Investigating focal adhesion substructures by localization microscopy and expectation maximization of a Gaussian mixture

H. Deschout[1], I. Platzman[2], D. Sage[3], L. Feletti[1], J. P. Spatz[2] and A. Radenovic[1]

[1]*Laboratory of Nanoscale Biology, Institute of Bioengineering, School of Engineering, EPFL, Lausanne, Switzerland*

[2]*Max-Planck-Institute for Medical Research, Dept. of Cellular Biophysics, University of Heidelberg, Dept. of Biophysical Chemistry, Heidelberg, Germany*

[3]*Biomedical Imaging Group, School of Engineering, EPFL, Lausanne, Switzerland*

Abstract

Cells rely on focal adhesions (FAs) to carry out a variety of important tasks, including motion, environmental sensing, and adhesion to the extracellular matrix. Although attaining a fundamental characterization of FAs is a compelling goal, their extensive complexity and small size, which can be below the diffraction limit, have hindered a full understanding. In this study we have used single-molecule localization microscopy (SMLM) to investigate integrin β3 and paxillin in rat embryonic fibroblasts growing on two different extracellular matrix-representing substrates (i.e. fibronectin-coated substrates and specifically bio-functionalized nano-patterned substrates). To quantify the substructure of FAs, we developed a clustering method based on expectation maximization of a Gaussian mixture that accounts for localization uncertainty and background. Analysis of our SMLM data indicates that the structures within FAs, characterized as a Gaussian mixture, typically have areas between 0.01 and 1 μm^2, contain 10 to 100 localizations, and can exhibit substantial eccentricity. Our approach based on SMLM opens new avenues for studying structural and functional biology of molecular assemblies that display substantial varieties in size, shape, and density.

1. Introduction

Focal adhesions (FAs) are subcellular macromolecular assemblies consisting of dynamic protein complexes that are localized near the cell membrane. FAs affect nearly all aspects of a cell's life, including, but not limited to, adhesion, directional migration, cell proliferation, differentiation, survival, and gene expression (1). Despite having been studied for several decades, the inner architecture of FAs is still not completely understood. In part, this is due to the limitations of conventional fluorescence microscopy for FA analysis. FAs are molecularly diverse structures, containing a large number of proteins (2). Therefore, their investigation requires imaging techniques that offer sufficient multiplexing capabilities (3). Moreover, FAs have a size that is typically in the order of a micron or less, and therefore their internal spatio-temporal organization is not fully resolvable with conventional microscopy.

During the last decade, several super-resolution microscopy techniques have been employed to image FAs (4-9). An important insight from these studies was that FAs are not homogeneous spatial structures. Initially, photo-activated localization microscopy (PALM) was used to reveal that FAs can consist of patches of proteins with submicron dimensions (4, 9). Later on, Bayesian localization microscopy and structured illumination microscopy showed that many FAs exhibit discontinuous elongated (or fiber-like) substructures (5, 6). Moreover, single-particle tracking demonstrated that proteins can diffuse within FAs (7, 8), which again suggests that they have an internal spatial organization. However, dedicated tools that allow a systematic quantitative analysis of the FA substructure are still lacking.

For quantitative analysis of the internal spatial organization of FAs, single-molecule localization microscopy (SMLM) can potentially be implemented (10, 11). SMLM data consist of the localizations of individual photo-activatable or photo-switchable fluorescent molecules. Therefore, a variety of methods have been developed to identify and characterize clusters of such localizations (12, 13). These methods are often applied to investigate clusters of receptors in the cell membrane. Such clusters are usually radially symmetric, spatially well separated, and homogeneous in size and density. FA substructures, on the other hand, cannot be characterized as simply. Indeed, adhesions structures can vary from sub-diffraction entities composed of a couple of different proteins (e.g. focal complexes or nascent adhesions) to assemblies of many proteins measuring several microns (e.g. FAs) (14). Moreover, FA subunits are densely packed, since they cannot be resolved using a conventional microscope. Finally, FAs usually have an elongated shape, and the same is possibly true for their subcomponents. Therefore, it is not clear if established SMLM clustering methods are suitable for the identification of FA substructures.

In this study we have designed, to the best of our knowledge, a novel approach to investigate the FA substructure. We used expectation maximization of a Gaussian mixture (EMGM) (15) to interpret SMLM data in terms of spatial probability distributions. EMGM allows to quantify the properties of closely packed localization patterns that exhibit substantial varieties in size, density, and shape, and is therefore well suited for studying the inner architecture of FAs. Importantly, we improved the classical EMGM framework to account for localization uncertainties and the presence of a localization background, both being ubiquitous in SMLM data.

The other goal of this study was to quantify the properties of the subunits of which FAs are composed. For this purpose we used PALM, an implementation of SMLM that is popular for imaging FAs (4, 9, 16-18), since it makes use of photo-activatable fluorescent proteins that can be genetically expressed. More in particular, we used PALM to image integrin β3 and paxillin in fixed rat embryonic fibroblasts (REFs), a well-known cell line for FA investigation. Cell experiments were performed using fibronectin-coated substrates and specifically bio-functionalized nano-patterned substrates, on which ordered patterns of nanoscale adhesive spots were provided (19, 20). Such nano-patterned substrates have already been used to indirectly probe the behavior of FAs on the nanoscale (21). In this way, the spatial organization of integrin binding sites is precisely controlled, ensuring that the observed substructures are innate to FAs. Application of our improved version of EMGM on the PALM data allowed us to determine that FAs are composed of structures with areas between 0.01 and 1 μm^2, containing 10 to 100 localizations, and exhibiting substantial eccentricities.

2. Materials and Methods

2.1 Microscope

PALM imaging was carried out on a custom built microscope (22, 23). A 50 mW 405 nm laser (Cube, Coherent), a 100 mW 488 nm laser (Sapphire, Coherent), and a 100 mW 561 nm laser (Excelsior, Spectra Physics) were used for excitation/activation. The three lasers were focused into the back focal plane of the objective mounted on an inverted optical microscope (IX71, Olympus). We used a 100× objective (UApo N 100×, Olympus) with a numerical aperture of 1.49 configured for TIRF. A dichroic mirror (493/574 nm BrightLine, Semrock) and an emission filter (405/488/568 nm StopLine, Semrock) were used to separate fluorescence and illumination light. The fluorescence light was detected by an EMCCD camera (iXon DU-897, Andor). An adaptive optics system (Micao 3D-SR, Imagine Optics) and an optical system (DV2, Photometrics) equipped with a dichroic mirror (T565lpxr, Chroma) were placed in front of the EMCCD camera.

2.2 Imaging procedure

Cells were imaged in PBS at room temperature. Prior to imaging, 100 nm gold fiducial markers (C-AU-0.100, Corpuscular) were added to the sample for lateral drift monitoring. Axial drift correction was ensured by a nanometer positioning stage (Nano-Drive, Mad City Labs) driven by an optical feedback system (22). Excitation of the mEos2 was done at 488 nm or 561 nm with ~10 mW power (as measured in the back focal plane of the objective). The mEos2 was activated at 405 nm with ~2 mW power. The gain of the EMCCD camera was set to 100 and the exposure time to 50 ms. For each experiment 10,000 camera frames were recorded.

2.3 Substrate preparation

Quasi-hexagonal patterns of AuNPs were prepared on 25 mm diameter microscope cover slips (#1.5 Micro Coverglass, Electron Microscopy Sciences) by means of block-copolymer micelle nanolithography (BCML) as previously described (19, 20, 24) (Supporting Material). Fibronectin-coated cover slips were prepared by first cleaning with an oxygen plasma and then incubating with PBS containing 50 μg/ml fibronectin (Bovine Plasma Fibronectin, Invitrogen) for 30 minutes at 37 °C. To remove the excess of fibronectin, the cover slip was washed with PBS before seeding the cells.

2.4 Cell culture and fixation

The REFs (CRL-1213, ATCC) were grown in DMEM supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin, 1% non-essential amino acids and 1% glutamine, at 37 °C with 5% CO_2. The cells were transfected by electroporation (Neon Transfection System, Invitrogen), which was performed on ~10^6 cells using 1 pulse of 1350 V lasting for 35 ms. The amount of DNA used for the transfection was 4 μg for both the mEos2-paxillin-22 vector and the mEos2-Integrin-β3-N-18 vector. Around 2.10^5 transfected cells were seeded on individual cover slips and grown in cell culture medium without penicillin-streptomycin, at 37 °C with 5% CO_2. The cells were washed with PBS around 20h after transfection (Fig. S1 in the Supporting Material), and then incubated in PBS with 2.5%

paraformaldehyde at 37 °C for 10 minutes. After removing the fixative, the cells were again washed with PBS, and the cover slip was placed into a custom made holder.

2.5 PALM data analysis

The recorded images were analyzed by a custom written algorithm (Matlab, The Mathworks) that was adapted from a previously published algorithm (4, 23). First, peaks were identified in each camera frame by filtering and applying an intensity threshold. Only peaks with an intensity at least 4 times the background were considered to be emitters. Subsequently, each emitter was localized by maximum likelihood estimation of a 2D Gaussian distribution (25). When peaks appeared during several consecutive frames within the same pixel, they were assumed to correspond to the same emitter, and the emitter images in these frames were summed prior to maximum likelihood estimation. Drift was corrected in each frame by subtracting the average position of the fiducial markers from the positions of the emitters in that frame. The localization uncertainty for each emitter was obtained from the Cramér-Rao lower bound of the maximum likelihood procedure (26). PALM images were generated by plotting a 2D Gaussian centered on each fitted position with a standard deviation equal to the corresponding localization uncertainty. Only positions with a localization uncertainty below 40 nm were used.

2.6 EMGM procedure

The EMGM procedure (Supporting Material) was implemented in Matlab (The Mathworks). The initial values of the parameters that describe a mixture consisting of K components were estimated by deleting a component from the previously estimated mixture consisting of K-1 components and adding two new components that were generated from the deleted one (Supporting Material). Additionally, one new Gaussian component was generated from the background component of the previously estimated mixture. This was done 3 times for each of the original K-1 Gaussian components and the background component, resulting in a total of $3K$ initializations. In the case of K=1, the initialization was done randomly 3 times. The procedure is stopped when the null hypothesis that the previously estimated K-1 component mixture is the correct one was fulfilled (Supporting Material). For this purpose, we simulated the distribution of likelihood increments when comparing the K-1 and K component models under the null hypothesis. This distribution is obtained by simulating 100 datasets assuming the K-1 solution, and applying EMGM on each dataset, for both K-1 and K mixture components. If the real likelihood increment had a p-value below 0.01 under the null hypothesis, it was assumed that the K-1 component solution is the correct one. Prior to analysis, the PALM data was split into overlapping 2×2 µm areas, and the EMGM analysis was performed on each area separately (Fig. S2 in the Supporting Material). Afterwards, identical mixture components in different EMGM results were combined according to a criterion based on the correlation between their posterior probabilities (Supporting Material).

3. Results

3.1 Expectation Maximization of a Gaussian Mixture (EMGM)

FAs display a substantial variety in size, shape, and density, and their substructure potentially as well. Quantifying the properties of the FA substructure with SMLM clustering methods is therefore challenging. Clusters in SMLM data are often characterized using the pair correlation function (27) or Ripley's $K(r)$ or $L(r)$ function (28). These functions describe the density around a certain point as a function of the distance r from that point. As an illustration, we used PALM to image integrin β3 in a REF cell (Fig. 1A). We used Ripley's $L(r)$-r function (29) to analyze a subset of the data (Fig. 1, B-D). This function shows a peak around 0.2 μm, indicating that the degree of clustering is highest on this length scale. However, it is difficult to interpret this result in terms of FA substructure properties, especially considering the heterogeneity in size and shape of the FAs themselves.

Such difficulties can be avoided by clustering methods that identify individual clusters based on criteria related to the local density of localizations, such as the nearest neighbor method (30) or density-based spatial clustering of applications with noise (DBSCAN) (31). We applied DBSCAN (32) to the same subset of the PALM data mentioned above (Fig. 1E). One value for the DBSCAN search radius identified several substructures in the FA, while a larger value did not. However, the large search radius identified two clusters that were considered to be background by the small search radius. It is clear that DBSCAN can handle the heterogeneity in size and shape of FAs, but identification of FA substructures largely depends on the values used for parameters that are related to a localization density threshold. Such a threshold is challenging to define, since FA substructures exhibit a variety of localization densities and can be closely packed (Fig. 1, A and B).

The difficulties related to established SMLM clustering methods prompted us to develop an approach based on EMGM (15). The main assumption of EMGM is that FAs can be modeled by a mixture of bivariate Gaussian probability distributions (Supporting Material). After choosing initial values for the parameters of each Gaussian component, the posterior probability that a certain localization was generated from a certain Gaussian component is evaluated (i.e. the expectation step). The Gaussian component parameters are then re-estimated using the new posterior probabilities (i.e. the maximization step) and the likelihood of the updated Gaussian mixture is calculated and checked for convergence.

In order to apply EMGM on SMLM data, we used a "greedy learning" approach (33) to initialize the parameters of the Gaussian components, and a model selection procedure based on hypothesis testing (34) to determine the number of components in the mixture (Supporting Material). However, the specific nature of SMLM data poses some additional challenges for EMGM. One problem is that not necessarily all localizations are part of the structure of interest, but can instead belong to a background. In case of a simple uniform background, the EMGM algorithm can readily be adjusted (Supporting Material). Moreover, the localizations in SMLM data contain measurement uncertainties (35). This localization uncertainty can be described by a spatial probability distribution that is usually modeled as a Gaussian. The EMGM can therefore be adapted by convolving the probability distributions that describe the mixture and the localization uncertainties (Supporting Material).

3.2 Evaluation of EMGM on simulations

The performance of the EMGM algorithm adapted for SMLM data was evaluated and validated by applying it to simulated data. We simulated mixtures consisting of K closely spaced Gaussian components described by identical spatial probability distributions (i.e. 2D symmetric Gaussians with standard deviation $\sigma_x = \sigma_y = 20$ nm) and containing an identical number of positions (i.e. 100) (Fig. 2A, and Supporting Material). Such components have similar characteristics as nascent adhesions or, more speculatively, as the substructure of larger FAs.

First, we verified the performance of our proposed initialization scheme and model selection procedure. The results show that the simulated mixtures are correctly identified, provided K is smaller than 10 (Fig. 2B, and Fig. S3 in the Supporting Material). Interestingly, simulations of random Gaussian mixtures that are closer to the experimental reality confirm this finding (Fig. S4 in the Supporting Material). We used $3K$ initializations for a mixture with K components (Supporting Material). Increasing the number of initializations does not substantially improve the EMGM performance (Fig. S5 in the Supporting Material).

Next, we simulated the effect of a uniform localization background density bg and a localization uncertainty s. The results indicate that the adapted EMGM correctly predicts $\sigma_{x,y}$ for values of bg up to 25,000 #/μm^2 when $K = 4$ (Fig. 2C, and Fig. S6 in the Supporting Material). For larger values of K, the method performs well for bg values up to 10,000 #/μm^2 (Fig. S7 in the Supporting Material). Our EMGM approach also captures the effect of the apparent increase in $\sigma_{x,y}$ due to localization uncertainties for values of s up to 30 nm (Fig. 2D, Fig. S8 in the Supporting Material). Unlike for the localization background, this limit does not seem to depend on the number of components (Fig. S9 in the Supporting Material). Note that the largest values of s and bg included in these simulations are typically not encountered in "good quality" SMLM data.

Since one cannot assume that the substructures of FAs are radially symmetric, the component shape should be accounted for by the EMGM algorithm. We simulated mixture components with decreasing σ_x and simultaneously increasing σ_y (Supporting Material). The results (Fig. 2E) clearly show that the algorithm correctly predicts the changing eccentricity σ_x/σ_y. The adapted EMGM should also be able to distinguish closely spaced substructures inside FAs. Towards this end, we simulated Gaussian mixtures with a decreasing spacing $d_{x,y}$ between the component centers (Fig. 2F, and Supporting Material). The adapted EMGM performs well when $d_{x,y}$ is larger than 70 nm, or more generally when the relative spacing $d_{x,y}/\sigma_{x,y}$ is larger than 4 (Fig. S10 in the Supporting Material). A smaller $d_{x,y}$ (or $d_{x,y}/\sigma_{x,y}$) results in a significant overlap in the spatial probability distribution of two adjacent components.

It should be noted that the results (Fig. 2) depend on the number of localizations that are contained by the components. The sensitivity of the EMGM algorithm strongly decreases for components containing around 10 localizations (Fig. S11 in the Supporting Material).

3.3 Application of EMGM on experimental data

To demonstrate the application of our EMGM algorithm, we made use of the SMLM data of a REF cell expressing mEos2-labelled integrin β3 (Fig. 1, B and C). Similar to DBSCAN applied with the small search radius (Fig. 1E), EMGM also finds several FA substructures (Fig. 1F). Moreover, EMGM identifies two structures on the right as well, as indicated by the DBSCAN result using the large search radius (Fig. 1E).

We next proceeded to apply the EMGM algorithm on the whole PALM dataset (Fig. 1A). Since the simulation results (Fig. 2B) indicate that our algorithm works best for a small number of components, we reduce their number by applying a scanning procedure, consisting of splitting the original field of view into smaller overlapping areas, and by subsequently applying EMGM to each of these areas (Fig. S2). The size of these areas has to be chosen carefully, as clipping of mixture components should be avoided, while ensuring that only a few are included. Afterwards, the results are combined, by merging identical Gaussian components in overlapping regions based on the correlation between their posterior probabilities, while excluding Gaussian components that belong to structures that were clipped during the splitting procedure (Supporting Material).

EMGM characterizes FA substructures in terms of bivariate Gaussian probability distributions. The properties of such a distribution can be translated into more intuitive properties using the error ellipse, i.e. the line that describes a constant probability density. The major axis a and the minor axis b of an ellipse define its area and shape (Fig. S12 in the Supporting Material). We therefore describe the FA substructure shape by the eccentricity b/a (similar to the definition above). To calculate the area, we choose the 2σ error ellipse, corresponding to twice the standard deviation of the Gaussian distribution. This error ellipse defines the area in which there is a probability to find ~95% of all localizations belonging to the mixture component. We pooled the area and eccentricity values of all identified components in our PALM data set (Fig. 1G). Most components have an area below 0.5 μm^2 with a peak around 0.1 μm^2, and many exhibit some degree of eccentricity, with most values lower than 0.8. The EMGM algorithm also returns the posterior probability of each localization belonging to a specific Gaussian distribution, which gives the total number of localizations of each FA substructure (Supporting Material). Making the simplifying assumption that the localizations are uniformly distributed within the 2σ error ellipse, this leads to a characteristic localization density. Most FA substructures have a localization density below 2000 #/μm^2, and contain less than 100 localizations (Fig. 1G).

3.4 Integrin and paxillin

Following the evaluation of the adapted EMGM, we applied our method to investigate the substructure of FAs in cells growing on often-used fibronectin-coated substrates. We used PALM to image fixed REF cells (n = 10) expressing paxillin or integrin β3 labelled with mEos2 (Fig. 3, A and B). To identify the FA substructure, we applied the adapted EMGM to each of these PALM datasets (Fig. 3C). As discussed above, the properties of individual mixture components, defined as bivariate Gaussians, can be described by three parameters: eccentricity, area, and number of localizations. We plotted these quantities as a function of each other, for both paxillin and integrin β3 (Fig. 3, D-F). Most mixture components contain between 10 and 100 localizations, and have an area between 0.01 and 1 μm^2 (Fig. 3D). The components with the lowest number of localizations are mainly located outside the FA structure (Fig. S13 in the Supporting Material). The paxillin case displays a slightly more pronounced tail towards components that contain more localizations (up to 1000 localizations). These components are situated within the FA structure (Fig. S14 in the Supporting Material), explaining the visual difference between paxillin and integrin β3 (Fig. 3B). When plotting the eccentricity as a function of the number of localizations (Fig. 3E), it is again apparent that the paxillin FA substructures can contain more localizations than the integrin ones. Furthermore, the mixture components in both cases appear to be eccentric, with most values below 0.7. The FA substructures containing less localizations appear

to be somewhat more eccentric, a tendency that is more apparent in the paxillin case. A similar observation can be made when plotting the eccentricity as a function of the area (Fig. 3F). The larger the FA substructure, the more eccentric it seems to be. Interestingly, both paxillin and integrin objects seem to have similar areas, with a peak around 0.1 μm^2.

3.5 Nano-patterned substrates

The FA substructure properties (Fig. 3) have been obtained from REF cells growing on fibronectin-coated substrates, which do not have well controlled binding sites (especially considering the presence of extracellular matrix proteins in the cell culture medium). It can therefore not be guaranteed that the observed FA substructure is innate, it might simply be reflecting how the integrin binding sites on the fibronectin-coated substrate are organized on the nanoscale level. Such difficulties in interpretation of the data can be avoided by making use of a substrate where the integrin binding site locations are precisely controlled. We have therefore made use of BCML to pattern substrates with a quasi-hexagonal grid of 8 nm diameter gold nanoparticles (AuNPs) (19, 20) (Supporting Material). The AuNPs are functionalized with cyclic arginyl-glycyl-aspartic acid (cRGD) peptides, using a flexible polyethylene glycol (PEG) spacer. The area between the AuNPs is passivated with a PEG layer, ensuring that integrins can only adhere to the AuNPs. This enables a more unambiguous interpretation of the observed FA substructure. We chose a 56 nm spacing between the AuNPs, which was shown to result in normal cell adhesion (19). Furthermore, we also tested a 119 nm spacing, which poses more challenges for adhering cells (20).

We again imaged fixed REFs (n = 10) expressing integrin β3 labelled with mEos2 (Fig. 4, A and B). Next, we applied the adapted EMGM to each of the PALM datasets, in order to investigate the FA substructure (Fig. 4C). We plotted the number of localizations as a function of the area, for both the 56 and 119 nm AuNP spacings (Fig. 4, E and F). The fibronectin case (Fig. 4D) was added for comparison. It is clear that the objects on the fibronectin-coated substrate can contain up to 100 localizations, while the localization numbers on the 56 nm spacing substrate are generally below that level (Fig. 4, D and E). Interestingly, the FA substructure areas are very similar between both types of substrates, mostly between 0.01 and 1 μm^2 (Fig. 4, E and F). The FA substructure observed on the nano-patterned substrates does not appear in contradiction with the results obtained from fibronectin-coated substrates.

3.6 Isolated and overlapping mixture components

The interpretation of the EMGM results can be complicated (Fig. 3C, and Fig. 4C). Especially inside dense and large structures, which visually appear to be FAs, one can observe several components that overlap, based on their 2σ error ellipses. The isolated mixture components, on the other hand, seem to correspond with smaller structures that could be nascent adhesions or focal complexes. We, therefore, performed a post-analysis step on EMGM results (Fig. 5A, and Supporting Material). We split the mixture components into two categories: the ones whose 1σ error ellipse overlaps with at least one other 1σ error ellipse, called the “overlapping” components, and the ones whose 1σ error ellipse does not overlap with another one, called the “isolated” components. A new object can be calculated from a set of overlapping components, giving rise to a third category, called the “merged” components (Fig. 5A, and Supporting Material). Application of this merging procedure on a previously

obtained EMGM result (Fig. 3C) shows that there are indeed several components that overlap (Fig. 5, B and C).

We applied the merging procedure on the EMGM results of REFs (n = 10) expressing integrin β3 labelled with mEos2, growing on fibronectin-coated (Fig. 3D) and 56 nm spacing nano-patterned (Fig. 4E) substrates. As expected, on both types of substrate, the merged objects tend to have a larger area (up to 1 μm^2) and contain more localizations (up to 1000 localizations) than the isolated and overlapping objects (Fig. 5D). The isolated components exhibit a similar behavior on both substrate types (Fig. 5E). Both cases exhibit FA substructures with an area between 0.01 and 0.1 μm^2, containing less than 100 localizations. The overlapping components are also not showing much difference between both substrate types, although the ones on the fibronectin-coated substrate can contain more localizations (Fig. 5F). Interestingly, the isolated and overlapping objects on the nano-patterned substrate also behave quite similarly (Fig. 5, E and F). The overlapping FA substructures are therefore not necessarily artifacts found by EMGM in a dense localization environment.

4. Discussion

We propose, to the best of our knowledge, a new way to explore the properties of unknown structures as observed by SMLM. Using EMGM, we interpret patterns in SMLM data as a mixture of bivariate Gaussians. This approach allows to describe densely packed structures that can display strong heterogeneities in size, shape, and density, and is therefore well suited for investigation of the substructure of FAs.

However, application of EMGM to SMLM data is not without challenges. The result can be influenced by the choice of the initial values for the mixture component properties, and the number of components needs to be chosen as well. We identified an initialization procedure and a selection criterion for the number of components that gives good results for mixtures consisting of a small number of components (e.g. less than 10 for our simulated data). To allow analysis of larger numbers of components, we used a scanning procedure that consists of splitting the SMLM data into smaller overlapping areas, and performing EMGM on each area separately. It is important to note that, unlike some SMLM clustering methods, the EMGM approach essentially does not depend on the choice of a free parameter (except for the area size of the scanning procedure).

The properties of SMLM data pose challenges to the classic EMGM algorithm. One complication is the localization uncertainty, which leads to an overestimation of the standard deviation of the Gaussian mixture components. An important contribution of this work is that we improved the EMGM approach to account for this effect. For “reasonable” localization uncertainties (e.g. below 30 nm for our simulated data) we found that the adapted EMGM worked well. We would like to point out that the effect of localization uncertainties is ignored by most existing SMLM clustering methods. Besides localization uncertainty, we also adjusted the EMGM algorithm to account for the presence of a uniform localization background. The method was found to perform excellently for any realistic level of background (e.g. up to 10,000 #/μm^2 for our simulated data).

To investigate the inner architecture of FAs, we performed SMLM imaging of FAs in fixed REF cells. We first explored the use of points accumulation in nanoscale topography (PAINT) (36) for imaging integrin β3 (Supporting Material). Our PAINT data suggests that that not all integrins are accessible for antibodies (Fig. S15 in the Supporting Material). To avoid antibody labeling problems, we therefore opted for PALM. We imaged integrin β3 and paxillin in fixed REF cells on fibronectin-coated substrates. The EMGM algorithm allowed us to identify integrin β3 objects with a typical area in the range between 0.01 and 1 μm^2, and containing between 10 and 100 localizations. Paxillin objects were found to have a similar area, but can contain more localizations, up to 1000. We attribute this difference to a tree-like organization of the FAs, rooting from isolated integrin islands, and expanding towards the actin filaments due to crosslinking and multivalent binding of paxillin and other proteins to their recruiting components. The equivalent diameter of the smallest objects was found to be around 100 nm (using the 2σ error ellipse area, which is 0.01 μm^2 for the smaller objects). This indeed justifies the need for super-resolution microscopy to investigate the inner structure of FAs. Most objects were found to exhibit substantial eccentricity, with values down to 0.1. An algorithm that does not assume radial symmetry, such as EMGM, is therefore essential for the analysis of FA substructure.

A fibronectin coating is often used to ensure good cell adhesion to the substrate. However, it is important to rule out that the observed FA substructure is a mere artifact of the binding sites

presented by such fibronectin-coated substrates. We therefore repeated the experiments on substrates that were patterned with a quasi-hexagonal grid of functionalized AuNPs. Our EMGM algorithm identified integrin β3 objects with areas in the same range as on fibronectin-coated substrates, while the number of localizations was lower, typically not exceeding 100. The FA substructure observed on the nano-patterned and fibronectin-coated substrates do not contradict each other.

The EMGM results sometimes display strongly overlapping mixture components, which is mathematically perfectly possible, but difficult to interpret. One possibility is that the background within the FAs is more complex than a simple uniform distribution. This could lead to the background partially being characterized by some of the mixture components, while the others are actual FA substructures. Note that our scanning procedure already captures background heterogeneities on the scale of the scanned areas. Another possibility is that a bivariate Gaussian is not the most accurate model for the FA subunits. To a certain extent, a post-analysis step can provide more insight. We performed a merging procedure that describes FA substructures either as an isolated Gaussian component, or a combination of several overlapping components. We hypothesize that a substantial set of the isolated components (areas between 0.001 and 0.01 μm^2, and number of localizations between 10 and 100) correspond to focal complexes or nascent adhesions. The overlapping mixture components, which appear to belong to FAs, have areas and localization numbers in the same range as the isolated components. This suggests that the observed objects are indicative of the real FA substructure. The merged components have a maximal area around 1 μm^2 and contain up to 1000 localizations, which can be interpreted as an upper limit for the FA substructure.

We envisage several ways in which our EMGM approach could be extended or adapted to allow a systematic and detailed study of the inner architecture of FAs. Several FA proteins could be investigated in multi-color mode to assess their spatial relationship. In this context, it could be of interest to develop an extension of EMGM that allows to investigate the “co-localization” of the mixture components. It would also be interesting to develop a 3D implementation of EMGM for the investigation of FA substructure in both the lateral and axial direction, as observed for instance by iPALM (18). It seems worthwhile to explore the possibility of incorporating other models than the Gaussian bivariate distribution, and other types of background besides the uniform one. Note that the effect of repetitive localizations on EMGM should be investigated, since photo-activatable fluorescent proteins can be localized more than once due to a phenomenon called “photoblinking” (37). Using transient transfection, a population of endogenous proteins will not be fluorescently labeled, and the labelled proteins might be overexpressed. Techniques such as CRISPR/cas9 can bring solutions to this problem (38).

5. Conclusion

We have used PALM to investigate FAs in REF cells growing on fibronectin-coated substrates and specifically bio-functionalized nano-patterned substrates, on which ordered patterns of nanoscale adhesive spots were provided. To quantify the FA subunit properties, we developed a method based on EMGM that accounts for localization uncertainty and background. Analysis of our PALM data indicates that integrin β3 and paxillin structures within FAs have areas between 0.01 and 1 μm^2, contain 10 to 100 localizations, and can exhibit substantial eccentricities. We believe that our EMGM based approach is generic enough for the investigation of various other SMLM imaged nanoscale structures as well, especially for closely packed protein structures, or objects that display strong radial asymmetries and differences in size and density.

Acknowledgements

The mEos2-paxillin-22 vector and the mEos2-Integrin-β3-N-18 vectors were kindly provided by Dr. Michael Davidson. H.D., J.P.S., and A.R. acknowledge the support of the Max Planck-EPFL Center for Molecular Nanoscience and Technology.

Supporting citations

References (39, 40) appear in the Supporting Material.

References

1. Zamir, E., and B. Geiger. 2001. Molecular complexity and dynamics of cell-matrix adhesions. J Cell Sci 114:3583-3590.
2. Zaidel-Bar, R., S. Itzkovitz, A. Ma'ayan, R. Iyengar, and B. Geiger. 2007. Functional atlas of the integrin adhesome. Nat Cell Biol 9:858-868.
3. Harizanova, J., Y. Fermin, R. S. Malik-Sheriff, J. Wieczorek, K. Ickstadt, H. E. Grecco, and E. Zamir. 2016. Highly Multiplexed Imaging Uncovers Changes in Compositional Noise within Assembling Focal Adhesions. Plos One 11:e0160591.
4. Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. Science 313:1642-1645.
5. Hu, S. Q., Y. H. Tee, A. Kabla, R. Zaidel-Bar, A. Bershadsky, and P. Hersen. 2015. Structured illumination microscopy reveals focal adhesions are composed of linear subunits. Cytoskeleton 72:235-245.
6. Morimatsu, M., A. H. Mekhdjian, A. C. Chang, S. J. Tan, and A. R. Dunn. 2015. Visualizing the Interior Architecture of Focal Adhesions with High-Resolution Traction Maps. Nano Lett 15:2220-2228.
7. Rossier, O., V. Octeau, J. B. Sibarita, C. Leduc, B. Tessier, D. Nair, V. Gatterdam, O. Destaing, C. Albiges-Rizo, R. Tampe, L. Cognet, D. Choquet, B. Lounis, and G. Giannone. 2012. Integrins beta(1) and beta(3) exhibit distinct dynamic nanoscale organizations inside focal adhesions. Nat Cell Biol 14:1057-1067.
8. Shibata, A. C. E., T. K. Fujiwara, L. M. Chen, K. G. N. Suzuki, Y. Ishikawa, Y. L. Nemoto, Y. Miwa, Z. Kalay, R. Chadda, K. Naruse, and A. Kusumi. 2012. Archipelago architecture of the focal adhesion: Membrane molecules freely enter and exit from the focal adhesion zone. Cytoskeleton 69:380-392.
9. Shroff, H., C. G. Galbraith, J. A. Galbraith, H. White, J. Gillette, S. Olenych, M. W. Davidson, and E. Betzig. 2007. Dual-color superresolution imaging of genetically expressed probes within individual adhesion complexes. P Natl Acad Sci USA 104:20308-20313.
10. Changede, R., X. C. Xu, F. Margadant, and M. P. Sheetz. 2015. Nascent Integrin Adhesions Form on All Matrix Rigidities after Integrin Activation. Dev Cell 35:614-621.
11. Tabarin, T., S. V. Pageon, C. T. T. Bach, Y. Lu, G. M. O'Neill, J. J. Gooding, and K. Gaus. 2014. Insights into Adhesion Biology Using Single-Molecule Localization Microscopy. Chemphyschem 15:606-618.

12. Deschout, H., A. Shivanandan, P. Annibale, M. Scarselli, and A. Radenovic. 2014. Progress in quantitative single-molecule localization microscopy. Histochem Cell Biol 142:5-17.
13. Nicovich, P. R., D. M. Owen, and K. Gaus. 2017. Turning single-molecule localization microscopy into a quantitative bioanalytical tool. Nat Protoc 12:435-461.
14. Gardel, M. L., I. C. Schneider, Y. Aratyn-Schaus, and C. M. Waterman. 2010. Mechanical Integration of Actin and Adhesion Dynamics in Cell Migration. Annu Rev Cell Dev Bi 26:315-333.
15. Bishop, C. M. 2006. Pattern recognition and machine learning. Springer.
16. Shroff, H., C. G. Galbraith, J. A. Galbraith, and E. Betzig. 2008. Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. Nat Methods 5:417-423.
17. Fuchs, J., S. Bohme, F. Oswald, P. N. Hedde, M. Krause, J. Wiedenmann, and G. U. Nienhaus. 2010. A photoactivatable marker protein for pulse-chase imaging with superresolution. Nat Methods 7:627-630.
18. Kanchanawong, P., G. Shtengel, A. M. Pasapera, E. B. Ramko, M. W. Davidson, H. F. Hess, and C. M. Waterman. 2010. Nanoscale architecture of integrin-based cell adhesions. Nature 468:580-584.
19. Arnold, M., E. A. Cavalcanti-Adam, R. Glass, J. Blummel, W. Eck, M. Kantlehner, H. Kessler, and J. P. Spatz. 2004. Activation of integrin function by nanopatterned adhesive interfaces. Chemphyschem 5:383-388.
20. Platzman, I., C. A. Muth, C. Lee-Thedieck, D. Pallarola, R. Atanasova, I. Louban, E. Altrock, and J. P. Spatz. 2013. Surface properties of nanostructured bio-active interfaces: impacts of surface stiffness and topography on cell-surface interactions. Rsc Adv 3:13293-13303.
21. Geiger, B., J. P. Spatz, and A. D. Bershadsky. 2009. Environmental sensing through focal adhesions. Nat Rev Mol Cell Bio 10:21-33.
22. Annibale, P., M. Scarselli, M. Greco, and A. Radenovic. 2012. Identification of the factors affecting co-localization precision for quantitative multicolor localization microscopy. Optical Nanoscopy 1:9.
23. Deschout, H., T. Lukes, A. Sharipov, D. Szlag, L. Feletti, W. Vandenberg, P. Dedecker, J. Hofkens, M. Leutenegger, T. Lasser, and A. Radenovic. 2016. Complementarity of PALM and SOFI for super-resolution live-cell imaging of focal adhesions. Nat Commun 7:13693.
24. Pallarola, D., I. Platzman, A. Bochen, E. A. Cavalcanti-Adam, H. Axmann, H. Kessler, B. Geiger, and J. P. Spatz. 2017. Focal adhesion stabilization by enhanced integrin-cRGD binding affinity. BioNanoMaterials https://doi.org/10.1515/bnm-2016-0014.
25. Mortensen, K. I., L. S. Churchman, J. A. Spudich, and H. Flyvbjerg. 2010. Optimized localization analysis for single-molecule tracking and super-resolution microscopy. Nat Methods 7:377-U359.
26. Ober, R. J., S. Ram, and E. S. Ward. 2004. Localization accuracy in single-molecule microscopy. Biophys J 86:1185-1200.
27. Sengupta, P., T. Jovanovic-Talisman, D. Skoko, M. Renz, S. L. Veatch, and J. Lippincott-Schwartz. 2011. Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. Nat Methods 8:969-975.
28. Owen, D. M., C. Rentero, J. Rossy, A. Magenau, D. Williamson, M. Rodriguez, and K. Gaus. 2010. PALM imaging and cluster analysis of protein heterogeneity at the cell surface. J Biophotonics 3:446-454.
29. Kiskowski, M. A., J. F. Hancock, and A. K. Kenworthy. 2009. On the Use of Ripley's K-Function and Its Derivatives to Analyze Domain Size. Biophys J 97:1095-1103.
30. Baddeley, D., I. D. Jayasinghe, L. Lam, S. Rossberger, M. B. Cannell, and C. Soeller. 2009. Optical single-channel resolution imaging of the ryanodine receptor distribution in rat cardiac myocytes. P Natl Acad Sci USA 106:22275-22280.

31. Endesfelder, U., K. Finan, S. J. Holden, P. R. Cook, A. N. Kapanidis, and M. Heilemann. 2013. Multiscale Spatial Organization of RNA Polymerase in Escherichia coli. Biophys J 105:172-181.
32. Ester, M., H. P. Kriegel, J. Sander, and X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96):226-231.
33. Verbeek, J. J., N. Vlassis, and B. Krose. 2003. Efficient greedy learning of Gaussian mixture models. Neural Comput 15:469-485.
34. Punzo, A., R. P. Browne, and P. D. McNicholas. 2014. Hypothesis testing for parsimonious Gaussian mixture models. arXiv 1405.0377.
35. Deschout, H., F. C. Zanacchi, M. Mlodzianoski, A. Diaspro, J. Bewersdorf, S. T. Hess, and K. Braeckmans. 2014. Precisely and accurately localizing single emitters in fluorescence microscopy. Nat Methods 11:253-266.
36. Sharonov, A., and R. M. Hochstrasser. 2006. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. P Natl Acad Sci USA 103:18911-18916.
37. Annibale, P., S. Vanni, M. Scarselli, U. Rothlisberger, and A. Radenovic. 2011. Identification of clustering artifacts in photoactivated localization microscopy. Nat Methods 8:527-528.
38. Ratz, M., I. Testa, S. W. Hell, and S. Jakobs. 2015. CRISPR/Cas9-mediated endogenous protein tagging for RESOLFT super-resolution microscopy of living human cells. Sci Rep-Uk 5:9592.
39. Busemeyer, J. R., and Y. M. Wang. 2000. Model comparisons and model selections based on generalization criterion methodology. J Math Psychol 44:171-189.
40. Vinga, S., and J. S. Almeida. 2004. Renyi continuous entropy of DNA sequences. J Theor Biol 231:377-388.

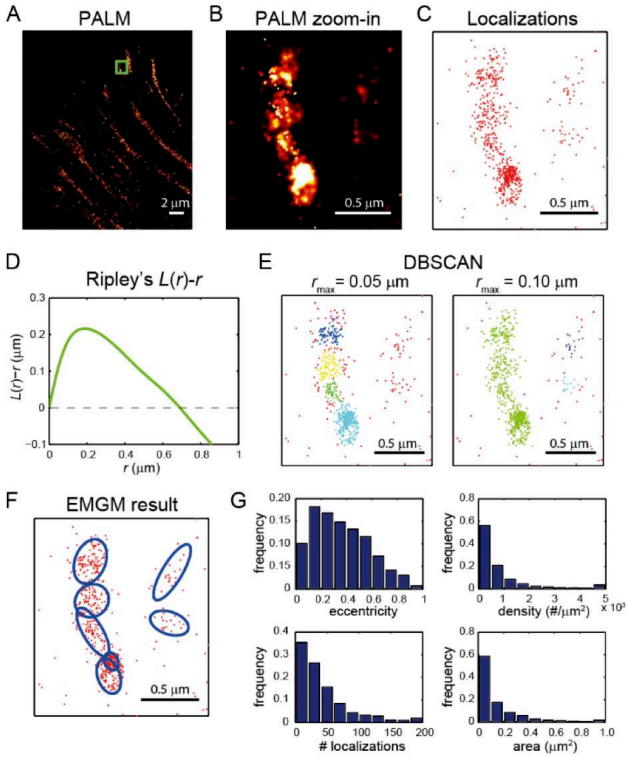

FIGURE 1 Application of SMLM clustering algorithms to PALM data of focal adhesions. (A) PALM image of a fixed REF cell expressing integrin β3 labelled with mEos2. (B) Zoom-in PALM image corresponding to the green rectangle in (A). (C) Scatter plot of the mEos2 localizations corresponding to the green rectangle in (A). (D) Ripley's $L(r)$-r as a function of r, obtained from the localizations in (C). (E) Clusters obtained from the localizations in (C) by DBSCAN. The minimum number of localizations was set to 10, and two values were chosen for the maximum search radius r_{max}: 0.05 and 0.10 µm. The different colors of the localizations indicate to which cluster they belong, the background localizations are red. (F) Result of EMGM analysis of the localizations in (C). The red dots symbolize the localizations, and the blue ellipses the 2σ error ellipses of the components. (G) Histograms showing the eccentricity b/a, localization density, number of localizations, and area πab of the 2σ error ellipses of the components obtained by EMGM from the complete PALM data set in (A). The rightmost bins in each histogram (except for the eccentricity histogram) contain all values within that bin and larger.

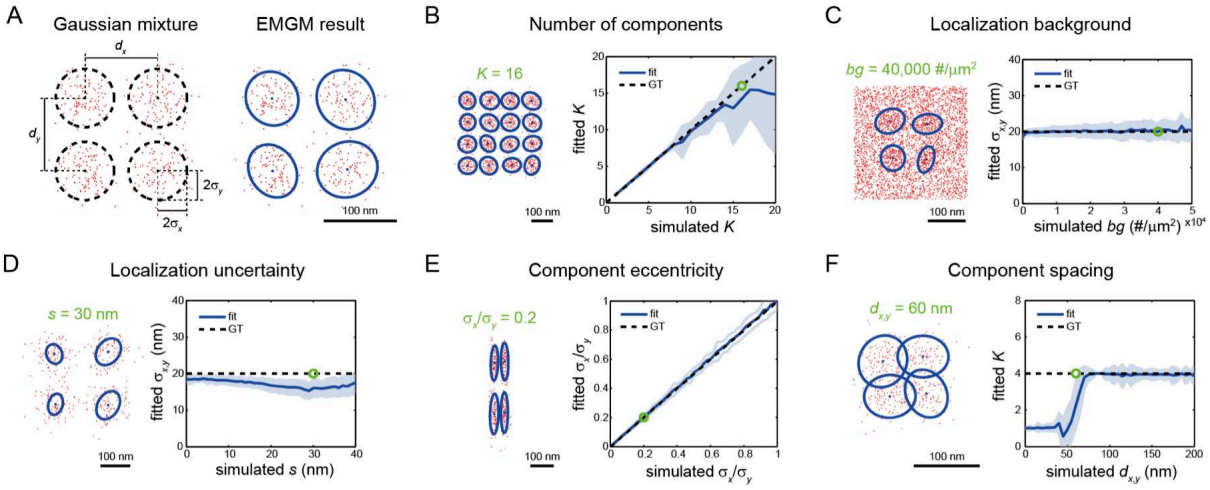

FIGURE 2 Evaluation of EMGM using simulated data. (A) On the left, an example of a simulated Gaussian mixture consisting of K = 4 components, each containing 100 localizations, described by a symmetric 2D Gaussian distribution with a standard deviation $\sigma_x = \sigma_y$ = 20 nm. The Gaussian centers are placed in a square grid with spacing $d_{x,y}$ = 100 nm. On the right, the EMGM result. The red dots symbolize the localizations. The blue dots symbolize the center positions and the blue ellipses the 2σ error ellipses of the components. (B) On the right, the average number of mixture components correctly identified by EMGM as a function of the simulated K. On the left, an example EMGM result for K = 16. (C) On the right, the average standard deviation $\sigma_{x,y}$ of the mixture components calculated by EMGM as a function of the simulated localization background density bg. On the left, an example EMGM result for bg = 40,000 #/µm^2. (D) On the right, the average $\sigma_{x,y}$ calculated by EMGM as a function of the simulated localization uncertainty s. On the left, an example EMGM result for s = 30 nm. (E) On the right, the average eccentricity σ_x/σ_y of the mixture components calculated by EMGM as a function of the simulated σ_x/σ_y. On the left, an example EMGM result for σ_x/σ_y = 0.2. (F) On the right, the average number of mixture components correctly identified by EMGM as a function of the simulated spacing $d_{x,y}$. On the left, an example EMGM result for $d_{x,y}$ = 60 nm. The simulated Gaussian mixtures in (C-F) consist of K = 4 components, similar to (A). The dashed lines in (B-F) represent the ground truth (GT) and the shaded areas the standard deviation (n = 100).

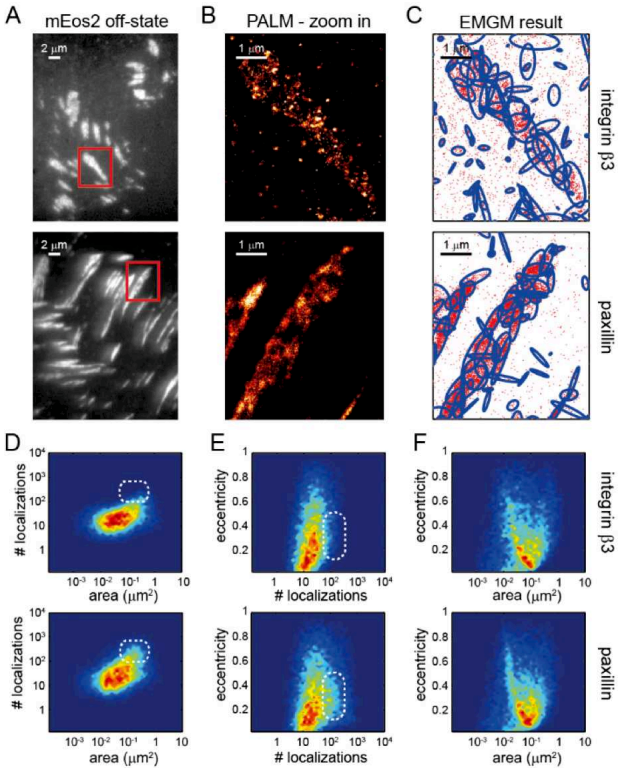

FIGURE 3 EMGM analysis of PALM data of integrin β3 or paxillin on fibronectin-coated substrates. (A) Summed TIRF images of the mEos2 off-state of fixed REF cells expressing integrin β3 or paxillin labelled with mEos2, growing on fibronectin-coated substrates. (B) Zoom-in PALM images corresponding to the red rectangles in (A). (C) Result of the EMGM analysis of the PALM data shown in (B). The red dots symbolize the localizations, and the blue ellipses the 2σ error ellipses of the mixture components. (D-F) Result of the EMGM analysis of PALM data corresponding to different REF cells (n = 10): (D) number of localizations in each mixture component as a function of the area of its 2σ error ellipse, (E) eccentricity of the 2σ error ellipse of each mixture component as a function of its number of localizations, (F) eccentricity of the 2σ error ellipse of each mixture component as a function of its area. The dashed white rounded rectangles in (D) and (E) are visual guides.

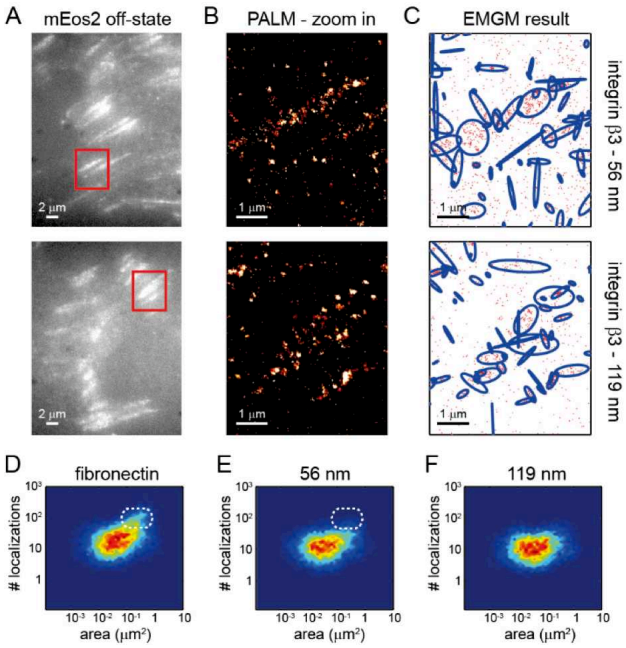

FIGURE 4 EMGM analysis of PALM data of integrin β3 on nano-patterned substrates. (A) Summed TIRF images of the mEos2 off-state of fixed REF cells expressing integrin β3 labelled with mEos2, growing on nano-patterned substrates with 56 nm or 119 nm spacing between the AuNPs. (B) Zoom-in PALM images corresponding to the red rectangles in (A). (C) Result of the EMGM analysis of the PALM data shown in (B). The red dots symbolize the localizations, and the blue ellipses the 2σ error ellipses of the mixture components. (D-F) Result of the EMGM analysis of PALM data corresponding to different REF cells (n = 10). The number of localizations in each mixture component is shown as a function of the area of its 2σ error ellipse, for (D) fibronectin coated substrates (Fig. 3D), (E) nano-patterned substrates with 56 nm spacing, (F) nano-patterned substrates with 119 nm spacing. The dashed white rounded rectangles in (D) and (E) are visual guides.

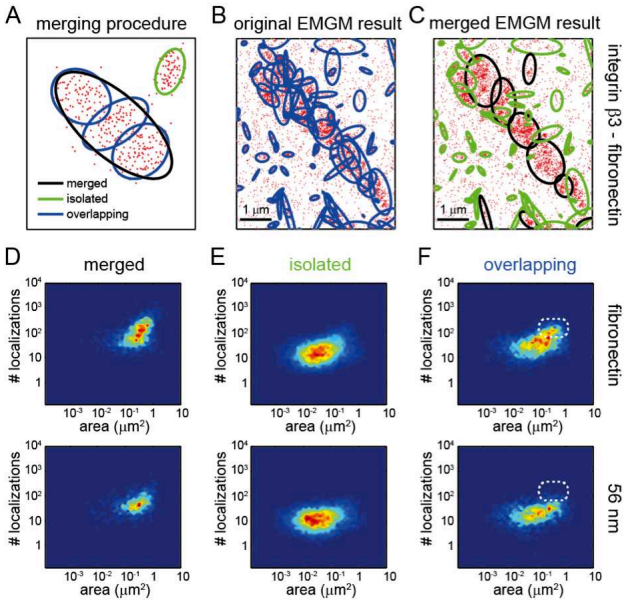

FIGURE 5 Merging procedure applied on EMGM results for integrin β3. (A) Illustration of the concept of merging overlapping mixture components based on overlapping error ellipses. The red dots symbolize the localizations. The black/green/blue ellipses represent the 2σ error ellipses of the merged/isolated/overlapping mixture components. (B) EMGM result for PALM data of a fixed REF cell growing on a fibronectin-coated substrate expressing integrin β3 labelled with mEos2 (Fig. 3C). (C) Result of the merging procedure applied on the EMGM result in (B). (D-F) Result of the merging procedure applied on EMGM results for integrin β3 (Fig. 3D, and Fig. 4E). The number of localizations in each mixture component is shown as a function of the area of its 2σ error ellipse, for (D) the merged components, (E) the isolated components, and (F) the overlapping components. The dashed white rounded rectangles in (F) are visual guides.