Patterns

CellPress
OPEN ACCESS

Review

A Review of Super-Resolution Single-Molecule Localization Microscopy Cluster Analysis and Quantification Methods

Ismail M. Khater,[1,*] Ivan Robert Nabi,[2,3,*] and Ghassan Hamarneh[1,3,*]
[1]Medical Image Analysis Lab, School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
[2]Department of Cellular and Physiological Sciences, Life Sciences Institute, University of British Columbia, Vancouver, BC V6T 1Z3, Canada
[3]These authors contributed equally
*Correspondence: ikhater@sfu.ca (I.M.K.), irnabi@mail.ubc.ca (I.R.N.), hamarneh@sfu.ca (G.H.)
https://doi.org/10.1016/j.patter.2020.100038

THE BIGGER PICTURE Recent developments in super-resolution SMLM imaging techniques enable researchers to study macromolecular structures at the nanometer scale. However, SMLM data quantification and interpretation methods have yet to keep pace with the rapid advancement of SMLM imaging. This article provides a balanced and comprehensive review of state-of-the-art SMLM image analysis methods and ties disparate approaches together in a cohesive manner. Researchers are actively exploring new computational methods to analyze SMLM data, including recent approaches to use data-driven and machine-learning approaches. However, the validation of the SMLM clustering methods remains an open challenge. Potential future directions using multi-modality imaging (e.g., SMLM and electron microscopy) might help validate quantitative SMLM image analysis methods.

Single-molecule localization microscopy (SMLM) is a relatively new imaging modality, winning the 2014 Nobel Prize in Chemistry, and considered as one of the key super-resolution techniques. SMLM resolution goes beyond the diffraction limit of light microscopy and achieves resolution on the order of 10–20 nm. SMLM thus enables imaging single molecules and study of the low-level molecular interactions at the subcellular level. In contrast to standard microscopy imaging that produces 2D pixel or 3D voxel grid data, SMLM generates big data of 2D or 3D point clouds with millions of localizations and associated uncertainties. This unprecedented breakthrough in imaging helps researchers employ SMLM in many fields within biology and medicine, such as studying cancerous cells and cell-mediated immunity and accelerating drug discovery. However, SMLM data quantification and interpretation methods have yet to keep pace with the rapid advancement of SMLM imaging. Researchers have been actively exploring new computational methods for SMLM data analysis to extract biosignatures of various biological structures and functions. In this survey, we describe the state-of-the-art clustering methods adopted to analyze and quantify SMLM data and examine the capabilities and shortcomings of the surveyed methods. We classify the methods according to (1) the biological application (i.e., the imaged molecules/structures), (2) the data acquisition (such as imaging modality, dimension, resolution, and number of localizations), and (3) the analysis details (2D versus 3D, field of view versus region of interest, use of machine-learning and multi-scale analysis, biosignature extraction, etc.). We observe that the majority of methods that are based on second-order statistics are sensitive to noise and imaging artifacts, have not been applied to 3D data, do not leverage machine-learning formulations, and are not scalable for big-data analysis. Finally, we summarize state-of-the-art methodology, discuss some key open challenges, and identify future opportunities for better modeling and design of an integrated computational pipeline to address the key challenges.

Introduction

Cells are the structural and functional units of living organisms. Studying the cell requires an understanding of its different compartments and their relationship to one other inside and outside the cell. With the aid of microscopes, researchers can visualize, identify, and study cell organelles and molecular components, which is critical to understanding cell function in health and malfunction in different diseases. The recent advent of super-resolution microscopy, which provides an order-of-magnitude improvement in resolution compared with light microscopy,

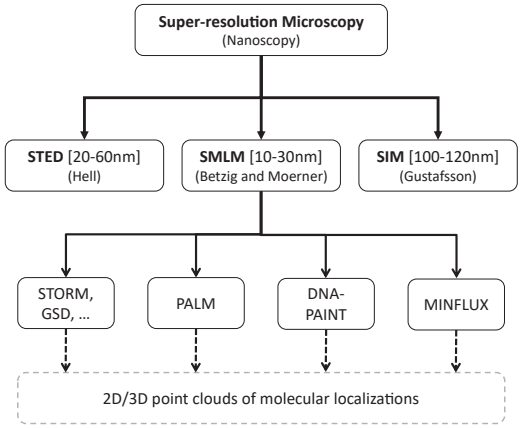


Figure 1. Classification of Super-Resolution Nanoscopy Methods
All SMLM methods generate localizations as 2D or 3D point clouds.

allows visualization and quantification of the organization of proteins to form macromolecular complexes *in situ* in intact cells, facilitating our understanding of molecular interactions in different biological structures that drive cell behavior. This revolutionary discovery advances science to better understand cellular function and the machinery of its subcellular compartments.

Resolution Limit

Optical microscopy (also known as a light microscopy) uses visible light and a series of lenses to image and magnify cell and tissue samples. Light is usually used in non-invasive imaging of cells *in vitro* and *in vivo* for various imaging applications, which helps researchers magnify and visualize biological structures within the optical resolution of the imaging system. The resolution of the imaging system is affected by physical phenomena such as lens misalignment and the diffraction of light, known as the diffraction limit of the imaging system (i.e., the microscope). The diffraction limit of the light microscope, which was theoretically described by Abbe in 1873,[1] is proportional to the wavelength of the light (λ) being observed and inversely proportional to the numerical aperture (NA) of the objective lens. The formulation of Abbe's diffraction limit is given in Equation 1:[2,3]

$$d = \frac{\lambda}{2\mathrm{NA}_{\mathrm{obj}}}. \qquad \text{(Equation 1)}$$

For example, if we use visible light with a wavelength of 500 nm (green light) and an objective lens with $\mathrm{NA}_{\mathrm{obj}}$ of 1.0, the minimum theoretical distance d will be 250 nm that defines two separable objects.[2] However, practically, the resolution can be less than that due to various experimental reasons. Abbe's diffraction limit of light microscopy (i.e., 250 nm) is a major barrier that has prevented researchers from studying multiple biological structures and macromolecular complexes below the diffraction barrier. Breaking this barrier has been achieved finally with development of super-resolution microscopy (next section) and improvements in fluorescent probes and labeling techniques for super-resolution microscopy.[4]

Super-Resolution Nanoscopy Methods

The invaluable contribution of super-resolution microscopy was acknowledged by the 2014 Nobel Prize in Chemistry awarded jointly to three scholars for their contribution to bringing light microscopy into the nanometer scale (i.e., nanoscopy). Eric Betzig, William E. Moerner, and Stefan W. Hell shared the prize.[3,5,6] Betzig and Moerner developed the principles of single-molecule localization microscopy (SMLM)[7,8] and Hell worked on stimulated emission depletion (STED) microscopy.[9] Another microscopy method, structured illumination microscopy (SIM),[10] developed by Mats Gustafsson, lately gained rapid popularity.[6] Another super-resolution method, super-resolution optical fluctuation imaging (SOFI),[11] has been developed to overcome the diffraction limit of light. SOFI is faster than SMLM but has lower resolution. Schidorsky et al.[12] combined SOFI and SMLM to improve the overall imaging performance. They showed that by rejecting common background sources, SOFI-assisted SMLM can be used to improve image reconstruction. Figure 1 depicts the various super-resolution methods developed to break the diffraction limit barrier of light microscopy.

SMLM methods include photoactivated localization microscopy (PALM),[7,13] stochastic optical reconstruction microscopy (STORM),[14] direct STORM (dSTORM),[15] ground state depletion (GSD),[16] DNA-based point accumulation for imaging in nanoscale topography (DNA-PAINT),[17] and MINFLUX.[18] SMLM achieved the highest resolution among the super-resolution methods (Figure 1). The lateral resolution of SMLM could be from 10 to 30 nm (MINFLUX achieves 2 nm resolution). The STED lateral resolution reaches 60–100 nm, while it is about 100–120 nm for SIM. On the other hand, the analysis complexity of SMLM is ranked as the most complex according to Owen and Gaus,[19] followed by SIM as intermediate in the analysis complexity and, finally, STED as the simplest. Wegel et al.[20] experimentally studied the super-resolution methods, including SMLM, and applied them to image various subcellular structures. They showed the weakness and strength of each method on the studied structures (e.g., vesicles and filaments).

Super-resolution microscopy has allowed for unprecedented high-resolution visualization of various biological structures such as microtubules, actin, clathrin-coated pits, mitochondria, chromatin complexes, neurons, ER, and focal adhesion complexes.[21] However, the initial demand for high-resolution images of biological structures has been replaced by a need for quantitative methods and analysis.[22] SMLM imaging methods produce spatial coordinates of molecular localizations, called "point clouds" in this survey, that are ideally suited for the application of cluster analysis algorithms and tools. We focus this survey on state-of-the-art super-resolution SMLM cluster analysis methods and their capabilities and shortcomings. Note that we do not intend to survey all possible data-clustering methods but rather limit the presented works to those methods that have been applied to SMLM data clustering.

From Imaging to Quantification

Figure 2 gives an overview of the imaging-to-quantification pipeline for SMLM, which starts with fluorescent labeling of the target molecule, then determining molecular localization from the acquired SMLM images, and ends with post-processing and quantification of the imaged and localized proteins. Our focus in this

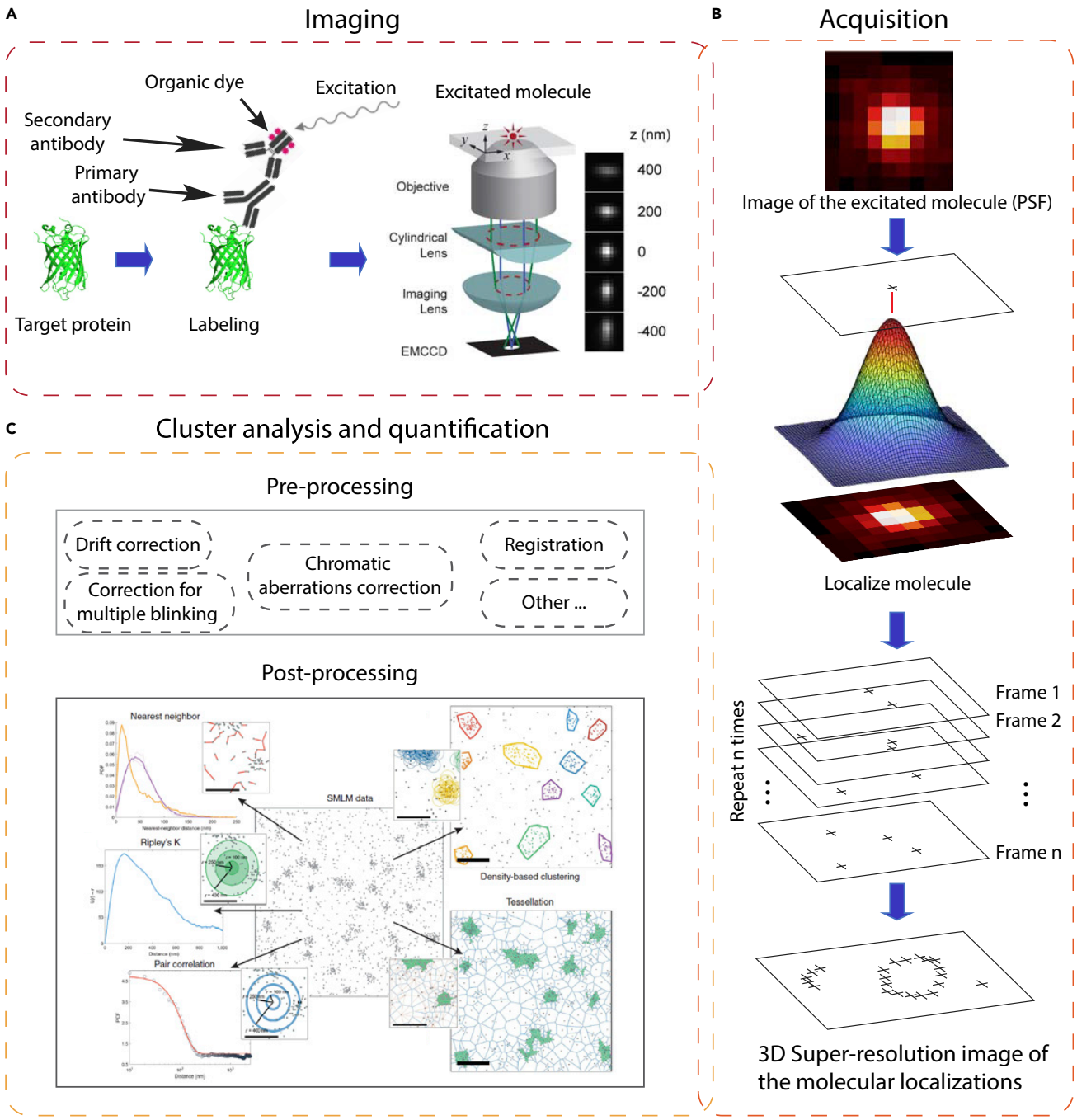


Figure 2. Overview of SMLM Quantification Pipeline

An overview of the whole framework from imaging to quantification of super-resolution STORM SMLM data: (A) 3D SMLM imaging of the target protein,[23] (B) acquiring the protein localizations and getting a map for the molecular coordinates, and (C) analyzing the super-resolved image to quantify the SMLM clusters. (C) is produced with permission from Nicovich et al.[24]

survey will be on cluster analysis and quantification of SMLM data.

Acquisition and Localization

The first step in imaging or tracking a protein is to label the target protein with a fluorescent dye. The labeling process varies depending on the SMLM imaging technique. For example, for STORM imaging the target molecule is labeled via antibodies conjugated to organic dyes (e.g., Alexa 647 molecules) (Figure 2). In PALM, genetically modified fluorescent proteins (FPs) are used in the labeling of the target proteins (e.g., mEos2). Also, the labeling strategy may depend on the binding proteins/antibodies used in the labeling. We show the primary-secondary labeling strategy in Figure 2A as an example. Other labeling strategies, including using fragment antigen-binding (Fab) antibodies

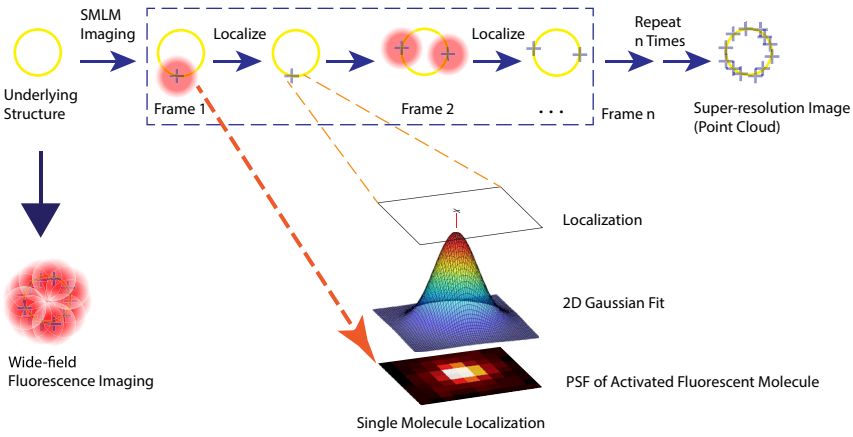


Figure 3. Illustration of the SMLM Imaging Principle

Labeling the yellow circle (i.e., biological structure below the diffraction limit) efficiently with fluorescent dye to be imaged with a fluorescence microscope. The conventional diffraction-limited wide-field microscope produces a blurred image. The SMLM imaging produces a super-resolved image that is constructed from a set of time-separated images, wherein each time frame image a sparse set of excited labeled proteins can be localized using Gaussian PSF to form the final point-cloud super-resolution image for the structure.

(that can be obtained for both primary and secondary labeling) and nanobody labeling, might be used to reduce the size of the fluorescent probe and improve the resolution.[25–28]

Although the exact implementation may vary, all SMLM methods fundamentally rely on temporal separation of the emissions of the excited fluorophores,[4,29] where the fluorophores are sparsely activated and forced to switch between **on** "bright state" and **off** "dark state" stochastically during the imaging session. Stochastic "blinking" of non-overlapping point spread functions (PSFs), formed due to diffraction of light, are recorded by the imaging system. Positional localization of individual fluorophores is approximated to be at the center of a Gaussian fitting of the PSF, as shown in Figure 3, resulting in significantly improved (~10×) resolution. By repeating this process thousands of times and compiling the fluorophore localizations from all the acquired frames, we obtain a high-resolution image. This is in contrast to diffraction-limited fluorescence microscopy in which, due to the single-shot approach, the PSFs of molecules at distances below the diffraction limit overlap (Figure 3), resulting in reduced resolution of the image.

dSTORM is based on the use of standard fluorophores that are commercially available conjugated to a wide range of antibodies, and are therefore applicable to common immunofluorescent labeling of multiple cellular constituents. For dSTORM, fluorophores are induced to enter a weakly emissive or dark state by high-powered laser illumination from which fluorophores will spontaneously return to the ground state and emit fluorescence.[2] Choice of fluorophore is based on photon output, as higher photon output improves localization accuracy as well as the relative time the fluorophore spends in the dark and bright states (duty cycle) and how many times the fluorophore can cycle between the dark and bright states (switching cycle). Fluorophore blinking is enhanced using buffers containing thiol reducing agents and oxygen scavengers. The dSTORM dye of choice is Alexa 647, which exhibits high levels of blinking and photon yield that are critical for analysis approaches described later in this review.

In addition to localizing the photon events of the excited fluorophores in the plane, i.e., x and y coordinates, introducing a cylindrical lens in the light path of the imaging system will deform the PSF according to the depth (i.e., z) of the molecule within the imaged sample. Arriving at the depth value of a single molecule involves fitting a multi-variate Gaussian PSF to the deformed shape of PSF.[23] Other three-dimensional (3D) localization methods include biplane,[30] PSF engineering as in (1) double-helix PSF,[31] (2) phase ramp,[32] and (3) Zernike optimized localization approach in 3D (ZOLA-3D),[33] and dual opposed objective interferometry as in (1) iPALM[34] and (2) 4Pi detection scheme,[35] and supercritical-angle fluorescence recovery.[36]

The imaging process is repeated thousands of times using a Gaussian PSF fit to determine the localization of the individual molecules at high resolution. Every molecular location might have other information such as the localization uncertainty (fitting error of the Gaussian PSF), frame number, and number of photons. The final super-resolved image is formed by recombining all the localizations from all of the imaged frames. A number of methods have been designed for this purpose, such as ThunderSTORM,[37] QuickPALM,[38] RapidSTORM,[39] and RainSTORM[40] (see EPFL SMLM Software Directory, http://bigwww.epfl.ch/smlm/software/index.html). As shown in Figure 2B, the localization of the molecules is obtained from every frame image individually enabling the user to map locations in two-dimensional (2D) or 3D coordinate space. The set of molecular localizations and their associated metadata are known as point clouds, events-list, pointillist, and so forth. The point clouds representation is used as input to the cluster analysis and quantification as shown in Figure 2C. We guide the reader's attention to many of the excellent references and reviews on super-resolution microscopy, especially the SMLM imaging techniques.[2,7,13,14,22,26] In Figure 2A, we show 3D STORM imaging as an example of imaging a single molecule.[23]

Imaging Artifacts

Quantification of super-resolution SMLM data might be biased due to some imaging artifacts. Some of the artifacts are challenging[41] and should be accounted for before analyzing the data. There are common pitfalls in super-resolution microscopy specimen preparation and imaging acquisition that should be avoided and optimized to ensure data reproducibility.[42] However, there are computational methods to address some of the artifacts and mitigate their effects to produce artifact-free super-resolution images. To enlighten readers about the super-resolution imaging artifacts and challenges that facing the cluster analysis and quantification, we guide them to further references.[41–47] We list here the main artifacts as they appear in recent papers.[41–43]

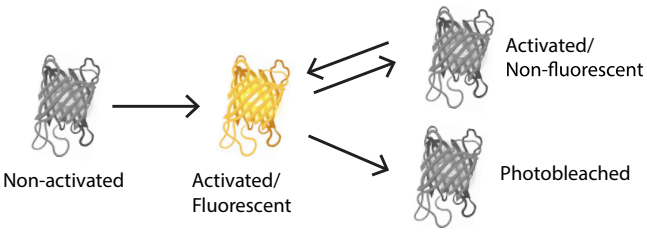


Figure 4. The Four-State Photokinetics Model for Photoswitchable Fluorescent Proteins
The image used in this illustration is adapted from Frick et al.[50]

Labeling Errors. Labeling of the protein of interest in SMLM is done primarily by expression of a photoactivatable FP directly linked to the protein of interest or via an antibody-conjugated fluorescent organic dye by means of immunolabeling. The photoactivatable FP, used in PALM, has a large size and may alter the localization and function of the protein of interest. For the immunolabeling approach adopted for STORM, the dye is conjugated to an antibody specific for the protein of interest. This method might include unspecific labeling, and antibody specificity for the protein of interest should be validated.[41] In both labeling methods, the location of the imaged dye/FP can differ from the true location of the protein of interest and in a random direction. These localization errors create limitations to the quantification methods such that the protein clusters appear enlarged.

Detection Efficiency. Several methods have been proposed to quantify the percentage of proteins that are properly active. Not all the photoconvertible FP and molecules/fluorophores used in protein labeling are mature or successfully photoconvert. Hence, no algorithm can count the proteins that never appear.[43]

Localization Uncertainty. Several methods[48] have been used to determine the position of the emitting molecule. The localization algorithms estimate the localization of the formed PSF of the fluorescent molecule. For example, in the Gaussian PSF model, the localization uncertainty is inversely proportional to the square root of the number of collected photons from the molecule.

Blinking. The blinking artifact, also known as multiple blinking of a single fluorophore, is considered a serious artifact and has been studied extensively.[45,49–53] Multiple blinking affects molecular counting and creates pseudoclusters. For example, in PALM imaging, according to the four-state photokinetic model for photoswitchable fluorescent protein (Figure 4), once the fluorescent probe is activated, it can switch between non-fluorescent and fluorescent state before photobleaching irreversibly occurs,[50] resulting in overcounting.

Drift. Super-resolution SMLM images consist of thousands of stacked frames collected over time. Changes in temperature, the vibration of the microscope base, or air current, among others, might cause sample drift in both lateral and axial directions.[41] Hence, the consequent drift can introduce localization (or spatial translation) errors by dozens of nanometers for different molecules relative to each other during the data acquisition.

Chromatic Aberrations. This occurs in multi-color imaging whereby light undergoes wavelength-dependent distortions. Motion artifacts (e.g., mechanical movements) and imperfections in the optical imaging system are the main sources of the chromatic aberration artifacts that affect co-localization of fluorophores of different colors.[54] The chromatic aberration correction is thus required on top of the drift.

Cluster Analysis and Quantification

Using point-cloud representation for SMLM data analysis is not trivial. Point-cloud representation is fundamentally different from the intensity grid valued pixel or voxel image representation used in conventional microscopy. Consequently, the computational tasks such as image processing, segmentation, and registration applied for SMLM data analysis are different[55] from the ones applied for conventional microscopy data analysis. Researchers have been working for decades to develop computational methods designed for conventional microscopy image analysis. However, these methods are not necessarily applicable to the point-cloud data. Analysis of SMLM point-cloud data is more complex[19,55] and therefore requires new approaches. In particular, cluster analysis methods are most appropriate for analysis of super-resolution SMLM data point clouds generated by the localization methods that produce the data.

The SMLM data analysis literature uses pre-processing and post-processing terms interchangeably. We believe that a clear distinction between the two terms should be made. Hence, in Figure 2C, we have included the methods used to correct for imaging artifacts in a pre-processing sub-box and the analysis methods used to quantify the biological clusters in a post-processing sub-box. For better quantification and interpretation of biological clusters, as opposed to artifactual pseudoclusters, we believe that pre-processing should be applied first to obtain artifact-free data. In this survey, we focus on the post-processing methods used for SMLM cluster analysis as discussed in the next section. Specifically, this work focuses on reviewing the post-processing methods used to cluster and analyze point-cloud SMLM data. However, some other methods, i.e., image-based cluster analysis, have been utilized to analyze the SMLM data and do not utilize the intrinsic pointillist properties of SMLM data. Examples of image-based cluster analysis methods that include extracting statistical measures are provided by several studies.[56–60]

SMLM Cluster Analysis Methods

Protein-to-protein interactions produce heterogeneous and dynamic multi-molecular protein complexes.[61] The complex arrangements might consist of multiple molecules that vary in sizes; ranging from few to tens of nanometers. Studying protein cluster structure and organization is important to determine their function in the cell. Figure 5 shows how the protein molecules could cluster together in many ways to form more complex structures. In this section, we focus on the cluster analysis methods used to specifically understand the molecular clusters in super-resolution SMLM data.

Not all the clusters in SMLM data are related to biological structures. Some of the clusters in SMLM data are due to imaging artifacts (i.e., pseudoclusters) caused by the uncertainty of the photophysical properties of the fluorescent reporters[27] as mentioned in Imaging Artifacts, as well as labeling of the target molecule by more than one antibody probe. Pseudoclusters could bias the quantification and the interpretation of detected

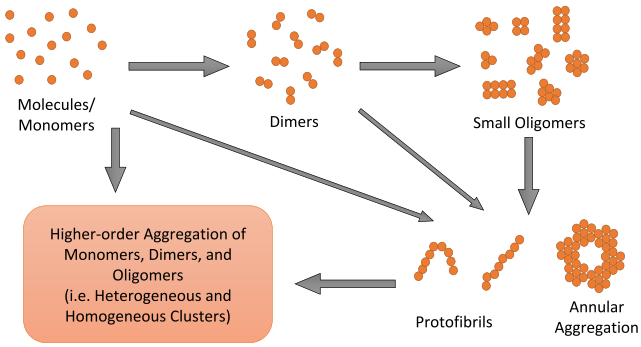


Figure 5. Illustration of How the Protein Molecules Cluster Together to Form Complexes

Monomers aggregate to form dimers which aggregate to form the small oligomers. Monomers could also cluster directly to form the large mutants and oligomers.

molecular clusters. Hence, pre-processing analysis is required to correct for multi-blinking artifacts.

SMLM Molecular Localizations Clustering Task

SMLM data are a point cloud in 2D or 3D coordinate space where fluorophore events, or localizations, of the labeled target protein are output as an eventlist or pointillist $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_L\}$, where L is the total number of localizations. Each $\mathbf{p}_i$records the reconstructed spatial location of the event as a coordinate vector (x_i, y_i)in 2D or (x_i, y_i, z_i)in 3D. Additionally, metadata associated with each fluorophore event are recorded, which can vary from SMLM imaging technique to another (e.g., STORM, PALM, GSD) or from one microscope to another. Examples of fluorophore event $\mathbf{p}_i$ metadata are $photon - count_i$, $frame - ID_i$, $error - in - x_i$, and $error - in - y_i$. These metadata may be leveraged during the pre-processing of the eventlist. For example, $photon - count$ could be used as a confidence measure for an event where events with low photon counts are discarded. Following attempts to correct for imaging artifacts (e.g., diminishing the multiple blinking that causes pseudoclusters and the offsets due to the length of antibody chain), one can obtain a reconstructed point cloud that provides estimates of molecular localizations. Each reconstructed molecular localization $\mathbf{m}_i$ is represented as a 2D or 3D spatial coordinate. Mathematically, we write $\mathbf{M} = \mathcal{R}(\mathbf{P})$, where $\mathbf{M}$ is a set of reconstructed molecular localizations, $\mathcal{R}$ represents the correction process, and $\mathbf{m}_i \in \mathbf{M}$, $i = 1, 2, \cdots, N$, where N is the total number of reconstructed molecular localizations (i.e., $N = |\mathbf{M}|, L = |\mathbf{P}|$). $\mathbf{m}_i$ has coordinate vector $(x_i, y_i) \in \mathbb{R}^2$ for a 2D point or $(x_i, y_i, z_i) \in \mathbb{R}^3$ in 3D. Notice that typically $N \leq L$.

The molecular cluster analysis task is to apply various mathematical operators to find relations, patterns, curves (e.g., Ripley's H function that shows a cluster of molecular localizations) over the entire set $\mathbf{M}$ to detect, segment, and classify the molecular clusters that represent the different protein complexes in the imaged SMLM data. Formally, a cluster of molecular localizations is a dense group of molecules that are, loosely speaking, closer in some way to each other than to the localizations outside the cluster. Clustering analysis is the task of decomposing a given point cloud of molecular localizations $\mathbf{M}$ into smaller disjoint (non-overlapping) subsets, such that their union covers the whole set $\mathbf{M}$. Each subset (i.e., cluster/blob b_i) implies some underlying macromolecular biological structure. Sometimes the process of isolating a given set of point clouds of molecular localizations $\mathbf{M}$ into smaller disjoint subsets is called segmentation.

Formally, let $\mathcal{C}$ represent the clustering operation applied to the set of localizations $\mathbf{M}$. $\mathcal{C}$ takes every localization in $\mathbf{M}$ and bins into one of K bins or clusters denoted $b_j, j = 1, 2, \cdots, K$. If the set of K clusters or bins is denoted $\mathbf{B}$, and when $\mathbf{m}_i$ can only belong to one cluster (i.e., disjoint, non-intersecting bins), we write the mapping $\mathcal{C} : \mathbf{M} \rightarrow \mathbf{B}$ as $\mathcal{C}(\mathbf{m}_i) = b_j \in \mathbf{B}$, $\cup_j b_j = \mathbf{M}$, $b_j \cap b_k = \varphi, \forall j, k$, where $\cup$ and $\cap$ are union and intersection, respectively.

Sometimes, the cluster analysis is applied directly to the raw SMLM data (i.e., $\mathbf{P}$).

In this section, we dig deeper into the cluster analysis methods used to quantify the molecular clusters (i.e., post-processing) obtained through super-resolution SMLM. The methods include statistical, Bayesian, density-based, Voronoi tessellation-based, and graph-based approaches.

Statistical Methods

Over the past few years, researchers have started to apply statistical methods that are based on second-order statistics and spatial point analysis methods to quantify the SMLM clusters. The statistical methods have been applied to ecological spatial data and adopted for SMLM analysis. In this section, we cover the main statistical methods used in the literature to analyze super-resolution SMLM data as listed in Table 2. Given its large popularity in analyzing SMLM data, we describe Ripley's functions next as well as some of its variants. We then describe a class of statistical methods that are based on correlation techniques.

Ripley's Functions. Ripley's K, H, and L functions are gaining popularity in cluster analysis of the SMLM membrane proteins. These functions are used increasingly due to the point-cloud nature of SMLM data (localization of molecules). Ripley[62] studied the stochastic models that have been proposed for spatial point patterns. Ripley's K function is a tool used for analyzing spatial point process data.[62,63] It is usually used for analyzing 2D data, but may be used to analyze locations along a line (one dimension) or may be extended to 3D spatial data.[63,64]

The density of points in an area (number of points per unit area) is known as the first moment property.[65] The second moment property (also known as a bivariate or multi-variate generalization) is used to describe the relationships between two or more point patterns by finding the expected number of points N within a distance r of another point.[63,65] Ripley's K function[62] is a second moment property (second-order statistics). Theoretically, the K function is given in Equation 2:[63]

$$K(r) = \lambda^{-1} E[\text{number of points within distance } r \text{ of randomly chosen point}], \quad \text{(Equation 2)}$$

where λ is the density normalization of points (number of points per area A, $\lambda = N/A$). Formally, the K function is given in Equation 3:[62,65]

$$K(r) = \frac{1}{n} \sum_{i=1}^{n} N_{p_i}(r) / \lambda, \quad \text{(Equation 3)}$$

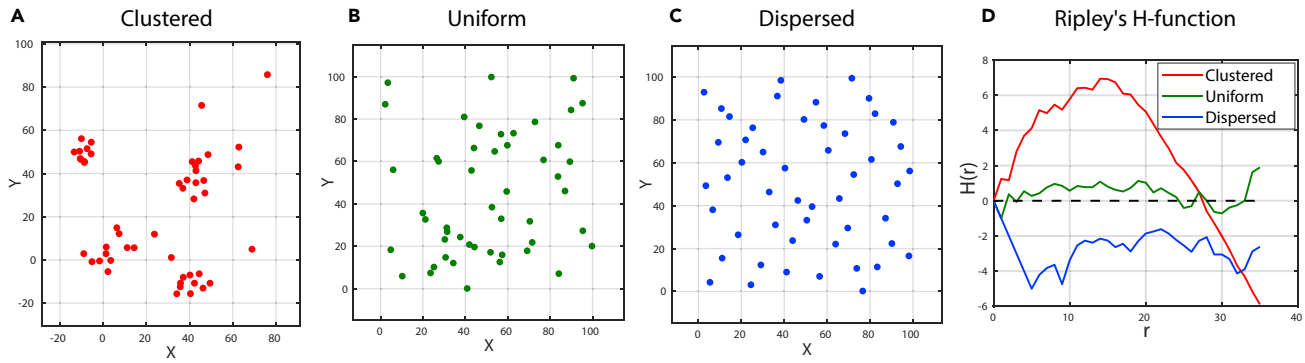


Figure 6. An Example Showing the Behavior of the $H(r)$ Function for the Different Distributions of Spatial Point Patterns
The $H(r)$ function (D) has positive values for clustered points (A), fluctuates around 0 for uniformly distributed (random) points (B), and has negative values for dispersed points (C). The generated data consist of 50 points for each one of the patterns shown in (A) to (C).

where p_i is the ith point in the data and the sum is taken over n points. For a homogeneous Poisson process, which is known as complete spatial randomness (CSR), the expected value of function $K(r)$ is given in Equation 4:[63]

$$\mathrm{K}(r) = \pi r^2. \qquad \text{(Equation 4)}$$

Note that deviation from the CSR expected value indicates scales of clustering and dispersion. So Ripley's K function is typically used to find the level of clustering by comparing a given distribution with a random distribution.

Other Ripley's functions can be derived from the K function. The complete derivation for all the other functions can be found in other papers.[63,65] The L function was proposed by Besag[66] as a normalization for the K function, as seen in Equation 5:

$$L(r) = \sqrt{K(r)/\pi}. \qquad \text{(Equation 5)}$$

The L function and its derivative can be used to identify the radius of the clusters.[65] Normalizing the L function will produce another function called the H function.[67] Hence, the H function is a further normalization of the original K function. The H function is given in Equation 6:

$$H(r) = L(r) - r. \qquad \text{(Equation 6)}$$

Note that the H function may result in a positive value, which indicates clustering over the spatial scale; on the other hand, the negative value indicates dispersion. The value is zero when we have CSR distributed points (not clustered or dispersed points). This is because for a CSR distribution, $L(r) = r$ for all values of r. Figure 6 shows three cases of spatial point patterns and the corresponding Ripley's H-function behavior. The pattern of the H function fluctuates around zero for uniformly distributed points, above zero for clustered points, and below zero for dispersed points.

To estimate $K(r)$, the numerator of Equation 3 can be written as $N^{-1}\sum_i\sum_{j\neq i} I(d_{ij} < r)$,[63,68] where d_{ij} is the distance between the ith and jth points. $I(.)$ is an indicator function that is equal to 1 if $d_{ij} \leq r$ and is zero otherwise.

In its current formulation, the K function does not consider the effect of the points close to the border of the study area. This issue, which is called the edge effect of Ripley's K function, causes underestimation of K.[63,68–70] Hence, Ripley's K function requires more elaborate methods for edge correction. Many methods have been proposed to correct the edge effect of Ripley's K function.[69–72] Generally, the corrected K function ($\widehat{K}(r)$) can be written as Equation 7:

$$\widehat{K}(r) = \frac{A}{N}\sum_i\sum_{j\neq i}\frac{I(d_{ij}<r)}{w_{ij}}, \qquad \text{(Equation 7)}$$

where w_{ij} is a weight function that provides the edge correction.

Ripley's functions are becoming increasingly popular in analyzing SMLM data. The functions have been utilized in many biological applications to find the level of molecular clustering. They are used either alone or in combination with the other cluster analysis methods.[53,73–97] We summarize how the methods are adopted for SMLM cluster analysis of the different biological applications in Table 2.

Getis and Franklin's Local Point Pattern. Getis and Franklin published a paper and proposed a new $L(r)$ function. Their function is a variant of Ripley's K function, called second-order neighborhood analysis. The goal of their $L(r)$ function is to quantify the clustering of the points (molecules in SMLM context) at various spatial scales.[98] The values of $L(r)$ function are calculated for each point as described in Equation 8:

$$L(r)_j = \sqrt{A\sum_{i=1}^{n}\frac{\delta_{ij}}{n-1}\Big/\pi}, \qquad \text{(Equation 8)}$$

where A is the region area (e.g., rectangular region) under study and n is the total number of points in the region. The indicator function delta $\delta_{ij} = 1$ if the distance between point i and point j is $<r$ and zero otherwise. $\sum_{i=1}^{n}\delta_{ij}$ is the summation over all points within distance r from point j (i.e., all points within a circle of radius r centered at localization j). Thus, $L(r)$ is another way of normalizing Ripley's K function by finding the local point patterns normalized by the average point density in the whole analyzed

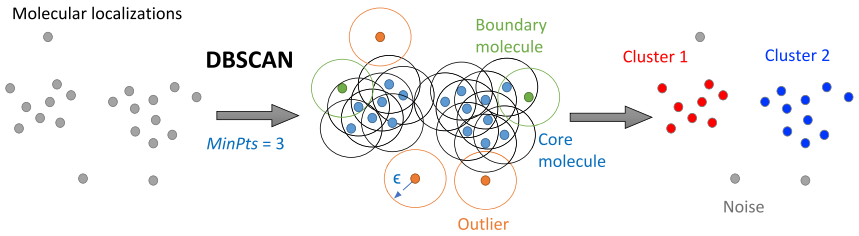


Figure 7. An Example Illustrating the Density-Based DBSCAN Clustering Method Applied to SMLM Data

For instance, DBSCAN algorithm is applied when using ε and *MinPts* = 3 parameters. Sometimes the subjectivity of selecting the parameters might change the clustering results dramatically. For example, pseudoclusters in SMLM complicate the selection of the algorithm parameters.

region. The property of $L(r) = r$ for all values of r when having a CSR distribution still holds.

The Getis and Franklin (G&F) function could be used to find the local descriptor (patterns) of every localization, while Ripley's function is used to describe all the points in the region globally. Therefore, the G&F function has been used in combination with Ripley's K function to analyze the localization of molecules from SMLM data.[75,80,83,85,87–93,99] For example, a G&F point pattern was used in double protein-labeling analysis[99] to investigate the co-clustering of membrane proteins. It was also used for generating a topographic map of the level of clustering to determine the heights of peaks in the map across a region[90] and then using the relative heights of the peaks to determine the clustering characteristics and avoid inaccurate thresholding.

Correlation-Based Methods. Correlation-based analysis methods (pair correlation, autocorrelation, cross-correlation, co-localization) have been applied to super-resolution SMLM data for both pre-processing and post-processing quantification (Figure 2C). Pre-processing methods[53,100–102] address imaging artifacts such as the multiple blinking of a single fluorophore that may cause molecular overcounting. Overcounting, as well as the other implications of imaging artifacts, might bias SMLM cluster analysis and should be corrected before post-processing the SMLM data for quantifying the biological clusters. For example, Malkusch et al.[101] used a correlation coefficient framework, coordinate-based co-localization, to analyze every single localization within a certain radial distance and assign to it a score ranging from –1 to 1. In their formulation, –1 is assigned to perfectly segregated localizations, 0 for uncorrelated (randomly distributed), and +1 for perfect co-localization. The same coordinate-based co-localization idea has also been used to analyze the protein localizations of biological clusters.[82,88,96] Many other correlation functions have been used to quantify the localization clustering for post-processing SMLM data.[82,88,96,99,102–104] The proposed correlation functions are used to analyze biological clusters rather than the biologically irrelevant pseudoclusters (also known as nanoclusters). For example, Schnitzbauer et al.[102] derived a cross-correlation function for the localization coordinates inspired by the translational cross-correlation function in pixel-based image representation. They mathematically showed that the point-to-point distance distribution in super-resolution SMLM is equivalent to the pixel-based correlation function. They then extended cross-correlation to quantify the spatial relationship between complicated structures by considering the point-to-set distance.

Bayesian Methods

The SMLM cluster analysis methods usually depend on a set of user-defined parameters. Sometimes, the subjectivity and the ambiguity of selecting the parameters affect the performance of the SMLM clustering task. The main goal of the Bayesian approach for super-resolution SMLM data is to design a clustering method that alleviates the need for arbitrary user-selected analysis parameters. A well-defined prior Bayesian will replace the arbitrary user-selected parameters. Bayesian is a model-based approach which is used for spatial point clustering generated by SMLM. The model is used to evaluate the assignment of every molecule to clusters by its marginal posterior probability. The posterior probability is computed based on a specified model for the molecular data and their uncertainties.[78] Therefore, the mechanism is to select clusters from a set of generated clustering proposals. Usually, clustering proposals are generated with variable spatial scale and threshold using statistical methods such as Ripley's K function[83] or the G&F function.[78] After generating thousands of candidate proposals per region of interest (ROI), the optimum number of proposals is then selected by scoring them against the Bayesian model.[55,78,83,105] For example, in their generative model, Griffié et al.[83], considered an ROI containing clustered and non-clustered localizations. The user sets the probability that localization is non-clustered, and this is the prior parameter for the model. They also assume that the molecular positions in the cluster are following spherical Gaussian distribution. The radius (Gaussian standard deviation) of the cluster is drawn from a user-specified histogram of sizes. They claim that the aforementioned model reflects the a priori knowledge of the molecular distribution. The Bayesian approach is not limited to quantifying 2D, and it was extended by Griffié et al. to analyze 3D SMLM data.[83]

Density-Based Methods

Density-based clustering methods are popular in data mining and spatial data clustering. Ester et al. proposed density-based spatial clustering of applications with noise (DBSCAN),[106] a density-based clustering algorithm that is capable of discovering clusters of arbitrary shapes. This can be used to filter out noisy events from the SMLM data when its parameters are set correctly.

DBSCAN is based on two parameters for detecting and segmenting the clusters in SMLM data. It requires a neighborhood radius ε and the minimum number of localizations/points (*MinPts*) within ε to qualify as a cluster (Figure 7). The algorithm can start from any molecular localization that has not been visited. The connectivity of the molecules of the qualified clusters should maintain the *MinPts* condition within ε while propagating from one molecule to another within the same cluster until reaching the boundary molecules, where the *MinPts* condition no longer holds. Otherwise, the cluster is considered an outlier. It

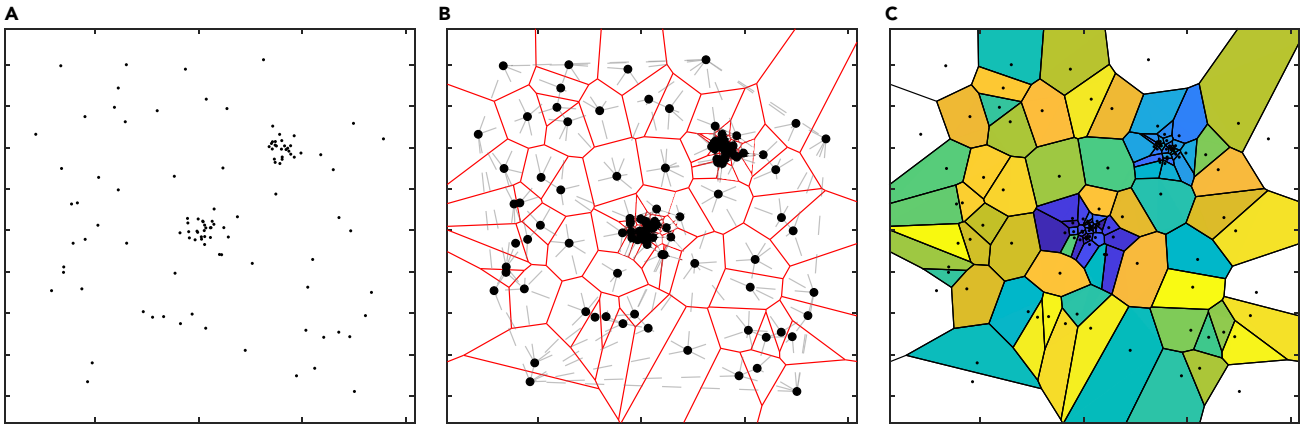


Figure 8. Voronoi Tessellation-Based Method Used to Segment the Clustered SMLM Molecular Localizations
(A) The input space of molecular localization. It has two clusters and noisy/background localizations.
(B) Voronoi tessellation and partitioning the space into polygonal regions (Voronoi cells) in red. The Delaunay triangulation (dual of Voronoi) is shown by gray dashed connections.
(C) The Voronoi cells colored with different colors. The white regions are the Voronoi cells with open regions.

is clear that the clustering is conditioned on the minimum density of molecules within neighborhood radius ε. Figure 7 shows an example of how the DBSCAN clustering method works and the required parameters to cluster the localizations.

According to Mazouchi and Milstein,[107] leveraging DBSCAN to analyze super-resolution SMLM data has certain limitations. The algorithm is slow scaling with the number of localizations and it has $O(n\log(n))$ at best. The ambiguity and subjectivity in selecting the algorithm parameters affect its performance and makes the algorithm general. The imaging artifacts and the multiple blinking of a single fluorophore cause the formation of pseudoclusters. DBSCAN may not be sufficient to differentiate between protein clusters (i.e., biologically relevant clusters) and the non-biologically relevant pseudoclusters.[107] To address the limitations of DBSCAN, Mazouchi and Milstein[107] propose a density-based clustering algorithm, fast optimized cluster algorithm for localizations (FOCAL). FOCAL is a grid-based method optimized for fast analysis of SMLM data. It has one parameter that needs to be optimized, density threshold (*minL*). It has a linear time complexity ($O(n)$). FOCAL has limitations in dealing with small clusters and requires the setting of fine grids. The problem becomes more severe with high levels of noise. Moreover, FOCAL has issues with high-density SMLM data and the overlapped clusters. FOCAL3D[108] is an extension for the FOCAL method that is capable of analyzing 3D SMLM data.

In SMLM data cluster analysis, DBSCAN is used either alone[107,109–111] or in combination with other clustering algorithms[82,95,96,102,112,113] to quantify the SMLM clusters.

Voronoi Tessellation-Based Methods

The Voronoi diagram, or tessellation of point clouds, has been used in many applications for various goals, including computational geometry, computational physics, astrophysics, computational chemistry, and biology. In SMLM, a Voronoi diagram is a method used to partition the input space of molecular localizations into regions according to the Euclidean distance between the seed points (i.e., molecules). The resultant polygonal regions are called Voronoi cells, where each cell is centered around one of the molecules. Figure 8A shows a set of points in 2D space, where each point might represent a molecular localization. Figure 8B depicts the Voronoi diagram for the molecular localizations shown in Figure 8A. The Voronoi cells are shown in different colors in Figure 8C. Notice that the Voronoi edges are equidistant from the two nearest molecules. Specifically, the projected perpendicular line from every molecule to any one of its Voronoi cell edges is the shortest distance between every neighboring pair of molecules. Hence, there is no intersection between any Voronoi cells. To learn more about the methods used to find the Voronoi polygons, we draw the reader's attention to the work of Okabe et al.[114]

Segmenting SMLM molecular clusters using geometric properties (e.g., area, shape) of the Voronoi cell was done by Levet et al. and Andronov et al.[115,116] The Voronoi geometric characteristics could be used to describe the neighborhood of the molecules. The Voronoi geometric characteristics for every molecule might be different based on the density and data organization. Both of these studies[115,116] depended on the Voronoi cell area to segment the SMLM molecular clusters, where the Voronoi cell area is inversely proportional to the density of the molecules (regions with high molecular densities are composed of Voronoi cells with smaller areas). Both works designed a method to cluster SMLM localizations by comparing their Voronoi cell areas with a reference distribution chosen to be either a spatially uniform[115] or CSR distribution.[116] Figure 8 explains the concept of the Voronoi diagram method when applied to point-cloud data. Note that the clustered points (Figure 8A) have Voronoi cell areas (Figure 8C) that are smaller than the non-clustered points.

Voronoi tessellation for clustering molecular localization has been applied to structures with various shapes, such as tubular-shaped structures (e.g., microtubules, filamentous, fibrous).[115,116] Voronoi tessellation was also adopted in other super-resolution SMLM cluster analysis applications.[84,86,97]

Delaunay triangulation is the companion of the Voronoi tessellation (Figure 8B). It has also been used for analyzing and quantifying super-resolution SMLM data.[86,117] However, Voronoi tessellation methods are different from the Delaunay

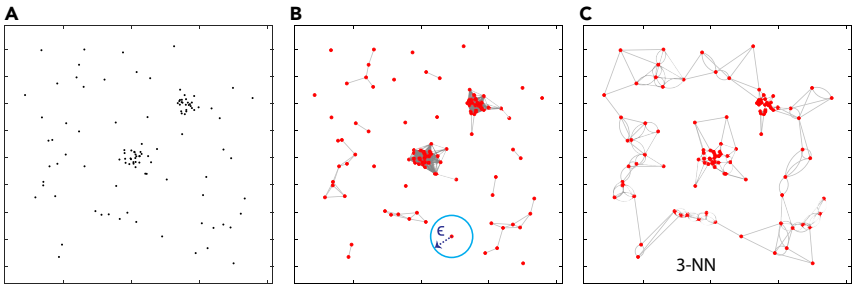


Figure 9. Network/Graph-Based Method Used to Model the SMLM Molecular Localizations for Cluster Analysis

(A) The input space of molecular localization. It has two clusters and noisy/background localizations.

(B) The ε-graph used to construct the network, where every node is connected to all the other nodes within the proximity distance ε.

(C) The kNN graph used to construct the network, where every node is connected to only the k closest neighboring nodes. We constructed the 3-NN graph for illustration.

triangulation with respect to the former's ability to provide a direct estimation of the region of influence,[118] and hence is preferable for analyzing SMLM data. Recently, a 3D extension of the Voronoi tessellation method has been proposed by Andronov et al.,[118] who used Voronoi volumes as a characteristic of 3D Voronoi cells to segment 3D SMLM clusters. Andronov et al.[118] claim that the Voronoi-based methods are able to handle the edge effect (see Ripley's Functions). This is because the border molecules have larger or infinite Voronoi cell areas that prevent them from contributing to the clustering.

Clustering molecular localizations is therefore based on the geometrical properties of the Voronoi cells such as the cell area. Some molecules are considered as part of a cluster based on their individual Voronoi cell areas. This leads to crude segmentation of the clusters. Moreover, some of the border molecules might be excluded from being part of a cluster as they might have very large Voronoi cell areas (e.g., white cells in Figure 8C) compared with inner molecules. In addition, the Voronoi tessellation-based methods might fail in extracting the true molecular clusters from SMLM data with multiple blinking of single fluorophore artifact. Leveraging Voronoi cell area for segmenting clusters from SMLM data with varying cluster densities might be another problem in such methods because the cell areas hugely depend on the underlying molecular densities and the closeness of the nearby clusters.

Graph-Based Methods

Graphs are strong mathematical structures employed to model the interaction between objects or entities of a system. The entities are represented as graph nodes and their interactions are represented as edges.[119] Hence, the graphs are considered powerful and rich data structures that encode the connectivity relationships between the different entities of a system. In real-world problems, graphs are frequently complex networks because they have many properties that make them different from other types of graphs such as random graphs.[120] For example, real-world networks have many subgraphs, modules, patterns, and small-worldness that are not frequent in other types of graphs. Networks are ubiquitous and they are used to study and model many real-world problems effectively. Recently, network analysis methods have been successfully adopted in many fields of study such as the brain, social, computer, road, metabolic, and Internet.[121–125]

Leveraging graph theory to analyze SMLM data is infrequent in the surveyed literature. Few recent works have been proposed to utilize complex networks and graphs for cluster analysis of super-resolution SMLM data. Various neighborhood networks/graphs can be constructed from the spatial SMLM data. Figure 9 shows two types of neighborhood graphs that can be adopted for analyzing the SMLM data.

Networks have been adopted to analyze large datasets of prostate cancer data as well as cardiac data from various SMLM modalities.[126,127] A graph-based network method has been proposed by Khater et al. to model SMLM data,[126] where the nodes represent the molecules and the molecular interactions are represented as edges connecting the nodes. Various network features have been leveraged to pre-process the SMLM data (e.g., correcting for multiple blinking of a single fluorophore artifact) as well as post-processing (e.g., denoising the SMLM data and extracting its constituent molecular clusters). A combination of network analysis and the mean-shift algorithm are leveraged to segment the clusters and obtain 3D representation of molecular localization of diffraction-limited cellular structures, in this case plasma membrane invagination called caveolae. Khater et al.[128] also proposed a graph-based method to extract graphlet features from SMLM molecular clusters for automatic identification and quantification of various biological structures. Network community detection and modularity analysis have been proposed to decipher the architecture of the molecular clusters.[129] Communities represent subclusters of molecules within a larger cluster. An example of modular detection within caveolae is shown in Figure 10.

Some other very recent graph-based methods to analyze and quantify SMLM data are posted as preprints and are still unpublished. We cover them briefly in this survey and categorize them under the graph-based methods. Researchers are exploring new computational methods to analyze the SMLM data. For example, community detection has also been exploited for extracting SMLM clusters.[130] A segmentation protocol based on persistence homology and DBSCAN has been employed to segment and quantify the topological structure within SMLM data.[131] In this persistence homology method, the density modes were constructed from a graph that connects all the localizations within the same search radius.

Machine-Learning-Based Methods

Machine-learning algorithms (including deep learning) are data-driven approaches. Deep-learning approaches typically require relatively large data that could capture the variations in the dataset. Supervised machine-learning approaches require ground-truth data for training the different learning models, which is difficult to obtain in SMLM data. Machine-learning models can be trained to perform various computational tasks such as predicting, segmenting, and classifying the molecular complexes.

We have not witnessed a large amount of work leveraging machine learning for super-resolution SMLM data for the

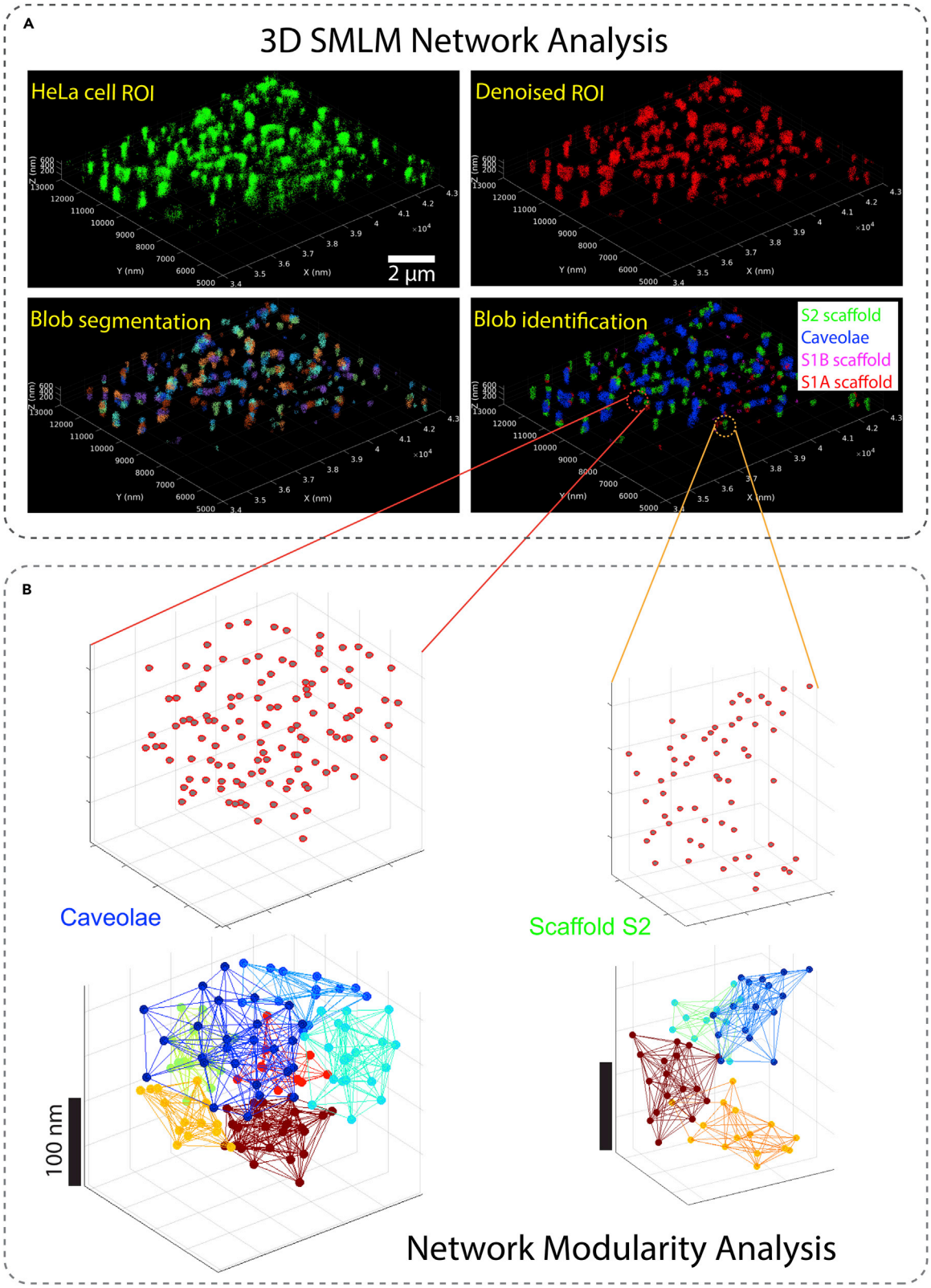


Figure 10. Graph-Based Network Analysis Methods for SMLM Data Proposed by Khater et al.

(A) Khater et al.[126,129] proposed the 3D SMLM Network Analysis pipeline[126] to correct for multiple blinking of a single fluorophore, filter out noisy localizations, segment the biological structures into clusters/blobs, and identify the cluster/blob classes.

(B) Network community/modularity analysis[129] detecting the modules within caveola and S2 scaffold domains.

aforementioned computational tasks. However, some works have applied deep learning for the localization and data acquisition[132–134] but not for cluster analysis.

Khater et al.[126,128] designed a graph-based machine-learning method to automatically identify the class of the molecular complexes from super-resolution SMLM data. They leveraged machine learning for many computational tasks such as determining the scale of clustering, finding the biosignatures for several biological structures, and identifying patterns of the isolated and multiple antibody proteins. They also leveraged deep learning for the biological structures classification task[135] applied to several SMLM data representations. Sieben et al.[110] used machine learning to identify the class of the biological structures from SMLM data. Another recent work that utilized machine learning to detect clustered and unclustered (background) molecules was proposed by Tobin et al.[103] Williamson et al.[136] proposed a supervised machine-learning method that is capable of classifying all the localizations from microscopy datasets into clustered or non-clustered classes. Their model is trained on several simulated clustered datasets.

Validation

Validation of the cluster analysis method is critical when applying the various algorithms to find the biological clusters from SMLM data. There is no publicly available dataset with ground-truth class labels for the membership of the localizations to the various cluster types. Hence, most of the methods are unsupervised approaches whereby the ground truth is not provided along with the data. The ground truth might include information such as the number of clusters and their features (e.g., sizes and densities). Also, comparing the different clustering methods requires having benchmark SMLM data with known cluster features. We summarize the main methods used to validate the super-resolution SMLM clustering methods in the following subsections.

Computer Simulations (In Silico). *Synthetic Data*. Generating synthetic data with known cluster features (e.g., density, size/volume, shape) has been widely used to mimic SMLM data. Background and noise signals with known distributions are also generated along with the synthetic clusters. Some methods generate synthetic data that is based on specific assumptions, such as generating Gaussian clusters, and the minimum distance between the generated clusters should be greater than some threshold value.[65,118]

Given synthetic data, a clustering method is tested on extracting the synthetic clusters first. It could then be used for cluster analysis in experimental SMLM data. To assess the quality of a clustering method, the extracted clusters and their features are compared with the known clusters used in the data generation.

Simulated Data. Simulation could be used to mimic super-resolution SMLM imaging for known biological structures. Simulators have the ability to imitate the SMLM imaging by varying several parameters (e.g., labeling strategy, labeling efficiency, epitope length, number of frames, imaging time, density, background) that might be useful for optimizing the imaging of the experimental sample. Hence, simulation gives more control to study all the possible scenarios that might lead to less imaging artifacts in the data. Moreover, simulation could help in assessing the quality of the adopted clustering analysis methods.

Recently, many SMLM simulators have been developed and posted as publicly available software tools. The simulators facilitate the generation of data for use in cluster analysis applied to various biological structures. Popular SMLM simulators software includes SuReSim,[137] TestSTORM,[138] and SMeagol.[139] Synthetic data generation offers controlled creation of clusters and background with known density distributions (e.g., Gaussian, uniform) and cluster shapes (e.g., circular, tubular). For example, Levet et al.[140] generated synthetic two-color 2D and 3D clusters of circular and square shapes. They also simulated multiple scenarios by varying the number of clusters, their relative positions, their diameters, their density ratios, and background/noise levels.

Data simulators, on the other hand, are designed to mimic realistic labeling and imaging conditions.[46,47] SMLM data simulation considers the inner workings of the SMLM imaging technique and labeling parameters (e.g., epitope length, labeling efficiency, localization precision, number of frames, blinking events per frame) in the data generation, but does not give direct control of the resulting data. For example, Spahn et al.[141] used the SuReSim simulator[137] with specific SMLM imaging parameters (e.g., cluster diameter of 100 nm and various numbers of epitopes per cluster) to generate an image with a field of view of 15 × 15 μm^2, of some biological structures (e.g., clathrin-coated pits). Sieben et al.[110] also used simulation to validate their work. They mimicked their real experimental conditions to generate ground-truth models by controlling the labeling efficiency, localization positions, noise molecules, and fluorophore parameters (e.g., distributions for photon count, localization precision). To evaluate their ERGO emitter density estimation method, Cardoen et al.[134] used the *in silico* sequence of 2,500 frames (each 64 × 64 grayscale pixels corresponding to a 2D view of 100 × 100 nm^2) from Sage et al.,[48] which simulates a realistic acquisition of microtubules labeled with the commonly used Alexa 647 fluorophore. They verified their approach on real-world data[142] with a markedly different microscope configuration whereby they showed that aligning the intensity distribution between training and real-world data is sufficient to obtain consistent results without retraining.

Validation via Physical Phantoms. *DNA Origami and Nanorulers*. DNA origami and nanorulers have been developed to validate many of the SMLM imaging and analysis methods. They are used in super-resolution imaging and microscope calibration.[47] The DNA origami is designed to allow placing of a known number of fluorescent molecules to nanostructures in defined geometries.[143] In addition, DNA origami has been used to quantify the protein copy number in the cells using super-resolution microscopy.[144]

Validation via Knowledge of Biology and Other Imaging Modalities. Real experimental super-resolution SMLM data can be used in clustering methods validation if the studied clusters of biological complexes have been studied before with other imaging modalities. Biological structures, imaged using electron microscopy (EM), with known size and number of molecules, can be used as ground truths for super-resolution cluster analysis methods. Generally, researchers use simulation or synthetic data to validate their methods and then apply their methods to real experimental data. For example, Sieben et al.[110,145] used

Table 1. Summary of the Main Categories of the Super-Resolution Cluster Analysis and Quantification Methods

Method	Pros (+)	Cons (–)
Statistical	simple and easy to implement; could be used to detect the level of clustering; could be used for both pre- and post-processing	restricted to analyze homogeneous clusters; edge effect; some normalization based on Poisson point process assumption
Bayesian	can handle SMLM localizations and their associated uncertainties; parameter-free model	very slow; sensitive to the prior settings; used in combination with other methods; requires generation of thousands of cluster proposals; very sensitive to imaging artifacts
Density-based	efficient in noise removal; could be used to discover clusters with various shapes	clustering is conditioned on the minimum density of molecules within neighborhood radius; slow scaling with the number of localizations; ambiguity and subjectivity in selecting the algorithm parameters affect its performance; cannot deal with varying cluster densities and hollow clusters
Voronoi-based	fast and scalable to handle big data; efficient in noise removal; sensitive to clusters of specific geometry (e.g., tubular-like structures).	might cause problems in segmenting data with non-isotropic distribution; limited multi-scale capabilities; might not be good for segmenting hollow clusters
Graph-based	fast and scalable to handle big data; easy to be integrated with machine/deep learning; robust to noise; capable to extract per-point and per-cluster features; topological graph is invariant to the dimensionality of the data; extracting heterogeneous clusters capability; can be used for both pre- and post-processing; all the other methods can be derived from the graph-based method	graph construction is not straightforward for big data; might cause problems in segmenting data with non-isotropic distribution; clustering results are highly dependent on the graph construction method

The cons (+) and pros (–) for each of the clustering analysis methods are also shown for comparison.

EM imaging to validate their multi-color 3D SMLM reconstruction and analysis method. They used dual-color SMLM to image around 300 centrioles per field of view. They then used masking to segment the localizations and DBSCAN to separate adjacent particles. The 3D volumes were reconstructed by EM routines and classified by applying 2D clustering. Khater et al.[126,129] used known information about the cell surface invagination, caveolae, to validate their work. For example, they compared their findings with known topology, size, and number of predicted proteins per segmented structure.

Summary and Discussion

In this paper, we surveyed the state-of-the-art cluster analysis and quantification works applied to super-resolution SMLM. We depended on various criteria to study the papers and tabulate them in Table 2 according to: (1) the biological application of the study; (2) the data acquisition; and (3) the data analysis technique adopted. We then categorized the different clustering methods for easy reference and comparison and identified the pros and cons of these categories in Table 1. Looking at the various methods/algorithms listed in Table 2, we note the following:

- 2D or 3D analysis. Some algorithms have been used only for 2D super-resolution SMLM data analysis, while some other algorithms were used for 2D and then extended to 3D. Dealing with 3D SMLM data is challenging because the axial resolution is usually poorer than the lateral resolution. Also, some biological structures depict structural properties evident in 3D (e.g., hollow structures), so the analysis methods should be designed with care to handle such 3D structures in the denoising, clustering, and identification stages.
- Pre-processing. Few methods could effectively handle some of the imaging artifacts, such as the multiple blinking of a single fluorophore artifact (e.g., graph-based, statistical methods), while some other methods could not (e.g., Voronoi tessellation-based, density-based methods).
- Localization uncertainties. Few methods utilized the localization uncertainties (e.g., Bayesian methods) in the analysis, while the majority of the methods did not.
- Parameterization. The majority of the methods have parameters, while the Bayesian methods are claimed to be parameter-free models. However, Bayesian methods are relatively much slower (e.g., Griffié et al.[105] reported that the processing time for one dataset consisting of 30 small 2D ROIs is ~19 h with user input). Voronoi tessellation-based clustering is parameter-free method if the segmentation threshold is determined by Monte Carlo simulations.[116]
- Intra-cluster analysis. The intra-cluster features (features of molecular interaction within a cluster and its subclusters such as network analysis of the molecules, modularity analysis, and subnetworks) lead to understanding the architecture of the biological complexes. Very few methods are equipped with capabilities to extract the intra-cluster features (e.g., graph-based), while the majority of the methods do not have this capability.
- Machine-learning integration. Most of the methods are not equipped to be integrated with machine-learning approaches for further analysis. Machine-learning approaches require associating features with samples/

Table 2. Super-Resolution SMLM Cluster Analysis and Quantification Methods (Post Processing)

Study			Acquisition						Analysis								
Ref	Year	App	ImgMeth	Dim	Res (*nm*)	#FPI	#Loc	DataSz	CAM	A/V ($\mu m^2/\mu m^3$)	WAV	SRIC	CP	ICA	ML & CC	MSA	SW
Owen et al.[73]	2010	Lck and Src in T cells	PALM dSTORM	2D	~20	15K	1,500/μm^2	NR	Ripley	2 × 2	✓		2–5			✓	
Lillemeier et al.[74]	2010	TCR and Lat in T cells	hsPALM	2D	~25	1K	140–150/μm^2	5–10 cells per exp. (3 exp.)	Ripley	NR	✓		2–5			✓	
Williamson et al.[93]	2011	Lat in T cells	PALM dSTORM	2D	NR	150–200K	NR	3–25 exp.	Ripley, G&F	3 × 3			3–4			✓	
Pereira et al.[75]	2012	HIV-1 amtrix in HIV-1 virus	dSTORM	2D	15–20	20K	NR	5–6 cells per exp. (2 exp.)	Ripley, G&F	10 × 10	✓		2–5			✓	
Owen et al.[92]	2012	LAT in T; HeLa cells	PALM	2D	NR	15K	NR	7 cells per cond. (4 cond.)	Ripley, G&F	3 × 3			2			✓	
Pageon et al.[76]	2013	NKG2D in NK T cells	PALM GSD	2D	20–30	20–25K	1,140–1,920/μm^2	16–23 cells (2–4 exp.)	Ripley	2 × 2 3 × 3	✓		2–5			✓	
Rossy et al.[77]	2013	Lck and CD45 in T cells	PALM dSTORM	2D	21	15–20K	NR	10–13 cells (3 exp.)	Ripley	3 × 3 4 × 4	✓		2–5			✓	
Owen et al.[80]	2013	DiI, Lat vesicles in T cells HIV-1 Gag	HILO-PALM	3D	NR	NR	NR	NR	Ripley, G&F	2 × 2	✓	✓	3–5				
Malkusch et al.[81]	2013	polyprotein in T cells	dSTORM	2D[a]	~20	4–10K	NR	NR	Ripley, NN	2 × 2	✓		2–3			✓	
Rossy et al.[99]	2014	Lck and CD45 in T cells	PALM dSTORM	2D	20–30	NR	NR	NR	G&F	2 × 2[b]	✓		2			✓	
Wee et al.[89]	2015	CD37,β2-integrin in HL-60 cells	dSTORM	2D	NR	8K	NR	20 cells	Ripley, G&F	3 × 3	✓						
Stone and Veatch[104]	2015	Lyn kinase, BCR in CH27 cells	STORM	2D	NR	NR	NR	5 cells	steady-state cross-correlation	NR	✓					✓	MATLAB func.
Gao et al.[85]	2016	STAT1, STAT3 in HeLa cells; CENP-A in U2OS cells	dSTORM	2D	29	5K	NR	20 cells per cond. (5 exp.)	Ripley, G&F	4 × 4	✓		2–3			✓	
Oszmiana et al.[91]	2016	KIR, KIR2DL1, KIR2DS1 in NKL cells	GSD	2D	NR	20K	NR	14–35 cells per exp. (24 exp.)	Ripley, G&F	3 × 3	✓		8				
Krüger et al.[149]	2017	TLR4 in HEK 293 cells	PALM	2D	50	NR	NR	9–10 cells per cond.	NN histogram	NR	✓						
Lopes et al.[88]	2017	FcγRI, FcγRII, SIRPα in T, B cells	dSTORM	2D	25	5K	NR	10–30 cells	Ripley, G&F, CBC	5 × 5	✓		2–3			✓	

(*Continued on next page*)

Table 2. ***Continued***

Study			Acquisition						Analysis								
Ref	Year	App	ImgMeth	Dim	Res (*nm*)	#FPI	#Loc	DataSz	CAM	A/V (μm^2/μm^3)	WAV	SRIC	CP	ICA	ML & CC	MSA	SW
Bálint et al.[87]	2018	NKG2D in T cells	dSTORM	2D	NR	5K	NR	12–500 cells	Ripley, G&F	5 × 5	✓		2–3			✓	
Peters et al.[94]	2018	actin cytoskeleton in T cells	iPALM dSTORM	2D, 3D	NR	50–100K	NR	NR	angular Ripley	3 × 3			2–3			✓	MATLAB func.
Rubin-Delanchy et al.[78]	2015	CD3 in T cells	PALM dSTORM	2D	10–30	20K	2 × 2[b]	30 ROIs (per cond.)	Bayesian; Ripley	3 × 3	✓		4			✓	R func.
Griffié et al.[90]	2015	LFA-1 in T cells	dSTORM	2D	NR	NR	NR	10 cells	Ripley, G&F, Bayesian	2 × 2[b]			4				
Griffié et al.[105]	2016	ZAP-70 in T cells	PALM dSTORM	2D	20–30	NR	15–20K	12 cells	Bayesian	2 × 2 3 × 3			3			✓	MATLAB, R
Griffié et al.[83]	2017	LAT vesicles in T cells	iPALM	3D	10–30	30K	2 × 2[b]	5 cells per cond.	Ripley, Bayesian	2 × 2[b]			5			✓	
Griffié et al.[55]	2018	CD4 in T cells	Live-cell PALM	2D	NR	NR	4K	6 cells	Ripley, Bayesian	2 × 2[b]			3				
Pengo et al.[109]	2014	GaG data HIV, Nef	PALM	2D	NR	NR	NR	NR	DBSCAN	NR	✓	✓					PALMsiever
Caetano et al.[79]	2015	PACS-1, LAMP1, etc. in HeLa cells	GSD	2D	20	5–30K	NR	5 cells per exp. (3 exp.)	density–based; Ripley	1.5 × 1.5 3.5 × 3.5 4 × 4	✓		4			✓	MIiSR
Mazouchi and Milstein[107]	2015	RNAP II in cortex cells; H-NS in *E. coli*	PALM dSTORM	2D	~10	NR	NR	NR	density-based; DBSCAN	2 × 2[b]	✓		2			✓	FOCAL
Pageon et al.[82]	2016	TCR, VD45 in T cells	dSTORM	2D	20–30	20K	NR	NR	Ripley, DBSCAN	4 × 4	✓		13			✓	Clus-DoC
Malkusch and Heilemann[96]	2016	HIV, gag, env in T cells	SMLM	2D	NR	NR	NR	1 cell	DBSCAN, Ripley, OPTICS	NR			5			✓	LAMA
Barna et al.[111]	2016	mitochondrial protein Tom20 in neuroblast brain cells	STORM	3D	40	20K	NR	NR	DBSCAN	NR	✓		3				VividSTORM
Mollazade et al.[112]	2017	RGD peptides	dSTORM	2D	16	20K	NR	NR	DBSCAN, NND	2 × 2[b]	✓		3			✓	

(*Continued on next page*)

Table 2. *Continued*

Study			Acquisition						Analysis								
Ref	Year	App	ImgMeth	Dim	Res (*nm*)	#FPI	#Loc	DataSz	CAM	A/V ($\mu m^2/\mu m^3$)	WAV	SRIC	CP	ICA	ML & CC	MSA	SW
Zhang et al.[113]	2017	*Salmonella typhimurium* mutants in bacterial cells	FPALM	2D, 3D	~35	6–18K	NR	58–600 cells	DBSCAN wavelet	NR	✓		3			✓	
Lukeš et al.[150]	2017	CD4 glycoprotein mutants in T cells	SOFI; simulated PALM	2D	NR	5K	NR	20 cells per cond.	SOFI density analysis	3 × 3	✓		3				code[c]
Schnitzbauer et al.[102]	2018	*cis*-Golgi, GRASP65, GM130, *trans*-Golgi, TGN46 in RPE cells	STORM	2D	NR	NR	NR	NR	DBSCAN, histogram correlation	NR	✓		2–3				Python func.
Lagache et al.[95]	2018	synapsin, VGLUT in primary hippocampal neurons	SIM STORM	3D	NR	30K	0.5M	NR	DBSCAN, Ripley, SODA	20 × 20 × 2	✓		3–5				Icy Plugin
Sieben et al.[110]	2018	proteins within centrioles and procentrioles in KE37 cells	STORM	3D	32–65	30–60K	NR	NR	DBSCAN	2 × 2	✓	✓	8		✓		SPARTAN
Tobin et al.[103]	2018	trastuzumab (HER2 receptor) in breast cancer cell lines (BT-474, SK-BR-3, MDA-MB-468)	dSTORM	2D	NR	20–40K	230–360 locs/μm^2	17–23 cells	density pair correlation	20 × 20	✓		6		✓		Cluster Occupancy
Nino et al.[108]	2019	nuclear pore complex in U-2 human OS cells	dSTORM	3D	NR	NR	2 × 2[b]	NR	density-based; DBSCAN	2 × 2[b]	✓		2			✓	FOCAL3D
Paul et al.[151]	2019	DSB foci in human (U2Os) cells	dSTORM	2D	NR	NR	NR	NR	KDE DBSCAN, Voronoi	NR			3				SMoLR
Levet et al.[115]	2015	microtubules in COS7 cells; GluA1, tubulin, integrin-β3 in neuronal cells	PALM dSTORM	2D	33.9	NR	0.024–0.277M	3 cells per cond.	Voronoi	2 × 2[b]	✓		4				SR-Tesseler
Andronov et al.[116]	2016	microtubules, chromatin in HeLa cells	GSD	2D	NR	NR	0.230M	NR	Voronoi	2 × 2[b]	✓		5				ClusterVisu

(*Continued on next page*)

Table 2. *Continued*

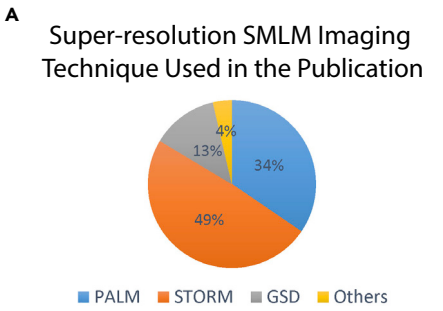
Study			Acquisition						Analysis								
Ref	Year	App	ImgMeth	Dim	Res (*nm*)	#FPI	#Loc	DataSz	CAM	A/V ($\mu m^2/\mu m^3$)	WAV	SRIC	CP	ICA	ML & CC	MSA	SW
Andronov et al.[84]	2016	β-tubulin in HeLa cells TPR	GSD	2D	20	NR	NR	NR	Ripley, Voronoi	1.5 × 1.5	✓					✓	SharpViSu
Andronov et al.[118]	2018	β-tubulin in HeLa cells; CENP-A in U2OS cells	dSTORM	3D	NR	28–50K	NR	3 exp.	Voronoi	ROIs from 18 × 18	✓	✓	2–4				3DClusterViSu
Haas et al.[86]	2018	RAD51, RPA in HeLa cells; HPNE, LN9, EUFA423	dSTORM	2D[a]	~30–40	25K	NR	5 exp.	Ripley, Delaunay, Voronoi	NR	✓		2–5				Grafeo
Peters et al.[97]	2018	F-actin in T cells; microtubule network in fixed HeLa cell	dSTORM	2D	60	100K	NR	3–5 cells	Voronoi; angular Ripley	3 × 3	✓		4				
Levet et al.[140]	2019	nuclear pore complex; microtubules; actin cytoskeleton regulators	DNA-PAINT dSTORM PALM	2D, 3D	20–60	40K	20K-8.3M	3–18 cells per cond.	Voronoi	NR	✓		4				Coloc-Tesseler
Khater et al.[126]	2018	Cav1, Cavin-1 in PC3 cells	GSD	3D	20–50	32–40K	0.45–1.2M	9–11 cells per cond. 2 cond. (4 exp.)	network graph mean-shift	18 × 18 × 0.8	✓	✓	28	✓	✓	✓	
Khater et al.[128]	2019	Cav1, Cavin-1 in PC3 cells	GSD	3D	20–50	32–40K	1.17–1.43M	10 cells per cond. (2 cond.)	network graph mean-shift graphlet	18 × 18 × 0.8	✓		11–47	✓	✓	✓	

The nomenclature used in this table is as follows. Ref, reference to the method; App, application; ImgMeth, imaging method; Dim, dimensionality; Res, resolution; #FPI, number of frames per image; #Loc, number of localizations; DataSz, dataset size; CAM, cluster analysis method; A/V, area/volume; WAV, whole-area visualization; SRIC, surface reconstruction for individual clusters; CP, cluster properties; ICA, intra-cluster analysis; ML & CC, machine learning and cluster classification; MSA, multi-scale analysis; SW, software; NR, information not reported.

[a]Data acquisition is 3D but the analysis is applied to the projected 2D data.

[b]Information is reported for simulated data.

[c]Code available upon request.



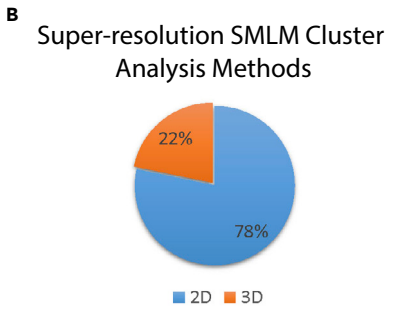


Figure 11. SMLM Imaging Techniques and Dimensionality Used in Various Publications

(A) The distribution of the publications based on the super-resolution SMLM imaging technique used in the study.

(B) The distribution of the publications based on the dimensionality of the super-resolution SMLM cluster analysis method.

clusters to train a model. Subtle features could be extracted from graph-based methods (e.g., network measures) and used to train a machine-learning model. Also, some recent deep-learning approaches for graphs[146,147] could be leveraged for analysis of SMLM data.

- Big-data analysis. The majority of the methods do not scale well to handle the big data generated from the super-resolution SMLM imaging techniques, while some methods are highly efficient and scale up efficiently with big data (e.g., graph-based and Voronoi tessellation-based methods).
- Cluster shape variation. Some of the methods could discover the clusters with various shapes (e.g., density-based methods). Some methods are more suitable for identifying tubular-like shapes (e.g., Voronoi tessellation-based methods).
- FOV/ROI analysis. The majority of the surveyed methods were used to analyze small ROIs rather than the whole field of view (FOV). Also, in most methods the ROIs were either selected manually or randomly from the whole FOV. We believe that selecting a small ROI is not a good strategy and will bias the cluster analysis. Analyzing ROI is dependent on its location in the cell. For example, selecting an ROI very close to the periphery of the cell could reveal structures that are different to those in an ROI in the middle of the cell, because the structures at the periphery might have different functions (e.g., focal adhesion) than the structures in the middle of the cell.
- Software. Some published software is designed to visualize SMLM data with very limited analysis capabilities, such as ViSP.[148] Some published software is limited to analyzing 2D regions of the SMLM data, is unstable, cannot handle the whole FOV but is limited to small ROIs, and is not robust to noise. Some software packages implement more than one method for analyzing SMLM data. In general, we noted limited work on automatic quantification and analysis methods applied to super-resolution SMLM data.
- 2D methods for 3D data. We noted that some methods acquire 3D super-resolution SMLM images; then, in order to leverage existing 2D super-resolution SMLM cluster analysis methods, they project the 3D data to 2D.[81,86] Projecting 3D data to 2D for analysis is not a good idea. Processing data in its 3D native format is much better to (1) avoid artifacts and (2) filter the noisy background localizations. About 78% of the super-resolution SMLM cluster analysis methods are 2D, as shown in Figure 11B. Around 49% of the SMLM imaging used STORM-based techniques as depicted in Figure 11A.
- Validation. Various ways to validate the methods were used, such as using DNA origami and nanorulers, synthetic data generation, SMLM simulators, and, finally, using real experimental cellular data with known biosignatures. Validation and evaluation of the different methods remains a challenging task for all surveyed methods, as no public dataset is available for benchmarking and assessing the performance of different (post-localization) analysis methods.

We summarize the number of publications (listed in Table 2) per year categorized according to super-resolution SMLM cluster analysis methods used in the study in Figure 12. Our survey shows that before 2014, only a couple of methods (i.e., statistical) and few studies addressed analysis of SMLM data. After 2014, researchers started exploring new clustering methods and, at the same time, the number of publications per year started growing, except for 2017. Graph-based cluster analysis methods applied to super-resolution SMLM data started appearing in 2018. We expect that more clustering methods based on graphs will appear in 2020 and onward. Ripley's functions are the most popular methods used for super-resolution SMLM cluster analysis over the years, as depicted in Figure 12. Furthermore, given their successes in analyzing other imaging modalities, we anticipate growth in the number of methods that leverage and adapt machine-learning (and particularly deep-learning) methods for SMLM analysis.

Conclusions

The SMLM imaging modality is relatively new and is creating exciting opportunities to help us understand the structure and function of many macromolecular complexes below the diffraction limit of fluorescence microscopy. The data it provides can enable discoveries, but we note that there is still a need and an opportunity to develop methods and tools that can (1) read data from different super-resolution microscopes and pre-process the data to handle image-acquisition artifacts, (2) provide different visualization alternatives, (3) analyze a large number of datasets in 3D, (4) extract and quantitatively describe the structural geometry and interaction of the underlying biological structures, and (5) do so in a way that is either robust to parameter settings or provide intuitive descriptions of parameters easily communicated to and understandable by the end user.

Our observations from conducting this review revealed that studying the various methods adopted for cluster analysis and quantification requires a benchmarking dataset and evaluation

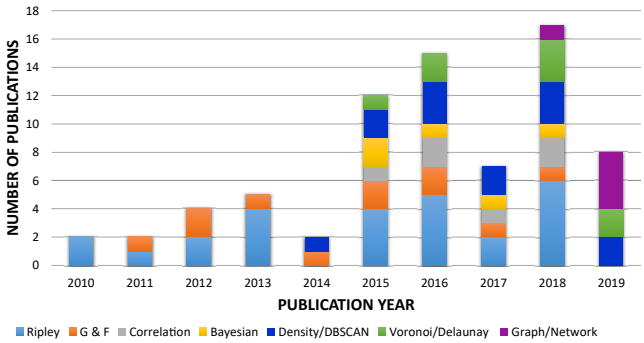


Figure 12. Number of Publications per Year (Starting from 2010) Categorized According to Super-Resolution SMLM Cluster Analysis Methods Used in the Study

measures for the assessment of the quality of clusters. The benchmark dataset should be available to validate both 2D and 3D cluster analysis methods. It should have several types of clusters with various densities, shapes, sizes, and noise levels.

Our review also motivated us to highlight the importance of strong interdisciplinary collaborations between computational scientists, biophysicists, biochemists, and biologists for novel breakthrough discoveries. Furthermore, artificial intelligence algorithms (e.g., machine learning) need to be incorporated in the analysis to (1) get rid of the subjectivity and bias, (2) robustly analyze the generated big data, and (3) automatically identify the distinct biological structures and their constituent biosignatures.

In this Review, we have summarized and compared the various computational methods for SMLM cluster analysis and quantification. We note that network/graph-based methods have more capabilities such that they could be used for pre-processing (e.g., correcting for multiple blinking of a single fluorophore artifact) and post-processing (e.g., filtering, segmenting, and identifying the biological structures) of the SMLM data. Graph-based methods could be applied to extract per-point and per-cluster features for analyzing 3D SMLM data, analyze the whole FOV, process big data (millions of localizations), and extract heterogeneous clusters. Graph-based methods are well suited to extract intra-cluster features and could be integrated with the machine-learning algorithms for automatic analysis and quantification of the underlying clusters.

In the end, we believe that synergy and harmony across research disciplines such as biophysics, biology, imaging, biochemistry, and artificial intelligence, among others, are required to improve our understanding of the underlying protein cluster structure and function. The extraordinary SMLM imaging modality and elegant computational methods will lead to a better understanding of protein interactions in several subcellular structures, consequently enhancing the modeling of antibodies and drug design to obtain better disease therapies.

Details of the criteria used in Table 2 are as follows:

- Application: the main biological application of the related studies. The imaged protein and biological model
- Imaging method: imaging technique applied in the study (the study may apply more than one imaging method; we are here focusing on the SMLM super-resolution methods only)
  - PALM: photoactivated localization microscopy
  - hsPALM: high-speed version of PALM
  - iPALM: interferometric PALM
  - FPALM: fluorescence PALM
  - STORM: stochastic optical reconstruction microscopy
  - dSTORM: direct STORM
  - GSD: ground state depletion
  - GSDIM: GSD followed by individual molecule return
  - HILO: highly inclined and laminated optical sheet
  - SOFI: super-resolution optical fluctuation imaging
  - SIM: structured illumination microscopy
- Dimensionality: is either 2D data analysis (for 2D acquisition) or 3D data analysis (for 3D acquisition) applied for the provided method?
- Resolution/localization precision: the resolution of the images on which the reported analysis is applied. (We report the localization precision if the resolution is missing.) The localization precision is much smaller than the spatial image resolution
- Area/volume: the total area or volume where the analysis is applied
- #frames per image: number of frames per super-resolution image (in thousands [K]).
- #localizations (blinks, events, pointillist, etc.): the total number or density of acquired blinks/molecules (in PALM-based methods)/localization events as described in the study
- Dataset size: dataset size (how many cells, how many experiments used in the work)
- Cluster analysis method: the clustering analysis method(s) used in the work
- Whole-area visualization: is cluster visualization of the whole analyzed area provided (after cluster analysis)?
- Surface reconstruction for individual clusters: is surface reconstruction for individual clusters provided?
- Cluster properties/descriptors: does the study provide analysis on any cluster properties? Yes or no. If yes, how many features have been used in the analysis?
- Intra-cluster analysis: does the study provide detailed analysis at intra-cluster levels (network analysis of the molecules, modularity analysis, subnetworks, etc.)?
- Machine learning and cluster classification: is machine learning used to automatically classify the clusters?
- Multi-scale analysis: is multi-scale analysis supported and used by the method?
- Software: is software available for the method/algorithm? Yes (software name) or no.

ACKNOWLEDGMENTS

Supported by grants from the CIHR (PJT-156424, PJT-159845), CHRP, NSERC, and CFI/BCKDF (I.R.N., G.H.). We thank the anonymous reviewers for their extensive feedback and insightful suggestions that were invaluable for improving this Review.

AUTHOR CONTRIBUTIONS

Conceptualization, I.M.K., I.R.N., and G.H.; Methodology, I.M.K., I.R.N., and G.H.; Software, I.M.K.; Formal Analysis, I.M.K, I.R.N., and G.H.; Investigation, I.M.K., I.R.N., and G.H.; Writing – Original Draft, I.M.K.; Writing – Review &

Editing, I.M.K., I.R.N., and G.H.; Visualization, I.M.K.; Funding Acquisition, I.R.N. and G.H.; Resources, I.R.N. and G.H.; Supervision, I.R.N. and G.H.

DECLARATION OF INTERESTS

An international patent PCT/CA2018/051553 covering the material presented herein has been submitted by the authors: "Methods for Analysis of Single Molecule Localization Microscopy to Define Molecular Architecture," US Patent Application No. 62/594,642, December 5, 2018.

REFERENCES

1. Abbe, E. (1873). Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. Arch. Mikrosk. Anat. *9*, 413–418.

2. Sezgin, E. (2017). Super-resolution optical microscopy for studying membrane structure and dynamics. J. Phys. Condens. Matter *29*, 273001.

3. Shashkova, S., and Leake, M.C. (2017). Single-molecule fluorescence microscopy review: shedding new light on old problems. Biosci. Rep. *37*, https://doi.org/10.1042/BSR20170031.

4. Klein, T., Proppert, S., and Sauer, M. (2014). Eight years of single-molecule localization microscopy. Histochem. Cell Biol. *141*, 561–575.

5. Choquet, D. (2014). The 2014 Nobel Prize in Chemistry: a large-scale prize for achievements on the nanoscale. Neuron *84*, 1116–1119.

6. Laine, R.F., Kaminski Schierle, G.S., van de Linde, S., and Kaminski, C.F. (2016). From single-molecule spectroscopy to super-resolution imaging of the neuron: a review. Methods Appl. Fluoresc. *4*, 022004.

7. Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. Science *313*, 1642–1645.

8. Moerner, W.E., and Kador, L. (1989). Optical detection and spectroscopy of single molecules in a solid. Phys. Rev. Lett. *62*, 2535.

9. Hell, S.W., and Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. Opt. Lett. *19*, 780–782.

10. Gustafsson, M.G.L. (2000). Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J. Microsc. *198*, 82–87.

11. Dertinger, T., Colyer, R., Iyer, G., Weiss, S., and Enderlein, J. (2009). Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). Proc. Natl. Acad. Sci. U S A *106*, 22287–22292.

12. Schidorsky, S., Yi, X., Razvag, Y., Sajman, J., Hermon, K., Weiss, S., and Sherman, E. (2018). Synergizing superresolution optical fluctuation imaging with single molecule localization microscopy. Methods Appl. Fluoresc. *6*, 045008.

13. Hess, S.T., Girirajan, T.P.K., and Mason, M.D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. Biophys. J. *91*, 4258–4272.

14. Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat. Methods *3*, 793–796.

15. Heilemann, M., Van De Linde, S., Schüttpelz, M., Kasper, R., Seefeldt, B., Mukherjee, A., Tinnefeld, P., and Sauer, M. (2008). Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. Angew. Chem. Int. Ed. *47*, 6172–6176.

16. Fölling, J., Bossi, M., Bock, H., Medda, R., Wurm, C.A., Hein, B., Jakobs, S., Eggeling, C., and Hell, S.W. (2008). Fluorescence nanoscopy by ground-state depletion and single-molecule return. Nat. Methods *5*, 943–945.

17. Schnitzbauer, J., Strauss, M.T., Schlichthaerle, T., Schueder, F., and Jungmann, R. (2017). Super-resolution microscopy with DNA-PAINT. Nat. Protoc. *12*, 1198.

18. Balzarotti, F., Eilers, Y., Gwosch, K.C., Gynnå, A.H., Westphal, V., Stefani, F.D., Elf, J., and Hell, S.W. (2017). Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. Science *355*, 606–612.

19. Owen, D.M., and Gaus, K. (2013). Imaging lipid domains in cell membranes: the advent of super-resolution fluorescence microscopy. Front. Plant Sci. *4*, 503.

20. Wegel, E., Göhler, A., Lagerholm, B.C., Wainman, A., Uphoff, S., Kaufmann, R., and Dobbie, I.M. (2016). Imaging cellular structures in super-resolution with SIM, STED and localisation microscopy: a practical comparison. Sci. Rep. *6*, 1–13.

21. Huang, B., Babcock, H., and Zhuang, X. (2010). Breaking the diffraction barrier: super-resolution imaging of cells. Cell *143*, 1047–1058.

22. Baddeley, D., and Bewersdorf, J. (2018). Biological insight from super-resolution microscopy: what we can learn from localization-based images. Annu. Rev. Biochem. *87*, 965–989.

23. Huang, B., Wang, W., Bates, M., and Zhuang, X. (2008). Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. Science *319*, 810–813.

24. Nicovich, P.R., Owen, D.M., and Gaus, K. (2017). Turning single-molecule localization microscopy into a quantitative bioanalytical tool. Nat. Protoc. *12*, 453.

25. Hell, S.W., Sahl, S.J., Bates, M., Zhuang, X., Heintzmann, R., Booth, M.J., Bewersdorf, J., Shtengel, G., Hess, H., Tinnefeld, P., et al. (2015). The 2015 super-resolution microscopy roadmap. J. Phys. D Appl. Phys. *48*, 443001.

26. Sauer, M., and Heilemann, M. (2017). Single-molecule localization microscopy in eukaryotes. Chem. Rev. *117*, 7478–7509.

27. Golfetto, O., Wakefield, D.L., Cacao, E.E., Avery, K.N., Kenyon, V., Jorand, R., Tobin, S.J., Biswas, S., Gutierrez, J., Clinton, R., et al. (2018). A platform to enhance quantitative single molecule localization microscopy. J. Am. Chem. Soc. *140*, 12785–12797.

28. Sahl, S.J., Hell, S.W., and Jakobs, S. (2017). Fluorescence nanoscopy in cell biology. Nat. Rev. Mol. Cell Biol. *18*, 685.

29. Turkowyd, B., Virant, D., and Endesfelder, U. (2016). From single molecules to life: microscopy at the nanoscale. Anal. Bioanal. Chem. *408*, 6885–6911.

30. Juette, M.F., Gould, T.J., Lessard, M.D., Mlodzianoski, M.J., Nagpure, B.S., Bennett, B.T., Hess, S.T., and Bewersdorf, J. (2008). Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples. Nat. Methods *5*, 527–529.

31. Prasanna Pavani, S.R., Thompson, M.A., Biteen, J.S., Lord, S.J., Liu, N., Twieg, R.J., Piestun, R., and Moerner, W.E. (2009). Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. Proc. Natl. Acad. Sci. U S A *106*, 2995–2999.

32. Baddeley, D., Cannell, M.B., and Soeller, C. (2011). Three-dimensional sub-100 nm super-resolution imaging of biological samples using a phase ramp in the objective pupil. Nano Res. *4*, 589–598.

33. Aristov, A., Lelandais, B., Rensen, E., and Zimmer, C. (2018). Zola-3d allows flexible 3D localization microscopy over an adjustable axial range. Nat. Commun. *9*, 1–8.

34. Shtengel, G., Galbraith, J.A., Galbraith, C.G., Lippincott-Schwartz, J., Gillette, J.M., Manley, S., Sougrat, R., Waterman, C.M., Kanchanawong, P., Davidson, M.W., et al. (2009). Interferometric fluorescent super-resolution microscopy resolves 3D cellular ultrastructure. Proc. Natl. Acad. Sci. U S A *106*, 3125–3130.

35. Aquino, D., Schönle, A., Geisler, C., Middendorff, C.V., Wurm, C.A., Okamura, Y., Lang, T., Hell, S.W., and Egner, A. (2011). Two-color nanoscopy of three-dimensional volumes by 4pi detection of stochastically switched fluorophores. Nat. Methods *8*, 353.

36. Bourg, N., Mayet, C., Dupuis, G., Barroca, T., Bon, P., Lécart, S., Fort, E., and Lévêque-Fort, S. (2015). Direct optical nanoscopy with axially localized detection. Nat. Photon. *9*, 587.

37. Ovesnỳ, M., Křížek, P., Borkovec, J., Švindrych, Z., and Hagen, G.M. (2014). ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. Bioinformatics *30*, 2389–2390.

38. Henriques, R., Lelek, M., Fornasiero, E.F., Valtorta, F., Zimmer, C., and Mhlanga, M.M. (2010). QuickPALM: 3D real-time photoactivation nanoscopy image processing in mageJ. Nat. Methods *7*, 339.

39. Wolter, S., Löschberger, A., Holm, T., Aufmkolk, S., Dabauvalle, M.C., Van De Linde, S., and Sauer, M. (2012). rapidSTORM: accurate, fast open-source software for localization microscopy. Nat. Methods *9*, 1040.

40. Rees, E.J., Erdelyi, M., Kaminski Schierle, G.S., Knight, A., and Kaminski, C.F. (2013). Elements of image processing in localization microscopy. J. Opt. *15*, 094012.

41. Shivanandan, A., Deschout, H., Scarselli, M., and Radenovic, A. (2014). Challenges in quantitative single molecule localization microscopy. FEBS Lett. *588*, 3595–3602.

42. Lambert, T.J., and Waters, J.C. (2017). Navigating challenges in the application of superresolution microscopy. J. Cell Biol. *216*, 53–63.

43. Lee, A., Tsekouras, K., Calderon, C., Bustamante, C., and Pressé, S. (2017). Unraveling the thousand word picture: an introduction to super-resolution data analysis. Chem. Rev. *117*, 7276–7330.

44. Endesfelder, U., and Heilemann, M. (2014). Art and artifacts in single-molecule localization microscopy: beyond attractive images. Nat. Methods *11*, 235.

45. Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U., and Radenovic, A. (2011). Quantitative photo activated localization microscopy: unraveling the effects of photoblinking. PLoS One *6*, e22678.

46. Culley, S., Albrecht, D., Jacobs, C., Pereira, P.M., Leterrier, C., Mercer, J., and Henriques, R. (2018). Quantitative mapping and minimization of super-resolution optical imaging artifacts. Nat. Methods *15*, 263.

47. Marsh, R.J., Pfisterer, K., Bennett, P., Hirvonen, L.M., Gautel, M., Jones, G.E., and Cox, S. (2018). Artifact-free high-density localization microscopy analysis. Nat. Methods *15*, 689–692.

48. Sage, D., Pham, T.A., Babcock, H., Lukes, T., Pengo, T., Chao, J., Velmurugan, R., Herbert, A., Agrawal, A., Colabrese, S., et al. (2019). Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software. Nat. Methods *16*, 387–395.

49. Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U., and Radenovic, A. (2011). Identification of clustering artifacts in photoactivated localization microscopy. Nat. Methods *8*, 527.

50. Fricke, F., Beaudouin, J., Eils, R., and Heilemann, M. (2015). One, two or three? probing the stoichiometry of membrane proteins by single-molecule localization microscopy. Scientific Rep. *5*, 14072.

51. Karathanasis, C., Fricke, F., Hummer, G., and Heilemann, M. (2017). Molecule counts in localization microscopy with organic fluorophores. ChemPhysChem *18*, 942–948.

52. Baumgart, F., Arnold, A.M., Leskovar, K., Staszek, K., Fölser, M., Weghuber, J., Stockinger, H., and Schütz, G.J. (2016). Varying label density allows artifact-free analysis of membrane-protein nanoclusters. Nat. Methods *13*, 661.

53. Sengupta, P., Jovanovic-Talisman, T., Skoko, D., Renz, M., Veatch, S.L., and Lippincott-Schwartz, J. (2011). Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. Nat. Methods *8*, 969.

54. Erdelyi, M., Rees, E., Metcalf, D., Kaminski Schierle, G.S., Dudas, L., Sinko, J., Knight, A.E., and Kaminski, C.F. (2013). Correcting chromatic offset in multicolor super-resolution localization microscopy. Opt. Express *21*, 10978–10988.

55. Griffié, J., Burn, G.L., Williamson, D.J., Peters, R., Rubin-Delanchy, P., and Owen, D.M. (2018). Dynamic Bayesian cluster analysis of live-cell single molecule localization microscopy datasets. Small Methods. https://doi.org/10.1002/smtd.201800008.

56. Lelek, M., Di Nunzio, F., Henriques, R., Charneau, P., Arhel, N., and Zimmer, C. (2012). Superresolution imaging of HIV in infected cells with flash-palm. Proc. Natl. Acad. Sci. U S A *109*, 8564–8569.

57. Szymborska, A., De Marco, A., Daigle, N., Cordes, V.C., Briggs, J.A.G., and Ellenberg, J. (2013). Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. Science *341*, 655–658.

58. Salvador-Gallego, R., Mund, M., Cosentino, K., Schneider, J., Unsay, J., Schraermeyer, U., Engelhardt, J., Ries, J., and García-Sáez, A.J. (2016). Bax assembly into rings and arcs in apoptotic mitochondria is linked to membrane pores. EMBO J. *35*, 389–401.

59. Mund, M., van der Beek, J.A., Deschamps, J., Dmitrieff, S., Hoess, P., Monster, J.L., Picco, A., Nédélec, F., Kaksonen, M., and Ries, J. (2018). Systematic nanoscale analysis of endocytosis links efficient vesicle formation to patterned actin nucleation. Cell *174*, 884–896.

60. Andronov, L., Ouararhni, K., Stoll, I., Klaholz, B.P., and Hamiche, A. (2019). CENP-A nucleosome clusters form rosette-like structures around HJURP during G1. Nat. Commun. *10*, 1–8.

61. Sherman, E. (2016). Resolving protein interactions and organization downstream the T cell antigen receptor using single-molecule localization microscopy: a review. Methods Appl. Fluoresc. *4*, 022002.

62. Ripley, B.D. (1977). Modelling spatial patterns. J. R. Stat. Soc. Series B Methodol. *39*, 172–192.

63. Dixon, P.M. (2014). Ripley's K Function (Wiley StatsRef: Statistics Reference Online).

64. Hansson, K., Jafari-Mamaghani, M., and Krieger, P. (2013). RipleyGUI: software for analyzing spatial patterns in 3D cell distributions. Front. Euroinformatics *7*, 5.

65. Kiskowski, M.A., Hancock, J.F., and Kenworthy, A.K. (2009). On the use of Ripley's K-function and its derivatives to analyze domain size. Biophys. J. *97*, 1095–1103.

66. Besag, J.E. (1977). Comments on Ripley's paper. J. R. Stat. Soc. B *39*, 193–195.

67. Ehrlich, M., Boll, W., Van Oijen, A., Hariharan, R., Chandran, K., Nibert, M.L., and Kirchhausen, T. (2004). Endocytosis by random initiation and stabilization of clathrin-coated pits. Cell *118*, 591–605.

68. Wiegand, T., and Moloney, K.A. (2004). Rings, circles, and null-models for point pattern analysis in ecology. Oikos *104*, 209–229.

69. Haase, P. (1995). Spatial pattern analysis in ecology based on Ripley's K-function: introduction and methods of edge correction. J. vegetation Sci. *6*, 575–582.

70. Marcon, E., and Puech, F. (2009). Generalizing Ripley's K Function to Inhomogeneous Populations. https://halshs.archives-ouvertes.fr/halshs-00372631/document.

71. Baddeley, A.J., Moyeed, R.A., Howard, C.V., and Boyde, A. (1993). Analysis of a three-dimensional point pattern with replication. J. R. Stat. Soc. Series C Appl. Stat. *42*, 641–668.

72. Goreaud, F., and Pélissier, R. (1999). On explicit formulas of edge effect correction for Ripley's K-function. J. Veg. Sci. *10*, 433–438.

73. Owen, D.M., Rentero, C., Rossy, J., Magenau, A., Williamson, D., Rodriguez, M., and Gaus, K. (2010). Palm imaging and cluster analysis of protein heterogeneity at the cell surface. J. Biophotonics *3*, 446–454.

74. Lillemeier, B.F., Mörtelmaier, M.A., Forstner, M.B., Huppa, J.B., Groves, J.T., and Davis, M.M. (2010). TCR and lat are expressed on separate protein islands on T cell membranes and concatenate during activation. Nat. Immunol. *11*, 90.

75. Pereira, C.F., Rossy, J., Owen, D.M., Mak, J., and Gaus, K. (2012). HIV taken by STORM: super-resolution fluorescence microscopy of a viral infection. Virol. J. *9*, 84.

76. Pageon, S.V., Cordoba, S.P., Owen, D.M., Rothery, S.M., Oszmiana, A., and Davis, D.M. (2013). Superresolution microscopy reveals nanometer-scale reorganization of inhibitory natural killer cell receptors upon activation of NKG2D. Sci. Signal. *6*, ra62.

77. Rossy, J., Owen, D.M., Williamson, D.J., Yang, Z., and Gaus, K. (2013). Conformational states of the kinase lck regulate clustering in early T cell signaling. Nat. Immunol. *14*, 82.

78. Rubin-Delanchy, P., Burn, G.L., Griffié, J., Williamson, D.J., Heard, N.A., Cope, A.P., and Owen, D.M. (2015). Bayesian cluster identification in single-molecule localization microscopy data. Nat. Methods *12*, 1072.

79. Caetano, F.A., Dirk, B.S., Tam, J.H.K., Cavanagh, P.C., Goiko, M., Ferguson, S.S.G., Pasternak, S.H., Dikeakos, J.D., de Bruyn, J.R., and Heit, B. (2015). MIiSR: molecular interactions in super-resolution imaging enables the analysis of protein interactions, dynamics and formation of multi-protein structures. PLoS Comput. Biol. *11*, e1004634.

80. Owen, D.M., Williamson, D.J., Boelen, L., Magenau, A., Rossy, J., and Gaus, K. (2013). Quantitative analysis of three-dimensional fluorescence localization microscopy data. Biophysical J. *105*, L05–L07.

81. Malkusch, S., Muranyi, W., Müller, B., Kräusslich, H.G., and Heilemann, M. (2013). Single-molecule coordinate-based analysis of the morphology of HIV-1 assembly sites with near-molecular spatial resolution. Histochem. Cell Biol. *139*, 173–179.

82. Pageon, S.V., Nicovich, P.R., Mollazade, M., Tabarin, T., and Gaus, K. (2016). Clus-DoC: a combined cluster detection and colocalization analysis for single-molecule localization microscopy data. Mol. Biol. Cell *27*, 3627–3636.

83. Griffié, J., Shlomovich, L., Williamson, D.J., Shannon, M., Aaron, J., Khuon, S., Burn, G., Boelen, L., Peters, R., Cope, A.P., et al. (2017). 3D Bayesian cluster analysis of super-resolution data reveals LAT recruitment to the T cell synapse. Sci. Rep. *7*, 4077.

84. Andronov, L., Lutz, Y., Vonesch, J.L., and Klaholz, B.P. (2016). SharpViSu: integrated analysis and segmentation of super-resolution microscopy data. Bioinformatics *32*, 2239–2241.

85. Gao, J., Wang, F., Chen, J., Wang, J., Cai, M., Xu, H., Jiang, J., and Wang, H. (2016). Super-resolution imaging of STAT3 cellular clustering during nuclear transport. RSC Adv. *6*, 54597–54607.

86. Haas, K.T., Lee, M., Esposito, A., and Venkitaraman, A.R. (2018). Single-molecule localization microscopy reveals molecular transactions during RAD51 filament assembly at cellular DNA damage sites. Nucleic Acids Res. *46*, 2398–2416.

87. Bálint, Š., Lopes, F.B., and Davis, D.M. (2018). A nanoscale reorganization of the IL-15 receptor is triggered by NKG2D in a ligand-dependent manner. Sci. Signal. *11*, eaal3606.

88. Lopes, F.B., Bálint, Š., Valvo, S., Felce, J.H., Hessel, E.M., Dustin, M.L., and Davis, D.M. (2017). Membrane nanoclusters of FcγRI segregate from inhibitory SIRPα upon activation of human macrophages. J. Cell Biol. *216*, 1123–1141.

89. Wee, J.L., Schulze, K.E., Jones, E.L., Yeung, L., Cheng, Q., Pereira, C.F., Costin, A., Ramm, G., van Spriel, A.B., Hickey, M.J., et al. (2015). Tetraspanin CD37 regulates β2 integrin-mediated adhesion and migration in neutrophils. J. Immunol. *195*, 5770–5779.

90. Griffié, J., Boelen, L., Burn, G., Cope, A.P., and Owen, D.M. (2015). Topographic prominence as a method for cluster identification in single-molecule localisation data. J. Biophotonics *8*, 925–934.

91. Oszmiana, A., Williamson, D.J., Cordoba, S.P., Morgan, D.J., Kennedy, P.R., Stacey, K., and Davis, D.M. (2016). The size of activating and inhibitory killer Ig-like receptor nanoclusters is controlled by the transmembrane sequence and affects signaling. Cell Rep. *15*, 1957–1972.

92. Owen, D.M., Williamson, D.J., Magenau, A., and Gaus, K. (2012). Sub-resolution lipid domains exist in the plasma membrane and regulate protein diffusion and distribution. Nat. Commun. *3*, 1256.

93. Williamson, D.J., Owen, D.M., Rossy, J., Magenau, A., Wehrmann, M., Gooding, J.J., and Gaus, K. (2011). Pre-existing clusters of the adaptor lat do not participate in early T cell signaling events. Nat. Immunol. *12*, 655.

94. Peters, R., Griffié, J., Williamson, D.J., Aaron, J., Khuon, S., and Owen, D.M. (2018). Development of 2-colour and 3D SMLM data analysis methods for fibrous spatial point patterns. J. Phys. D Appl. Phys. *52*, 014005.

95. Lagache, T., Grassart, A., Dallongeville, S., Faklaris, O., Sauvonnet, N., Dufour, A., Danglot, L., and Olivo-Marin, J.C. (2018). Mapping molecular assemblies with fluorescence microscopy and object-based spatial statistics. Nat. Commun. *9*, 698.

96. Malkusch, S., and Heilemann, M. (2016). Extracting quantitative information from single-molecule super-resolution imaging data with LAMA—LocAlization Microscopy Analyzer. Sci. Rep. *6*, 34486.

97. Peters, R., Griffié, J., Burn, G.L., Williamson, D.J., and Owen, D.M. (2018). Quantitative fibre analysis of single-molecule localization microscopy data. Sci. Rep. *8*, 10418.

98. Getis, A., and Franklin, J. (1987). Second-order neighborhood analysis of mapped point patterns. Ecology *68*, 473–477.

99. Rossy, J., Cohen, E., Gaus, K., and Owen, D.M. (2014). Method for co-cluster analysis in multichannel single-molecule localisation data. Histochem. Cell Biol. *141*, 605–612.

100. Veatch, S.L., Machta, B.B., Shelby, S.A., Chiang, E.N., Holowka, D.A., and Baird, B.A. (2012). Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting. PLoS One *7*, e31457.

101. Malkusch, S., Endesfelder, U., Mondry, J., Gelléri, M., Verveer, P.J., and Heilemann, M. (2012). Coordinate-based colocalization analysis of single-molecule localization microscopy data. Histochem. Cell Biol. *137*, 1–10.

102. Schnitzbauer, J., Wang, Y., Zhao, S., Bakalar, M., Nuwal, T., Chen, B., and Huang, B. (2018). Correlation analysis framework for localization-based superresolution microscopy. Proc. Natl. Acad. Sci. U S A *115*, 3219–3224.

103. Tobin, S.J., Wakefield, D.L., Jones, V., Liu, X., Schmolze, D., and Jovanović-Talisman, T. (2018). Single molecule localization microscopy coupled with touch preparation for the quantification of trastuzumab-bound HER2. Sci. Rep. *8*, 15154.

104. Stone, M.B., and Veatch, S.L. (2015). Steady-state cross-correlations for live two-colour super-resolution localization data sets. Nat. Commun. *6*, 7347.

105. Griffié, J., Shannon, M., Bromley, C.L., Boelen, L., Burn, G.L., Williamson, D.J., Heard, N.A., Cope, A.P., Owen, D.M., and Rubin-Delanchy, P. (2016). A Bayesian cluster analysis method for single-molecule localization microscopy data. Nat. Protoc. *11*, 2499.

106. Ester, M., Kriegel, H.P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In KDD-96 Proceedings, pp. 226–231.

107. Mazouchi, A., and Milstein, J.N. (2015). Fast optimized cluster algorithm for localizations (FOCAL): a spatial cluster analysis for super-resolved microscopy. Bioinformatics *32*, 747–754.

108. Nino, D., Djayakarsana, D., and Milstein, J.N. (2019). Focal3D: a 3-dimensional clustering package for single-molecule localization microscopy. bioRxiv. https://doi.org/10.1101/777722.

109. Pengo, T., Holden, S.J., and Manley, S. (2014). PALMsiever: a tool to turn raw data into results for single-molecule localization microscopy. Bioinformatics *31*, 797–798.

110. Sieben, C., Banterle, N., Douglass, K.M., Gönczy, P., and Manley, S. (2018). Multicolor single-particle reconstruction of protein complexes. Nat. Methods *15*, 777.

111. Barna, L., Dudok, B., Miczán, V., Horváth, A., László, Z.I., and Katona, I. (2016). Correlated confocal and super-resolution imaging by VividSTORM. Nat. Protoc. *11*, 163.

112. Mollazade, M., Tabarin, T., Nicovich, P.R., Soeriyadi, A., Nieves, D.J., Gooding, J.J., and Gaus, K. (2017). Can single molecule localization microscopy be used to map closely spaced RGD nanodomains? PLoS One *12*, e0180871.

113. Zhang, Y., Lara-Tejero, M., Bewersdorf, J., and Galán, J.E. (2017). Visualization and characterization of individual type III protein secretion machines in live bacteria. Proc. Natl. Acad. Sci. U S A *114*, 6098–6103.

114. Okabe, A., Boots, B., Sugihara, K., and Chiu, S.N. (2009). Spatial Tessellations: Concepts and Applications of Voronoi Diagrams (John Wiley).

115. Levet, F., Hosy, E., Kechkar, A., Butler, C., Beghin, A., Choquet, D., and Sibarita, J.B. (2015). SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data. Nat. Methods *12*, 1065.

116. Andronov, L., Orlov, I., Lutz, Y., Vonesch, J.L., and Klaholz, B.P. (2016). ClusterViSu, a method for clustering of protein complexes by Voronoi tessellation in super-resolution microscopy. Sci. Rep. *6*, 24084.

117. Baddeley, D., Jayasinghe, I.D., Lam, L., Rossberger, S., Cannell, M.B., and Soeller, C. (2009). Optical single-channel resolution imaging of the ryanodine receptor distribution in rat cardiac myocytes. Proc. Natl. Acad. Sci. U S A *106*, 22275–22280.

118. Andronov, L., Michalon, J., Ouararhni, K., Orlov, I., Hamiche, A., Vonesch, J.L., and Klaholz, B.P. (2018). 3DClusterViSu: 3D clustering analysis of super-resolution microscopy data by 3D Voronoi tessellations. Bioinformatics *1*, 9.

119. Newman, M.E.J. (2003). The structure and function of complex networks. SIAM Rev. *45*, 167–256.

120. Kim, J., and Wilhelm, T. (2008). What is a complex graph? Phys. A Stat. Mech. Appl. *387*, 2637–2652.

121. Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101.

122. Baronchelli, A., Ferrer-i Cancho, R., Pastor-Satorras, R., Chater, N., and Christiansen, M.H. (2013). Networks in cognitive science. Trends Cogn. Sci. *17*, 348–360.

123. Bassett, D.S., and Sporns, O. (2017). Network neuroscience. Nat. Neurosci. *20*, 353.

124. Sporns, O. (2013). Structure and function of complex brain networks. Dialogues Clin. Neurosci. *15*, 247.

125. Newman, M. (2010). Networks: An Introduction (Oxford University Press).

126. Khater, I.M., Meng, F., Wong, T.H., Robert Nabi, I., and Hamarneh, G. (2018). Super resolution network analysis defines the molecular architecture of caveolae and caveolin-1 scaffolds. Sci. Rep. *8*, 9009.

127. Khater, I.M., Scriven, D.R.L., Moore, E.D.W., and Hamarneh, G. (2016). Sub-cellular network analysis of ryanodine receptor positioning in control and phosphorylated states. In 2016 Computing in Cardiology Conference (CinC) (IEEE), pp. 821–824.

128. Khater, I.M., Meng, F., Robert Nabi, I., and Hamarneh, G. (2019). Identification of caveolin-1 domain signatures via machine learning and graphlet analysis of single molecule super-resolution data. Bioinformatics *35*, 3468–3475.

129. Khater, I.M., Liu, Q., Chou, K.C., Hamarneh, G., and Robert Nabi, I. (2019). Super-resolution modularity analysis shows polyhedral caveolin-1 oligomers combine to form scaffolds and caveolae. Sci. Rep. *9*, 9888.

130. Scurll, J., Abraham, L., Zheng, D.W., Tafteh, R., Chou, K., Gold, M.R., and Coombs, D. (2019). StormGraph: an automated graph-based algorithm for quantitative clustering analysis of single-molecule localization microscopy data. bioRxiv. https://doi.org/10.1101/515627.

131. Pike, J.A., Khan, A.O., Pallini, C., Thomas, S.G., Mund, M., Ries, J., Poulter, N.S., and Styles, I.B. (2019). Topological data analysis quantifies biological nano-structure from single molecule localization microscopy. Bioinformatics *36*, 1614–1621.

132. Nehme, E., Weiss, L.E., Michaeli, T., and Shechtman, Y. (2018). Deep-STORM: super-resolution single-molecule microscopy by deep learning. Optica *5*, 458–464.

133. Boyd, N., Jonas, E., Babcock, H.P., and Recht, B. (2018). DeepLoco: fast 3D localization microscopy using neural networks. bioRxiv. https://doi.org/10.1101/267096.

134. Cardoen, B., Ben Yedder, H., Sharma, A., Chou, K.C., Robert Nabi, I., and Hamarneh, G. (2019). Ergo: efficient recurrent graph optimized emitter density estimation in single molecule localization microscopy. IEEE Trans. Med. Imaging. https://doi.org/10.1109/TMI.2019.2962361.

135. Khater, I.M., Aroca-Ouellette, S.T., Meng, F., Nabi, I.R., and Hamarneh, G. (2019). Caveolae and scaffold detection from single molecule localization microscopy data using deep learning. PLoS One *14*, https://doi.org/10.1371/journal.pone.0211659.

136. Williamson, D.J., Burn, G.L., Simoncelli, S., Griffié, J., Peters, R., Davis, D.M., and Owen, D.M. (2020). Machine learning for cluster analysis of localization microscopy data. Nat. Commun. *11*, 1493.

137. Venkataramani, V., Herrmannsdörfer, F., Heilemann, M., and Kuner, T. (2016). SuReSim: simulating localization microscopy experiments from ground truth models. Nat. Methods *13*, 319.

138. Novák, T., Gajdos, T., Sinkó, J., Szabó, G., and Erdélyi, M. (2017). TestSTORM: versatile simulator software for multimodal super-resolution localization fluorescence microscopy. Sci. Rep. *7*, 951.

139. Lindén, M., Ćurić, V., Boucharin, A., Fange, D., and Elf, J. (2016). Simulated single molecule microscopy with SMeagol. Bioinformatics *32*, 2394–2395.

140. Levet, F., Julien, G., Galland, R., Butler, C., Beghin, A., Chazeau, A., Hoess, P., Ries, J., Giannone, G., and Sibarita, J.B. (2019). A tessellation-based colocalization analysis approach for single-molecule localization microscopy. Nat. Commun. *10*, 1–12.

141. Spahn, C., Herrmannsdörfer, F., Kuner, T., and Heilemann, M. (2016). Temporal accumulation analysis provides simplified artifact-free analysis of membrane-protein nanoclusters. Nat. Methods *13*, 963.

142. Li, Y., Mund, M., Hoess, P., Deschamps, J., Matti, U., Nijmeijer, B., Jimenez Sabinina, V., Ellenberg, J., Schoen, I., and Ries, J. (2018). Real-time 3D single-molecule localization using experimental point spread functions. Nat. Methods *15*, 367.

143. Raab, M., Jusuk, I., Molle, J., Buhr, E., Bodermann, B., Bergmann, D., Bosse, H., and Tinnefeld, P. (2018). Using DNA origami nanorulers as traceable distance measurement standards and nanoscopic benchmark structures. Sci. Rep. *8*, 1780.

144. Zanacchi, F.C., Manzo, C., Alvarez, A.S., Derr, N.D., Garcia-Parajo, M.F., and Lakadamyali, M. (2017). A DNA origami platform for quantifying protein copy number in super-resolution. Nat. Methods *14*, 789.

145. Sieben, C., Douglass, K.M., Guichard, P., and Manley, S. (2018). Super-resolution microscopy to decipher multi-molecular assemblies. Curr. Opin. Struct. Biol. *49*, 169–176.

146. Kipf, T.N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv, 1609.02907.

147. M Defferrard, X Bresson, and P Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In 30th Conference on Neural Information Processing Systems (NIPS 2016) pages 3844–3852, 2016.

148. El Beheiry, M., and Dahan, M. (2013). ViSP: representing single-particle localizations in three dimensions. Nat. Methods *10*, 689.

149. Krüger, C.L., Zeuner, M.T., Cottrell, G.S., Widera, D., and Heilemann, M. (2017). Quantitative single-molecule imaging of TLR4 reveals ligand-specific receptor dimerization. Sci. Signal. *10*, eaan1308.

150. Lukeš, T., Glatzová, D., Kvíčalová, Z., Levet, F., Benda, A., Letschert, S., Sauer, M., Brdička, T., Lasser, T., and Cebecauer, M. (2017). Quantifying protein densities on cell membranes using super-resolution optical fluctuation imaging. Nat. Commun. *8*, 1731.

151. Paul, M.W., de Gruiter, H.M., Lin, Z., Baarends, W.M., van Cappellen, W.A., Houtsmuller, A.B., and Slotman, J.A. (2019). SMoLR: visualization and analysis of single-molecule localization microscopy data in R. BMC Bioinformatics *20*, 30.