bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Super-resolution modularity analysis shows that polyhedral caveolin-1 oligomers combine to form scaffolds and caveolae

Ismail M. Khater[1], **Qian Liu**[2], **Keng C. Chou**[2], **Ghassan Hamarneh**[1*], **Ivan Robert Nabi**[3*#]

[1]Medical Image Analysis Lab, School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

[2]Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

[3]Department of Cellular and Physiological Sciences, Life Sciences Institute, University of British Columbia, Vancouver, BC V6T 1Z3, Canada

*Correspondence: irnabi@mail.ubc.ca (IRN), hamarneh@sfu.ca (GH)

*Equal contribution

[#]Lead Contact

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT

Network analysis of single molecule super resolution data of caveolin-1 (Cav1) identifies caveolae and 3 distinct non-caveolar Cav1 scaffolds. Here, modularity analysis of endogenous caveolae and scaffolds in HeLa cells shows that small scaffolds combine to form larger scaffolds and caveolae. We leveraged a spectral decomposition algorithm to divide the blob's network, based on the eigenvector representation of its adjacency matrix, into modules or communities. The optimization problem of finding modules of a blob involves maximizing intra-connectivity between Cav1 molecules within a module and minimizing inter-connectivity between Cav1 molecules across modules. Features of modules are then matched with intact blobs by finding the similarity between the module-blob pairs of group centers. Our results show that smaller S1A and S1B scaffolds are made up of small polygons, that S1B scaffolds correspond to S1A scaffold dimers and that caveolae and hemi-spherical S2 scaffolds are complex, modular structures formed from S1B and S1A scaffolds, respectively. Polyhedral interactions of Cav1 oligomers therefore leads progressively to the formation of larger and more complex scaffold domains and the biogenesis of caveolae.

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Of the various super-resolution microscopy approaches, the best resolution is obtained using single-molecule localization microscopy (SMLM), based on the repeated activation (blinking) of small numbers of discrete fluorophores, usually through *d*STORM. Precise localization of these blinks is determined from a Gaussian fit of the point spread function (PSF) providing ~10 nm X-Y (lateral) resolution and ~30 nm Z (axial) resolution for astigmatic lens 3D SMLM (Huang et al., 2008; Shroff et al., 2008). SMLM generates point coordinates in 3D space that can then be used to reconstruct localizations with significantly improved resolutions. An alternate approach to study point distributions is to visualize them as a graph or network. Graphs are mathematical structures used to model interactions between entities for many systems, with the entities represented as graph nodes and the connections between them as edges (Newman, 2003). Real world graphs are frequently complex networks that have many different subgraphs or modules (Kim and Wilhelm, 2008). Networks with high modularity have dense connections (edges) between the nodes within modules (sub-networks) and sparse connections between nodes in different modules. The optimization problem of finding divisions within a network (i.e. modules or communities) has been solved via various methods such as normalized-cut graph partitioning and spectral algorithms (Newman, 2006b; Newman, 2013). Network and subgraph (module) analysis is therefore ideally suited to define molecular and subgroup organization between labeled molecules within the 3D SMLM point cloud of macromolecular complexes.

A major challenge to determining molecular structure by SMLM is defining molecular localizations from the millions of blinks generated by SMLM, many of which derive from the same labeled molecule, particularly when the labeling approach is based on antibody labelling (i.e. *d*STORM). The same fluorophore can blink twice in succeeding acquisition frames dependent on the on-off duty cycle (Dempsey et al., 2011a) and the same molecule can be labeled by different fluorophores, either on the same secondary antibody or on different secondaries bound to the same target protein, introducing error in blink localization relative to the actual antigen. Each of these blinks is in addition subject to localization error due to drift and to Gaussian fitting of the PSF. Multiple, distinct localizations therefore derive from the same molecule and generate a dense non-biological network (NBN) with high degree nodes centred around the actual molecule (Khater et al., 2017; Khater et al., 2018). Network analysis of the biological network, composed of nodes corresponding to molecular localizations of the labeled proteins, requires reduction/consolidation of the NBNs. To do so, we developed a merge algorithm, in which blinks within a defined merge proximity threshold (MPT) are iteratively combined into one average position (Khater et al., 2017; Khater et al., 2018). Blinks in closest proximity are combined first such that merging is initiated within the dense NBNs and continues progressively until no localizations within the point cloud are closer than the MPT.

Caveolae are smooth 50-80 nm plasma membrane invaginations whose formation requires the coat protein Cav1 and the adaptor protein CAVIN1 (also called PTRF)(Hill et al., 2008). Functional roles of caveolae include: mechanoprotective membrane buffers; mechanosensors; signaling hubs; and endocytic transporters (Parton and del Pozo, 2013). Cryo-electron microscopy (cryoEM) analysis of caveolae has reported that the Cav1 coat is polygonal, formed of distinct edges and suggested to form a dodecahedral cage (Ludwig et al., 2016; Stoeber et al., 2016). CryoEM analysis of Cav1 protein distribution in the caveolae coat, in either mammalian

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cells or following heterologous Cav1 expression in bacteria (h-caveolae), show that Cav1 exhibits a highly regular distribution of repeating polygons (Ludwig et al., 2013; Ludwig et al., 2016; Walser et al., 2012). CAVIN1 forms an outer filamentous coat layer whose filamentous structure likely corresponds to the striations observed on caveolae by deep-etch EM (Kovtun et al., 2014; Ludwig et al., 2013; Ludwig et al., 2016; Rothberg et al., 1992; Stoeber et al., 2016).

In the absence of CAVIN1, Cav1 is localized to non-caveolar membrane domains known as Cav1 scaffolds (Lajoie et al., 2009). Previously, application of machine learning and network analysis to SMLM data sets for Cav1 in PC3 prostate cancer cells (Khater et al., 2018), that express Cav1 but no CAVIN1 and therefore no caveolae (Hill et al., 2008), identified two classes of Cav1 scaffolds corresponding to small Cav1 homo-oligomers (S1 scaffolds) (Monier et al., 1995; Sargiacomo et al., 1995) as well as larger hemispherical S2 scaffolds. The formation of curved Cav1 structures in the absence of CAVIN1 is consistent with Cav1 induction of invaginated h-caveolae in bacteria and supports a role for Cav1 in membrane curvature (Ariotti et al., 2015; Walser et al., 2012). Larger hollow caveolae were only detected upon transfection of PC3 cells with CAVIN1 (PC3-PTRF cells) (Khater et al., 2018) and their modular nature supported the polyhedral Cav1 coat structure observed by cryoEM (Ludwig et al., 2016; Stoeber et al., 2016).

We have now applied SMLM network analysis to endogenous caveolae and scaffolds in HeLa cells labeled using primary anti-Cav1 antibodies and Alexa647 conjugated secondary antibodies. To enhance localization precision, we used an SMLM microscope equipped with real-time nanometer-scale drift correction hardware (Tafteh et al., 2016a) and included only high-precision localizations to improve localization accuracy (Foi et al., 2008; Thompson et al., 2002). With this in-house built SMLM microscope, the localization precision approaches 10 nm (Liu et al., 2018), and the drift is limited to 1 nm in the x-y plane and 3 nm in the z axis (Tafteh et al., 2016a; Tafteh et al., 2016b). Improved resolution of the approach has identified sub-modules within S1 scaffolds that correspond to the polygons observed in the Cav1 coat by cryoEM (Ludwig et al., 2013; Ludwig et al., 2016; Walser et al., 2012). S1A scaffolds correspond to stable Cav1 homo-oligomers detected biochemically (Monier et al., 1995; Sargiacomo et al., 1995). Modularity analysis and group matching shows that S1A scaffolds can dimerize to form S1B scaffolds and oligomerize to form hemispherical S2 scaffolds. S1B scaffolds match the modules that make up the caveolae coat suggesting that the caveolae coat is built progressively by dimerization of S1A scaffolds, composed of the basic polygonal Cav1 units, that then combine to form a polyhedral caveolae coat.

RESULTS AND DISCUSSION

3D point clouds of Cav1 from 10 *d*STORM images of HeLa cells were processed using the *3D SMLM Network Analysis* computational pipeline (Khater et al., 2018). Figure 1A shows the 3D point cloud of one of the HeLa cells in our dataset at various stages of the pipeline: 1) Iterative merging of the acquired localizations for molecular reconstruction and noise filtering comparing degree network measure with a random network to filter out noisy localizations and monomeric Cav1, retaining clustered Cav1; 2) Segmentation into separate blobs and extraction of a 28

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

feature/descriptor vector for every blob; 3) Unsupervised machine learning to learn the various Cav1 domains from the extracted blobs and their descriptor features. Machine learning identified four groups of Cav1 domains (H1, H2, H3, and H4) in HeLa cells. We used the Euclidean distance in 28 dimensions to encode similarity of HeLa groups with groups previously identified in PC3 and PC3-PTRF cells (Khater et al., 2018), with similarity proportional to the inverse Euclidean distance. As seen in Figure 1B, for the groups with larger blobs, H2 matches PP2, corresponding to caveolae, and H1 matches PP1, corresponding to the larger hemispherical S2 scaffolds. For the smaller S1 scaffolds, H4 matches PP3 and H3 matches PP4. Distribution of the different classes of blobs in the different cell types shows that HeLa and PC3-PTRF cells present a similar distribution of Cav1 blobs with slightly more caveolae detected in HeLa cells (Fig. 1C).

Tunable iterative merge algorithm

We previously applied a fixed MPT of 20 nm (Khater et al., 2018) and now applied MPTs from 10-20 nm in steps of 1 nm. Importantly, 4 classes were learnt at each MPT from 10-20 nm. Further, tuning the MPT from 10-20 nm minimally impacted classification, size, modularity, characteristic path and hollowness of all 4 classes of blobs (Fig. 2A). Machine learning blob classification is therefore independent of the merge algorithm for MPTs from 10-20 nm. Shape, topology, hollowness, and modularity features, as well as, most importantly, blob classification, were minimally impacted by changing the MPT for iterative merging of localizations from 10-20 nm, except, not unexpectedly, for molecular localization number per blob and the other features that are closely related to the number of localizations feature.

We set the MPT based on the reported 145 Cav1 proteins per caveolae (Pelkmans and Zerial, 2005). An MPT of 19 nm resulted in an average of 142 localizations for the H2 blobs (Fig. 2A,B), that match the PP2 caveolae blobs from PC3-PTRF cells (Fig. 1B). Feature analysis after group matching shows that the four HeLa groups match the four PC3-PTRF groups as well as the S2 and S1A scaffolds present in PC3 cells with high degree (Fig. 2B). Relative to the PC3 data (Khater et al., 2018), we observed a doubling in molecular localizations for S1B scaffolds relative to S1A scaffolds and increased modularity of S1B scaffolds in the HeLa data set that we attribute to the improved resolution obtained with the real-time drift control SMLM (Tafteh et al., 2016a; Tafteh et al., 2016b). Increased Cav1 localization number in S1B scaffolds parallels the increased size (X-range) and reduced network density of these clusters relative to S1A scaffolds, reflecting differences between these structures that led to their classification as distinct cluster groups in this and our previous analysis (Khater et al., 2018). Indeed, we observe a progressive increased number of localizations (Caveolae > S2 Scaffolds > S1B scaffolds > S1A scaffolds) associated with increased modularity and decreased network density (Fig. 2B). Caveolae are the most modular structures (modularity > 0.4), then S2, and S1B. S1A scaffolds have the least tendency to form modules (modularity < 0.04).

Small Cav1 S1 scaffolds combine to build larger scaffolds and caveolae

The variable modularity of the different classes of Cav1 blobs led us to extract the blobs' modules and study their features. To make sure that we correctly construct the modules within the blob's network, we used multi-proximity threshold (PT) network analysis (Fig. 3A) to decompose the blobs' networks into modules using spectral analysis. For all the groups, any PT greater than 60 nm renders every blob a single connected component; the average connected

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

component size plateaus and equals the blob size for the different groups at PTs greater than 60 nm. This indicates that at PT > 60 nm, all post-merge Cav1 localizations in a cluster are within 60 nm of each other. The number of modules and of Cav1 localizations per module are stable across the PT range from 60 to 170 nm. This range is therefore suitable to determine the number of modules and of Cav1 localizations per module. HeLa caveolae were found to be highly modular containing 6-7 modules of ~29 Cav1 localizations, S2 scaffolds 5 modules of ~14 localizations and S1B scaffolds ~4 modules of 7-8 localizations each (Fig. 3A). S1A scaffolds have the minimum average number of modules of ~2 modules per blobs of ~5-6 localizations each.

Visualization of blobs from the identified groups (Fig. 3B) highlights the modular nature of the various Cav1 structures. At 80 nm, each blob forms one connected component network and extracted modules for every blob are shown in different colors. The presence of small modules (~5-8 molecules) within both S1A and S1B scaffolds is indicative of an additional degree of suborganization within these small scaffold domains. 3D cryoEM tomography identified a network of 3-way junctions and polygonal arrangements of Cav1 protein densities within the caveolae coat (Ludwig et al., 2016). Similarly, cryoEM analysis of Cav1-induced vesicles in bacteria (h-caveolae) present distinct polygonal repeating units on the h-caveolae cage (Walser et al., 2012). We propose that the sub-modules that we detect in S1A and S1B scaffolds correspond to these polygonal repeating units that comprise the caveolae coat. The fact that S1A scaffolds form one connected component unit and that the number of localizations of S1A scaffolds matches that of Cav1 homo-oligomers (~14-15 Cav1s) (Monier et al., 1995; Sargiacomo et al., 1995) suggests that interaction between these polygonal sub-modules forms more stable structural units. This is supported by the identification of larger modules in both S2 scaffolds and caveolae (Fig. 3A,4A).

Most interestingly, the decomposed modules from the different Cav1 cluster groups show a much higher degree of similarity in terms of the shape, topology and network features than the clusters from which they originate (Fig. 4A). For instance, while Cav1 blob classes show a progressive reduction in network density from S1A scaffolds to caveolae, modules from the different blob classes show a similar network density. This suggests that differential interaction between modules is responsible for the changes in network density of the different classes of Cav1 blobs and that these modules form fundamental building blocks of larger Cav1 structures.

Indeed, many features of caveolae modules match S1B scaffolds while S2 and S1B modules match S1A scaffolds. For instance, #localizations, hollowness, characteristic path, modularity, size, and network density of the caveolae modules (blue bars to right of graphs) are very similar to their corresponding features in the S1B blobs (magenta bars to left) (Fig. 4A). However, we quantitatively assessed module-blob similarity across all features using the matching matrix of the features for the various blobs and modules group center using Euclidean distance (Fig. 4B). The column-wise (i.e. the modules) similarity shows that: S2 scaffold modules match S1A blobs, caveolae modules match S1B blobs, and S1B modules match S1A blobs. The close matching of S1B modules with S1A blobs and doubling in number of modules and localizations of S1B modules relative to S1A blobs suggests that S1B scaffolds represent dimers of S1A scaffolds. Further, PC3 cells that lack caveolae have only S1A and S2 scaffolds (Fig. 1 B,C) (Khater et al., 2018) supporting the matching between S1A blobs and S2 modules reported here. The

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

dissimilarity between caveolae and S2 scaffolds and the modules of any other blob types suggests that these are complex structures made up of primitive S1A and S1B scaffolds.

Overall, our data support a model in which Cav1 is organized into smaller units of 5-8 Cav1 localizations that correspond to the polygonal base units observed by cryoEM analysis of the Cav1 caveolar coat (Ludwig et al., 2016; Walser et al., 2012). These base units combine to form larger stable structures of which the smallest is S1A scaffolds, that we propose correspond to the previously identified ~14-15 Cav1 homo-oligomers (Monier et al., 1995; Sargiacomo et al., 1995). We also identify S1B scaffolds, previously classified as distinct from S1A scaffolds (Khater et al., 2018), as larger structures that may correspond to S1A dimers. Modularity analysis and group matching show that S1A scaffolds combine to form both S1B dimers and the larger hemispherical S2 scaffold structures. Caveolae modules show better matching and correspond in size to S1B and not S1A modules suggesting that caveolae formation may be a two-step process in which S1A scaffolds first combine to form dimers that then interact to form the caveolae coat (Fig. 5, Video S1). Consistent with a role for S1B in caveolae formation, PC3 cells that lack caveolae and CAVIN1 do not contain S1B scaffolds (Khater et al., 2018). The progressive reduction in density of Cav1 domains, but not of their constituent modules, as cluster size increases suggests that interaction between smaller S1 scaffolds to form larger structures, including caveolae, is associated with changes in how modules interact and are organized.

Importantly, our analysis based on TIRF microscopy argues that all 4 Cav1 domains, from S1A scaffolds to caveolae are present at the plasma membrane. Based on the reported kiss-and-run association of caveolae with the plasma membrane and the role of caveolae as membrane buffers that flatten in response to mechanical stretching (Pelkmans and Zerial, 2005; Sinha et al., 2010), we suggest that these modular interactions are dynamic and reversible.

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

MATERIALS AND METHODS

Cell culture and immunofluorescent labeling

HeLa cells were tested for mycoplasma by PCR (Applied Biomaterial, Vancouver, BC, Canada) and cultured in Dulbecco's Modified Eagle's medium (DMEM; Invitrogen) containing 10% fetal bovine serum (Invitrogen). For SMLM imaging, cells were plated on fibronectin coated coverslips (No. 1.5H) for 24 h prior to fixation with 3% paraformaldehyde (PFA) for 15 min at room temperature. Fixed cells were rinsed with PBS/CM (phosphate buffered saline complemented with 1 mM $MgCl_2$ and 0.1 mM $CaCl_2$), permeabilized with 0.2% Triton X-100 in PBS/CM, and blocked with 10% goat serum and 1% bovine serum albumin (BSA; Sigma-Aldrich Inc.) in PBS/CM before incubation with rabbit anti-caveolin-1 (BD Transduction Inc.) for 12 h at 4°C and then Alexa Fluor 647-conjugated goat anti-rabbit (Thermo-Fisher Scientific Inc.) for 1 h at room temperature. Primary and secondary antibodies were diluted in SSC (saline sodium citrate) buffer containing 1% BSA, 2% goat serum and 0.05% Triton X-100. Cells were washed extensively after each antibody incubation with SSC buffer containing 0.05% Triton X-100 and post-fixed using 3% PFA for 15 min followed by extensive washing with PBS/CM. Near-infrared fiducial markers (diameter 100 nm; Thermo Fisher Scientific) were added for real-time drift correction. Immediately prior to imaging, cells were mounted and sealed on glass depression slides in freshly prepared imaging buffer (10% glucose, 0.5 mg/ml glucose oxidase, 40 µg/mL catalase, 50 mM Tris, 10 mM NaCl and 50 mM β-mercaptoethylamine (Sigma-Aldrich Inc.) in double-distilled water (Dempsey et al., 2011b; Huang et al., 2008).

SMLM Imaging

Imaging of Hela cells was performed on a in-house built SMLM system equipped with an apochromatic TIRF oil immersion objective lens (60x/1.49; Nikon Instruments) and a real-time drift correction system which limits the lateral drift to ~1 nm and the axial drift to ~3 nm. A 639 nm laser line (Genesis MX639, Coherent Inc., USA) was used to excite Alexa Fluor 647 fluorophores and near-infrared fiducial markers. A 405 nm laser line (Laserglow Technologies) was used to activate Alexa Fluor 647. The detailed optical setup and the imaging acquisition procedure were described previously (Tafteh et al., 2016a; Tafteh et al., 2016b).

The dataset used in this work consists of 10 fields of view (FOV) of Cav1-labeled HeLa cells. Each HeLa FOV is 54×54×0.8 μm^3 which is 9 times larger than the FOV in the PC3 cell study (18×18×0.8 μm^3), acquired using a Leica GSDIM microscope equipped with a 160X objective (Khater et al., 2018). Each HeLa FOV therefore included multiple cells and this study analysed a larger number of cells than the PC3 study. We collected 40,000 frames per super resolution image. The total number of collected localizations that we processed per image ranged from 1.6 to 6.7 million. The *3D SMLM Network Analysis* method (Khater et al., 2018) was able to process the whole FOV. Details of the *3D SMLM Network Analysis* approach can be found in (Khater et al., 2018).

Multi-proximity threshold network modularity analysis

For every blob of 3D localizations, more than one network can be constructed; one per each proximity threshold in the set $\{PT_1, PT_2,\ldots,PT_T\}$ (i.e. $blob_i$ has T networks $\{G_i^1, G_i^2,\ldots, G_i^T\}$,

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where G_i^t is composed of a set of nodes V_i and edges E_i^t to form $G_i^t(V_i, E_i^t)$). V_i, unaffected by PT_t, represents the molecules of $blob_i$ and E_i^t is the set of edges connecting all pairs of molecules interacting within PT_t nm.

Given G_i^t, a $blob_i$'s network at PT_t, we find the optimal number of modules (communities) using eigenvectors of the network adjacency matrix. At small PTs, the molecules of a blob might not form one connected network (i.e. the network might consist of more than one connected component). A blob network containing non-dense and non-connected regions cannot be used to extract modules (i.e. as per definition, networks with high modularity have dense connection between the nodes within modules and sparse connections between nodes in different modules). Hence, PTs that generate networks with more than one connected component should be avoided when extracting the modules. We leveraged the spectral algorithm to decompose the adjacency matrix into modules. Specifically, we utilized the Newman method (Newman, 2006a; Newman, 2006b), where the optimization problem of finding the modules of one blob is to maximize the intra-connectivity between the Cav1 molecules within a module and minimize the inter-connectivity between the Cav1 molecules across the modules.

Features extraction and module-blob similarity

For every segmented Cav1 blob, we extracted 28 features (Khater et al., 2018). The features are then used to group the blobs' into classes using the *3D SMLM Network Analysis* pipeline (Khater et al., 2018). The learned classes from HeLa dataset are then matched with the previously identified Cav1 domains from PC3 and PC3-PTRF datasets (Khater et al., 2018). The modules of the HeLa Cav1 blobs are then extracted using the multi-proximity threshold network modularity analysis described in the previous subsection. We extracted 28 features for every module. To find the similarity/dissimilarity among the extracted modules and the various blobs, we leveraged the matching analysis to match blob modules with intact blobs using the Euclidean distance of group centers.

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ACKNOWLEDGMENTS

Supported by grants from the CIHR (PJT-156424, PJT-159845), NSERC and CFI/BCKDF (IRN, GH, KCC).

AUTHOR CONTRIBUTIONS

Conceptualization, I.M.K., I.R.N. and G.H.; Methodology, I.M.K., I.R.N. and G.H; Software, I.M.K. and G.H.; Sample Preparation, Q. L.; SMLM Imaging, Q. L. and K. C. C.; Formal Analysis, I.M.K., I.R.N. and G.H.; Investigation, I.M.K., I.R.N. and G.H.; Resources, K.C.C., I.R.N. and G.H.; Writing – Original Draft, I.M.K., I.R.N. and G.H.; Writing – Review & Editing, I.M.K., I.R.N. and G.H.; Visualization, I.M.K., I.R.N. and G.H.; Supervision, I.R.N. and G.H.; Project Administration, I.R.N. and G.H.; Funding Acquisition, K.C.C., I.R.N. and G.H.

CONFLICT OF INTEREST

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

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FIGURE LEGENDS

Figure 1. *3D SMLM Network Analysis* of the HeLa cells dataset. A. 3D point clouds of Cav1 of a representative HeLa cell imaged with drift-corrected *d*STORM (Tafteh et al., 2016a; Tafteh et al., 2016b) before (green) and after (red) iterative blink merging and filtering out noisy localizations. Color-coded representations of blobs after segmentation and after identification by machine learning using *3D SMLM Network Analysis* (Khater et al., 2018) pipeline are shown. We identified four groups of blobs representing different Cav1 domains in HeLa cells. **B.** Matching HeLa Cav1 groups with the previously identified Cav1 domains in PC3 and PC3-PTRF (Khater et al., 2018). The numbers are the Euclidean distances between the groups that capture the similarity/dissimilarity. We matched learned groups from PC3-PTRF and HeLa cells and show distances among the feature vector of group centers and present color matching of HeLa groups with previously identified P1 and P2 Cav1 domains in PC3 cells and PP1, PP2, PP3, and PP4 Cav1 domains in PC3-PTRF cells (Khater et al., 2018). **C.** Distribution of the matched groups from HeLa, PC3-PTRF and PC3 datasets are presented for comparison.

Figure 2. MPT tuning does not impact blob identification. A. Biological signatures of HeLa Cav1 blobs at different MPTs (10-20 nm) were obtained by *3D SMLM Network Analysis* (Khater et al., 2018). We learn 4 groups/classes of Cav1 domains at each MPT. Cav1 blob shape, topology, hollowness and network features are minimally affected by MPT tuning while number of localizations and related features are affected by MPT tuning. Error bars represent standard deviation. **B.** Signatures of matched groups from HeLa, PC3 and PC3-PTRF cells show a high degree of correspondence of the individual group features.

Figure 3. Modularity analysis of Cav1 blobs. A. Multi-proximity threshold modularity analysis shows the number of connected components, number of modules and localizations per module (at 19 nm MPT) for HeLa blobs at different proximity thresholds. **B.** Representative blobs from the different HeLa Cav1 domains are shown. Visualization shows the blob's localizations, the localizations' connections, and the blob's modules.

Figure 4. Module-blob matching between Cav1 domains. A. Signatures of Cav1 blobs and blob modules shows that some module features are similar to blob features. For example, the right bars that represent the caveolae modules (blue) are very similar to the left bars that represent the S1B blobs (magenta). **B.** We extracted 28 features (e.g. shape, topology, hollowness, network…) for every blob and module. The table encodes the module-blob similarity between the different Cav1 domains (blobs) and the modules of each type as Euclidean distances between every pair of group centres.

Figure 5. Modular interaction of Cav1 S1A scaffolds forms larger scaffolds and caveolae. Based on the module-blob matching results (Fig. 4B), S1A blobs are stable primitive structures (simplex) that are used to build up more complex, modular S1B and S2 scaffolds. S1B scaffolds correspond to S1A dimers and are used to build the caveolae coat complex (see Video S1). The figure also shows the hemispherical shape of S2 blobs and the hollow caveolae blobs.

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supp. Video S1. Animated rotating structures of S1A scaffolds (H3), S1B scaffolds (H4), S2 scaffolds (H1), and caveolae (H2) blobs in HeLa cells. The blob's localizations, connections between localizations (edges), and modules (for S1B, S2 scaffolds and caveolae) are shown from different angles with rotation. The video illustrates Cav1 domain formation as described in Figure 5.

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

Ariotti, N., J. Rae, N. Leneva, C. Ferguson, D. Loo, S. Okano, M.M. Hill, P. Walser, B.M. Collins, and R.G. Parton. 2015. Molecular Characterization of Caveolin-induced Membrane Curvature. *J. Biol. Chem.* 290:24875-24890.

Dempsey, G.T., J.C. Vaughan, K.H. Chen, M. Bates, and X. Zhuang. 2011a. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nature methods*. 8:1027-1036.

Dempsey, G.T., J.C. Vaughan, K.H. Chen, M. Bates, and X. Zhuang. 2011b. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nature methods*. 8:1027.

Foi, A., M. Trimeche, V. Katkovnik, and K. Egiazarian. 2008. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Trans Image Process*. 17:1737-1754.

Hill, M.M., M. Bastiani, R. Luetterforst, M. Kirkham, A. Kirkham, S.J. Nixon, P. Walser, D. Abankwa, V.M. Oorschot, S. Martin, J.F. Hancock, and R.G. Parton. 2008. PTRF-Cavin, a conserved cytoplasmic protein required for caveola formation and function. *Cell*. 132:113-124.

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

Khater, I.M., Y. Liu, Q. Liu, F. Meng, K.C. Chou, I.R. Nabi, and G. Hamarneh. 2017. Distinguishing Biological and Non-Biological Networks in Single Molecule Localization Super Resolution Microscopy. *In* American Society for Cell Biology and European Molecular Biology Organization (ASCB-EMBO). Vol. 28. Molecular Biology of the Cell Philadelphia, USA. page 1094 (Abstract P2728).

Khater, I.M., F. Meng, T.H. Wong, I.R. Nabi, and G. Hamarneh. 2018. Super Resolution Network Analysis Defines the Molecular Architecture of Caveolae and Caveolin-1 Scaffolds. *Scientific Reports*. 8:9009.

Kim, J., and T. Wilhelm. 2008. What is a complex graph? *Physica A: Statistical Mechanics and its Applications*. 387:2637-2652.

Kovtun, O., V.A. Tillu, W. Jung, N. Leneva, N. Ariotti, N. Chaudhary, R.A. Mandyam, C. Ferguson, G.P. Morgan, W.A. Johnston, S.J. Harrop, K. Alexandrov, R.G. Parton, and B.M. Collins. 2014. Structural insights into the organization of the cavin membrane coat complex. *Dev. Cell*. 31:405-419.

Lajoie, P., J.G. Goetz, J.W. Dennis, and I.R. Nabi. 2009. Lattices, rafts, and scaffolds: domain regulation of receptor signaling at the plasma membrane. *J. Cell Biol.* 185:381-385.

Liu, Q., L. Chen, H.C. Aguilar, and K.C. Chou. 2018. A stochastic assembly model for Nipah virus revealed by super-resolution microscopy. *Nature communications*. 9:3050.

Ludwig, A., G. Howard, C. Mendoza-Topaz, T. Deerinck, M. Mackey, S. Sandin, M.H. Ellisman, and B.J. Nichols. 2013. Molecular composition and ultrastructure of the caveolar coat complex. *PLoS biology*. 11:e1001640.

Ludwig, A., B.J. Nichols, and S. Sandin. 2016. Architecture of the caveolar coat complex. *J. Cell Sci.* 129:3077-3083.

Monier, S., R.G. Parton, F. Vogel, J. Behlke, A. Henske, and T.V. Kurzchalia. 1995. VIP21-caveolin, a membrane protein constituent of the caveolar coat, oligomerizes in vivo and in vitro. *Mol. Biol. Cell*. 6:911-927.

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Newman, M.E.J. 2003. The structure and function of complex networks. *SIAM review*. 45:167-256.

Newman, M.E.J. 2006a. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*. 74:036104.

Newman, M.E.J. 2006b. Modularity and community structure in networks. *Proceedings of the national academy of sciences*. 103:8577-8582.

Newman, M.E.J. 2013. Spectral methods for community detection and graph partitioning. *Physical Review E*. 88:042822.

Parton, R.G., and M.A. del Pozo. 2013. Caveolae as plasma membrane sensors, protectors and organizers. *Nat. Rev. Mol. Cell Biol.* 14:98-112.

Pelkmans, L., and M. Zerial. 2005. Kinase-regulated quantal assemblies and kiss-and-run recycling of caveolae. 436:128-133.

Rothberg, K.G., J.E. Heuser, W.C. Donzell, Y.-S. Ying, J.R. Glenney, and R.G.W. Anderson. 1992. Caveolin, a protein component of caveolae membrane coats. *Cell*. 68:673-682.

Sargiacomo, M., P.E. Scherer, Z. Tang, E. Kubler, K.S. Song, M.C. Sanders, and M.P. Lisanti. 1995. Oligomeric structure of caveolin: implications for caveolae membrane organization. *Proc. Natl. Acad. Sci. USA*. 92:9407-9411.

Shroff, H., C.G. Galbraith, J.A. Galbraith, and E. Betzig. 2008. Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nat Meth*. 5:417-423.

Sinha, B., D. Koster, R. Ruez, P. Gonnord, M. Bastiani, D. Abankwa, R.V. Stan, G. Butler-Browne, B. Vedie, L. Johannes, N. Morone, R.G. Parton, G. Raposo, P. Sens, C. Lamaze, and P. Nassoy. 2010. Cells respond to mechanical stress by rapid disassembly of caveolae. *Cell*. 144:402-413.

Stoeber, M., P. Schellenberger, C.A. Siebert, C. Leyrat, A. Helenius, and K. Grunewald. 2016. Model for the architecture of caveolae based on a flexible, net-like assembly of Cavin1 and Caveolin discs. *Proc. Natl. Acad. Sci. USA*. 113:E8069-E8078.

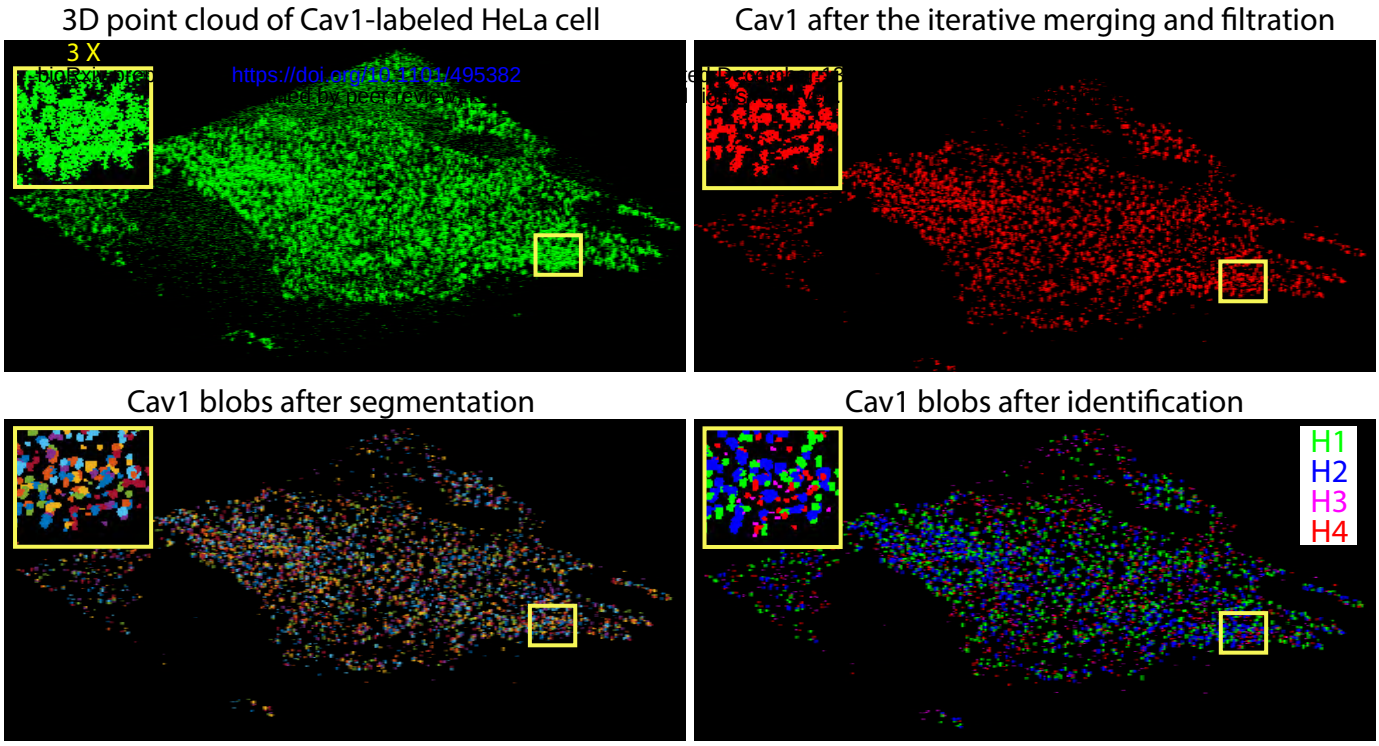
Tafteh, R., L. Abraham, D. Seo, H.Y. Lu, M.R. Gold, and K.C. Chou. 2016a. Real-time 3D stabilization of a super-resolution microscope using an electrically tunable lens. *Opt. Express*. 24:22959-22970.

Tafteh, R., D.R. Scriven, E.D. Moore, and K.C. Chou. 2016b. Single molecule localization deep within thick cells; a novel super-resolution microscope. *J Biophotonics*. 9:155-160.

Thompson, R.E., D.R. Larson, and W.W. Webb. 2002. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* 82:2775-2783.

Walser, P.J., N. Ariotti, M. Howes, C. Ferguson, R. Webb, D. Schwudke, N. Leneva, K.J. Cho, L. Cooper, J. Rae, M. Floetenmeyer, V.M. Oorschot, U. Skoglund, K. Simons, J.F. Hancock, and R.G. Parton. 2012. Constitutive formation of caveolae in a bacterium. *Cell*. 150:752-763.

A



B

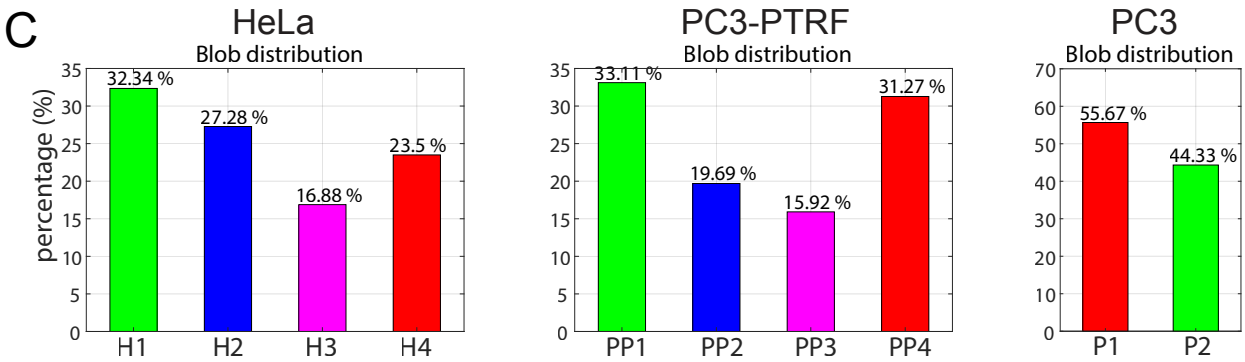
HeLa groups

PC3-PTRF groups	H1	H2	H3	H4
PP1	**26.869**	123.27	107.18	45.268
PP2	74.54	**35.535**	192.44	134.82
PP3	90.559	186.51	43.483	**30.717**
PP4	110.84	204.22	**12.241**	52.278

Matched groups

PC3	PC3-PTRF	HeLa	Blobs
P2	PP1	H1	Scaffold S2
	PP2	H2	Caveolae
	PP3	H4	Scaffold S1B
P1	PP4	H3	Scaffold S1A

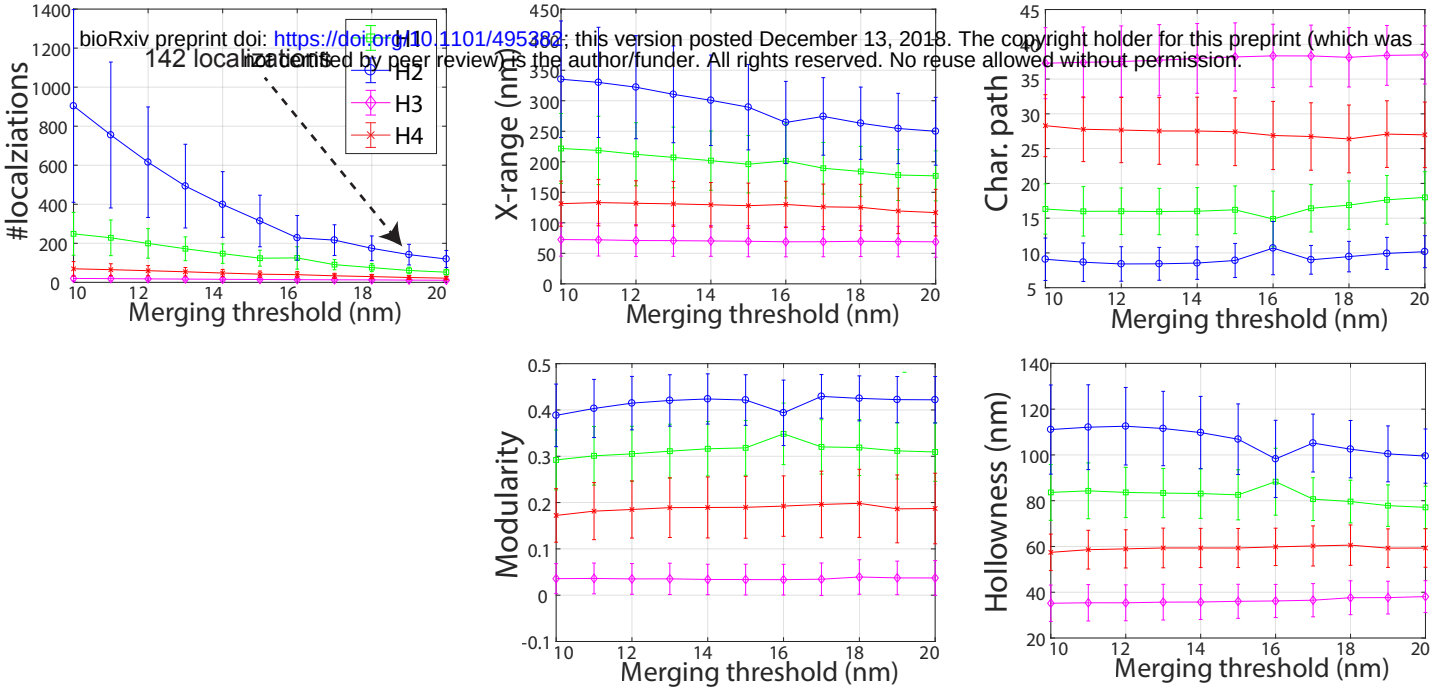
C



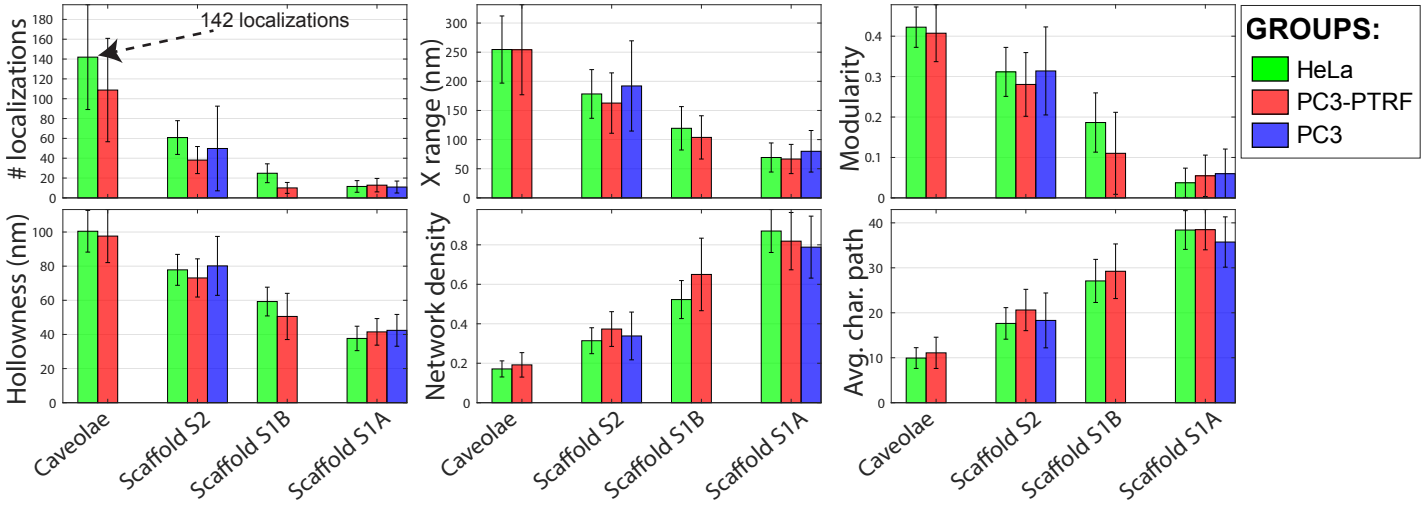
Khater et al, Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A **Biological signatures with varying merge proximity thresholds (MPTs)**

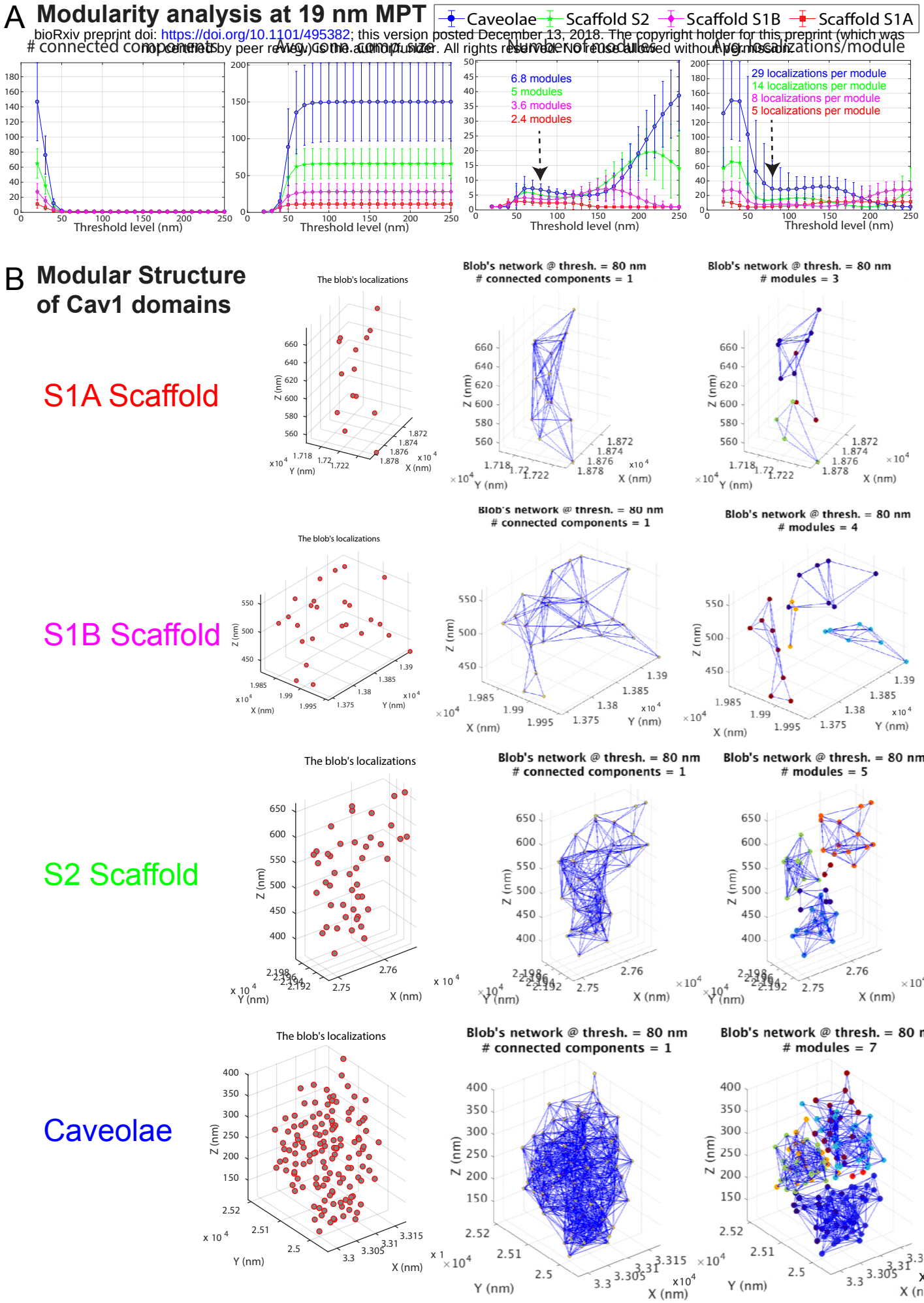


B **Signatures after group matching**



Khater et al, Figure 2

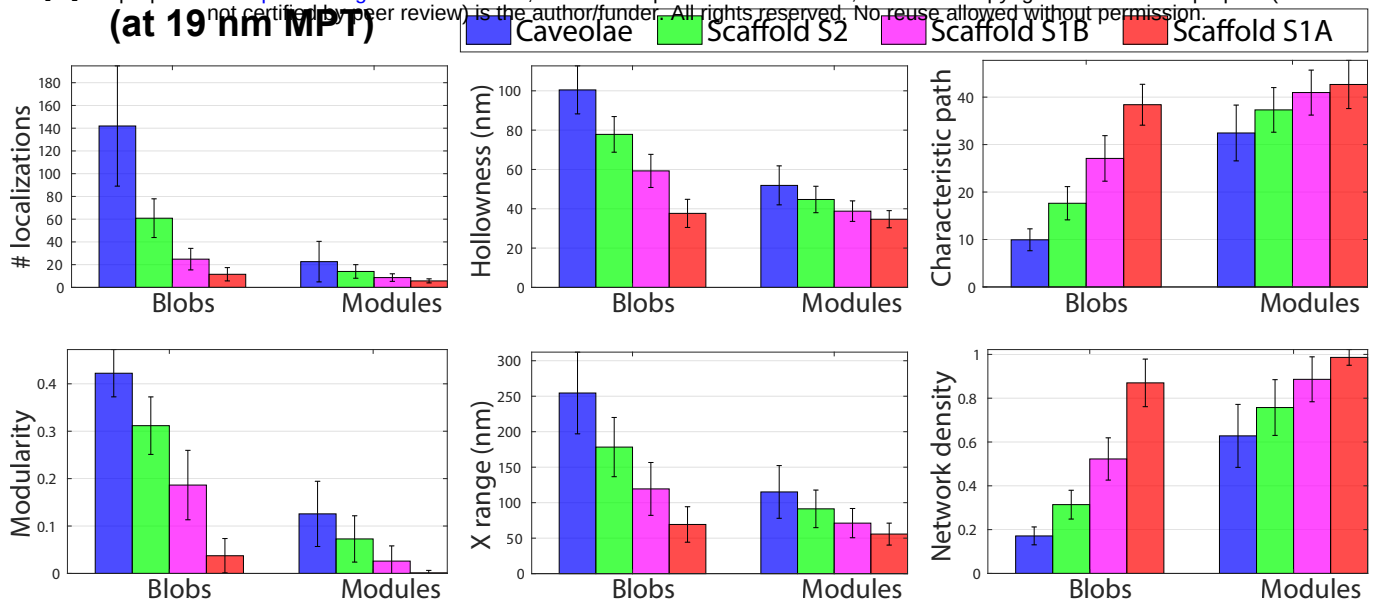
bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Khater et al, Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A **Modules vs Blobs (at 19 nm MPT)**



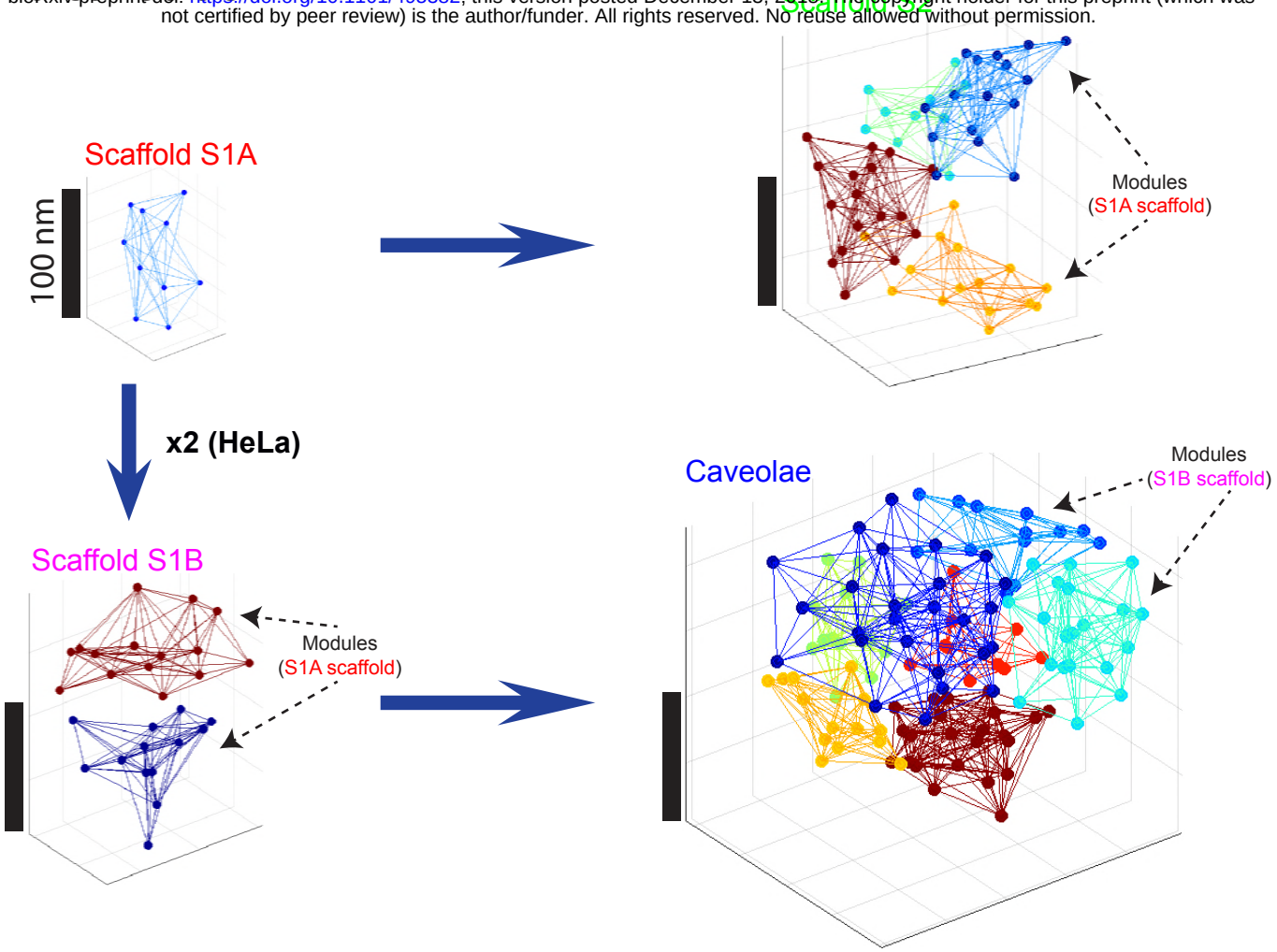
B **Module and blob matching**

Blobs \ Modules	Scaffold S2	Caveolae	Scaffold S1A	Scaffold S1B
Scaffold S2	102.73	79.955	137.91	122.2
Caveolae	198.42	176.92	230.48	216.39
Scaffold S1A	**19.519**	41.233	**23.483**	**9.4866**
Scaffold S1B	47.184	**26.064**	83.455	66.677

Khater et al, Figure 4

Scaffolds are modules that interact to form larger scaffolds and caveolae

bioRxiv preprint doi: https://doi.org/10.1101/495382; this version posted December 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Khater et al, Figure 5