scientific reports

Check for updates

OPEN

Attention-guided enhanced deconvolution enables reference-free cell type estimation in spatial transcriptomics

Xiao Yang, Yujiao Wang✉ & Xiaozhou Chen✉

Spatial transcriptomics technologies profile gene expression across tissue sections while retaining spatial information, yet most platforms capture signals from multiple cells per measurement location, requiring computational methods to determine the underlying cellular composition. Current deconvolution approaches either depend on single-cell reference atlases—which may be unavailable for rare tissues or specific disease contexts—or employ unsupervised methods that struggle with complex spatial patterns and typically need manual specification of how many cell types to identify. We developed Attention-Guided Enhanced Deconvolution (AGED), a two-stage framework combining probabilistic modeling with neural attention architectures for reference-free analysis. The first stage uses a Performer-based network with linear-complexity attention to systematically evaluate models with different numbers of cell types, automatically selecting the optimal configuration through composite scoring of reconstruction quality, cluster separation, and topic diversity. The second stage, Attention-Guide, stars with hierarchical probabilistic initialization. It progressively refines cell type features through multiple attention mechanisms: cross-attention based on statistical priors, spatial attention aggregating neighborhood information, and collaborative attention capturing theme-gene associations. A dynamic gating mechanism enables the model to balance global statistical patterns with locally data-driven features, while regularization promotes sparse solutions consistent with biological reality—each position containing only a few dominant cell types. Testing on Mouse Olfactory Bulb (MOB) tissue, AGED automatically identified four anatomical structures and achieved superior reconstruction performance (r=0.86) with characteristic sparsity patterns. When similarly applied to human pancreatic ductal adenocarcinoma (PDAC) and human thymus tissue, it revealed detailed relationships between dissected structures and cell types. The learned cell type distributions aligned well with established neuroanatomical boundaries and known molecular markers, indicating the attention-based refinement maintains biological interpretability throughout training. This framework offers a practical solution for spatial transcriptomics analysis across diverse experimental systems without requiring matched single-cell data.

Keywords Spatial transcriptomics, Reference-free deconvolution, Attention mechanisms, Topic modeling, Deep learning

Spatial transcriptomics technologies have revolutionized our ability to profile gene expression while preserving spatial context, enabling unprecedented insights into tissue architecture and cellular organization[1,2]. Platforms such as 10× Genomics Visium[3], Slide-seq[4], and MERFISH[5] now routinely capture transcriptomic information at resolutions approaching single-cell scale. However, most current technologies measure gene expression at spatial locations that encompass multiple cells, necessitating computational deconvolution to infer cell composition within heterogeneous tissue regions[6,7].

Existing deconvolution methods broadly fall into two categories. Reference-based approaches leverage single-cell RNA sequencing (scRNA-seq) atlases to define cell type signatures and decompose spatial measurements as linear combinations of these references[8–10]. Methods including Cell2location[11], RCTD[12], Stereoscope[13], SpatialDWLS[14], and AdRoit[15] have demonstrated success when comprehensive references are available. More recently, deep learning methods such as Tangram[16] and DestVI([17] have shown promise by learning complex

School of Mathematics and Computer Science, Yunnan Minzu University, Kunming 650500, Yunnan, China. ✉email: 877145043@qq.com; ch_xiaozhou@163.com

mappings between modalities. However, these approaches face fundamental limitations: they require high-quality matched scRNA-seq data, which may not exist for rare tissues or disease states[18,19]; batch effects between platforms can introduce systematic biases[10]; and reference atlases may fail to capture tissue-specific cell states[7].

Reference-free methods offer an alternative by directly learning cellular composition from spatial data without external references. STdeconvolve employs latent Dirichlet allocation (LDA) to model spots as mixtures of cell types[20], while matrix factorization approaches decompose expression matrices into interpretable basis vectors[21,22]. SpatialDE and SPARK identify spatially variable genes to guide decomposition[23,24], and BayesSpace performs Bayesian clustering with spatial priors[22]. However, these methods struggle to capture nonlinear gene interactions and complex spatial dependencies inherent in tissue architecture[1,25].

Deep learning has emerged as a powerful framework for spatial transcriptomics analysis, with transformer architectures proving particularly effective at modeling high-dimensional biological data[26,27]. Graph neural networks exploit spatial neighborhood structures[28,29], while vision transformers have been adapted for spatial omics[30]. However, purely data-driven models require large training datasets and can overfit on moderate-sized samples typical of spatial experiments[31]. Furthermore, black-box approaches often lack the interpretability crucial for biological discovery[32].

The integration of probabilistic priors with neural architectures represents a promising direction, combining interpretability of statistical models with representational power of deep learning[33,34]. Variational autoencoders have successfully incorporated priors for single-cell analysis[35,36], yet no existing approach systematically integrates the global structure captured by topic models with local pattern recognition of attention mechanisms for spatial deconvolution. Moreover, determining the optimal number of cell types (K) remains a critical challenge, with most methods requiring manual specification or applying single metrics that provide conflicting signals[37,38].

Here, we present a two-stage computational framework addressing these challenges. In the first stage, we employ a Performer-based architecture with FAVOR+ linear attention[39] to efficiently evaluate candidate models across K values, integrating three complementary metrics into a composite selection criterion using knee detection. In the second stage, Attention-Guided initializes with hierarchical LDA priors and refines them through a sophisticated attention architecture comprising LDA-guided cross-attention, spatial neighborhood attention, and topic-gene co-attention modules. This design bridges global statistical structure of LDA with local feature learning of transformers through learnable gating mechanisms that dynamically balance prior knowledge and data-driven insights. We demonstrate that our framework achieves superior reconstruction accuracy, improved spatial coherence, and greater interpretability compared to existing reference-free methods.

Results

Overview of the Reference-free deconvolution framework

We developed a computational framework that integrates attention-based neural networks with probabilistic topic modeling for reference-free spatial transcriptomics deconvolution. The workflow comprises two sequential stages: (1) automatic determination of optimal cell type numbers via Performer model selection (Fig.1 Stage 1); (2) spatially aware deconvolution utilizing attention mechanisms guided by latent Dirichlet allocation (LDA) prior (Fig.2 Stage 2).

AGED captures cell type proportions and expression profiles within ST data from simulated data

We employed a Poisson sampling strategy for reference-free deconvolution of spatial transcriptomics data from the MOB, using simulated cell type proportion matrices as ground truth. The MOB data comprise four major anatomical layers organized from inner to outer based on H&E staining annotation: the granular cell layer (GCL), mitral cell layer (MCL), glomerular layer (GL), and outer nuclear layer (ONL) (Fig. 2a). We successfully reconstructed all four anatomical layers of the MOB using AGED (Fig. 2b). Visual inspection and validation using marker genes confirmed the accurate reconstruction of the spatial landscape (Fig. 2c). Performance validation against other reference-based[40] deconvolution tools—Redeconve[41], SPOTlight [42] and the reference-free strategy STdeconvolve[20]—showed AGED accurately captures cell composition within spots (Fig. 2d). Figure 2e shows the evolution of loss values during training with AGED. It can be observed that the KL divergence loss and total loss exhibit consistent changes in the early stages of training, indicating that the model relies on the probabilistic prior to establish an initial representation, while other loss terms have not yet been fully optimized. Furthermore, to validate AGED performance, we performed correlation analysis between ground truth data and AGED deconvolution results (Fig. 2f). achieving a maximum correlation of 0.86. However, without reference expression profiles, performance was less than ideal for cell types with low abundance in spatial transcriptomics data. Comparing box plots of RMSE against other tools, AGED outperformed mainstream deconvolution methods (RMSE = 0.182). Additionally, we conducted a systematic analysis to assess whether the attention mechanisms in AGED capture biologically meaningful patterns. The results indicate that the learned attention weights indeed reflect genuine biological structures rather than arbitrary computational artifacts (Supplementary Figure 1).

AGED accurately localizes tissue regions in human pancreatic ductal adenocarcinoma

We applied the AGED framework (Fig. 1) to a human pancreatic ductal adenocarcinoma (PDAC) spatial transcriptomics dataset (n=428 spots, m=25,753 genes), contain four regions annotated by histologists based on H&E staining: the tumor region, pancreatic region, ductal region, and stroma region[43](Fig. 3a). Models were evaluated across K. A distinct inflection point emerged at K*=20 in the composite weighted scores, confirmed by maximum curvature analysis. This choice was validated by the observation that K=20 minimized the occurrence rate of rare themes (those with a global proportion <5%) while maintaining theme entropy values above 0.8×log(K). Visual inspection of spatial distributions confirmed that K=20 successfully captured

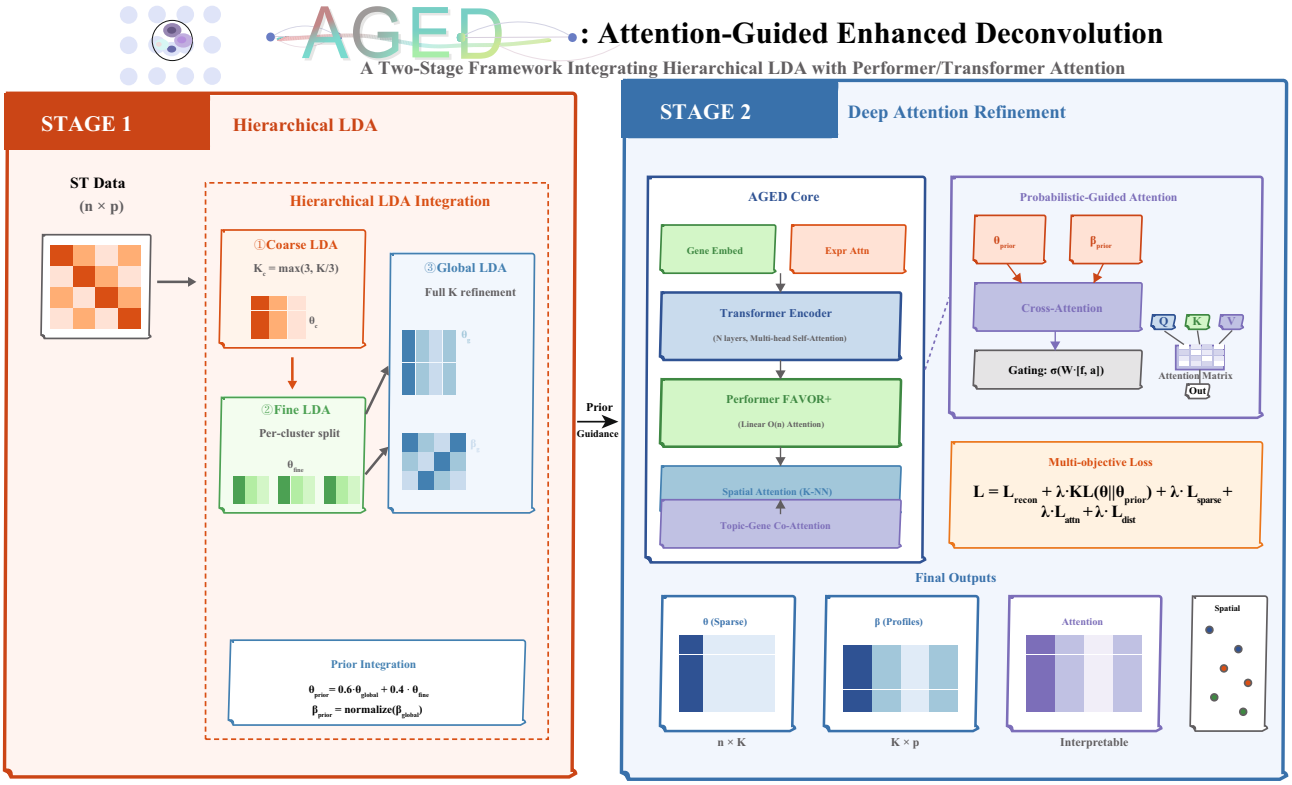

Fig. 1. Overview of AGED. AGED takes as input a spatial transcriptomics (ST) gene counts matrix of spot (rows) by N genes(columns). A matrix of spatial coordinates for each of the D pixels can also be used for visualization. Stage 1. Performer-based architecture is employed to determine the optimal number of cell types K*. Each spot is processed independently: (**a**) Hierarchical LDA probability initialization for coarse-grained decomposition; (**b**) Fine-grained decomposition for further subdivision of each cluster; (**c**) Global refinement integration to generate the thematic distribution θ and gene reconstruction results. Stage 2. (**a**) The core architecture of AGED: gene embedding → Transformer encoder → spatial attention → theme-gene collaborative attention; (**b**) incorporating prior information into the probabilistic attention guidance module via cross-attention and gating mechanisms, ultimately yielding sparse outputs θ (cell proportion), β (gene expression profile), and attention weights.

distinct anatomical regions without excessive fragmentation (Fig. 3b). Comparing the final spatial landscape with H&E sections and clusters derived from tissue segmentation reveals that captured anatomical regions align with histologist-annotated areas in H&E sections. Using AGED to project the three cell types in PDAC onto a spatial landscape aligns with the expression patterns of ground truth and corresponding marker genes (Fig. 3c). Correlation analysis between deconvolutions and regions indicates that most cell types can be assigned to the correct tissue regions (Fig. 3d). The specificity scores between cell types obtained via AGED reference-free deconvolution and four tissue regions (Fig. 3e). Finally, we found that when evaluating performance using the structural similarity index (SSIM) to measure the discrepancy between deconvolved data processed by other deconvolution tools and the clustering results, the AGED tool outperformed the others (SSIM = 0.564; Fig. 3f), while the highest SSIM value achieved by the other tools was only 0.502.

AGED can recognize the cellular boundaries between cTECs and mTECs in human thymus tissue

We applied AGED to thymus tissue, which includes: DN cells (CD4CD8), DP cells (CD4CD8), and SP cells (CD4CD8CD3 or CD8CD4CD3) are the predominant cell types involved in T cell differentiation. The DN cells included early DN cells (DN_early), DN blast cells (DN_blast), and DN thymocytes undergoing rearrangement (DN_re), and the DP cells consisted of the DP_blast and DP_re subsets. The subpopulations of SP cells include CD4 T cells, regulatory T cells (Treg), CD8 T cells, and CD8aa T cells[44]. H&E-stained sections clearly reveal distinct boundaries between cTECs and mTECs (Fig. 4a). We first applied reference-based deconvolution methods-CARD, SPOTlight, and Redeconve—and found only CARD could relatively clearly delineate cell boundaries (Fig. 4b). Applying AGED to ST data also accurately separated cell boundaries. However, when projecting cTECs and mTECs separately into space, we found that CARD did not identify these two cell types. By correlating reference-free deconvolution results with ground truth data, we determined CellType 21 corresponds to mTECs and CellType 24 to cTECs. Similarly, spatial projections closely resemble both the deconvoluted results and structures observed in H&E sections. To bolster our findings, we further analyzed the relationship between the two cell types. By identifying marker genes for each, we preliminarily established

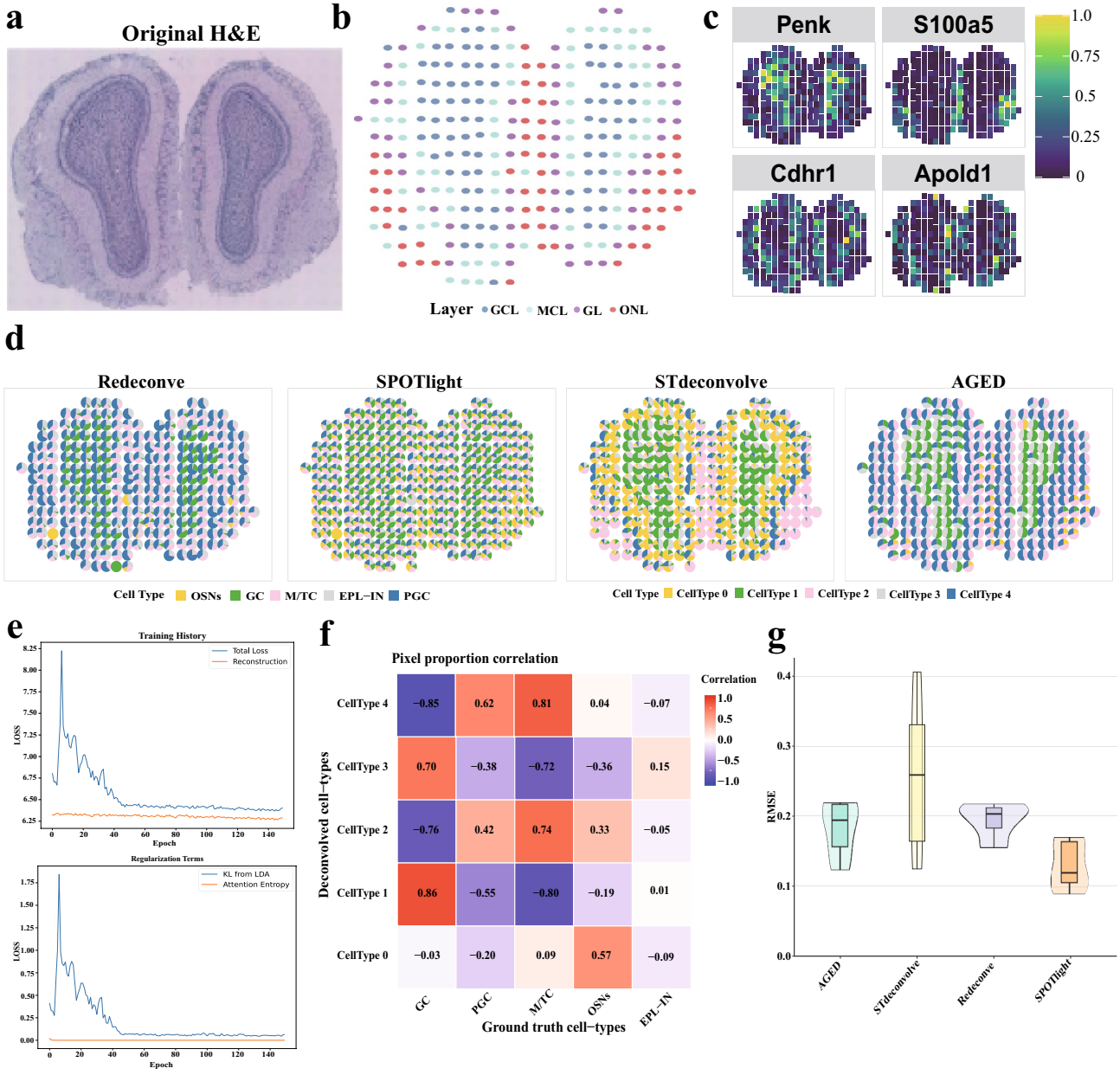

Fig. 2. AGED Deconvolution of MOB Data. **a**. H&E staining of the olfactory bulb. **b**. The AGED displays four anatomical layers arranged from inner to outer: the ganglion cell layer (GCL), the middle cell layer (MCL), the ganglion cell layer (GL), and the outer ganglion cell layer (ONL). **c**. Expression levels of cell type-specific marker genes in the four structural types. **d**. The spatial scatter plot displays the cell type composition inferred at each spatial location using reference-based deconvolution methods and AGED. **e**. Loss Changes in the Training History of AGED. **f**. Correlation between cell types inferred by AGED and Ground Truth cell types. Color is scaled by the correlation value. **g**. RMSE of AGED and other deconvolution tools compared to Ground Truth.

a negative correlation between mTECs and cTECs. This is because the abundance of cTECs and mTECs plays a crucial role in the selection and regulation of T cell development. The abundance of cTECs and mTECs exhibits an opposite trend during the transition from the distal cortical zone to the medullary center. cTECs dominate in the cortical zone and rapidly decline at the M-C boundary, while mTEC numbers significantly increase in the medullary zone, thereby causing this phenomenon[45,46]. Based on this biological insight, we performed a Pearson correlation analysis (Fig. 4c). The results revealed that only the AGED output (Pearson = -0.401) approximated the negative correlation observed with marker genes. This finding conclusively demonstrates that AGED outperforms mainstream reference-based deconvolution tools in reference-free spatial transcriptomics.

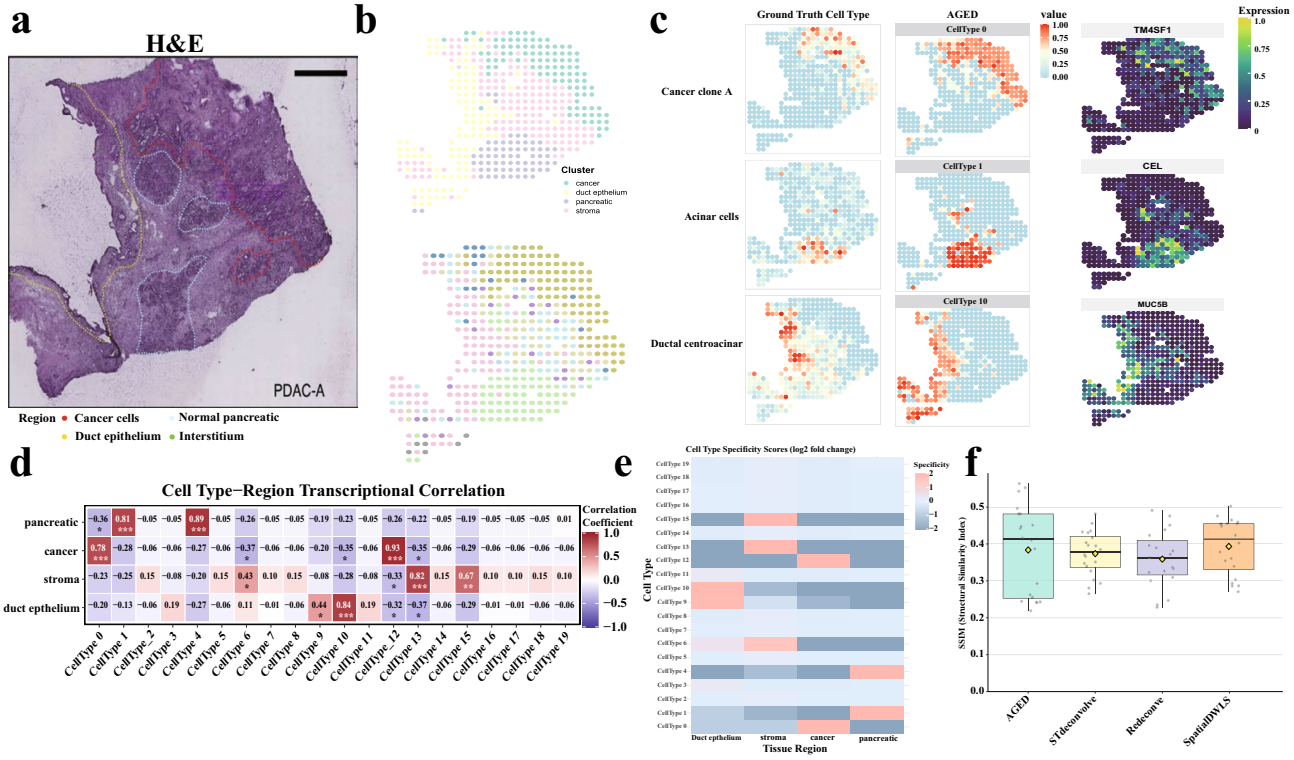

Fig. 3. Analysis of AGED on PDAC Spatial Transcriptome Data. **a**, **b**. H&E staining of the PDAC (left) displays four regions (right) annotated from the original publication. Spatial Landscape of AGED Deconvolution (bottom). **c**. Spatial projections of the three major cell types in PDAC. Left, Ground truth projection. Middle, AGED projection. Right, Expression levels of cell type-specific marker genes. **d**. Correlation between cell types in AGED reference-free deconvoluted samples and four regions. **e**. Cell types and specificity scores for AGED reference-free deconvolution across four regions. **f**. Structural similarity between AGED and deconvolution tools and Ground Truth.

Discussion

We have developed a two-stage computational framework that addresses fundamental challenges in reference-free spatial transcriptomics deconvolution by systematically integrating probabilistic topic modeling with hierarchical attention mechanisms. Our approach makes several methodological advances: automated selection of cell type number through performer, explicit incorporation of probabilistic priors into neural attention architectures, and hierarchical modeling of spatial dependencies at multiple scales. These innovations collectively enable more accurate, interpretable, and biologically meaningful deconvolution of spatial transcriptomics data without requiring external single-cell references.

A central contribution of our work lies in the principled integration of probabilistic-derived priors into deep learning frameworks. Traditional probabilistic models like LDA provide interpretable global structure but lack flexibility to capture complex nonlinear patterns, while purely data-driven neural networks offer representational power but require large datasets and risk overfitting[34,47]. Our probabilistic-guided attention mechanism bridges this gap by encoding probabilistic priors as learnable embeddings that participate in gradient optimization. The gating mechanism adaptively balances prior knowledge with data-driven features, allowing the model to rely on probabilistic guidance when data are noisy or ambiguous while learning to override priors when strong empirical evidence exists. This dynamic weighting represents a departure from fixed hyperparameter-based regularization common in existing methods.

The hierarchical Probabilistic initialization strategy further enhances this integration. By performing coarse-to-fine decomposition before neural refinement, we provide the attention mechanism with structured starting points that reflect both major cell type categories and subtle subtype variations. This initialization proved crucial in our experiments, substantially accelerating convergence compared to random initialization or single-level LDA.

Determining optimal cell type numbers remains challenging in unsupervised deconvolution. Single metrics often conflict: perplexity favors complex models, silhouette scores prefer coarse partitions, and entropy may increase monotonically. Our multi-metric composite scoring recognizes that different measures capture complementary aspects—reconstruction accuracy, cluster separability, and topic diversity. The Performer architecture's O (N log N) complexity makes systematic exploration of K-space computationally feasible.

Our hierarchical attention design explicitly models spatial structure at multiple scales. The spatial attention layer captures local correlation by aggregating information from K-nearest neighbors weighted by distance,

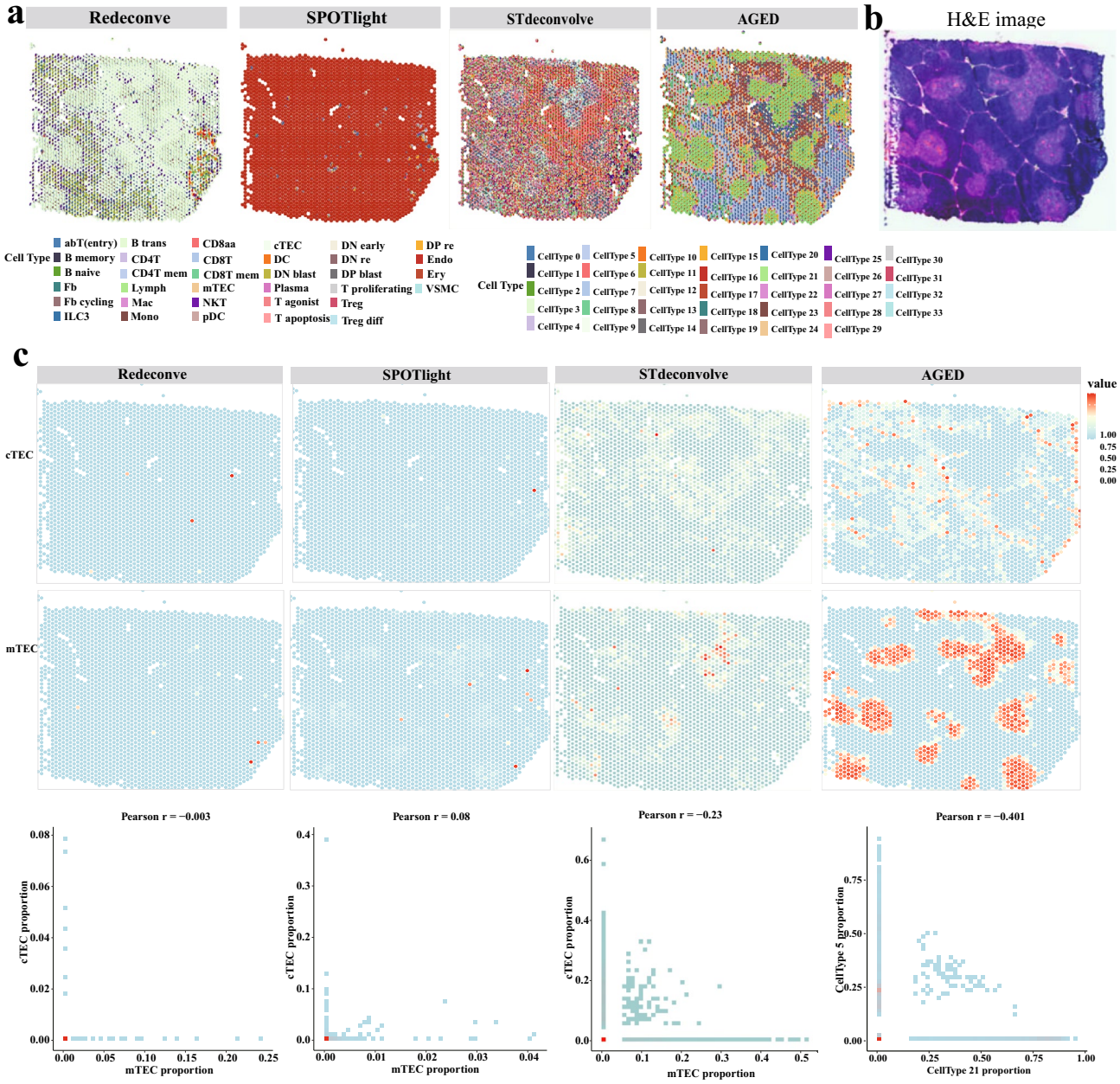

Fig. 4. Analysis of human thymus data. **a**. H&E staining of the Human Thymus. **b**. Spatial scatter pie charts show the cell type composition inferred at each spatial location by different deconvolution methods. The deconvolution methods compared include Redeconve, SPOTlight, STdeconvolve, and AGED. **c**. The negative correlation between cTEC and mTEC was verified using density plots to determine the effectiveness of each deconvolution tool. Using the correlation between marker genes of cTEC and mTEC as a benchmark, only AGED obtained negative correlations among the four deconvolutional.

consistent with the biological principle that adjacent tissue regions exhibit continuity in cellular composition, the topic-gene co-attention identifies distributed expression programs, and Probabilistic-guided attention provides global context. This contrasts with methods applying uniform smoothing—learned attention weights adapt to local tissue properties, preserving sharp boundaries while maintaining continuity in homogeneous regions.

Due to the structured sparsity inherent in biological tissues, each spatial location typically harbors 2-3 dominant cell types[48,49]. Our framework reinforces this characteristic through complementary mechanisms: L1 regularization, top-k thresholding, and attention-based feature selection. Attention-guided sparsification simultaneously integrates local context and global structure, yielding interpretable decompositions where the number of active themes naturally fluctuates with tissue architecture. Compared to Probabilistic initialization, the trained β-matrix exhibits significantly enhanced enrichment effects and specificity for known markers, indicating genuine biological learning rather than noisy overfitting.

Aged also has several limitations. First, this method treats spatial coordinates as fixed values without accounting for technical variability in point localization. Second, despite its reference-free framework, selectively integrating local information such as labeled gene lists could further enhance its performance. Fourth, despite the efficiency gains from the Performer algorithm, processing ultra-large datasets still requires substantial GPU resources.

Future research may extend to other spatial omics modalities, explore ensemble attention mechanisms for joint spot modeling, and develop distributed training strategies for emerging ultra-high-resolution platforms. The success of Probabilistic -guided attention demonstrates that domain-specific probabilistic priors, when appropriately encoded, can significantly enhance neural network architecture performance in data-constrained biological applications.

Method

Aged workflow

For spatial transcriptomics data $X \in \mathbb{R}^{(n\times m)}$ comprising n spots and m genes, we employ a Performer-based architecture to determine the optimal cell type count K*. The independent processing workflow for each probe spot includes: (1) Hierarchical LDA probability initialization for coarse-grained decomposition; (2) Fine-grained decomposition for further subdivision of each cluster; (3) Global refinement integration to generate the thematic distribution θ and gene reconstruction results. (Fig. 1 Stage 1). The second stage of deep refinement of attention guidance (Fig. 1 Stage 2) comprises: (1) the core architecture of AGED: gene embedding → Transformer encoder → spatial attention → theme-gene collaborative attention; (2) incorporating prior information into the probabilistic attention guidance module via cross-attention and gating mechanisms, ultimately yielding sparse outputs θ (cell proportion), β (gene expression profile), and attention weights.

For each candidate value K, we evaluate three metrics: perplexity (reconstruction quality), contour score (clustering separation), and topic entropy (diversity). By applying inflection point detection to the weighted composite score, we balance model complexity and representational power at the point of maximum curvature to select the optimal K* value.

Based on the K* value, we employ hierarchical LDA for initialization, obtaining the prior distributions $\theta\_init \in \mathbb{R}^{(n\times K^*)}$ and $\beta\_init \in \mathbb{R}^{(K^*\times m)}$. For each gene locus, we select the top 200 most highly expressed genes and process them through a hierarchical attention encoder; (2) Guided cross-attention with learnable gating; (3) K-Nearest Neighbor(KNN)[50] Spatial Attention; (4) Topic-Gene Collaborative Attention; (5) Three-layer Transformer architecture (with 8 self-attention heads). Attention pooling mechanisms compress the 200 gene sequences into a fixed 256-dimensional representation.

Two parallel generators produce the final output: θ generator (256→K*) generates sparse cell type proportions via thresholding, while β generator maintains learnable $K^* \times m$ parameters for gene features. Training optimizes a multi-objective loss function:

$$L\_total = \alpha L\_recon + \beta L\_KL + \gamma L\_sparse + \delta L\_distinct \tag{1}$$

where L_recon denotes Poisson negative log-likelihood between reconstructed and observed gene counts; L_KL measures Kullback-Leibler(KL) divergence from probabilistic priors with annealing weight decreasing from 1.0 to 0.1 over 50 epochs; L_sparse applies L1 regularization encouraging each spot to contain few dominant cell types; $L\_distinct$ enforces topic separation by minimizing MSE. The AdamW optimization algorithm (lr=5×10^{-4}) was employed, incorporating cosine annealing and early stopping mechanisms (Fig. 1B). The training history on ST data revealed that Probabilistic guidance dominated during the initial training phase (first 30 epochs), while subsequently learned features took precedence, demonstrating the successful integration of prior knowledge with data-driven refinement. Data details (Supplementary Table 1).

Stage1 Hierarchical LDA

Probabilistic prior

Latent Dirichlet Allocation (LDA) is a classic probabilistic topic model originally developed for text analysis. Within the LDA framework, each document is modeled as a mixture distribution across multiple topics, where each topic represents a probability distribution over vocabulary. Specifically, LDA assumes that the document-topic distribution θ follows a Dirichlet prior Dir (α), while the topic-word distribution β follows Dir (η). For each word w in document d, the generation process is: (1) sample a topic z from θ_d; (2) sample word w from the topic's word distribution β_z.

Spatial transcriptomics analogy. We extend this framework to spatial transcriptomics data: each spatial spot is analogous to a "document," gene expression serves as the "vocabulary," and cell types correspond to "topics." Under this correspondence:

$\theta \in \mathbb{R}^{(n\times K)}$: Spot-cell type distribution matrix, where θ_ik denotes the proportion of cell type k in spot i.

$\beta \in \mathbb{R}^{(K\times m)}$: Cell Type-Gene Distribution Matrix, where β_kj denotes the expression profile of gene j in cell type k.

The observed gene expression matrix can be approximated as: $X \approx \theta\beta$.

This modeling assumes each spot contains a mixture of multiple cell types, each with characteristic gene expression patterns, aligning with the biological reality of tissue heterogeneity.

Standard LDA faces limitations when applying single-layer LDA directly to the entire dataset, presenting two issues: (1) susceptibility to local optima, especially when K is large; (2) difficulty in capturing hierarchical cell type structures (e.g., major categories and subtypes). To circumvent these issues, we employ a three hierarchical strategy:

Step 1 (Coarse-grained clustering): First, perform LDA with a smaller number of themes $K\_coarse = max(3, K/3)$ to identify major cell clusters. This step uses batch learning (max_iter=100) to ensure convergence. The resulting θ_coarse matrix identifies the primary affiliation for each spot.

Step 2 (Fine-Grained Decomposition): For each coarse-grained topic t, select spots with topic weights >0.3 (or top 30% if insufficient). Run LDA with $K\_fine = K/K\_coarse$ topics on this subset. This achieves hierarchical topic discovery—first identifying broad categories, then subdividing subtypes within each category.

Step 3 (Global Integration): Finally, run LDA with K topics on the full dataset, using sparsity priors to encourage each spot to focus on a small number of cell types. This step integrates information from the first two phases, producing the final initializations θ_init and β_init.

Optimal topic number selection with performer

Given the characteristics of spatial transcriptomics data and practical considerations of computational efficiency, we employed a Performer-based topic model in the first phase to automatically determine the optimal number of cell types (K*) present within the tissue. Given spatial transcriptomics data $X \in \mathbb{R}^{(n\times m)}$, where n denotes the number of spots and m denotes genes, we independently process each spot through the following architecture:

Gene and spatial feature encoding. Raw gene expression vectors are first embedded into a 256-dimensional latent space via linear transformation, followed by batch normalization and application of the GELU activation function. Concurrently, spatial coordinates (x, y) undergo independent embedding at the same dimensionality. We integrate the learned two-dimensional position encodings to capture spatial relationships, generating an initial representation of shape [B, 1, 256], where B denotes the batch size and the sequence length is 1 (for single-point processing).

Performer Attention Layers. Utilizing the FAVOR+ (Fast Attention Based on Orthogonal Random Features) algorithm, which approximates standard softmax attention with O (N log N) complexity instead of $O(N^2)$ complexity. Performer's FAVOR+ algorithm offers an ingenious solution. Its core observation is that softmax attention can essentially be viewed as a kernel function:

$$softmax(q^T k) = \frac{exp(q^T k)}{\sum_j exp(q^T k_j)} \propto k(q,k) \tag{2}$$

Here, $q^T k$ is an n×n matrix, where n represents the sequence length (corresponding to the number of spots in the study). For typical spatial transcriptomics data, the number of spots ranges from thousands to tens of thousands. Direct computation of this matrix is impractical both in terms of memory and computational time. But if a feature mapping $\phi(\cdot)$ can be found such that:

$$k(q,k) = exp(q^T k) \approx \phi(q)^T \phi(k) \tag{3}$$

Attention computation can then be restructured. FAVOR+ employs orthogonal random features:

$$\phi(x) = \frac{1}{\sqrt{m}} exp(-\frac{\| x\|^2}{2})[exp({\omega_1}^T x), \ldots, exp({\omega_m}^T x)]^T \tag{4}$$

Among these, $\{w_i\}_{i=1}^m$ are random vectors sampled from $\mathrm{N}(0, \mathrm{I}_d)$ and subsequently orthogonalized, where m denotes the number of random features (typically chosen as $m = O(dlogd)$).

With this approximation, attention can be written as:

$$Attention(Q,K,V) \approx \hat{D}^{-1}(\phi(Q)(\phi(K)^T V)) \tag{5}$$

Here, $\hat{D} = diag(\phi(Q)(\phi(K)^T 1_n))$ is the normalization factor.

The key lies in the order of computation: first compute $\phi(K)^T V$ ((complexity $O(nmd)$), then multiply by $\phi(Q)$, rather than first calculating $\phi(Q)(\phi(K)^T$. This reduces the complexity from $O(n^2 d)$ to $O(nmd)$. Significant acceleration is achieved when $m \ll n$.

Each layer contains 8 attention heads, each with a dimension of 32. The 256-dimensional hidden state undergoes multi-head self-attention processing, followed by a positional feedforward network with residual connections and layer normalization. This architecture maintains a constant dimension [B, 1, 256] across all layers. We chose Performer due to its excellent scalability, enabling processing of large-scale spatial transcriptomics datasets (>10,000 spots) where standard Transformers encounter memory overflow. Additionally, training a 12-layer Performer takes approximately 30-40% of the time required for a standard Transformer of comparable scale. This efficiency stems from the need to evaluate multiple K values when selecting the optimal number of cell types, necessitating training multiple models within the K parameter range. Performer's efficiency enables this process to be completed within a reasonable timeframe.

Dual prediction heads

The final Performer layer outputs two parallel prediction heads: (1) The topic prediction head employs a GELU activation function and dropout for dimensionality reduction (256→128→K), followed by softmax normalization to generate the topic probability distribution $\theta \in \mathbb{R}^{(B\times 1\times K)}$; (2) The gene reconstruction head restores the original gene expression space through feature expansion (256→512→m).

Multi-metric model selection

For each candidate K value within the defined range, we train the full model and evaluate three complementary metrics: (i) Perplexity, calculated via the MSE index between reconstructed and observed gene expression; (ii) Profiling coefficient, measuring clustering separation quality; (iii) Topic entropy, assessing learned topic diversity. The optimal K* value is determined by applying inflection point detection with maximized linear descent deviation on the composite score curve.

Stage 2 attention-guided reference-free deconvolution

Deconvolution is performed via a hierarchical attention architecture, which utilizes probabilistic prior initialization and employs a selected K* number of themes.

Probabilistic Initialization. We first apply a hierarchical latent Dirichlet allocation with K* topics to obtain initial estimates of the topic-spot distribution $\theta\_init \in \mathbb{R}^{(n\times K^*)}$ and the topic-gene distribution $\beta\_init \in \mathbb{R}^{(K^*\times m)}$. This initialization serves a dual purpose: providing reasonable starting values and supplying a prior distribution for regularization.

Sequence Construction. To enhance computational efficiency, we select the top 200 highly expressed genes for each spot based on their count values and construct fixed-length gene index sequences. This generates input data in the form of gene indices paired with corresponding count values, with a shape of [B, 200].

Multi-scale Attention Encoder. This encoder comprises five sequentially arranged attention modules: (a) Gene Embedding Layer, Converts the gene index into a 256-dimensional embedding vector. It enhances the representation by performing expression-weighted attention calculations on log-transformed count values through two network layers (1→64→256), generating a representation vector [B, 200, 256]. (b) Probabilistic-guided attention, by infusing prior knowledge through LDA, we establish a connection between the probabilistic model framework and neural networks. The topic structure (θ and β) learned across the entire dataset via variational inference contains valuable global statistical information: which genes tend to co-express (column patterns in the β matrix) and the distribution patterns of cell types among spots (row patterns in the θ matrix). The core idea of Probabilistic-guided attention is to explicitly inject this prior knowledge into the attention computation, enabling the neural network to learn from an "informed starting point" rather than blindly exploring high-dimensional parameter space. Since LDA outputs θ and β as discrete probability distributions while neural networks operate on continuous vector representations, we need to establish a mapping: Prob (topic | spot, gene) → Embedding $\in \mathbb{R}^d$. This bridge is achieved through a parameterized embedding layer, translating LDA's probabilistic semantics into differentiable vector representations that can participate in gradient optimization. Input features are projected linearly to generate the query vector Q, representing what information the current spot seeks based on feature representations learned from the neural network's earlier layers. The Key and Value come from LDA embeddings. Unlike self-attention, K and V are not generated from the current features but extracted from pre-initialized LDA embeddings. This design ensures each spot queries a fixed set of K global topics rather than other spots' features. This forces the model to learn "spot-topic associations" instead of arbitrary inter-spot dependencies. (c) Spatial Attention Module, encodes pairwise distances between spots using a three-layer network (1→64→128→256), followed by multi-head attention (4 heads) based on K-nearest neighbors to aggregate neighborhood information. (d) Topic-Gene Collaborative Attention Layer, facilitates bidirectional information flow between topic and gene representations through a two-way attention mechanism, enabling topic-aware gene selection. (e) Transformer Encoder Module, employs three standard Transformer layers with 8-head self-attention and feedforward expansion (256→1024→256), maintaining dimensions [B, 200, 256].

Attention pooling

To compress variable-length gene sequences into a fixed representation, we employ learned attention pooling: Attention scores are computed through a two-layer network (256→128→1), normalized via softmax, and weighted-summed to generate a [B, 256] vector.

θ Generator: A three-layer network (256→2K*→K*) with ReLU activation generates unnormalized topic log-likelihood values. We apply temperature-scaled softmax with a learnable temperature parameter, followed by thresholding the top k terms to enforce sparsity, yielding the proportion $\theta \in \mathbb{R}^{(B\times K^*)}$.

β Generator: Maintains a learnable parameter matrix of shape $[K^*, m]$, initialized close to LDA's β_init. After softmax normalization, it directly represents each topic's gene expression features.

Metric for selecting the optimal cell type

a. Perplexity Score

$$Perplexity(K) = \left(exp\frac{1}{nm}\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{m}(X_{ij} - \hat{X}_{ij})^2 \right) \tag{6}$$

Where, X_{ij} is observed gene expression for spot i, gene j. $\hat{X}_{ij} = (\theta\beta)_{ij}$ is reconstructed expression from topic model. n is number of spots, m is number of genes.

b. Silhouette Score

For each spot i:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{7}$$

where, $a(i)$ denotes the average distance between spots within the same cluster, while $b(i)$ represents the average distance to the nearest spot in a different cluster.

Overall Silhouette Score:

$$S_{silhouette}(K) = \frac{1}{n}\sum_{i=1}^{n} s(i) \tag{8}$$

c. Topic Entropy

Average topic distribution across all spots:

$$\overline{\theta_k} = \frac{1}{n}\sum_{i=1}^{n} \theta_{ik} \tag{9}$$

where θ_{ik} is the proportion of topic k in spot i.

Next, calculate Shannon entropy:

$$H(K) = -\sum_{k=1}^{K} \overline{\theta_k} log\overline{\theta_k} \tag{10}$$

Finally, the theme entropy is obtained:

$$S_{entropy}(K) = \frac{H(K)}{logK} \tag{11}$$

Conclusion

We propose a computational framework, AGED, that advances spatial transcriptomics deconvolution by fundamentally integrating probabilistic topic modeling with hierarchical attention mechanisms. This approach embeds probabilistic prior information into learnable neural network components and automates model selection through composite metrics, demonstrating superior accuracy and interpretability compared to existing reference-free methods. The framework's modular design ensures broad applicability across diverse spatial omics platforms and tissue types. Performance validation on MOB, PDAC, and Thymus datasets demonstrates that AGED accurately reconstructs spatial landscapes without reference expression profiles, exhibiting high similarity to H&E tissue sections and outperforming some mainstream reference-based deconvolution tools in accuracy. Notably, its performance on thymus data is remarkable: AGED not only successfully identifies cell boundaries but also identifies a negative correlation between cTECs and mTECs through correlation assignment—result unattainable by other reference-based tools. To evaluate the contribution of each attention mechanism to overall performance, we conducted ablation experiments. Experimental details are provided in Supplementary Figure 2. To demonstrate the applicability of AGED across different resolution platforms, we applied AGED to single-cell resolution MERFISH data from the mouse medial prefrontal cortex and organogenesis seqFISH+ data from mouse organs. We evaluated the deconvolution results using Ground Truth with the Adjusted Rand Index (ARI) and purity metrics. AGED demonstrated outstanding performance on both datasets (Supplementary Figure 3).

Data availability

This study made use of publicly available datasets. These include the MOB dataset (http://www.spatialtranscriptomicsresearch.org), human PDAC dataset (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111672), Human Thymus dataset spatial transcriptomics data obtained from (https://www.nature.com/articles/s41467-024-51767-y#data-availability).

Code availability

AGED is available as an open-source Python software package with the source code available as Supplemental Software and on GitHub at https://github.com/njjyxl/AGED.

Received: 25 November 2025; Accepted: 6 February 2026
Published online: 10 February 2026

Reference

1. Asp, M., Bergenstrahle, J. & Lundeberg, J. Spatially resolved transcriptomes-next generation tools for tissue exploration. *Bioessays* **42**(10), e1900221. https://doi.org/10.1002/bies.201900221 (2020).
2. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat. Methods* **19**(5), 534–546. https://doi.org/10.1038/s41592-022-01409-2 (2022).
3. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**(6294), 78–82. https://doi.org/10.1126/science.aaf2403 (2016).
4. Rodriques, S. G. et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**(6434), 1463–1467. https://doi.org/10.1126/science.aaw1219 (2019).
5. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**(6233), aaa6090. https://doi.org/10.1126/science.aaa6090 (2015).
6. Longo, S. K., Guo, M. G., Ji, A. L. & Khavari, P. A. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat. Rev. Genet.* **22**(10), 627–644. https://doi.org/10.1038/s41576-021-00370-8 (2021).
7. Zeng, H. What is a cell type and how to define it?. *Cell* **185**(15), 2739–2755. https://doi.org/10.1016/j.cell.2022.06.031 (2022).
8. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**(4), 517–526. https://doi.org/10.1038/s41587-021-00830-w (2022).

9. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat. Rev. Genet.* **20**(5), 257–272. https://doi.org/10.1038/s41576-019-0093-7 (2019).
10. Tran, H. T. N. et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* **21**(1), 12. https://doi.org/10.1186/s13059-019-1850-9 (2020).
11. Kleshchevnikov, V. et al. Cell 2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**(5), 661–671. https://doi.org/10.1038/s41587-021-01139-4 (2022).
12. Cable, D. M. et al. Cell type-specific inference of differential expression in spatial transcriptomics. *Nat. Methods* **19**(9), 1076–1087. https://doi.org/10.1038/s41592-022-01575-3 (2022).
13. Andersson, A. et al. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun. Biol.* **3**(1), 565. https://doi.org/10.1038/s42003-020-01247-y (2020).
14. Dong, R. & Yuan, G. C. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biol.* **22**(1), 145. https://doi.org/10.1186/s13059-021-02362-7 (2021).
15. Yang, T. et al. AdRoit is an accurate and robust method to infer complex transcriptome composition. *Commun. Biol.* **4**(1), 1218. https://doi.org/10.1038/s42003-021-02739-1 (2021).
16. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nat. Methods* **18**(11), 1352–1362. https://doi.org/10.1038/s41592-021-01264-7 (2021).
17. Lopez, R. et al. DestVI identifies continuums of cell types in spatial transcriptomics data. *Nat. Biotechnol* **40**(9), 1360–1369. https://doi.org/10.1038/s41587-022-01272-8 (2022).
18. Lahnemann, D. et al. Eleven grand challenges in single-cell data science. *Genome Biol.* **21**(1), 31. https://doi.org/10.1186/s13059-020-1926-6 (2020).
19. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res* **25**(10), 1491–1498. https://doi.org/10.1101/gr.190595.115 (2015).
20. Miller, B. F., Huang, F., Atta, L., Sahoo, A. & Fan, J. Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. *Nat. Commun.* **13**(1), 2339. https://doi.org/10.1038/s41467-022-30033-z (2022).
21. Dries, R. et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**(1), 78. https://doi.org/10.1186/s13059-021-02286-2 (2021).
22. Zhao, E. et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol* **39**(11), 1375–1384. https://doi.org/10.1038/s41587-021-00935-2 (2021).
23. Sun, S., Zhu, J. & Zhou, X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat. Methods* **17**(2), 193–200. https://doi.org/10.1038/s41592-019-0701-7 (2020).
24. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat. Methods* **15**(5), 343–346. https://doi.org/10.1038/nmeth.4636 (2018).
25. Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent dirichlet allocation. *J. Mach. Learn Res.* **3**, 993–1022 (2003).
26. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S An image is worth 16x16 words: transformers for image recognition at scale. *arXiv preprint arXiv:201011929* (2020)
27. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I: Attention is all you need. In: *Proc. of the 31st International Conference on Neural Information Processing Systems.* Long Beach, California, USA: Curran Associates Inc. 6000–6010. (2017)
28. Long, Y. et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun.* **14**(1), 1155. https://doi.org/10.1038/s41467-023-36796-3 (2023).
29. Zeng, Y. et al. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Brief Bioinform.* https://doi.org/10.1093/bib/bbac297 (2022).
30. Pham, D. et al. Robust mapping of spatiotemporal trajectories and cell-cell interactions in healthy and diseased tissues. *Nat. Commun.* **14**(1), 7739. https://doi.org/10.1038/s41467-023-43120-6 (2023).
31. Geirhos, R. et al. Shortcut learning in deep neural networks. *Nat. Mach. Intell.* **2**(11), 665–673. https://doi.org/10.1038/s42256-020-00257-z (2020).
32. Lipton, Z. C. The mythos of model interpretability. *Commun. ACM* **61**(10), 36–43. https://doi.org/10.1145/3233231 (2018).
33. Blei, D. M., Kucukelbir, A. & McAuliffe, J. D. Variational inference: a review for statisticians. *J. Am. Stat. Assoc.* **112**(518), 859–877 (2017).
34. Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **521**(7553), 452–459. https://doi.org/10.1038/nature14541 (2015).
35. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat Methods* **15**(12), 1053–1058. https://doi.org/10.1038/s41592-018-0229-2 (2018).
36. Lotfollahi, M., Wolf, F. A. & Theis, F. J. scGen predicts single-cell perturbation responses. *Nat. Methods* **16**(8), 715–721. https://doi.org/10.1038/s41592-019-0494-8 (2019).
37. Arun R, Suresh V, Madhavan CEV, Murthy MNN On finding the natural number of topics with latent dirichlet allocation: some observations. In: *Proc. of the 14th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining - Volume Part I.* Hyderabad, India: Springer-Verlag 391–402. (2010)
38. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7 (1987).
39. Choromanski K, Likhosherstov V, Dohan D, Song X, Gane A, Sarlos T, Hawkins P, Davis J, Mohiuddin A, Kaiser L Rethinking attention with performers. *arXiv preprint arXiv:200914794* 2020 (2020)
40. Tepe, B. et al. Single-Cell RNA-seq of mouse olfactory bulb reveals cellular heterogeneity and activity-dependent molecular census of adult-born neurons. *Cell Rep* **25**(10), 2689–2703. https://doi.org/10.1016/j.celrep.2018.11.034 (2018).
41. Zhou, Z., Zhong, Y., Zhang, Z. & Ren, X. Spatial transcriptomics deconvolution at single-cell resolution using redeconve. *Nat. Commun.* **14**(1), 7930. https://doi.org/10.1038/s41467-023-43600-9 (2023).
42. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**(9), e50. https://doi.org/10.1093/nar/gkab043 (2021).
43. Moncada, R. et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**(3), 333–342. https://doi.org/10.1038/s41587-019-0392-8 (2020).
44. Xu, H. et al. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Med.* **16**(1), 12. https://doi.org/10.1186/s13073-024-01283-x (2024).
45. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat. Rev Immunol.* **14**(6), 377–391. https://doi.org/10.1038/nri3667 (2014).
46. Li, Y. et al. Development of double-positive thymocytes at single-cell resolution. *Genome Med.* **13**(1), 49. https://doi.org/10.1186/s13073-021-00861-7 (2021).
47. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**(7553), 436–444. https://doi.org/10.1038/nature14539 (2015).
48. Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**(7124), 168–176. https://doi.org/10.1038/nature05453 (2007).
49. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**(4), 999–1014. https://doi.org/10.1016/j.cell.2018.06.021 (2018).

50. Cover, T. & Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory* **13**(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964 (1967).

Acknowledgements
We thank all the authors involved in this study for data collection, preparation, quality control and manuscript writing. This work was carried out on the High-performance Computing platform of Yunnan Minzu University.

Author contributions
X.Y. conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft. J.Y.W. conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft. X.Z.C. conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Funding
This research was funded by the National Natural Science Foundation of China of XZC, grant number 31460297. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-39703-0.

Correspondence and requests for materials should be addressed to Y.W. or X.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.