

Nanoscale spatial-omics via contrastive embedding of single-molecule localisation data

S. Shirgill^{1,2}, L.G. Jensen³, D.J. Nieves¹, M.C. Wales², A. Kaminer^{4,5}, K. Savoye^{1,2,6}, R. Peters⁷, M. Heilemann^{4,5}, R. Henriques⁸, S.F. Lee⁹, P. Rubin-Delanchy¹⁰ and D.M. Owen^{1,2,11*}

¹*Department of Immunology and Immunotherapy, School of Infection, Inflammation and Immunology, College of Medicine and Health, University of Birmingham, UK*

²*School of Mathematics, College of Engineering and Physical Sciences, University of Birmingham, UK*

³*Department of Mathematics, Aarhus University, Denmark*

⁴*Institute of Physical and Theoretical Chemistry, Goethe-University Frankfurt, Germany*

⁵*International Max Planck Research School (IMPRS) on Cellular Biophysics, Frankfurt am Main, Germany*

⁶*School of Physics and Astronomy, College of Engineering and Physical Sciences, University of Birmingham, UK*

⁷*School of Mathematical and Physical Sciences, University of Sheffield, UK*

⁸*Instituto de Tecnologia Química e Biológica António Xavier, Universidade NOVA de Lisboa, Portugal*

⁹*Yusuf Hamied Department of Chemistry, University of Cambridge, UK*

¹⁰*School of Mathematics, University of Edinburgh, UK*

¹¹*Centre of Membrane Proteins and Receptors, University of Birmingham, UK*

*correspondence to d.owen@bham.ac.uk

Abstract: Omics approaches have revolutionised biology, and cells can now be routinely characterised on the genomic, transcriptomic and proteomic levels. However, there is an additional pillar; the (nanoscale) spatial organisation of molecules in the cell – information now accessible through super-resolution microscopy. We present a contrastive learning framework for nanoscale spatial-omics that embeds single-molecule localisation microscopy data into a latent space representing protein architecture directly to enabling comparative analysis. Using simulated and experimental data, we demonstrate its ability to enable new bioanalysis capabilities including assessing changes to cellular nanoscale architecture arising from pharmacological treatments, cell type, fluorophore selection or data-processing workflows. The approach supports downstream tasks such as clustering proteins by nanoscale organisation, mapping dose–response trajectories and identifying batch effects in replicate datasets, establishing contrastive learning as a scalable foundation for nanoscale spatial-omics and providing a platform for comparative phenotyping, quality control, and hypothesis generation.

Introduction

Omics approaches enable comprehensive characterisation of biological systems at the molecular level. One of the most well-established is genomics, where rapid sequencing technology and curated public databases have enabled large-scale GWAS studies, personalised medicine, insights into human evolution, microbiome research and forensic science [1, 2]. Single-cell transcriptomics has provided insights into developmental biology and the discovery of new cell types and states [3], while protein sequence databases enabled AlphaFold's breakthrough in structural biology [4].

Beyond describing cells at the genomic, transcriptomic and proteomic levels, there is a fourth pillar of cellular characterisation – the spatial arrangement of those molecules in the cell. Understanding this organisation is ultimately the motivation behind fluorescence microscopy – one of the most ubiquitous research techniques in cell biology. Despite its utility however, conventional fluorescence microscopy provides no information on length scales shorter than around 200 nm. Developments in super-resolution imaging, especially point-localisation approaches such as single-molecule localisation microscopy (SMLM), MINFLUX and the use of gold nanoparticles with EM have enabled protein organisation to be studied at much finer lengths. Within this landscape, SMLM is widely applied and well suited to examining the spatial structure of proteins at the nanoscale and on the level of individual molecules [5-8] and wide range of analysis methods have been deployed to describe nanoscale protein organisation. SMLM analysis approaches include spatial statistics (e.g., Ripley's K-function, pair-correlation), clustering algorithms (e.g., DBSCAN, Voronoi tessellation), and convolutional neural network (CNN)-based approaches for cluster detection, drift correction, and single-particle tracking [9–20]. However, these methods often hypothesis-driven, requiring users to decide which features of the data to analyse and interpret. They are also always applied to one dataset at a time, making them unsuitable for comparative, global or exploratory analysis (**Figure 1a**).

Contrastive learning bypasses these limitations by using a hypothesis-agnostic approach and embedding datasets into a unified feature space that captures nanoscale protein architecture directly (**Figure 1b**). Contrastive learning is a self-supervised approach that embeds structurally similar data closer together and dissimilar data further apart, thereby capturing underlying organisational patterns without requiring manual annotation [21]. This makes it particularly appropriate for SMLM, where ground-truth labels are rarely available. The framework generates a compact numerical representation of each dataset (an embedding) in the context of all other datasets being evaluated. In contrast to supervised CNNs, contrastive learning extracts generalisable features of nanoscale organisation (e.g., spatial arrangements, densities, fibre-like vs. clustered motifs) that transfer across experiments.

We use contrastive learning to embed SMLM datasets into a high-dimensional latent space where structurally similar protein distributions naturally cluster, enabling systematic comparisons at scale. The approach parallels strategies in single-cell transcriptomics, where high-dimensional expression profiles are first embedded into a latent space and subsequently explored with dimensionality reduction tools such as UMAP for visualisation [22]. Our framework therefore provides a functional foundation for *nanoscale spatial-omics* by offering a scalable, machine learning-based approach to compare nanoscale protein architectures across conditions. The framework is integrated with community-driven databases to support accessibility, reproducibility, and benchmarking [23]. Nanoscale spatial-omics therefore offers a foundation for the systematic, comparative analysis of protein spatial organisation, opening new opportunities for cell phenotyping, drug discovery, and the development of integrative models of cellular organisation. A key motivation for using contrastive learning is that it produces a vector representation of point patterns in which Euclidean geometry is meaningful, allowing

biologically relevant similarities and differences to be quantified directly in latent space. Traditional approaches such as DBSCAN or Ripley's K provide only scalar descriptors that are difficult to interpret geometrically and confound cross-experiment comparisons. By embedding ROIs into a unified feature space, our framework overcomes these limitations and establishes a scalable, generalisable approach for comparative nanoscale spatial-omics.

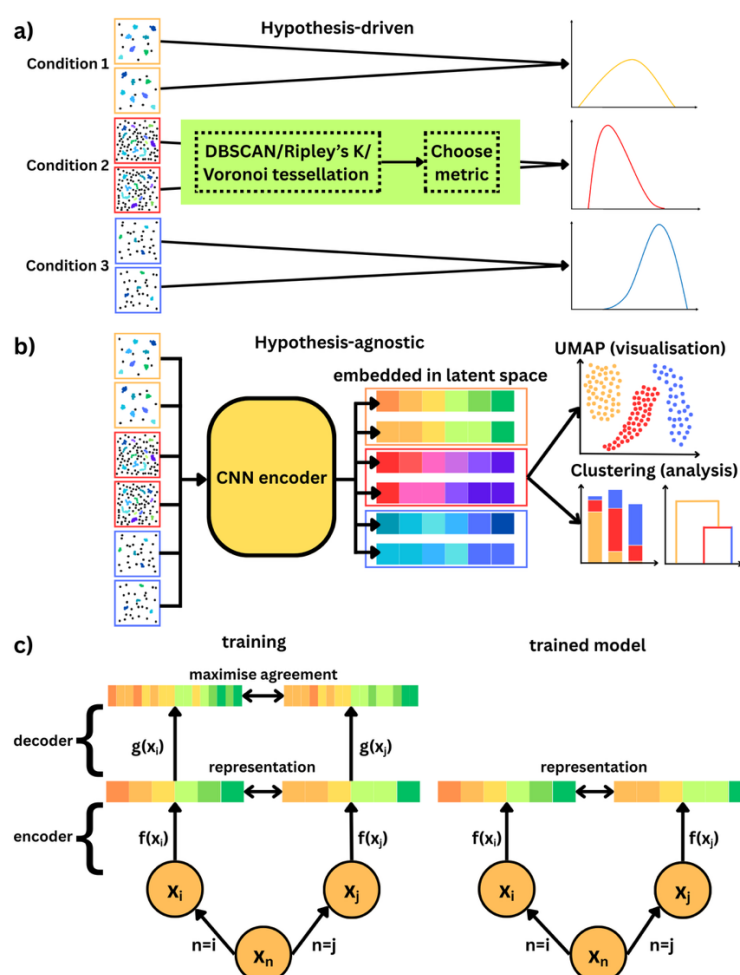


Figure 1: Overview of the contrastive learning framework *a)* Traditional SMLM analysis pipelines often rely on clustering or spatial statistics applied to each dataset individually. Comparisons across datasets are limited and usually restricted to comparing histograms of selected descriptive metrics. *b)* Contrastive learning is hypothesis-agnostic and global: It treats all datasets collectively, embedding them into a shared high-dimensional latent space that captures structural similarity *c)* Machine-learning workflow: during training, structurally similar dataset pairs are passed through the encoder and decoder networks, and their representations are optimised to maximise agreement (minimise distance). After training, the decoder is removed, and the encoder alone is used to embed new datasets for downstream analysis.

Results

To study the nanoscale organisation of proteins, we pre-process SMLM datasets into standardised $3 \times 3 \mu\text{m}$ regions of interest (ROIs) for input into a contrastive learning framework. Each ROI is passed through a convolutional neural network (CNN) encoder and embedded into a latent space. In this space, ROIs with similar protein architectures are positioned closer together, while greater distances reflect larger differences in nanoscale organisation. This latent

representation enables a range of downstream analyses, including clustering or ROIs in the latent space to assess whether ROIs from different experimental conditions form distinct groups, or whether overlap suggests shared structural characteristics. Additionally, distances between clusters can be computed to quantify differences in protein organisation across conditions. To visualise the embedded ROIs, we apply UMAP [24], to reduce the latent space to two dimensions, where random seeds were fixed for reproducible results throughout the training and analysis. Compared to PCA, UMAP performs non-linear dimensionality reduction, enabling it to capture complex relationships in the data while providing a faithful low-dimensional representation for visualisation.

The CNN encoder was trained using an adapted SimCLR framework (**Figure 1c**) [21], which is well suited for learning structural representations from unlabelled point-cloud data, as it avoids the need for manual annotations and encourages the embeddings to emphasise underlying organisational patterns rather than irrelevant variation such as rotation. We chose SimCLR over alternative contrastive approaches (e.g., MoCo) because of its simplicity and proven performance across diverse domains [21]. Training data was generated using a Perlin noise-based simulator in Python, which produces pseudo-random point patterns from a fixed set of parameters. Perlin noise was chosen because it efficiently generates a wide range of spatial structures, such as fibrous and clustered patterns, that reflect the types of protein organisations commonly observed in SMLM. Pairs of point patterns generated from the same parameters (termed positive pairs) are passed through the encoder, followed by a decoder (projection head). During training, the contrastive loss is minimised (see methods). This explicitly reduces the distance between embeddings of positive pairs while maximising their separation from negative pairs. After training, the projection head is discarded, and the encoder alone is used to embed experimental ROIs into the latent space.

To evaluate the model's ability to embed ROIs based on spatial similarity, we simulated 50 $3 \times 3 \mu\text{m}$ ROIs with varying numbers of randomly-placed Gaussian clusters which were passed through the contrastive learning framework and embedded into a 128-dimension latent space. The 2-dimensional UMAP projection (used for visualisation only) was coloured by condition and showed that most conditions formed distinct clusters, indicating successful separation by the model (**Figure 2a**). To quantify clustering performance, we applied K-means clustering in the 128-dimensional latent space, specifying the number of clusters to match the number of conditions and illustrated this analysis with a second UMAP plot that shows the data coloured by cluster ID. We then analysed the condition composition of each cluster and calculated their normalised Shannon entropies (0–1, where 0 indicates a pure cluster composed of a single condition and higher values reflect increasing diversity). ROIs with 5 or 20 Gaussian clusters were well grouped and dominated their respective latent-space clusters, with normalised entropies of 0.135 and 0 respectively. By contrast, ROIs with 9 and 10 Gaussian clusters largely merged into a shared K-means cluster, with an entropy of 0.43.

To assess structural differences between conditions, we computed pairwise distances between K-means cluster centroids in the 128-dimensional latent space. Each cluster was labelled according to its dominant condition. Hierarchical clustering was performed using average linkage, which provides a balanced representation of both local and global relationships between clusters. The dendrogram in **Figure 2a** shows that clusters corresponding to ROIs with 2 Gaussian clusters were tightly grouped and furthest in distance from all other conditions whereas those with 5, 10, and 20 clusters were more similar to each other. This suggests that the contrastive learning framework captures meaningful gradients in spatial organisation and is especially effective at distinguishing coarse structural differences. Additional simulations were

performed to explore how well the framework distinguishes between other parameters, including cluster size and localisation density (Supplementary Figure S1).

To further assess the model's ability to distinguish differences in nanoscale spatial organisation, we simulated ROIs containing both Gaussian clusters and linear fibre-like structures. The relative fibre density, defined as the number of localisations per fibre relative to each cluster, was varied. The ROIs were embedded into the contrastive learning latent space and visualised using UMAP (Figure 2b). As fibre density increased, the embeddings formed a smooth, continuous trajectory across the UMAP space, reflecting a gradient in spatial organisation. K-means clustering was again applied in the high-dimensional latent space, and the condition composition of each cluster was analysed. Finally, we computed pairwise distances between K-means cluster centroids and visualised their relationships using a dendrogram. Fibre density conditions grouped according to structural similarity: lower-density ROIs clustered together, as did the high-density conditions. These findings demonstrate the model's ability to encode continuous variation in spatial composition.

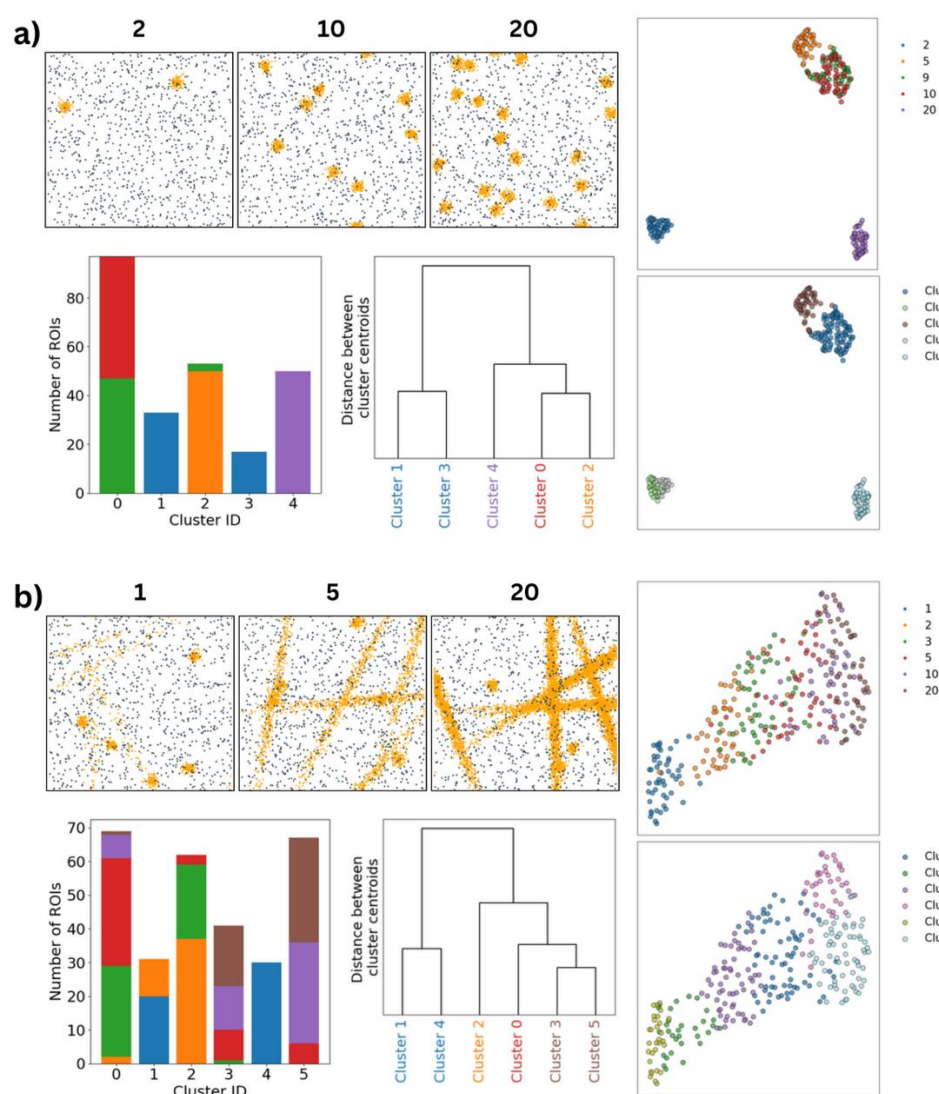


Figure 2: Contrastive learning enables discrimination of simulated point patterns based on nanoscale organisation. a) 50 Simulated $3 \times 3 \mu\text{m}$ ROIs were generated with varying numbers of Gaussian clusters per ROI, embedded into a latent space using the contrastive learning

framework. Data shows exemplar ROIs, cluster composition in the latent space and dendrograms showing the distance between clusters. Also shown are the UMAP 2D embedding for visualisation, coloured by condition and cluster ID for comparison. b) ROIs were generated with 5 clusters and 5 fibres. Fibre density was varied from 1× (equal localisations per fibre and cluster) to 20× (fibres having 20× more localisations). The same contrastive embedding and clustering analysis was applied.

Next, we evaluated the ability of the contrastive learning framework to distinguish between experimental SMLM data corresponding to different proteins (**Figure 3a**). A selection of datasets was obtained from the *nano-org* public database of curated SMLM data [23] and included a range of fluorophores and cell types. The UMAP showed clear separation by protein type, with tubulin and actin forming distinct clusters on one side of the latent space, and KIR2DL1, TIGIT, and Lck clustering on the opposite side (**Figure 3b**). This indicates that the model effectively distinguishes fibrous from non-fibrous protein architectures in real data. K-means clustering achieved an Adjusted Rand Index (ARI) of 0.74 (95% CI [0.71, 0.78]) indicating high-performance. In this test the number of clusters for k-means was set to match the number of experimental conditions. To relax this assumption, we also applied the elbow method to the same data to estimate the optimal number of clusters, followed by K-means clustering (Figure S2). This approach identified three clusters: two composed predominantly of microtubules and actin, and a third containing the non-fibrous proteins KIR2DL1, Lck, and TIGIT.

To assess whether traditional approaches also captured the differences detected by the contrastive learning framework, we applied DBSCAN and Ripley's K-function to the non-fibrous datasets (TIGIT, KIR2DL1, and Lck). Both methods revealed large differences between TIGIT and the other proteins across most cluster descriptors, and only modest differences between Lck and KIR2DL1 (Figure S3). Using a two-sided Mann-Whitney test with correction for multiple comparisons, all comparisons involving TIGIT were highly significant ($p < 0.0005$), with the exception of average points per cluster when comparing TIGIT and Lck ($p = 0.0006$). In contrast, comparisons between Lck and KIR2DL1 showed no significant differences for number of clusters, average points per cluster, or peak radius. The average cluster area was only marginally significant ($p = 0.038$), while noise fraction and maximum deviation were both strongly significant ($p < 0.0005$). While these results are broadly consistent with the contrasts revealed by our framework, traditional metrics produce a disparate set of outputs (*e.g.* cluster counts, areas, radii, noise fractions) that can be contradictory and are difficult to synthesise into a coherent picture. By contrast, our embedding captures these multi-faceted differences within a single, unified representation, providing a holistic view of nanoscale organisation that offers a clear conceptual advantage.

In SMLM, users may (or may not) have controlled the expression, labelling or imaging parameters during the experiment and so the overall localisation density in the final datasets may (or may not) be biologically meaningful. To determine whether the differences we observed were driven purely by localisation density, we repeated the analysis after normalising all ROIs to a fixed density of 100 localisations/ μm^2 (**Figure 3c**). As expected, this increased condition overlap, with the ARI decreasing to 0.55 (95% CI [0.51, 0.59]), but the framework still successfully distinguished protein architectures based solely on nanoscale distributions. Although subsampling to fixed density can in principle introduce artefacts or bias against dense structures, we performed repeated random subsampling and found the overall K-means clusters and UMAP embeddings were stable.

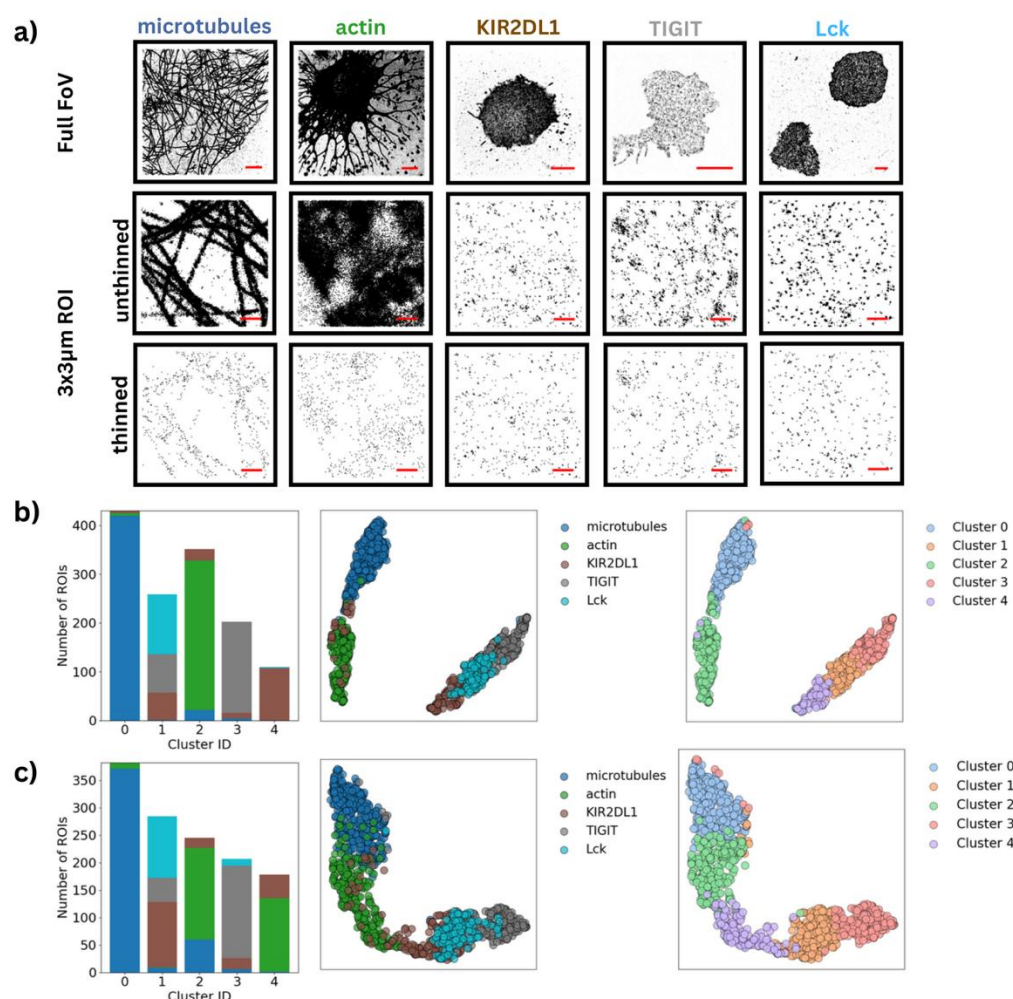


Figure 3: Contrastive learning distinguishes experimental protein distributions based on nanoscale organisation. *a)* Experimental SMLM data grouped by protein type. For each protein, a representative full field-of-view (FoV) image and example 3×3µm region of interest (ROI) are shown. Scalebars in full FoV and ROI SMLM images are 5µm and 0.5µm respectively. *b)* K-means cluster composition in the 128-dimensional space and UMAP plots for visualisation; each data point represents a ROI passed through the contrastive learning framework. *c)* Repeat analysis in which ROIs were normalised to 100 localisations/µm².

Next, we evaluated whether the contrastive learning framework could distinguish differences in protein organisation when varying the sample of interest (cell type) and experimental protocol (choice of fluorophore) (**Figure 4**). As a contrast condition, we used a structurally distinct protein, Lck, imaged in Jurkat E6.1 cells with AF647 as a “standard” condition to be included in all analysis. When comparing microtubules across different cell types (COS-7, HeLa, HEK), the model was able to separate conditions to some extent. ROIs from the Lck condition formed a distinct latent cluster with minimal overlap, as expected (normalised Shannon entropy of 0.11). In contrast, microtubules from all three cell types showed only partial separation were distributed across all three clusters. This pattern indicates that the fundamental architecture of microtubule networks is largely conserved across these adherent cell lines, with only subtle variations. The model is therefore correctly reporting a high degree of structural similarity, while still capturing fine-grained differences where they exist. This mixing is reflected in the high normalised Shannon entropies of the three clusters representing microtubules (0.57, 0.75, and 0.99 for clusters 0, 1, and 3, respectively) and in the relatively low ARI of 0.25 (95% CI [0.21, 0.3]),

despite the clear separation of Lck. When comparing microtubules stained with different fluorophores (AF647, CF568, CF660C), separation was even less pronounced. Clusters containing fluorophore data showed high normalised Shannon entropies (0.94, 0.97, and 0.56 for clusters 0, 2, and 3, respectively), and the ARI was lower still at 0.2 (95% CI [0.17, 0.24]). Excluding Lck, the ARI dropped further for both comparisons, by 0.12 (95% CI [0.08, 0.17]) for cell type and 0.11 (95% CI [0.08, 0.14]) for fluorophore, indicating that Lck's clear separation contributes substantially to overall clustering accuracy. This suggests that fluorophore choice has relatively little influence on nanoscale organisation compared with cell type, and both factors are substantially less impactful than the protein itself.

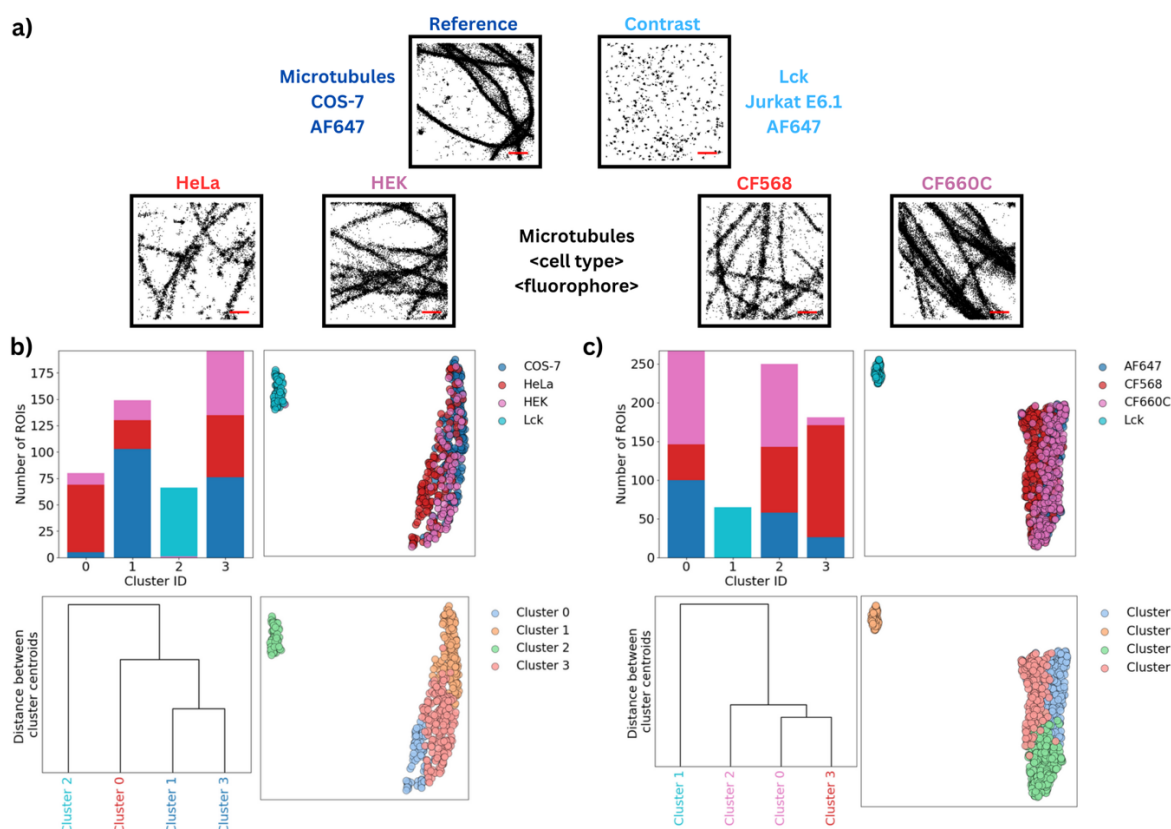


Figure 4: Contrastive learning distinguishes experimental conditions based on nanoscale protein organisation. Experimental SMLM data of microtubules were analysed across different conditions (representative ROIs are shown in **a**), where scalebars in SMLM images are 0.5 μm) including, cell types (COS-7, HeLa, HEK) and fluorophores (AF647, CF568, CF660C), where analyses are shown in **b**) and **c**) respectively. Microtubules in COS-7 cells stained with AF647 were used as the reference condition, while Lck in Jurkat E6.1 cells (AF647) served as a contrast condition representing a structurally distinct protein. ROIs were embedded using the contrastive learning framework, and K-means clustering was performed with the number of clusters set to match the number of conditions.

Additional analysis is presented in Supplementary Figure S4 where biological replicates of the same experimental condition are compared. This revealed potential anomalies, with one cluster composed primarily of data from a single biological replicate, suggesting possible batch effects or experimental variation. For clusters containing biological replicate data, all clusters showed high normalised Shannon entropies, however cluster 9 had the lowest entropy (0.59), suggesting a potential batch effect or experimental anomaly. This highlights an important strength of the

framework: beyond comparative analysis, it also functions as a quality control tool, capable of flagging anomalous replicates or batch effects. Such built-in QC provides a valuable safeguard for ensuring the robustness and reproducibility of SMLM studies. Anomalies can also arise from differences in data processing workflows. For example, processing the same raw data with different localisation algorithms in SMAP or ThunderSTORM showed that the radial symmetry method produced outputs that diverged from centroid, Gaussian, or integrated Gaussian fitting. Across clusters containing data from multiple algorithms, most had high normalised Shannon entropies (>0.9), whereas cluster 4, dominated by the radial method, had a markedly lower entropy of 0.27. These differences likely stem from the radial algorithm's geometric assumptions. This example illustrates another strength of the framework: its ability to act as a benchmark for analytical reproducibility, highlighting when outputs from alternative workflows diverge and flagging cases where methodological choices may bias downstream interpretation.

Next, we investigated the spatial organisation of microtubules in COS-7 cells treated with increasing concentrations of nocodazole (**Figure 5a**), a drug known to disrupt microtubule polymerisation. As nocodazole concentration increased from 0 to 1 $\mu\text{g/ml}$, the fibre networks became increasingly fragmented, with the highest concentration leaving only sparse, disconnected fibres. The UMAP projection reveals separation between untreated cells, intermediate (0.1 $\mu\text{g/ml}$) and high-dose treatment (1 $\mu\text{g/ml}$), suggesting that the contrastive learning framework captures a graded change in nanoscale organisation. Cluster 1 was composed almost entirely of untreated cells (normalised Shannon entropy = 0.03). Cells treated with 0.1 $\mu\text{g/ml}$ predominantly populated cluster 0, with some overlap with untreated cells (entropy = 0.42). Cells treated with 1 $\mu\text{g/ml}$ were the dominant condition in cluster 2, although this cluster showed substantial overlap with other conditions, particularly 0.1 $\mu\text{g/ml}$ cells, reflected in its higher entropy (0.88), ARI: 0.49 (95% CI [0.44, 0.56]). To explore whether the contrastive learning framework can resolve subtle differences in actin organisation, we applied it to experimental SMLM data of actin acquired in Jurkat E6.1 T cells forming early or late immunological synapses (**Figure 5b**). ROIs were stratified by both synapse stage (early vs late) and spatial region (centre vs periphery), yielding four experimental conditions. The UMAP projection revealed partial separation of these conditions in latent space, with greater distinction observed between synapse stages than between spatial regions. K-means clustering in the latent space supported this trend, with normalised Shannon entropies exceeding 0.8 for all clusters except cluster 3 which was dominated by late: periphery with an entropy of 0.57. These findings indicate that the contrastive learning framework can sensitively detect changes in actin architecture associated with both synapse maturation and subcellular localisation.

Finally, we explored whether the contrastive learning framework could track subtle changes in the nanoscale organisation of membrane receptors. We focused on the epidermal growth factor receptor (EGFR), a receptor tyrosine kinase that is predominantly monomeric in the resting state and dimerises upon binding its ligand EGF. We recorded SMLM data from HeLa cells labelled for EGFR under resting conditions and following EGF stimulation for 5 or 15 minutes. Separation was observed between resting and EGF-treated states: cluster 0 contained predominantly resting EGFR (normalised Shannon entropy = 0.11), while cluster 1 was enriched for EGF-treated cells (entropy = 0.64; **Figure 5c**). Cluster 2 comprised a mixture of all three EGFR conditions (entropy = 0.94), but with a higher proportion of resting and 5-minute treatments than 15-minute treatments. Overall clustering accuracy was modest (ARI = 0.33, 95% CI [0.3, 0.37]), but these results nonetheless demonstrate that the framework can detect treatment-dependent shifts in EGFR organisation. Crucially, the framework quantitatively captures the progression of receptor reorganisation. The dendrogram shows that the largest structural change is untreated cells compared to the treated cells, confirming the model's ability to resolve subtle, treatment-dependent shifts in molecular patterning.

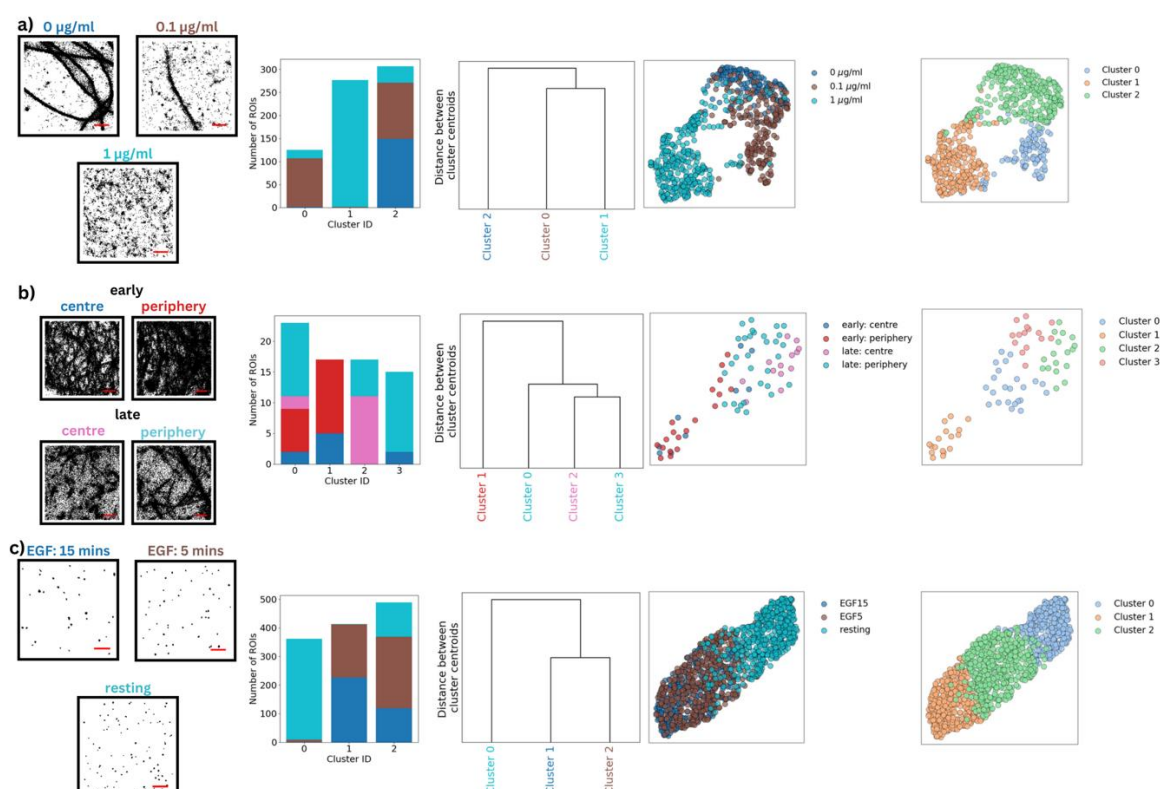


Figure 5: Contrastive learning distinguishes between experimental conditions based on nanoscale organisation. *a)* Experimental SMLM data of microtubules in COS-7 cells treated with increasing concentrations of nocodazole. *b)* Experimental SMLM data of actin imaged in early or late immunological synapses of Jurkat E6.1 cells. *c)* Experimental SMLM data of EGFR in HeLa cells, either untreated (resting) or treated with EGF for 5 or 15 minutes. ROIs were embedded into latent space using the contrastive learning framework. K-means clustering was applied with the number of clusters set to match the number of treatment conditions. Clustering performance was visualised via UMAP projections and assessed by calculating the condition composition per cluster, and inter-cluster distances to quantify separation in the latent space. Scalebars in SMLM images are 0.5 μm .

Discussion

In this study, we present a contrastive learning framework for embedding SMLM point cloud data into a latent space that reflects nanoscale spatial organisation. By using a self-supervised approach, our method learns meaningful representations of protein architecture without requiring manual annotation or prior knowledge of biological labels. We demonstrate that this framework can sensitively and systematically distinguish both coarse and subtle structural features across a variety of conditions. We have shown that the framework can consistently distinguish fibrous and clustered conditions, separate datasets based on protein nanoscale organisation, evaluate the effects of cell type, fluorophore choice and localisation algorithm on output data and accurately monitor drug dose-responses. Despite these strengths, there are several limitations to consider. The current analysis is restricted to two-dimensional ROIs. Results may also be influenced by fluorophore-specific properties and labelling strategies, which can bias apparent nanoscale distributions. In addition, the fixed ROI size ($3 \times 3 \mu\text{m}$) balances resolution and localisation counts but may not be optimal for all protein architectures.

Embeddings generated by the framework could serve as features for cell-type classification, suggesting that nanoscale protein maps may contribute to cell phenotyping in a way comparable to transcriptomic profiles. In principle, composite embeddings from multiple key proteins could be combined to define a cell's nanoscale spatial state. This would support a multidimensional classification of cells that extends beyond traditional markers, providing a complementary layer of information alongside gene expression-based approaches.

The framework provides a foundation for a new field of nanoscale spatial-omics: the systematic study of protein nanoscale organisation. Along side the three fundamental pillars of cellular characterisation: genomics, transcriptomics and proteomics, nanoscale spatial-omics has long been the ultimate goal of fluorescence microscopy. This is now becoming feasible due to a) SMLM allowing nanoscale data on protein architecture, with single-protein specificity b) curated, public databases of high-volume SMLM data and c) the systematic analytical and comparative tools presented here, enabled by statistical advances and machine learning.

Together with data standardisation, imaging automation, quality control, and high-throughput SMLM, the presented framework provides a foundation for nanoscale cell-atlasing. This will advance our understanding of cellular spatial organisation and help the community generate hypotheses, test predictions, and model cellular processes more effectively. At the same time, the framework is sensitive to technical differences arising from biological replicates, localisation algorithms, and imaging setups, underscoring the need for rigorous standardisation in nanoscale spatial-omics. Establishing robust acquisition protocols, metadata reporting, benchmarking practices, and normalisation strategies (e.g., density equalisation, subsampling, or calibration with reference datasets) will ensure that the framework not only drives biological discovery but also safeguards data quality by identifying and flagging technical biases that might otherwise confound interpretation.

Methods

Contrastive learning

Contrastive learning was developed based on the SimCLR framework [21]. A convolutional neural network (CNN) was used as the encoder in a contrastive learning framework. The encoder consisted of four convolutional layers followed by one fully connected layer. Convolutional layers used kernel sizes of 3×3 or 2×2 with increasing feature depths (from 1 to 8 to 16 channels), and all layers employed the Mish activation function. Layer normalisation was applied to the fully connected layers, while batch normalisation was used after convolutional layers to stabilise feature distributions and improve training convergence. Dropout with a probability of 0.2 was applied after selected convolutional and linear layers to reduce overfitting. Average pooling and flattening were used after the final convolutional block to reduce dimensionality while preserving spatial structure. The encoder outputs a fixed-dimensional latent embedding that captures the nanoscale organisation of each ROI, which can then be used for downstream analysis.

The decoder, used only during training, comprised three linear layer, projecting from the 128-dimensional latent space to a 512-dimensional output. A sigmoid activation was used to scale the output. This projection encouraged the encoder to learn embeddings that capture fine-grained structural differences among point patterns.

The model was trained on simulated point pattern data generated with Perlin noise using the *FractalPerlin2D* function from the *pyperlin* Python package. Similar pairs were created by randomly sampling from shared generative parameters such as point density and morphological

features. No explicit data augmentations were applied; instead, following the SimCLR framework, two point patterns generated from the same Perlin noise parameters were treated as augmented views of the same condition. In total, 15,000 pairs (30,000 point patterns) were generated, with 80% used for training and 20% reserved for evaluation, assigned at random.

Training was conducted using the Adam optimiser starting with a learning rate of 3×10^{-5} and a weight decay of 1×10^{-6} . The model was trained with a batch size of 32 for up to 100 epochs. We monitored validation loss each epoch and used plateau-based early stopping with patience = 3 in conjunction with a learning-rate reduction on plateau (factor 0.5, floor 1×10^{-6}). In our data, validation loss decreased rapidly within the first 10–15 epochs and then stabilised; stopping typically occurred at 30–40 epochs, well into the plateau, where additional training yielded <1% loss change and no material change in downstream metrics (UMAP structure, k-means composition, ARI, entropy). Model hyperparameters: learning rate, batch size and weight decay, as well as model architecture were optimised to minimise the loss. An adapted version of the NT-Xent loss function was used, in which the similarity scores of negative pairs were weighted by the difference in their generative parameters. This weighting reduces the contribution of harder-to-distinguish (i.e., parameter-similar) negatives during training, allowing the model to focus on separating truly distinct patterns in the embedding space. The loss function used was:

$$L(z_i, z_j) = -\log \left(\frac{\exp\left(\frac{\text{sim}(z_i, z_j)}{\tau}\right)}{\exp\left(\frac{\text{sim}(z_i, z_j)}{\tau}\right) + \sum_{k \neq i, j}^{2N} \exp\left(\frac{\text{sim}(z_i, w_{ik} z_k)}{\tau}\right) + \sum_{k \neq i, j}^{2N} \exp\left(\frac{\text{sim}(z_j, w_{jk} z_k)}{\tau}\right)} \right),$$

Where:

z_i, z_j : vectors associated with similar pairs of images i and j respectively.

z_k : vector associated with dissimilar images k .

$w_{ik} = \text{par}_i - \text{par}_k$, where $\text{par}_{i,k}$ are the parameters used to generate point patterns for i and k respectively. Note that point pattern j has the same parameters as i .

$\text{sim}(\cdot)$: is the cosine similarity between two vectors.

Analysis

Cluster analysis was performed on data in the latent space using KMeans, where the number of clusters was set to the number of conditions. Distances between cluster centroids, normalised Shannon entropies and ARI's were calculated. Adjusted Rand Index (ARI) confidence intervals were estimated using nonparametric bootstrapping. For each bootstrap replicate, paired sets of condition and cluster labels were resampled with replacement, preserving their correspondence. The ARI was recalculated for each resampled dataset across 1,000 replicates, and the 2.5th and 97.5th percentiles of the bootstrap distribution were taken as the bounds of the 95% confidence interval. For visualisation, embeddings were projected into two dimensions using UMAP with parameters $\text{min_dist} = 0.1$ and $n_neighbors = 15$. A range of parameter values was tested, but results were stable within the recommended ranges.

Additional cluster analysis was performed on the non-fibrous proteins shown in Figure 3 (TIGIT, NKp30, and Lck). DBSCAN was applied with parameters *minimum points* = 5 and $\epsilon = 25$, and the following descriptors were calculated: number of clusters, noise fraction, average number of points per cluster, and average cluster area. Ripley's K analysis was also performed, from which peak radius and maximum deviation were extracted. Statistical comparisons of metrics from both methods were performed using two-sided Mann-Whitney tests with Bonferroni correction for multiple comparisons.

Data Simulations

To model protein distributions that exhibit clustering behaviour, we generate Gaussian-distributed clusters within a $3 \times 3 \mu\text{m}^2$ ROI. Each simulation follows these steps:

1. Cluster Generation: A specified number of clusters (n) are randomly positioned within the ROI.
2. Point Distribution: Each cluster contains p localisations, which are sampled from a Gaussian distribution centred at the cluster position.
3. Cluster Variability: The spread of each cluster is controlled by the standard deviation (σ), determining the tightness of clustering.
4. Background Density: Unclustered localisations are present with a density of b , which is proportional to the number of point per cluster, p .

This approach ensures that clusters of varying sizes and densities can be systematically compared using the dissimilarity algorithm.

To simulate a mixture of protein distributions with clustering behaviour and nanoscale organisation of cytoskeletal-like structures, we generate linear fibre distributions as well as Gaussian-distributed clusters within an ROI. The process involves both defining the fraction of clusters to fibres as well as defining the density of fibres.

Software

All analyses were performed in Python using PyTorch for contrastive learning.

Experimental methods

dSTORM microtubule data in Figures 4,5, and S3 as well as Lck data were acquired in-house. DNA-PAINT EGFR data in Figure 5 was kindly provided by Alexandra Kaminer from the Heilemann lab. Experimental data for these datasets are described below, all other experimental data is available on nano-org, with experimental methods available through their DOIs.

dSTORM experimental data

Cell culture

COS-7, HeLa, and HEK293 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, high glucose; Sigma-Aldrich), supplemented with 10% fetal bovine serum (FBS; Gibco, Life Technologies), 1% penicillin-streptomycin (Gibco, Life Technologies), and 1% L-glutamine (Gibco, Life Technologies). Jurkat E6.1 cells were cultured in RPMI 1640 (Sigma-Aldrich) with the same supplement concentrations. All cells were maintained at 37 °C in a humidified incubator with 5% CO₂.

For imaging, adherent cells were seeded at a density of 1×10^4 cells per well into eight-well μ -slides (Ibidi, glass bottom) one day prior to fixation. HEK wells were pre-treated with fibronectin (Sigma-Aldrich) (1:100 dilution in phosphate-buffered saline (PBS); Gibco) for 30 minutes before cell seeding. Jurkat cells were seeded at 2×10^5 cells per chamber in six-channel μ -slides (Ibidi, glass bottom) pre-coated with anti-CD3/CD28 (1mg/ml; Invitrogen). Cells were activated at 37 °C

for 5 minutes, then fixed in 4% paraformaldehyde (PFA; Sigma-Aldrich) for 10 minutes and washed three times with PBS. These samples were used for Lck imaging.

Sample preparation

For microtubule disruption experiments, COS-7, HeLa, and HEK cells were incubated in DMEM containing 0.1 $\mu\text{g/mL}$ or 1 $\mu\text{g/mL}$ nocodazole (Tocris) for 30 minutes at 37 °C. Cells were then washed three times with PBS. Untreated controls were PBS-washed directly.

To preserve microtubule architecture, cells were subjected to sequential extraction and fixation. Cells were first extracted for 90 seconds at 37 °C using a pre-warmed solution of 0.25% Triton X-100 (Sigma-Aldrich) and 0.1% glutaraldehyde (Sigma-Aldrich) in PEM buffer (80 mM PIPES (Bioworld), 5 mM EGTA (ThermoFisher Scientific), 2 mM MgCl_2 (Invitrogen), pH 6.8; (Invitrogen)), followed by fixation for 10 minutes at 37 °C in 0.25% Triton X-100 and 0.5% glutaraldehyde in PEM. After fixation, cells were quenched with 1 mg/mL sodium borohydride (Sigma-Aldrich) for 7 minutes and washed three times with PBS.

dSTORM imaging

Cells were permeabilised with 0.1% Triton X-100 in PBS for 3 minutes at room temperature and blocked with 5% bovine serum albumin (BSA; Sigma-Aldrich) for 30 minutes. For microtubule staining, cells were incubated with mouse monoclonal anti- β -tubulin IgG₃ (200 $\mu\text{g/mL}$; Santa Cruz Biotechnology), diluted 1:50 in 5% BSA, for 30 minutes at room temperature. After three PBS washes, cells were incubated in Alexa Fluor™ 647-conjugated goat anti-mouse IgG (2 mg/mL; Life Technologies), diluted 1:1000 in 5% BSA, for 30 minutes in the dark. If specified, CF568-conjugated goat anti-mouse IgG (2 mg/mL; Sigma-Aldrich) was used under the same conditions. For Lck staining, cells were incubated with Lck (D88) XP® Rabbit mAb (51 $\mu\text{g/mL}$; Cell Signalling Technology), diluted 1:400 in 5% BSA, for 30 minutes at room temperature. After washing three times with PBS, samples were incubated in Alexa Fluor™ 647-conjugated goat anti-rabbit IgG (2 mg/mL; Life Technologies), diluted 1:1000 in 5% BSA, for 30 minutes in the dark. Finally, all samples were washed ten times with PBS.

Immediately before dSTORM imaging, PBS was replaced with an imaging buffer containing 18% glucose, 10 mM Tris (pH 8), 50 mM NaCl, 0.8 mg/mL glucose oxidase, 50 mM cysteamine, and 40 $\mu\text{g/mL}$ catalase (all from Sigma-Aldrich).

Prior to dSTORM imaging, PBS was replaced with an imaging buffer consisting of 18% glucose, 10 mM Tris (pH 8), 50 mM NaCl (Sigma-Aldrich), 0.8 mg/mL glucose oxidase, 50 mM cysteamine (Sigma-Aldrich), and 40 $\mu\text{g/mL}$ catalase (Sigma-Aldrich).

All experiments were performed using an ONI Nanoimager S microscope unless otherwise stated. Where indicated, a Nikon N-STORM microscope was used for comparison.

Data analysis

Data analysis was conducted using the Super resolution Microscopy Analysis Platform (SMAP) [26], with default settings applied unless stated otherwise. Single-molecule localisations were fitted using the *PSF free* algorithm. To assess the impact of different localisation algorithms on similarity metrics, localisations were also fitted using the *ellipt:PSFx PSFy* or *PSF fix* algorithm in SMAP, as well as several alternative fitting methods available in ThunderSTORM [26], including *Gaussian*, *integrated Gaussian*, *centroid*, and *radial* fitters.

Localisations with an estimated precision >30 nm were excluded, consistent with common SMLM practice to remove low-confidence points and ensure reliable nanoscale resolution. Drift correction and grouping were performed to mitigate the effects of sample drift and multiple blinking of fluorophores.

ROIs were generated from full field-of-view localisation files by first defining a cell-bounding polygon and then dividing this region into 3 × 3 μm grids.

DNA-PAINT experimental data

Cell culture

The human cervical cancer cell line HeLa (# ACC 57, DSMZ, Braunschweig, Germany) was cultured in Dulbecco's modified Eagle's medium (DMEM) (# 11574486, Gibco, Life Technologies, Waltham, MA, USA) supplemented with 10% fetal bovine serum (# 35-079-CV, Corning Inc., Corning, NY, USA), 1 unit/mL penicillin and 1 µg/mL streptomycin (Gibco, Life Technologies) and 1% v/v GlutaMAX (# 35050-038, Gibco, Life Technologies). The cells were incubated at 37 °C with 5% CO₂ and were passaged every 3–4 days. HeLa cells were seeded in ibidi µ-Slide (# 80607, ibidi GmbH, Gräfelfing, Germany) coated with PLL-PEG-RGD (poly-L-lysine-grafted polyethylene glycol modified with a CGRGDS peptide) at the density of 1x10⁵ cells/mL for one day growth.

Sample preparation

The cells were starved for 2 hours prior stimulation using the introduced growth media without fetal bovine serum. After starvation, EGF (# AF-100-15, PeproTech, Thermo Fisher Scientific, Waltham, MA, USA) was diluted in the serum free media to end concentration of 100 ng/mL and then incubated for either 5 or 15 min. Both, resting and stimulated cells were fixed using 3% formaldehyde (FA) (# 28908, Sigma-Aldrich, St. Louis, MO, USA) with 0.25% glutaraldehyde (GA) (# G5882, Sigma) and incubated for 15 min at 37 °C. For EGFR labeling, a monoclonal primary antibody (1:50 dilution, #sc-120, Santa Cruz Biotechnology, Texas, USA) was pre-incubated with a secondary nanobody (2-fold excess) conjugated to the R3 docking strand (Massive Photonics, München, Germany) for 1 h at 4 °C in the blocking buffer (1 mM EDTA, 0.02% Tween 20, 0.05% NaN₃, 2% bovine serum albumin (BSA, # 9048-46-8, Carl Roth GmbH & Co. KG, Karlsruhe, Germany), 0.05 mg/mL salmon sperm DNA in PBS). Fixed cells were blocked in the blocking buffer for 20 min and incubated with the pre-incubated antibody/nanobody mixture for 2 h at room temperature. Gold beads (# A11-100-NPC-DIH-1-100, Nanopartz Inc., Loveland, US) were diluted to 1:5 in PBS and added to each well for 15 min, followed by washing with PBS. The cells were post-fixed using 4% FA.

DNA-PAINT imaging

DNA-PAINT imaging was performed on a home-built widefield setup based on a custom-built widefield microscope based on a Nikon Eclipse Ti inverted microscope. Excitation was provided by a 561 nm laser (200 mW Sapphire, Coherent Inc., Santa Clara, CA, USA) with laser power modulated via an acousto-optic tunable filter (AOTFnc-400.650-TN, AA Opto Electronic, France). To ensure a clean beam profile, the laser was fiber-coupled using a collimator (60FC-4-M6.2-33) into a polarization-maintaining single-mode optical fiber (PMC-E-400RGB), and then re-collimated to a 6 mm full width at half maximum (FWHM) beam (60FC-T-4-M50L-01; all from Schäfter & Kirchhoff GmbH, Germany). The collimated beam was expanded using a telescope (AC255-030-A-ML and AC508-150-A-ML, Thorlabs GmbH, Germany) and focused onto the back focal plane of a 100× TIRF oil immersion objective (CFI Apochromat TIRF 100XC Oil, Nikon, Japan). A motorized mirror (MTS50-Z8, Thorlabs) enabled adjustment of the illumination angle for widefield, HILO, or TIRF imaging modes. Axial focus was stabilized using Perfect Focus System (Ti-PFS, Nikon), while lateral sample positioning was controlled by a motorized stage (Ti-S-ER, Nikon) in combination with a piezo stage (Nano-Drive, MadCityLabs, USA). Excitation light was introduced into the microscope via a multiband dielectric beamsplitter (zt405/488/561/640rpc, AHF Analysentechnik, Germany), which also directed emission light into the detection path. Fluorescence was spectrally filtered with a bandpass filter (610/60 ET, Chroma) and imaged with an EMCCD camera (iXon Ultra DU-897U-CS0, Andor, Northern Ireland).

Prior to imaging, the R3 imager strand (conjugated to Cy3B; Massive Photonics) was diluted to a final concentration of 2 nM in an imaging buffer containing 5 mM Tris/HCl (pH 8.0), 75 mM MgCl₂, and 0.05% Tween-20, freshly supplemented with an oxygen scavenging and triplet-state

quenching system (1× PCA, 1× PCD, and 1× Trolox), following a published protocol [27]. DNA-PAINT imaging was conducted using a 561 nm laser at an excitation intensity of 0.1 kW/cm². All microscope components were controlled via the μManager software [28]. Image stacks of 20,000 frames were recorded with the following parameters: 100 ms exposure time, EM gain of 150, 3× preamplifier gain, 10 MHz readout rate, image size of 256×256 pixels, and frame transfer mode activated. Cell positions were saved within μManager to facilitate subsequent multi-target imaging. Bright-field images were acquired before and after each DNA-PAINT acquisition.

Data analysis

Image processing was performed using Picasso software [27]. First, single emitters in each frame were localized by fitting the Maximum Likelihood Estimation for Integrated Gaussian parameters. Next, drift correction was performed using fiducial markers. Localized single molecule events were filtered for width of the point spread function (PSF), localizations which appeared in multiple consecutive frames were merged with parameters based on the NeNA (nearest neighbor based analysis) value which represents experimental localization precision: radius of 1 times NeNA and 4 min. dark frame. Next DBSCAN (density-based spatial clustering and application with noise) clustering was performed using 1xNeNA and 12min.sample. The identified clusters were further filtered based on the mean frame time within a range of $\mu - 2\delta$ to $\mu + 2\delta$, where μ represents the average mean frame time and δ is the standard deviation, and the δ within a range from 1500 to 8000 frames.

Acknowledgements

The research described in this paper was carried out with the assistance of Advanced Research Computing at the University of Birmingham. This included support from the Research Software Group to convert the original academic code into a reusable Python package, support with the development of the database and website, data storage on the Research Data Store, computations on the BlueBEAR HPC service, and use of a BEAR Cloud Virtual Machine to host the website." Data was acquired in the COMPARE Advanced Imaging Facility at the University of Birmingham. DMO and SS acknowledge funding from the Biotechnology and Biological Sciences research Council (BBSRC) grant BB/X018644/1. RH received funding from the European Research Council (ERC) through grant 101001332-SelfDriving4DSR and Horizon Europe through grants 101057970-AI4LIFE and 101099654-RTSuperES. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This work was also supported by a European Molecular Biology Organization (EMBO) installation grant (EMBO-2020-IG-4734 to RH).

Author Contributions

LGJ, SS and MCW developed the contrastive learning framework. SS, DJN, MCW, AK, KS, MH and RP provided simulations, experimental data and testing. LGJ, PR-D, RH, SFL and DMO conceived the work. SS and DMO wrote the manuscript.

Code and data availability

All experimental data is stored and available for download on <https://nano-org.bham.ac.uk>. The core analysis functionality and algorithms, including trained models, and simulated data used are available at <https://gitlab.bham.ac.uk/owendz-protein-databank/smlm-contrastive-learning>.

References

1. Gish, W., and States, D.J. Identification of protein coding regions by database similarity search. *Nature Genetics* **3**(3), p. 266-272 (1993).
2. Berman, H.M., *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**(1), p. 235-242 (2000).
3. Tang, F., *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods* **6**(5), p.377-382 (2009).
4. Tabula Muris Consortium, *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**(7727), p.367-372 (2018).
5. Lelek, M., *et al.* Single-molecule localization microscopy. *Nature Reviews Methods Primers* **1**(1), p. 39 (2021).
6. Zhou, Y., *et al.* Membrane potential modulates plasma membrane phospholipid dynamics and K-Ras signaling. *Science* **349**(6250), p. 873-876 (2015).
7. Truong Quang, B.A., *et al.* Extent of myosin penetration within the actin cortex regulates cell surface mechanics. *Nature communications* **12**(1), p. 6511 (2021).
8. Svitkina, T.M. Actin cell cortex: structure and molecular organization. *Trends in cell biology* **30**(7), p. 556-565 (2020).
9. Speiser, A., *et al.* Deep learning enables fast and dense single-molecule localization with high accuracy. *Nature methods* **18**(9), p.1082-1090 (2021).
10. Nehme, E., *et al.* DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning. *Nature methods* **17**(7), p.734-740 (2020).
11. Khater, I.M., *et al.* Super resolution network analysis defines the molecular architecture of caveolae and caveolin-1 scaffolds. *Scientific reports* **8**(1), p.9009 (2018).
12. Fordjour, F.K., *et al.* Exomap1 mouse: A transgenic model for in vivo studies of exosome biology. *Extracellular Vesicle* **2**, p.100030 (2023).
13. Bender, S.W.B., *et al.* SEMORE: SEgmentation and MORphological fingErprinting by machine learning automates super-resolution data analysis. *Nature Communications* **15**(1), p.1763 (2024).
14. Khater, I.M., *et al.* Caveolae and scaffold detection from single molecule localization microscopy data using deep learning. *PLoS One* **14**(8), e0211659 (2019).
15. Williamson, D.J., *et al.* Machine learning for cluster analysis of localization microscopy data. *Nature communications* **11**(1): p.1493 (2020).
16. Saavedra, L.A., Mosqueira, A. and Barrantes, F.J. A supervised graph-based deep learning algorithm to detect and quantify clustered particles. *Nanoscale* **16**(32): p.5308-15318 (2024).
17. Nieves, D.J., *et al.* A framework for evaluating the performance of SMLM cluster analysis algorithms. *Nature methods* **20**(2), p.259-267 (2023).
18. Kowalek, P., *et al.* Boosting the performance of anomalous diffusion classifiers with the proper choice of features. *Journal of Physics A: Mathematical and Theoretical* **55**(24): p.244005 (2022).
19. Martinez, Q., *et al.* Sequence-to-sequence change-point detection in single-particle trajectories via recurrent neural network for measuring self-diffusion. *Transport in Porous Media* **147**(3), p.679-701 (2023).
20. You, B. and Yang G. International Joint Conference on Neural Networks (IJCNN) (pp. 1-6). *IEEE* (2021).
21. Chen, Ting, *et al.* A simple framework for contrastive learning of visual representations. *International conference on machine learning*. PmlR (2020).
22. Becht, E., *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* **37**(1), p. 38-44 (2019).
23. Shirgill, S. *et al.* Nano-org, a functional resource for single-molecule localisation microscopy data. *Nature Communications* **16**, 8674 (2025).
24. McInnes, L., Healy, J. and Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv, 1802.03426* (2018).
25. Ries, J. SMAP: a modular super-resolution microscopy analysis platform for SMLM data. *Nature Methods* **17**(9), p. 870-872 (2020).

26. Ovesný, M., *et al.* ThunderSTORM: a comprehensive ImageJ plugin for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**(16), p. 2389-2390 (2014).
27. Schnitzbauer, Joerg, et al. "Super-resolution microscopy with DNA-PAINT." *Nature protocols* 12.6 (2017): 1198-1228.
28. Edelstein, Arthur D., et al. "Advanced methods of microscope control using µManager software." *Journal of biological methods* 1.2 (2014): e10.