SpatioCell: A Deep Learning Algorithm for High-resolution Single-cell Mapping through Deep Integration of Histology Image and Sequencing Data

Naiqiao Hou[#1,2], Zhaorun Wu[#5], Yue Yu[#1,2], Yanqi Zhang[1,6,7], Jiayun Wu[1,2], Zhixing Zhong[5], Fangfang Xu[5], Zeyu Wang[2], Chaoyong Yang*[2,5], Weihong Tan*[2,3,4], Jia Song*[1,2]

[1]Institute for Developmental and Regenerative Cardiovascular Medicine, MOE-Shanghai Key Laboratory of Children's Environmental Health, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China

[2]Institute of Molecular Medicine, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China

[3]Molecular Science and Biomedicine Laboratory (MBL), State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, College of Biology, Aptamer Engineering Center of Hunan Province, Hunan University, Changsha, Hunan 410082, P. R. China

[4]Zhejiang Cancer Hospital, Key Laboratory of Zhejiang Province for Aptamers and Theranostics, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences，Hangzhou, Zhejiang 310022, P. R. China

[5]Institute of Artificial Intelligence, Department of Chemical Biology, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, 361102, China.

[6]School of Life Sciences, Inner Mongolia University, Hohhot, 010000, China

[7]School of Life Sciences, Fudan University, Shanghai, 200438, China

Corresponding authors: cyyang@xmu.edu.cn (Chaoyong Yang), tan@hnu.edu.cn (Weihong Tan), songjiajia2010@shsmu.edu.cn (Jia Song)

These authors contributed equally: Naiqiao Hou, Zhaorun Wu, Yue Yu

Abstract

Spatial transcriptomics (ST) enables gene expression analysis within spatial context of tissues. Since most data are limited to multicellular resolution, current computational methods can only estimate cell proportions. To address this gap, we introduce SpatioCell, a framework that integrates imaging data-derived morphological features with transcriptomic measurements for precise single-cell annotation. SpatioCell combines automated adaptive prompting with an optimized vision model for accurate cell segmentation and morphological analysis. Dynamic programming then optimizes cell-type assignments by integrating H&E-derived morphology and transcriptomic constraints, improving resolution within and beyond ST spots. Extensive benchmarking on histopathological, simulated, and real ST data from five distinct cancer types shows that SpatioCell outperforms state-of-the-art methods in both nuclear segmentation and cell annotation accuracy. Notably, SpatioCell reveals tumor microenvironment details, such as tumor boundaries, blood vessel structures, and immune infiltration, that were missed by other methods, while also showing strong potential in correcting deconvolution errors to further increase annotation accuracy. By redefining single-cell spatial mapping with unprecedented accuracy and resolution, SpatioCell enables precise analysis of tissue heterogeneity, offering new insights into cancer and tissue biology.

Introduction

Deciphering the cellular composition of tissues is crucial for understanding how spatial organization dictates cell interactions, functional specialization, and microenvironmental dynamics. These structural complexities shape physiological processes and influence disease progression, immune responses, and therapeutic outcomes[1, 2]. Accurate spatial mapping at the cellular level is therefore essential for advancing mechanistic research and precision medicine.

Spatial transcriptomics (ST) has emerged as a powerful tool for profiling gene expression within the spatial context of tissues, enabling the characterization of cell compositions and molecular landscapes in situ[3, 4]. Existing ST technologies can be broadly categorized into image-based and sequencing-based approaches[5]. Image-based methods, such as MERFISH[6] and seqFISH+[7], achieve subcellular resolution, but they are limited to predefined gene sets. In contrast, sequencing-based methods, including 10x Visium[8], Decoder-seq[9], and Stereo-seq[10], provide whole-transcriptome coverage with higher throughput and simpler protocols, making them more widely adopted. However, their resolution remains constrained, with platforms like Visium featuring spot sizes around 55 μm, often capturing multiple cells within each spatial unit (hereinafter termed as spot). As a result, most available ST datasets operate at multicellular resolution, restricting the ability to resolve fine-scale cellular organization and interactions[11].

To address these resolution constraints, computational methods have been developed. Deconvolution algorithms leverage single-cell RNA sequencing (scRNA-seq) references to infer cellular composition within each spatial spot[12, 13]. However, these approaches only estimate cell-type proportions rather than achieving genuine single-cell resolution, consequently limiting precise spatial mapping and impeding the comprehensive characterization of microenvironmental heterogeneity.

Notably, ST data is often accompanied by histology images, mostly through hematoxylin and eosin (H&E) staining, which provide additional spatial context that could boost resolution. Histology image analysis, a key focus in computational pathology, has led to the development of several cell/nuclear segmentation algorithms[14, 15]. When integrated with ST data, these methods can achieve analysis at single-cell resolution. However, current studies (such as STdGCN[16], SpatialScope[17] and Spotiphy[18]) are largely limited to estimating cell counts within each spot using nuclear segmentation, because few models can simultaneously perform both nuclear

segmentation and cellular morphological analysis on H&E-stained images with sufficient accuracy. Moreover, no effective methods are available to integrate image features with sequencing information.

Recently, a large number of models have been developed in the field of image segmentation. These include, for example, vision foundation models (VFMs) that comprise pretrained AI models for computer vision tasks. The development of VFMs like SAM (Segment Anything Model) has significantly improved segmentation accuracy[19, 20]. SAM was trained on a dataset called SA-1B, which contains over 1 billion segmentation masks across 11 million images. It is designed to segment any object or region with high precision and adaptability, making it a promising tool for scalable, high-precision cell/nuclear segmentation in complex tissue architectures. However, challenges remain in adapting these models to the specific needs of ST. SAM's reliance on user-provided prompts—such as clicks, bounding boxes, or polygons—becomes a limitation when managing large-scale, densely packed objects like cells in histological sections wherein numerous similar, or overlapping, targets require more efficient and automated segmentation strategies. Furthermore, cell morphological analysis in ST images demands careful consideration of tissue-specific patterns and cellular contexts.

To address these challenges, we developed SpatioCell, an integrative algorithm designed to construct high-resolution single-cell spatial maps. SpatioCell adapts SegFormer[21], a hierarchical transformer architecture, to generate prompts for SAM, enabling precise nuclei segmentation and analysis in complex tissues. It further models the integration of histology-derived morphological features and deconvolution-based predictions as a knapsack problem, employing a dynamic programming algorithm to achieve precise single-cell annotation. In extensive evaluations on six public datasets, including both H&E-stained images both from Formalin-Fixed Paraffin-Embedded (FFPE) tissues and fresh frozen (FF) tissues, and DAPI-stained images, SpatioCell outperformed state-of-the-art methods in nuclear segmentation and classification. When applied to tumor samples with multicellular-resolution STs (55–100μm), SpatioCell successfully reconstructed single-cell-level annotations, revealing intricate tumor microenvironment (TME) features such as vasculature, ducts, tumor boundaries, and immune infiltration sites—details obscured by other methods. Additionally, SpatioCell optimizes the integration of imaging and sequencing data, providing a mechanism for correcting deconvolution errors and enabling more accurate cell-type

proportion estimation and annotation. This framework overcomes the resolution limitations of STs, offering a powerful tool for precise tissue microenvironmental characterization and advancing single-cell spatial analysis.

Results

Overview of SpatioCell

To achieve single-cell resolution cell annotation in multicellular-resolution STs, SpatioCell comprises two key modules: (1) a deep learning module for histopathological image processing that performs nuclear segmentation and analysis through detailed morphological feature extraction, and (2) a dynamic programming module for cell annotation that incorporates both histopathological image-derived features and transcriptomic data to achieve high precision and resolution in cell annotation (**Fig. 1A**).

Existing algorithms extract cell counts from images within each spot, but they overlook the rich morphological information essential for accurate cell classification. To address this, we developed a novel nuclear segmentation and analysis model that combines a multi-scale feature encoding automatic prompt generator (Prompter) with a fine-tuned SAM model adapted for H&E images. This approach leverages H&E images for precise cell localization and counting, while incorporating cell morphology to generate classification probability profiles.

SAM is a foundational model for image segmentation, pretrained on large-scale image datasets[19]. Although SAM is capable of zero-shot learning, its performance on H&E-stained images remains suboptimal owing to indistinct nuclear boundaries and complex tissue backgrounds[22]. Moreover, SAM's original automatic mask generation (AMG) method relies on grid-based prompts, which, however, fail to achieve optimal segmentation accuracy for densely clustered and morphologically complex nuclei. Additionally, the large number of targets in stained images makes manual prompting impractical. To overcome these limitations, we introduced three key modifications to update SAM's adaptability to H&E images (**Fig. 1B-C**). First, we fine-tuned SAM for H&E-stained images using prompts from ground truth nuclear masks, improving its accuracy in segmenting individual nuclei. Second, we designed an automatic Prompter. The Prompter consists of a feature encoder and two decoders. The feature encoder, derived from SegFormer's multi-scale encoding backbone, enables it to capture both coarse high-resolution features and fine-grained detailed features[21]. The decoders

replace SegFormer's original multi-layer perceptron (MLP) with convolutional layers, improving the ability to extract structured spatial information from histopathological images. This enables SAM to receive accurate positive prompts and carry out automated cell segmentation. Finally, to effectively utilize the cell morphological information provided by the images, the Prompter also includes an additional classification head designed for multi-class semantic segmentation. By integrating the fine-tuned SAM with this Prompter, we developed a histopathology-specific model that combines SegFormer's multi-scale semantic recognition with SAM's precise boundary delineation, providing a robust and scalable solution for accurate nuclear segmentation and detailed morphological analysis, as detailed in **Methods**. Ultimately, this SpatioCell module generates two primary outputs: (1) cell locations and masks, and (2) cell identity (supertype) probability profiles derived from morphological cues.

While H&E images reveal cell localization and potential types, their resolution is insufficient for distinguishing fine-grained subtypes, and these images often contain noise, necessitating integration with sequencing data. ST is typically processed using deconvolution algorithms to estimate the high-resolution composition of cell types, including subtypes, at each spot. SpatioCell's annotation module employs dynamic programming to integrate these data. It also provides an interface for seamless incorporation of any deconvolution algorithm to determine cell composition ratios per spot. The cell identity assignment can be modeled as a 'knapsack' problem where the constraint is deconvolution-inferred cell type compositions for which dynamic programming optimally assigns identities by maximizing total probability based on H&E-derived profiles. This approach leverages morphological insights for high-resolution single-cell annotation in spots, while minimizing errors through deconvolution constraints, thereby augmenting spatial transcriptomic resolution to the single-cell level, as detailed in **Methods**.

Above all, SpatioCell introduces a novel approach for integrating histological and transcriptomic data, while it improves spatial transcriptomics resolution to the single-cell level. It refines nuclear segmentation with an adapted SAM model and optimally assigns cell identities using dynamic programming, ensuring precise and robust annotation.

SpatioCell Achieves Major Improvement in Cell Segmentation and Analysis on Histopathology Images

To assess the effectiveness of SpatioCell's H&E image processing module, we first

evaluated its performance on nuclear segmentation. We conducted benchmark comparisons between SpatioCell and one classic instance segmentation model, Mask R-CNN[23], as well as four widely used cell/nuclear segmentation algorithms, including HoverNet[24], Cellpose3[25], StarDist[26], and CellProfiler[27]. These models were tested on five public H&E-stained histopathology datasets, as detailed in Methods. To comprehensively evaluate segmentation performance, we calculated Aggregated Jaccard Index (AJI), Panoptic Quality (PQ), along with its two components: Detection Quality (DQ) and Segmentation Quality (SQ) (see **Methods** for details). AJI measures the overall instance segmentation accuracy, penalizing over- and under-segmentation[28]. PQ is a unified metric for semantic and instance segmentation[29], with DQ assessing detection quality and SQ evaluating consistency with the ground truth.

SpatioCell achieves the best performance across most evaluation metrics in every test dataset (**Fig. 2A-B**). The CPM17 dataset[30] is primarily composed of glioblastoma (GBM), head and neck squamous cell carcinoma (HNSCC), and non-small cell lung cancer (NSCLC) samples. SpatioCell achieves a significantly higher DQ than other methods, indicating fewer false-positive segmentations (**Fig. 2A**, **Supplementary Figure 1A-B**). The Kumar dataset[28], characterized by a higher density of overlapping and crowded nuclei, further demonstrates SpatioCell's robustness and superior performance in nuclear segmentation (**Fig. 2A**, **Supplementary Figure 1C-D**). The PanNuke dataset is one of the most challenging public nuclear segmentation benchmarks. It contains over 200,000 annotated nuclei across 19 tissue types[31]. In three-fold cross-validation, SpatioCell consistently outperforms all other models in both nuclear detection and segmentation metrics (**Fig. 2B, Supplementary Figure 2A-C**). Furthermore, SpatioCell both successfully detects nuclei with weaker staining and accurately excludes unstained mature red blood cells (**Fig. 2C**), demonstrating its ability to remove background artifacts and accurately segment true nuclei. In comparison, Cellpose, Stardist, and CellProfiler exhibit significant false-positive detections in dense blood cell regions. While Hover-Net and Mask R-CNN also face challenges in precise nuclear boundary delineation, highlighting the advantage of SpatioCell's segmentation module in capturing nuclear morphology with higher precision.

The H&E image processing module of SpatioCell further performs broad nuclear type classification based on morphological features, providing essential input for subsequent modular cell annotation. To evaluate the effectiveness of this module, we

compared SpatioCell with Hover-Net and Mask R-CNN, two widely used semantic segmentation models, across three benchmark datasets: PanNuke, MoNuSAC, and CoNSeP. For classification performance evaluation, we employed mean Panoptic Quality (mPQ) across different cell types and overall accuracy across all types, as detailed in **Methods**. These metrics collectively assess both nuclear detection accuracy and classification performance. SpatioCell outperformed Hover-Net and markedly exceeded Mask R-CNN across most metrics on three datasets. The PanNuke dataset annotates nuclei from a wide range of tissues, such as malignant epithelial cells, normal epithelial cells, inflammatory cells, connective and soft tissue cells, and dead cells. In three-fold cross-validation, SpatioCell achieved the highest overall accuracy compared to the other two methods. We also examined the average multi-class panoptic quality (mPQ) per tissue type, and SpatioCell outperformed both methods in most tissues (18 out of 19 tissues) (**Fig. 2D**, **Supplementary Figure 2D**). **Fig. 2E** exhibits nuclear segmentation and classification results from SpatioCell in part of tissues, which illustrates its robust classification capability across diverse histopathological images.

Overall, equipped with SAM and an automatic prompt generator, our extensive benchmarking shows that SpatioCell consistently outperforms state-of-the-art nuclear segmentation and classification models in both accuracy and robustness, through effectively extracting highly reliable morphological information from H&E images.

SpatioCell Achieves Superior Nuclear Segmentation on H&E and DAPI Images from ST Data

To further assess the adaptability of SpatioCell's image analysis model to ST datasets, we applied it to the CellbinDB dataset, which includes microscopic images from ST experiments, such as H&E- and DAPI-stained images, covering both normal and diseased tissues from human and mouse samples[32]. The dataset contains both FFPE and FF tissue images, which are the most commonly used preservation methods in current ST studies.

We conducted benchmark experiments on CellBinDB with the same algorithms and metrics as those in the previous section, and detailed parameters for each model are provided in the **Methods**. Across both FFPE and FF H&E images, SpatioCell consistently achieved the highest scores across all evaluation metrics, demonstrating its robustness in diverse histopathological images (**Supplementary Figure 3A-C**). Notably, SpatioCell exhibited a significant advantage in nuclear segmentation on H&E images of FF samples (**Fig. 2F, G**). Since the accuracy of nuclear detection is critical

for SpatioCell's downstream cell annotation, we specifically focused on evaluating its ability to correctly identify true nuclei. To this end, we also calculated the proportion of correctly detected nuclei across various IoU (Intersection over Union) thresholds for different algorithms. As shown in **Fig. 2G**, SpatioCell again demonstrates superior performance in detecting nuclei from H&E-stained images of FF samples. The segmentation results in **Supplementary Figure 3B** further validate this observation in that SpatioCell accurately identifies small nuclei, while other methods segment noticeable false negatives, failing to detect many nuclei. These results suggest that SpatioCell's image processing module is more adaptable to H&E images from ST.

DAPI staining is another widely used nuclear marker in ST. CellbinDB provides two datasets with DAPI images from 10x Genomics and Stereo-seq. To evaluate whether SpatioCell can be extended to other types of cellular images while maintaining accurate segmentation performance, we benchmarked its performance on DAPI-stained images from 10x Genomics. As SpatioCell, Hover-Net, and Mask R-CNN do not have pretrained models specific to DAPI images, all three were trained on DAPI-stained images from Stereo-seq, and other detailed parameters are provided in the **Methods**. Unlike H&E images, DAPI staining highlights only the nuclei, ignoring overall tissue structure, making the segmentation task more straightforward. As expected, all models achieved high detection and segmentation scores, with SpatioCell consistently outperforming other methods (**Fig. 2G, Supplementary Figure 3A, D**).

In conclusion, these results demonstrate that SpatioCell can robustly perform nuclear segmentation on H&E-stained images in ST, serving as a critical foundation for the subsequent cell annotation module. Moreover, its adaptability highlights strong potential for extension to a broader range of imaging modalities in ST datasets.

SpatioCell Enables Single Cell-resolution Cell Type Annotation on Spatial Transcriptomics Datal

To overcome the limitation of multicellular resolution, the cell annotation module of SpatioCell combines H&E-derived and expression-derived information, enabling the construction of high-resolution single-cell spatial maps and improving the accuracy of cell-type assignment in STs. (**Fig. 3A，Methods**). To comprehensively evaluate on SpatioCell's annotation capability in ST data, we simulated multicellular-resolution ST datasets using single-cell resolution ST (10x Xenium) on four cancer datasets, including breast cancer (BRCA), ovarian cancer (OVCA), pancreatic cancer (PDAC) and lung cancer (LUCA). We compared annotation results from SpatioCell with two baseline

approaches: an H&E image-based classification (adapted from SpatioCell's image analysis module) and a deconvolution-based method that estimates cell type proportions using Cell2location and then randomly assigns cell identities to the segmented cells within each spot, as detailed in **Methods**. Notably, all cell segmentation was performed using SpatioCell's image analysis module. This comparison allows us to assess whether the deep integration of H&E-derived and expression-derived information in SpatioCell outperforms the use of either data type alone.

Cell type annotations were tested at different granularities, with supertypes broadly categorized into tumor cells, normal epithelial cells, connective and soft tissue cells, and inflammatory cells, while subtypes provide more detailed classifications. Our findings showed that SpatioCell surpassed alternative cell annotation methods that rely solely on a single source of information (**Fig. 3B-C**, see results of LUCA in **Supplementary Figure 4**). At a commonly adopted 55μm resolution, the annotation accuracy for supertype annotation across four datasets exhibited a significant improvement over the other two strategies. For example, in the BRCA dataset, SpatioCell achieved 82% supertype annotation accuracy, improving upon the deconvolution-based and H&E-based methods by 13% and 10.7%, respectively (**Figure 3B**). We further assessed the applicability of SpatioCell across different spot sizes, and SpatioCell consistently outperformed both image-based classification and deconvolution-based cell assignment, with its advantage becoming more pronounced in larger spots (**Figure 3B**). For further evaluation on subtype cell identification, we only compared SpatioCell with the deconvolution-based strategy since image-based methods alone lack the resolution needed for precise subtype differentiation. In 55μm-resolution BRCA data, SpatioCell achieved an accuracy of 70.2% across 10 cell subtypes, whereas the deconvolution-based method reached only 59%, demonstrating SpatioCell's superior ability in fine-grained cell annotation. Similarly, in the OVCA and PDAC datasets, SpatioCell exhibited a comparable boost in cell subtype annotation accuracy (**Fig. 3D**). The consistent advantages reinforce its robustness and generalizability across different tissue types.

Moreover, in many multicellular-resolution STs, only partial regions have been sequenced, leading to discontinuities in spatial cell annotations. Therefore, we artificially classified simulated spots into capture regions and spot intervals to mimic the lack of sequencing coverage in inter-spot regions. SpatioCell's performance on

capture regions consistently surpassed deconvolution-based method across multiple datasets (**Fig. 3E**). A key advantage of incorporating H&E features is the ability to impute cell identities in spot intervals using cell-type probability profiles derived from nuclear morphology. These profiles, generated by the image processing module, are compared with those of neighboring cells to infer the most likely identity based on morphological similarity and annotation of adjacent cells. (**Fig. 3F**, **Supplementary Figure 5A**). To validate SpatioCell's ability for cell type imputation at interval region, we measured annotation accuracy across entire tissue sections and within spot interval regions in four datasets. The results demonstrate that the annotation accuracy of SpatioCell outperforms that of deconvolution-based method, when inferred in the same manner as SpatioCell in both whole-section and spot interval regions (see results in 55μm-resolution, as shown in **Fig. 3G**, others detailed in **Supplementary Figure 5B-C**). Notably, across all four tissues, even with intervals, SpatioCell achieved higher overall annotation accuracy than the deconvolution-based method applied to complete (gap-free) sections (**Supplementary Figure 5D**).

We applied SpatioCell to the OVCA dataset with simulated spot intervals to generate a single-cell resolution annotation across the whole slide (**Fig. 3H**). The result indicates that SpatioCell excels in delineating tissue boundaries and tumor structures (such as tubular-shaped tumor) compared to deconvolution-based method, providing a more structured tissue architecture and cellular spatial distribution. Notably, SpatioCell accurately captures the cell-type distribution at tumor margins (as in **Fig. 3H region I-IV**). Beside one follicular cyst, SpatioCell successfully identified tumor cells surrounded by T and NK cells, a key feature in the OVCA microenvironment[33] (**Fig. 3H**). In contrast, deconvolution results merely provided a coarse estimate of cell-type proportions, and random cell type assignments, according to the proportions within a spot, tended to overlook or blur critical cellular compositions.

Above all, by integrating H&E-derived morphological features with sequencing data, SpatioCell provides more biologically meaningful annotation of tissues, thereby enabling deeper insights into tissue architecture.

SpatioCell Facilitates the Accurate Depiction of Tumor Microenvironment in Multicellular-resolution ST.

In the TME, biologically significant regions, such as vascular and glandular structures, tumor invasion fronts, and lymphocyte infiltration zones, are often located at the margins where multiple cell types converge. To further validate the applicability

of SpatioCell in revealing these subtle, complex, and crucial structures, we applied it to real multicellular-resolution ST datasets from various tumor samples. Given that the resolution of multicellular-resolution ST can vary across different sequencing techniques, we selected ST datasets with spatial spot sizes of 55 µm and 100 µm to comprehensively assess SpatioCell's performance at different resolutions, as detailed in the **Data Availability** section. In a 100 µm-resolution ST dataset of triple negative breast cancer (TNBC), the larger spot size leads to higher cell density per spot (up to 294 cells/spot; **Fig. 4A**), while also capturing more cells for sequencing. The intervals between spots are narrower, with fewer cells in those regions (**Fig. 4B**). In contrast, 55 µm-resolution ST, such as the OVCA dataset obtained from 10x Visium, results in fewer cells per spot (**Fig. 4C**). However, only 19.5% of the nuclei fall within sequenced regions (**Fig. 4D**). Overall, these datasets, with their varying resolutions, cell densities within spots, and interval coverage, provide a comprehensive framework by which to assess SpatioCell's adaptability in different scenarios.

After processing these datasets with SpatioCell, as detailed in **Methods**, we obtained a comprehensive single-cell spatial atlas of the TNBC data, closely aligned with the manual annotation labeled by a pathologist (**Fig. 4E**, *i-ii*, **Supplementary Figure 6**). In contrast, the deconvolution-based method produced diffuse and scattered cell-type distributions, obscuring key structural features. Moreover, immune cell infiltration is a critical prognostic marker in breast cancer[34]. In another TNBC slide from the same dataset, SpatioCell successfully identified a few lymphocyte-enriched regions, with abundant T cells and B cells accumulated at the border between tumor and stroma, whereas traditional methods overlooked this characteristic (**Supplementary Figure 7A**).

In a 55 µm-resolution ST data from OVCA, SpatioCell accurately reconstructed conspicuous boundaries between tumor cells and proliferative cells (**Fig. 4F**). In contrast, the deconvolution-based annotation method yielded highly disordered cell-type distributions within each spot, resulting in blurry structures that concealed the original tumor-stroma interfaces (**Fig. 4F**, *i-ii*). These findings confirm that SpatioCell surpasses traditional methods in tissue structure reconstruction and spatial organization analysis.

Overall, in these datasets, deconvolution methods only exhibit a rough picture of cell type composition, which limits the spatial analysis ability (**Fig. E–F**, panels *I-II*). Conventional annotation methods, such as CytoSpace[35], attempt to refine this by

integrating segmentation-based cell positions and randomly assigning identities, but they often fail to effectively reconstruct these intricate cellular compositions because of low accuracy. SpatioCell offers superior resolution and accuracy, which establishes the foundation for precise characterization of tumor microenvironment features.

SpatioCell Improves Deconvolution Results with a Competitive Balanced Index

SpatioCell relies on deconvolution-based constraints for cell annotation, but molecular diffusion, sequencing errors, and uneven sampling cause the inherent noise of ST and single-cell reference data. Consequently, cell-type abundance estimates from deconvolution sometimes deviate from true composition in complex regions. For example, a deconvolution result of the ST data from a lung squamous cell carcinoma (LUSC) sample displays widespread tumor cell distribution, even in stromal regions, whereas H&E images reveal distinct stromal structures and necrotic zones (**Fig. 5A**). This observation highlights the potential of leveraging H&E-derived morphological features to refine cell annotations by correcting deconvolution errors. To achieve this, we introduce Competitive Balance Index (CBI, as detailed in **Methods**) which is a parameter that quantifies the discrepancy between deconvolution results and H&E-derived nuclear classification, with the aim of correcting errors that arise from deconvolution. Specifically, when CBI exceeds a preset threshold, SpatioCell progressively refines the deconvolution result using H&E-derived classification probabilities until the discrepancy falls below the threshold.

To evaluate the impact of CBI, we introduced controlled errors into the simulated multicellular-resolution ST dataset (described in the previous section) and applied CBI-based correction, comparing cell composition accuracy before and after adjustment, using mean absolute error (MAE) , which measures the average magnitude of the errors, and mean squared error (MSE), which measures the average squared magnitude of the errors (see **Methods**). In BRCA data, results showed that CBI significantly improved annotation accuracy under various perturbation levels, with the optimal performance observed when CBI was set to 0.5 (**Fig. 5B**). A similar trend was observed in OVCA, where the optimal CBI setting was around 10 (**Fig. 5C**). The improvement observed in OVCA was slightly lower than that in BRCA, likely owing to the lower resolution of its H&E image, half that of BRCA. Our findings suggest that CBI possesses a strong potential for correcting deconvolution errors and that the CBI threshold should be adapted according to the quality of the H&E image. When image quality is high and nuclear classification is reliable, a smaller CBI allows for more drastic correction of

deconvolution errors. Conversely, for lower-resolution images, a larger CBI helps preserve annotation stability. In practice, we recommend starting with a higher threshold (e.g., 10).

After applying CBI to the LUSC data above, we observed a notable improvement in the annotation of stromal regions embedded within the tumor, which were not clearly delineated in the original deconvolution results. The cell composition also showed corresponding changes (**Fig. 5D**). Notably, by effectively correcting deconvolution errors, SpatioCell reconstructs key tumor microenvironment structures, such as small vessels, stroma in tumor areas, well-defined tumor borders, and immune infiltrations, with annotations closely matching the pathologist's manual labels (**Fig. 5E, Supplementary Figure 8**).

Overall, by leveraging CBI, SpatioCell can balance nuclear classification from H&E images with deconvolution-derived composition estimates, enabling adaptive correction of deconvolution results and revealing key structural features of the TME, thereby enhancing our understanding of TME dynamics.

Discussion

ST offers critical insights into gene expression within the tissue architecture, yet its resolution is inherently constrained by the size of spatial capture spots. Existing deconvolution methods have sought to infer cell compositions within spots, but these methods lack the spatial precision required to determine the exact localization of individual cells. SpatioCell goes further by integrating imaging data-derived morphological features with transcriptomic measurements to reconstruct single-cell spatial maps from multicellular-resolution ST data. Through detailed analysis of cell distribution and tissue architecture, it holds the potential to advance research in developmental biology, tumor microenvironment studies, and disease pathology.

To achieve this, SpatioCell integrates a novel deep learning model for cell segmentation, combining the fine-tuned SAM model with a Prompter built upon a revised SegFormer. The Prompter autonomously detects nuclear centroids, which serve as prompts for the SAM model, while also utilizing a shared feature extractor to capture multi-scale features for semantic classification. This integration ensures robust and accurate segmentation. To evaluate its performance, we conducted tests across six public datasets. SpatioCell outperformed other widely used nuclear segmentation and classification algorithms. Additionally, leveraging SAM’s strong generalizability, SpatioCell excelled in adapting to H&E images from FF tissue, which are commonly

used in ST, and demonstrated rapid adaptation and robust generalization to DAPI-stained datasets, even with limited training data. These results highlight SpatioCell's versatility and its potential for broad application in diverse ST contexts.

Building on this foundation, SpatioCell integrates H&E-derived cell morphology with transcriptomics data, enabling a more biologically meaningful reconstruction of single cell distributions. Algorithms, such as CytoSpace[35], Spotiphy[18] and Tangram[36], which combine traditional cell segmentation methods (e.g., Cellpose[37]) with deconvolution results to estimate cell counts, overlook valuable morphological information from H&E images. SpatioCell, on the other hand, achieves a deeper integration of morphology and transcriptomics for more accurate annotations since different cell types exhibit distinct nuclear staining characteristics. To address this, SpatioCell generates a probabilistic profile for each mask based on its morphological features, rather than assigning a single most likely cell type. This enables SpatioCell to model the integration as a knapsack problem, as described in Results, balancing conflicting information from image and sequence data and ensuring a more coherent fusion. Additionally, based on these probabilistic profiles, SpatioCell extends cell identity assignment beyond the sequencing area, enabling more comprehensive spatial annotation. In this way, SpatioCell achieves high-resolution single-cell annotation and recovers fine-scale structural details that traditional deconvolution methods often miss. For example, in a 100 μm resolution TNBC ST dataset, SpatioCell vividly restored a normal breast lobule structure heavily infiltrated by T and B cells, alongside neighboring tumor cells. Similarly, in OV data, it accurately delineated the tumor-stroma boundary, revealing a stroma-rich region densely populated with proliferative cells, subtle, yet biologically significant, features otherwise obscured by conventional annotation methods.

What's more, SpatioCell demonstrates the effective use of image information to further optimize deconvolution results, as shown in both simulated and real datasets. Although numerous deconvolution algorithms have been developed in recent years, such as Cell2location[38], CARD[39] and RCTD[40], significant errors still persist owing to the complexity of the task and the variability in single-cell reference datasets. To address this, SpatioCell introduced a "CBI" parameter, effectively balancing the reliance on different data types during the annotation process. This integration allows SpatioCell to correct deconvolution errors by leveraging image data. For example, in simulated datasets, we observed significant improvements in deconvolution results; for

example, in BRCA, the MAE was reduced by 13.5% compared to the deconvolution results with 10% perturbations. Similarly, in real LUSC data, SpatioCell corrected errors in tumor-stroma cell annotations caused by deconvolution inaccuracies, resulting in clearer tissue structures. Future directions include using image annotations as prior knowledge in deconvolution analyses, which might lead to more flexible and precise corrections.

Potential future directions for SpatioCell also include further training of the model to unlock its broader application potential. While SpatioCell excels in nuclear segmentation and single-cell annotation, its performance is still limited by the size of training datasets for cell and nuclear segmentation and classification tasks. To expand the training datasets, we can both annotate more stained images and incorporate deconvolution results from ST data, similar to HistoCell, as ST data provide local cell proportions that can be used as labels. Furthermore, SpatioCell's success with DAPI-stained images indicates its broader applicability to other imaging modalities. With expanded training, SpatioCell could be effectively applied to subcellular-resolution STs, where cell contours are often provided in forms like DAPI images, facilitating more precise annotations in both subcellular and single-molecule resolution platforms.

In conclusion, SpatioCell marks a significant advancement in ST by integrating image and transcriptomic data from ST to achieve single-cell resolution annotations, enabling a deeper understanding of tissue architecture and its connection to disease.

Methods

Module I of SpatioCell：H&E image segmentation and analysis

The H&E processing module is designed to achieve precise nuclear segmentation and analysis on H&E-stained images. The algorithm is implemented in Python, based on the PyTorch deep learning library and the MMSegmentation framework. It outputs nuclear segmentation masks, centroid and contour coordinates, and cell type classification probability profiles. Pretrained weights for our segmentation and classification models are available for download, allowing users to perform direct inference on new H&E images.

Architecture of Prompter

The encoder adopts the MixTransformer-b3 (mit-b3) backbone from the Segformer, consisting of an Overlap Patch Embedding layer followed by four hierarchical Transformer blocks. Each Transformer block includes Efficient Self-Attention, a Mix-

Feedforward Network (Mix-FFN), and Overlap Patch Merging, which together extract multi-scale features while preserving local continuity and spatial resolution. Feature maps at each level are downsampled progressively (1, 1/2, 1/4 and 1/8).

The extracted feature maps are first processed through 1×1 convolutions and upsampled to the highest resolution. They are then concatenated along the channel dimension and passed into two decoder branches: the prompt branch generates heatmaps of nuclear centroids, while the classification branch produces semantic segmentation maps representing cell-type probabilities. Both branches employ atrous spatial pyramid pooling (ASPP) with dilation rates of 1, 2, 4, and 8 to capture multi-scale contextual information. The resulting features are concatenated and fed into their respective decoder heads.

In the prompt head, high-value points in the heatmap are extracted as prompt points for SAM. In the classification branch, a softmax layer is applied to generate class probability profiles, with each value indicating a cell-type probability. The Prompter thus provides both spatial prompts for SAM and type-specific probabilities for downstream annotation.

Architecture of SAM

SpatioCell employs SAM for nuclear segmentation. SAM includes an image encoder (ViT), a prompt encoder, and a mask decoder. The image encoder extracts global features from image patches with positional encoding and self-attention. Prompt embeddings, generated from nuclear centroids by the prompter, are fused with image features in the decoder through self- and cross-attention mechanisms. The decoder refines these features using Multilayer Perceptrons (MLPs) and produces segmentation masks by integrating them with upsampled image features.

Prompter Training

SpatioCell adopts a two-stage training strategy, in which the prompter and SAM are trained independently. The Prompter is optimized through two supervised learning objectives: nuclear centroid localization and cell-type classification. To enable accurate localization of nuclear centroids, the prompt branch is trained using Gaussian kernel-based heatmaps as supervision, which is generated on the heatmap to represent the probability of a pixel being at the center of a nucleus. Owing to the imbalance distribution on the heatmap, we used Focal loss[41] in this branch, defined as:

$$L_{prompt} = -\frac{1}{N}\sum_{x,y}\begin{cases}(1-\hat{Y}_{xy})^{\alpha}\, log(\hat{Y}_{xy}) & if\ \hat{Y}_{xy} = 1\\ (1-\hat{Y}_{xy})^{\beta}(\hat{Y}_{xy})^{\alpha}\, log(1-\hat{Y}_{xy}) & otherwise\end{cases}$$

where $\hat{Y}_{xy}$ is defined as the heatmap predict value of pixel (x, y), α and β are hyperparameters, and N denotes the number of total prompt points.

The classification branch is trained as a semantic segmentation task, and total loss is defined as the sum of Dice loss and cross-entropy loss:

$$L_{nc} = L_{ce} + \lambda L_{Dice}$$

where λ represents the hyperparameter that weighs the impact of the two losses, and in this study, λ = 3. Here, we use the mit_b3 weight pretrained on the ImageNet dataset[21, 42].

SAM fine-tuning

To improve its performance in nuclear segmentation, SAM was fine-tuned it on a nucleus instance segmentation dataset. For each nucleus, the coordinates of one pixel within the ground truth mask were randomly selected as a positive prompt. Additionally, with a probability of 50%, a negative prompt was added by selecting the coordinates of a pixel from the nearest neighboring nucleus. This strategy helps the model better distinguish between closely adjacent nuclei. The predicted mask was trained using a combination of Dice loss and Focal loss between the predicted and the ground truth mask:

$$L_{sam} = L_{Dice} + \alpha L_{Focal}$$

where α=1 is set as the default value during training.

Nuclear segmentation inference

During inference, the Prompter generates a predictive heatmap of nuclear centroids along with a semantic map indicating nuclear types. Pixels with heatmap values exceeding a threshold τ are selected as positive prompts, and their corresponding locations in the semantic mask are used to assign cell type labels. The selected prompts, along with the input image patches, are passed to SAM for instance segmentation. Finally, overlapping predictions resulting from the sliding-window tiling strategy are filtered using Matrix NMS.

Module II of SpatioCell: Single-cell resolution Annotation of ST

Estimating cell type compositions within spots

The cell annotation module is compatible with outputs from various spatial deconvolution tools (e.g., Cell2location, RCTD, CARD). In this study, Cell2location was employed using scRNA-seq data derived from the same tissue type. Counts from ST data are preprocessed with mitochondrial and uninformative genes were filtered out. Median number of segmented nuclei within spots was used in 'N_cells_per_location', and 'detection_alpha' was set to 200. All other parameters were kept at their default settings. (Deconvolution results are provided in **Supplementary Figure 6-8**).

Cell type annotation within spots

Cell type annotation within a spot involves two primary inputs: (1) cell nuclei center coordinates and cell type probability profiles, which are output by SpatioCell's H&E image processing module, and (2) cell type compositions within the spot, which are provided by deconvolution approaches. These two inputs are then combined to estimate the distribution of cell types across different spots in a tissue section.

Step 1: Estimating the number of cells in each spot

Each spot is defined by its spatial location and size, which enables the calculation of the total number of cells it contains. Let the total number of cells within a given spot be N_{spot}, and the corresponding cell type proportions obtained from deconvolution be represented as p_{type_i} for each cell type i. The total number of cells of each type in the spot can be estimated as:

$$N_{type_i} = N_{spot} \times p_{type_i}, \forall i \in cell\ types$$

where $\mathrm{N}_{\mathrm{type_i}}$ represents the estimated number of cells of type i in the spot.

Step 2: Dynamic programming for cell type assignment

After estimating the number of cells of each type in the spot, we aim to assign these cell types to each nucleus such that the total sum of assigned probabilities across all nuclei within the spot is maximized. This problem can be formulated as a variant of the knapsack problem, where each segmented nucleus in a spot is considered an "item", associated with cell type probability profile as "values".

SpatioCell performs the cell annotation through a dynamic programming (DP) algorithm that recursively searches all feasible label assignments (i.e. states) across nuclei to identify the globally optimal outcome.

Cell type imputation in spot interval

Segmented cells are first located within each interval region based on spot coordinates and size. For each interval, neighboring spots are identified, and cell types

are assigned by comparing type probability profiles between interval and neighboring cells. The most frequent type among the top-k most similar neighbor nuclei is selected. Similarity is measured using KL divergence, with k set to 5 in this study. The similarity is calculated as follows:

$$similarity = KL_{(p||q)} = -\sum_x p(x) \log \frac{q(x)}{p(x)}$$

where $p(x)$ and $q(x)$ denote cell type probability profile of the query cell at interval and annotated cell in neighbor spots respectively.

Competitive Balance Index (CBI) for deconvolution improvement

CBI quantifies the relative discrepancy between image-based nuclear classification and deconvolution-derived cell-type assignments within a spot. Specifically, after annotation, we compute the total cell-type probability $P_{assigned}$ and compare it to the sum of the maximum class probabilities of nuclei P_{images}, defined as:

$$CBI = \frac{P_{images} - P_{assigned}}{P_{assigned} + 10^{-4}}$$

where 10^{-4} is added as smoothing factor to prevent exaggerated CBI values.

If the discrepancy exceeds the predefined CBI threshold, SpatioCell iteratively updates cell-type assignments by identifying nuclei whose H&E-based type differ from the deconvolution-based assignment. These cells are ranked by their highest classification probability across all types, and their assigned supertypes are replaced with H&E-predicted types progressively. After each update, $P_{assigned}$ is recalculated, and the process continues until the CBI falls below the threshold.

Subtypes for adjusted cells are further assigned from the most similar unmodified cell within the same spot that shares the same supertype (similarity defined as in interval imputation).

Benchmarks on nuclear segmentation and classification

We benchmarked the image processing module of SpatioCell with five algorithms including Hover-Net, Cellpose, Stardist, Mask-RCNN and CellProfiler, on six public datasets: Kumar, CPM 17, PanNuke, CoNSeP, MoNuSAC, and CellbinDB. Nuclear segmentation performance was evaluated on all except MoNuSAC datasets since Stardist was pretrained on it. Cell type classification was benchmarked on PanNuke, CoNSeP, and MoNuSAC using SpatioCell, Hover-Net and Mask-RCNN.

For each dataset, we used its original training, validation, and testing splits. In

PanNuke dataset, we conducted a three-fold cross-validation on its predefined partitions to ensure robust evaluation. We used AJI, DQ, SQ and PQ to comprehensively validate nuclear segmentation. AJI and SQ measures the segmentation quality while DQ evaluates models' ability of nuclei detection (F1-score for detection). PQ considered SQ and PQ together.

Segmentation performance on DAPI-stained images is tested using 10x Genomics dataset in CellBinDB. Since SpatioCell, Hover-Net, and Mask R-CNN do not offer pre-trained weights for DAPI-stained images, we trained these models from scratch using DAPI-stained images from Stereo-seq dataset in CellBinDB. In contrast, CellPose3 and StarDist were evaluated using their respective pre-trained weights, cyto3 and fluoro.

Data availability

All datasets used in this study have already been published and were obtained from public data repositories.

References

1. Rao A, Barkley D, Franca GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211-220 (2021).
2. Palla G, Fischer DS, Regev A, Theis FJ. Spatial components of molecular tissue biology. *Nat Biotechnol* **40**, 308-318 (2022).
3. Marx V. Method of the Year: spatially resolved transcriptomics. *Nat Methods* **18**, 9-14 (2021).
4. Asp M, Bergenstrahle J, Lundeberg J. Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration. *Bioessays* **42**, e1900221 (2020).
5. Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, Haque A. An introduction to spatial transcriptomics for biomedical research. *Genome Med* **14**, 68 (2022).
6. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
7. Eng CL*, et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235-239 (2019).
8. Stahl PL*, et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82 (2016).
9. Cao J*, et al.* Decoder-seq enhances mRNA capture efficiency in spatial RNA sequencing. *Nat Biotechnol* **42**, 1735-1746 (2024).
10. Chen A*, et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777-1792 e1721 (2022).
11. Moses L, Pachter L. Museum of spatial transcriptomics. *Nat Methods* **19**, 534-546 (2022).
12. Hou N*, et al.* Artificial intelligence in cell annotation for high-resolution RNA

sequencing data. *TrAC Trends in Analytical Chemistry* **178**, (2024).
13. Li H, *et al.* A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. *Nat Commun* **14**, 1548 (2023).
14. Wang Y, *et al.* A systematic evaluation of computational methods for cell segmentation. *Brief Bioinform* **25**, (2024).
15. Ma J, *et al.* The multimodality cell segmentation challenge: toward universal solutions. *Nat Methods* **21**, 1103-1113 (2024).
16. Li Y, Luo Y. STdGCN: spatial transcriptomic cell-type deconvolution using graph convolutional networks. *Genome Biol* **25**, 206 (2024).
17. Wan X, *et al.* Integrating spatial and single-cell transcriptomics data using deep generative models with SpatialScope. *Nat Commun* **14**, 7848 (2023).
18. Yang J, *et al.* Spotiphy enables single-cell spatial whole transcriptomics across an entire section. *Nat Methods*, (2025).
19. Kirillov A, *et al.* Segment Anything. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*) (2023).
20. Ma J, Wang B. Towards foundation models of biological image segmentation. *Nat Methods* **20**, 953-955 (2023).
21. Xie E, Wang W, Yu Z, Anandkumar A, Alvarez JM, Luo P. SegFormer: simple and efficient design for semantic segmentation with transformers. In: *Proceedings of the 35th International Conference on Neural Information Processing Systems*). Curran Associates Inc. (2021).
22. Ruining D, *et al.* Segment Anything Model (SAM) for Digital Pathology: Assess Zero-shot Segmentation on Whole Slide Imaging. In: *Medical Imaging with Deep Learning, short paper track*) (2023).
23. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN.) (2017).
24. Graham S, *et al.* Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal* **58**, 101563 (2019).
25. Stringer C, Pachitariu M. Cellpose3: one-click image restoration for improved cellular segmentation. *Nat Methods* **22**, 592-599 (2025).
26. Schmidt U, Weigert M, Broaddus C, Myers G. Cell Detection with Star-Convex Polygons. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (eds Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G). Springer International Publishing (2018).
27. Carpenter AE, *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7**, R100 (2006).
28. Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A, Sethi A. A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology. *IEEE Transactions on Medical Imaging* **36**, 1550-1560 (2017).
29. Kirillov A, He K, Girshick R, Rother C, Dollár P. Panoptic Segmentation.) (2019).
30. Vu QD, *et al.* Methods for Segmentation and Classification of Digital Microscopy Tissue Images. *Front Bioeng Biotechnol* **7**, 53 (2019).
31. Gamper J, *et al.* PanNuke Dataset Extension, Insights and Baselines. Preprint at https://ui.adsabs.harvard.edu/abs/2020arXiv200310778G (2020).

32. Shi C, *et al.* CellBinDB: A Large-Scale Multimodal Annotated Dataset for Cell Segmentation with Benchmarking of Universal Models. *bioRxiv*, 2024.2011.2020.619750 (2024).
33. Yang B, *et al.* Spatial heterogeneity of infiltrating T cells in high-grade serous ovarian cancer revealed by multi-omics analysis. *Cell Rep Med* **3**, 100856 (2022).
34. Karn T, *et al.* Association Between Genomic Metrics and Immune Infiltration in Triple-Negative Breast Cancer. *JAMA Oncol* **3**, 1707-1711 (2017).
35. Vahid MR, *et al.* High-resolution alignment of single-cell and spatial transcriptomes with CytoSPACE. *Nat Biotechnol* **41**, 1543-1548 (2023).
36. Biancalani T, *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods* **18**, 1352-1362 (2021).
37. Pachitariu M, Stringer C. Cellpose 2.0: how to train your own model. *Nat Methods* **19**, 1634-1641 (2022).
38. Kleshchevnikov V, *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol* **40**, 661-671 (2022).
39. Ma Y, Zhou X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat Biotechnol* **40**, 1349-1359 (2022).
40. Cable DM, *et al.* Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat Biotechnol* **40**, 517-526 (2022).
41. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**, 318-327 (2020).
42. Russakovsky O, *et al.* ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vision* **115**, 211–252 , numpages = 242 (2015).

Acknowledgements

We are grateful to Dr. Zaoyu Wang and Dr. Dengfeng Cao from the Department of Pathology, Renji Hospital, for their expert guidance and contributions to the annotation of histopathological slides.

Author contributions

Conceptualization, J.S.; Investigation, N.H., Z.W, Y.Y., Z.Z., F.X., and Z.W.; Writing, J.S., N.H., and Y.Y.; Visualization: N.H., Y.Y., Y.Z., J.W., and Z.W.; Supervision, J.S., W.T., and C.Y..

Competing interests

The authors declare no competing interests.

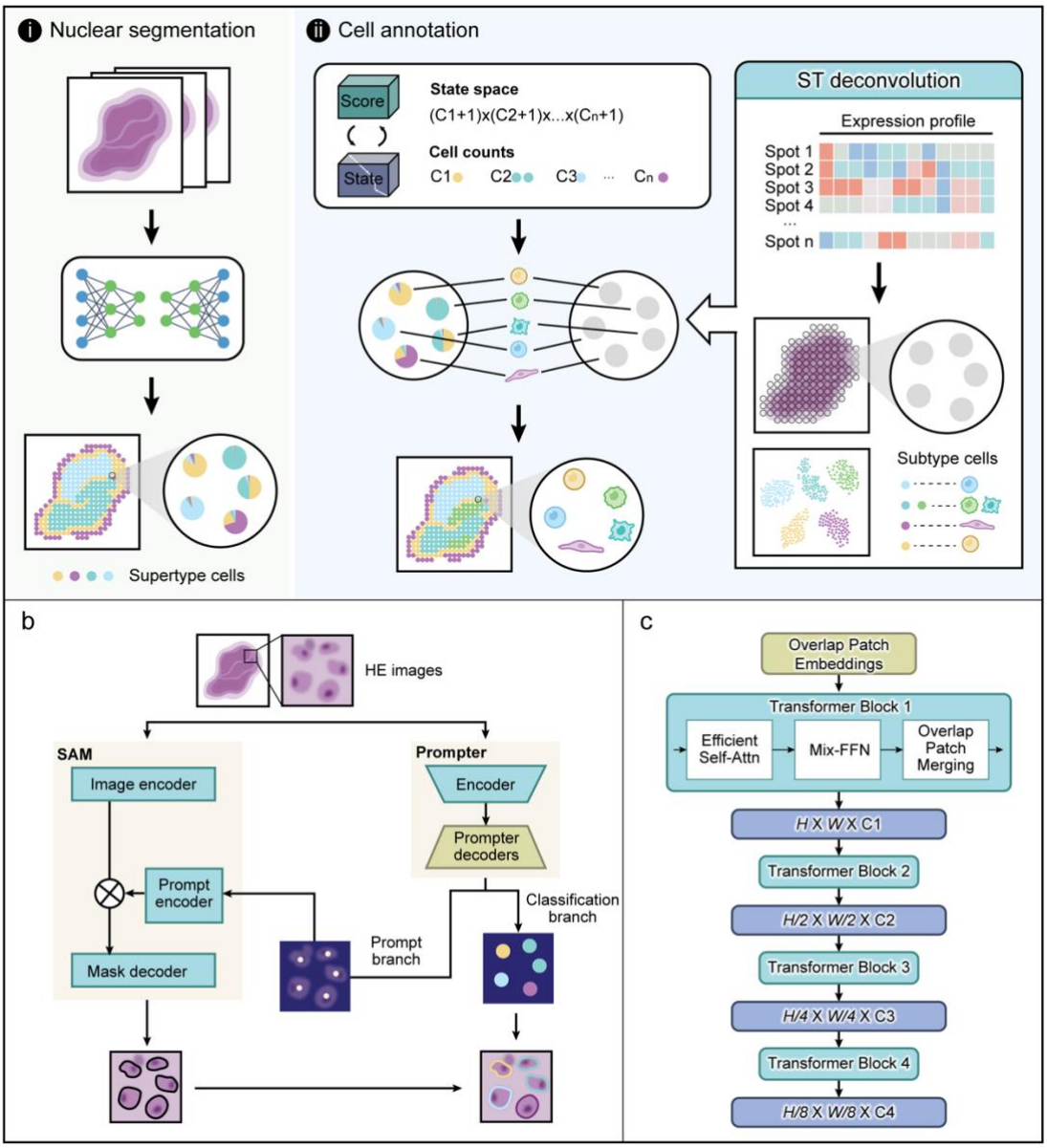

Figure 1. Overview of SpatioCell.

A. SpatioCell computational framework. SpatioCell consists of two main modules: (1) a deep learning module for histopathological image processing that performs nuclear segmentation and analysis through detailed morphological feature extraction, and (2) a dynamic programming module for cell annotation that integrates supertype and subtype annotations derived from morphological and transcriptomic data to achieve high precision and resolution.

B. Deep learning architecture of module 1. The framework includes (1) a fine-tuned SAM-based segmentation model and (2) an adaptive prompting mechanism, which allows SAM to receive accurate positive prompts for automated cell segmentation and analysis.

C. Detailed architecture of the Prompter. The Prompter's encoder, derived from Segformer, uses overlapping patch embedding to extract hierarchical features. This includes (1) Efficient Self-Attention for long-range dependencies, (2) Mix-FFN for

feature transformation, and (3) Overlap Patch Merging for spatial dimension reduction. The output feature maps enable nuclear center detection and semantic segmentation.

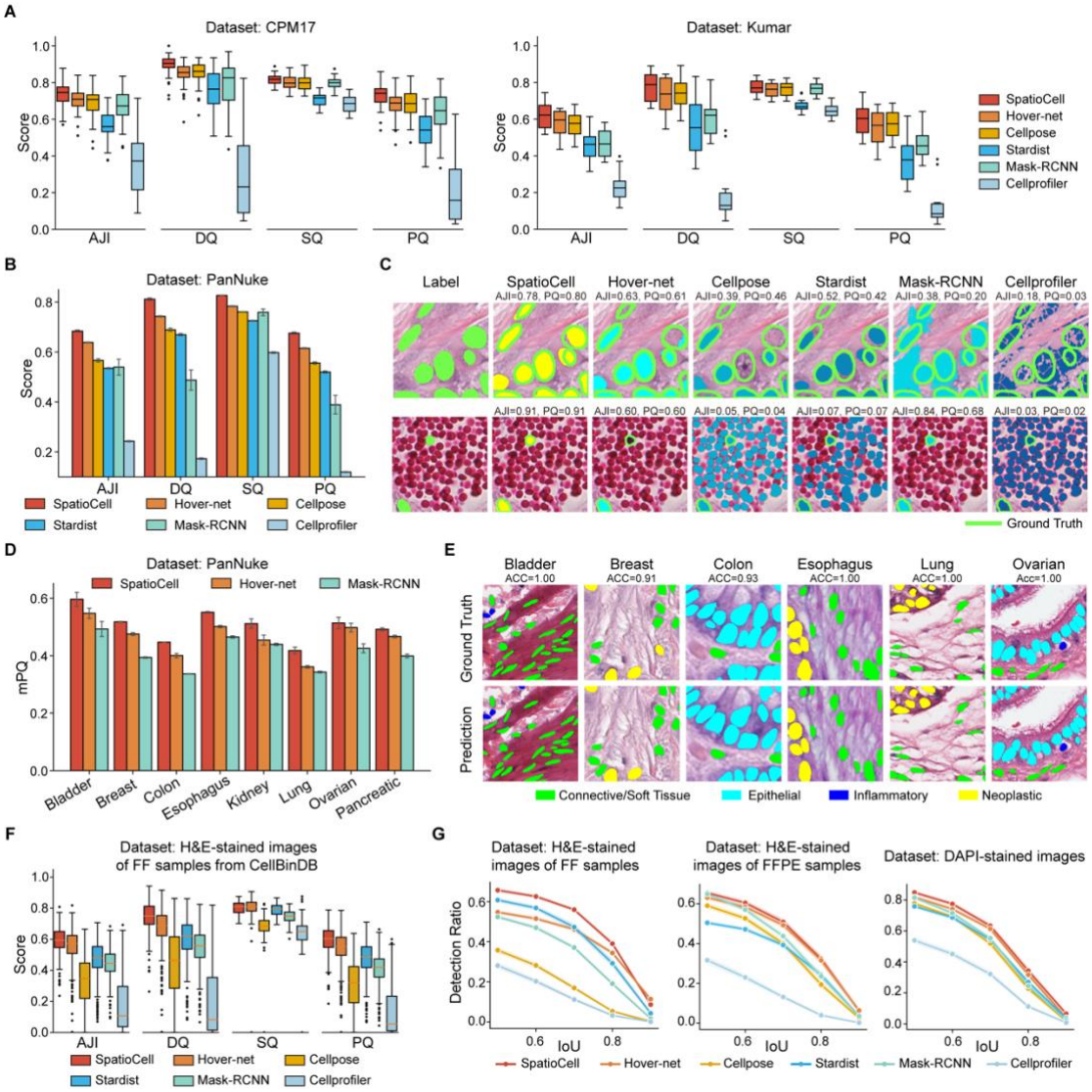

Figure 2. Performance evaluation of SpatioCell in nuclear segmentation and analysis.

A. Segmentation performance comparison of SpatioCell with other 5 methods on CPM17 (left) and Kumar (right) datasets. Each data point represents a metric calculated from a sample image in the test datasets.

B. Segmentation performance of various methods on the PanNuke dataset via three-fold cross-validation. Error bars represent standard error.

C. Visual examples of nuclear segmentation on the PanNuke dataset, with AJI and PQ scores annotated.

D. Distribution of mPQ scores of SpatioCell, Hover-net, and Mask-RCNN on the PanNuke dataset through three-fold cross-validation. Cell type classification for SpatioCell is based on the maximum probability in the classification probability profile.

E. Visual examples of nuclear classification by SpatioCell on the PanNuke dataset.

F. Nuclear segmentation performance of different methods on H&E-stained images of

FF samples from the CellbinDB dataset.

G. Detection ratio of various methods on the CellbinDB dataset under different IoU thresholds.

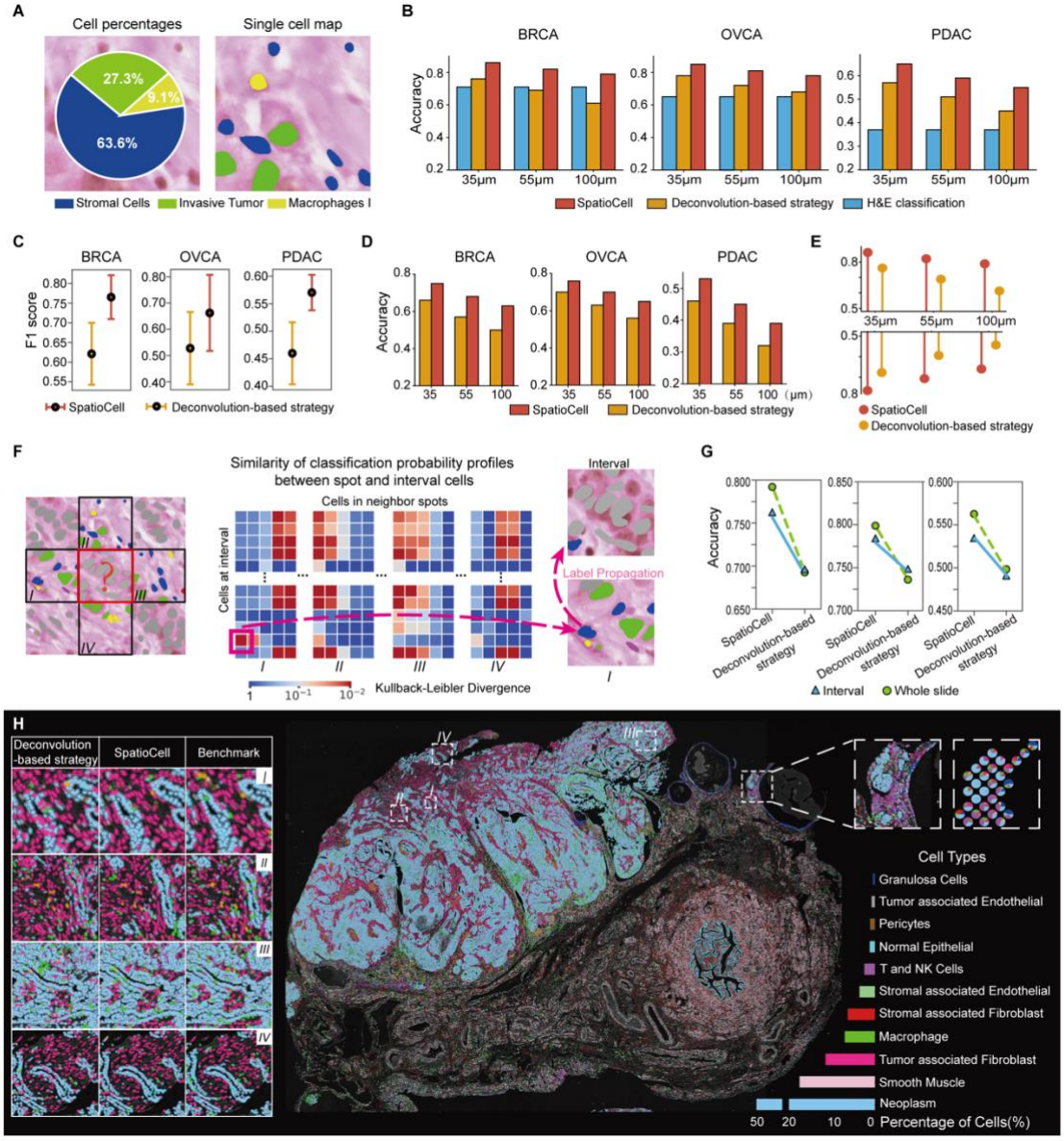

Figure 3. Benchmark and visualization of SpatioCell cell annotation using simulated multicellular-resolution ST data.

A. Schematic of SpatioCell's cell annotation, which transfers deconvolution-inferred cell type composition into a single-cell spatial map.

B. Comparison of supertype cell annotation accuracy between SpatioCell, deconvolution-based annotation, and H&E classification on simulated datasets (including BRCA, OV, and PDAC tissues) at 35 μm, 55 μm, and 100 μm spot resolutions.

C. Comparison of supertype annotation accuracy (measured by F1-score) between SpatioCell and deconvolution-based methods across three simulated datasets, with

circles representing the median and error bars indicating the interquartile range (IQR). Data from the 55 μm spot size is shown.

D. Comparison of subtype annotation accuracy between SpatioCell and deconvolution-based methods on simulated datasets at 35 μm, 55 μm, and 100 μm spot sizes.

E. Comparison of supertype (top) and subtype (bottom) annotation accuracy within spots across three datasets and various spot sizes between SpatioCell and deconvolution-based methods when interval regions are present.

F. Schematic of SpatioCell's imputation strategy for spot interval regions. Gray cells represent query cells, and black-framed squares indicate annotated neighboring spots. The heatmap visualizes the KL divergence between query and annotated cells based on probability profiles from module 1, with the darkest red square indicating the lowest divergence.

G. Comparison of supertype annotation accuracy across three datasets at a 55 μm spot size between SpatioCell and deconvolution-based methods, showing accuracy within spot interval regions (blue) and across the entire slide (green).

H. Single-cell spatial map of a simulated OV dataset with interval region imputation using SpatioCell. The left panel compares deconvolution-based, SpatioCell, and benchmark annotations. The top-right inset highlights SpatioCell's annotation in a lymphocyte-infiltrated tumor region, with deconvolution-derived cell composition pie charts for the same area.

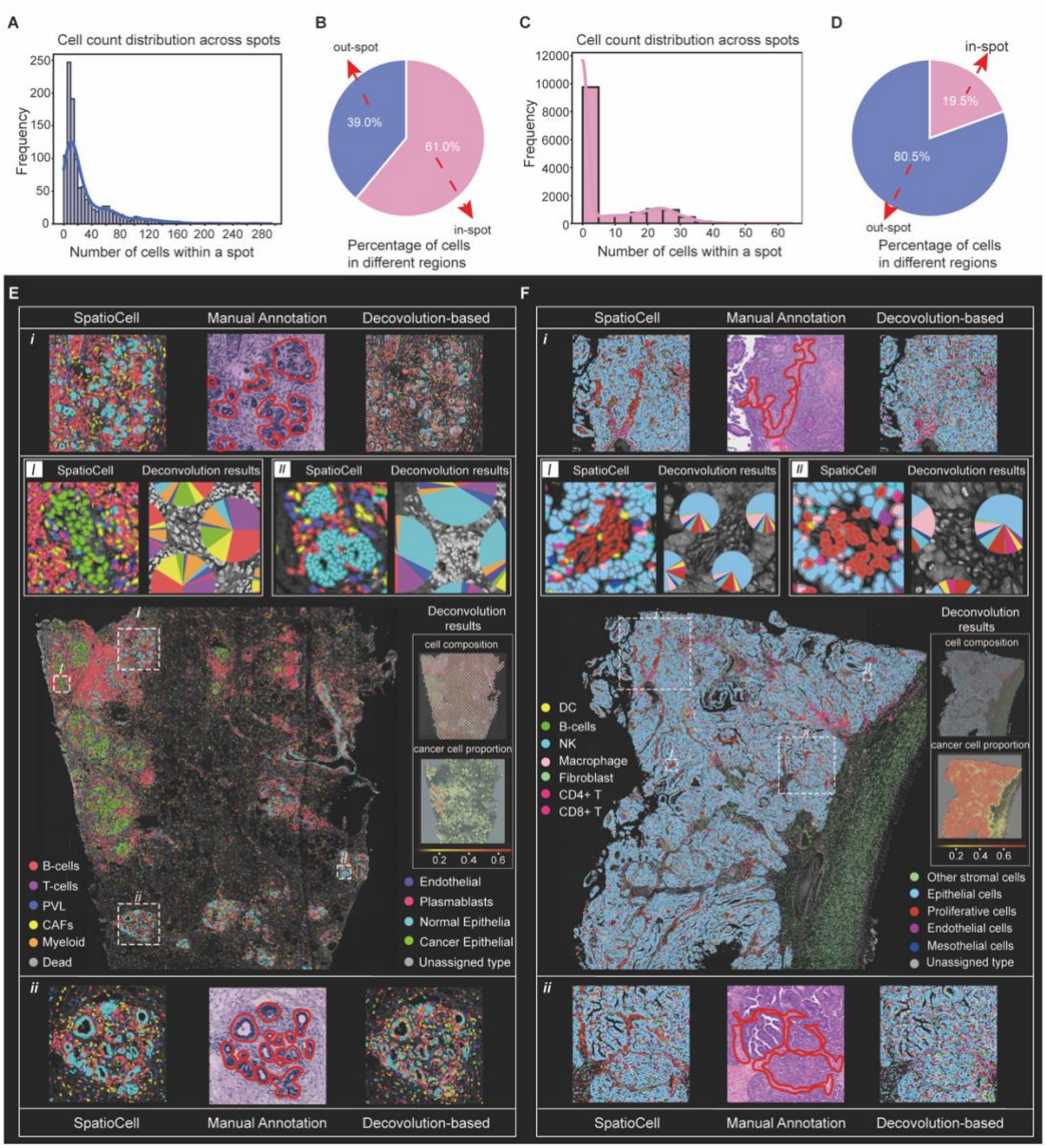

Figure 4. SpatioCell identifies characteristic tissue microenvironment in multicellular-resolution STs.

A. Distribution of cell counts in captured regions of a TNBC sample sequenced with 100 μm spot size, with a density curve showing cell number variability per spot.

B. Proportion of total cell counts between spots and inter-spot regions in the TNBC sample.

C. Distribution of cell counts in captured regions of an ovarian cancer (OV) sample sequenced with 55 μm spot size.

D. Proportion of total cell counts between captured and inter-spot regions in the OV sample.

E. Single-cell spatial map of the TNBC sample resulted from SpatioCell, with the right panel showing cell composition and cancer cell proportion from deconvolution. (I, II) Comparison of cellular annotation by SpatioCell and deconvolution results, with SpatioCell significantly improving resolution. (i-ii) Comparison of cellular annotation

by SpatioCell, manual annotation on H&E, and deconvolution-based methods.

F. Single-cell spatial map of the OV sample resulted from SpatioCell, with details similar to (E).

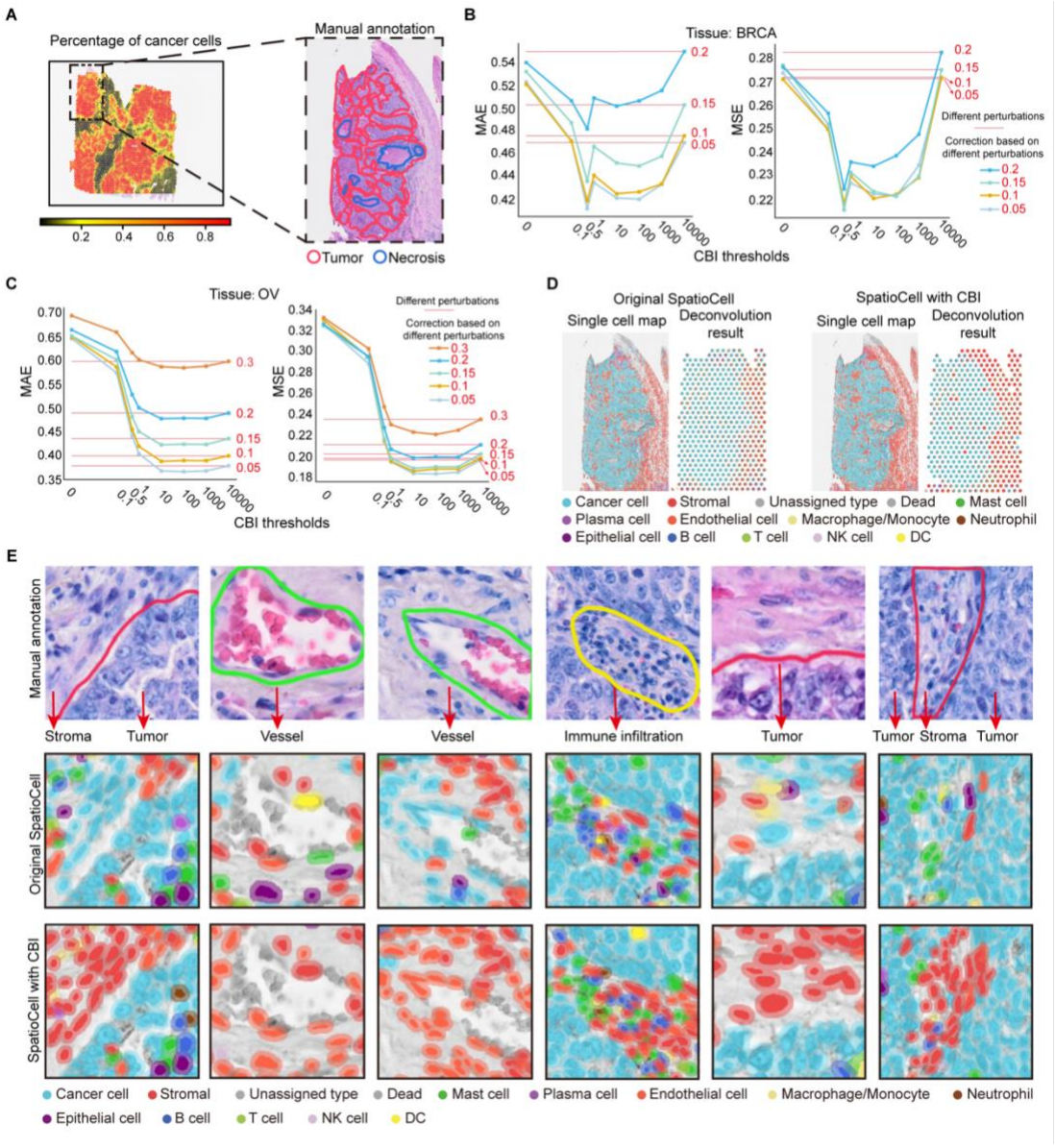

Figure 5. Validation and visualization of SpatioCell with Competitive Balance Index (CBI).

A. Cancer cell proportion estimated by Cell2location in an LUSC sample from 10x Visium (left), with a zoomed-in H&E image (right) highlighting tumor (red) and necrosis (blue) areas as annotated by a pathologist.

B. Distribution of mean absolute error (MAE, left) and mean squared error (MSE, right) of corrected deconvolution results for BRCA spatial transcriptomics (ST) data across different CBI thresholds. Colored lines represent correction results under various perturbation levels (0.2, 0.15, 0.1, 0.05). Red horizontal lines indicate the corresponding MAE and MSE values without CBI thresholding for each perturbation level.

C. Distribution of MAE (left) and MSE (right) of corrected deconvolution results for OV ST data across different CBI thresholds. Colored lines represent correction results under various perturbation levels (0.3, 0.2, 0.15, 0.1, 0.05). Red horizontal lines indicate the corresponding MAE and MSE values without CBI thresholding for each perturbation level.

D. Comparison of cell annotation results and deconvolution before and after correction based on CBI.

E. Comparison of cell annotations in 10x Visium-sequenced LUSC ST data: Top row shows manual H&E annotations, second row shows original SpatioCell annotations, and third row shows SpatioCell annotations refined with CBI. The refined annotations align more closely with true tissue structures.