nature communications

Article

https://doi.org/10.1038/s41467-026-76552-x

Enhancing pan-cancer spatial transcriptomics at single-cell resolution with stPainter

Received: 9 April 2026

Accepted: 3 August 2026

Published online: 10 August 2026

Check for updates

Yuhang Yang[1,4], Yiming Luo[2,4], Kai Zhang[1] ✉, Zaixi Zhang[3], Haoxin Peng[2], Chenlin Cao[2], Qi Liu[1], Bin Ma[1], Yang Chen[2] ✉, Lin Shen[2] ✉ & Enhong Chen[1] ✉

Subcellular spatial transcriptomics can resolve tissue architecture at cellular scale, but sparse gene panels and limited detection sensitivity constrain downstream analysis. Existing enhancement methods often require tissue-matched single-cell RNA sequencing (scRNA-seq) references and dataset-specific retraining. Here we show that stPainter, a conditional generative model pretrained on a pan-cancer scRNA-seq atlas, can enhance spatial transcriptomics data without matched references or retraining. Using a latent diffusion architecture guided by Stochastic Differential Equations (SDE), stPainter reconstructs expanded expression profiles from sparse measurements and produces latent representations for clustering and cell-state analysis. When we apply stPainter upon 6 spatial transcriptomics datasets of different cancer types, we demonstrate that our model empowers downstream biological analyses, including fine-grained subpopulation clustering and pathway enrichment. Comparison with spatially resolved proteomics (CODEX) provided independent support for regional agreement between imputed cellular compositions and protein-level tissue organization. These results establish stPainter as a scalable approach for analyzing tumor microenvironments without auxiliary sequencing data.

The advent of spatial transcriptomics (ST) has enabled the systematic interrogation of gene expression within intact tissues, preserving spatial organization that is lost in dissociative single-cell RNA sequencing (scRNA-seq) workflows[1,2]. While scRNA-seq provides a molecular census of cellular identities and states, it lacks the spatial context that is essential for understanding tissue architecture and cell-cell interactions. Recent advances in high-resolution ST technologies, particularly image-based platforms such as 10 × Genomics Xenium and NanoString CosMx[3,4], have pushed spatial profiling to true single-cell and subcellular resolution with single-molecule sensitivity[5]. In contrast to sequencing-based ST methods that measure transcript abundance at multi-cellular spot levels[6–8], image-based ST enables localization of transcripts within individual cells, producing cell-level expression matrices that are compatible with many single-cell analytical frameworks[4]. This convergence allows established single-cell analytical frameworks to be extended into the spatial domain, establishing image-based ST as a modality for in situ dissection of complex tissue ecosystems such as the tumor microenvironment.

Despite these advances, single-cell resolution ST data suffer from technical limitations that hinder downstream analysis. A primary challenge is limited RNA capture, resulting in high sparsity and technical noise, which obscures subtle biological signals[9,10]. Furthermore, image-based probe hybridization techniques used for single-cell resolution ST often rely on targeted gene panels, capturing only a

[1]State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China, Hefei, China. [2]Department of Gastrointestinal Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing, China. [3]Princeton University, Princeton, NJ, USA. [4]These authors contributed equally: Yuhang Yang, Yiming Luo. ✉e-mail: kkzhang08@ustc.edu.cn; yang_chen@bjcancer.org.cn; shenlin@bjmu.edu.cn; cheneh@ustc.edu.cn

fraction of the transcriptome, with panels containing up to 5000 genes in Xenium 5 k. Consequently, the sparsity and limited coverage compromise unsupervised clustering and annotation, the identification of differentially expressed genes (DEGs) and spatial niche characteristics, particularly for low-abundance transcripts.

To address these limitations of ST data, researchers have developed computational strategies for gene imputation, which integrate transcriptomic profiles from scRNA-seq to enhance spatial fidelity. These approaches broadly bifurcate into alignment-based[11–18] and generative-based[19–24] paradigms. The former focuses on establishing direct correspondence between modalities; examples include Seurat[11], which utilizes probabilistic transfer anchors to align reference phenotypes, Tangram[12], which applies a deep learning framework to map scRNA-seq cells to spatial spots by minimizing feature-level alignment costs, NovoSpaRc[15], which postulates a structural correspondence between transcriptomic similarity and physical proximity to probabilistically map cells via Gromov-Wasserstein optimal transport, and SpaOTsc[16], which employs structured and unbalanced optimal transport to address population mismatches, deriving a spatial metric to reconstruct cell locations and infer intercellular communication networks. Conversely, generative-based methods go beyond simple mapping by leveraging deep architectures to explicitly learn and internalize intrinsic cellular gene expression patterns. This category encompasses diverse advanced frameworks: gimVI[19] pioneers the use of variational autoencoders for probabilistic integration, stDiff[22] harnesses diffusion models to reconstruct spatial gradients, and the recently introduced SpaIM[23] utilizes style transfer learning to effectively disentangle modality-specific styles from shared biological content. Ultimately, this generative paradigm can model expression patterns from learned data distributions rather than relying solely on rigid reference retrieval.

Notwithstanding the significant strides achieved by prior gene imputation strategies, critical limitations persist that impede their broad applicability and downstream efficacy. Firstly, the majority of existing models operate under a strictly supervised paradigm, necessitating scRNA-seq from the same samples references. This dependency renders them incapable of generalizing to unseen tissues, compelling computationally expensive retraining for every new biological scenario. Secondly, although these methods successfully expand the gene panel, the resulting imputed profiles often retain the high dimensionality and stochastic noise characteristic of ST data. Consequently, the enhanced data remains suboptimal for direct usage, necessitating secondary dimensionality reduction (e.g., PCA or encoders) to extract usable biological signals for downstream analysis. Finally, most existing approaches are optimized for coarser spot-level resolutions. When applied to cell-level data characterized by higher throughput but a limited number of measured genes, these methods frequently suffer from computational inefficiencies and noise amplification.

In this work, we introduce stPainter, a latent diffusion model pretrained on a large pan-cancer scRNA-seq dataset to enhance ST data across diverse histological landscapes. Leveraging a pretraining atlas, stPainterenables direct application to diverse tumors without retraining. Methodologically, stPainter employs a conditional latent diffusion framework[25] to capture shared and distinct gene expression patterns across malignancies. It compresses high-dimensional gene profiles into a compact latent space, where we simulate a Stochastic Differential Equation (SDE)[26] to enhance features before decoding them back to the gene space. Ultimately, our model yields dual outputs for each cell: a latent embedding for downstream clustering and subpopulation analysis, and an imputed gene matrix for differential expression analysis. In the evaluated benchmarks, latent-space inference reduced computational costs relative to several comparison methods, and stPainter achieved higher performance on multiple evaluation metrics. Subsequent analyses further substantiate its utility for downstream tasks. stPainteris available as open-source software on GitHub (https://github.com/yyh030806/stPainter), with detailed but user-friendly tutorials demonstrating its capabilities in enhancing the utility of single-cell resolution ST.

Results

Overview of stPainter

We introduce stPainter, a generative framework designed to extend conventional spatial imputation by establishing a scalable pretrained and inference paradigm (Fig. 1a). Unlike existing methods that necessitate strictly paired references and supervised mapping, thereby severely restricting generalizability, stPainter leverages a massive pan-cancer scRNA-seq atlas, encompassing 1.3 million cells across 21 diverse cancer types, to learn a broad pan-cancer cellular prior (Fig. 1b). This data-centric approach enables the model to capture a broad pan-cancer transcriptional manifold, facilitating reference-free generalization to diverse histological landscapes without the need for dataset-specific retraining or matched reference data. By learning from this vast heterogeneity, stPainter effectively bridges the gap between the high-resolution ST and the gene coverage of single-cell sequencing.

To model this complex distribution, stPainter employs a latent diffusion architecture (Fig. 1c). First, a Variational Autoencoder (VAE) maps high-dimensional expression profiles into a compact latent manifold, filtering out technical noise while preserving biological variance. Within this manifold, we train a Gene Diffusion Transformer (GiT) as the conditional generative model, which explicitly conditions the generation process on cancer type. Mathematically, the inference is formulated as a guided generation problem via the SDE. Instead of generating cells from pure noise, we conceptualize the ST measurement as a corrupted state. The model initializes the reverse diffusion process with the latent representation of the observed ST data and iteratively refines it using the learned generative priors. This process effectively projects partial spatial measurements onto the biologically informed cellular manifold, recovering missing information consistent with the global biological context.

Ultimately, stPainter supports downstream characterization through its dual outputs: a denoised latent embedding and a fully imputed genome-wide expression matrix (Fig. 1d). The topology-preserving latent embedding enhances unsupervised clustering and subpopulation analysis, resolving fine-grained cellular heterogeneity often obscured by sparsity. Simultaneously, the imputed biologically informed matrix enables the precise identification of differentially expressed genes (DEGs), facilitating the discovery of cell-group transcriptomic markers. By expanding gene signatures within their spatial context, stPainter facilitates downstream analysis of TME organization and spatial niches.

stPainter improves recovery of spatial gene expression

To evaluate the capability of stPainter in recovering spatial gene expression, we implemented two variants of our model: stPainter-100 (using a latent dimension of 100) and stPainter-50 (using a latent dimension of 50). We conducted comparative experiments on three cancer types with paired scRNA-seq references: Colorectal Cancer (COAD), Ovarian Cancer (OV), and Liver Hepatocellular Carcinoma (LIHC). We benchmarked performance against nine existing baseline methods: Tangram[12], NovoSpaRc[15], SpaOTsc[16], PASTA[18], gimVI[19], stDiff[22], SpatialScope[20], VISTA[24] and SpaIM[23]. Performance was comprehensively assessed using four quantitative metrics: Pearson Correlation Coefficient (PCC), Structural Similarity Index Measure (SSIM), Root Mean Square Error (RMSE), and Jensen-Shannon divergence (JS).

To further assess reference-free generalization, we applied the pretrained stPainter model directly to independent Xenium datasets without using a dataset-matched scRNA-seq reference from the same

https://doi.org/10.1038/s41467-026-76552-x

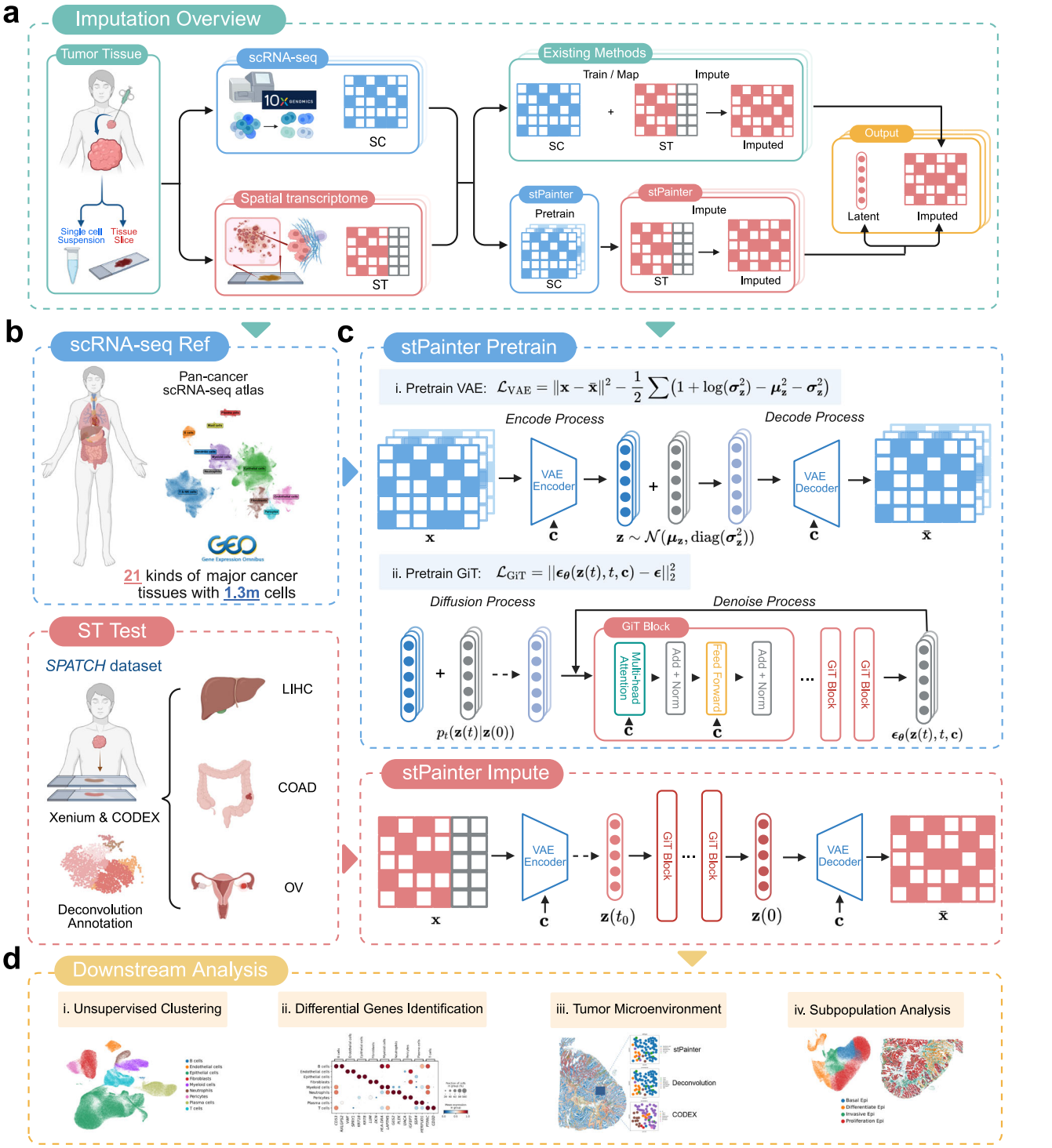


Fig. 1 | Overview of stPainter. a stPainter adopts a pretraining paradigm using unpaired single-cell RNA sequencing (scRNA-seq) data to generate imputed expression profiles and latent embeddings, unlike supervised methods that require paired references. **b** Pretrained on a pan-cancer atlas, the model is applied to spatial transcriptomics (ST) data from different cancer types without dataset-specific retraining. **c** stPainter separates pretraining from imputation; the training phase sequentially optimizes the variational autoencoder (VAE) and Gene Diffusion Transformer (GiT) to learn generative priors for subsequent inference. **d** The enhanced data support downstream analyses, including unsupervised clustering, differential gene expression analysis and tumor microenvironment characterization. SC denotes single-cell data, GEO denotes Gene Expression Omnibus, and CODEX denotes co-detection by indexing. Colors distinguish the data modalities, model stages and downstream analysis modules shown in the workflow. Created in BioRender. Luo, Y. (2026) https://BioRender.com/ixk4is0.

sample and without dataset-specific retraining or fine-tuning. This reference-free evaluation included Cervical Squamous Cell Carcinoma (CESC), Non-Small Cell Lung Cancer (NSCLC), and Prostate Adenocarcinoma (PRAD) datasets (Supplementary Fig. 4). In contrast to the paired-reference benchmarking experiments performed on COAD, OV, and LIHC, these analyses were designed to evaluate whether stPainter can support downstream ST enhancement and cell-state characterization when no matched scRNA-seq reference is available. Beyond reference-free inference within the training domain, we evaluated stPainter on a Multiple Myeloma Xenium 5 k dataset, a cancer

type entirely absent from the pretraining atlas due to the substantial biological differences between blood cancers and solid tumors. As shown in Supplementary Fig. 15, stPainter successfully recovered gene expression profiles and resolved cell-type clusters in this unseen pathological context, providing evidence for generalizability beyond the training distribution. Finally, to assess cross-platform transferability, we applied stPainterto datasets acquired on MERFISH and CosMx platforms, including MERFISH-COAD, CosMx-GBM, and MERFISH-BRCA comprising five tissue regions (Supplementary Fig. 16). Consistent performance across these platforms demonstrates that stPainter is applicable to multiple single-cell resolution spatial transcriptomics platforms.

Quantitative benchmarking across varying numbers of highly variable genes (HVGs) further underscores the robust performance and scalability of stPainter (Fig. 2a). Specifically, on the COAD dataset, both stPainter-100 and stPainter-50 consistently surpassed all baseline methods across the four evaluation metrics. As the number of imputed genes scaled from 10 to 300, stPainter-100 maintained a superior Pearson Correlation Coefficient (PCC). This metric peaked at 0.31 for the top 100 genes, which represents a margin over the leading generative baseline gimVI (PCC = 0.24) and alignment-based Tangram (PCC = 0.24). The advantage of stPainter was most pronounced in structural fidelity as measured by the SSIM. While stPainter variants sustained SSIM scores above 0.30 across the entire gene range, other methods, particularly stDiff and SpaIM, yielded values below 0.01. These near-zero results indicate an inability to reconstruct coherent spatial expression manifolds. Furthermore, stPainterachieved the lowest global error (RMSE = 1.16) and statistical divergence (JS = 0.48) at the 100 HVG scale. These results suggest that the imputed profiles closely mirror the absolute abundance and distribution of the ground-truth transcripts. Notably, the performance of stPainterremained stable as the complexity of the gene panel increased, whereas baseline methods often exhibited diminishing returns. This consistency highlights the effectiveness of the broad pan-cancer cellular prior in regularizing the imputation process.

We further examined the imputation quality of biologically significant marker genes to assess spatial consistency. As illustrated in Fig. 2b, stPainter improved the recovery of the spatial distributions of key immune and structural markers, including *CD4* and *CD8A* (T cells), *CD68* (macrophages), *EPCAM* (epithelial cells), *ENG* (endothelial cells), and *POSTN* (fibroblasts). The imputed gene expression patterns generated by stPainter were broadly consistent with both the ground-truth Xenium measurements and the cell-type spatial distributions derived from CODEX protein imaging. To further assess whether the imputed marker genes faithfully reflected the in situ tumor microenvironment, we quantified the correlation of key gene expression profiles within corresponding 100μm × 100μm spatial grids. As shown in Fig. 2c, stPainter achieved the highest correlation with the ground-truth data, particularly for T-cell markers *CD4* and *CD8A*, as well as the vascular marker *ENG*. In contrast, baseline methods frequently produced blurred spatial signals or artifacts, which compromised the precise localization of these biologically relevant markers. Finally, we extended this evaluation to the OV and LIHC datasets to ensure robustness across different tissue types. Consistent with the findings in COAD, stPainter demonstrated robust performance advantages in these cancer types as well, the detailed comparative results of which are presented in Supplementary Fig. 3a–d.

Across datasets from different cancer types, we observed substantial variability in gene imputation performance. We therefore examined the factors associated with imputation quality and found that performance was primarily linked to raw dataset quality, including UMI counts, the median number of detected genes, and sparsity (Supplementary Fig. 17a, b). In contrast, imputation performance showed no clear association with gene-level spatial aggregation, as measured by Top50 gene Moran's I, or with the representation of each cancer type in the pretrained pan-cancer dataset (Supplementary Fig. 17b–d).

stPainter enhances cell-type identification and spatial mapping

Accurate cell type identification is a prerequisite for interpreting ST data. We evaluated whether the latent representations and imputed profiles generated by stPainter could improve unsupervised clustering performance compared to raw data and baseline methods.

To quantitatively evaluate the capacity of stPainter in resolving spatial domains, we benchmarked its two variants (stPainter-100 and stPainter-50) against six existing methods using four complementary clustering metrics: Adjusted Rand Index (ARI), Adjusted Mutual Information (AMI), Homogeneity (Homo), and Normalized Mutual Information (NMI). The SPATCH dataset provides transfer labels derived from paired scRNA-seq data[27]. We validated the biological fidelity of the expression profiles and used these labels as reference annotations for clustering evaluation (Supplementary Fig. 17). The quantitative results on the COAD dataset (Fig. 3a) demonstrate that stPainter consistently outperforms all baseline methods across all four metrics. Specifically, stPainter-100 achieved the highest scores, with an ARI of 0.85, AMI of 0.70, Homo of 0.63, and NMI of 0.70. Even with a reduced latent dimension, stPainter-50 maintained competitive performance (ARI = 0.80, AMI = 0.65, Homo = 0.55, NMI = 0.65), still exceeding the best-performing generative baseline, SpaIM (ARI = 0.57, AMI = 0.52, Homo = 0.37, NMI = 0.52), and alignment-based methods like Tangram (ARI = 0.56, AMI = 0.50, Homo = 0.35, NMI = 0.50). In contrast, other baseline tools such as gimVI and stDiff showed substantially lower scores, while NovoSpaRc and SpaOTsc failed to provide biologically meaningful clustering results, yielding near-zero values across the board. These consistent gains across multiple metrics suggest that by leveraging a broad pan-cancer cellular prior, stPaintereffectively reconstructs a biologically informed manifold that is essential for accurate and robust tissue segmentation in complex cancer landscapes.

To assess the concordance between unsupervised clusters and biological identities, we used Sankey diagrams to map Leiden clusters derived from the imputed data to transfer-annotation labels (Fig. 3b–e). At low resolution ($r = 0.02$), clustering results from stPainter-100 and stPainter-50 showed clear and coherent mappings for major cell types (Fig. 3b, c). In contrast, under the same resolution, baseline methods such as gimVI and Tangram produced more fragmented clustering patterns. Their Sankey diagrams exhibited pronounced entanglement, with individual cell types split across multiple noisy clusters or, conversely, distinct cell types merged into a single cluster (Fig. 3d, e).

stPainter resolves spatial cellular heterogeneity from imputed transcriptomes

Single-cell resolution ST platforms, such as Xenium 5k, pose substantial challenges for downstream analyses, including unsupervised clustering and cell type annotation[10]. We evaluated whether gene imputation with stPainter could improve the downstream interpretability of Xenium data while preserving biologically meaningful spatial and transcriptional structure. Applying stPainter-100 to Xenium COAD datasets, we reconstructed transcriptomes comprising 10,000 genes together with compact latent representations (Fig. 4a). Latent-driven UMAP embeddings revealed well-separated and structured cell clusters that resembled major lineage organization observed in high-quality scRNA-seq references. These clusters exhibited strong concordance with cell type labels derived from transfer annotation. Moreover, the selective and robust expression of canonical marker genes, such as *KRT19* in epithelial cells and *PDGFRB* in smooth muscle cells, supporting the ability of stPainter to enhance marker-supported transcriptomic resolution without disrupting major biological structure.

https://doi.org/10.1038/s41467-026-76552-x

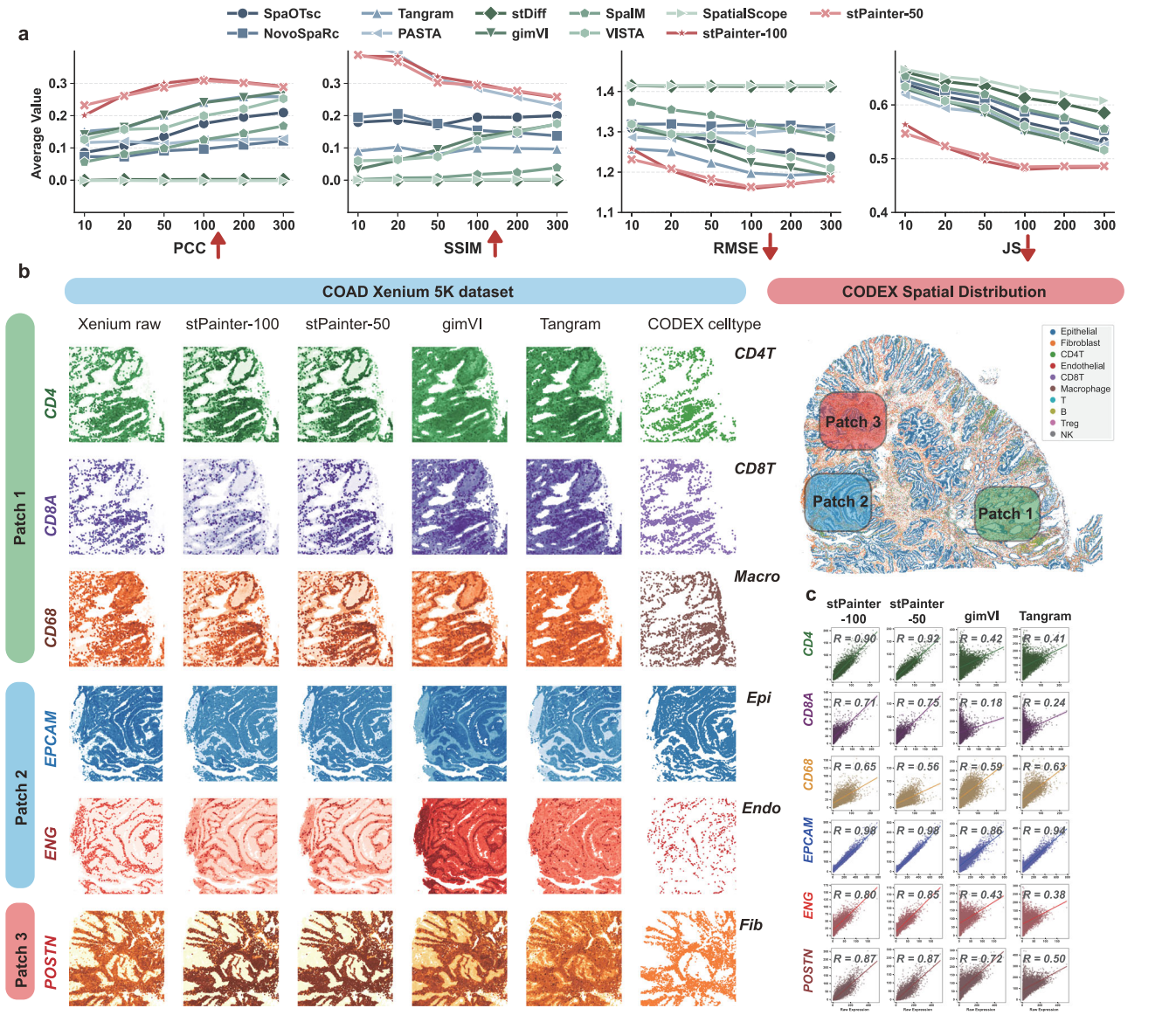


Fig. 2 | Biologically informed recovery of spatial gene expression by stPainter. **a** Quantitative benchmarking of gene expression recovery on the colon adenocarcinoma (COAD) dataset. Performance was evaluated using Pearson correlation coefficient (PCC), structural similarity index measure (SSIM), root mean square error (RMSE) and Jensen--Shannon divergence (JS). Upward and downward arrows indicate whether higher or lower values represent better performance, respectively. Colors and marker shapes denote the indicated computational methods. **b** Visual comparison of spatially resolved marker-gene expression across three representative tissue patches. Color intensity represents relative gene expression, and colors in the CODEX maps denote the indicated protein-defined cell types. CODEX denotes co-detection by indexing; Macro, Epi, Endo and Fib denote macrophage, epithelial, endothelial and fibroblast annotations, respectively. **c** Gene-wise scatter plots showing the correlations between imputed expression levels and raw Xenium measurements for selected marker genes. Pearson's *r* values indicate correlations calculated within 100 μm × 100 μm spatial regions. Source data are provided as a Source Data file.

We further benchmarked stPainter against baseline tools and the standard Xenium processing workflow (PCA-based clustering utilizing transfer-annotation labels)[28]. In contrast to the raw data and other imputation methods, stPainter produced UMAP topologies with clearer separation of major cellular populations and stronger marker-supported biological interpretability (Fig. 4b). Notably, competing methods frequently exhibited over-smoothed trajectories in UMAP space. Such continuous structures likely represent technical artifacts resulting from the generation of dense matrices without adequately modeling the inherent sparsity of ST data, rather than discrete biological populations. Consequently, these artifacts obscured distinct cell populations, leading to lower concordance with reference-supported annotations compared with stPainter.

Finally, we assessed whether the imputed data preserved spatial organization within the tissue microenvironment (Fig. 4c). We found that cell type annotations derived from stPainter exhibited high consistency with transfer-annotation results. To quantitatively assess this, we segmented the tissue sections into 100 μm × 100 μm patches and compared the cellular composition against orthogonal CODEX protein imaging from adjacent serial sections. At the patch level, our analysis revealed that stPainter-derived cellular compositions achieved a high correlation with CODEX-derived protein-level cellular patterns, comparable to transfer-annotation. This high degree of concordance was particularly evident in key lineages, including epithelial cells and T cells, supporting the use of stPainterfor recovering biologically interpretable spatial cellular distributions.

https://doi.org/10.1038/s41467-026-76552-x

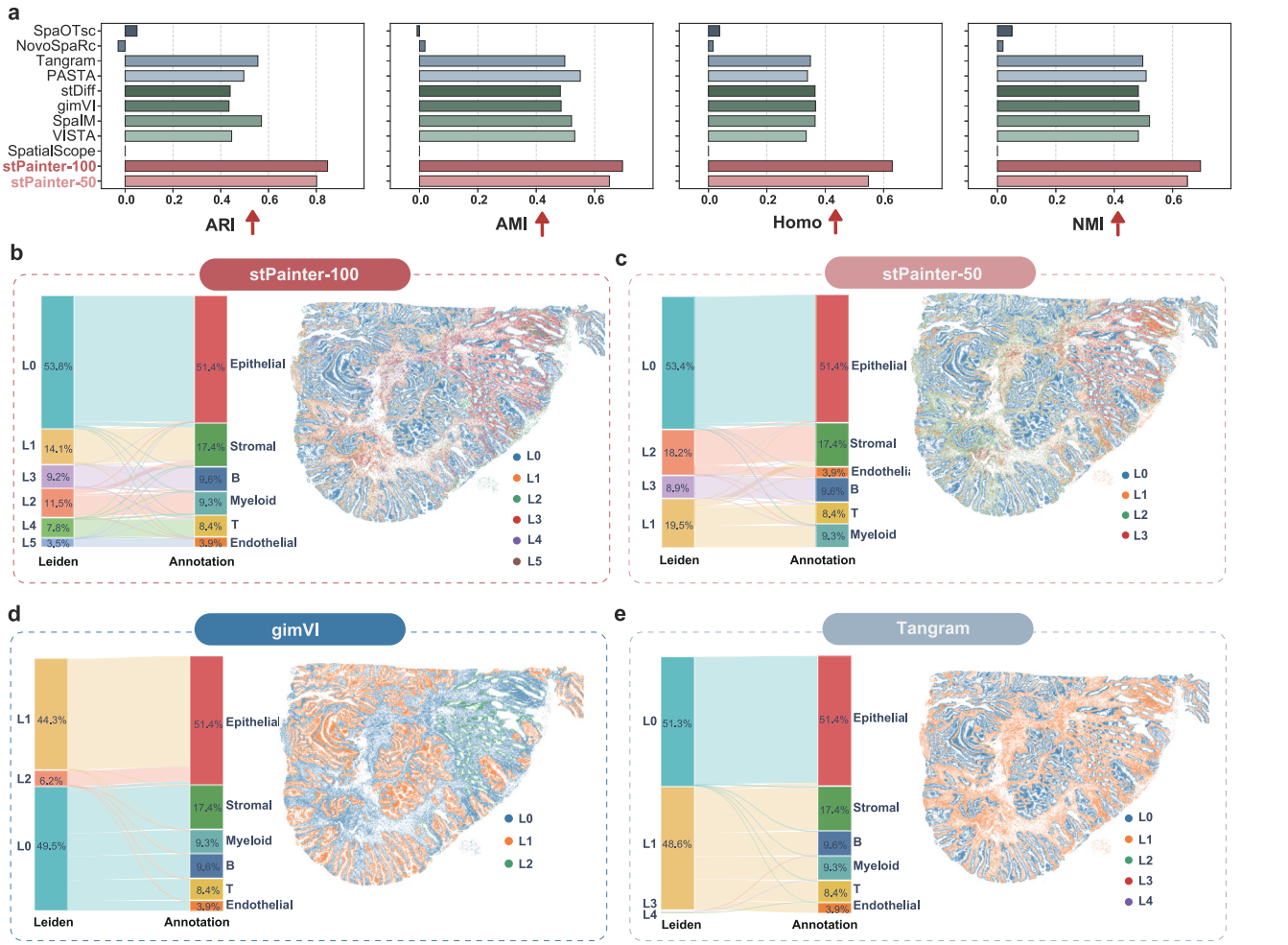


Fig. 3 | Benchmarking of unsupervised clustering with imputed Xenium data. **a** Quantitative evaluation of clustering performance on the COAD Xenium dataset using adjusted Rand index (ARI), adjusted mutual information (AMI), homogeneity (Homo), and normalized mutual information (NMI). Upward arrows indicate that higher values represent better performance. Bar colors denote the indicated computational methods. **b**–**e** Sankey diagrams and corresponding spatial cluster maps illustrating the correspondence between unsupervised Leiden clusters (left nodes) and transfer-annotation cell types (right nodes). Colors denote the Leiden clusters and transfer-annotation cell types indicated by the corresponding keys. **b**, **c** Clustering results derived from the latent embeddings of stPainter-100 (**b**) and stPainter-50 (**c**). **d**, **e** Clustering results obtained using gimVI (**d**) and Tangram (**e**), showing greater mixing between Leiden clusters and transfer-annotation cell types. COAD denotes colon adenocarcinoma. Source data are provided as a Source Data file.

stPainter enables fine-grained dissection of cell subpopulations

Fine-grained cell annotation is pivotal for dissecting tissue heterogeneity in solid tumors, particularly within the immune and epithelial compartments[29]. However, identifying rare subpopulations in raw ST data remains challenging, where lineage-specific marker genes often suffer from low detection efficiency[27]. To overcome this limitation, we isolated T cells, myeloid cells, and epithelial cells from the stPainter-imputed COAD Xenium dataset and performed high-resolution subclustering with stPainter-latent.

Within the T cell compartment, subclustering based on the stPainter latent representations separated six T cell states, including CD4+ memory T cells (Tm), naive T cells (Tn), regulatory T cells (Treg), CD8+ effector T cells (Teff), exhausted T cells (Tex), and proliferating T cells (Tprolif) (Fig. 5a). stPainter increased the detectability of canonical marker genes associated with these states, as *FOXP3* in Tregs, *CCR7* in $CD4^+$ Tn, and *CXCR5* in $CD8^+$ Tex, thereby improving state interpretation of T cell heterogeneity[30]. To assess spatial concordance at the tissue-region level, we compared these annotations with protein-based T cell groups derived from adjacent CODEX sections. Pearson correlation analysis of cell densities within 100 µm × 100 µm patches revealed regional concordance between stPainter-derived T cell compositions and CODEX-derived protein-level patterns (Fig. 5b, d). Similarly, subclustering also identified fine-grained populations within the myeloid compartment. Beyond major populations such as macrophages, monocytes, neutrophils, and mast cells, the model identified marker-supported rare or less abundant subsets like regulatory dendritic cells (mregDC), characterized by the specific expression of *CCR7* and *LAMP3* (Fig. 5e–h)[31].

Epithelial cells are fundamental to understanding tumor heterogeneity. We stratified epithelial cells into five distinct functional states: Differentiated, Basal, Stem, Proliferative, and Invasive Epithelial cells (Fig. 5i). Differential expression analysis supported the biological annotation of these epithelial states: Differentiated cells expressed *CDX2*[32], Basal cells were marked by *ITGA6*[33], while Stem populations showed elevated *OLFM4*[34]. Notably, the Invasive cluster was enriched for genes associated with Epithelial-Mesenchymal Transition (EMT), including *TGFB1* and *MACC1* (Fig. 5k). Spatial mapping revealed distinct morphological contexts: Differentiated cells formed organized glandular structures, whereas Invasive cells were preferentially localized at the invasive tumor front (Fig. 5j, l)[35]. These findings demonstrate that stPainter-derived epithelial states are spatially interpretable, providing hypotheses about functional organization during tumor progression.

https://doi.org/10.1038/s41467-026-76552-x

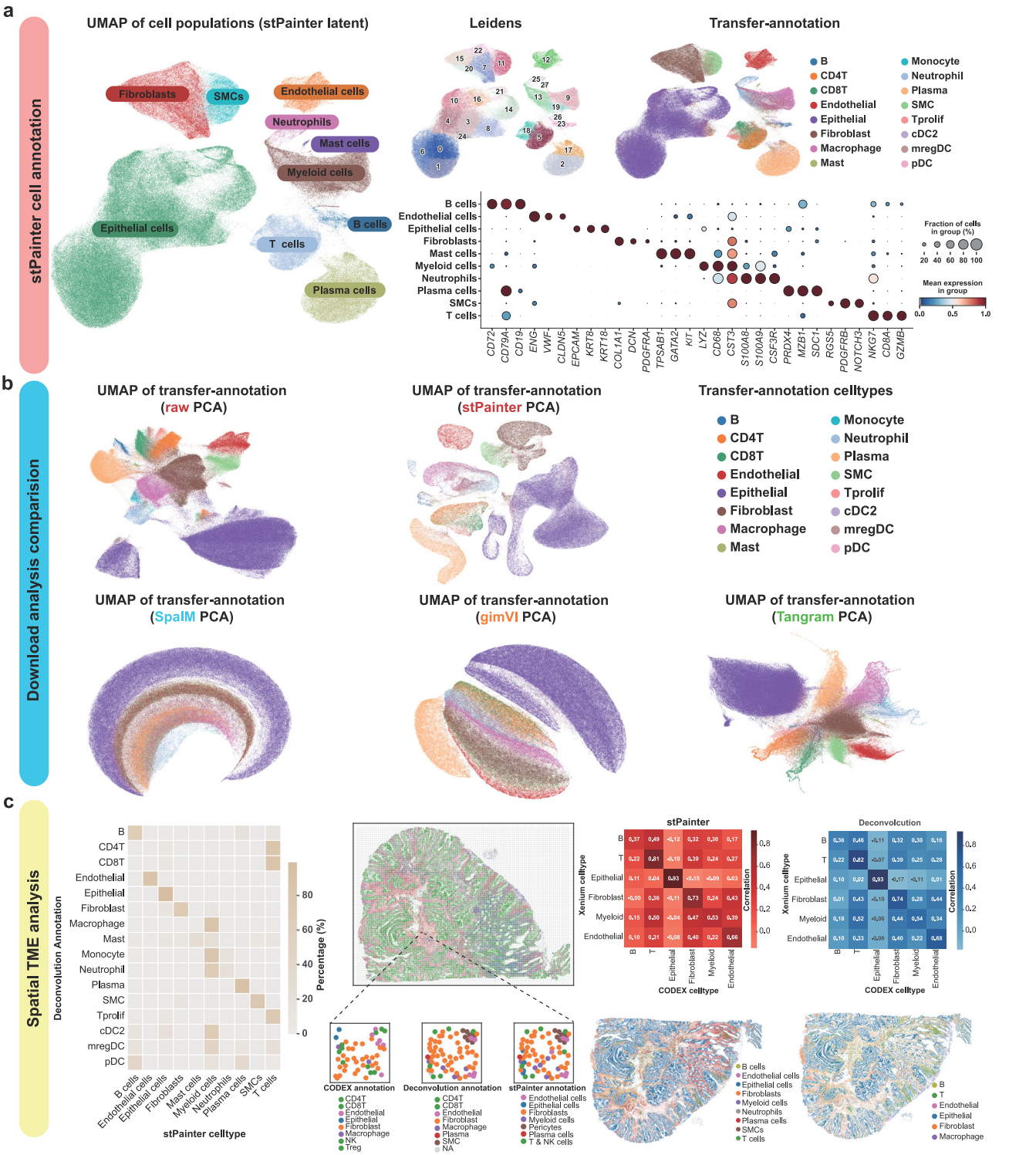
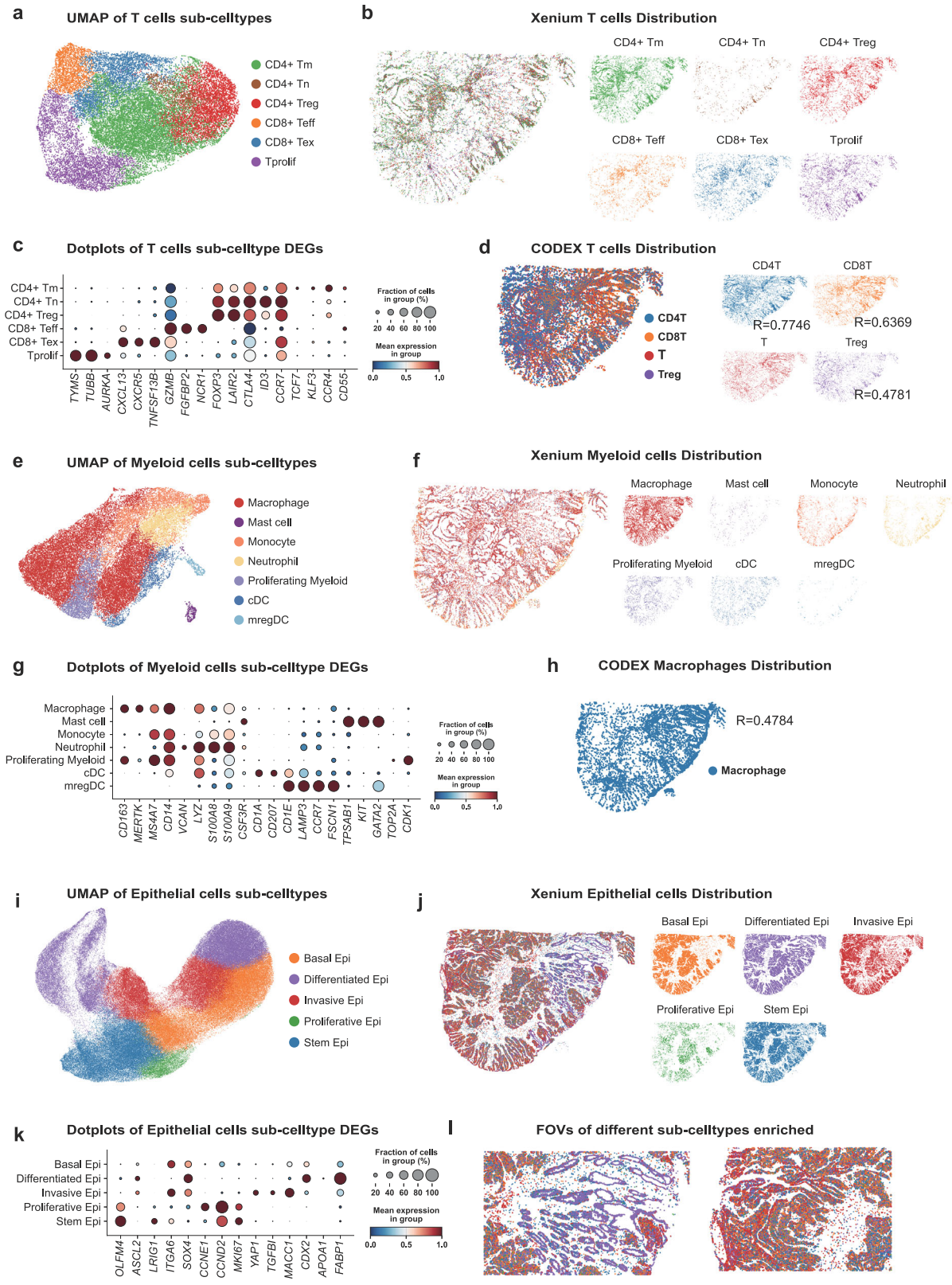


Fig. 4 | Resolution of spatial cellular heterogeneity and tumor microenvironment architecture. a Uniform manifold approximation and projection (UMAP) of cell populations derived from stPainter latent embeddings and the corresponding dot plot of lineage-defining markers. Dot size represents the fraction of cells expressing each gene, and dot color represents the mean expression level. **b** Comparative UMAP representations generated by different computational workflows using raw principal component analysis (PCA), stPainter-100, SpaIM, gimVI and Tangram. **c** Spatial coherence and lineage-level comparison with adjacent-section co-detection by indexing (CODEX) imaging. Heatmaps of cell-type proportions and Pearson's *r* assess regional concordance between stPainter-imputed cellular compositions and CODEX-derived protein-level tissue organization across spatial patches. Heatmap colors represent the corresponding proportions or correlation values according to the displayed scales. Colors in the UMAP and spatial maps denote the indicated cell types or computational methods. SMC denotes smooth muscle cell, cDC2 denotes conventional dendritic cell type 2, mregDC denotes mature regulatory dendritic cell, and pDC denotes plasmacytoid dendritic cell. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-026-76552-x



Finally, we evaluated the generalizability of stPainter-100 across an expanded spectrum of malignancies by extending our analyses to the remaining SPATCH datasets (OV and LIHC) and multiple independent public Xenium datasets, including PRAD, NSCLC, and CESC (Supplementary Figs. 8–11). In all cases, stPainter facilitated robust downstream workflows, achieving a level of resolution in major and fine-grained subpopulation identification comparable to that observed in COAD. These results support the robustness and broad applicability of stPainter for facilitating downstream analysis across diverse cancer histologies.

Discussion

High-resolution ST[28] is fundamentally constrained by stochastic RNA capture and restricted gene panels, necessitating the integration of

Fig. 5 | Multimodal characterization of fine-grained cell sublineages and their spatial niches. a–d T cell lineage analysis using UMAP visualization (**a**), Xenium-based spatial mapping (**b**), a dot plot of differentially expressed genes (DEGs) (**c**) and comparison with adjacent-section CODEX protein imaging (**d**). Tm, Tn, Treg, Teff, Tex and Tprolif denote memory, naive, regulatory, effector, exhausted and proliferating T cells, respectively. **e–h** Myeloid cell lineage analysis using UMAP visualization (**e**), Xenium-based spatial mapping (**f**), a dot plot of myeloid-cell DEGs (**g**) and comparison with adjacent-section CODEX protein imaging (**h**). cDC and mregDC denote conventional dendritic cells and mature regulatory dendritic cells, respectively. **i–l**, Epithelial cell lineage analysis using UMAP visualization (**i**), Xenium-based spatial mapping (**j**), a dot plot of epithelial-cell DEGs (**k**) and representative spatial enrichment within selected fields of view (FOVs) (**l**). Epi denotes epithelial cells. In the dot plots, dot size represents the fraction of cells expressing each gene and dot color represents the mean expression level. Colors in the UMAP and spatial maps denote the indicated cell sublineages or protein-defined cell types. Reported r values are Pearson correlation coefficients calculated across matched spatial regions. UMAP denotes uniform manifold approximation and projection, and CODEX denotes co-detection by indexing. Source data are provided as a Source Data file.

whole-transcriptome scRNA-seq to restore biological signals and enhance data fidelity. However, existing imputation frameworks[12,19] are hampered by a sample-specific supervised paradigm that requires tissue-matched references and repetitive retraining for every new experiment, a dependency that severely restricts their scalability and clinical applicability. Here, we developed stPainter, a conditional latent diffusion model that shifts the field toward a pretrained paradigm. By leveraging a massive pan-cancer atlas of 1.3 million cells across 21 diverse malignancies, stPainter shifts ST imputation from a sample-specific supervised setting toward a pretrained inference paradigm without the need for auxiliary matched data or retraining, providing a convenient and effective (Supplementary Fig. 7) tool for ST data analysis.

The technical efficacy of stPainter is grounded in its conditional latent diffusion architecture. By encoding cancer-type metadata as a conditional prior (c), the model constrains the generative dynamics to a biologically plausible subspace. In contrast to conventional approaches that operate directly in the high-dimensional gene expression space, or require tissue-matched references and dataset-specific retraining, stPainter performs SDE-guided generation within a highly compressed, denoised latent manifold (z)[26], iteratively mapping sparse spatial measurements back onto the learned distribution from a pretrained pan-cancer atlas (Supplementary Fig. 5). By compressing transcriptomic variability into a low-dimensional latent embedding (e.g., 50 or 100 dimensions), the model establishes a stable basis for unsupervised manifold learning and high-resolution cluster delineation[36]. In parallel, the generative decoder reconstructs a continuous expression matrix spanning the full pretraining gene space, enabling comprehensive downstream transcriptomic analyses. Although these individual components originate from the computer vision community, the core contribution of stPainter is the integration of these components into a reference-free pretraining-and-inference paradigm for single-cell resolution spatial transcriptomics, whose component-level necessity is validated by our ablation experiments (Supplementary Figs. 4 and 5).

Single-cell resolution ST provides gene expression profiles analogous to scRNA-seq, yet critical differences in data characteristics necessitate analytical approaches[28]. Unlike scRNA-seq, where enzymatic dissociation can disproportionately impact fragile cell types like epithelial cells[10,37,38], Xenium technologies often exhibit higher dropout rates for less abundant populations, such as immune cell genes (*CD8A*, *CD14*) (Supplementary Fig. 8). Furthermore, signal leakage from abundant neighboring cells can mask the transcriptomic signatures of rarer cell types, complicating their accurate identification[10]. stPainter is designed to address these fundamental data quality issues. By enhancing the raw measurements, it expands the assayed gene space to 10,000 genes and mitigates sparsity, generating profiles that are better suited for downstream clustering, marker evaluation, and spatial-domain analysis.

The high levels of noise and sparsity in Xenium data also severely compromise downstream computational tasks, particularly the clustering and identification of fine-grained cell subtypes. Conventional methods like PCA often fail to effectively separate biological signals from technical noise, resulting in poor cluster resolution[10]. While many imputation tools exist, they are often not optimized for clustering tasks and can fail to remove technical artifacts[22]. stPainter provides an alternative representation for clustering. We demonstrate that performing clustering directly on the latent representations learned by our model, rather than on raw or imputed gene expression, produced clearer separation of major cell populations and stronger concordance with marker-supported and reference-supported annotations. The resulting 50- to 100-dimensional latent space robustly captures major transcriptional variation relevant to lineage organization and downstream clustering. This enables the resolution of fine-grained subtypes with clusters that show strong concordance with known marker genes, and provides a useful representation for downstream exploratory analysis, particularly when matched scRNA-seq references are unavailable.

Although stPainter showed consistent performance across the evaluated datasets, several avenues remain for further optimization and exploration. The current framework addresses data sparsity based on patterns derived from scRNA-seq. Future iterations could explore more adaptive and learnable strategies for sparsity processing that dynamically account for sample-specific technical biases and varying capture efficiencies. Furthermore, while the latent representations produced by our model provide a representation for downstream clustering, the biological interpretability of these latent dimensions warrants deeper investigation. Uncovering the specific gene-program alignments within this manifold could provide additional insights into the regulatory logic of the tumor microenvironment. Moreover, transfer-label and CODEX analyses provided complementary evidence for the downstream interpretability of stPainter-enhanced data. Transfer labels were used as reference-supported annotations to assess agreement with established cell-type organization, while adjacent-section CODEX imaging offered orthogonal protein-level support for regional spatial concordance. We therefore interpret these results as evidence of marker- and domain-level biological consistency, rather than definitive same-cell or same-section ground-truth validation. Finally, the current pan-cancer atlas faces an inherent challenge of compositional imbalance, with rare malignancies being underrepresented compared to common cancer types. Integrating broader, more balanced multi-omic datasets may further refine stPainter's reference-free generalization and enhance its capacity to resolve the intricate biological granularity essential for downstream mechanistic interrogation.

Methods

Ethical statement

This study complies with all relevant ethical regulations. All datasets analyzed in this study were obtained from publicly available, de-identified sources, and no new experiments involving human participants or animals were conducted. Therefore, approval from an institutional review board or ethics committee was not required.

Problem formulation

Formally, we represent a single-cell resolution transcriptomic dataset as a tuple $\mathcal{T} = (\mathbf{X}, \mathbf{C})$, where $\mathbf{X} \in \mathbb{R}^{N \times G}$ denotes the gene expression matrix for N cells and G genes, and $\mathbf{C} \in \mathbb{R}^{N \times K}$ constitutes the one-hot

encoding for K distinct cancer types. Accordingly, the i-th cell is characterized by the pair $(\mathbf{x}_i, \mathbf{c}_i)$. We aim to capture the heterogeneous gene expression patterns intrinsic to different cancer types by pre-training our latent diffusion model stPainter on a comprehensive pan-cancer scRNA-seq atlas reference $\mathcal{T}_{sc} = (\mathbf{X}_{sc}, \mathbf{C}_{sc})$, which spans a 10,000 genes feature space G_{sc}. Conversely, we define a target single-cell resolution spatial transcriptomic dataset as $\mathcal{T}_{st} = (\mathbf{X}_{st}, \mathbf{C}_{st})$, typically restricted to a specific cancer type with a limited gene panel such that $G_{st} \ll G_{sc}$. Our primary objective is to leverage the learned generative priors to transform the raw $\mathcal{T}_{st}$ into an enhanced $\bar{\mathcal{T}}_{st} = (\bar{\mathbf{X}}_{st}, \mathbf{Z}_{st}, \mathbf{C}_{st})$. This output comprises the spatially informed expression matrix $\bar{\mathbf{X}}_{st} \in \mathbb{R}^{N_{st} \times G_{sc}}$ imputed to the full reference gene space, and a compact latent embedding $\mathbf{Z}_{st} \in \mathbb{R}^{N_{st} \times L}$ that encapsulates cellular manifolds.

stPainterarchitecture

The architecture of stPainter is composed of two synergistic components: a VAE for latent space embedding and a GiT for conditional generative modeling. The VAE is first trained to compress high-dimensional transcriptomic profiles into a compact, structured latent manifold that captures essential biological variance. Subsequently, the GiT functions as our core generative module by operating within this learned manifold. It is designed as a conditional diffusion model to learn the complex probability distribution of cellular states based on auxiliary information, such as cancer type. This decoupled, two-stage approach allows stPainter to efficiently handle broad gene-panel data while leveraging the power of Transformer models for precise, context-aware generation.

VAE architecture. Our VAE is inspired by the probabilistic framework of scVI[36] to learn a latent representation $\mathbf{z}$ while disentangling it from technical nuisance factors. It consists of an encoder-decoder architecture:

- The encoder network maps the raw gene expression profile $\mathbf{x}$ and its cancer type $\mathbf{c}$ to the parameters of the two factorized latent distributions via distinct output heads:

$$(\boldsymbol{\mu}_{\mathbf{z}}, \log \boldsymbol{\sigma}_{\mathbf{z}}^2) = \text{Encoder}_{\mathbf{z}}(\mathbf{x}, \mathbf{c}), \quad (\mu_l, \log \sigma_l^2) = \text{Encoder}_l(\mathbf{x}, \mathbf{c}). \tag{1}$$

  These parameters define the distributions for the biological latent vector $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}}, \text{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2))$ and the scalar library size factor $l \sim \text{LogNormal}(\mu_l, \sigma_l^2)$. The latent variables are then sampled using the reparameterization trick.

- The decoder network maps the latent state back to the parameters of the Zero-Inflated Negative Binomial (ZINB) distribution via two distinct output heads:

$$\boldsymbol{\rho} = \text{Decoder}_{\boldsymbol{\rho}}(\mathbf{z}, \mathbf{c}), \quad \boldsymbol{\pi} = \text{Decoder}_{\boldsymbol{\pi}}(\mathbf{z}, \mathbf{c}). \tag{2}$$

  These outputs define the normalized gene frequencies $\boldsymbol{\rho}$ and the dropout probabilities π. The reconstructed gene count $\bar{\mathbf{x}}$ is then drawn from the distribution ZINB($\boldsymbol{\mu}$, $\boldsymbol{\theta}$, π), where the mean is given by $\boldsymbol{\mu} = l \cdot \boldsymbol{\rho}$ and the dispersion $\boldsymbol{\theta}$ is a globally learnable parameter.

The hyperparameters of the VAE architecture are detailed in Supplementary Table 3.

GiT architecture. The GiT is a conditional diffusion model based on the Diffusion Transformer (DiT)[39]. We specifically adapted its architecture to process 1D latent transcriptomic embeddings, enabling it to model intricate dependencies within the VAE-learned cellular manifold:

- The input to the GiT is the noisy latent vector $\mathbf{z}(t)$. We implement a feature-wise tokenization strategy by treating the latent vector as a sequence of L tokens. Each one is projected into a hidden dimension D via a learnable linear embedding. Learnable 1D positional embeddings are added to preserve the topological semantics of the latent dimensions, yielding the initial sequence $\mathbf{H}_0 \in \mathbb{R}^{L \times D}$.
- The core network comprises a stack of Transformer blocks that iteratively denoise the latent representation. We employ an adaptive layer norm zero (adaLN-Zero) mechanism to rigorously condition the process on the cancer type $\mathbf{c}$ and timestep t. For each block, a shared MLP processes the fused condition vector to predict modulation parameters ($\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, α) that are applied to the Multi-Head Self-Attention (MSA) and Feed-Forward Network (FFN) modules:

$$\mathbf{H}'_k = \mathbf{H}_{k-1} + \alpha_1 \cdot \text{MSA}(\text{adaLN}(\mathbf{H}_{k-1}; \boldsymbol{\gamma}_1, \boldsymbol{\beta}_1)), \tag{3}$$

$$\mathbf{H}_k = \mathbf{H}'_k + \alpha_2 \cdot \text{FFN}(\text{adaLN}(\mathbf{H}'_k; \boldsymbol{\gamma}_2, \boldsymbol{\beta}_2)). \tag{4}$$

  The regression MLP is initialized to zero, rendering each block an identity function at the start of training to stabilize learning dynamics.
- After processing through n blocks, the output sequence $\mathbf{H}_n$ is projected back to the latent space dimension. The tokens undergo a final adaLN normalization, followed by a linear projection back to the scalar dimension. The sequence is then reshaped to yield the predicted noise required for the diffusion process, formally denoted as:

$$\boldsymbol{\epsilon} = \text{GiT}(\mathbf{z}(t), t, \mathbf{c}). \tag{5}$$

The hyperparameters of the GiT architecture are detailed in Supplementary Table 4.

Training and inference of stPainter

The application of stPainter involves two phases: a training phase to learn generative priors from the pan-cancer atlas, and an inference phase to enhance target ST data. The training is sequential: we first train the VAE to create a robust latent representation of cellular states. Then, holding the VAE fixed, we train the GiT to model the conditional distribution of these latent states. For inferencing, we employ a guided generation process based on score-based diffusion, which iteratively refines an initial noisy estimate of the target cell's latent state to be consistent with both the learned priors and the original measurements.

Generative Modeling with SDE. The generative process of stPainter is formulated through the lens of score-based generative modeling via SDEs[26]. This framework conceptualizes generation as a time-reversed diffusion process. A forward process, defined by a pre-specified SDE, systematically perturbs data samples $\mathbf{z}(1)$ from the complex data distribution p_1 towards a simple prior distribution p_0 (e.g., an isotropic Gaussian) at $t = 0$ over a continuous time interval $t \in [0, 1]$. A standard choice for this forward SDE is the Variance Preserving (VP) SDE:

$$\mathrm{d}\mathbf{z} = -\frac{1}{2}\beta(t)\mathbf{z}\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\mathbf{w}, \tag{6}$$

where $\beta(t)$ is a positive, monotonically increasing variance schedule defined on $t \in [0, 1]$, and $\mathbf{w}$ is a standard Wiener process.

The reverse process is defined as follows. The trajectory of this diffusion can be reversed by solving a corresponding reverse-time SDE[40]:

$$\mathrm{d}\mathbf{z} = \left[-\frac{1}{2}\beta(t)\mathbf{z} - \beta(t)\nabla_{\mathbf{z}} \log p_t(\mathbf{z})\right]\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\bar{\mathbf{w}}, \tag{7}$$

where $d\bar{\mathbf{w}}$ is a Wiener process running backwards in time, and $\nabla_{\mathbf{z}} \log p_t(\mathbf{z})$ is the score function of the perturbed data distribution p_t at time t. Simulating this SDE from a noise sample $\mathbf{z}(0) \sim p_0$ to $t = 1$ effectively generates a sample from the original data distribution p_1. The central challenge is thus to accurately estimate the time-dependent score function, which we accomplish by training our GiT.

Training of stPainter. The training of stPainter is conducted in two sequential stages, ensuring a stable latent manifold is learned before the generative priors are established.

- The VAE is first trained on the pan-cancer scRNA-seq atlas $\mathcal{T}_{sc}$ to learn a compressed and robust latent representation of single-cell transcriptomes. The objective is to maximize the evidence lower bound (ELBO) of the data log-likelihood, which balances reconstruction fidelity with regularization of the latent space:

$$\mathcal{L}_{\text{VAE}} = \mathbb{E}_{q(\mathbf{z}|\mathbf{x},\mathbf{c})}[\log p(\mathbf{x}|\mathbf{z},\mathbf{c})] - D_{\text{KL}}(q(\mathbf{z}|\mathbf{x},\mathbf{c})||p(\mathbf{z})), \tag{8}$$

  where the first term is the ZINB-based reconstruction loss and the second is the Kullback-Leibler divergence that regularizes the posterior q towards a standard normal prior $p(\mathbf{z})$. Upon convergence, the VAE encoder provides a high-quality mapping from gene expression space to the latent manifold.
- Following VAE convergence, its encoder is used to project the entire pan-cancer atlas into the latent space, creating a dataset of latent vectors $\mathbf{z}_{sc}$. The GiT, parameterized by $\boldsymbol{\theta}$ and denoted as $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}$, is then trained on this latent dataset. Its objective is to predict the initial noise $\boldsymbol{\epsilon}$ added to a latent vector $\mathbf{z}(0)$ to produce its noisy version $\mathbf{z}(t)$, conditioned on the cancer type $\mathbf{c}$ and timestep t. This is equivalent to learning the score function and is achieved by minimizing the following objective:

$$\mathcal{L}_{\text{GiT}} = \mathbb{E}_{t,\mathbf{z}(0),\boldsymbol{\epsilon}\sim\mathcal{N}(0,\mathbf{I})}\left[||\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{z}(t),t,\mathbf{c}) - \boldsymbol{\epsilon}||_2^2\right], \tag{9}$$

  where $\mathbf{z}(t)$ is sampled from the forward process perturbation kernel $p_t(\mathbf{z}(t)|\mathbf{z}(0))$ and t is sampled uniformly from [0, 1].

The training hyperparameters are detailed in Supplementary Table 5.

Inferencing with stPainter. During inference with stPainter, we employ a guided generative process based on Stochastic Differential Editing (SDEdit)[41]. This process ensures that the generated profiles are both biologically realistic and faithful to the original, partially measured data. The procedure unfolds in five steps:

- First, we align the spatial transcriptomics gene panel G_{st} with the pretraining gene space G_{sc}. A binary imputation mask $\mathbf{m} \in \{0,1\}^{G_{sc}}$ is constructed, where $m_g = 1$ if gene g is measured in the ST panel and $m_g = 0$ otherwise. The raw ST expression vector $\mathbf{x}_{st}$ is zero-padded to the full pretraining dimensionality, and the masked input $\mathbf{x}_{st} \odot \mathbf{m}$ is used as the starting point for inference.
- Second, we obtain a latent representation $\mathbf{z}_{st}$ of the masked input $\mathbf{x}_{st} \odot \mathbf{m}$ by passing it through the frozen, pretrained VAE encoder. This maps the partially observed measurement onto the learned cellular manifold.
- Third, instead of initializing the generative process from pure noise (i.e., at $t = 1$), we add a controlled amount of noise to the initial latent state $\mathbf{z}_{st}$. This is achieved by simulating the forward SDE (Eq. (6)) starting from $\mathbf{z}(0) = \mathbf{z}_{st}$ for a specified duration, up to a time $t_0 \in (0, 1]$. This yields a perturbed latent vector $\mathbf{z}(t_0)$. The hyperparameter t_0 critically governs the trade-off between faithfulness to the original measured expression (lower t_0) and the realism of the imputed profile (higher t_0).
- Fourth, starting from the perturbed latent vector $\mathbf{z}(t_0)$, we then numerically solve the reverse-time SDE (Eq. (7)) from $t = t_0$ down to $t = 0$. This is typically performed using a predictor-corrector sampler or a simpler Euler-Maruyama scheme, where each step is guided by the score estimated by our trained GiT, $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}$. This step effectively projects the noisy, partial information onto the manifold of complete, realistic cellular states learned from the pan-cancer atlas.
- Fifth, the denoised latent vector $\mathbf{z}(0)$ represents the imputed cellular state that optimally balances prior knowledge and measured data. It is then passed through the frozen VAE decoder to deterministically generate the full-dimensional expression profile $\bar{\mathbf{x}}_{st}$. Finally, a sparsity-aware postprocessing step is applied: for each gene, values below the S-th percentile are set to zero, where S is the gene-specific zero-value ratio estimated from the pretraining scRNA-seq atlas. This step enforces biologically realistic sparsity patterns in the imputed profiles. The effect of this postprocessing is validated by the `wo_sparsity` ablation in Supplementary Fig. 4a, b, which demonstrates consistent performance degradation when this step is removed.

The sampling hyperparameters are detailed in Supplementary Table 6.

Pan-cancer scRNA-seq atlas processing

We aggregated publicly available scRNA-seq datasets encompassing 21 cancer types, including both primary and metastatic tumors. To ensure broad genomic representation across cancer studies, only genes present in all datasets were retained for integration[42]. Cells were excluded if they exhibited any of the following: mitochondrial gene fraction > 20%, ribosomal gene fraction > 40%, fewer than 300 detected genes, or total UMI counts < 500. Highly variable genes were identified using the *Scanpy* implementation of the normalized dispersion method; the top 10,000 most variable genes were selected as targets for downstream imputation. To evaluate the biological fidelity of the integrated atlas, we applied scVI for batch-corrected integration after quality control[36]. The resulting embedding resolved canonical cell subtypes, confirming the atlas's suitability for pan-cancer analyses. For the tested multiple myeloma dataset, we used an external scRNA-seq reference recommended in the original study and performed label transfer with TACCO to support subsequent downstream analyses[43,44].

Downstream analysis of ST datasets

To facilitate robust downstream analysis, we first performed quality control by excluding the bottom 15% of cells based on total UMI counts. Gene expression imputation was then carried out using stPainter. Subsequent analyses followed the standard *Scanpy* workflow: after log-normalization, low-dimensional embeddings were derived using both stPainter's latent representation and PCA. Neighborhood graphs were constructed, followed by UMAP dimensionality reduction and Leiden clustering. Given that cell-type annotations were initially assigned using a transfer-annotation approach, we assessed the concordance between these transfer-annotation labels and the Leiden clusters. To refine annotation resolution, we performed sub-clustering and re-annotation of major cell populations and validated their identities by comparing marker gene expression profiles with those from reference single-cell transcriptomic datasets of the same tissue type. Major cell populations were annotated based on the expression of canonical marker genes: *CD8A* (T cells), *CD19* (B cells), *MZB1* (plasma cells), *CD68* (myeloid cells), *FCGR3B* (neutrophils), *PLVAP* (endothelial cells), *COL1A1* (fibroblasts), and *PDGFRB* (SMCs).

To evaluate the impact of stPainter imputation on the spatial architecture of the TME, we conducted a spatial coherence analysis

based on cell-type annotations. Specifically, we partitioned each tissue section into 100 μm × 100 μm patches and quantified the cellular composition within each patch. Using spatially resolved CODEX proteomics data from adjacent sections as orthogonal protein-level support, we computed patch-wise cell-type proportions and assessed regional concordance using Pearson correlation. This analysis demonstrates that stPainter preserves biologically plausible spatial organization of the TME.

Major lineages, including T cells, epithelial cells, and myeloid cells, were further subclustered and manually annotated via cluster-specific DEGs. The resulting cell type assignments were cross-validated with adjacent CODEX sections by calculating Pearson's r of cell counts within matched 100 μm × 100 μm regions.

Evaluation metrics

To evaluate model performance, we employed a suite of metrics categorized into gene-level and cluster-level.

Gene Metrics. These metrics quantify the similarity between predicted gene expression ($\hat{\mathbf{X}}$) and ground-truth observations ($\mathbf{X}$) across n cells.

- **Pearson correlation coefficient (PCC):** PCC measures the linear relationship between the predicted and observed expression profiles for each gene:

$$\mathrm{PCC}(\mathbf{x}_i, \hat{\mathbf{x}}_i) = \frac{\sum_{j=1}^{n}(x_{ij} - \mu_i)(\hat{x}_{ij} - \hat{\mu}_i)}{\sigma_i \hat{\sigma}_i}, \tag{10}$$

  where μ and σ denote the mean and standard deviation of expression values, respectively.

- **Structural similarity index measure (SSIM):** SSIM assesses the preservation of spatial patterns by integrating luminance, contrast, and structural information:

$$\mathrm{SSIM}(\mathbf{x}_i, \hat{\mathbf{x}}_i) = \frac{(2\mu_i\hat{\mu}_i + C_1)(2\mathrm{cov}(\mathbf{x}_i, \hat{\mathbf{x}}_i) + C_2)}{(\mu_i^2 + \hat{\mu}_i^2 + C_1)(\sigma_i^2 + \hat{\sigma}_i^2 + C_2)}, \tag{11}$$

  where $\mathrm{cov}(\mathbf{x}_i, \hat{\mathbf{x}}_i)$ represents the covariance between $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$, and C_1, C_2 are stabilization constants.

- **Root mean square error (RMSE):** RMSE quantifies the average prediction error magnitude. To account for cross-platform technical variation, it is calculated on z-score normalized expression values:

$$\mathrm{RMSE}(\mathbf{x}_i, \hat{\mathbf{x}}_i) = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(\hat{z}_{ij} - z_{ij})^2}, \tag{12}$$

  where z_{ij} and $\hat{z}_{ij}$ are normalized intensities. A lower RMSE indicates superior imputation accuracy.

- **Jensen-Shannon divergence (JS):** JS evaluates the similarity of spatial distributions, computed here as the Jensen-Shannon divergence between normalized spatial probability distributions $\mathbf{P}_i$ and $\hat{\mathbf{P}}_i$:

$$\mathrm{JS}(\mathbf{x}_i, \hat{\mathbf{x}}_i) = \frac{1}{2}\left[\mathrm{KL}(\mathbf{P}_i \parallel \bar{\mathbf{P}}_i) + \mathrm{KL}(\hat{\mathbf{P}}_i \parallel \bar{\mathbf{P}}_i)\right], \tag{13}$$

  where $\bar{\mathbf{P}}_i = (\mathbf{P}_i + \hat{\mathbf{P}}_i)/2$ and KL denotes the Kullback-Leibler divergence.

Cluster Metrics. These metrics assess the model's capacity to preserve the spatial organization of cell types and tissue domains.

- **Adjusted Rand Index (ARI):** ARI evaluates the similarity between the clustering of imputed data and ground-truth annotations, corrected for chance:

$$\mathrm{ARI} = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]/\binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}\right]/\binom{n}{2}}, \tag{14}$$

  where n_{ij} is the number of shared elements between clusters, and a_i, b_j are the marginal sums of the contingency table.

- **Adjusted Mutual Information (AMI):** AMI quantifies the information shared between two partitions, adjusted for the expected mutual information due to random chance:

$$\mathrm{AMI} = \frac{\mathrm{MI}(\mathcal{A}, \mathcal{B}) - \mathbb{E}[\mathrm{MI}(\mathcal{A}, \mathcal{B})]}{\mathrm{avg}(H(\mathcal{A}), H(\mathcal{B})) - \mathbb{E}[\mathrm{MI}(\mathcal{A}, \mathcal{B})]}, \tag{15}$$

  where MI is the mutual information, H is the entropy, and $\mathbb{E}$ represents the expectation.

- **Normalized Mutual Information (NMI):** NMI scales the mutual information score to the range [0, 1] based on the entropy of the partitions:

$$\mathrm{NMI}(\mathcal{A}, \mathcal{B}) = \frac{\mathrm{MI}(\mathcal{A}, \mathcal{B})}{\sqrt{H(\mathcal{A})H(\mathcal{B})}}, \tag{16}$$

- **Homogeneity (Homo):** Homogeneity measures the degree to which each cluster contains only members of a single class:

$$\mathrm{Homo} = 1 - \frac{H(\mathcal{B}|\mathcal{A})}{H(\mathcal{B})}, \tag{17}$$

  where $H(\mathcal{B}|\mathcal{A})$ is the conditional entropy of ground-truth categories $\mathcal{B}$ given predicted assignments $\mathcal{A}$. A value of 1.0 indicates perfect homogeneity.

Statistics and reproducibility

No statistical method was used to predetermine sample size. Sample sizes were determined by the availability of publicly accessible datasets that met the predefined quality-control criteria. All dataset inclusion, exclusion and quality-control procedures are described in the Methods. Randomization and blinding were not applicable because this was a computational study involving no experimental allocation or subjective outcome assessment. Pearson correlation analyses in Supplementary Fig. 17 were performed across six independent cancer datasets ($n = 6$) using two-sided tests. Pearson's r and exact P-values are reported, and no correction for multiple comparisons was applied. The datasets, preprocessing procedures, model settings and evaluation metrics required to reproduce the analyses are described in the Methods and Supplementary Information.

Software and computational environment

Data processing, model training, spatial gene-expression imputation and downstream analyses were performed using stPainter (https://github.com/yyh030806/stPainter). Software packages used in the analyses included PyTorch 2.9.1, torchvision 0.24.1, torchaudio 2.9.1, torchdiffeq 0.2.5, Lightning 2.5.3, JAX 0.6.2, jaxlib 0.6.2, Flax 0.10.7, Optax 0.2.5, Scanpy 1.11.4, AnnData 0.11.4, Squidpy 1.6.5, scvi-tools 1.3.3, tangram-sc 1.0.4, NumPy 2.2.6, pandas 2.3.2, SciPy 1.15.3, scikit-learn 1.7.1, Matplotlib 3.10.5 and seaborn 0.13.2. The complete software environment and exact dependency versions are specified in the `requirements_lock.txt` file included in the stPainter v1.0.0 release.

The release associated with this study has been archived in Zenodo and is available at https://doi.org/10.5281/zenodo.21502447.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The pan-cancer scRNA-seq datasets used for pretraining stPainter were obtained from publicly available studies deposited in the GEO database. Detailed accession numbers are provided in Supplementary Table 1.

Spatial transcriptomics datasets analyzed in this study include the SPATCH dataset (COAD, OV, and LIHC) as well as additional publicly available Xenium datasets (PRAD, NSCLC, and CESC). The SPATCH dataset is publicly available at https://spatch.pku-genomics.org[27]. Xenium 5 k spatial transcriptomics datasets were obtained from the 10x Genomics public data repository, including datasets available at: https://www.10xgenomics.com/datasets/xenium-prime-ffpe-human-prostate, https://www.10xgenomics.com/datasets/xenium-human-lung-cancer-post-xenium-technote, and https://www.10xgenomics.com/datasets/xenium-prime-ffpe-human-cervical-cancer. The MM Xenium dataset is available under accession number GSE299207[43]. For cross-platform evaluation, the MERFISH V2 dataset was obtained from the official Vizgen MERFISH 2.0 resource, and the GBM CosMx 6 K dataset was obtained from Zenodo record 17792860[45].

All data generated during this study are publicly available via Google Drive at https://drive.google.com/drive/folders/1ROePTWXUtuZWIrbLF6HLf34cRaLjGTCw. Source data are provided in this paper.

Code availability

The source code for stPainter is publicly available at https://github.com/yyh030806/stPainter under the MIT License. The version used in this study (v1.0.0) has been archived in Zenodo and is available at https://doi.org/10.5281/zenodo.21502447[46].

References

1. Gulati, G. S., D'Silva, J. P., Liu, Y., Wang, L. & Newman, A. M. Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics. *Nat. Rev. Mol. Cell Biol.* **26**, 11–31 (2025).
2. Liu, Y., Dai, Y. & Wang, L. Spatial omics at the forefront: emerging technologies, analytical innovations, and clinical applications. *Cancer Cell* **44**, 24–49 (2025).
3. He, S. et al. High-plex imaging of rna and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol.* **40**, 1794–1806 (2022).
4. Janesick, A. et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat. Commun.* **14**, 8353 (2023).
5. Tian, L., Chen, F. & Macosko, E. Z. Moving genomics into tissues. *Nat. Biotechnol.* **41**, 773 (2022).
6. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
7. Asp, M. et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* **179**, 1647–1660 (2019).
8. Moncada, R. et al. Integrating microarray-based spatial transcriptomics and single-cell rna-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**, 333–342 (2020).
9. Kwok, A. W. C. et al. Denoising image-based spatial transcriptomics data with denoist. Preprint at https://doi.org/10.1101/2025.11.13.688387 (2025).
10. Bilous, M. et al. Resolving sensitivity, specificity and signal contamination in Xenium spatial transcriptomics. *Nat. Methods* **23**, 1152–1162 (2026).
11. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
12. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nat. Methods* **18**, 1352–1362 (2021).
13. Abdelaal, T., Mourragui, S., Mahfouz, A. & Reinders, M. J. Spage: spatial gene enhancement using scrna-seq. *Nucleic Acids Res.* **48**, e107–e107 (2020).
14. Shengquan, C., Boheng, Z., Xiaoyang, C., Xuegong, Z. & Rui, J. stplus: a reference-based method for the accurate enhancement of spatial transcriptomics. *Bioinformatics* **37**, i299–i307 (2021).
15. Moriel, N. et al. Novosparc: flexible spatial reconstruction of single-cell gene expression with optimal transport. *Nat. Protoc.* **16**, 4177–4200 (2021).
16. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* **11**, 2084 (2020).
17. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887 (2019).
18. Li, R. et al. Accurate imputation of pathway-specific gene expression in spatial transcriptomics with pasta. *Nat. Commun.* **17**, 726 (2025).
19. Lopez, R. et al. A joint model of unpaired data from scrna-seq and spatial transcriptomics for imputing missing gene expression measurements. Preprint at https://doi.org/10.48550/arXiv.1905.02269 (2019).
20. Wan, X. et al. Integrating spatial and single-cell transcriptomics data using deep generative models with spatialscope. *Nat. Commun.* **14**, 7848 (2023).
21. Sun, E. D., Ma, R. & Zou, J. Sprite: improving spatial gene expression imputation with gene and cell networks. *Bioinformatics* **40**, i521–i528 (2024).
22. Li, K., Li, J., Tao, Y. & Wang, F. stdiff: a diffusion model for imputing spatial transcriptomics through single-cell transcriptomics. *Brief. Bioinform.* **25**, https://doi.org/10.1093/bib/bbae171 (2024).
23. Li, B. et al. Spaim: Single-cell spatial transcriptomics imputation via style transfer. *Nat Commun.***16**, https://doi.org/10.1038/s41467-025-63185-9 (2025).
24. Liu, T., Lin, Y., Luo, X., Sun, Y. & Zhao, H. Vista uncovers missing gene expression and spatial-induced information for spatial transcriptomic data analysis. *Commun. Biol.* **9**, 203 (2026).
25. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022).
26. Song, Y. et al. Score-based generative modeling through stochastic differential equations. In *Proceedings of the International Conference on Learning Representations*. https://mlanthology.org/iclr/2021/song2021iclr-scorebased/ (2021).
27. Ren, P. et al. Systematic benchmarking of high-throughput subcellular spatial transcriptomics platforms across human tumors. *Nat. Commun.* **16**, 9232 (2025).
28. Marco Salas, S. et al. Optimizing xenium in situ data utility by quality assessment and best-practice analysis workflows. *Nat. Methods* **22**, 813–823 (2025).
29. Chang, J. et al. Single-cell multi-stage spatial evolutional map of esophageal carcinogenesis. *Cancer Cell* **43**, 380–397 (2025).
30. Zheng, L. et al. Pan-cancer single-cell landscape of tumor-infiltrating t cells. *Science* **374**, abe6474 (2021).
31. You, S. et al. Lymphatic-localized treg-mregdc crosstalk limits antigen trafficking and restrains anti-tumor immunity. *Cancer Cell* **42**, 1415–1433 (2024).

32. Loi, P. et al. Epigenetic and oncogenic inhibitors cooperatively drive differentiation and kill kras-mutant colorectal cancers. *Cancer Discov.* **14**, 2430–2449 (2024).
33. Brooks, D. L. P. et al. Itga6 is directly regulated by hypoxia-inducible factors and enriches for cancer stem cell activity and invasion in metastatic breast cancer models. *Mol. Cancer* **15**, 26 (2016).
34. Van der Flier, L. G., Haegebarth, A., Stange, D. E., Van de Wetering, M. & Clevers, H. Olfm4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. *Gastroenterology* **137**, 15–17 (2009).
35. Marteau, V. et al. Single-cell integration and multi-modal profiling reveals phenotypes and spatial organization of neutrophils in colorectal cancer. *Cancer Cell* **44**, 146–165 (2026).
36. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
37. Denisenko, E. et al. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus rna-seq workflows. *Genome Biol.* **21**, 130 (2020).
38. van den Brink, S. C. et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017).
39. Peebles, W. & Xie, S. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2023).
40. Anderson, B. D. Reverse-time diffusion equation models. *Stoch. Process. Appl.* **12**, 313–326 (1982).
41. Meng, C. et al. Sdedit: Guided image synthesis and editing with stochastic differential equations. In *Proceedings of the International Conference on Learning Representations*. (2022).
42. Wolf, F. A., Angerer, P. & Theis, F. J. Scanpy: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
43. Yip, R. K. H. et al. Profiling the spatial architecture of multiple myeloma in human bone marrow trephine biopsy specimens with spatial transcriptomics. Blood **146**, 1837–1849 (2025).
44. Mages, S. et al. Tacco unifies annotation transfer and decomposition of cell identities for single-cell and spatial omics. *Nat. Biotechnol.* **41**, 1465–1473 (2023).
45. Migliozzi, S. et al. Restraint of cancer cell plasticity by spatial homotypic clustering. *Cancer cell* **43**, 2206–2223 (2025).
46. Yang, Y. yyh030806/stPainter: stPainter v1.0.0 – nature communications release. *Zenodo* https://doi.org/10.5281/zenodo.21502447 (2026).

Acknowledgements

The authors thank all members of the State Key Laboratory of Cognitive Intelligence at the University of Science and Technology of China and the Department of Gastrointestinal Oncology at Peking University Cancer Hospital and Institute for their helpful discussions and support during this work.

Author contributions

Conceptualization: Y.H.Y. and Y.M.L.; Methodology: Y.H.Y. and Y.M.L.; Formal Analysis: Y.H.Y. and Y.M.L.; Software: Y.H.Y. and B.M.; Investigation: Y.H.Y. and Y.M.L.; Data Curation: Y.M.L.; Validation: Y.H.Y. and Y.M.L.; Visualization: Y.H.Y. and Y.M.L.; Writing - Original Draft: Y.H.Y. and Y.M.L.; Writing - Review & Editing: Y.H.Y., Y.M.L., K.Z., Z.X.Z., H.X.P., C.L.C., Q.L., and Y.C.; Supervision: K.Z.; Project Administration: K.Z.; Resources: K.Z., Y.C., L.S., and E.H.C.; Funding Acquisition: K.Z., Y.C., L.S., and E.H.C. All authors reviewed the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (62406303, 62525606, 92459302, and U22A20327), the Non-communicable Chronic Diseases-National Science and Technology Major Project (2024ZD0534800), the Beijing Nova Program (202604841241), the Beijing Natural Science Foundation (L2609074), the Clinical Medicine Plus X-Young Scholars Project of Peking University (PKU2026LCXQ021), the Capital's Funds for Health Improvement and Research (2026-1Q-1093), and the Science Foundation of Peking University Cancer Hospital (BJCH2025GG03).

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/s41467-026-76552-x.

Correspondence and requests for materials should be addressed to Kai Zhang, Yang Chen, Lin Shen or Enhong Chen.

Peer review information *Nature Communications* thanks Zeyu Fu and Baijian Yang for their contribution to the peer review of this work. A peer review file is available.