SpatioCell: Deep Integration of Histology and Spatial Transcriptomics for Profiling the Cellular Microenvironment at Single-Cell Level.

Naiqiao Hou, [#1,2] Yue Yu, [#1,2] Zhaorun Wu, [#5] 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, Institutes of Biomedical Sciences, Inner Mongolia University, Hohhot 010000, China.

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

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

Keywords: Spatial Transcriptomes, Artificial Intelligence, Single-cell Annotation, Tumor Microenvironment.

Abstract

Spatial transcriptomics (ST) is a powerful assay to capture gene expression in tissue context. However, due to the limitation of resolution, most existing ST datasets remain at multicellular resolution which hinders comprehensive understanding of the spatial organization. We propose SpatioCell, a computational algorithm to automatically extract both cell type and expression information at single-cell resolution from ST data, through a morpho-transcriptomic spatial reconstruction framework solved via dynamic programming, integrating morphological and transcriptomic information. This framework enables deterministic single-cell spatial reconstruction, assigning cell identities to precise locations rather than inferring spot-level cell-type composition and expression, and further uncovers overlooked microenvironment features while correcting deconvolution errors. Using the ST data from triple negative breast cancer as an example, SpatioCell reveals the significance of the distance between cancer-associated fibroblast (CAF) and tumor or immune cells for disease progression. The establishment of SpatioCell will broaden the biomedical applications of ST and facilitate investigations of single-cell spatial organization.

Background

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].

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 a spot). As most current ST datasets are derived from these sequencing-based platforms, the growing volume of data, while offering valuable insights into tissue organization, still largely remains at multicellular resolution, hindering single-cell level analysis [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 estimate only cell-type proportions rather than achieving actual single-cell resolution, which limits precise spatial mapping and hinders comprehensive characterization of microenvironmental heterogeneity, especially in terms of the fine-scale spatial distribution, intercellular distances, and cell interactions.

Notably, ST data is often accompanied by histology images, mostly through hematoxylin and eosin (H&E) staining, which provide additional spatial context that could enhance resolution. Histology image analysis, a key focus in computational pathology, has led to the development of several cell/nuclear segmentation models[14, 15]. Current studies (such as SpatialScope[16] and Spotiphy[17]) have integrated these segmentation models with ST data analysis. However, these approaches are mainly limited to either counting the number of nuclei within each spot or using histology-derived information to heuristically distribute spot-level gene expression in pursuit of super-resolution profiles, but they still cannot precisely assign cell types to spatial positions. This limitation largely arises because few models can accurately perform both nuclear segmentation and cellular morphological analysis on H&E-stained images. Moreover, no effective annotation methods are available to integrate image-derived morphological features with sequencing information.

To address these challenges, we developed SpatioCell, a morpho-transcriptomic spatial reconstruction framework that produces high-resolution, single-cell–resolved maps of both cell types and their gene expression. SpatioCell re-engineers the Segment Anything Model (SAM) [18], an instance

segmentation foundation model, by redesigning its prompting[19] and decoding mechanisms to achieve robust and accurate nuclear segmentation and morphological analysis in complex tissues. It further models the cell-type assignment, based on histology-derived cell-identity features and deconvolution predictions, as a constrained combinatorial optimization problem, and employs a dynamic programming (DP) algorithm to achieve precise single-cell annotation. In extensive evaluations on six public datasets comprising thousands of tissue-stained images or ST slides, SpatioCell outperformed state-of-the-art methods in nuclear segmentation and classification. Applied to tumor samples with multicellular-resolution ST (55–100 μm), it corrected deconvolution errors and reconstructed single-cell maps of cell type and expression, revealing tumor microenvironment (TME) features undetectable by existing approaches. In triple-negative breast cancer (TNBC), SpatioCell identified cancer-associated fibroblast (CAF) subtypes with distinct spatial proximities to tumor and immune cells. By integrating intercellular distance, gene expression, and clinical outcome data, it further revealed functionally and prognostically relevant CAF states shaped by the local microenvironment. These findings demonstrate the utility of SpatioCell as a powerful framework for decoding microenvironment-driven cellular heterogeneity, and further suggest its potential to advance the biomedical applications of ST by enabling accurate analysis of cell–cell distances, interactions, functions, and their implications in disease.

Results

Overview of SpatioCell

To achieve single-cell–level cell annotation in multicellular-resolution ST data, existing algorithms use histopathological images to locate and count cells within each spot but often overlook the rich morphological cues, leaving precise single-cell mapping beyond the reach of current methods. Here we developed SpatioCell, by integrating morphological information from histopathological images, it enables single-cell–resolved cell-type assignment and elevating gene expression profiles to the single-cell level. SpatioCell comprises two key modules: (1) a specialized spatio-morphological learner for histopathological image processing, which performs nuclear segmentation and detailed morphological

77 feature extraction, and (2) a morpho-transcriptomic spatial reconstruction module for cell annotation
78 that integrates image-derived morphological features with transcriptomic data to achieve high-precision,
79 single-cell–level annotation (**Fig. 1A**).

80 Conventional SAM, while powerful in natural image segmentation[18], struggles to generalize to
81 H&E-stained pathology images due to blurred nuclear boundaries, heterogeneous cell morphologies,
82 and densely packed tissue architectures[20]. To address these limitations, we propose a spatio-
83 morphological learner within SpatioCell, a pathology-tailored segmentation framework that
84 systematically re-engineers SAM's prompting and decoding mechanisms. Unlike prior adaptations that
85 only fine-tune SAM, SpatioCell introduces three key innovations. First, it leverages nuclear mask–
86 derived prompts during fine-tuning, aligning the foundation model with histopathology-specific object
87 scales and boundaries. Second, we design an automatic Prompter, built upon a multi-scale transformer
88 encoder[19], that simultaneously encodes global tissue context and fine-grained nuclear details to generate
89 precise positive prompts without manual intervention. Third, we augment the Prompter with a semantic
90 classification head, enabling joint modeling of nuclear segmentation and cell-type identification within
91 a unified architecture (**Fig. 1B**). Through this design, SpatioCell outputs both (i) accurate nuclear masks
92 and (ii) cell identity (supertype) probability profiles. This reconfiguration of SAM into a specialized
93 spatio-morphological learner establishes a new paradigm for foundation models in computational
94 pathology.

95 While H&E images provide valuable insights into cell localization and coarse cell supertypes, they
96 lack the resolution to resolve fine-grained subtypes and are often noisy, necessitating integration with
97 ST. Conventional ST analysis relies on deconvolution algorithms to estimate the aggregate composition
98 of cell subtypes within each spot, but these approaches operate only at spot resolution. SpatioCell
99 introduces a novel optimization-based framework to bridge this gap by coupling morphology-derived
100 probability profiles with deconvolution-derived compositional constraints. Specifically, the spatio-
101 morphological learner outputs a probability distribution across cell types for each nucleus, based on
102 which we reformulate the cell identity assignment as a constrained combinatorial optimization problem

(i.e. a knapsack problem) (**Fig. 1C**): the objective is to maximize the global likelihood of type assignments under limited capacity defined by deconvolution-inferred cell-type proportions. This formulation enables a dynamic programming (DP) solver to jointly optimize assignments across all cells in a spot, ensuring consistency with global transcriptomic priors while leveraging fine-grained morphological cues. Lastly, by mapping the cell identities with DP and modeling the spot-level counts as a multinomial distribution of each gene's reference profiles weighted by the cell-type proportions[17], we further resolve the inferred gene expression for each cell at specific spatial location, thereby generating a high-precision and high-resolution atlas of cell identity and gene expression. In contrast to prior mapping approaches, SpatioCell's optimization-driven integration achieves both fidelity to transcriptomic constraints and resolution at the single-cell level, establishing a new paradigm for cross-modal reconstruction in ST.

Above all, SpatioCell introduces a framework that integrates histological and transcriptomic data to enhance ST resolution to the single-cell level. It refines nuclear segmentation with a specialized spatio-morphological learner and assigns cell identities and gene expression using a constrained optimization strategy, ensuring precise and robust annotation while achieving high-resolution spatial mapping of transcriptomic profiles.

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

To assess the effectiveness of SpatioCell's image processing module, we benchmarked it against five widely used nuclear segmentation and classification algorithms, including Hover-Net[21], Cellpose3[22], StarDist[23], Mask R-CNN[24], and CellProfiler[25]. Evaluation metrics included the Aggregated Jaccard Index (AJI), Panoptic Quality (PQ) and its components (Detection Quality, DQ and Segmentation Quality, SQ), as well as the detection ratio for instance segmentation. For nuclear type classification, mean PQ (mPQ) per class and a classification-aware F-score F_C, proposed by Graham et al., were used[21]. These metrics provided a comprehensive assessment of both instance segmentation accuracy and semantic segmentation performance (see **SI Note 1**. Evaluation Metrics for details).

128 These models were first tested on four public H&E-stained histopathology datasets for the instance
129 segmentation task (see **SI Note 1**. Dataset for details). SpatioCell achieved the best performance across
130 most evaluation metrics in CPM17, Kumar and CoNSeP datasets, including in challenging cases with
131 densely packed and overlapping nuclei, further demonstrating its robustness and reliability (**Fig. 2A**,
132 **Fig. S1**, **Table S1** to **3**). In PanNuke, one of the most complex benchmarks with over 200,000 annotated
133 nuclei across 19 tissue types[26], SpatioCell consistently outperformed all other models in both nuclear
134 detection and segmentation metrics across three cross-validation folds (**Fig. 2B**, **Table S4**). Furthermore,
135 SpatioCell both successfully detected nuclei with weaker staining and accurately excluded unstained
136 mature red blood cells (**Fig. 2C**), demonstrating its ability to remove background artifacts and accurately
137 segment true nuclei.

138 To further assess its generalizability to ST datasets, we conducted benchmark experiments on the
139 CellBinDB dataset, which includes H&E-stained images from both formalin-fixed, paraffin-
140 embedded(FFPE) and fresh frozen(FF) preparations, covering both normal and diseased tissues from
141 human and mouse samples[27]. SpatioCell consistently achieved the highest scores across all evaluation
142 metrics, demonstrating its robustness in diverse histopathological images, particularly excelling in the
143 segmentation of FF H&E images, where nuclei are more faint and morphologically variable (**Fig. 2D**
144 and **E**, **Fig. S2A** to **C**). The accuracy of nuclear detection is critical for SpatioCell's downstream cell
145 annotation and quantitative analysis across multiple IoU thresholds confirmed SpatioCell's superior
146 nuclear detection capabilities, especially on small and faint nuclei (**Fig. 2E**, **Fig. S2B**). Furthermore, we
147 tested SpatioCell on DAPI-stained images from the dataset, a common imaging type in ST, and found
148 it generalizes well across different staining methods (**Fig. 2E**, **Fig. S2A** and **D**).

149 To incorporate additional morphological information and enhance cell annotation, SpatioCell was
150 further developed to include a nuclear classification module, providing essential input for the subsequent
151 cell annotation step. To evaluate its effectiveness, we compared SpatioCell with Hover-Net and Mask
152 R-CNN, two widely used semantic segmentation models, across three benchmark datasets: PanNuke,
153 MoNuSAC, and CoNSeP. SpatioCell outperformed both methods across all three datasets (see results

154 of CoNSeP and MoNuSAC in **Table S5-6**). Specifically, in three-fold cross-validation on PanNuke,
155 SpatioCell achieved the highest F_C compared to the other two methods (**Table S7**). We also examined
156 the mPQ per tissue type, and SpatioCell outperformed both methods in18 out of 19 tissues (**Fig. 2F**, **Fig.**
157 **S3**). As shown in **Fig. 2G**, SpatioCell demonstrates robust classification capability across diverse
158 histopathological images.

159 **High-Precision Single-Cell Annotation and Gene Expression Mapping in Spatial Transcriptomics**

160 To overcome the limitation of multicellular resolution, the cell annotation module of SpatioCell
161 combines H&E-derived and expression-derived information, enabling the construction of high-
162 resolution single-cell spatial maps and improving the accuracy of cell-type assignment in STs. (**Fig. 3A，**
163 **Methods**). To comprehensively evaluate SpatioCell's annotation capability in ST data, we simulated
164 multicellular-resolution ST datasets based on single-cell resolution ST (10x Xenium) from breast cancer
165 (BRCA), ovarian cancer (OVCA), pancreatic cancer (PDAC) and lung cancer (LUCA). We compared
166 annotation results from SpatioCell with two widely used approaches: an H&E image-based
167 classification and a deconvolution-based method that randomly assigns identities to the segmented cells
168 within each spot based on cell type proportions, representing advanced deconvolution-based strategies
169 such as those used in Spotiphy[17], as detailed in **Methods**. This comparison allows us to assess whether
170 the deep integration of H&E-derived and expression-derived information in SpatioCell outperforms the
171 use of either data type alone.

172 Cell type annotations were tested at different granularities: four broad supertypes and more detailed
173 subtypes. SpatioCell surpassed alternative cell annotation methods that rely solely on a single source of
174 information (**Fig. 3B** and **C**, results of LUCA are in **Fig. S4**). At a commonly adopted 55μm resolution,
175 the annotation accuracy for supertype annotation across four datasets exhibited a significant
176 improvement over the other two strategies. For example, in the BRCA dataset, SpatioCell achieved a
177 supertype annotation accuracy of 82%, improving upon the deconvolution-based and image-based
178 methods by 13% and 10.7%, respectively (**Fig. 3B**). It consistently outperformed both methods across
179 different spot sizes, with advantages more pronounced in larger spot (**Fig. 3B**). For further evaluation

180 on subtype cell identification, we only compared SpatioCell with the deconvolution-based strategy since
181 image-based methods alone lack the resolution needed for precise subtype differentiation. In 55μm-
182 resolution BRCA data, SpatioCell achieved an accuracy of 70.2% across 10 cell types, whereas the
183 deconvolution-based method reached only 59%, demonstrating SpatioCell's superior ability in fine-
184 grained cell annotation. Similarly, in the OVCA and PDAC datasets, SpatioCell exhibited a comparable
185 boost in cell subtype annotation accuracy (**Fig. 3D**). Furthermore, we integrated SpatioCell with
186 multiple deconvolution algorithms and compared the results to the original methods, showing that
187 SpatioCell consistently improves the accuracy of single-cell assignments **(Fig. 3E).**

188 Moreover, we divided simulated spots into capture regions and spot intervals to mimic the lack of
189 sequencing coverage, a common issue in many multicellular-resolution ST. SpatioCell's performance
190 on capture regions consistently surpassed the deconvolution-based method (**Fig. 3F**). Meanwhile, a key
191 advantage of integrating H&E features is that SpatioCell uses cell-type probability profiles from H&E
192 images (Module 1) to infer cell identities in spot intervals. This is achieved by comparing these profiles
193 with those of neighboring cells within the sequencing spots based on morphological similarity and then
194 transferring cell identity information from these spots to the intervals (**Fig. 3G, Fig. S5A**). To validate
195 SpatioCell's ability for cell type imputation at interval regions, we measured annotation accuracy across
196 entire tissue sections and within spot interval regions in four datasets. SpatioCell consistently
197 outperformed deconvolution-based methods in both regions (**Fig. 3H, Fig. S5B-C**). Notably, across all
198 four tissues, even with intervals, SpatioCell achieved higher overall annotation accuracy than the
199 deconvolution-based method applied to complete (gap-free) sections (**Fig. S5C**).

200 As shown in **Fig. 3I**, SpatioCell reconstructs the single-cell distribution in OVCA, effectively
201 revealing its complex TME. SpatioCell more accurately delineates tissue boundaries and tumor
202 structures (such as tubular-shaped tumor) compared to the deconvolution-based method, providing a
203 more structured tissue architecture and cellular spatial distribution. Notably, SpatioCell accurately
204 captured the cell-type distribution at tumor margins (as in **Fig. 3I** *regions I-II*) and identified tumor cells
205 surrounded by T and NK cells adjacent to a follicular cyst, a key feature in the OVCA

206 microenvironment[28] (**Fig. 3I**). In contrast, deconvolution results provided only a coarse estimate of cell-
207 type proportions, overlooking or blurring critical cellular compositions. Meanwhile, SpatioCell also
208 achieves accurate single-cell–level gene expression mapping based on the high-resolution annotation.
209 As illustrated in **Fig. 3I**, SpatioCell reconstructed a single-cell–resolved expression maps for *EPCAM*,
210 thereby enhancing the spatial resolution of transcriptomic profiles.

211 **SpatioCell Facilitates Fine-Resolution Mapping of Tissue Architecture Within the Tumor**
212 **Microenvironment from Multicellular ST**

213 In the TME, biologically significant regions, such as vascular and glandular structures, tumor
214 invasion fronts, and lymphocyte infiltration zones, are often located at the margins where multiple cell
215 types converge. However, these small and spatially intricate structures are difficult to detect or resolve
216 using existing deconvolution methods. To validate SpatioCell's applicability in revealing these subtle,
217 complex, and crucial structures, we applied it to real multicellular-resolution ST datasets from various
218 tumor samples with spatial spot sizes of 55 μm and 100 μm, representing common sequencing
219 resolutions. In a 100 μm-resolution ST dataset of TNBC, the larger spot size leads to higher cell density
220 per spot (up to 294 cells/spot; **Fig. 4A**), while also capturing more cells for sequencing. The intervals
221 between spots are narrower, with fewer cells in those regions (**Fig. 4B**). In contrast, the 55 μm-resolution
222 10x Visium ST of OVCA results in fewer cells per spot (**Fig. 4C**). Only 19.5% of the nuclei fall within
223 sequenced regions (**Fig. 4D**). Overall, these datasets, with varying resolutions, cell densities, and
224 interval coverage, provide a comprehensive framework to assess SpatioCell's adaptability in different
225 scenarios.

226 After processing these datasets with SpatioCell, we obtained a comprehensive single-cell spatial
227 atlas of the TNBC data that closely aligned with the manual annotation labeled by a pathologist (**Fig.**
228 **4E** and **Fig. S6**). In another TNBC slide, SpatioCell successfully identified a few lymphocyte-enriched
229 regions, with abundant T cells and B cells accumulating at the border between tumor and stroma (**Fig.**
230 **S7**). In the 55 μm OVCA data, SpatioCell accurately reconstructed conspicuous boundaries between
231 tumor cells and proliferative cells (**Fig. 4F**). However, the deconvolution-based annotation method

232 yielded highly disordered cell-type distributions that concealed these key structural features (**Fig. 4E**
233 and **F**, *i-ii*). These findings confirm that SpatioCell surpasses existing methods in tissue structure
234 reconstruction and spatial organization analysis.

235 Overall, in these datasets, deconvolution methods only exhibit a rough picture of cell type
236 composition, which limits their ability to perform spatial analysis (**Fig. 4E–F**, panels *I-II*). Prevalent
237 annotation methods, attempt to refine this by integrating segmentation-based cell positions and
238 randomly assigning identities, but they often fail to effectively reconstruct these intricate cellular
239 compositions because of low accuracy. SpatioCell offers superior resolution and accuracy, which
240 establishes the foundation for precise characterization of tumor microenvironment features.

241 **SpatioCell Enables Deconvolution Errors Correction via a Competitive Balanced Index**

242 Cell-type abundance estimation from deconvolution can be erroneous due to noises introduced by
243 molecular diffusion, sequencing errors, and uneven sampling in ST and single-cell reference data,
244 especially in complex regions. For example, a deconvolution result of a lung squamous cell carcinoma
245 (LUSC) ST data displays widespread tumor cell distribution, even in stromal regions, whereas the H&E
246 image reveals distinct stromal structures and necrotic zones (**Fig. 5A, Fig. S8**). This observation
247 highlights that, by integrating H&E-derived morphological features, SpatioCell not only achieves
248 precise single-cell mapping but also has the potential to refine cell annotations by correcting
249 deconvolution errors. SpatioCell corrects deconvolution-induced errors by calculating Competitive
250 Balance Index (CBI, as detailed in **Methods**), a parameter that quantifies the discrepancy between
251 deconvolution results and H&E-derived nuclear classification. Specifically, when CBI exceeds a preset
252 threshold, SpatioCell progressively refines the deconvolution result using H&E-derived classification
253 probabilities until the discrepancy falls below the threshold.

254 To evaluate the impact of CBI, we introduced controlled errors into the simulated multicellular-
255 resolution ST dataset (described in the previous section) and compared cell composition accuracy before
256 and after CBI-based correction, using mean absolute error (MAE) and mean squared error (MSE).
257 Results show that CBI significantly improves annotation accuracy under various perturbation levels in

258 BRCA and OVCA, with the optimal performance observed at CBI threshold of 0.5 and 10, respectively
259 (**Fig. 5B** to **C**). The improvement observed in OVCA was slightly lower than that in BRCA, likely
260 owing to the lower resolution of its H&E image, half that of BRCA. Our findings suggest that CBI
261 possesses a strong potential for correcting deconvolution errors and that the CBI threshold should be
262 adapted according to the quality of the H&E image. When image quality is high and nuclear
263 classification is reliable, a smaller CBI threshold allows for more drastic correction of deconvolution
264 errors. Conversely, for lower-resolution images, a larger CBI threshold helps preserve annotation
265 stability. In practice, we recommend starting with a higher threshold (e.g., 10).

266 After applying CBI to the LUSC data above, we observed a notable improvement in the annotation
267 and cell composition of stromal regions embedded within the tumor, which were not clearly delineated
268 in the original deconvolution results (**Fig. 5D**). By correcting deconvolution errors, SpatioCell
269 reconstructs key TME structures, such as small vessels, well-defined tumor borders, and immune
270 infiltrations, closely matching the pathologist's manual labels (**Fig. 5E**).

271 Overall, by leveraging CBI, SpatioCell can balance nuclear classification from H&E images with
272 deconvolution-derived composition estimates, allowing for adaptive correction of deconvolution results
273 and revealing key structural features of the TME, thereby enhancing our understanding of TME
274 dynamics. As a result of this correction, single-cell–level cell-type assignments become more accurate,
275 and reliable single-cell–resolved gene expression profiles can be obtained.

276 **SpatioCell Uncovers Associations between Spatial Heterogeneity and Functional States of CAFs**
277 **in TNBC**

278 CAFs, as key regulators in the TME, contribute to tumor growth, immune regulation, and
279 metastasis, and show marked heterogeneity in function and spatial distribution[29]. Recent studies have
280 shown that the composition of the cellular neighborhood can influence cell behavior and molecular
281 characteristics through intercellular interactions[30, 31]. However, the associations between CAFs' spatial
282 neighborhoods and their functional states remain poorly understood, yet deciphering these relationships
283 is crucial for understanding the TME and advancing therapeutic strategies. To address this challenge,

284 SpatioCell provides single-cell–resolved maps of cell types and gene expression, allowing precise
285 analysis of intercellular distances, spatial neighborhoods, and gene expression patterns.

286 We applied SpatioCell to annotate CAFs in 83 TNBC ST samples[32] (spot resolution: 100 μm) and
287 stratified them into five spatial neighborhood subtypes based on the compositions of their six nearest
288 neighboring cells (**Fig. 6A**, detailed in **Methods**). Each cluster exhibited distinct neighborhood features,
289 among which Cluster 0 and Cluster 1 showed particularly notable characteristics. Cluster 0 was
290 positioned significantly closer to tumor cells (average ~22 μm) than other CAF clusters (~35–40 μm;
291 **Fig. 6B, Fig. S9**), suggesting enhanced capacity to influence local tumor behavior, such as promoting
292 invasion and matrix remodeling[33]. In contrast, Cluster 1 was found nearer to B and T cells (averaging
293 ~21 μm and ~41 μm, respectively) than other CAF subtypes (~43–45 μm to B cells and ~55–58 μm to
294 T cells; **Fig. 6C**), suggesting a greater likelihood of participating in immune modulation through
295 chemokine signaling. Moreover, Cluster 1 exhibited greater spatial separation from other CAF clusters
296 (**Fig. 6D**), suggesting its localization within a distinct microenvironment. Collectively, these spatial
297 features support functional divergence among CAF subtypes: Cluster 0's proximity to tumor cells may
298 enable tumor-supportive behavior such as invasion and matrix remodeling, whereas Cluster 1's closer
299 spatial association with lymphocytes points to a potential role in coordinating localized immune
300 responses.

301 Consistent with these spatial-functional distinctions, single-cell atlas statistics showed that the
302 prevalence of each CAF subtype varied across samples, with Cluster 0 being the most abundant and
303 Cluster 1 the least (**Fig. 6E**). Survival analysis further revealed that patients with a higher proportion of
304 CAFs from Cluster 0 had significantly shorter overall survival, whereas those with a higher proportion
305 of Cluster 1 exhibited longer survival durations (**Fig. 6F, Fig. S10-11**). These findings further emphasize
306 the critical role of CAF spatial heterogeneity in modulating CAF function, thereby affecting tumor
307 progression and patient outcomes.

308 To investigate the intercellular communication between CAF subtypes and their neighboring cells,
309 we classified ST spots according to the dominant CAF subtype within each spot and examined ligand–

310 receptor interactions using CellChat[34]. Cluster 0 showed enrichment of pro-invasive and
311 immunomodulatory interactions, such as FN1-CD44, APP-CD74, PPIA-BSG, and JAG–NOTCH pairs.
312 These interactions drive ECM remodeling, barrier breakdown, immune suppression, and EMT,
313 collectively shaping a tumor-promoting and immune-evasive microenvironment[35, 36, 37, 38]. In contrast,
314 Cluster 1 displayed characteristics of immune-activated CAFs, with strong ligand–receptor signaling
315 activity related to antigen presentation, leukocyte adhesion, and also innate immunity, including HLA-
316 CD4 family, ICAM1/2–ITGAL/ITGB2, and C3–ITGAX/ITGB2 interactions[39, 40, 41] **(Fig. 6G)**.

317 Building on the precise single-cell annotation, SpatioCell decomposed the spot-level expression
318 and mapped gene expression to each cell location. The gene expression profiles of Cluster 0 and Cluster
319 1 CAFs were further analyzed at single-cell resolution to explore their distinct roles within the TME.
320 Cluster 0 CAFs showed high expression of immunosuppressive genes, whereas Cluster 1 CAFs were
321 enriched for immune-activating genes (**Fig. S12A**). Noteworthily, Cluster 1 highly expressed CCL19,
322 while its neighboring B and T cells expressed high level of CCR7, a chemokine-receptor pair associated
323 with TLSs and known to guide lymphocyte trafficking via the CCL19/CCL21–CCR7 axis[42] (**Fig. 6H,**
324 **Fig. S12B**). This expression profile aligns well with the spatial proximity to B and T cells (~21 μm and
325 ~41 μm, respectively) identified by SpatioCell, reinforcing the functional and spatial integration of this
326 immune-modulating CAF subtype. SpatioCell-based annotation and inferred cell expression of CCL19
327 and CCR7 also revealed that Cluster 1 CAFs are frequently positioned near TLS regions (**Fig. 6I**).
328 Consistent with these findings, analysis of the TCGA BRCA cohort demonstrated that high CCL19
329 expression was associated with improved survival outcomes (**Fig. 6J**).

330 In summary, SpatioCell enables single-cell annotation and precise analysis of intercellular
331 distances, spatial neighborhoods, and gene expression patterns, revealing fine-grained spatial
332 heterogeneity of CAFs in TNBC. This analysis uncovers clinically relevant CAF subtypes and
333 underscores the key role of spatial organization in shaping their behavior and influencing tumor
334 progression.

335 **Discussion**

336 ST offers critical insights into gene expression within the tissue architecture, and the growing
337 accumulation of ST datasets presents valuable opportunities for spatial biology. However, the resolution
338 of most available datasets remains inherently constrained by the size of spatial capture spots. Existing
339 deconvolution methods[17 43,] have sought to infer cell compositions within spots, but these methods lack
340 the spatial precision required to determine the exact localization of individual cells, which makes it
341 difficult to recover detailed structural features and gene expression pattern within the tissue. SpatioCell
342 goes further by integrating imaging data-derived morphological features with transcriptomic
343 measurements to reconstruct single-cell spatial maps from multicellular-resolution ST data. Through
344 detailed analysis of cell distribution, expression and tissue architecture, it holds the potential to advance
345 research in developmental biology, tumor microenvironment studies, and disease pathology.

346 To achieve this, SpatioCell integrates a novel spatio-morphological learner, re-engineering SAM's
347 prompting and decoding mechanisms for robust and accurate nuclear segmentation and classification.
348 To evaluate its performance, we conducted tests across six public datasets. SpatioCell outperformed
349 other widely used nuclear segmentation and classification algorithms. Building on this foundation,
350 SpatioCell integrates H&E-derived cell morphology with transcriptomics data, enabling a more
351 biologically meaningful reconstruction of single-cell distributions. Detailly, SpatioCell generates a
352 probabilistic cell-type profile for each nucleus based on its morphology, rather than assigning a hard
353 label. This enables SpatioCell to model the integration as a constrained combinatorial optimization
354 problem, balancing conflicting information from image and sequence data and ensuring a more coherent
355 fusion. In this way, SpatioCell achieves high-resolution single-cell annotation and recovers fine-scale
356 structural details that deconvolution methods often miss. For example, in a 100 μm resolution TNBC
357 ST dataset, SpatioCell vividly restored a normal breast lobule structure heavily infiltrated by T and B
358 cells, alongside neighboring tumor cells. Similarly, in OVCA data, it accurately delineated the tumor-
359 stroma boundary, revealing a stroma-rich region densely populated with proliferative cells, subtle yet
360 biologically significant features otherwise obscured by existing annotation methods.

361 To validate the usefulness of SpatioCell for ST studies, we use its high-resolution spatial annotation
362 capability to dissect the spatial heterogeneity of CAFs in TNBC. The identification of distinct CAF
363 neighborhood subtypes with divergent transcriptional and functional profiles, along with their
364 characteristic intercellular distances to tumor cells and immune populations, underscores the complex
365 role of spatial context in shaping CAF behavior. Particularly, the contrasting pro-tumorigenic versus
366 immune-activating phenotypes of Cluster 0 and Cluster 1 highlight how CAFs can differently influence
367 tumor progression and immune response depending on their microenvironment. These insights not only
368 advance our understanding of CAF biology but also provide a framework to better integrate spatial
369 distribution with analyses of cellular functional heterogeneity.

370 SpatioCell marks a significant advancement in ST by integrating image and transcriptomic data to
371 achieve single-cell resolution annotations, enabling a deeper understanding of tissue architecture and its
372 connection to disease. However, its capabilities are still limited by the size of available training datasets
373 for segmentation and classification. Future directions include expanding the training data by annotating
374 more stained images and applying novel computational methods for more precise cell-type classification
375 on H&E images—as demonstrated in ROSIE[44]. Notably, SpatioCell's success with DAPI-stained
376 images highlights its potential applicability across diverse imaging modalities. With further training,
377 SpatioCell could be extended to subcellular-resolution ST platforms (such as Stereo-seq[10] and
378 Xenium[45]), where nuclear contours (e.g., DAPI-stained) are readily available, enabling more precise
379 annotations in subcellular and single-molecule resolution contexts. These advancements expand
380 SpatioCell's applicability to all spatial omics platforms and further drive the translation of spatial omics
381 into biomedical research.

382 Materials and Methods

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

384 The H&E processing module is designed to achieve precise nuclear segmentation and analysis on H&E-
385 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 (1/4, 1/8, 1/16 and 1/32).

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 multiple convolutional layers 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.

411 The decoder refines these features using MLPs and produces segmentation masks by integrating them
412 with upsampled image features. Full architectural details are provided in **SI Note 2**.

413 ***Prompter Training***

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

420 $$L_{prompt} = -\frac{1}{N}\sum_{x,y}\begin{cases}(1-\hat{Y}_{xy})^{\alpha}\, log\,(\hat{Y}_{xy}) \;\; if\; \hat{Y}_{xy} \;=\; 1 \\ (1-\hat{Y}_{xy})^{\beta}(\hat{Y}_{xy})^{\alpha}\, log\big(1-\hat{Y}_{xy}\big) \;\; otherwise\end{cases} \quad (1)$$

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

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

425 $$L_{nc} \;=\; L_{ce} \;+\; \lambda L_{dice} \quad (2)$$

426 where λ represents the hyperparameter that weighs the impact of the two losses, and in this study, λ =
427 3. Here, we use the mit_b3 weight pretrained on the ImageNet dataset[19]. Additional details are available
428 in **SI Note 2**.

429 ***SAM fine-tuning***

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

436 $$L_{sam} = L_{Dice} + \alpha L_{Focal} \quad (3)$$

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

438 ***Nuclear segmentation inference***

439 During inference, the Prompter generates a predictive heatmap of nuclear centroids along with a
440 semantic map indicating nuclear types. Pixels with heatmap values exceeding a threshold τ are selected
441 as positive prompts, and their corresponding locations in the semantic mask are used to assign cell type
442 labels. The selected prompts, along with the input image patches, are passed to SAM for instance
443 segmentation. Finally, overlapping predictions resulting from the sliding-window tiling strategy are
444 filtered using Matrix NMS. Additional details are available in **SI Note 2: *Inference and Postprocessing***.

445 **Module II of SpatioCell: Single-cell resolution annotation and gene expression mapping of ST**

446 ***Estimating cell type compositions within spots***

447 The cell annotation module is compatible with outputs from various spatial deconvolution tools
448 (e.g., Cell2location, Spotiphy, Tangram). In this study, Cell2location was employed using scRNA-seq
449 data derived from the same tissue type. Counts from ST data are preprocessed with mitochondrial and
450 uninformative genes were filtered out. Median number of segmented nuclei within spots was used in
451 'N_cells_per_location', and 'detection_alpha' was set to 200. All other parameters were kept at their
452 default settings. Deconvolution results are provided in **Fig. S6** to **8**. Details on implementation of other
453 deconvolution algorithms are available in **SI Note 3: *Deconvolution algorithms applied in benchmarks***.

454 ***Cell type annotation within spots***

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

460 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_i for each cell type i. The total number of cells of each type in the spot can be estimated as:

$$N_i = N_{spot} \times P_i, \forall i \in \{1,2,\dots T\} \tag{4}$$

where n_i represents the estimated number of cells of type i in the spot.

Step 2: DP for cell type assignment

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

SpatioCell performs the cell annotation through a DP algorithm that recursively searches all feasible label assignments (i.e. states) across nuclei to identify the globally optimal outcome. Let $P_{n,i}$ denote the probability that nucleus n belongs to cell type i, the DP algorithm proceeds as follows:

(1) State Space Initialization: Define the state space for the cell type assignment problem. The state s is represented as a vector of cell counts for each type, i.e. $s = (n_1, n_2, \dots, n_T), n_i \leq N_i$, where n_i is the number of cells of type i that have been assigned so far.

(2) Transition Between States: For each state s, the algorithm explores all possible directions of state transitions, $e^{(i)}$. For each transition, the goal is to maximize the sum of the probabilities of assigning a cell to a particular type while respecting the total number of cells for each type, as constrained by the deconvolution results.

The transition function is given by:

$$s_{next} = s + e^{(i)}, \text{where } i \in \{1, \dots, T\} \tag{5}$$

and for each transition, the objective is to maximize the total probability:

$$f(s_{next}) = \max\left(f(s_{next}), f(s) + P_{n,i}\right) \quad (6)$$

where $f(s_{next})$ represents the current maximum score for a given state and $P_{n,i}$ is the probability of the nth cell belonging to type i.

(3) Backtracking the Optimal Path: The DP starts from $s_0 = (0, 0, \ldots, 0)$, and iteratively updates states until reaching $s_{\text{final}} = (N_1, N_2, \ldots, N_T)$. Once all possible transitions have been explored, the optimal assignment is found by backtracking from the final state to the initial state, determining the optimal direction (i.e., cell type assignment) at each step.

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 (k is set to 5 in the study). Similarity is measured using KL divergence as follows:

$$similarity = KL_{(p||q)} = -\sum_x p(x) \log \frac{q(x)}{p(x)} \quad (7)$$

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

Gene expression mapping at single-cell level

Gene expression within individual spots was decomposed into cell-type–specific contributions following the strategy implemented in Spotiphy[17]. It assumed that for each spot s and gene g, the observed counts $x_{s,g}$ was modeled as a multinomial distribution across cell types t:

$$u_{s,g} | x_{s,g}, P, \Phi, \sim Multinomial\ (x_{s,g}, (\omega_{s,g,1}, \ldots, \omega_{s,g,T})) \quad (8)$$

where the allocation weight across cell types t is defined by:

$$\omega_{s,t,g} = \frac{\tilde{p}_{s,t} \cdot \phi_{t,g}}{\sum_{t'} \tilde{p}_{s,t'} \cdot \phi_{t',g}} \quad (9)$$

Here, $\tilde{p}_{s,t}$ denotes the proportion of cell type t in spot s estimated by SpatioCell, and $\phi_{t,g}$ represents the expression profile of gene g in cell type t obtained from scRNA-seq data. The expected expression assigned to each cell type is then given by:

$$\hat{u}_{s,t,g} = x_{s,g} \cdot \omega_{s,g,t} \tag{10}$$

The inferred cell-type–specific expression profiles were precisely assigned to individual cells according to the single-cell annotations derived from SpatioCell, thereby generating a high-resolution spatial map of cell identities and gene expression. For each cell at interval, gene expression was imputed by borrowing information from the spatially nearest neighbor cell whose morphological probabilistic profile exhibited the highest similarity.

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 nucleus (P_{images}), defined as:

$$CBI = \frac{P_{images} - P_{assigned}}{P_{assigned} + 10^{-4}} \tag{11}$$

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

533 We benchmarked the image processing module of SpatioCell with five algorithms including Hover-Net,
534 Cellpose3, StarDist, Mask R-CNN and CellProfiler, on six public datasets: Kumar, CPM 17, PanNuke,
535 CoNSeP, MoNuSAC, and CellBinDB. Nuclear segmentation performance was evaluated on all except
536 MoNuSAC dataset since StarDist was pretrained on it. Cell type classification was benchmarked on
537 PanNuke, CoNSeP, and MoNuSAC to compare SpatioCell, Hover-Net and Mask R-CNN.
538 For each dataset, we used its original training, validation, and testing splits. In PanNuke dataset, we
539 conducted a three-fold cross-validation on its predefined partitions to ensure robust evaluation. We used
540 AJI, DQ, SQ and PQ to comprehensively validate nuclear instance segmentation. AJI and SQ measure
541 the segmentation quality while DQ evaluates models' ability in nuclei detection (F1-score for detection).
542 PQ considered SQ and PQ together. For cell type classification, we used mPQ and F_C, defined as the
543 product of the detection F1-score and overall classification accuracy for correctly detected nuclei.
544 Segmentation performance on DAPI-stained images is tested using 10x Genomics dataset in CellBinDB.
545 Since SpatioCell, Hover-Net, and Mask R-CNN do not offer pre-trained weights for DAPI-stained
546 images, we trained these models from scratch using DAPI-stained images from stereo-seq dataset in
547 CellBinDB. In contrast, Cellpose3 and StarDist were evaluated using their respective pre-trained
548 weights, "cyto3" and "fluoro". Additional details in the ***compared methods***, ***image preprocessing***,
549 ***dataset*** and ***evaluation metrics*** **are available in SI Note 1**.

550 **Evaluation of annotation using Xenium data**

551 To simulate multicellular-resolution ST data, we obtained Xenium datasets from 10x Genomics website,
552 including BRCA, OVCA, PDAC, and LUCA samples. Each dataset provides H&E and DAPI images,
553 DAPI-based nuclear information (regarded as ground truth nuclei), affine transformation matrices and
554 cell annotations (except for LUCA, which we generated marker-based annotations using Seurat). H&E
555 images were aligned to DAPI by '*skimage*' according to affine transformation matrices and then
556 partitioned into 35, 55, and 100 μm square to simulate spots. Squares located at odd rows with even
557 columns and even rows with odd columns were designated as intervals. Segmented and ground truth
558 nuclei were assigned to a spot based on their centroid coordinates. Simulated deconvolution results were

generated by calculating the proportion of annotated cell types within each spot based on the provided annotations.

To evaluate annotation accuracy, we used a mutual nearest neighbor (MNN) strategy to align segmented and ground truth nuclei based on their centroids. Cell pairs identified as mutual nearest neighbors were considered aligned. Unmatched ground truth nuclei and uncertain labels were excluded from computing accuracy. Additional details in ***Xenium Human Lung Cancer Cell Type Annotation*** and ***Mapping H&E image segmented cells with Xenium annotation*** are available in **SI Note 3.**

Evaluation of CBI impact on deconvolution results using Xenium data

We used BRCA and OVCA datasets to evaluate robustness by introducing perturbations into the simulated deconvolution results within each spot. Specifically, 5%, 10%, 15%, 20%, and 30% of cell assignments were randomly perturbed per spot (rounded down, with at least one cell modified per spot). For each perturbation level, we applied CBI-based correction using thresholds of 0, 0.1, 0.5, 1, 10, 100, 1000, and 10000 (representing no correction), and compared the corrected compositions and the perturbed compositions against the ground truth. MAE and MSE were used as evaluation metrics, defined as follows:

$$AE_i = \sum_{k=1}^{K} \left| p_{i,k} - t_{i,k} \right| \tag{12}$$

$$\text{MAE} = \frac{1}{S} \sum_{i=1}^{S} AE_i \tag{13}$$

$$SE_i = \sum_{k=1}^{K} (p_{i,k} - t_{i,k})^2 \tag{14}$$

$$\text{MSE} = \frac{1}{S} \sum_{i=1}^{S} SE_i \tag{15}$$

where $p_{i,k}$ and $t_{i,k}$ denote the estimated and ground truth proportion of cell type k in spot i respectively.

CAF neighborhood clustering and spot classification

580 We analyzed ST data from 83 breast cancer patients with ER-, PR-, and HER2- status from the Wang
581 et al. cohort[32]. Single-cell annotation of the ST data was performed using SpatioCell. For each annotated
582 cell, the six nearest neighboring cells within a 20 μm radius were identified, and the cell-type
583 compositions were recorded to construct a local neighborhood profile.
584 Focusing on CAFs, we applied K-means clustering to six-cell neighborhood profiles and identified five
585 spatial subtypes based on the elbow plot.
586 For downstream transcriptomic analysis, we quantified the number and proportion of each CAF
587 neighborhood subtype within every ST spot. Each ST spot was assigned a dominant CAF subtype if
588 over 80% of its CAFs belonged to the same cluster. Only these subtype-dominant spots were retained,
589 while spots without a clear dominant population were excluded to avoid confounding effects from
590 cellular heterogeneity.

591 **Spatial proximity measurements**

592 To characterize the spatial proximity between CAF clusters and key cellular components of the tumor
593 microenvironment, two types of distances were calculated based on cell centroids using scikit-learn's
594 *'NearestNeighbors'*: (1) CAF-to-cell-type distances—for each CAF in a given cluster, we computed the
595 average distance to its six nearest neighboring B cells, T cells, or epithelial cells, respectively; and (2)
596 inter-cluster distances—to assess spatial relationships between CAF clusters, we calculated the average
597 minimum distance from cells in cluster i to their six nearest neighbors in cluster j. The final pairwise
598 distance matrix was obtained by averaging across samples.

599 **Inferring cell communication networks and analyzing cell neighbors**

600 Cell communication analysis was conducted using the R package *'CellChat'* (version 2.2.0) on ST data
601 stratified by dominant CAF subtypes. Intra-cluster ligand–receptor interactions were visualized using
602 the '*netVisual_chord_gene*' function, and only interactions with a communication probability greater
603 than 0.01 were included in the chord diagrams.

604 **Survival analysis**

605 The *'survival'* (version 3.8.3) and *'survminer'* (version 0.5.0) R packages were used to perform survival
606 analysis. For each CAF neighborhood subtype and for CCL19 expression (in TCGA-BRCA), patients
607 were divided into high and low groups based on median values. Kaplan–Meier survival curves were
608 generated using the *'survfit'* function, and differences in overall survival between groups were evaluated
609 using the log-rank test via the *'survdiff'* function. Visualization of survival curves was performed using
610 the *'ggsurvplot'* function.

611 **Statistical analysis**

612 The highly differential genes in each cluster of spatial transcriptomics data were identified using the
613 Python package *'scanpy'* and tested using Wilcoxon rank-sum test. Differences in overall survival in
614 both TNBC cohort and TCGA cohort were evaluated using the log-rank test. All tests were two sided,
615 and p values < 0.05 were considered statistically significant.

616 **Paper Writing**

617 To improve the clarity and fluency of the manuscript, we utilized OpenAI's ChatGPT-4o model for
618 English language polishing. The model assisted in refining grammar, style, and readability, while all
619 scientific content was reviewed and validated by the authors.

620 Data availability

621 All data used in this study were previously published and publicly available. CPM17 and Kumar datasets
622 are available from https://github.com/vqdang/hover_net, CoNSeP dataset and PanNuke dataset can be
623 downloaded from https://warwick.ac.uk/fac/cross_fac/tia/data/. MoNuSAC dataset is available from
624 https://monusac-2020.grand-challenge.org, CellBinDB dataset is available from
625 https://db.cngb.org/search/project/CNP0006370.
626 Xenium datasets of FFPE human breast cancer, FFPE human lung cancer, FFPE human ovarian cancer,
627 and FFPE human pancreatic cancer are downloaded from 10x Genomics
628 (https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast;
629 https://www.10xgenomics.com/datasets/ffpe-human-lung-cancer-data-with-human-immuno-

oncology-profiling-panel-and-custom-add-on-1-standard;
https://www.10xgenomics.com/datasets/xenium-prime-ffpe-human-ovarian-cancer;
https://www.10xgenomics.com/datasets/pancreatic-cancer-with-xenium-human-multi-tissue-and-cancer-panel-1-standard).

Visium datasets of human ovarian cancer and lung cancer are available on the 10x Genomics websites (https://www.10xgenomics.com/datasets/human-ovarian-cancer-11-mm-capture-area-ffpe-2-standard; https://www.10xgenomics.com/datasets/human-lung-cancer-ffpe-2-standard). The TNBC ST data and the associated brightfield H&E images can be downloaded via Zenodo: https://doi.org/10.5281/zenodo.8135721.

The human breast cancer scRNA-seq dataset is available from the Gene Expression Omnibus under accession number GSE176078 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE176078). The human ovarian cancer scRNA-seq dataset is available on Mendeley Data (https://doi.org/10.17632/rc47y6m9mp.1). The non-small cell lung cancer scRNA-seq dataset is available on Zenodo (https://zenodo.org/records/7227571).

Acknowledgments

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.

Funding: This work was supported by the National Key R&D Program of China 2024YFC3405600 and 2022YFA1104200, and National Natural Science Foundation of China 22474075 (J.S.).

Author Contributions: Conceptualization: J.S. Methodology: N.H., ZR.W., and J.S. Investigation: N.H., ZR.W., Y.Y., Z.Z., F.X., and ZY.W. Visualization: N.H., Y.Y., Y.Z., and J.W.

Competing interests: Authors declare that they have no competing interests.

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. Wan X*, et al.* Integrating spatial and single-cell transcriptomics data using deep generative models with SpatialScope. *Nat Commun* **14**, 7848 (2023).
17. Yang J*, et al.* Spotiphy enables single-cell spatial whole transcriptomics across an entire section. *Nat Methods*, (2025).
18. Kirillov A*, et al.* Segment Anything. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*) (2023).
19. 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).
20. 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).
21. Graham S*, et al.* Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal* **58**, 101563 (2019).
22. Stringer C, Pachitariu M. Cellpose3: one-click image restoration for improved

cellular segmentation. *Nat Methods* **22**, 592-599 (2025).
23. 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).
24. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. Preprint at https://ui.adsabs.harvard.edu/abs/2017arXiv170306870H (2017).
25. Carpenter AE*, et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7**, R100 (2006).
26. Gamper J*, et al.* PanNuke Dataset Extension, Insights and Baselines. Preprint at https://ui.adsabs.harvard.edu/abs/2020arXiv200310778G (2020).
27. 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).
28. 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).
29. Costa A*, et al.* Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. *Cancer Cell* **33**, 463-479 e410 (2018).
30. Chen H*, et al.* Integrative spatial analysis reveals tumor heterogeneity and immune colony niche related to clinical outcomes in small cell lung cancer. *Cancer Cell* **43**, 519-536 e515 (2025).
31. Pei G*, et al.* Spatial mapping of transcriptomic plasticity in metastatic pancreatic cancer. *Nature*, (2025).
32. Wang X*, et al.* Spatial transcriptomics reveals substantial heterogeneity in triple-negative breast cancer with potential clinical implications. *Nat Commun* **15**, 10232 (2024).
33. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
34. Jin S*, et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021).
35. Volker SE, Hedrick SE, Feeney YB, Clevenger CV. Cyclophilin A Function in Mammary Epithelium Impacts Jak2/Stat5 Signaling, Morphogenesis, Differentiation, and Tumorigenesis in the Mammary Gland. *Cancer Res* **78**, 3877-3887 (2018).
36. Tsang JYS*, et al.* Proteolytic cleavage of amyloid precursor protein by ADAM10 mediates proliferation and migration in breast cancer. *EBioMedicine* **38**, 89-99 (2018).
37. Briot A, Iruela-Arispe ML. Blockade of specific NOTCH ligands: a new promising approach in cancer therapy. *Cancer Discov* **5**, 112-114 (2015).
38. Chen C, Zhao S, Karnad A, Freeman JW. The biology and role of CD44 in cancer progression: therapeutic implications. *J Hematol Oncol* **11**, 64 (2018).
39. Song J*, et al.* Antigen-presenting cancer associated fibroblasts enhance antitumor immunity and predict immunotherapy response. *Nat Commun* **16**, 2175 (2025).
40. Figenschau SL*, et al.* ICAM1 expression is induced by proinflammatory cytokines and associated with TLS formation in aggressive breast cancer subtypes. *Sci Rep* **8**,

11720 (2018).
41. Cords L, *et al.* Cancer-associated fibroblast classification in single-cell and spatial proteomics data. *Nat Commun* **14**, 4294 (2023).
42. Grout JA, *et al.* Spatial Positioning and Matrix Programs of Cancer-Associated Fibroblasts Promote T-cell Exclusion in Human Lung Tumors. *Cancer Discov* **12**, 2606-2625 (2022).
43. Vahid MR, *et al.* High-resolution alignment of single-cell and spatial transcriptomes with CytoSPACE. *Nat Biotechnol* **41**, 1543-1548 (2023).
44. Wu E, *et al.* ROSIE: AI generation of multiplex immunofluorescence staining from histopathology images. *Nat Commun* **16**, 7633 (2025).
45. 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).

Figures and Tables

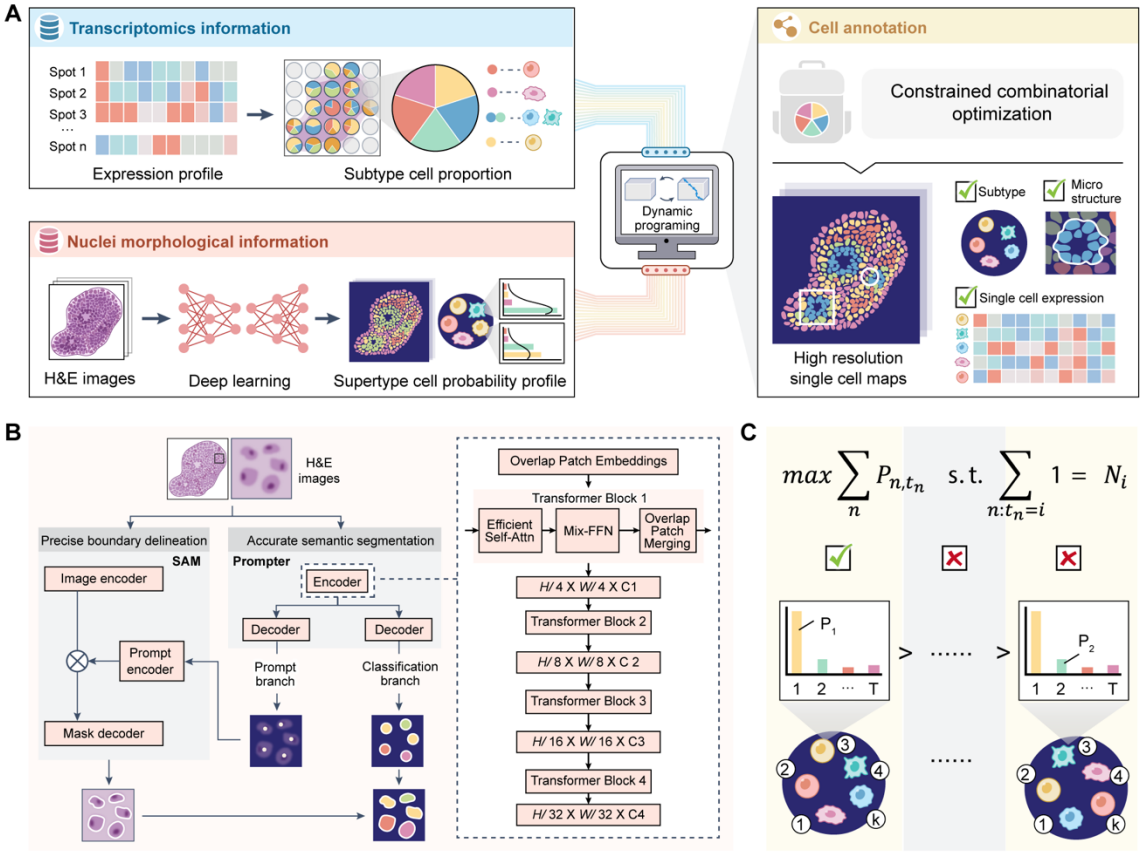


Figure 1. Overview of SpatioCell

761 **(A)**, SpatioCell computational framework. SpatioCell consists of two main modules: (1) a specialized
762 spatio-morphological learner, which is a deep learning module for histopathological image processing
763 that performs nuclear segmentation and analysis through detailed morphological feature extraction, and
764 (2) a morpho-transcriptomic spatial reconstruction module，which is a DP module for cell annotation
765 and gene expression mapping that integrates supertype and subtype annotations derived from
766 morphological and transcriptomic data to achieve high precision and resolution. Module 1 generates a
767 probabilistic profile for each mask based on its morphological features, allowing SpatioCell to model
768 the integration as a knapsack problem. **(B)** Deep learning architecture of module 1. Left: The framework
769 includes (1) a fine-tuned SAM-based segmentation model and (2) an adaptive prompting mechanism,
770 which allows SAM to receive accurate positive prompts for automated nuclear segmentation. The final
771 outputs include cell locations, masks, and cell identity (supertype) probability profiles derived from
772 morphological cues. Right: Detailed architecture of the Prompter. The Prompter's encoder uses
773 overlapping patch embedding to extract hierarchical features. This includes (1) Efficient Self-Attention
774 for long-range dependencies, (2) Mix-FFN for feature transformation, and (3) Overlap Patch Merging
775 for spatial dimension reduction. The output feature maps enable nuclear center detection and semantic
776 segmentation. **(C),** Schematic illustration of the optimization problem of module 2. The goal of cell type
777 assignment is to maximize the total type probability within each spot, where P_{n,t_n} denotes the
778 probability that cell n is classified as type t_n, subject to the constraint that the number of cells assigned
779 to each type i equals a predefined value N_i. The left panel shows an optimal assignment maximizing the
780 summed probabilities, while the right panels illustrate suboptimal assignments under the same
781 constraints.
782

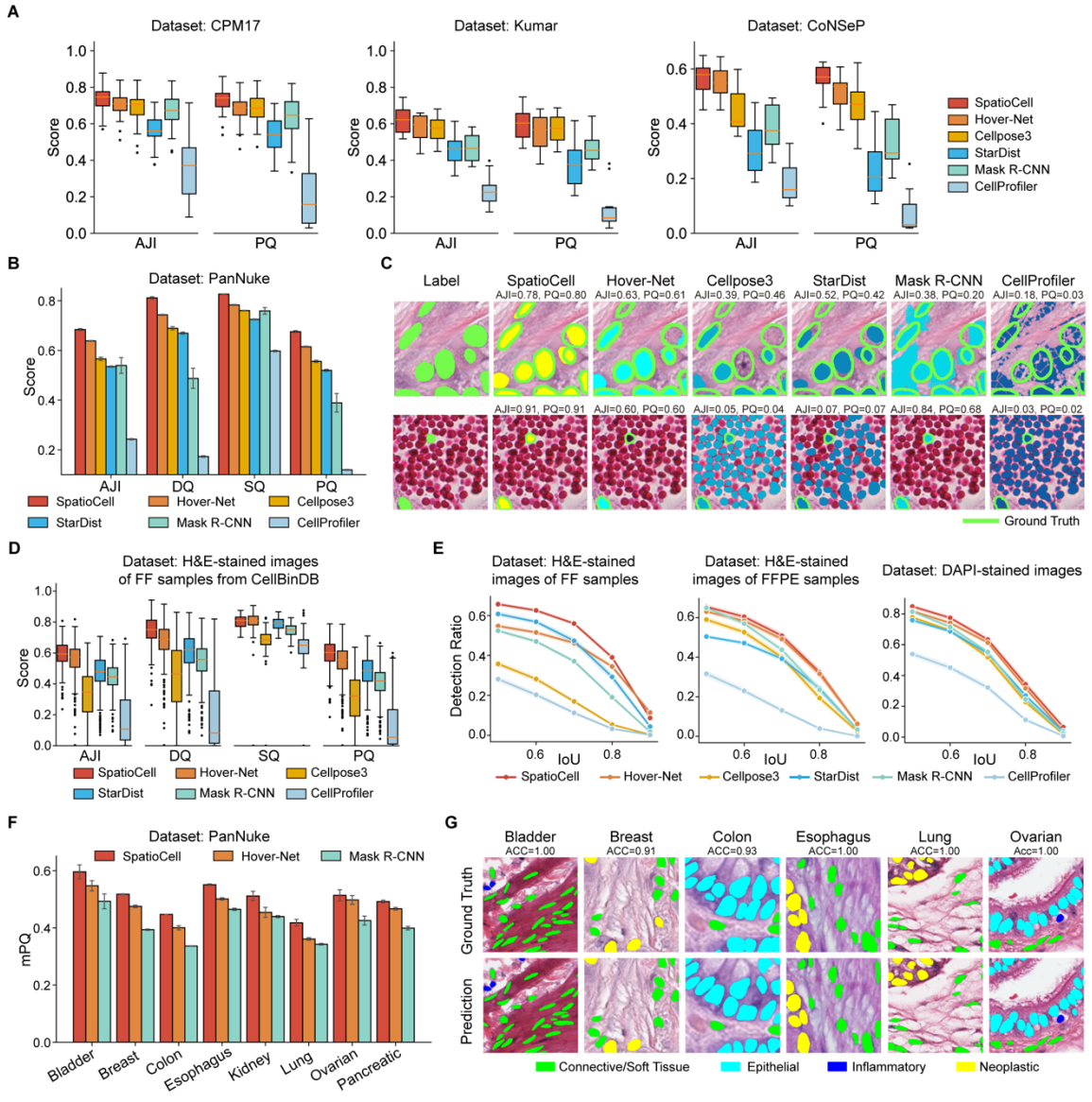
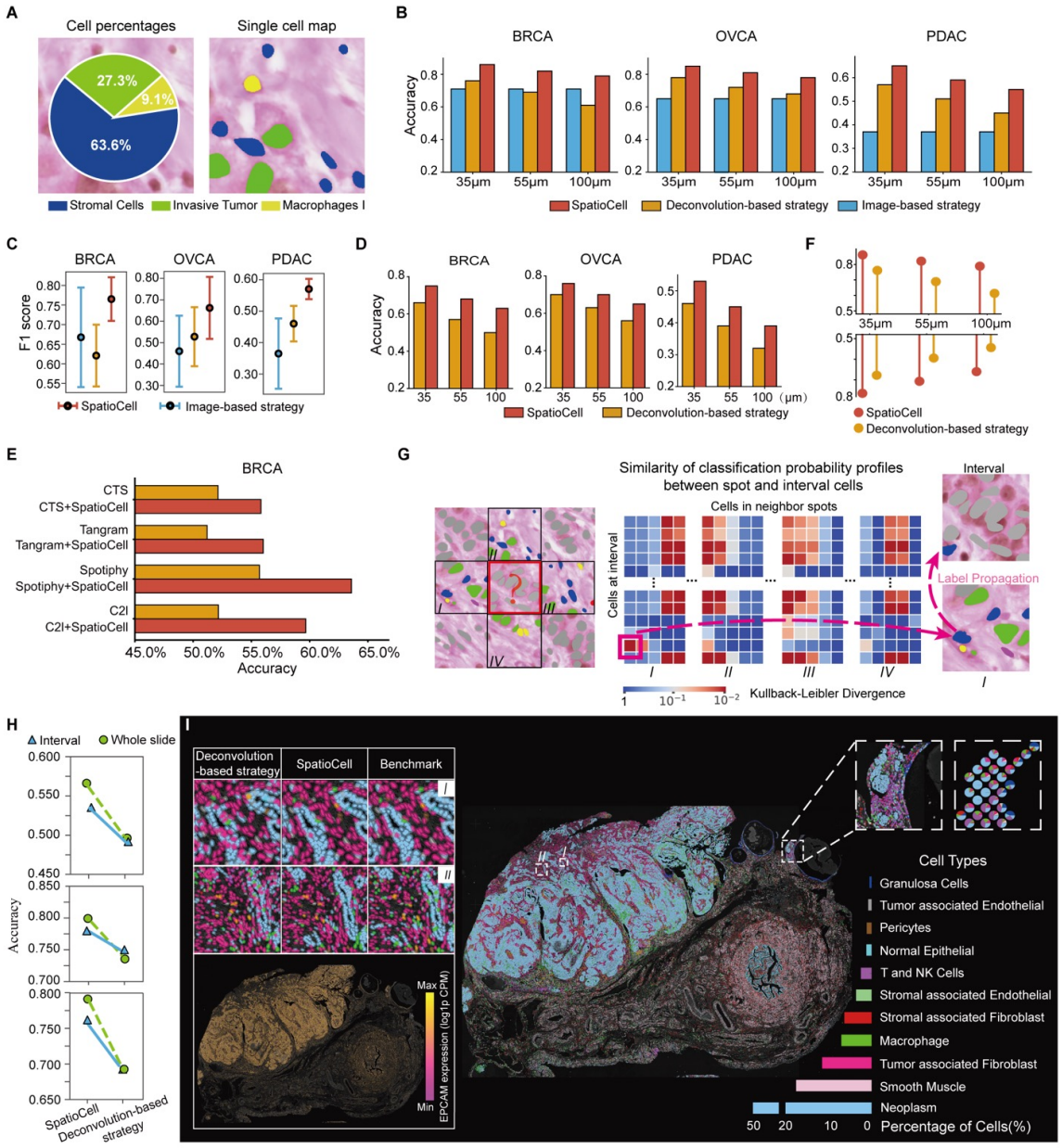


Fig. 2. Performance evaluation of SpatioCell in nuclear segmentation and analysis.

(A), Segmentation performance comparison of SpatioCell with other 5 methods on CPM17 (left), Kumar (middle) and CoNSeP (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)**, Nuclear segmentation performance of different methods on H&E-stained images of FF samples from the

CellBinDB dataset. **(E)**, Detection ratio of various methods on the CellBinDB dataset under different IoU thresholds. **(F)**, Distribution of mPQ scores of SpatioCell, Hover-Net, and Mask R-CNN on the PanNuke dataset through three-fold cross-validation. Nuclear type classification for SpatioCell is based on the maximum probability in the classification probability profile. **(G)**, Visual examples of nuclear classification by SpatioCell on the PanNuke dataset.



797 **Figure 3. Benchmark and visualization of SpatioCell cell annotation using simulated**
798 **multicellular-resolution ST data.**
799 **(A)**, Schematic of SpatioCell's cell annotation, which transfers deconvolution-inferred cell type
800 composition into a single-cell spatial map. **(B)**, Comparison of supertype cell annotation accuracy
801 among SpatioCell, deconvolution-based annotation, and H&E classification on simulated datasets
802 (including BRCA, OV, and PDAC tissues) at 35 μm, 55 μm, and 100 μm spot resolutions. **(C)**,
803 Comparison of supertype annotation accuracy (measured by F1-score) among SpatioCell,
804 deconvolution-based methods and H&E classification across three simulated datasets, with circles
805 representing the median and error bars indicating the interquartile range (IQR). Data from the 55 μm
806 spot size is shown. **(D)**, Comparison of subtype annotation accuracy between SpatioCell and
807 deconvolution-based methods on simulated datasets at 35 μm, 55 μm, and 100 μm spot sizes, using
808 ground-truth cell-type compositions as constraints. (**E**), Comparison of within-spot subtype annotation
809 accuracy for deconvolution algorithms alone vs. deconvolution + SpatioCell at 55 μm spot size (CTS,
810 CytoSpace; C2l, Cell2location). **(F)**, Comparison of supertype (top) and subtype (bottom) annotation
811 accuracy within spots across three datasets and various spot sizes between SpatioCell and
812 deconvolution-based methods when interval regions are present. **(G)**, Schematic of SpatioCell's
813 imputation strategy for spot interval regions. Gray cells represent query cells, and black-framed squares
814 indicate annotated neighboring spots. The heatmap visualizes the KL divergence between query and
815 annotated cells based on probability profiles from module 1, with the darkest red square indicating the
816 lowest divergence. **(H)**, Comparison of supertype annotation accuracy across three datasets at a 55 μm
817 spot size between SpatioCell and deconvolution-based methods, showing accuracy within spot interval
818 regions (blue) and across the entire slide (green). **(I)**, Single-cell spatial map of a simulated OVCA
819 dataset with interval region imputation using SpatioCell. The left panel compares deconvolution-based,
820 SpatioCell, and benchmark annotations (top) and illustrates the inferred EPCAM expression (log1p
821 CPM) distribution generated by SpatioCell (bottom). 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.

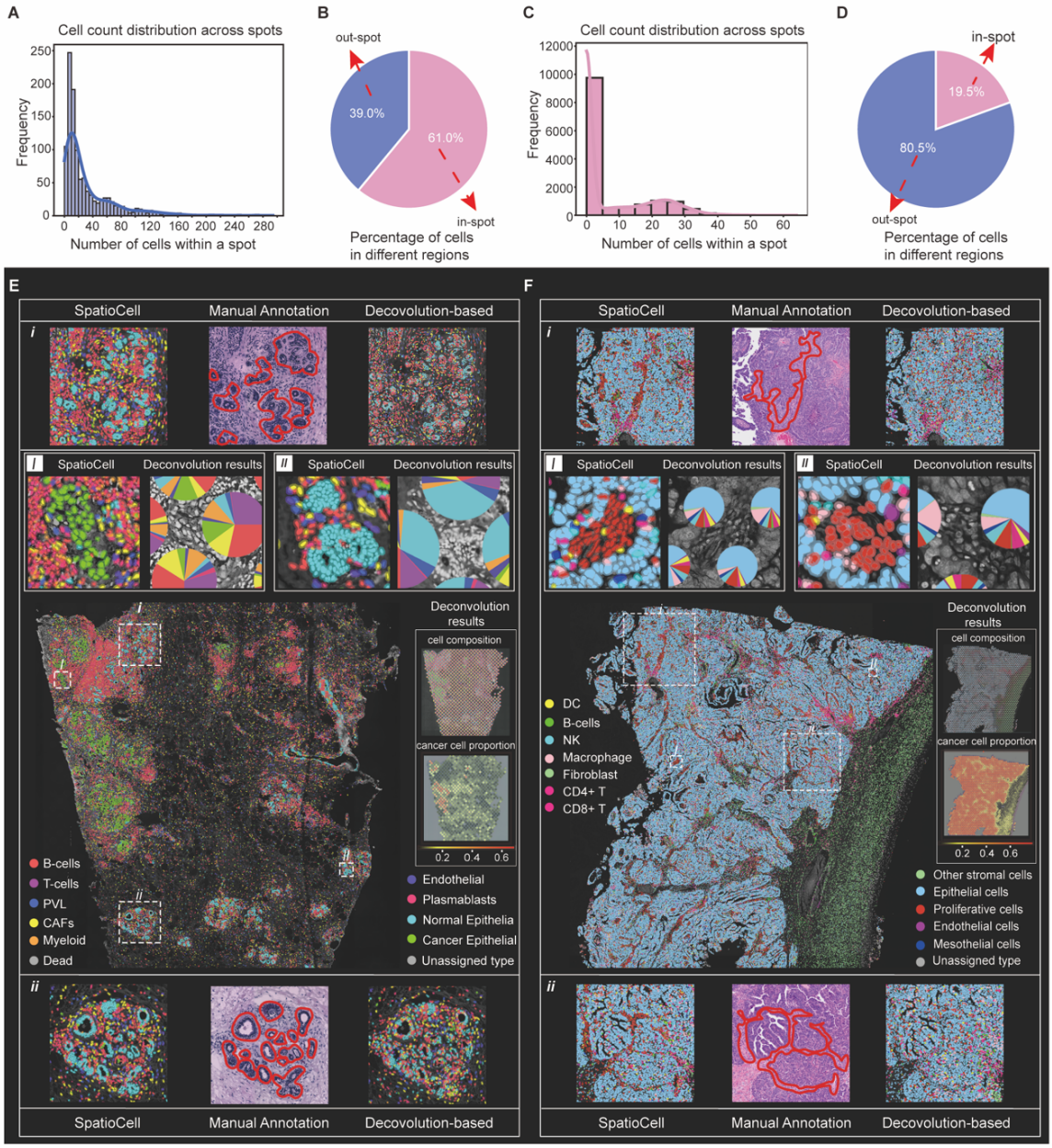


Figure 4. SpatioCell identifies characteristic tissue microenvironment in multicellular-resolution STs.

(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

829 between spots and inter-spot regions in the TNBC sample. **(C)**, Distribution of cell counts in captured
830 regions of an ovarian cancer (OVCA) sample sequenced with 55 μm spot size. **(D)**, Proportion of total
831 cell counts between captured and inter-spot regions in the OVCA sample. **(E)**, Single-cell spatial map
832 of the TNBC sample resulted from SpatioCell, with the right panel showing cell composition and cancer
833 cell proportion from deconvolution. (I, II) Comparison of cellular annotation by SpatioCell and
834 deconvolution results, with SpatioCell significantly improving resolution. (i-ii) Comparison of cellular
835 annotation by SpatioCell, manual annotation on H&E, and deconvolution-based methods. **(F)**, Single-
836 cell spatial map of the OVCA sample resulted from SpatioCell, with details similar to (E).

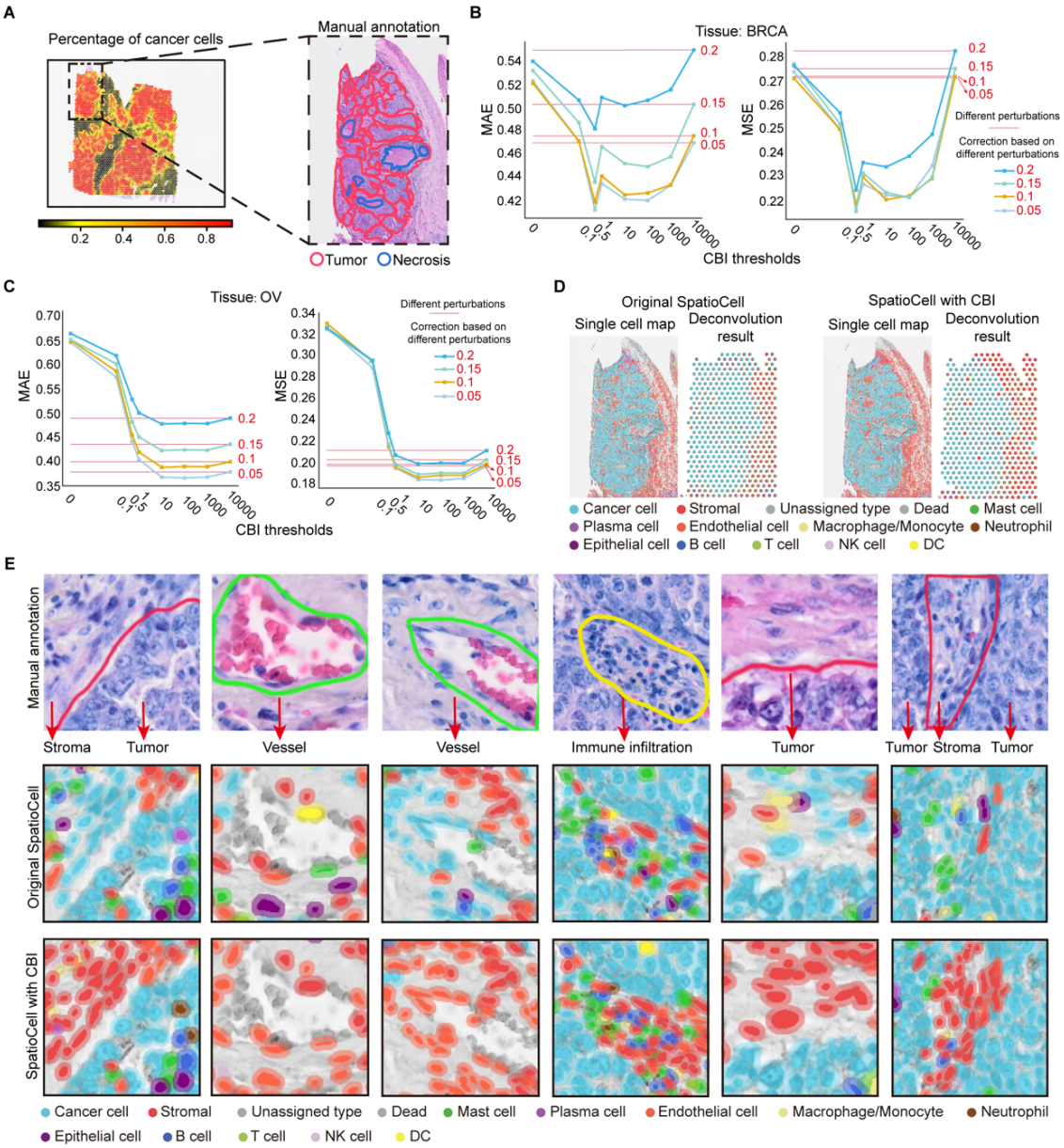


Figure 5. Validation and visualization of SpatioCell with Competitive Balance Index (CBI). **(A)**, Cancer cell proportion estimated by Cell2location in an LUSC sample from 10x Visium (left), with a zoomed-in H&E image (right) highlighting tumor (red) and necrosis (blue) areas as annotated by a pathologist. **(B)**, Distribution of mean absolute error (MAE, left) and mean squared error (MSE, right) of corrected deconvolution results for BRCA 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

844 indicate the corresponding MAE and MSE values without CBI for each perturbation level. **(C)**,
845 Distribution of MAE (left) and MSE (right) of corrected deconvolution results for OVCA ST data across
846 different CBI thresholds, with details similar to (B). **(D)**, Comparison of cell annotation results and
847 deconvolution before and after correction based on CBI. **(E)**, Comparison of cell annotations in 10x
848 Visium-sequenced LUSC ST data: Top row shows manual H&E annotations, second row shows original
849 SpatioCell annotations, and third row shows SpatioCell annotations refined with CBI. The refined
850 annotations align more closely with true tissue structures.

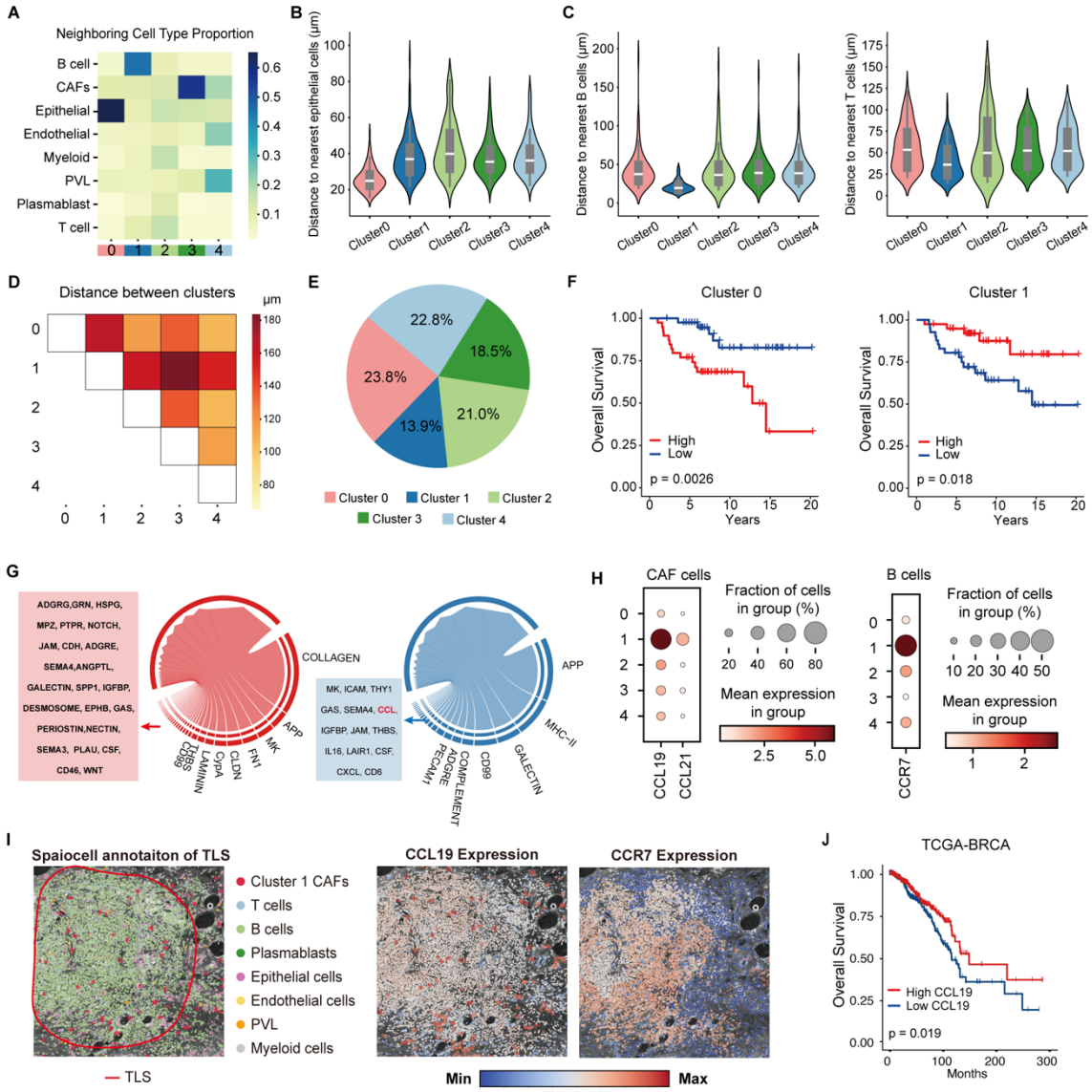


Figure 6. SpatioCell uncovers the spatial heterogeneity and functional states of CAFs in TNBC. **(A)**, Average neighboring cell type composition for each CAF cluster, calculated across all TNBC samples. **(B)**, Comparisons of the average distances from CAFs in each cluster to their six nearest epithelial cells across samples. Distances are measured in microns. **(C)**, Comparisons of the average distances from CAFs in each cluster to their six nearest B cells (left) and T cells (right) across samples respectively. Distances are measured in microns. **(D)**, Heatmap of pairwise spatial distances between CAF clusters. Each value represents the average distance from CAFs in one cluster to their six nearest

859 CAFs in another cluster. **(E)**, Total proportion of CAF clusters across 83 ST samples in the TNBC
860 dataset. **(F)**, Kaplan–Meier survival curves for patients stratified by high (red) vs. low (blue) abundance
861 of CAF Cluster 0 (left) and Cluster 1 (right). High and low groups were defined based on the median
862 proportion of each cluster. P-values were calculated using the log-rank test. **(G)**, Ligand–receptor
863 interaction networks between cells in spots dominated by Cluster 0 (left) and Cluster 1 (right). Band
864 width indicates interaction strength. **(H)**, Inferred cell expression of CCL19–CCR7 axis components
865 across CAF subtypes (CCL19 and CCL21, left) and neighboring B cells (CCR7, right), with dot size
866 indicating the proportion of expressing spots and color intensity representing mean expression per
867 cluster. **(I)**, Single-cell spatial annotation and gene expression mapping of a tertiary lymphoid structure
868 (TLS) region in a TNBC sample resulted from SpatioCell. Left: Spatial distribution of Cluster 1 CAFs
869 and other cell types within and around the TLS. Middle: Spatial distribution of CCL19 expression in
870 the same region. Right: Spatial distribution of CCR7 expression in the same region. **(J)**, Kaplan–Meier
871 survival curve for TCGA-BRCA patients stratified by high (red) vs. low (blue) CCL19 expression. High
872 and low groups were stratified based on the median expression level of CCL19. P-value was calculated
873 using the log-rank test.