Clus-DoC: a combined cluster detection and colocalization analysis for single-molecule localization microscopy data

Sophie V. Pageon, Philip R. Nicovich, Mahdie Mollazade, Thibault Tabarin, and Katharina Gaus*
EMBL Australia Node in Single Molecule Science, School of Medical Sciences, and ARC Centre of Excellence in Advanced Molecular Imaging, University of New South Wales, Sydney, NSW 2052, Australia

ABSTRACT Advances in fluorescence microscopy are providing increasing evidence that the spatial organization of proteins in cell membranes may facilitate signal initiation and integration for appropriate cellular responses. Our understanding of how changes in spatial organization are linked to function has been hampered by the inability to directly measure signaling activity or protein association at the level of individual proteins in intact cells. Here we solve this measurement challenge by developing Clus-DoC, an analysis strategy that quantifies both the spatial distribution of a protein and its colocalization status. We apply this approach to the triggering of the T-cell receptor during T-cell activation, as well as to the functionality of focal adhesions in fibroblasts, thereby demonstrating an experimental and analytical workflow that can be used to quantify signaling activity and protein colocalization at the level of individual proteins.

INTRODUCTION
Most biological processes are regulated by complex networks of protein–protein interactions that ultimately lead to specific cellular behaviors. The process of signal transduction is often mediated by heterogeneous and dynamic protein complexes rather than through linear cascades of biochemical interactions as frequently depicted in textbooks. An in-depth understanding of the molecular mechanisms underlying the regulation of cell behavior requires better knowledge of the molecular composition and molecular interactions of multiscale molecular structures and signaling complexes. Conventional biochemical, genetic, structural, and imaging approaches have provided important insights into biological processes but do not provide a molecular picture of the key nanometer-scale organization. During the past 30 yr, fluorescence microscopy has emerged as a valuable tool to unravel biomolecular interactions in cellular processes by facilitating the systematic study of protein colocalization in cells. However, the spatial resolution of conventional fluorescence microscopy is limited by diffraction to ∼250 nm, whereas molecular interactions occur at a scale of ∼10 nm. Superresolution and single-molecule imaging techniques have revolutionized the way in which we can address these complex biological questions. Constant improvements in methodology are increasingly pushing the limits of resolution, turning the focus back on how biologists can use these tools. Several techniques can locate single fluorescent molecules in intact cells down to a few nanometers (Moerner, 2012). Collectively known as single-molecule localization microscopy (SMLM), these techniques include photoactivated localization microscopy (PALM; Betzig et al., 2006) and direct stochastic optical reconstruction microscopy (dSTORM; Heimann et al., 2008; van de Linde et al., 2011). SMLM techniques reconstruct a molecular image of the sample by exploiting fluorophores that exist in both a dark and a bright state, such as photoactivatable or photoswitchable fluorescent proteins or dyes. A sparse subset of fluorophores can be photoconverted to the bright state, imaged, and then bleached, and the location of these molecules can be determined with high precision using center-of-mass algorithms (Henriques et al., 2010). Repeating this cycle over several thousand frames yields the localizations of all of the imaged molecules in the sample, making SMLM an effective tool in the study of protein interactions.
approach for quantifying the distribution of proteins in cells (Owen et al., 2010).

Unlike conventional diffraction-limited microscopy or ensemble superresolution microscopy, SMLM techniques produce not just a “pretty image” but also a map of the coordinates of all detected individual points (Endesfelder and Heilemann, 2014). This allows a level of quantification that goes beyond the conventional way of analyzing images by measuring averages, such as intensity correlation-based colocalization of two proteins in an image using Pearson’s r (Pearson, 1896; Manders et al., 1992) or Manders’ overlap coefficient (Manders et al., 1993). By retaining the spatial information of each molecule, local parameters can be extracted to establish hierarchical information—at the level of the whole cell, at the level of protein clusters or complexes, and at the level of molecules.

Point pattern analysis methods evaluate the spatial distribution of points in an image through different approaches. A global overview of the distribution of points can be obtained using Ripley’s K (Ripley, 1979; Perry, 2004; Owen et al., 2010) or pair correlation analysis (Sengupta et al., 2011; Veatch et al., 2012), allowing researchers to determine whether the proteins are dispersed, clustered, or randomly distributed within the region of interest. Clusters are defined as regions of high density separated by regions of lower density. For the identification of clusters, hierarchical clustering methods such as density-based spatial clustering of applications with noise (DBSCAN; Ester et al., 1996) or OPTICS (Ankerst et al., 1999) provide additional segmentation of clusters, providing quantitative information at the level of individual clusters. Ripley’s K can also be extended to allow cluster detection (Owen et al., 2010).

Similarly, several methods have been developed to investigate the colocalization of two proteins imaged by SMLM, including coordinate-based colocalization (CBC) analysis (Malkusch et al., 2012) and coclustering methods (Rossy et al., 2014). CBC analysis has also been applied to single-color data to correlate localization distributions from the same protein with each other and determine the degree of clustering for each individual point (Tarancón Díez et al., 2014). Voronoi tessellation is another strategy that has recently been implemented for cluster detection and colocalization analysis in SMLM data (Levet et al., 2015; Andronov et al., 2016).

Whereas cell biologists are exploiting SMLM techniques and producing large amounts of finely detailed images, the development of analysis tools that are able to extract relevant quantitative information is lagging behind. Although many algorithms underlying point pattern analysis methods have been published, they have not often been implemented in biological applications. There is thus an urgent need to make these analytical methods more accessible in order to address specific biological questions. Malkusch and Heilemann (2016) released an analysis toolbox called Lama: The LocAlization Microscopy Analyzer, which comprises a collection of algorithms for postprocessing and data analysis for SMLM data sets. This is a useful tool because it provides a variety of previously published routines for extracting quantitative information on localization precision, spatial distributions, cluster size, cluster composition, and colocalization within the same package.

Here we propose a new approach that combines two analysis techniques of cluster detection (Ester et al., 1996) and colocalization (Malkusch et al., 2012) to develop a multilevel analysis that goes beyond what these methods can achieve separately. We call this combined analysis strategy cluster detection with degree of colocalization (Clus-DoC). Clus-DoC can subclassify the data for a deeper analysis at the level of different cluster populations. By unifying different tools into one analysis, we can achieve a more in-depth analysis of colocalization in the context of protein clustering. This type of approach is applicable to various biological systems; here we demonstrate its capabilities in relation to T-cell signaling and focal adhesions.

RESULTS AND DISCUSSION

Acquisition of SMLM data

To achieve nanometer-scale resolution, SMLM exploits the properties of photoactivatable or photoswitchable fluorophores to reconstruct a single-molecule image (Betzig et al., 2006; Rust et al., 2006). The basic principle behind SMLM is the stochastic photoactivation of small numbers of fluorescent molecules and the localization of each activated molecule using center-of-mass algorithms. Thus, in a typical SMLM acquisition, several thousand frames need to be captured in order to image sufficient molecules to produce a high-resolution image. We typically acquire 20,000 frames (Figure 1A) and achieve a localization precision of 20-30 nm. Reconstructed SMLM images contain single-molecule localizations, with each point representing one detected molecule (Figure 1, B and C). SMLM methods generate data sets of molecular coordinates at nanometer resolution rather than traditional fluorescence intensity images. One of the advantages of conventional fluorescence microscopy that is retained in SMLM is the capability for multicolor imaging. Because most cellular functions involve multiple components, multicolor SMLM is an essential strategy to address how interactions between different proteins determine biological function. Although several methods have been proposed to quantify dual-color SMLM data, none of them satisfyingly incorporates an evaluation of the spatial distribution with the quantification of colocalization between the two colors, making it difficult to evaluate the link between protein clustering and function.

FIGURE 1: SMLM acquisition and image reconstruction. (A) For SMLM, a large number of frames are acquired for each cell. In each frame, only sparse subsets of molecules are fluorescent and can be accurately localized. If sufficient frames are collected (e.g., 20,000 frames in our example), a superresolved image of the sample can be reconstructed from the localizations of all molecules detected during the acquisition. (B) Reconstructed images of a two-color SMLM acquisition. Scale bar, 10 μm. (C) Enlargement of the 4 μm × 4 μm region highlighted in white in B. Each individual dot represents one detected molecule.
It is important to note several challenges when using SMLM for quantitative measurements. For example, sources of error include overcounting of fluorophores due to photoblinking (Annibale et al., 2010; Lee et al., 2012; Veatch et al., 2012) or limited detection efficiency related to incomplete photoconversion of fluorochromes (Annibale et al., 2011; Durisic et al., 2012). Therefore individual detected localizations do not always represent biological molecules at a 1:1 ratio. Photoblinking occurs when a fluorophore goes into a long-lived dark state instead of getting photobleached and then returns to the bright state in a later frame. To account for the photoblinking phenomenon, several “grouping” methods have been proposed that use a threshold in time and space to assign several localizations to one molecule (Annibale et al., 2011). However, the optimal grouping parameters to correct for photoblinking vary between fluorophores and experimental settings, and quantification remains difficult. Here we used ungrouped data to avoid grouping artifacts and to retain high numbers of data points. For obtaining absolute numbers of molecules, grouped data should be used.

Single-molecule cluster detection and colocalization analysis

For cluster detection, we implemented DBSCAN (Ester et al., 1996) and applied it to the table of coordinates generated from SMLM. DBSCAN has been used in biology to investigate the role of RAF multimers in cell growth (Nan et al., 2013) and the spatial organization of RNA polymerase during bacterial transcription (Endesfelder et al., 2013). DBSCAN is a propagative cluster detection method that links points that are closely packed together and detects outliers. Points are considered part of a cluster if they fulfill two criteria of connectivity: they need to have a number of neighbors equal to or greater than MinPts (minimum number of neighbors) within a radius $r$ (Figure 2A). The analysis starts with an arbitrarily selected starting point in the image. If this point has MinPts neighbors or more within $r$, a cluster is started, and this point and its neighbors are part of the cluster. If the neighbors themselves have MinPts neighbors or more, the cluster is propagated; if not, these points are cluster boundary points. A new starting point that has not been visited is then arbitrarily selected to continue the analysis. If a point does not have more than MinPts neighbors and is not a neighbor for another point, it is classified as an outlier, that is, a nonclustered point, corresponding to isolated molecules or noise. Some of the advantages of using DBSCAN over other point pattern analysis methods are that DBSCAN can detect arbitrarily shaped clusters (rather than being biased toward circular clusters, as in Ripley’s K cluster analysis), it is rapid, and it is robust to outliers. In our implementation of the DBSCAN analysis, the outputs include a cluster map in which clustered molecules are in green, nonclustered molecules are in gray, and cluster contours are in black, as well as quantification of certain cluster properties, such as cluster size (Table 1). The user can select the values for the parameters MinPts and search radius, $r$. A good starting point for choosing $r$ can be the median localization precision ($\sigma$, in nanometers). In our experimental conditions, $\sigma = 21.3$ nm for PSCFP2 and 20.6 nm for Alexa Fluor 647. We therefore typically use MinPts = 3 and $r = 20$ nm and define a cluster as having 10 localizations or more.

To quantify the extent of colocalization between two proteins at the single-molecule level, we implemented a coordinate-based colocalization method (Malkusch et al., 2012). Intensity-based approaches can be used on SMLM data but require histogramming or blurring of localizations, thereby losing the resolution of single-molecule localizations. By applying colocalization algorithms directly to the coordinates of single molecules, information is retained at the single-molecule level. This approach relies on the comparison of the spatial distribution of surrounding localizations from both species around each individual point (Figure 2B). For each molecule of protein A, the number of molecules of protein A and protein B within circles of increasing radius is calculated, respectively, giving the density gradients of protein A and protein B around this molecule. These density gradients are then corrected by the density at the maximum radius respectively for A and B. The two distributions are compared by calculating a rank correlation coefficient (using Spearman correlation), in which the colocalization coefficient is weighted by a value proportional to the distance to the nearest neighbor to avoid long-distance effects (Malkusch et al., 2012). Each molecule is thus assigned a DoC score ranging from −1 (anticolocalized or segregated), through 0 for noncorrelated distributions (no colocalization), to +1 (perfectly colocalized). Our implementation of this analysis yields colocalization maps for protein A or protein B that retain the spatial location of molecules and color codes them according to their DoC score. The level of colocalization with the other species. Thus this method enables the visualization of the distribution of colocalization within an image and quantifies the DoC score on a per-molecule basis within the image. We take this analysis one step further by using a DoC threshold for colocalization, allowing us to calculate the percentage of colocalized molecules for protein A and for protein B (Table 1). We
method we call Clus-DoC (Figure 3). Here the first step in the analysis is to calculate the DoC score of each molecule for both channels using the coordinate-based colocalization method described earlier (green in Figure 3). In the next step, individual clusters are detected using DBSCAN for all localizations—this determines, for each given point, whether it is part of a cluster or not—and the cluster contours are defined using a two-dimensional (2D) histogram smoothing method (blue in Figure 3). From these two steps, localizations are simultaneously classified into clusters and assigned a DoC score. This allows us to perform a second level of classification (red in Figure 3) in which clusters are separated into colocalized clusters (containing a user-defined number and DoC threshold, i.e., >10 localizations with a DoC score >0.4) and noncolocalized clusters. These two populations of clusters are then further analyzed to extract cluster properties. Alternatively, clusters can be subclassified according to cluster characteristics, such as size, density, or number of molecules per cluster, and these cluster populations can be interrogated regarding average DoC scores or distribution of DoC scores. Thus this combined approach not only provides cluster parameters for individual clusters, but it also facilitates a cluster reclassification according to a specific parameter (such as the degree of colocalization). In all cases, it is also possible to compare the DoC distributions for clustered and nonclustered molecules.

### TABLE 1: Output parameters from DoC, DBSCAN, and Clus-DoC.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Output parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoC</td>
<td>DoC score for each localization for protein A and protein B</td>
</tr>
<tr>
<td></td>
<td>Percentage of colocalized points for protein A and protein B (threshold based)</td>
</tr>
<tr>
<td>DBSCAN plus contour mapping</td>
<td>Percentage of localizations in clusters</td>
</tr>
<tr>
<td></td>
<td>Number of clusters in ROI</td>
</tr>
<tr>
<td></td>
<td>Number of localizations per cluster</td>
</tr>
<tr>
<td></td>
<td>Circularity of clusters</td>
</tr>
<tr>
<td></td>
<td>Size of clusters</td>
</tr>
<tr>
<td></td>
<td>Absolute molecular density within clusters</td>
</tr>
<tr>
<td></td>
<td>Relative molecular density within clusters</td>
</tr>
<tr>
<td>Clus-DoC</td>
<td>Distinguishes colocalized versus noncolocalized clusters for protein A or protein B:</td>
</tr>
<tr>
<td></td>
<td>- Number of molecules per cluster</td>
</tr>
<tr>
<td></td>
<td>- Circularity of clusters</td>
</tr>
<tr>
<td></td>
<td>- Size of clusters</td>
</tr>
<tr>
<td></td>
<td>- Absolute or relative density within clusters</td>
</tr>
<tr>
<td></td>
<td>- Distribution of total or colocalized molecules between colocalized and noncolocalized clusters</td>
</tr>
<tr>
<td></td>
<td>Distinguishes cluster populations based on density/size/number of localizations per cluster:</td>
</tr>
<tr>
<td></td>
<td>- Average DoC score</td>
</tr>
<tr>
<td></td>
<td>- Distribution of DoC scores</td>
</tr>
<tr>
<td></td>
<td>- Percentage of colocalized molecules within each cluster population</td>
</tr>
<tr>
<td></td>
<td>- Distribution of colocalized molecules between the subpopulations</td>
</tr>
<tr>
<td></td>
<td>- Cluster morphology parameters</td>
</tr>
<tr>
<td></td>
<td>Isolates nonclustered molecules:</td>
</tr>
<tr>
<td></td>
<td>- Average DoC score</td>
</tr>
<tr>
<td></td>
<td>- Proportion of total molecules and colocalized molecules that are not in clusters</td>
</tr>
<tr>
<td></td>
<td>- Percentage of nonclustered molecules that are colocalized</td>
</tr>
</tbody>
</table>

chose the threshold based on the distribution of DoC scores when a single-molecule data set was doubled and shifted in one direction by 10 nm (Malkusch et al., 2012). This resulted in the majority of the localizations (>90%) having a DoC score >0.4, with a peak of the DoC distribution at 1, indicating high colocalization, despite the 10-nm shift. Because in practice SMLM at best achieves 10-nm precision, it is therefore reasonable to assume that a 10-nm shift recapitulates experimental conditions. Therefore a threshold of 0.4 was chosen to distinguish colocalized from noncolocalized molecules. The DoC threshold is a user-specified variable that can be changed to suit the experimental conditions. The choice of the DoC threshold value can be confirmed through simulations that match the experimental conditions and specific biological scenario. We can therefore characterize the extent of protein A colocalization with protein B and the extent of protein B colocalization with protein A separately.

**Clus-DoC: a combined cluster detection and colocalization analysis**

Protein clustering is known to be a common strategy for cells to spatially concentrate signaling proteins for efficient signal transduction. To assess the influence of cluster properties (such as cluster geometry or cluster density) on signaling, we combined the DoC analysis with the DBSCAN cluster detection analysis in a
Traditionally, biological approaches have viewed on/off processes as occurring at the level of a whole cell. However, it is becoming increasingly obvious that cellular processes are localized to specific sites within a cell and that signaling can be turned on or off differentially at different locations. The potential of Clus-DoC lies in its ability to address the functional state of individual structures (e.g., clusters, fibers) to determine which spatial parameters drive a particular structure to be active.

**Application to T-cell signaling**

The complexity of signal transduction is well illustrated by the example of T-cell signaling, in which even the first step in signal initiation remains controversial. Phosphorylation of the T-cell receptor (TCR) complex by the kinase Lck is one of the earliest detectable biochemical events in T-cell signaling, but the precise mechanism by which extracellular ligation of the TCR results in intracellular phosphorylation remains contested. Several SMLM studies reported that at least a proportion of TCRs were organized into small clusters that were 30–300 nm in diameter, termed “nanoclusters,” that were remodelled upon T-cell activation (Lillemeier et al., 2010; Sherman et al., 2011). However, little is known about the signaling activity of these clusters and whether they are equivalent in terms of function.

Using two-color SMLM, we simultaneously imaged the distribution of total TCR and phosphorylated TCR (pTCR) in activated T-cells. We imaged CD3ζ (a marker of the TCR complex) fused to the photoswitchable fluorescent protein PSCFP2 and pTCR using an antibody conjugated to Alexa Fluor 647 (Figure 6A). We also imaged TCR-pTCR data with Clus-DoC. We detected TCR clusters and, as expected, found that the TCR showed similar spatial patterns in both experiments, as quantified in Supplementary Figure S1. However, visual inspection of the maps suggested that TCR colocalized much more with pTCR than with CD45 (Figure 6, B and C). Using the DoC analysis, we next quantified the degree of colocalization of pTCR and CD45 with TCR and found that CD45 molecules had a much lower average DoC score than with pTCR (Figure 6D). In fact, pTCR molecules colocalized quite highly with TCR (as expected), whereas the average DoC score for CD45 molecules was negative, indicating no colocalization and perhaps a tendency toward segregation. Similarly, by using a threshold of 0.4 to distinguish colocalized from noncolocalized molecules, we found that the percentage of CD45 molecules that were colocalized with TCR was significantly lower than the percentage of pTCR molecules that colocalized with TCR (Figure 6E). Using these two colocalization parameters, we are able to distinguish between proteins that do and do not colocalize with a given protein cluster or structure.

We then proceeded to analyze the TCR-pTCR data with Clus-DoC to separate TCR clusters into two populations: clusters that were colocalized with pTCR (i.e., phosphorylated clusters) and the phosphatase CD45 (Figure 6, B and C). Using DBSCAN, we detected TCR clusters and, as expected, found that the TCR showed similar spatial patterns in both experiments, as quantified in Supplementary Figure S1. However, visual inspection of the maps suggested that TCR colocalized much more with pTCR than with CD45 (Figure 6, B and C). Using the DoC analysis, we next quantified the degree of colocalization of pTCR and CD45 with TCR and found that CD45 molecules had a much lower average DoC score than with pTCR (Figure 6D). In fact, pTCR molecules colocalized quite highly with TCR (as expected), whereas the average DoC score for CD45 molecules was negative, indicating no colocalization and perhaps a tendency toward segregation. Similarly, by using a threshold of 0.4 to distinguish colocalized from noncolocalized molecules, we found that the percentage of CD45 molecules that were colocalized with TCR was significantly lower than the percentage of pTCR molecules that colocalized with TCR (Figure 6E). Using these two colocalization parameters, we are able to distinguish between proteins that do and do not colocalize with a given protein cluster or structure.

We then proceeded to analyze the TCR-pTCR data with Clus-DoC to separate TCR clusters into two populations: clusters that were colocalized with pTCR (i.e., phosphorylated clusters) and
clusters that were not colocalized with pTCR. We then compared the morphology of these two subpopulations of TCR clusters and found that phosphorylated TCR clusters were larger and less circular than nonphosphorylated clusters (Figure 6, F and G), suggesting that phosphorylation and thus signal initiation may be dependent on the geometry of the clusters.

With Clus-DoC, we can also isolate nonclustered molecules and assess their DoC scores (Supplemental Figure S2 for TCR and CD45). We found that for TCR and CD45, nonclustered molecules were randomly distributed relative to each other (DoC scores peaking around 0; Supplemental Figure S2A). We also assessed the distribution of total or colocalized molecules between colocalized clusters, noncolocalized clusters, and outside clusters and found that the majority of TCR molecules were present in noncolocalized clusters (Supplemental Figure S2B), whereas colocalized molecules were as abundant in colocalized clusters as outside clusters (Supplemental Figure S2C). Clus-DoC also quantifies the average DoC score between these different populations (Supplemental Figure S2D).

Application to focal adhesion complexes

Focal adhesions are integrin-based structures that form at the cell-substrate interface, where they serve to connect the cell membrane to the extracellular matrix (ECM). Focal adhesions function as signaling hubs and traction points to help cells to sense or anchor to the ECM. Tyrosine phosphorylation of the proteins that make up these adhesions, such as paxillin, regulates adhesion dynamics from formation to maturation and turnover (Nobes and Hall, 1995; Wozniak et al., 2004). The phosphorylation of paxillin appears as a major switch in the assembly and turnover of focal adhesions. SMLM studies revealed that phosphorylated paxillin (p-paxillin) is abundant in the adhesion structures located at the cell periphery but is mainly absent from adhesions in the cell center (Zaidel-Bar et al., 2007). However, the distribution of p-paxillin nanoclusters within the adhesion structures remains unknown.

To examine the distribution of paxillin and p-paxillin within focal adhesion complexes, we imaged paxillin fused to the photoswitchable fluorescent protein mEos2 and p-paxillin using an antibody conjugated to Alexa Fluor 647 in fibroblasts on fibronectin (Figure 7, A–C). As expected, paxillin and p-paxillin were highly colocalized within focal adhesions, with 53 and 62% of paxillin and p-paxillin being colocalized with the other, respectively (Figure 7, D and E). We also quantified the distribution of DoC scores for nonclustered molecules (Supplemental Figure S3). The average DoC scores and the percentage of colocalized molecules were lower for nonclustered molecules than for total molecules (Supplemental Figure S3, A and B, and Figure 7, D and E), suggesting a preference for paxillin to be phosphorylated when in clusters. We then divided paxillin clusters into those that were or were not colocalized with p-paxillin (Figure 7F) and found that colocalized clusters were denser.

With the application of Clus-DoC to two biological systems, we showed that Clus-DoC provides a more in-depth characterization of signaling complexes and protein colocalization than DBSCAN or DoC alone would allow.

Conclusion

We demonstrated that a combined cluster detection and colocalization analysis method (Clus-DoC) is a powerful tool to extract quantitative and biologically relevant information from SMLM data sets. Clus-DoC provides several advantages over existing analysis methods. It is fast (6-min processing time on a standard desktop PC for a data set of 10 regions corresponding to ~500,000 localizations) and retains single-molecule information rather than providing ensemble measurements. Clus-DoC provides outputs at the level of molecules (e.g., distribution of DoC scores), at the level of clusters (e.g., size, density, morphology), and at the level of cells (e.g., comparisons across different conditions). In addition to being quantitative, Clus-DoC also generates maps for visualization of the data. Of importance, it has the potential to subclassify clusters according to different parameters (colocalization status, size, density, number of molecules, etc.), which provides deeper insight into the links between protein organization and function. By providing a user-friendly GUI to run Clus-DoC, we make this method easily accessible to any biologist with access to an SMLM-capable instrument. In addition, analysis parameters such as the region selection, region size, and cluster detection variables are controllable by the user. In the future, Clus-DoC will be extended to the analysis of multicolor SMLM data with three or more channels or with three-dimensional information. The software project, with new features and issue fixes as added, is open-source and continually maintained in the authors’ Git repository.

We believe that Clus-DoC can be applied to address a variety of biological questions in different fields. Because it quantifies the molecular structure and the degree of colocalization at the same time, this approach is well suited to characterizing the molecular composition and organization of higher-order molecular complexes or comparing the affinity of different proteins for a common structure (such as actin-binding proteins; Malkusch et al., 2012). Furthermore,
FIGURE 5: GUI for Clus-DoC analysis. Image of the GUI for the Clus-DoC analysis, which can be used to load data sets and then run a global clustering analysis (Ripley’s K), cluster detection (DBSCAN), or Clus-DoC analysis. Here the two-color data of one cell are displayed (in green and red), and one ROI is selected (blue square). The number of points from each channel within that ROI and the dimensions of the ROI (in nanometers) are displayed at the top.

Clus-DoC facilitates the identification of active signaling complexes (as shown here for TCR clusters) to probe the functional aspects of a molecular system. Similar mechanisms might be at play across different types of immune cells, such as natural killer cells, for which nanoclusters of inhibitory receptors were found to become smaller and denser upon cell activation (Pageon et al., 2013). This type of approach would also be beneficial for comparing the association of specific proteins with vesicles of different sizes (Caetano et al., 2015) or studying the mechanism of viral budding from infected cells (Malkusch et al., 2013).

In conclusion, we have developed a combined cluster detection and colocalization analysis (Clus-DoC) to probe the links between protein clustering and...
Sample preparation for imaging focal adhesions
Mouse embryonic fibroblasts (MEFs) were cultured in high-glucose DMEM (Life Technologies) supplemented with 10% FCS at 37°C and 5% CO₂. MEF cells were transfected by Lipofectamine 3000 (Invitrogen) according to the manufacturer’s instructions to express paxillin-mEos2. Transfected cells were plated onto 25 μg/ml fibronectin-coated clean cover glasses and incubated for 3 h at 37°C. The cells were then fixed for 15 min in freshly made warm 4% paraformaldehyde. For immunostaining against phosphorylated paxillin, cells were permeabilized for 10 min in 0.5% Triton X-100, followed by blocking with 5% FCS for an additional 40 min. Cells were then labeled with primary antibody from rabbit against mouse paxillin phosphorylated at Tyr-118 (Life Technologies), followed by staining with Alexa Fluor 647–conjugated secondary antibody specific to the rabbit F(ab’)2 fragment (111-606-047; Jackson ImmunoResearch). Cells were washed in 10 mM cysteamine in PBS three times, for 5 min each time, to decrease the floating background for dSTORM imaging.

SMLM imaging and dual-color image acquisition
SMLM image sequences were acquired on a total internal reflection fluorescence (TIRF) microscope (ELYRA; Zeiss, Jena, Germany) with a 100x oil-immersion objective (numerical aperture 1.46) and 1.6x Optivar. Photoconversion of PSCFP2 or mEos2 was achieved with 8 μW of 405-nm laser radiation, and the green-converted PSCFP2/red-converted mEos2 was imaged with 18 mW of 488/561-nm light, respectively. For Alexa Fluor 647, 15 mW of 633-nm laser illumination was used for imaging. An oxygen-scavenging PBS-based buffer (containing 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, 25 mM glucose, 5% glycerol, 0.05 mg/ml glucose oxidase, and 0.025 mg/ml horseradish peroxidase, supplemented with 50 mM cysteamine; all from Sigma-Aldrich, St. Louis, MO) was used for dSTORM imaging. For each cell, 20,000 images were acquired with a cooled, electron-multiplying charge-coupled device camera (iXon DU-897; Andor, Belfast, UK) using an exposure time of 30 ms and a pixel size of 100 nm in the image space. Raw fluorescence intensity images were analyzed with the software Zen 2011 SP3 (Zeiss Microimaging), generating reconstructed SMLM images, as well as tables containing the x, y-coordinates of each molecule detected during the acquisition.

Data processing
Data in the form of an ASCII text file containing spatial coordinate and channel identification information for each detected localization are loaded into the GUI application. An additional file containing user-defined ROIs is also loaded. These ROIs define subregions within the cell area within which each analysis will be applied. ROI area was typically 4000 nm × 4000 nm, and care was taken that the ROI border fell entirely within the cell boundary.

An average clustering value across an entire ROI is measured on a single user-given square ROI, with the start, end, and step size of r also user defined.

FIGURE 7: Clus-DoC analysis applied to focal adhesions. (A) Two-color TIRF image of paxillin (green) and phosphorylated paxillin (p-paxillin; red) in MEF cells on fibronectin. Scale bar, 10 μm. (B) SMLM data of paxillin (green) and p-paxillin (red) from 6 μm × 6 μm region highlighted in white in A. (C) Zoom of 1 μm × 1 μm region highlighted in black in B. (D) Frequency distributions of DoC scores of all molecules for paxillin (top) and p-paxillin (bottom), with mean DoC scores indicated. (E) Percentage of p-paxillin molecules colocalized with p-paxillin and percentage of p-paxillin molecules colocalized with paxillin. (F) Relative density of clusters. *p ≤ 0.05 (paired t test).

MATERIALS AND METHODS
Sample preparation for activated T-cells
Jurkat E6.1 T-cells (American Type Culture Collection, Manassas, VA) were maintained in RPMI 1640 (Life Technologies, Carlsbad, CA) supplemented with 10% fetal calf serum (FCS), 2 mM l-glutamine, and 1 mM penicillin and streptomycin (all from Invitrogen, Carlsbad, CA) and transfected by electroporation (NEON; Invitrogen) to express CD3ζ fused to PSCFP2. Clean glass coverslips were coated with anti-CD3ε (16-0037; eBioscience, San Diego, CA) and anti-CD28 (16-0289; eBioscience, San Diego, CA) antibodies for at least 1 h at 37°C. Cells were activated on antibody-coated coverslips for 10 min at 37°C and then fixed with 4% paraformaldehyde (vol/vol) in phosphate-buffered saline (PBS) for 20 min at room temperature. Cells were permeabilized with 0.1% (vol/vol) Triton X-100 for 4 min at room temperature, blocked in 5% bovine serum albumin (wt/vol) in PBS, and then labeled with primary antibody against human CD3ζ phosphorylated at Tyr-142 directly conjugated to Alexa Fluor 647 (S55489; BD Biosciences, San Jose, CA) overnight at 4°C or against CD45 (ab10559; Abcam, Cambridge, UK) for 1 h at room temperature, followed by staining with Alexa Fluor 647–conjugated goat antibody specific to the rabbit F(ab’)2 fragment (111-606-047; Jackson ImmunoResearch).

Sample preparation for imaging focal adhesions
Mouse embryonic fibroblasts (MEFs) were cultured in high-glucose DMEM (Life Technologies) supplemented with 10% FCS at 37°C and 5% CO₂. MEF cells were transfected by Lipofectamine 3000 (Invitrogen) according to the manufacturer’s instructions to express paxillin-mEos2. Transfected cells were plated onto 25 μg/ml fibronectin-coated clean cover glasses and incubated for 3 h at 37°C. The cells were then fixed for 15 min in freshly made warm 4% paraformaldehyde. For immunostaining against phosphorylated paxillin, cells were permeabilized for 10 min in 0.5% Triton X-100, followed by blocking with 5% FCS for an additional 40 min. Cells were then labeled with primary antibody from rabbit against mouse paxillin phosphorylated at Tyr-118 (Life Technologies), followed by staining with Alexa Fluor 647–coupled secondary antibody specific to the rabbit F(ab’)2 fragment (111-606-047; Jackson ImmunoResearch). Cells were washed in 10 mM cysteamine in PBS three times, for 5 min each time, to decrease the floating background for dSTORM imaging.
An additional parameter is supported to randomly subsample the points within a ROI to a user-defined maximum value. This improves processing speed for dense point fields and allows clustering behaviors between ROIs to be compared at identical point densities.

Points are segmented into clusters using the DBSCAN algorithm. Before this clustering, the user has the option of segmenting noise localizations from the ROI. Here a “noise” point is defined as a point whose $L(r) - r$ value at a user-given radius (typically 50 nm) is below that of the value expected for a spatially random distribution of that density of points. Given user-specified values for the search radius and minimum number of points per cluster, a custom-written DBSCAN implementation (written in C++ and compiled using MEX) segments the points within each ROI into clusters or as the outliers or noise component. Data are extracted on a per-cluster basis, including number of points per cluster, cluster border, area, and circularity based on the cluster contour, cluster density, and, where included with DoC scores, the number of points within the cluster above and below the DoC threshold. Those clusters containing at least one point with a DoC score above the threshold are further subsegmented into points above and below the threshold, with the number, area, and density measures for these sub-populations extracted.

The cluster contour is defined by generating a 2D image of the histogram of localization points within the cluster. This 2D histogram has bin widths of 1 nm in each dimension. This histogram rendering is smoothed by a user-defined value (typically 15 nm). The value of this smoothed rendering at each localization point is evaluated and a threshold taken at the lowest value associated with a data point. The outline of this thresholded region is taken as the contour of the cluster for measuring cluster circularity, diameter, and area and for counting the number of molecules per cluster.

A DoC score is calculated for each point within a ROI for a range of discrete search radii from 0 to a maximum value. Both the step size and maximum radius ($r_{\text{Max}}$) are user specified, typically 10 and 500 nm, respectively. The search for points within this radius is accelerated by a MEX function performing a k-dimensional tree search algorithm. The density of points at each search radius for channel $i$ is correlated against the values from channel $j$ using a pairwise linear correlation of Spearman’s $\rho$ coefficient. This correlation, $S_{\rho ij}$, is converted into a DoC score by first calculating the cross-channel nearest-neighbor distance at each point in $i$ to $j$, $N_{ij}$, by the expression

$$\text{DoC}_{ij} = S_{\rho ij} \exp \left( -\frac{N_{ij}}{r_{\text{Max}}} \right)$$

This is repeated for both channels, performing the correlation and nearest-neighbor searches in the opposite direction.

Once calculated, this DoC score can be compared with a user-defined threshold (typically 0.4) to segment points that are colocalized, clusters that are or are not colocalized (given a minimum number of points per cluster with a DoC score above the threshold), or to segment points within a cluster based on DoC score.

**Generation of GUI**

A MATLAB GUI was written to provide a graphical front end to the underlying clustering and DoC functions. This interface allows users to load and visualize SMLM dual-color data sets, executes the described functions, and exports plots and tabulated results. Additional functionalities, such as further data analysis, the ability to specify ROIs and imported binary masks and to do batch processing, are supported.

Codes are written for 64-bit MATLAB R2014b or above under a Windows operating system. All processing completes within minutes on a standard desktop PC for a single SMLM data set of ~10^6 points.

The latest version of the source codes for the underlying functions and GUI application, complete with new functionalities and bug fixes, are available at the authors’ Git repository (https://github.com/PRNicovici/ClusDoC).

**Statistical analysis**

All statistical analysis was performed using GraphPad software (Prism, San Diego, CA). Statistical significance between data sets was determined by performing two-tailed Student’s t tests. Graphs show mean values, and error bars represent the SEM. In statistical analysis, $p > 0.05$ is indicated as not significant (n.s.), whereas statistically significant values are indicated as follows: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

**REFERENCES**


