GRAPH ATTENTION NETWORK GENERATES SUPER-RESOLUTION SPATIAL TRANSCRIPTOMIC DATA

Luis Alonso
Department of Electrical Engineering
University of Navarra

Mikel Hernaez*
Centro de Investigación Médica Aplicada (CIMA)
Instituto de Ciencia de los Datos e Inteligencia Artificial (DATAI)
University of Navarra

Idoia Ochoa†
Department of Electrical Engineering
Instituto de Ciencia de los Datos e Inteligencia Artificial (DATAI)
University of Navarra

ABSTRACT

Spatial transcriptomics allow uncovering the spatial origin of RNA molecules within a tissue slide. However, acquiring informative signals remains a challenge, with a trade-off between sequencing depth and spatial resolution. Sequencing-based technologies provide unbiased transcriptional profiles, but their minimal workable spatial unit compromises more than one cell. While deconvolution methods have been proposed to estimate the cell-type composition of each spot, we believe that the resolution can be further improved by generating enhanced subpots in place of the original spots, such that cells in the original spot can be allocated into the newly generated subspots. We propose a computational method based on graph and attention learning, named Square, that leverages message passing for information sharing between neighboring spots. Even though this rearrangement of cells can be solely spatially approximated, a resolution enhancement is achieved. We show that the proposed approach is capable of deciphering the composition of spots in ST and Visium samples, whilst imputing sparse profiles and amplifying the signal in them. Newly generated subspots have been empirically and biologically validated. The gap between both spatial transcriptomic modalities is then closed, generating high-throughput cellular-scale outputs.

1 INTRODUCTION

The advent of spatial transcriptomics has enabled the precise mapping of gene expression profiles within tissue architecture (Moses & Pachter, 2022), providing key insights into cellular composition and thus, significantly advancing our understanding of the molecular signatures that drive disease progression and therapeutic responses (Piwecka et al., 2023; Arora et al., 2023). While imaging-based spatial transcriptomic technologies, which leverage the *in situ* hybridization of transcripts, can yield subcellular transcriptional resolution, they are constrained to a few thousand genes(Rao et al., 2021). Sequencing-based technologies, on the other hand, offer an unbiased view of the transcriptional architecture of the tissue at a cost of cellular resolution, as their minimal spatial unit can contain several cells. To address this, several computational models, termed deconvolution models, have been developed for deciphering the spot composition of different cell types (Ma & Zhou, 2022; Andersson et al., 2020; Zhou et al., 2023; Lopez et al., 2022). However, recent benchmarking works depict a lack of consistency among deconvolution methods leading to striking differences in their performance across different datasets (Yan & Sun, 2023; de la Fuente et al., 2023).

The very recent release of Visium HD (Oliveira et al., 2025) has challenged the current trade-off between (sub)cellular resolution and unbiased transcriptome profiling by moving from Visium's

*mhernaez@unav.es
†iochoal@unav.es

$55\mu m$-diameter spots to Visium HD's $2 \times 2\mu m$ bins. However, its cost has increased by more than 65% (Smith et al., 2024). Additionally, the number of Unique Molecular Identifiers (UMIs) captured per bin is significantly lower, reducing the transcriptional information captured within bins, and hence leading to an increased sparsity of transcriptional signals (Kamel et al., 2025).

To overcome the current scarce signal in sequencing-based technologies, we conceptually extended deconvolution to enable spatial reallocation of transcripts. Specifically, our proposed self-supervised model Square leverages graph attention networks to deconvolve the initial spots into smaller spatially-resolved subspots, leading to an enhanced spatial resolution, while only using the original (supra-cellular) spatial information. We show that the enhanced subspots generated by Square: i) exhibit lower entropy cell-type proportions than the original spots, ii) better capture the spatial location of cells within the original spot; and iii) uncover key biological insights that were hidden in the original data.

All in all, we envision that the proposed model will guarantee a minimal spatial resolution on which different technologies can be systematically compared, hence reducing the effects of a progressive obsolescence of available data.

2 Towards spatial super-resolution in spatial transcriptomics

Problem statement. A sequencing-based spatial transcriptomic experiment generates a gene expression matrix $\mathbf{X} \in \mathbb{R}^{G \times S}$, containing the expression of G genes across S spots, and a coordinate matrix $\mathbf{Z} \in \mathbb{R}^{2 \times S}$, containing the coordinates (x, y) of each spot. Our goal is to generate a *high resolution (hr)* sample by generating k new subspots for each spot, producing a new expression matrix $\mathbf{X}_{hr} \in \mathbb{R}^{G \times kS}$ and corresponding coordinate matrix $\mathbf{Z}_{hr} \in \mathbb{R}^{2 \times kS}$, yielding an enhanced spatial map. In what follows, k is set to 9, so that an isotropic expansion of both x and y axes is guaranteed.

Self-contained spatial inference. To facilitate the application of the developed model to complex tissues, our goal is to solely use the provided spatial sample, i.e., not requiring additional information such as paired single-cell RNA sequencing (scRNA-Seq) data or the number of expected cell-types in the dataset. In this context, datasets conventionally include a standard microscopic image, most commonly hematoxylin and eosin (H&E)-stained. The introduction of subcellular resolution with Visium HD has further elevated the relevance of these images in downstream analyses, as segmented cells constitute the primary processing units for this technology. Although the field has begun to explore the use of such images for resolution enhancement, the degree to which pixel intensities are directly correlated with gene expression remains uncertain (Li et al., 2024). Therefore, we chose to rely exclusively on spatial location information, thus improving the usability of our method.

2.1 Proposed model: Square

Hypothesis. Tissue arrangement is not stochastic and generally subjected to biological patterns, such as immune infiltration or cell proliferation, as cells belonging to the same cell type typically assemble together (Shah et al., 2016; Stoltzfus et al., 2020; Russ et al., 2021). Evidence from previous studies demonstrating spatial autocorrelation of marker genes (Ma & Zhou, 2022) further substantiates this hypothesis. Square builds on this phenomenon to increase the resolution of the spatial sample by predicting the transcriptional profile of newly generated subspots.

Square is trained through a self-supervised learning approach. To generate the training set, a *pooling mask* is applied by grouping $k = 9$ initial spots into super-spots by summing the transcriptional signals of the individual spots (App. Fig. 4A). Note that the *pooling mask* is not applied on the edges where neighbors are not available. Hence, each training instance consists of the super-spot and its m neighboring spots ($m = 16$ by default), creating training data with available ground truth: the initial spots used in the *pooling mask*. During training, virtual nodes are used as proxy for the *higher-resolution* subspots (App. Fig. 4B). Finally, an attention mechanism aims to predict spatially heterogeneous groups of virtual nodes based on the most relevant neighboring information for improved subspot prediction.

Square infers unobserved subspots. At inference time, an original spot is spatially deconvolved by using its m surrounding neighbors ($m = 8$ by default; App. Fig. 4C). This leads to a distribution shift between training and inference as during training the central node has a significantly

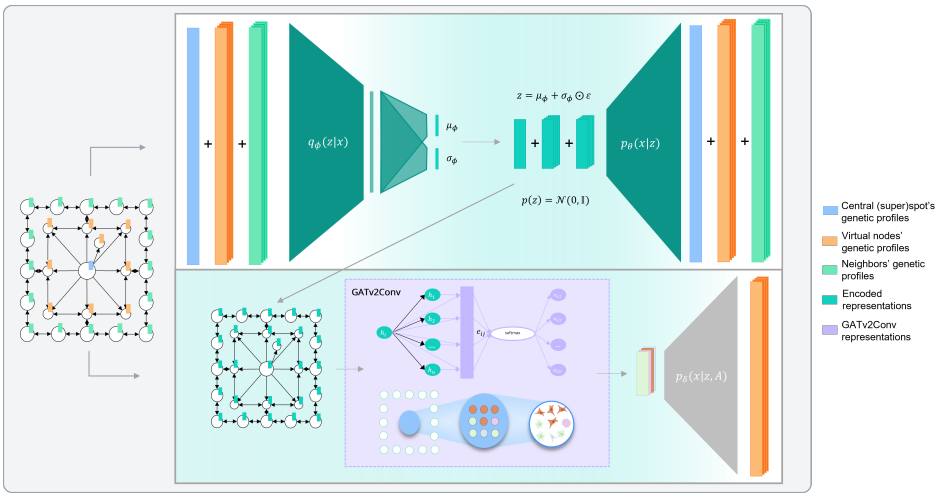

Figure 1: Overview of Square: the transcriptional profiles of the graph's nodes go through the VAE (top), and the encoded representations are used to initialize the node embeddings. Spatial dependencies are learned via message passing, generating meaningful embeddings (bottom); attention is implemented using GATv2Conv (purple box); the spatially-aware embeddings are decoded back to the gene space. The depicted graph is for training, but the same steps apply at inference time.

denser profile. Square addresses this through a variational autoencoder (VAE) (Kingma & Welling, 2013) that learns a latent representation of the spots on which the sparsity signal is regressed out (see below). Additionally, the difference in number of neighbors between training and inference is addressed via an inductive graph neural network that can handle a varying number of nodes (Fig. 1).

2.1.1 Spatial configuration using a graph-based representation

Each training data instance is a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ comprising a central spot and the surrounding m neighboring spots. Following recent works on learning on graphs (Qian et al., 2024; Pham et al., 2017), we introduce k virtual nodes, representing the enhanced subspots that are connected to both the central node and the original neighbors, as well as interconnected among themselves (App. Fig. 4B-C). To represent the original spatial structure, the adjacency matrix of $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is defined as:

$$A^*_{ij} = \begin{cases} d(s_i, s_j)^{-1} & \text{if spots } s_i \text{ and } s_j \text{ are both original spots or both virtual nodes} \\ t & \text{if } s_i \text{ and } s_j \text{ correspond to the central spot and a virtual node} \\ 1 & \text{if } s_i \text{ and } s_j \text{ correspond to a neighbouring spot and a virtual node} \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $d(s_i, s_j)$ is the Euclidean distance between the coordinates of (super)spots s_i and s_j. The coordinates of the virtual nodes are located between the central and neighboring nodes, specified in $\mathbf{Z} \in \mathbb{R}^{2\times S}$, to mirror the original topology. $t \geq 1$ is a model hyperparameter.

2.1.2 Addressing the distribution shift between training and inference data

The sparsity differences between (super)spots and the original spots may hinder the generalization capabilities of the model when inferring subspots. To address this distribution shift, we have implemented a variational autoencoder (VAE) that learns a smooth, low-dimensional representation of the input (super)spot data while accounting for biological and technical variability. Probabilistic frameworks are widely used in (single-cell) omics data to account for the uncertainty, sparsity, and noise inherent to these data (Lopez et al., 2018; 2022; Bergenstråhle et al., 2022).

Specifically, we use a standard VAE architecture (Lopez et al., 2018) where both encoder $q_\phi(\mathbf{z}|\mathbf{x})$ and decoder $p_\theta(\mathbf{x}|\mathbf{z})$ are parametrized as neural networks with non-linear activations, and the latent space is regularized towards a standard normal distribution. This architecture helps capture the dominant biological signal while smoothing over minor variations and batch effects that are not

critical for the task at hand (see App. A.4.3). In our set-up, the nodes from the graph $\mathcal{G}$ will have the VAE-encoded latent variables $\mathbf{z} \in \mathcal{R}^d$ as node features (d is set to 50% of the number of genes). Note that the encoding process is agnostic of the graph structure, so features of a spot are independent of its neighbors'. The graph $\mathcal{G}_{\mathbf{z}_\phi}$ is populated as follows:

$$\mathbf{z}_\phi = \{\mathbf{z}^{(i)}\}_{i=1}^{1+k+m} \sim q_\phi(\mathbf{z}^{(i)}|\mathbf{x}^{(i)}), \tag{2}$$

where k and m are the numbers of virtual nodes and the central spot's initial neighbors, respectively.

Virtual nodes' feature initialization. The transcriptomic profile of a virtual node is generated by averaging the transcriptomic profiles of its connected spots in the graph (this approach was selected experimentally, see App. A.4.1). The virtual nodes' profiles are then embedded into the latent space using the VAE encoder $q_\phi(\mathbf{z}|\mathbf{x})$.

2.1.3 Graph Attention Network

Square builds on the capacity of Graph Neural Networks (GNNs) to infer novel embeddings for each node by leveraging both the topology of the network as well as the features of neighboring nodes. Specifically, Square relies on a Graph Attention Network (GAT) (Brody et al., 2021), a specific type of GNN that computes an attention coefficient to infer neighbors' effect on the loss function. Using this architecture, uniform contribution of all neighbors $\mathcal{N}(i)$ of a node i is avoided, learning instead a weighted average of the representations of $\mathcal{N}(i)$. This upgrade is computed via an edge scoring function $e : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ that represents the importance of each one of the neighbors of a node:

$$e(\boldsymbol{h}_i, \boldsymbol{h}_j) = \text{LeakyReLU}\left(\boldsymbol{a}^{\text{T}} \cdot [\boldsymbol{W}_\delta \boldsymbol{h}_i \parallel \boldsymbol{W}_\delta \boldsymbol{h}_j]\right), \tag{3}$$

where $\boldsymbol{a} \in \mathbb{R}^{2d'}$ represents a new attention matrix, $\boldsymbol{W}_\delta \in \mathbb{R}^{d' \times d}$ stands for the standard GNN (learnable) parametric function, and $\parallel$ denotes concatenation. These edge scores (after normalization across all neighbors $j \in \mathcal{N}(i)$) are then multiplied by the matrix product of $\boldsymbol{W}_\delta$ and the neighbors' features $\boldsymbol{h}_j$. Ultimately, the update of node features after $l+1$ layers is computed as follows:

$$\boldsymbol{h}_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \text{softmax}_j\left(e\left(\boldsymbol{h}_i^{(l)}, \boldsymbol{h}_j^{(l)}\right)\right) \cdot \boldsymbol{W}_\delta \boldsymbol{h}_j\right). \tag{4}$$

As a last step, the updated embeddings of the virtual nodes are passed through a multi-layer perceptron (MLP) with weights δ to transform them back to the gene space.

2.2 Training

The VAE produces latent representations for all nodes (i.e., original spots, superspots, and virtual nodes), which are then used by the GAT to learn spatial dependencies across the graph. Given that the VAE (i.e., the encoder and decoder with parameters ϕ and θ, respectively) is trained jointly with the GAT (with parameters δ), the model needs to learn to maximize two log-likelihoods: one for the decoder distribution $p_\theta(\mathbf{x}|\mathbf{z})$, and another for the enhancement $p_\delta(\mathbf{x}|\mathbf{z}, \mathbf{A})$. Since the variational inference approach requires minimizing at the same time the distance between the encoded distribution $q_\phi(\mathbf{z}|\mathbf{x})$ and the prior $p(\mathbf{z})$, the function we maximize is:

$$\mathcal{L}_{\phi,\theta,\delta} = \mathcal{L}_{\phi,\theta}^{\text{recons}} + \mathcal{L}_{\phi,\delta}^{\text{deconv}} - \mathcal{L}_\phi^{\text{KL}}, \tag{5}$$

with

$$\mathcal{L}_{\phi,\theta}^{\text{recons}} = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})]; \mathcal{L}_{\phi,\delta}^{\text{deconv}} = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\delta(\mathbf{x}|\mathbf{z}, \mathbf{A})]; \mathcal{L}_\phi^{\text{KL}} = D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \| p(\mathbf{z})).$$

Note that $p_\delta(\mathbf{x})$ represents the probability of the observed spots' transcriptional profiles used to generate the superspot for ground truth availability. Instead, $p_\theta(\mathbf{x})$ is not restricted to a specific resolution level, but admits any genetic profile the model might process.

We use the Adam optimizer (Kingma & Ba, 2017) with a learning rate scheduler with a patience parameter of 10 epochs, and a batch size of 32.

Final configuration. Model architecture optimization was performed with Optuna (Akiba et al., 2019) for a total of 500 trials. The list of tested hyperparameters and the selected configuration are detailed in App. A.1.3. Square's pseudo-code is provided in App. A.1.2.

3 RELATED WORK

Current methods for increasing the resolution of spatial transcriptomic data—including only those that predict genetic profiles without using any reference—rely on different solutions:

BayesSpace (Zhao et al., 2021) is a Bayesian statistical method designed to perform clustering analysis of spatial transcriptomic data through the integration of spatial neighborhood information with a low-dimensional representation of gene expression. BayesSpace workflow starts with a spatial clustering; these clusters then undergo a refinement process, which aims to generate a more detailed spatial map. Subsequently, the refined clustering results are translated back into the gene expression space using regression models, with principal components serving as predictors.

CARD (Ma & Zhou, 2022) is a spatially-informed cell-type deconvolution method that, built upon a non-negative matrix factorization (NMF) framework, models cell-type distributions by incorporating spatial dependencies using a conditional autoregressive (CAR) model. This enables both the imputation of unobserved distributions at unmeasured locations and the construction of refined maps at higher resolution. The latter allows for a resolution enhancement, hence allowing a direct comparison to Square. Although it was crafted assuming the availability of a single-cell reference, a reference-free modality has been leveraged herein for the sake of fairness.

STAGE (Li et al., 2024) employs an autoencoder that maps the gene expression manifold to a location manifold, which can later be used to decode previously unseen locations. Initial spots are not altered, but new spots are generated in-between. This method was envisaged for several applications (e.g., recovery of down-sampled data and 3D generation) but is herein used for reconstructing non-sequenced regions (i.e., regions between spots) to mimic Square's operation. We note that the level of refinement cannot be modified, generating always 4 subspots where one was placed originally.

4 EVALUATION FRAMEWORK

4.1 DATASETS AND EVALUATION METRICS

Synthetic pseudo-phenotype. This dataset is derived from 5 distinct pseudo-phenotypes that emulate the role of cell-types (App. Fig. 8A,B). We generated a fully dense 90×90 square grid, mimicking that of ST. Each spot contains only one pseudo-phenotype, with its gene expression profile (dimension 100) sampled from the corresponding distribution (App. A.4.2). To generate a ground truth, groups of 3×3 spots (without overlap) are further aggregated into pseudospots. While this dataset may not be realistic in practice, it serves as a baseline for method evaluation and comparison.

Synthetic Mouse Olfactory Bulb (MOB). We use the statistical simulator scDesign3 (Song et al., 2024), using as reference a 10X Chromium scRNA-seq (Tepe et al., 2018) and a ST spatial RNA-seq (Ståhl et al., 2016) datasets, both from a MOB sample. The dataset comprises 278 spots, each with a gene expression vector of dimension 182, and containing a mixture of up to four cell-types: granule cell, periglomerular cell, mitral and tufted cell, and olfactory sensory neuron. While we do not have access to high-resolution spots in this case, we expect the reconstructed spots to maintain the cell-type proportions of the original ones.

Human Glioblastoma (GBM). We included a 10X Visium dataset of a human GBM tissue section (10x Genomics, 2020) compromising 3,468 spots with a median number of 4,326 genes.

Human Pancreatic Ductal Adenocarcinoma (PDAC). We considered a ST Human PDAC dataset (Moncada et al., 2020) containing 428 spots, each with approximately 20-70 cells and about 1,000 unique genes. The selected tissue section contains 20 cell-types and mainly four regions: cancer cells and desmoplasia, nonmalignant duct epithelium, stroma, and normal acini-rich pancreatic tissue.

Data preprocessing: Raw transcript counts undergo three main steps: i) normalization, to remove technical biases and adjust for differences in the sequencing depth of different spots; ii) logarithmization, to stabilize the variance across the range of expression values and reduce skewness; and iii) selection of the most highly variable genes (1,000 by default), to remove the inherent noise of house-keeping and uniformly-expressed genes, as well as to reduce the number of model parameters.

Evaluation metrics: We evaluated the performance of Square using a set of quantitative metrics across multiple scenarios, reporting results over five independent runs to ensure robustness. These

scenarios comprise: *i)* the synthetic pseudo-phenotype dataset, where comprehensive ground truth is available; *ii)* the MOB dataset, where only ground-truth proportions are available; and *iii)* real datasets, where no ground truth is available. Scenario *i)* is assessed using spot-wise distance metrics and the adjusted random index (ARI), computed with predicted and true cell-type assignments. The results of scenario *ii)* were evaluated applying k-means clustering to enhanced subspots and comparing the entropy of *cluster-frequency* vectors with that of true cell-type proportions; we further annotated clusters and assessed reconstruction accuracy via RMSE. For scenario *iii)*, we reproduced established biological findings and evaluated changes in cell-type purity produced by Square. See App. A.3 for more details.

5 RESULTS

Next, we provide an ablation study to showcase the contribution of the main components of Square and the results on the synthetic and GBM datasets. Results for PDAC are detailed in App. A.4.5.

5.1 ABLATION STUDY

To assess the contribution of three key components of Square to the overall performance, we conduct experiments: *i)* without virtual nodes, directly updating the original nodes and generating the reconstructed subspots by averaging the gene profiles (obtained by passing their updated embeddings through the MLP) of adjacent pairs of neighboring nodes; *ii)* without the VAE (not tackling the distribution shift between training and inference data), directly setting the gene expression values as the initial node features; and *iii)* without attention, using a standard message-passing GNN. For the assessment, we consider the *synthetic pseudo-phenotype* dataset, taking only superspots (for training) and spots (for inference) that have all neighbors available (i.e., edge spots are omitted). To better capture differences across models, we do not apply logarithmization to the data in this case.

All three ablated models underperformed Square across all metrics, except for the model without the VAE that obtained the lowest MSE (Table 9). Still, this model performs poorly on the other metrics, suggesting that the distribution shift produces a general worsening of the reconstruction if not considered thoroughly. It also exhibits the largest variance for all metrics. A notable result also emerges from visual inspection of the generated enhancements. The incorporation of attention learning introduces heterogeneity among subspot groups derived from an original spot. Without attention, the same weight matrix is propagated across all graph edges, leading to a lack of smoothness in the reconstructed expression (App. Fig. 8C). In contrast, with attention, the structural boundaries are preserved more faithfully, producing borders that closely resemble the original circular geometry rather than exhibiting a stepped or block-like pattern (App. Fig. 8D).

Table 1: **Ablation Study:** Performance comparison across model variants.

Model	MSE ↓	SCD ↓	$JD_{>\mu}$↓	JD_{30}↓	ARI ↑
w/o virtual nodes	1.74 ± 0.01	0.467 ± 0.011	0.475 ± 0.004	0.8140 ± 0.0003	0.52 ± 0.03
w/o VAE	$\mathbf{1.57 \pm 0.04}$	0.549 ± 0.030	0.550 ± 0.013	0.8169 ± 0.0017	0.44 ± 0.13
w/o attention	1.64 ± 0.03	0.517 ± 0.039	0.515 ± 0.026	0.8142 ± 0.0017	0.50 ± 0.16
Square	1.75 ± 0.01	$\mathbf{0.453 \pm 0.012}$	$\mathbf{0.469 \pm 0.005}$	$\mathbf{0.8133 \pm 0.0006}$	$\mathbf{0.58 \pm 0.07}$

5.2 SYNTHETIC PSEUDO-PHENOTYPE DATASET

We compare our model with BayesSpace, STAGE and CARD. Using the *synthetic pseudo-phenotype* data with the available ground truth, we compute the previously introduced (sub)spot-wise error and distance metrics, as well as the ARI between the predicted and true cell-types. Since STAGE generates four subspots per each initial spot but the ground truth consists of nine subspots, spot-wise metrics have been tailored by averaging the closest neighbors in the ground truth. Nevertheless, we do not compute ARI metrics since averaged profiles may not capture the spatial distribution of cells faithfully (e.g., boundaries between cell-type-specific regions would be oversmoothed).

Square generates more accurate enhanced subspots than the comparison methods, yielding a higher-resolution dataset (Table 2). Specifically, it obtains a significantly lower MSE, achieves the best

Table 2: **Benchmarking on the *synthetic pseudo-phenotype* dataset:** Performance comparison.

Method	MSE ↓	SCD ↓	$JD_{>\mu}$ ↓	JD_{30} ↓	ARI ↑
BayesSpace	2.33 ± 0.00	**0.410 ± 0.000**	0.410 ± 0.000	0.8133 ± 0.0002	0.00 ± 0.00
STAGE	40.18 ± 67.57	0.682 ± 0.325	0.512 ± 0.129	0.8194 ± 0.0074	-
CARD	9.66 ± 0.00	0.431 ± 0.002	0.450 ± 0.001	0.8132 ± 0.0005	0.00 ± 0.00
Square	**0.32 ± 0.08**	0.440 ± 0.012	**0.408 ± 0.005**	**0.8131 ± 0.0018**	**0.77 ± 0.08**

results in JD, albeit on par with other methods, and while it does not lead in SCD, its performance is comparable to the top-performing method. In terms of ARI, Square's inferred subspots showcase a partial agreement with the true cell-type assignments, while the other methods are no better than a random assignment. Delving more into the obtained results, we observed that CARD returns very low expression values that stand far apart from the initial data range, although they keep biological information in relative terms. The latter explains the high MSE it returns compared to BayesSpace and Square. STAGE also produces very high MSE values and lacks robustness overall, showing the highest standard deviation values for all metrics. This may be due in part to the disparity between the number of initial spots and the generated ones. BayesSpace is the method with the closest performance to Square's, but is still outperformed. The most remarkable result herein stands at the cell-type assignment comparison (ARI metric), which portrays the exclusive capability of Square to generate feasible subspots that accurately approximate the cell types in the ground truth, while at the same time reconstructing the spatial structures in the dataset.

5.3 Synthetic Mouse Olfactory Bulb (MOB) data

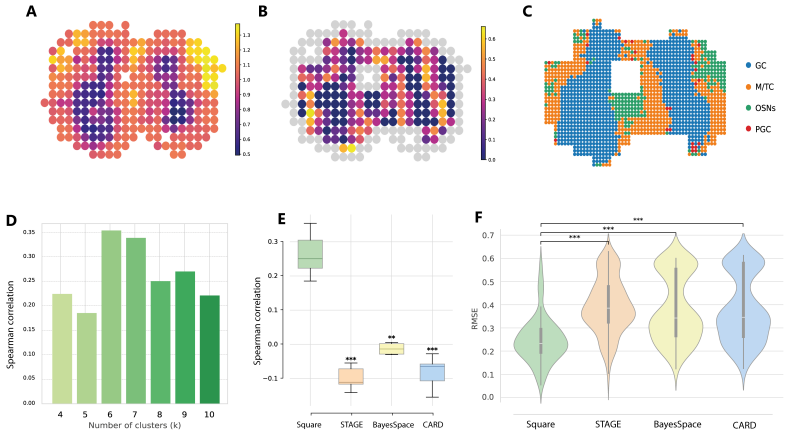

Figure 2: **A.** Cell-type abundance of the MOB synthetic dataset. **B.** Spatial entropy computed from true cell-type proportions. For Square: **C.** Cluster-frequency entropy for $k = 10$; **D.** Annotated enhanced sub-spots; **E.** Spearman correlation between the entropies of the *cell-type proportion* and *cluster-frequency* vectors for different levels of coarseness k. **F.** Box plots of the Spearman correlation for $k \in \{4, 10\}$. **G.** RMSE distribution of true cell-type and inferred proportions across spots.

The *synthetic MOB* dataset, which more realistically mimics a ST dataset (see App. A.2.1), contains two interior areas with lower entropy spots (computed from the cell-type proportions), corresponding to the areas with mainly GC cells (Fig. 2A), and higher entropy spots located in the exterior. The entropy of the spots after enhancement is computed on the cluster-frequency vectors after applying k-means (see Fig. 2C for spots' entropies with $k = 6$). Only spots for which all neighbors are available are enhanced. Spearman correlations between spot entropies across coarseness levels (k) show a maximum of 0.44 ($k = 6$; p-value $< 10^{-5}$; Fig. 2D). Higher correlations are limited in part by the nature of both variables: cell-type entropy is continuous, while the cluster simplex is discrete, so the vector entropy is categorical (with even fewer possible values). This is especially notable with low-entropy spots containing a dominant cell type. Nevertheless, Square obtains significantly higher correlation values compared to STAGE, BayesSpace and CARD (Fig. 2E).

We further annotated the enhanced subspots (Fig. 2C, Fig. A11), constructing for each spot a cell-type frequency vector that is compared to the true one. The RMSE values for Square are concentrated around 0.2, with the compared methods exhibiting distributions at higher values (Fig. 2F).

These results endorse the idea that Square is capable of untangling the mixture of different cells in the initial spots, and generating more informative phenotypes with higher-resolution spatial labels. This validation serves as a guarantee that Square is not oversmoothing or slightly jittering the initial data, but generating feasible subspots that approximate cells or smaller groups of them.

5.4 Human Glioblastoma (GBM) dataset

This section builds upon the findings of Greenwald et al. (2024), who proposed a hypoxia-driven organizational model of GBM comprising five distinct layers. Their analysis demonstrated that the heightened cellular organization associated with hypoxia is not confined to hypoxic cell populations alone, but is also observed in neighboring non-hypoxic cell types located in close proximity to hypoxic cells (referred to as *MES-Hyp*). To quantify the degree of organization, the authors introduced a global metric termed spatial coherence for individual metaprogrammes (MPs) or states, defined on the basis of the number of immediate neighboring spots sharing the same MP across all MP-assigned locations. By constructing subsamples as sets of concentric rings centered on each spot, the "spatial coherence" was computed at the spot level, enabling the classification of spots as either *structured* or *organized*. Using the malignancy level (computed from the CNA signal), the spots were further subdivided defining four classes: *struct-malig*, *disorg-malig*, *struct-norm* and *disorg-norm*.

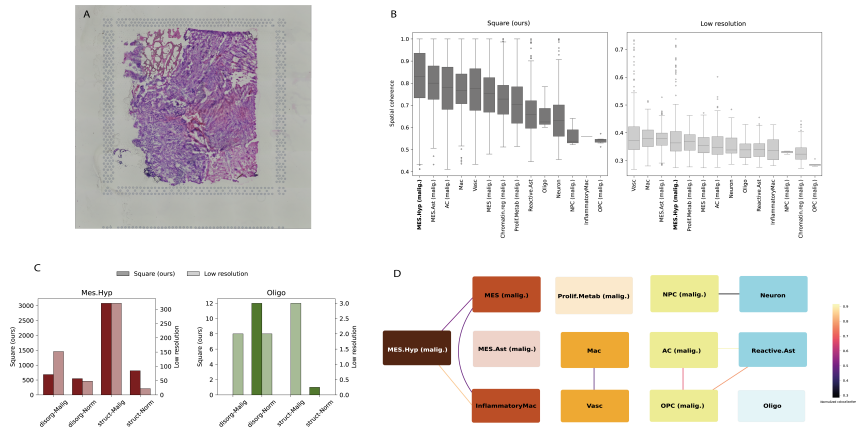

Figure 3: **A.** GBM H&E image. **B.** Mean spatial coherence for all metaprogrammes in descending order, for both the original and the Square-enhanced data. **C.** Proportions of organization + malignancy division for *MES-Hyp* and *Oligo*. **D.** Organizational model built from the consensus set of interactions inferred from the Square-generated data.

We ran Square on the GBM dataset (Fig. 3A) using only the 562 unique genes used to define the 14 proposed MPs, and their corresponding signatures were used to annotate the spots. We observed that the "spatial coherence" for the spots belonging to each of the states depicted bigger differences between them in terms of coherence (Fig. 3B). The latter aligns with the observed differences between the organizational behavior of different states. Besides, the increased organization associated with hypoxia was clearly visualized only after increasing the spatial resolution with Square.

Square-generated data also maintains the proportions consistently observed on other GBM samples (Fig. 3C). These proportions align with these previous findings: *disorg-norm* is composed of normal brain regions with a high degree of cancer infiltration and has the highest frequency of *Oligo*; and structured regions are found at the core of tumors (enriched with *MES-Hyp*). Note that the *Oligo* highest frequency was unobserved before running Square.

Furthermore, the authors quantified spatial relationships among cell states using three complementary coupling measures (see App. A.3 for details) ——regional composition, adjacency, and

colocalization—— and defined a consensus set of strong interactions supported by multiple metrics. This consensus revealed a layered organizational model, represented as a graph of *consistently strong* state–state interactions. Even though there are some interactions in the model that are not captured, the same layered structure is preserved in the Square-generated data (Fig. 3D), with no consensus interactions observed between non-consecutive layers, supporting the proposed model.

6 Conclusion

We proposed Square, which generates high-resolution spatial data using only the spots' transcriptional profiles. Our findings demonstrate that combining graph-based models with attention mechanisms effectively disentangles mixed spatial signals and reallocates them with high confidence. Enhancing resolution not only sharpens gene expression patterns but also reveals biological insights that remain obscured under coarse spatial measurements, ultimately enabling a more precise characterization of tissue architecture and cellular diversity.

References

10x Genomics. Human glioblastoma multiforme visium spatial gene expression dataset. Spatial Gene Expression dataset generated using Visium technology, October 2020.

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631, 2019.

Alma Andersson, Joseph Bergenstråhle, Michaela Asp, Ludvig Bergenstråhle, Aleksandra Jurek, José Fernández Navarro, and Joakim Lundeberg. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Communications biology*, 3(1):565, 2020.

Rohit Arora, Christian Cao, Mehul Kumar, Sarthak Sinha, Ayan Chanda, Reid McNeil, Divya Samuel, Rahul K Arora, T Wayne Matthews, Shamir Chandarana, et al. Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. *Nature communications*, 14(1):5029, 2023.

Rocky M Barilla, Brian Diskin, Raul Caso Caso, Ki Buom Lee, Navyatha Mohan, Chandan Buttar, Salma Adam, Zennur Sekendiz, Junjie Wang, Ruben D Salas, et al. Specialized dendritic cells induce tumor-promoting il-10+ il-17+ foxp3 neg regulatory cd4+ t cells in pancreatic carcinoma. *Nature communications*, 10(1):1424, 2019.

Ludvig Bergenstråhle, Bryan He, Joseph Bergenstråhle, Xesús Abalo, Reza Mirzazadeh, Kim Thrane, Andrew L Ji, Alma Andersson, Ludvig Larsson, Nathalie Stakenborg, et al. Super-resolved spatial transcriptomics by deep data fusion. *Nature biotechnology*, 40(4):476–479, 2022.

Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*, 2021.

Nicholas H Chakiryan, Gregory J Kimmel, Youngchul Kim, Joseph O Johnson, Noel Clark, Ali Hajiran, Andrew Chang, Ahmet M Aydin, Logan Zemp, Esther Katende, et al. Geospatial cellular distribution of cancer-associated fibroblasts significantly impacts clinical outcomes in metastatic clear cell renal cell carcinoma. *Cancers*, 13(15):3743, 2021.

Jesus de la Fuente, Naroa Legarra, Guillermo Serrano, Irene Marin-Goni, Aintzane Diaz-Mazkiaran, Markel Benito Sendin, Ana Garcia Osta, Krishna R Kalari, Carlos Fernandez-Granda, Idoia Ochoa, et al. Sweetwater: An interpretable and adaptive autoencoder for efficient tissue deconvolution. *arXiv preprint arXiv:2311.11991*, 2023.

Alissa C Greenwald, Noam Galili Darnell, Rouven Hoefflin, Dor Simkin, Christopher W Mount, L Nicolas Gonzalez Castro, Yotam Harnik, Sydney Dumont, Dana Hirsch, Masashi Nomura, et al. Integrative spatial analysis reveals a multi-layered organization of glioblastoma. *Cell*, 187 (10):2485–2501, 2024.

Rushikesh S Joshi, Samanvi S Kanugula, Sweta Sudhir, Matheus P Pereira, Saket Jain, and Manish K Aghi. The role of cancer-associated fibroblasts in tumor progression. *Cancers*, 13(6):1399, 2021.

Mena Kamel, Yiwen Song, Ana Solbas, Sergio Villordo, Amrut Sarangi, Pavel Senin, Mathew Sunaal, Luis Cano Ayestas, Clement Levin, Seqian Wang, et al. Enact: End-to-end analysis of visium high definition (hd) data. *Bioinformatics*, 41(3):btaf094, 2025.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. URL `https://arxiv.org/abs/1412.6980`.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Shang Li, Kuo Gai, Kangning Dong, Yiyang Zhang, and Shihua Zhang. High-density generation of spatial transcriptomics with stage. *Nucleic Acids Research*, 52(9):4843–4856, 2024.

Tianyi Liu, Linli Zhou, Danni Li, Thomas Andl, and Yuhang Zhang. Cancer-associated fibroblasts build and secure the tumor microenvironment. *Frontiers in cell and developmental biology*, 7:60, 2019.

Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12):1053–1058, 2018.

Romain Lopez, Baoguo Li, Hadas Keren-Shaul, Pierre Boyeau, Merav Kedmi, David Pilzer, Adam Jelinski, Ido Yofe, Eyal David, Allon Wagner, et al. Destvi identifies continuums of cell types in spatial transcriptomics data. *Nature biotechnology*, 40(9):1360–1369, 2022.

Ying Ma and Xiang Zhou. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nature biotechnology*, 40(9):1349–1359, 2022.

Reuben Moncada, Dalia Barkley, Florian Wagner, Marta Chiodin, Joseph C Devlin, Maayan Baron, Cristina H Hajdu, Diane M Simeone, and Itai Yanai. Integrating microarray-based spatial transcriptomics and single-cell rna-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nature biotechnology*, 38(3):333–342, 2020.

Lambda Moses and Lior Pachter. Museum of spatial transcriptomics. *Nature methods*, 19(5):534–546, 2022.

Michelli Faria de Oliveira, Juan Pablo Romero, Meii Chung, Stephen R Williams, Andrew D Gottscho, Anushka Gupta, Susan E Pilipauskas, Seayar Mohabbat, Nandhini Raman, David J Sukovich, et al. High-definition spatial transcriptomic profiling of immune cell populations in colorectal cancer. *Nature Genetics*, pp. 1–12, 2025.

Trang Pham, Truyen Tran, Hoa Dam, and Svetha Venkatesh. Graph classification via deep learning with virtual nodes. *arXiv preprint arXiv:1708.04357*, 2017.

Monika Piwecka, Nikolaus Rajewsky, and Agnieszka Rybak-Wolf. Single-cell and spatial transcriptomics: deciphering brain complexity in health and disease. *Nature Reviews Neurology*, 19(6): 346–362, 2023.

Chendi Qian, Andrei Manolache, Christopher Morris, and Mathias Niepert. Probabilistic graph rewiring via virtual nodes. *Advances in Neural Information Processing Systems*, 37:28359–28392, 2024.

Anjali Rao, Dalia Barkley, Gustavo S França, and Itai Yanai. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871):211–220, 2021.

Daniel E Russ, Ryan B Patterson Cross, Li Li, Stephanie C Koch, Kaya JE Matson, Archana Yadav, Mor R Alkaslasi, Dylan I Lee, Claire E Le Pichon, Vilas Menon, et al. A harmonized atlas of mouse spinal cord cell types and their spatial organization. *Nature communications*, 12(1):5722, 2021.

Sheel Shah, Eric Lubeck, Wen Zhou, and Long Cai. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron*, 92(2):342–357, 2016.

Kelly D Smith, David K Prince, James W MacDonald, Theo K Bammler, and Shreeram Akilesh. Challenges and opportunities for the clinical translation of spatial transcriptomics technologies. *Glomerular Diseases*, 4(1):49–63, 2024.

Dongyuan Song, Qingyang Wang, Guanao Yan, Tianyang Liu, Tianyi Sun, and Jingyi Jessica Li. scdesign3 generates realistic in silico data for multimodal single-cell and spatial omics. *Nature Biotechnology*, 42(2):247–252, 2024.

Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353 (6294):78–82, 2016.

Caleb R Stoltzfus, Jakub Filipek, Benjamin H Gern, Brandy E Olin, Joseph M Leal, Yajun Wu, Miranda R Lyons-Cohen, Jessica Y Huang, Clarissa L Paz-Stoltzfus, Courtney R Plumlee, et al. Cytomap: a spatial analysis toolbox reveals features of myeloid cell organization in lymphoid tissues. *Cell reports*, 31(3), 2020.

Leilei Tao, Guichun Huang, Haizhu Song, Yitian Chen, and Longbang Chen. Cancer associated fibroblasts: An essential role in the tumor microenvironment. *Oncology letters*, 14(3):2611–2620, 2017.

Burak Tepe, Matthew C Hill, Brandon T Pekarek, Patrick J Hunt, Thomas J Martin, James F Martin, and Benjamin R Arenkiel. Single-cell rna-seq of mouse olfactory bulb reveals cellular heterogeneity and activity-dependent molecular census of adult-born neurons. *Cell reports*, 25(10): 2689–2703, 2018.

Sara A Väyrynen, Jinming Zhang, Chen Yuan, Juha P Väyrynen, Andressa Dias Costa, Hannah Williams, Vicente Morales-Oyarvide, Mai Chan Lau, Douglas A Rubinson, Richard F Dunne, et al. Composition, spatial characteristics, and prognostic significance of myeloid cell infiltration in pancreatic cancer. *Clinical Cancer Research*, 27(4):1069–1081, 2021.

Maciej Wiewiora, Janusz Jopek, Elżbieta Świetochowska, Gregorczyn Sławomir, Jerzy Piecuch Jr, Mateusz Gaska, and Jerzy Piecuch. Blood-based protein biomarkers and red blood cell aggregation in pancreatic cancer. *Clinical Hemorheology and Microcirculation*, (Preprint):1–13, 2023.

Lulu Yan and Xiaoqiang Sun. Benchmarking and integration of methods for deconvoluting spatial transcriptomic data. *Bioinformatics*, 39(1):btac805, 2023.

Edward Zhao, Matthew R Stone, Xing Ren, Jamie Guenthoer, Kimberly S Smythe, Thomas Pulliam, Stephen R Williams, Cedric R Uytingco, Sarah EB Taylor, Paul Nghiem, et al. Spatial transcriptomics at subspot resolution with bayesspace. *Nature biotechnology*, 39(11):1375–1384, 2021.

Zixiang Zhou, Yunshan Zhong, Zemin Zhang, and Xianwen Ren. Spatial transcriptomics deconvolution at single-cell resolution using redeconve. *Nature Communications*, 14(1):7930, 2023.

A Appendix

A.1 Square's additional details

A.1.1 Training and inference graphs

Fig. 4 depicts the way in which superspots are generated from the original data, as well as the considered graphs at training and inference time.

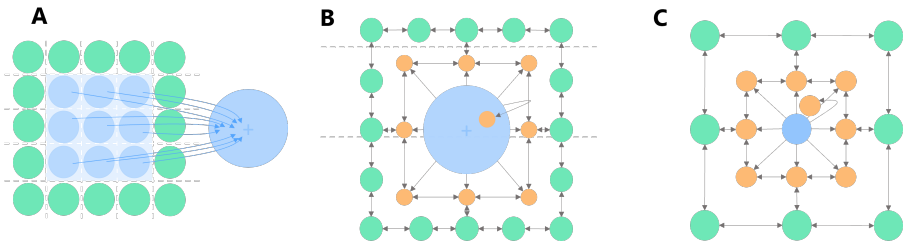

Figure 4: **A.** Training data: superspots are generated by shifting a pooling mask across the grid with overlap. **B.** Training mode: graph with a superspot (blue), its neighbors (green), and the virtual intermediate nodes (yellow). **C.** Inference mode: graph with a central spot (blue), its neighbors (green) and the virtual nodes (yellow). The virtual nodes eventually represent the enhanced subspots.

A.1.2 SQUARE'S PSEUDO-CODE

Algorithm 1 summarizes the steps followed by the proposed method Square.

Algorithm 1 Training and Inference Procedure for Square

Input: Spatial transcriptomics counts data $\mathbf{X}$, spots' coordinates array $\mathbf{Z}$, resolution factor $k = 9$, training neighbors $m_{\text{train}} = 16$, inference neighbors $m_{\text{inf}} = 8$
Output: Predicted high-resolution spot profiles ($\mathbf{X}_{hr}$ and $\mathbf{Z}_{hr}$)
// Data preprocessing
Normalize and logarithmize data
Select most highly variable genes
// Training Data Generation
for each non-edge region in X **do**
    Group k initial spots into a super-spot by summing their transcriptional signals
    Keep the k initial spots to serve as ground truth
    Identify m_{train} neighboring spots for each super-spot
    Generate training instance with paired matrices containing the super-spot and neighbors, and the ground-truth initial spots
end for
// Graph Construction for Each Training Instance
for each training instance **do**
    Initialize graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with central super-spot and neighbors
    Add k virtual nodes representing enhanced subspots
    Connect virtual nodes to central node, neighbors, and among themselves
    Set edge weights:
        $w = d(s_i, s_j)^{-1}$ for original neighbors or virtual neighbors
        $w = t$ for edges connecting central node to virtual nodes
        $w = 1$ for edges connecting original nodes to virtual nodes
    Initialize virtual nodes with the average of their closest neighbors
end for
// Model Training
for each training batch **do**
    Encode (with parameters ϕ) the super-spot, virtual nodes' and neighbors' profiles ($\mathbf{x}$) as Gaussian distributions ($\mathbf{z}$)
    $\mathcal{L}_{\phi}^{\text{KL}} = D_{KL}(q_{\phi}(\mathbf{z}||\mathbf{x}))$
    Reparametrization trick: $\mathbf{z} = \mu + \sigma \bigodot \epsilon$, where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$
    Decode (with parameters θ) latent representations back to gene space
    $\mathcal{L}_{\phi,\theta}^{\text{recons}} = \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}[\log p_{\theta}(\mathbf{x}|\mathbf{z})]$
    Populate the training graph with latent representations
    Encode graph embeddings using a Graph Attention Network (GAT)
    Decode (with parameters δ) GAT output back to gene space
    $\mathcal{L}_{\phi,\delta}^{\text{deconv}} = \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}[\log p_{\delta}(\mathbf{x}|\mathbf{z}, \mathbf{A})]$
    Compute total loss: $\mathcal{L}_{\phi,\theta,\delta} = \mathcal{L}_{\phi,\theta}^{\text{recons}} + \mathcal{L}_{\phi,\delta}^{\text{deconv}} - \mathcal{L}_{\phi}^{\text{KL}}$
    Update parameters $\{\phi, \theta, \delta\}$ using stochastic gradient descent:
        $\{\phi, \theta, \delta\} \leftarrow \{\phi, \theta, \delta\} - \eta \nabla_{\{\phi,\theta,\delta\}} \mathcal{L}_{\phi,\theta,\delta}$
end for
// Inference
for each spot in X **do**
    Identify m_{inf} neighboring spots
    Construct graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with central spot, neighbors, and k virtual nodes as in training
    Encode observed profiles into latent space using trained encoder (ϕ)
    Populate graph with latent embeddings
    Perform message passing using the trained GAT
    Decode (with parameters δ) the embeddings of virtual nodes into gene expression space ($\mathbf{X}_{hr}$)
    Assign each predicted virtual node its corresponding coordinate within the central spot to form the enhanced high-resolution map ($\mathbf{Z}_{hr}$)
end for

A.1.3 HYPERPARAMETER TUNING

We performed an optimization of model architecture and hyperparameters using Optuna[1], a framework for efficient hyperparameter search. We used its Tree-structured Parzen Estimator (TPE) sampler to suggest new hyperparameter configurations, and conducted a total of 500 trials. The list of considered hyperparameters and values is listed in Table 3.

Model optimization was conducted using the synthetic pseudo-phenotype dataset, for which the ground truth exists. Tables 4-7 show the top 10 trials based on MSE, cosine distance, correlation distance, and ARI, respectively. A visual summary of these results is also provided in Fig. 5 and 6. Based on the obtained results, we set the hyperparameters of the final Square model (see Table 8).

Hyperparameter	**Tried values**
Adjacency matrix weight between central node and virtual nodes (t)	$\{1.0, 1.2, \ldots, 3.0\}$
Embedded space dimension / Gene space dimension	$\{0.3, 0.4, \ldots, 0.8\}$
GAT layers	1, 2, 3
GAT dropout	$\{0.0, 0.1, \ldots, 0.5\}$
GAT aggregation function	*sum*, *mean*, *mul*, *min*, *max*
GAT learnable skip-connection	*True*, *False*
GAT LeakyReLU angle of the negative slope	log-uniform(0.1, 1.0)
$p_\delta(\mathbf{x}\|\mathbf{z}, \mathbf{A})$ decoder layers	1, 2, 3

Table 3: Hyperparameters tested in Square, together with the considered values. Hyperparameter tuning was conducted with Optuna.

Table 4: Top 10 trials optimized for MSE.

Trial	Value ↓	t	Dim. ratio	GAT layers	GAT dropout	Aggr. function	Skip-connections	LReLU neg. slope	$p_\delta(\mathbf{x}\|\mathbf{z}, \mathbf{A})$ layers
113	0.208	2.4	0.55	1	0.5	*sum*	*True*	0.118	2
407	0.209	2.6	0.55	1	0.5	*sum*	*False*	0.109	2
215	0.209	2.0	0.55	1	0.5	*sum*	*True*	0.116	2
65	0.209	1.8	0.55	1	0.5	*sum*	*False*	0.813	2
200	0.21	2.0	0.55	1	0.5	*sum*	*True*	0.125	2
336	0.21	2.2	0.55	1	0.5	*sum*	*False*	0.12	2
370	0.21	2.0	0.55	1	0.5	*sum*	*True*	0.124	1
483	0.21	2.4	0.55	1	0.5	*sum*	*False*	0.116	2
401	0.21	2.6	0.55	1	0.5	*sum*	*False*	0.101	2
204	0.21	2.0	0.55	1	0.5	*sum*	*True*	0.114	2

[1]T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework", Proceedings of the 25th ACM SIGKDD, 2019.

Table 5: Top 10 trials optimized for Cosine distance.

Trial	Value ↓	t	Dim. ratio	GAT layers	GAT dropout	Aggr. function	Skip-connections	LReLU neg. slope	$p_\delta(\mathbf{x}\|\mathbf{z},\mathbf{A})$ layers
65	0.005	1.4	0.55	1	0.5	*sum*	*False*	0.813	2
249	0.005	2.8	0.55	1	0.4	*sum*	*True*	0.222	2
485	0.005	2.8	0.55	1	0.4	*sum*	*False*	0.78	2
458	0.005	2.6	0.55	1	0.5	*sum*	*False*	0.633	2
152	0.005	1.2	0.55	1	0.4	*sum*	*True*	0.748	2
247	0.005	2.8	0.55	1	0.4	*sum*	*True*	0.22	2
187	0.005	2.6	0.55	1	0.4	*sum*	*True*	0.654	2
206	0.005	2.6	0.55	1	0.4	*sum*	*True*	0.671	2
372	0.005	2.4	0.55	1	0.4	*sum*	*True*	0.353	2
203	0.005	2.4	0.55	1	0.4	*sum*	*True*	0.66	2

Table 6: Top 10 trials optimized for Correlation distance.

Trial	Value ↓	t	Dim. ratio	GAT layers	GAT dropout	Aggr. function	Skip-connections	LReLU neg. slope	$p_\delta(\mathbf{x}\|\mathbf{z},\mathbf{A})$ layers
468	0.169	2.4	0.5	3	0.2	*mean*	*True*	0.648	2
232	0.17	1.0	0.5	3	0.2	*mean*	*True*	0.49	1
401	0.17	1.0	0.5	3	0.2	*mean*	*True*	0.534	1
399	0.173	2.0	0.5	3	0.2	*mean*	*True*	0.234	1
474	0.174	2.6	0.5	3	0.2	*mean*	*True*	0.62	1
229	0.174	1.0	0.5	3	0.2	*mean*	*True*	0.48	1
170	0.175	2.0	0.5	3	0.2	*mean*	*True*	0.46	1
451	0.175	2.0	0.5	3	0.2	*mean*	*True*	0.558	1
21	0.176	2.6	0.75	2	0.1	*min*	*False*	0.726	2
315	0.176	2.2	0.5	3	0.2	*min*	*True*	0.494	1

Table 7: Top 10 trials optimized for ARI.

Trial	Value ↑	t	Dim. ratio	GAT layers	GAT dropout	Aggr. function	Skip-connections	LReLU neg. slope	$p_\delta(\mathbf{x}\|\mathbf{z},\mathbf{A})$ layers
266	0.733	2.4	0.6	1	0.1	*sum*	*False*	0.239	3
420	0.733	1.8	0.6	1	0.1	*sum*	*False*	0.156	3
119	0.733	1.8	0.6	1	0.1	*sum*	*False*	0.197	3
280	0.732	2.2	0.6	1	0.1	*sum*	*False*	0.24	2
190	0.732	2.2	0.6	1	0.1	*sum*	*False*	0.2	3
398	0.732	1.8	0.6	1	0.1	*sum*	*False*	0.176	3
294	0.731	1.6	0.6	1	0.1	*sum*	*False*	0.261	3
218	0.731	2.0	0.6	1	0.1	*sum*	*False*	0.172	3
174	0.731	2.0	0.6	1	0.1	*sum*	*False*	0.179	3
390	0.731	2.0	0.6	1	0.1	*sum*	*False*	0.198	3

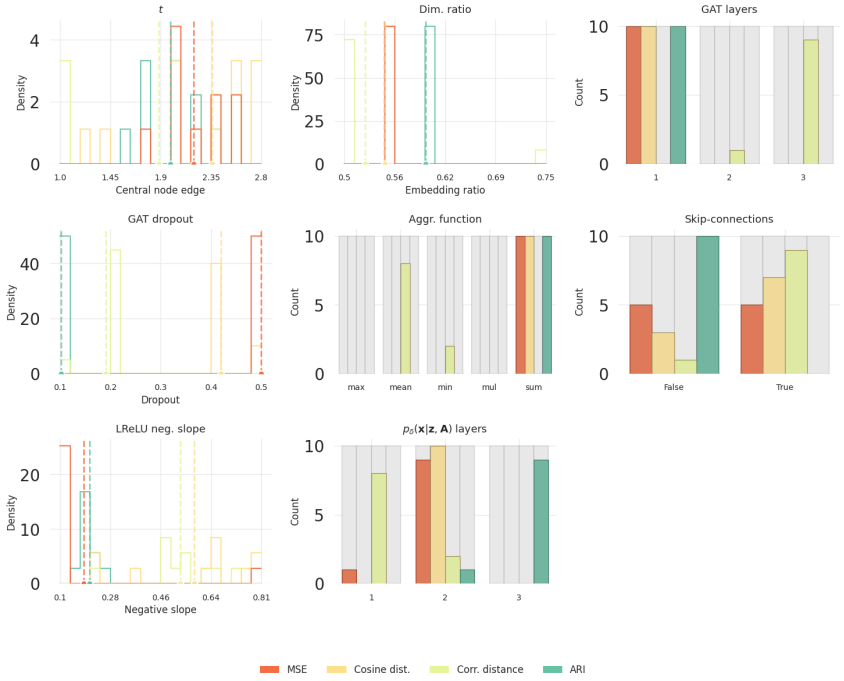

Figure 5: Results of the optimization process: density of the hyperparameters *central node edge, embedding ratio, dropout rate* and *negative slope* for the top 10 trials (dashed lines represent the mean value); and histograms of *GAT layers, Aggregation function, skip-connections* and $p_\delta(\mathbf{x}|\mathbf{z}, \mathbf{A})$ *decoder layers* also for the top 10 trials. Results are shown for MSE, cosine distance, correlation distance and ARI metrics.

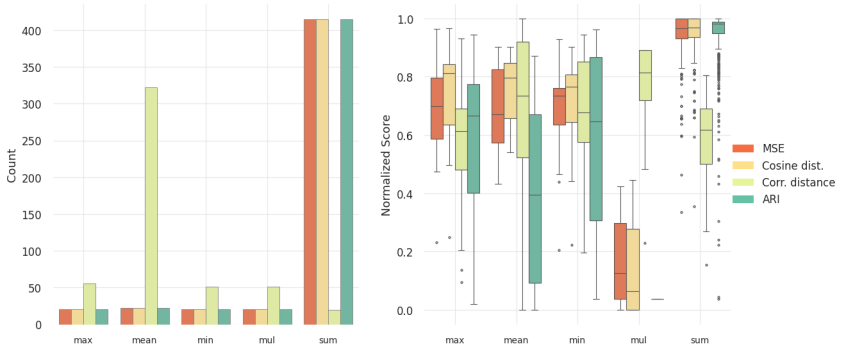

Figure 6: Results of the optimization process for hyperparameter *Aggregation function* for the 500 trials. Results are shown for MSE, cosine distance, correlation distance and ARI metrics. For the normalized score, the MSE is shown as $1 - MSE$ so that higher values are better.

Hyperparameter	**Optimal value**
Adjacency matrix weight between central node and virtual nodes (t)	2.0
Embedded space dimension / Gene space dimension	0.55
GAT layers	1
GAT dropout	0.2
GAT aggregation function	*sum*
GAT learnable skip-connection	*True*
GAT LeakyReLU angle of the negative slope	0.2
$p_\delta(\mathbf{x}\|\mathbf{z}, \mathbf{A})$ decoder layers	2

Table 8: Final hyperparameters' values for Square after Optuna optimization.

A.2 DATASETS: ADDITIONAL DETAILS

A.2.1 MOUSE OLFACTORY BULB (MOB)

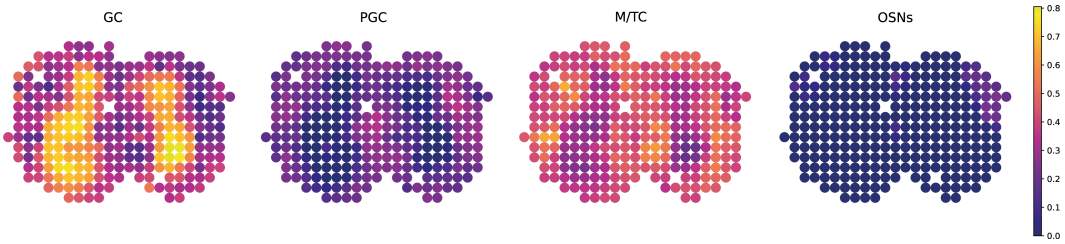

Figure 7: Mouse olfactory bulb dataset ground truth proportions.

A.3 EVALUATION METRICS

We devise several quantitative metrics that help in evaluating the effectiveness of Square in different scenarios. The metrics are reported across 5 runs to account for model's robustness.

Synthetic pseudo-phenotype dataset. We compute spot-wise metrics that directly compare the transcriptional profile of the generated subspots with that of the original ones: the Mean Squared Error (MSE) and the rank-based Spearman correlation distance (SCD), the latter to measure the preservation of the relative ranking of highly expressed genes. To further evaluate whether the most highly expressed genes are retained after resolution enhancement, we compute two metrics based on the Jaccard distance (JD) between the sets of top-expressed genes in the ground truth and the reconstruction: for the genes above the mean (JD_μ) and for the top 30 most expressed genes (JD_{30}).

Additionally, we consider a metric to capture whether the generated spots maintain the cell-type distribution of the original spots. For each cell-type, we generate a representative by computing the mean of the corresponding single-cell transcriptional profiles. Then, the enhanced subspots are assigned the cell-type whose representative is closer in Euclidean distance to their inferred transcriptional profile. Using these predicted and true cell-type assignments, we compute the Adjusted Rand Index (ARI) to quantify the agreement between them. This metric evaluates how well the enhanced data preserves both the spatial distribution and cell-type identity compared to the ground truth.

MOB dataset. Since only the cell-type proportions of the original spots are available, we consider a clustering-based assessment to determine whether the enhanced data accurately reflects the underlying cell-type distribution. We first apply k-means to the enhanced subspots with varying values of k, allowing for less and more coarse clustering. For a given k, we generate a *cluster-frequency* vector for every original spot, with the ith entry reflecting the proportion of subspots assigned to cluster i. Since cluster assignments may not reflect cell-types, we cannot directly compare this vector with the cell-type proportion vector. Instead, we compute the entropy of each of them, and then evaluate the Spearman correlation between them across spots. We expect the entropies between the vectors to be correlated, as the original cell-type distribution should be reflected in the clustering assignment.

To further evaluate the enhanced subspots, we annotate each of them by first clustering the enhanced spots, and then assigning to each cluster the cell-type with more overlap between its marker genes and the cluster significative genes. A cell-type frequency vector is then constructed for each spot and compared to the ground truth cell-type proportion vector via the RMSE.

Real datasets. For the GBM data, we recreate the main findings by Greenwald et al. (2024), who analyzed 13 GBM spatial samples (see Results). In order to quantify the spatial relationships between cell states, they used three complementary measures of cell type coupling at distinct spatial resolution: *i)* regional composition, defined within a neighborhood of spots determined by the number of concentric rings around a given spot r; *ii)* adjacency, assessed using a single spot and its immediate neighbors; and *iii)* colocalization, inferred from spot-level deconvolution using MP scores. A consensus set of state–state interactions can be defined by selecting strong interactions supported by multiple coupling measures.

On the PDAC data, we rely on deconvolution method CARD (Ma & Zhou, 2022) to estimate cell-type proportions for every original and enhanced spot. We expect the enhanced subspots to be more "pure" (fewer cell-types) than the original spot, as the original spot's cells are expected to be spatially distributed across the inferred subspots. The entropy of the cell-type proportion vector of the enhanced subspots is therefore expected to be smaller than that of the original spots. To evaluate this, we compute the entropies and apply a paired t-test.

A.4 ADDITIONAL RESULTS

A.4.1 ABLATION STUDY: VIRTUAL NODES' INITIALIZATION

We evaluated different approaches to initialize the virtual nodes' transcriptional profiles. Specifically, we tested: i) averaging the transcriptional profile a virtual node's neighbouring spots; ii) imputation with a GNN (see below); iii) an all zero-vector; and iv) a random initialization.

The GNN is designed to predict dataset-specific spots in between others. A prior training is performed by generating 100 different graphs with the whole dataset grid, randomly removing 20% of the spots, which are used to train the network with MSE as loss. The GNN contains 2 layers and the number of neurons at the final layer equals the number of genes. Training is conducted using SCD and a learning rate scheduler with a patience parameter of 5 epochs. This network is then used to populate the transcriptional profile of virtual nodes, which are also embedded into the latent space using the encoder $q_\phi(\mathbf{z}|\mathbf{x})$.

The ablation study is conducted with the synthetic pseudo-phenotype dataset, to have access to ground truth. The results are provided in Table 9. Overall, we observed the *mean* obtains the best results, and hence it is selected as the final approach to initialize the virtual nodes in Square.

Table 9: **Virtual nodes' initialization:** Performance comparison across different approaches.

Approach	MSE ↓	SCD ↓	$JD_{>\mu}$ ↓	JD_{30} ↓	ARI ↑
mean	1.72 ± 0.03	$\mathbf{0.438 \pm 0.030}$	$\mathbf{0.469 \pm 0.023}$	$\mathbf{0.8135 \pm 0.0012}$	$\mathbf{0.71 \pm 0.11}$
GNN	$\mathbf{1.01 \pm 0.19}$	1.049 ± 0.093	0.658 ± 0.027	0.8217 ± 0.0020	0.00 ± 0.00
zero	1.73 ± 0.04	0.468 ± 0.023	0.477 ± 0.026	0.8149 ± 0.0013	0.57 ± 0.12
random	1.49 ± 0.07	0.668 ± 0.087	0.541 ± 0.029	0.8172 ± 0.0014	0.23 ± 0.16

A.4.2 SYNTHETIC PSEUDO-PHENOTYPE DATASET

In this synthetic dataset, the pseudo-phenotypes are generated from a zero-inflated negative binomial (ZINB) distribution with different parameters, i.e., the expected number of events k occurring within the observed interval λ and the zero-inflated probability π which is fixed as the desired level of sparsity.

Fig. 8A depicts the marker genes across the five generated pseudo types, and Fig. 8B displays the spatial structure of the generated spatial dataset.

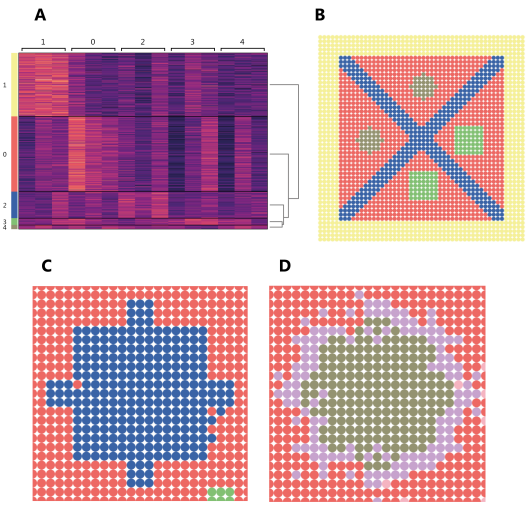

Figure 8: Synthetic pseudo-phenotype dataset. **A.** Heatmap of marker genes across the five generated pseudo-phenotypes. **B.** Spatial structure of pseudo-phenotypes displaying different geometries and patterns. **C.** Left circle enhancement by ablated model without attention. The colors represent cluster assignments of the enhanced subspots with k-means. **D.** Left circle enhancement by Square. The colors represent cluster assignments of the enhanced subspots with k-means.

A.4.3 EMBEDDED SPACE ANALYSIS

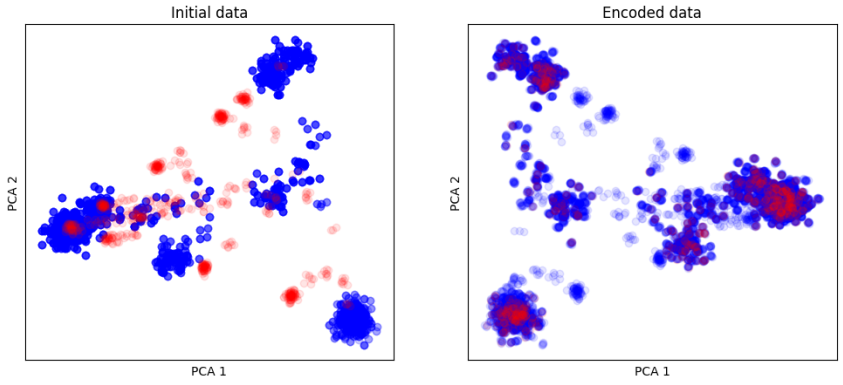

Figure 9: Projection on the two first principal components of PCA of the original transcriptional profile (left panel) and the generating embeddings after passing them through the encoder of the VAE (right panel) of spots (blue) and superspots (red) from the synthetic pseudo-phenotype dataset. As can be observed, there is a clear distribution shift on the original space, which is not visible on the embedding space.

A.4.4 SYNTHETIC MOUSE OLFACTORY BULB (MOB) DATA

In order to annotate the generated spots after resolution enhancement, the 50 most significative marker genes of each one of the cell types are leveraged. Figure 10 shows the overlap between them, which is relatively large. Exclusive marker genes for each cell type are: *CNTNAP4* for granule cells; *CPE* for ,itral and tufted cells; and *OMP*, *SPON1* and *AI593442* for olfactory sensory neurons.

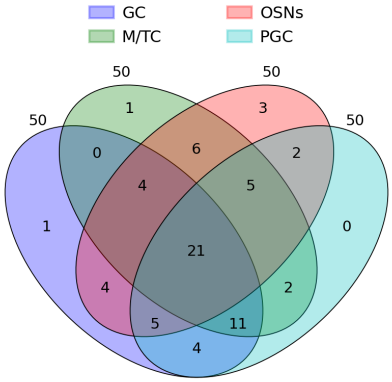

Figure 10: Venn diagram of the marker genes of MOB dataset cell types.

Figure 11 shows the annotation of the datasets generated by BayesSpace, CARD and STAGE. All three methods impose an excessively smooth, continuous spatial distribution of cell types across the tissue. Note that they overlook the presence of periglomerular cells, with a discrete appearance over the tissue, fact which makes its detection clearly challenging. In contrast, Square is able to recover the presence of these cells (Figure 2D), even when they are blurred by the rest.

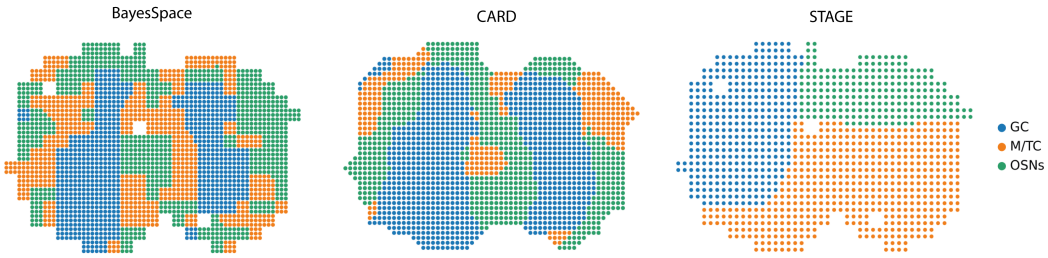

Figure 11: Annotation of datasets generated by BayesSpace, CARD and STAGE.

A.4.5 HUMAN PANCREATIC DUCTAL ADENOCARCINOMA (PDAC) DATASET

Given the lack of ground truth for this dataset, we first applied the deconvolution method CARD to the original and enhanced spots. An analysis of the entropies of the resulting cell-type proportion vectors revealed a significant decrease in the enhanced spots (Fig. 12A; p-value 0.004). The reduction in entropy suggests that Square successfully reduces the ambiguity in cell-type assignments, leading to a more accurate representation of cellular composition within the tissue.

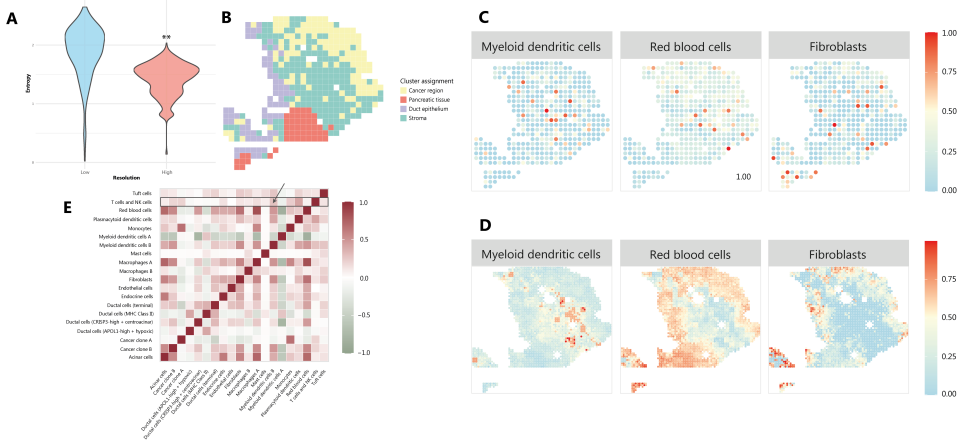

Figure 12: **A.** Entropy of the cell-type proportion vectors (computed with CARD) for the original and enhanced spots (low and high resolution, respectively). **B.** Four main regions of the spatial dataset. **C-D.** Proportion of Myeloid Dentric cells (mDCs), Red Blood cells (RBC), and Fibroblasts across the original and enhanced spots, respectively. Cell-type values do not represent absolute proportions, but normalized ones across all spots. **E.** Cell type proportion correlation computed with CARD.

Indeed, we further analyzed the spatial distribution of several cell-types and concluded that Square generates enhanced subspots that capture the cell-type distributions more precisely. For example, the presence of myeloid dendritic cells (mDCs) is mainly concentrated in the stroma region in both the high and low-resolution data (Fig. 12B-D). However, the enhanced spots exhibit a smoother distribution and the spatial clustering of cells with identical functions. Myeloid cells are known to be enriched within the stromal regions of PDAC, and to be highly heterogeneous across tumor regions, as observed in multiplex immunofluorescence assays (Väyrynen et al., 2021). This aligns perfectly with the observations of mDCs across the sequenced space (Fig. 12D). mDCs play a complex role in the tumor microenvironment of a PDAC. In healthy conditions, they serve as an immune modulator by their antigen-presentation task; otherwise, they promote tumor growth by inducing a unique regulatory T cell program that is associated with immune tolerance and reduced survival (Barilla et al., 2019). In fact, in the enhanced data, the presence of T cells is spatially positively correlated with the presence of one of the clusters of mDCs (Fig. 12E), being the second most correlated.

Furthermore, a recent study highlights the association of PDAC with disturbances in red blood cell (RBC) aggregation (Wiewiora et al., 2023). This study confirms that PDAC is linked to excessive aggregation of RBCs, including at the pancreas cancerous regions, but not in stroma regions. We note that the distribution of RBCs after Square enhancement follows a new spatial pattern, characterized broadly by the presence of RBCs at all regions except for the stroma (Fig. 12D). Finally, a significant enrichment of fibroblasts in the ductal epithelium and the cancer region has been reported for this dataset (Moncada et al., 2020). When gazing at the CARD proportions, we see that Square is indeed deciphering this enrichment (Fig. 12B,D). The enhanced spatial distribution aligns also perfectly with the well-studied association of cancer-associated fibroblasts (CAFs) with tumor growth and immune evasion in the tumor microenvironment (Liu et al., 2019; Tao et al., 2017; Joshi et al., 2021). Furthermore, the spatial distribution of CAFs driven by this enhancement has been previously observed in different tissues (Chakiryan et al., 2021).

These analyses showcase how Square can generate biologically driven enhanced subspots using only the transcriptomic profile of the original spot and its neighbors. Applying CARD to the enhanced data yields smoother and better-localized spatial distributions. Overall, these results not only support the effectiveness of the proposed resolution enhancement technique, but also underscore its potential utility in providing finer-scale insights into the spatial organization of cell types in complex tissues, thereby enhancing our understanding of tissue architecture and function at a more detailed level.