaningless results in persistent homology. Therefore, filtration is necessary that preserve biological



**Fig. 9.** The principle of persistent homology is explained graphically in the ‘Theory’ box. The left side of each panel (A – E) shows the point cloud and the right side shows the barcodes generated for this cloud. The structure is then successively analyzed for increasing radii (circles) around each point (A to E), starting from the zero radius (A). For each point in the cloud, a red bar begins in the right-hand bar graph. (B) As the radius increases, some of the circles begin to overlap, which leads to the connection of the involved points. Each connection means the end (death) of the component marked by the end of the red bar. (C) Once the connected points close some point-free space, we call this space a hole. Emerging holes are indicated by a starting blue bar. (D) Disappearing holes lead to the end (death) of the corresponding blue bar. (E) At a certain radius, all points merge, i.e., only one red bar remains, while all holes disappear, ending all blue bars. The final barcode representation is equivalent to the persistence diagram shown in the last image (E, right). The radius of birth of a component or hole is shown on the x-axis, the radius of death (disappearance) is then shown on the y-axis. Components are located on the y-axis, starting from radius zero. Holes are located over the angle bisector, as they vanish after they appear. The order of bars in the bar chart is arbitrary. (Parts of the Figure originally published under CC BY license in [95]). Around the box, the whole workflow with example images of a real measurement is shown. The time sequential image stack, measured by SMLM, is transformed to a 2D coordinate-matrix of points (point cloud) (1). To extract parameters, the persistent homology TDA method was applied. The point cloud data (for specific chromatin domains) is represented in the form of topological features, like hole- (blue) and component (red)-barcodes (2), and persistence diagrams (3). Overlaying the persistence diagram with a grid and counting the occurring hole-features leads to a persistent image (4).

properties. Because the measurement is performed by SMLM, some fluorophores do not bind to the target and most are bleached before measurement. This measurement effect manifests itself linearly and uniformly in the first order. This means that only a small fraction of target molecules is measured. Therefore, a random uniform selection of points is performed as filtering. This should not affect the overall biological structure, as it preserves exactly those properties that are accessible with SMLM. For our experiments, normalization to 1000 – 10,000 signals seem to be most useful.

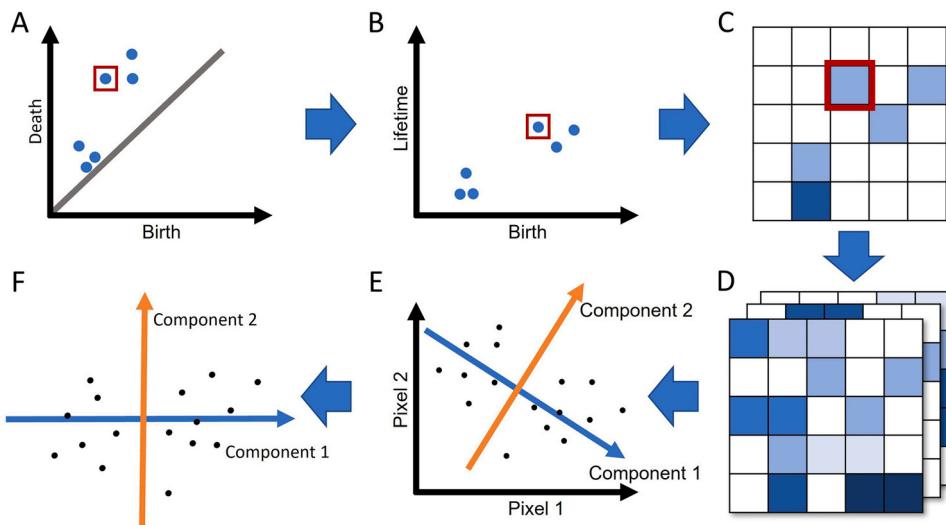
### 2.3.3. Comparison of multiple persistent diagrams

Different metrics are used to compare topologies obtained using persistent homology. This section explains which technique can be used for which application. The first method that comes into consideration is visual inspection of barcodes and persistent diagrams. This method is used to compare just two or three point clouds with a manageable number of signals. For comparing large groups of multiple point clouds, visual inspection becomes infeasible. Since this

critical number of measurements is needed to obtain statistically significant results, a different measure is required.

A metric that measures barcode overlap was developed by [35]. This similarity measure has been used in several recently described applications [36,54,91] showing the similarity of individual γH2AX and 53BP1 clusters measured by SMLM. For intercomparison of persistence diagrams, the bottleneck distance, the Wasserstein distance, and the Jaccard index provide a suitable metric [98,99]. What these methods have in common is that they compare one particular topological feature (hole) of one persistence diagram (or barcode) to the best fitting feature of the other persistence diagram (or barcode). Thereby it extracts the distance or similarity between these features. This approach is useful when comparing similar topological spaces, such as two circular point clouds with slightly different shapes and sizes. There, a single bar would be used for each circle with a larger overlap for more similar circles.

In SMLM measurements of the whole nucleus, it is not possible to find specific topological features that match the exact features in



**Fig. 10.** (A) Generation of persistence image [97]. In the first step the persistence diagram is folded down by  $45^\circ$  (B). The y-axis thus shows the lifetime instead of the disappearance (death) of the hole. In the next step, the diagram is converted to a grid (C). The number of points in each grid is represented by the color intensity. The red box shows the path for one hole in the persistence diagram. Based on persistent images, principal component analysis (PCA) is applied. Multiple persistence images (D) are transferred into a vector space where each pixel is represented as a dimension. Values for pixels 1 and 2 are shown in the first plot (E). In the next step, the basis vectors are rotated. The first component (blue) points into the direction of the largest variance. The next component (orange) must be perpendicular to all previous ones. Under this condition, it points into the direction of the largest variance. In the 2D case shown, there is only one possibility for the second component. Finally, the measurements are plotted with the new basis vectors (component 1 and 2) (F).

another nucleus. Rather, one has to compare the superposition of topological features, which means the entirety of bars. Therefore, a pictorial representation of persistence diagrams is used. Using the density of the persistent diagram, the superposition of topological holes is preserved (Fig. 10).

To further process and compare persistence diagrams it is useful to transform the barcode or persistence diagram to a point in a vector space. Such a vector can be displayed as a greyscale image. Each pixel represents one dimension, and the intensity represents a value in that dimension. The conversion of a persistence diagram to a persistent image [97] is shown in Fig. 10. Rotating the persistence diagram by  $45^\circ$  produces a persistence diagram that shows the lifetime of holes along the Y-axis instead of the death. The diagram is then converted to a grid of typically 500 – 3000 cells. The side length of a grid cell is typically set to 5 – 20 nm as this is the range of microscopic resolution. The sum of hole events in each grid cell results in the grey scale value of the persistent image. A good presentation is given by [100].

The persistent image was chosen instead of other possible vector representations, such as the landscape representation [101], because it reassembles typical features of SMLM data. A large-scale structure with repeating local structures is investigated. Thus, only the largest topological features (largest hole) should not be considered, as in the landscape representation. Also important is the often-irrelevant number of ‘short live’ structures closer to the diagonal of the persistence diagram. This is accounted for by the pixel value in the lowest row. The landscape representation is nonlinear (Section 2.2 in [102]). In contrast, the representation of persistent image is continuous. Slightly larger holes lead to small changes in the persistence diagram and small changes in the resulting image. This stability property of persistence diagrams has been demonstrated by [103]. Since the network structure of the whole nucleus is being investigated, the frequency of certain hole sizes matters. These differences in the occurrence of holes result in dense or loose regions within the persistence diagram. Finally, the persistent image preserves these superposition properties of the persistence diagram.

#### 2.3.4. Principal component analysis on persistent images

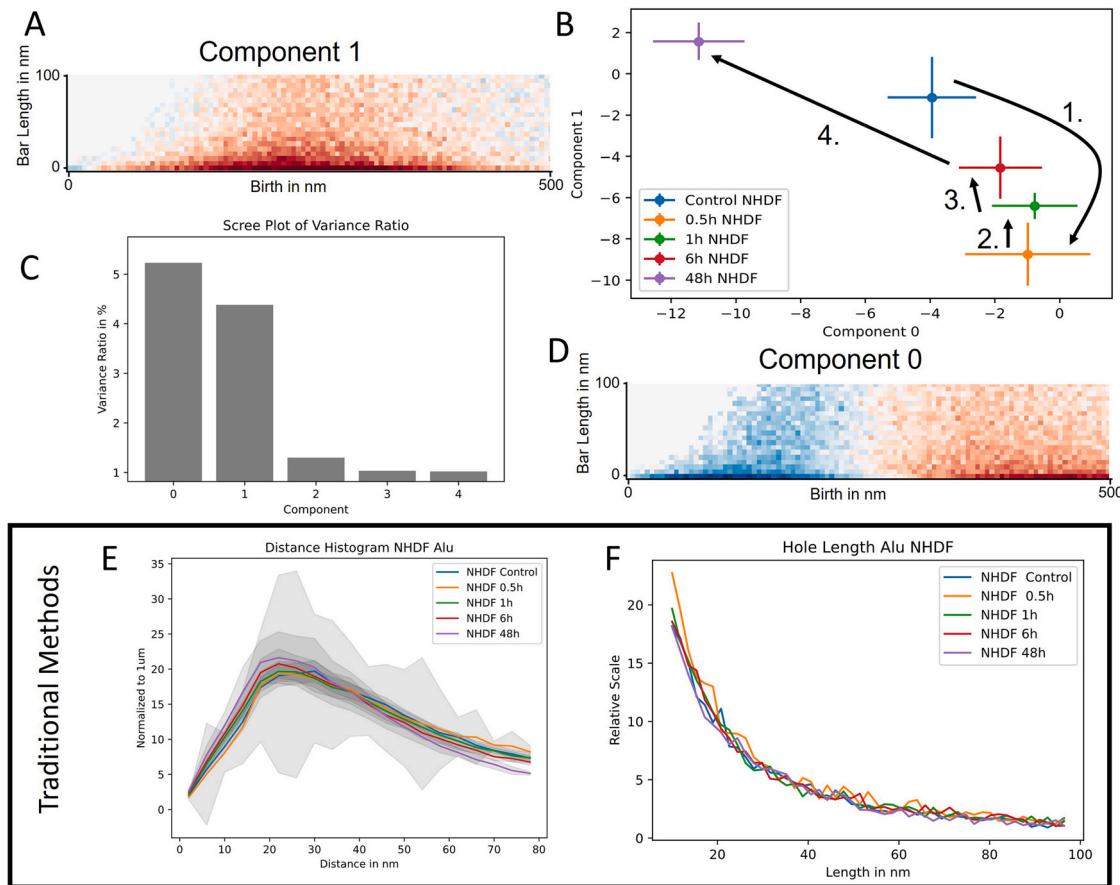
The goal of Principal component analysis (PCA) as presented here is to find the most important properties of persistent images. PCA is a multivariate statistical method developed by [37] that can also be considered an early approach to unsupervised machine learning. The principal components reveal only the important features and allow us to investigate the biological bases. In our case, we want to find the most important pixels of the persistent image, which represent important regions in the persistence diagram to separate different biological events. The persistence image can be viewed as a point in an N-dimensional vector space, where N is the total number of pixels. Generating a persistence image for each measurement of different nuclei leads to an N-dimensional point cloud in this vector space.

The PCA algorithm can be viewed as a rotation of the orthonormal basis that diagonalizes the covariance matrix. Furthermore, the new basis vectors (principal components) are sorted in descending order according to the proportion of total variance accounted for by each component. A graphical explanation is given in Fig. 10. By using only the first principal components, it is possible to reduce the dimensionality of the data and preserve most of the variance. Above all, however, it also holds the promise of finding accurate basic biological features that explain the structure of different cell types or different radiation treatments.

#### 2.3.5. Unveiling DNA damage repair pathway using PCA

The previously presented methods like distance distributions and pure persistent homology were found insufficient to describe cellular processes, such as the dynamics of DNA double-strand break (DSB) repair. The reason lies within the complex network structure of chromatin which is assumed to be affected during repair only in a small fraction. Thus, it is imperative to find a method with which differences in latent space can be extracted and brought back for biological interpretation through component images. This is offered by PCA (Fig. 11).

The scree plot (Fig. 11C) suggests a large influence of the first two principal components on the variability of the data. Visualizing the latent space of these two principal components (Fig. 11B) shows a significant variation for different post-irradiation times. The black



**Fig. 11.** Principal component analysis of Alu holes in the U87 cell line for different post-irradiation times. Cells were irradiated with low linear energy transfer (LET) photon radiation at a dose of 2 Gy. Aliquots of the same sample were prepared at 0.5 h, 1 h, 6 h and 48 h after irradiation. Slides were labeled with antibodies against  $\gamma$ H2AX to confirm DSB induction and repair and with oligonucleotides specifically visualizing Alu sequences in the DNA molecule. At least 20 nuclei were measured for each time point. The principal component 0 (A) and component 1 (D) of the persistence image space are shown. The relative variance, the variance in the principal component direction over the total variance, is presented as a scree plot for the first five principal components (C). (B) A projection onto the two most important principal components (0 and 1), known as the latent space, is plotted. The colored dots show the mean position of all measured nuclei for each post-irradiation time cohort. The error bars represent the standard error of the mean. The black arrows indicate the steps of the repair process. It is shown that traditional methods like the distance histogram (E) (variance in gray) and the length distribution of persistent holes (F) are insufficient to express the observed repair kinematics.

arrows indicate the repair pathway in the space of the first and second principal component. In the first step, 30 min after irradiation, a clear shift to the bottom right is visible (1. Orange). During the repair, at 1 h (2. Green) and 6 h (3. Red) after irradiation, the excitation is reversed back to the direction of the control slide. After 48 h (4. Purple) an overexcitation is detected that faces the opposite direction of excitation in latent space. Regarding the visualized principal components (Fig. 11A,D), the changes in hole distribution during repair can be investigated. A positive small deflection along component 0 is seen during repair, while component 1 becomes negative. Analyzing the image of component 0, it can be seen that more small holes in the range of 50–200 nm appear during repair (blue pixels), while the large holes (+400 nm) disappear (red pixels). The inverse deflection along component 1 reveals that also the region of holes between 200 nm and 300 nm is more frequent during the repair (red pixels). Overall, more holes appear in the region of 50–300 nm during repair.

These findings can be explained by DNA relaxation during repair (compare [76] and Fig. 8B), which leads to more open spaces in the smaller region that expands into the large empty spaces, resulting in a reduction of large holes as found in component 0. The reason for DNA unfolding could be the space occupied by phosphorylated H2AX ( $\gamma$ H2AX) or repair molecules, such as RAD51, which are required for different repair pathways. The contraction after repair could indicate a subsequent densely packed chromatin structure.

### 3. Discussion

Direct visualization of objects of interests by optical microscopy is an irreplaceable research method in cell biology. With this approach, fluorescently stained biomolecules form structured complexes and the cellular processes they participate in can be studied in the context of the overall natural environment under physiological conditions or even in living cells. Moreover, it is due to the nature of humans that they are usually more impressed by the image they see and can interpret it more easily than any other format of result presentation (one image is better than a thousand words). However, this is not the case for machines that rely on numbers as parameters to define structures within complexity.

The main advantage of microscopy—the image—thus confronts us with the myriad challenges when this data format has to be converted into numbers for computational analysis to enable the extraction of quantitative information. Thus, in reality, visual (manual) image evaluation usually still provides the most accurate results even in the age of advanced computing [104]. However, manual image analysis can only be relied upon in a very limited number of situations. It requires a great deal of analyst experience and can be extremely time-consuming. In addition, visual inspection of images alone does not allow quantitative assessment of cellular/biomolecule objects or object networks without further processing. Yet this capability is a prerequisite for research and diagnostic

applications of microscopy because, as is well known, the function of objects depends on their structure and vice versa [68]. Therefore, software detection, analysis, classification and matching of biological object architectures using image processing procedures is becoming a necessity [104]. Another major problem of visual assessment is the limited comparability of results between different laboratories.

Therefore, the analysis must be entrusted to software to ‘extract’ the most comprehensive and objective data from the microscopic images in a real time. However, in the absence of sufficiently sophisticated algorithms, computational image analysis is still very difficult or impossible. For example, we have shown this in our recent study [104], where we were unable to reliably detect even simple repair protein foci (IRIF) using available software (even dedicated software for this purpose) due to some variability in IRIF architecture and fluorescence over time and between samples. The need for new algorithms for image analysis and data interpretation becomes even more urgent with the shift of microscopy towards the superresolution (nanoscopy), where some image preprocessing is usually required [105], as well as the search for new ways to interpret previously unprecedented data. New AI-based algorithms may hold promise for the future [104], but commercial software development does not typically cover unique, rarely used microscopy techniques.

The image analysis tools using established algorithms, implemented for instance in the open-source program ThunderSTORM [106], have also become standard for Single Molecule Localization Microscopy (SMLM). The applications presented in [33] suggest their utility also for the cell nucleus research considered here, especially if the results thus obtained are supported by other experimental methods of molecular biology. However, many biophysical studies beyond molecular biology do not require visualization but quantification of parameters characteristic for a biological system, molecular arrangements, or parts of complex molecularly structured system components. For such parameters, statistic quantification, geometric features and topological measures are known to be relevant. To use such mathematical approaches, image analysis can be circumvented and a procedure to obtain pure coordinates, i.e., point values, would be more appropriate for abstract complexity reduction and evaluation.

In the present paper, we therefore introduce a suite of mathematical algorithms that enable comprehensive analysis and deep understanding of SMLM-generated data. For our approaches, the desired output of SMLM is not an image, but a matrix of spatial coordinates of each detected fluorophore leading to a point cloud of signals (see for instance [14,38,53,55,60,65,67,68,80,91,107]). This means that the coordinate matrix of all molecules generated by SMLM can be directly mathematically processed without the need for prior complex image analysis, which can be the most challenging and sometimes misleading step in microscopic data analysis, as already pointed out [104]. Thus, the presented tools lead directly to abstract but characteristic values of the system under investigation.

Another major advantage of SMLM is the resolution of this technique, given by the minimum measurable distances between signals on the order of 10 nm. This resolution allows a deep insight into the structure and function of molecular systems at the nanoscale, while the sample preparation is identical to confocal fluorescence microscopy, specifically immunofluorescent labeling of proteins with an antibody or labeling of DNA targets with fluorescently tagged oligonucleotides based on fluorescence in situ hybridization [65]. This means that the results can be obtained for the same samples in parallel at the micro- [82,83] and nanoresolution [53,54,85,89,92], and analyzed in correlation. Such a multidimensional analysis at different levels of resolution can be essential for understanding the functioning of complex cellular/biomolecular systems and their networks (reviewed in [4,5,55]). In this context, we would like to point out that the analysis of the architecture of

biological objects at different levels of resolution requires not only an appropriate combination of microscopic techniques, but also specific mathematical approaches. The persistent homology analysis, as discussed below, allows this in principle if density normalization is included. In many ways, SMLM also overperforms electron microscopy, which, while providing an order of magnitude better resolution, requires time-consuming and harmful cell processing. In contrast, SMLM samples can be prepared under near-physiological, i.e., 3D-conserving conditions and without the need for damaging labeling and sectioning.

The main advantage of SMLM—the mathematical analysis of the data independent of the image—can also be a major limiting factor in the use of this method, provided that suitable analysis algorithms are not available. On the other hand, with suitable approaches, SMLM data can not only be analyzed thoroughly and from different perspectives, but also secondarily produce artificial pointillist images without optical artifacts (see Fig. 6). These images can be used to visualize the desired parameters of the computational analysis such as intensity, clustering, colocalization of signals or their distance to other signals, similarity of signal clusters according to their architecture, or any other important properties of the system being analyzed.

The presented mathematical toolbox combines various existing and new models and algorithms and adapts them specifically for use on SMLM data:

1. A global overview of the object/network organization can be extracted from the point cloud by analyzing the distribution of distances between signals. Histograms of the distances between signals are based on Ripley's K function [71] because it is a simple and straightforward way to gain insights into the SMLM data represented by the signal point cloud. More sophisticated algorithms for studying point clouds can also be imagined. However, we have found this seemingly ‘simple’ frequency statistic of paired-point distances to be very robust and powerful, as it provides a straightforward way to gain insights into specific SMLM data that is well adaptable to very different types of cell exploration.
2. Clustering algorithms then allow the analysis of local events such as the formation of DNA repair foci (IRIFs), membrane receptors, various nuclear bodies, and changes of these molecular architectures in space and time. The DBscan algorithm [108] was chosen for its ability to summarize local accumulation of events as clusters.
3. Finally, new topological methods in combination with principal component analysis provide further detailed insights into the architecture of biomolecule clusters and biomolecular global networks. An invaluable benefit of rotation and deformation invariant persistent homology is that it allows the architecture of the molecular structures under study to be compared with each other at different locations within a single cell, between cells, and between different specimens. The persistent image representation [97] maps topological features of interest in a stable vector format. Based on this vector representation, principal component analysis [37] allows a traceable restructuring within cells. Moreover, when combined with vectorization, statistical methods such as the presented PCA allow deep insights into subtle geometrical and topological details that are only visible when analyzing large datasets. Converting individual SMLM measurements to a vector format opens the way to large-scale analyses using advanced statistical and machine learning methods that provide statistically significant and accurate results of biological events. This approach is particularly useful for investigating dynamic changes in these architectures over time and in response to various physiological or pathological stimuli or environmental stresses.

When jointly applied according to the proposed design, the above suite of algorithms provides a solid foundation for a brought set of possible applications to SMLM data and offers their detailed interpretation in terms of biological functions. A few examples of important questions that can be answered have been outlined in the results section, with chromatin and IRIF being the most frequent example of real experimental data. Compared to our previous studies [44,82,109,110], SMLM in combination with the proposed algorithms offered us the opportunity to compare the geometry and topology of IRIF foci, thus providing much deeper insight into the nature of DNA damage and repair under different conditions, e.g., in different cell types, in different chromatin domains, or after exposing cells to different types of ionizing radiation (e.g., [53,54,68,80,92,94,107]). We have shown that the architecture of IRIF is of functional importance but may differ in different (cancer) cells, which is probably related to the specifically affected ability of these cells to repair DSBs [53]. The IRIF architecture also depended on the local chromatin architecture. Thus, the IRIF architecture may represent an overarching physical mechanism that integrates myriad cellular factors contributing to the selection of a DSB repair mechanism, allowing the cell to rapidly decide on the most appropriate repair mechanism for a particular DSB [55].

However, the applications of the proposed procedures are far from being limited to the study of chromatin and cell nucleus. The arrangement of molecules in membranes, the transport of molecules in the cytosol, and essentially all molecular structures and processes in normal and pathologically altered cells can be studied [60–63,111]. This has been demonstrated by the two other examples—the molecular interactions and spatio-temporal behavior of PML/RAR $\alpha$  microspeckles in the APL pathogenesis, and the changes of membrane protein clustering in cancer cells [77,112]. The proposed mathematical toolbox has thus proved to be a powerful tool in many areas of research, including cell and tumor biology, cancer therapy and, for instance, in space exploration (manned spaceflight).

In the present manuscript, we offer a multi-angle and comprehensive view of the analysis of SMLM data. For this purpose, we considered it important to include not only the newest algorithms, but also those partially described in our previous papers (which also applies to some applications) [36,38,68]. Another important aspect of this manuscript is that the algorithms and the interpretation of possible results is first thoroughly explained through targeted simulations, and then the utility and robustness of these algorithms is demonstrated through challenging analyses of important cellular structures and processes. This approach, together with the presented improvements of the topological analysis, shows significant progress towards understanding of the translation of mathematical algorithms and their application to access structure–function relationships, structure-based control of interactions between biomolecules and/or cellular objects, the functioning of biomolecular networks as complex systems, and the biological relevance of structure-based (biophysical) regulation. Thus, the main advantage of these procedures lies in the reduction of complexity, so that the major characteristic and discriminative features of complex biological systems, which are usually subjected to biological variation, can be extracted. Finally, let it be mentioned that inappropriate labeling of molecules or signal registration can lead to certain data artefacts even in the case of SMLM [113]. Since these artefacts vary for different applications, it would not be appropriate to include any corrections in this direction in the proposed algorithms (except, for example, multiple registration of the same signal). Therefore, in the current manuscript, we assume correctly acquired data as detailed in [113].

## Acknowledgements

Funding of this research by the bilateral project of the Deutsche Forschungsgesellschaft (DFG, grant H1601/16-1) and the Czech

Science Foundation (project GACR 20-04109J) to Michael Hausmann and Martin Falk, and the German Federal Ministry of Education and Research (BMBF grant FKZ: 02 NUK 058A) to Michael Hausmann are gratefully acknowledged.

## Software

A tutorial for the topological analysis is provided on: [https://github.com/jonasw247/TDA\\_applied\\_on\\_SMLM](https://github.com/jonasw247/TDA_applied_on_SMLM).

The mathematical toolbox is available in the frame of collaboration.

## References

- [1] Bizzarri M, Naimark O, Nieto-Villar J, Fedeli V, Giuliani A. Complexity in biological organization: deconstruction (and subsequent restating) of key concepts. *Entropy* 2020;22:885. <https://doi.org/10.3390/e22080885>
- [2] Erenpreisa J, Giuliani A. Resolution of complex issues in genome regulation and cancer requires non-linear and network-based thermodynamics. *IJMS* 2019;21:240. <https://doi.org/10.3390/ijms21010240>
- [3] Hausmann M, Hildenbrand G, Pilarczyk G. Networks and islands of genome nano-architecture and their potential relevance for radiation biology: (a hypothesis and experimental verification hints). In: Kloc M, Kubiak JZ, editors. Nuclear, Chromosomal, and Genomic Architecture in Biology and Medicine, vol. 70. Cham: Springer International Publishing; 2022. p. 3–34. [https://doi.org/10.1007/978-3-031-06573-6\\_1](https://doi.org/10.1007/978-3-031-06573-6_1)
- [4] Erenpreisa J, Giuliani A, Yoshikawa K, Falk M, Hildenbrand G, Salmina K, et al. Spatial-temporal genome regulation in stress-response and cell-fate change. *IJMS* 2023;24:2658. <https://doi.org/10.3390/ijms24032658>
- [5] Erenpreisa J, Krigerts J, Salmina K, Gerashchenko BI, Freivalds T, Kurg R, et al. Heterochromatin networks: topology, dynamics, and function (a working hypothesis). *Cells* 2021;10:1582. <https://doi.org/10.3390/cells10071582>
- [6] Kloc M, Kubiak JZ, editors. Nuclear, chromosomal, and genomic architecture in biology and medicine. Cham: Springer; 2022.
- [7] Schermelleh L, Heintzmann R, Leonhardt H. A guide to super-resolution fluorescence microscopy. *J Cell Biol* 2010;190:165–75. <https://doi.org/10.1083/jcb.201002018>
- [8] Cremer C, Masters BR. Resolution enhancement techniques in microscopy. *Eur Phys J H* 2013;38:281–344. <https://doi.org/10.1140/epjh/e2012-20060-1>
- [9] Baddeley D, Bewersdorf J. Biological insight from super-resolution microscopy: what we can learn from localization-based images. *Annu Rev Biochem* 2018;87:965–89. <https://doi.org/10.1146/annurev-biochem-060815-014801>
- [10] Jacquemet G, Carisey AF, Hamidi H, Henriques R, Leterrier C. The cell biologist's guide to super-resolution microscopy. *J Cell Sci* 2020;133:jcs240713. <https://doi.org/10.1242/jcs.240713>
- [11] Jing Y, Zhang C, Yu B, Lin D, Qu J. Super-resolution microscopy: shedding new light on *in vivo* imaging. *Front Chem* 2021;9:746900. <https://doi.org/10.3389/fchem.2021.746900>
- [12] Lorat Y, Fleckenstein J, Görlinger P, Rübe C, Rübe CE. Assessment of DNA damage by 53PB1 and pKu70 detection in peripheral blood lymphocytes by immunofluorescence and high-resolution transmission electron microscopy. *Strahl Onkol* 2020;196:821–33. <https://doi.org/10.1007/s00066-020-01576-1>
- [13] Lorat Y, Reindl J, Isermann A, Rübe C, Friedl AA, Rübe CE. Focused ion microbeam irradiation induces clustering of DNA double-strand breaks in heterochromatin visualized by nanoscale-resolution electron microscopy. *IJMS* 2021;22:7638. <https://doi.org/10.3390/ijms22147638>
- [14] Eberle JP, Rapp A, Krufczik M, Eryilmaz M, Gunkel M, Erfle H, et al. Super-Resolution Microscopy Techniques and Their Potential for Applications in Radiation Biophysics. In: Erfle H, editor. Super-Resolution Microscopy, vol. 1663. New York, NY: Springer New York; 2017. p. 1–13. [https://doi.org/10.1007/978-1-4939-7265-4\\_1](https://doi.org/10.1007/978-1-4939-7265-4_1)
- [15] Lambert TJ, Waters JC. Navigating challenges in the application of super-resolution microscopy. *J Cell Biol* 2017;216:53–63. <https://doi.org/10.1083/jcb.201610011>
- [16] Dankovich TM, Rizzoli SO. Challenges facing quantitative large-scale optical super-resolution, and some simple solutions. *IScience* 2021;24:102134. <https://doi.org/10.1016/j.isci.2021.102134>
- [17] Zheng X, Zhou J, Wang L, Wang M, Wu W, Chen J, et al. Current challenges and solutions of super-resolution structured illumination microscopy. *APL Photonics* 2021;6:020901. <https://doi.org/10.1063/5.0038065>
- [18] Lemmer P, Gunkel M, Baddeley D, Kaufmann R, Urich A, Weiland Y, et al. SPDM: light microscopy with single-molecule resolution at the nanoscale. *Appl Phys B* 2008;93:1–12. <https://doi.org/10.1007/s00340-008-3152-x>
- [19] Lemmer P, Gunkel M, Weiland Y, Müller P, Baddeley D, Kaufmann R, et al. Using conventional fluorescent markers for far-field fluorescence localization nanoscopy allows resolution in the 10-nm range. *J Microsc* 2009;235:163–71. <https://doi.org/10.1111/j.1365-2818.2009.03196.x>
- [20] Kaufmann R, Lemmer P, Gunkel M, Weiland Y, Müller P, Hausmann M, et al. SPDM: single molecule superresolution of cellular nanostructures. In: Enderlein J, Gryczynski ZK, Erdmann R, editors. San Jose, CA: 2009. p. 71850J. <https://doi.org/10.1117/12.809109>

- [21] Cremer C, Kaufmann R, Gunkel M, Pres S, Weiland Y, Müller P, et al. Superresolution imaging of biological nanostructures by spectral precision distance microscopy. *Biotechnol J* 2011;6:1037–51. <https://doi.org/10.1002/biot.201100031>
- [22] Nieves DJ, Owen DM. Analysis methods for interrogating spatial organisation of single molecule localisation microscopy data. *Int J Biochem Cell Biol* 2020;123:105749. <https://doi.org/10.1016/j.biocel.2020.105749>
- [23] Nicovich PR, Owen DM, Gaus K. Turning single-molecule localization microscopy into a quantitative bioanalytical tool. *Nat Protoc* 2017;12:453–60. <https://doi.org/10.1038/nprot.2016.166>
- [24] van Leeuwen JM, Groeneveld J, de Boer J. New method for the calculation of the pair correlation function. I. *Physica* 1959;25:792–808. [https://doi.org/10.1016/0031-8914\(59\)90004-7](https://doi.org/10.1016/0031-8914(59)90004-7)
- [25] Khater IM, Nabi IR, Hamarneh G. A review of super-resolution single-molecule localization microscopy cluster analysis and quantification methods. *Patterns* 2020;1:100038. <https://doi.org/10.1016/j.patter.2020.100038>
- [26] Pike JA, Khan AO, Pallini C, Thomas SG, Mund M, Ries J, et al. Topological data analysis quantifies biological nano-structure from single molecule localization microscopy. *Bioinformatics* 2019;btz788. <https://doi.org/10.1093/bioinformatics/btz788>
- [27] Khater IM, Meng F, Wong TH, Nabi IR, Hamarneh G. Super resolution network analysis defines the molecular architecture of caveolae and caveolin-1 scaffolds. *Sci Rep* 2018;8:9009. <https://doi.org/10.1038/s41598-018-27216-4>
- [28] Sieben C, Banterle N, Douglass KM, Gönczy P, Manley S. Multicolor single-particle reconstruction of protein complexes. *Nat Methods* 2018;15:777–80. <https://doi.org/10.1038/s41592-018-0140-x>
- [29] Levet F, Hosy E, Kechkar A, Butler C, Beghin A, Choquet D, et al. SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data. *Nat Methods* 2015;12:1065–71. <https://doi.org/10.1038/nmeth.3579>
- [30] Andronov L, Orlov I, Lutz Y, Vonesch J-L, Klaholz BP. ClusterViSu, a method for clustering of protein complexes by Voronoi tessellation in super-resolution microscopy. *Sci Rep* 2016;6:24084. <https://doi.org/10.1038/srep24084>
- [31] Dlasková A, Engstová H, Špaček T, Kahancová A, Pavluch V, Smolková K, et al. 3D super-resolution microscopy reflects mitochondrial cristae alternations and mtDNA nucleoid size and distribution. *Biochim Et Biophys Acta (BBA) - Bioenerg* 2018;1859:829–44. <https://doi.org/10.1016/j.bbabiobio.2018.04.013>
- [32] Baddeley D, Cannell MB, Soeller C. Visualization of localization microscopy data. *Microsc Micro* 2010;16:64–72. <https://doi.org/10.1017/S143192760999122X>
- [33] Chapman KB, Filipsky F, Peschke N, Gelléri M, Weinhardt V, Braun A, et al. A comprehensive method to study the DNA's association with lamin and chromatin compaction in intact cell nuclei at super resolution. *Nanoscale* 2023;15:742–56. <https://doi.org/10.1039/D2NR02684H>
- [34] Zhang Y, Máté G, Müller P, Hillebrandt S, Krufczik M, Bach M, et al. Radiation induced chromatin conformation changes analysed by fluorescent localization microscopy, statistical physics, and graph theory. *PLoS ONE* 2015;10:e0128555. <https://doi.org/10.1371/journal.pone.0128555>
- [35] Máté G, Hofmann A, Wenzel N, Heermann DW. A topological similarity measure for proteins. *Biochim Et Biophys Acta (BBA) - Biomembr* 2014;1838:1180–90. <https://doi.org/10.1016/j.bbamem.2013.08.019>
- [36] Hofmann A, Krufczik M, Heermann D, Hausmann M. Using persistent homology as a new approach for super-resolution localization microscopy data analysis and classification of γH2AX foci/clusters. *Int J Mol Sci* 2018;19:2263. <https://doi.org/10.3390/ijims19082263>
- [37] Pearson III KL. On lines and planes of closest fit to systems of points in space. *Lond, Edinb, Dublin Philos Mag J Sci* 1901;2:559–72. <https://doi.org/10.1080/14786440109462720>
- [38] Hausmann M., Neitzel C., Hahn H., Winter R., Falkova I., Heermann D.W., et al. Space and Time in the Universe of the Cell Nucleus after Ionizing Radiation Attacks: A Comparison of Cancer and Non-Cancer Cell Response. The 1st International Electronic Conference on Cancers: Exploiting Cancer Vulnerability by Targeting the DNA Damage Response, MDPI; 2021, p. 15. <https://doi.org/10.3390/ECC2021-09219>
- [39] Pancaldi V. Chromatin network analyses: towards structure-function relationships in epigenomics. *Front Bioinform* 2021;1:742216. <https://doi.org/10.3389/fbinf.2021.742216>
- [40] Lohia R, Fox N, Gillis J. A global high-density chromatin interaction network reveals functional long-range and trans-chromosomal relationships. *Genome Biol* 2022;23:238. <https://doi.org/10.1186/s13059-022-02790-z>
- [41] van Mierlo G, Pushkarev O, Kribelbauer JF, Deplante B. Chromatin modules and their implication in genomic organization and gene regulation. *Trends Genet* 2023;39:140–53. <https://doi.org/10.1016/j.tig.2022.11.003>
- [42] Krigerts J, Salmina K, Freivalds T, Zayakin P, Rumnieks F, Inashkina I, et al. Differentiating cancer cells reveal early large-scale genome regulation by pericentric domains. *Biophys J* 2021;120:711–24. <https://doi.org/10.1016/j.bpj.2021.01.002>
- [43] Tsuchiya M, Wong ST, Yeo ZX, Colosimo A, Palumbo MC, Farina L, et al. Gene expression waves: Cell cycle independent collective dynamics in cultured cells. *FEBS J* 2007;274:2878–86. <https://doi.org/10.1111/j.1742-4658.2007.05822.x>
- [44] Falk M, Lukasova E, Gabrielova B, Ondrej V, Kozubek S. Chromatin dynamics during DSB repair. *Biochim Biophys Acta Mol Cell Res* 2007;1773:1534–45. <https://doi.org/10.1016/j.bbamcr.2007.07.002>
- [45] Sanders JT, Freeman TF, Xu Y, Goloshii R, Stallard MA, Hill AM, et al. Radiation-induced DNA damage and repair effects on 3D genome organization. *Nat Commun* 2020;11:6178. <https://doi.org/10.1038/s41467-020-20047-w>
- [46] Noon AT, Shibata A, Rief N, Löbrich M, Stewart GS, Jeggo PA, et al. 53BP1-dependent robust localized KAP-1 phosphorylation is essential for heterochromatic DNA double-strand break repair. *Nat Cell Biol* 2010;12:177–84. <https://doi.org/10.1038/ncb2017>
- [47] Goodarzi AA, Jeggo PA. The heterochromatic barrier to DNA double strand break repair: how to get the entry visa. *IJMS* 2012;13:11844–60. <https://doi.org/10.3390/ijms130911844>
- [48] Kumar R, Horikoshi N, Singh M, Gupta A, Misra HS, Albuquerque K, et al. Chromatin modifications and the DNA damage response to ionizing radiation. *Front Oncol* 2012;2:214. <https://doi.org/10.3389/fonc.2012.00214>
- [49] Nair N, Shoaib M, Sorensen CS. Chromatin Dynamics in genome stability: roles in suppressing endogenous dna damage and facilitating DNA repair. *IJMS* 2017;18:1486. <https://doi.org/10.3390/ijms18071486>
- [50] Falk M, Lukášová E, Štefančíková L, Baranová E, Falková I, Ježková L, et al. Heterochromatinization associated with cell differentiation as a model to study DNA double strand break induction and repair in the context of higher-order chromatin structure. *Appl Radiat Isot* 2014;83:177–85. <https://doi.org/10.1016/j.apradiso.2013.01.029>
- [51] Sanghi A, Gruber JJ, Metwally A, Jiang L, Reynolds W, Sunwoo J, et al. Chromatin accessibility associates with protein-RNA correlation in human cancer. *Nat Commun* 2021;12:5732. <https://doi.org/10.1038/s41467-021-25872-1>
- [52] Baskar R, Chen AF, Favaro P, Reynolds W, Mueller F, Borges L, et al. Integrating transcription-factor abundance with chromatin accessibility in human erythroid lineage commitment. *Cell Rep Methods* 2022;2:100188. <https://doi.org/10.1016/j.crmeth.2022.100188>
- [53] Bobkova E, Depes D, Lee J-H, Ježková L, Falková I, Pagacova E, et al. Recruitment of 53BP1 proteins for DNA repair and persistence of repair clusters differ for cell types as detected by single molecule localization microscopy. *Int J Mol Sci* 2018;19. <https://doi.org/10.3390/ijms19123713>
- [54] Hahn H, Neitzel Charlotte, Kopecná Olga, Heermann WD, Dieter, Falk Martin, Hausmann A. Topological analysis of γH2AX and MRE11 clusters detected by localization microscopy during X-ray induced DNA double- strand break repair. *Cancers* 2021;13:5561. <https://doi.org/10.3390/cancers13215561>
- [55] Falk M, Hausmann M. A paradigm revolution or just better resolution – will newly emerging superresolution techniques identify chromatin architecture as a key factor in radiation-induced DNA damage and repair regulation. *13:article Cancers* 2021;18:1–30. <https://doi.org/10.3390/cancers13010018>
- [56] Uversky VN, Giuliani A. Networks of networks: an essay on multi-level biological organization. *Front Genet* 2021;12:706260. <https://doi.org/10.3389/fgene.2021.706260>
- [57] Zimatore G, Tsuchiya M, Hashimoto M, Kasperski A, Giuliani A. Self-organization of whole-gene expression through coordinated chromatin structural transition. *Biophys Rev* 2021;2:031303. <https://doi.org/10.1063/5.0058511>
- [58] Schmidt U, Guigas G, Weiss M. Cluster formation of transmembrane proteins due to hydrophobic mismatching. *Phys Rev Lett* 2008;101:128104. <https://doi.org/10.1103/PhysRevLett.101.128104>
- [59] Guigas G, Weiss M. Membrane protein mobility depends on the length of extra-membrane domains and on the protein concentration. *Soft Matter* 2015;11:33–7. <https://doi.org/10.1039/C4SM01846J>
- [60] Boyd PS, Struve N, Bach M, Eberle JP, Gote M, Schock F, et al. Clustered localization of EGFRvIII in glioblastoma cells as detected by high precision localization microscopy. *Nanoscale* 2016;8:20037–47. <https://doi.org/10.1039/c6nr05880a>
- [61] Pilarczyk G, Nesnidal I, Gunkel M, Bach M, Bestvater F, Hausmann M. Localisation microscopy of breast epithelial ErbB-2 receptors and gap junctions: Trafficking after γ-irradiation, neuregulin-1β, and trastuzumab application. *Int J Mol Sci* 2017;18:362. <https://doi.org/10.3390/ijms18020362>
- [62] Pilarczyk G, Papenfuß F, Bestvater F, Hausmann M. Spatial arrangements of connexin43 in cancer related cells and re-arrangements under treatment conditions: investigations on the nano-scale by super-resolution localization light microscopy. *Cancers* 2019;11:301. <https://doi.org/10.3390/cancers11030301>
- [63] Bartosova M, Herzog R, Ridinger D, Levai E, Jenei H, Zhang C, et al. Alanyl-glutamine restores tight junction organization after disruption by a conventional peritoneal dialysis fluid. *Biomolecules* 2020;10:1178. <https://doi.org/10.3390/biom10081178>
- [64] Krufczik M, Sievers A, Hausmann A, Lee J-H, Hildenbrand G, Schaufler W, et al. Combining low temperature fluorescence DNA-hybridization, immunostaining, and super-resolution localization microscopy for nano-structure analysis of ALU elements and their influence on chromatin structure. *Int J Mol Sci* 2017;18:1005. <https://doi.org/10.3390/ijms18051005>
- [65] Hausmann M, Lee J.-H., Sievers A, Krufczik M, Hildenbrand G. COMBinatorial Oligonucleotide FISH (COMBO-FISH) with Uniquely Binding Repetitive DNA Probes. In: Hancock R editor. *The Nucleus*, vol. 2175, New York, NY: Springer US; 2020, p. 65–77. [https://doi.org/10.1007/978-1-0716-0763-3\\_6](https://doi.org/10.1007/978-1-0716-0763-3_6).
- [66] Grüll F, Kirchgessner M, Kaufmann R, Hausmann M, Kebisch U. Accelerating image analysis for localization microscopy with FPGAs. *Proc. - Int. Conf. Field Program. Logic Appl., FPL* 2011, p. 1–5. <https://doi.org/10.1109/FPL.2011.11>
- [67] Hausmann M, Ilić N, Pilarczyk G, Lee J-H, Logeswaran A, Borroni AP, et al. Challenges for super-resolution localization microscopy and biomolecular fluorescent nano-probing in cancer research. *Int J Mol Sci* 2017;18. <https://doi.org/10.3390/ijms18102066>
- [68] Hausmann M, Falk M, Neitzel C, Hofmann A, Biswas A, Gier T, et al. Elucidation of the clustered nano-architecture of radiation-induced DNA damage sites and surrounding chromatin in cancer cells: a single molecule localization

- microscopy approach. *Int J Mol Sci* 2021;22:3636. <https://doi.org/10.3390/ijms22073636>
- [69] Scipy community. *scipy.spatial.ConvexHull*. Website. Accessed on 2020-04-01. 2017. url: (<https://docs.scipy.org/doc/scipy-0.19.0/reference/generated/scipy.spatial.ConvexHull.html>).
- [70] Scipy community. *scipy.spatial.distance.cdist*. Website. Accessed on 2020-04-01. 2019. url: (<https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.cdist.html>).
- [71] Ripley BD. Modelling spatial patterns. *J R Stat Soc: Ser B (Methodol)* 1977;39:172–92. <https://doi.org/10.1111/j.2517-6161.1977.tb01615.x>
- [72] Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). Simoudis E, Han J, Fayyad UM (eds.), AAAI Press; n.d., p. 226–31.
- [73] Lagache T, Lang G, Sauvionnet N, Olivo-Marin JC. Analysis of the spatial organization of molecules with robust statistics. *PLoS ONE* 2013;8:e80914. <https://doi.org/10.1371/journal.pone.0080914>
- [74] Baumgärtner A, Binder K. Monte Carlo studies on the freely jointed polymer chain with excluded volume interaction. *J Chem Phys* 1979;71:2541–5. <https://doi.org/10.1063/1.438608>
- [75] Valle F, Favre M, De Los Rios P, Rosa A, Dietler G. Scaling exponents and probability distributions of DNA end-to-end distance. *Phys Rev Lett* 2005;95:158105. <https://doi.org/10.1103/PhysRevLett.95.158105>
- [76] Krufczik M. Reaktionen der Genomarchitektur auf ionisierende Strahlung: Quantitative Analyse mittels neuer Konzepte zur hochauflösenden Lokalisationsmikroskopie. Dissertation. 2017. <https://doi.org/10.11588/HEIDOK.00023243>
- [77] Kaufmann R, Müller P, Hildenbrand G, Hausmann M, Cremer C. Analysis of Her2/neu membrane protein clusters in different types of breast cancer cells using localization microscopy: ANALYSIS OF HER2/neu MEMBRANE PROTEIN CLUSTERS. *J Microsc* 2011;242:46–54. <https://doi.org/10.1111/j.1365-2818.2010.03436.x>
- [78] Kaufmann R, Müller P, Hausmann M, Cremer C. Imaging label-free intracellular structures by localisation microscopy. *Micron* 2011;42:348–52. <https://doi.org/10.1016/j.micron.2010.03.006>
- [79] Nakamura AJ, Rao VA, Pommier Y, Bonner WM. The complexity of phosphorylated H2AX foci formation and DNA repair assembly at DNA double-strand breaks. *Cell Cycle* 2010;9:389–97. <https://doi.org/10.4161/cc.9.2.10475>
- [80] Hausmann M, Wagner E, Lee J-H, Schrock G, Schaufler W, Krufczik M, et al. Super-resolution localization microscopy of radiation-induced histone H2AX phosphorylation in relation to H3K9-trimethylation in HeLa cells. *Nanoscale* 2018;10:4320–31. <https://doi.org/10.1039/c7nr08145f>
- [81] Scherthan H, Lee J-H, Maus E, Schumann S, Muhtadi R, Chojowski R, et al. Nanostructure of Clustered DNA Damage in Leukocytes after In-Solution Irradiation with the Alpha Emitter Ra-223. *Cancers (Basel)* 2019;11:1877. <https://doi.org/10.3390/cancers11121877>
- [82] Jezkova L, Zadnepranetic M, Kulikova E, Smirnova E, Bulanova T, Depes D, et al. Particles with similar LET values generate DNA breaks of different complexity and reparability: a high-resolution microscopy analysis of γH2AX/53BP1 foci. *Nanoscale* 2018;10:1162–79. <https://doi.org/10.1039/c7nr06829h>
- [83] Zadnepranetic M, Borejko A, Jezkova L, Falk M, Ryabchenko A, Hramco T, et al. Clustered DNA damage formation in human cells after exposure to low- and intermediate-energy accelerated heavy ions. *Phys Part Nucl Lett* 2022;19:440–50. <https://doi.org/10.1134/S1547477122040227>
- [84] Falk M, Hausmann M, Lukasova E, Biswas A, Hildenbrand G, Davidkova M, et al. Determining omics spatiotemporal dimensions using exciting new nanoscopy techniques to assess complex cell responses to DNA damage: part - structuromics. *Crit Rev Eukaryot Gene Expr* 2014;24:225–47. <https://doi.org/10.1615/CritRevEukaryotGeneExpr.v24.i3.40>
- [85] Dobešová L, Gier T, Kopečná O, Pagáčová E, Vičar T, Bestvater F, et al. Incorporation of low concentrations of gold nanoparticles: complex effects on radiation response and fate of cancer cells. *Pharmaceutics* 2022;14:166. <https://doi.org/10.3390/pharmaceutics14010166>
- [86] Lukášová E, Kořítek Z, Klabusay M, Ondrej V, Grigoryev S, Baćíková A, et al. Granulocyte maturation determines ability to release chromatin NETs and loss of DNA damage response; these properties are absent in immature AML granulocytes. *Biochim Biophys Acta Mol Cell Res* 2013;1833:767–79. <https://doi.org/10.1016/j.bbamcr.2012.12.012>
- [87] Hofer M, Falk M, Komůrková D, Falková I, Baćíková A, Klejdus B, et al. Two new faces of amifostine: protector from DNA damage in normal cells and inhibitor of DNA repair in cancer cells. *J Med Chem* 2016;59:3003–17. <https://doi.org/10.1021/acs.jmedchem.5b01628>
- [88] Štefančíková L, Lacombe S, Salado D, Porcel E, Pagáčová E, Tillement O, et al. Effect of gadolinium-based nanoparticles on nuclear DNA damage and repair in glioblastoma tumor cells. *J Nanobiotechnol*. 2016;14:63. <https://doi.org/10.1186/s12951-016-0215-8>
- [89] Pagáčová E, Štefančíková L, Schmidt-Kaler F, Hildenbrand G, Vičar T, Depeš D, et al. Challenges and contradictions of metal nano-particle applications for radio-sensitivity enhancement in cancer therapy. *Int J Mol Sci* 2019;20:588. <https://doi.org/10.3390/ijms20030588>
- [90] Natale F, Rapp A, Yu W, Maiser A, Harz H, Scholl A, et al. Identification of the elementary structural units of the DNA damage response. *Nat Commun* 2017;8:15760. <https://doi.org/10.1038/ncomms15760>
- [91] Hausmann M, Neitzel C, Bobkova E, Nagel D, Hofmann A, Chramko T, et al. Single molecule localization microscopy analyses of DNA-repair foci and clusters detected along particle damage tracks. *Front Phys* 2020;8:578662. <https://doi.org/10.3389/fphy.2020.578662>
- [92] Eryilmaz M, Schmitt E, Krufczik M, Theda F, Lee J-H, Cremer C, et al. Localization microscopy analyses of MRE11 clusters in 3D-conserved cell nuclei of different cell lines. *Cancers* 2018;10:25. <https://doi.org/10.3390/cancers10010025>
- [93] Lee Y, Wang Q, Shuryak I, Brenner DJ, Turner HC. Development of a high-throughput γ-H2AX assay based on imaging flow cytometry. *Radiat Oncol* 2019;14:150. <https://doi.org/10.1186/s13014-019-1344-7>
- [94] Bach M, Savini C, Krufczik M, Cremer C, Rösl F, Hausmann M. Super-resolution localization microscopy of γ-H2AX and heterochromatin after folate deficiency. *Int J Mol Sci* 2017;18:1726. <https://doi.org/10.3390/ijms18081726>
- [95] Chazal F, Michel B. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Front Artif Intell* 2021;4:667963. <https://doi.org/10.3389/frai.2021.667963>
- [96] Tinarrage R. Barcodes of the Čech filtration of a point cloud in Euclidean plane 2021. [https://www.youtube.com/watch?v=fKEBb190KQo&ab\\_channel=Rapha%C3%89ABiTinarrage](https://www.youtube.com/watch?v=fKEBb190KQo&ab_channel=Rapha%C3%89ABiTinarrage).
- [97] Adams H, Chepushtanova S, Emerson T, Hanson E, Kirby M, Motta F, et al. Persistence images: a stable vector representation of persistent homology. *J Mach Learn Res* 2017;18:1–35. <https://doi.org/10.48550/ARXIV.1507.06217>
- [98] Cohen-Steiner D, Edelsbrunner H, Harer J, Mileyko Y. Lipschitz functions have L<sup>p</sup>-stable persistence. *Found Comput Math* 2010;10:127–39. <https://doi.org/10.1007/s10208-010-9060-6>
- [99] Jaccard P. Etude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaud Sci Nat* 1901;37:547–79.
- [100] Wang B. Lecture 16: Tda, kernels, classification iii 2021. [https://www.youtube.com/watch?v=9V3iWrxTqLs&t=307s&ab\\_channel=UtahSoCComputationalTopology](https://www.youtube.com/watch?v=9V3iWrxTqLs&t=307s&ab_channel=UtahSoCComputationalTopology).
- [101] Bubenik P. Statistical topological data analysis using persistence landscapes. *J Mach Learn Res* 2015;16(1):77–102. <https://doi.org/10.48550/ARXIV.1207.6437>
- [102] Bubenik P. The persistence landscape and some of its properties. *Topological Data Analysis Abel Symposia Vol 15*. Springer; 2020. p. 97–117. <https://doi.org/10.48550/ARXIV.1810.04963>
- [103] Cohen-Steiner D, Edelsbrunner H, Harer J. Stability of persistence diagrams. *Discret Comput Geom* 2007;37:103–20. <https://doi.org/10.1007/s00454-006-1276-5>
- [104] Vicar T, Gumulec J, Kolar R, Kopečna O, Pagacova E, Falkova I, et al. DeepFoci: deep learning-based algorithm for fast automatic analysis of DNA double-strand break ionizing radiation-induced foci. *Comput Struct Biotechnol J* 2021;19:6465–80. <https://doi.org/10.1016/j.csbj.2021.11.019>
- [105] Aschenbrenner K.P., Butzek S., Guthier C.V., Krufczik M., Hausmann M., Bestvater F., et al. Compressed sensing denoising for segmentation of localization microscopy data. 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Chiang Mai, Thailand: IEEE; 2016. p. 1–5. <https://doi.org/10.1109/CIBCB.2016.7758097>.
- [106] Ovesný M, Křížek P, Borkovec J, Sivndrych Z, Hagen GM. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* 2014;30:2389–90. <https://doi.org/10.1093/bioinformatics/btu202>
- [107] Depes D, Lee J-H, Bobkova E, Jezkova L, Falkova I, Bestvater F, et al. Single-molecule localization microscopy as a promising tool for γH2AX/53BP1 foci exploration. *Eur Phys J D* 2018;72(9):158. <https://doi.org/10.1140/epjd/e2018-90148-1>
- [108] Schubert E, Sander J, Ester M, Kriegel HP, Xu X. DBSCAN revisited: why and how you should (still) use DBSCAN. *ACM Trans Database Syst* 2017;42:1–21. <https://doi.org/10.1145/3068335>
- [109] Falk M, Lukášová E, Kozubek S. Chromatin structure influences the sensitivity of DNA to gamma-radiation. *Biochim Biophys Acta* 2008;1783:2398–414. <https://doi.org/10.1016/j.bbapcr.2008.07.010>
- [110] Falk M, Lukášová E, Gabrielová B, Ondrej V, Kozubek S. Local changes of higher-order chromatin structure during DSB-repair. *J Phys Conf Ser* 2008;101:012018. <https://doi.org/10.1088/1742-6596/101/1/012018>
- [111] Müller P, Lemmermann NA, Kaufmann R, Gunzel M, Paech D, Hildenbrand G, et al. Spatial distribution and structural arrangement of a murine cytomegalovirus glycoprotein detected by SPDM localization microscopy. *Histochem Cell Biol* 2014;142:61–7. <https://doi.org/10.1007/s00418-014-1185-2>
- [112] Tobin SJ, Wakefield DL, Jones V, Liu X, Schmolze D, Jovanović-Talisman T. Single molecule localization microscopy coupled with touch preparation for the quantification of trastuzumab-bound HER2. *Sci Rep* 2018;8:15154. <https://doi.org/10.1038/s41598-018-33225-0>
- [113] Deschout H, Zanacchi FC, Mlodzianoski M, Diaspro A, Bewersdorf J, Hess ST, et al. Precisely and accurately localizing single emitters in fluorescence microscopy. *Nat Methods* 2014;11:253–66. <https://doi.org/10.1038/nmeth.2843>