CANCER

# SpaTopic: A statistical learning framework for exploring tumor spatial architecture from spatially resolved transcriptomic data

**Yuelei Zhang[1]†, Bianjiong Yu[1]†, Wenxuan Ming[1]†, Xiaolong Zhou[1], Jin Wang[1], Dijun Chen[1,2,3]***



**Tumor tissues exhibit a complex spatial architecture within the tumor microenvironment (TME). Spatially resolved transcriptomics (SRT) is promising for unveiling the spatial structures of the TME at both cellular and molecular levels, but identifying pathology-relevant spatial domains remains challenging. Here, we introduce SpaTopic, a statistical learning framework that harmonizes spot clustering and cell-type deconvolution by integrating single-cell transcriptomics and SRT data. Through topic modeling, SpaTopic stratifies the TME into spatial domains with coherent cellular organization, facilitating refined annotation of the spatial architecture with improved performance. We assess SpaTopic across various tumor types and show accurate prediction of tertiary lymphoid structures and tumor boundaries. Moreover, marker genes derived from SpaTopic are transferrable and can be applied to mark spatial domains in other datasets. In addition, SpaTopic enables quantitative comparison and functional characterization of spatial domains across SRT datasets. Overall, SpaTopic presents an innovative analytical framework for exploring, comparing, and interpreting tumor SRT data.**

## INTRODUCTION

Cells are the fundamental units of life, performing essential functions necessary for the survival of multicellular organisms (*1*). The spatial organization of cells within tissues is crucial for forming higher-order functional units (*2*). For instance, in the immune system, various immune cells interact with each other to coordinate pathogen responses (*3*). Similarly, in the nervous system, different types of neurons communicate to transmit signals and process information. In cancer, uncontrolled cell growth disrupts normal cell coordination, leading to the development of dysfunctional structures within the tumor microenvironment (TME). These structures interfere with physiological processes such as metabolism, circulation, and immune response. Therefore, understanding cellular ecosystems and their interactions within complex tissues is vital for comprehending multicellular organism biology and developing new treatments for diseases (*4*).

The advent of spatial transcriptomics technologies has revolutionized our understanding of cell organization within complex tissues by providing comprehensive information on the composition and spatial distribution of cell types (*5*, *6*). These methods involve sequencing the transcriptome on tissue sections and correlating transcriptomic data with spatial locations, offering insights into cellular interactions and spatial arrangements (*7*).

In recent years, notable progress has been made in the field of spatial transcriptomic analysis (*7*). Traditional spatial transcriptomics clustering methods involve clustering spots or cells based on gene expression patterns or on proximity (*8*–*10*), providing insight into molecularly distinct subgroups within tissues. However, such algorithms often overlook the functional correlations and interactions between cell types in the tissue microenvironment, failing to identify the cell populations involved in key biological processes in a specific region. To address these limitations, spatial transcriptomics deconvolution methods (*11*–*13*) have been introduced, aiming to infer the relative abundance of cell types in each discrete region to reconstruct the overall expression pattern of the tissue. Although deconvolution methods provide some structural tissue information, they often ignore the interactions and the overall organization between cell types, which are crucial for understanding functional structures in complex tissues.

In spatial genomics, a spatial domain comprises various cell types, similar to how a document consists of different words. Each spatial domain represents a unique functional region with a specific cell-type composition, much like how a document may encompass multiple main topics (fig. S1). To gain a comprehensive understanding of the impact of cellular composition and spatial arrangement on functional structures within complex tissues, we present SpaTopic—an innovative statistical learning approach based on topic modeling (*14*), which is commonly used in natural language processing. By integrating single-cell and spatial transcriptomic data, SpaTopic unveils the spatial functional units present in complex tissues, including the TME. By incorporating the latent Dirichlet allocation (LDA) models, SpaTopic facilitates the stratification of complex tissues into anatomical and functional structures, showcasing coherent cellular organization. This integration enables a holistic comprehension of cell-type colocalization and their spatial distribution within the tissue microenvironment.

The application of SpaTopic on diverse published datasets showcases its remarkable precision in delineating spatial domains. In the context of pancreatic cancer, SpaTopic adeptly annotated the spatial structure domains, aligning closely with histologists' annotations of tissue regions. Moreover, SpaTopic proves effective in elucidating spatially specific tertiary lymphoid structures (TLSs) and defining TLS-related markers that accurately characterizes TLSs in primary hepatocellular carcinoma (HCC). Furthermore, it

[1]Department of Gastroenterology, Nanjing Drum Tower Hospital, National Resource Center for Mutant Mice, State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, China. [2]Central Laboratory of Stomatology, Nanjing Stomatological Hospital, Medical School of Nanjing University, Nanjing, China. [3]Chemistry and Biomedicine Innovation Center, Nanjing University, Nanjing, China.
*Corresponding author. Email: dijunchen@nju.edu.cn
†These authors contributed equally in this work.

accurately predicts tumor boundaries (TBs) across various cancers, unveiling a conserved pattern of elevated fibroblast localization at these boundaries. The SpaTopic framework, integrating single-cell and spatial transcriptomic data to reveal spatial domains with unique cell-type organization in tumor tissues, holds significant promise for advancing research in both spatial and cancer biology.

## RESULTS

### The SpaTopic workflow

We have developed SpaTopic to predict, annotate, and compare spatial domains using spatially resolved transcriptomic (SRT) data and matched single-cell RNA sequencing (scRNA-seq) data with cell-type annotations (Fig. 1). In brief, SpaTopic aims to infer spatial clusters of spots characterized by coherent gene expression and cell-type organization within SRT data. Each spatial domain is considered as a distinct spot cluster with discernible patterns that differentiate it from other clusters. Specifically, SpaTopic first uses a deconvolution method [e.g., CARD (*12*)] to infer the cell-type composition of each spot. Meanwhile, it uses unsupervised clustering methods [e.g., STAGATE (*10*)] to aggregate the spots into initial clusters based on SRT data (Fig. 1). Next, SpaTopic applies the Kolmogorov-Smirnov (KS) test to determine the cell type–specific score for each cluster (the matrix **S**), leveraging the results from the above deconvolution and clustering step (Fig. 1B; see Materials and Methods). In the third step, SpaTopic uses the LDA model (*14*) to decompose the **S** matrix into two probability distribution matrices: (i) the probability distribution of a spot cluster belonging to a topic (cluster-topic distribution **C2**) and (ii) the contribution of a topic within a cell type (topic–cell type distribution **C1**) (Fig. 1A). The

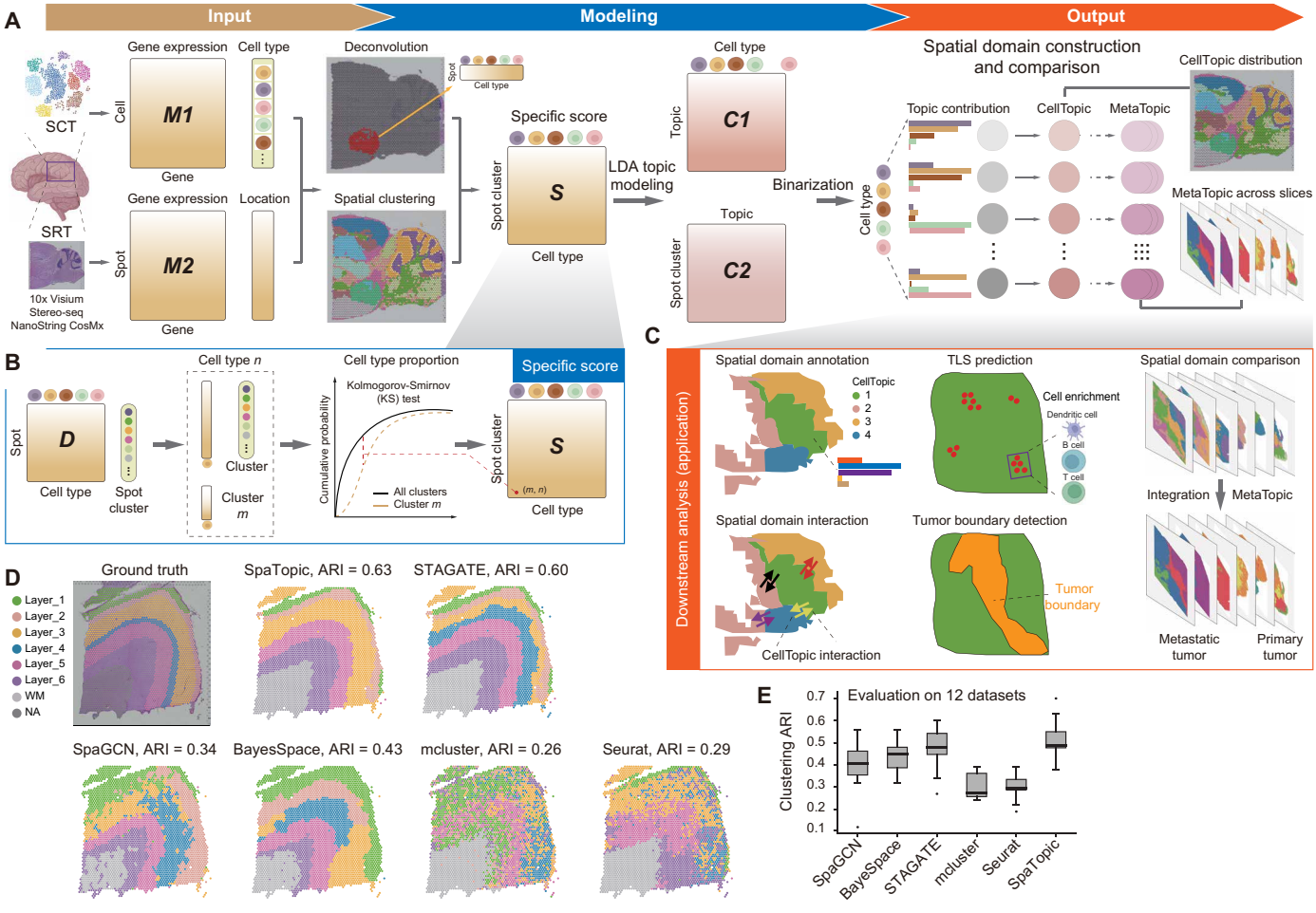


**Fig. 1. Schematic overview of SpaTopic.** (**A**) SpaTopic is designed to infer, annotate, and compare spatial domains based on single-cell transcriptomic (SCT) data with cell-type annotation and SRT data. In the modeling phase, spatial clustering and deconvolution analyses are performed based on the input SCT and SRT data. Then, a cell type–specific score matrix is calculated to reflect the significance of enrichment of cell types in different spatial clusters. Subsequently, topic modeling with LDA is applied to the specific score matrix to extract cell-type topics across different spot clusters. The cluster-topic matrix is binarized to refine the initial spot clusters into spatial domains, which show unique distribution of cell-type compositions (CellTopics). CellTopics from different SRT datasets can be quantitatively compare, in terms of "MetaTopic," based on their cell-type compositions. (**B**) Methodology used to calculate the cell type–specific score in different spot clusters. (**C**) Various downstream analyses of the SpaTopic output, including annotating spatial domains based on cell-type compositions, domain interactions, and spatial patterns of gene expression, and comparing spatial domains in terms of MetaTopics. SpaTopic can efficiently capture known spatial domains such as TBs and TLSs. (**D** and **E**) Performance of different methods for spatial domain calling. The evaluation is performed using 12 SRT datasets with manual annotation. *ARI*, adjusted Rand index.

inferred "cell-type topics" represent the predominant cell-type compositions in the given SRT data and thus implicate functional units. Last, the cluster-topic matrix is binarized, assigning each cluster to one or more specific topics (termed CellTopics). This step allows for the refinement of the initial spot clusters into spatial domains based on the learned cell-type topics derived from the SRT data. Therefore, SpaTopic facilitates the characterization of spatial domains by leveraging their cell-type topic contributions and quantitative comparison of spatial domains and identification of spatial gene expression programs across different SRT datasets (Fig. 1C).

To evaluate the performance of SpaTopic on identification of spatial domains, we applied it to 12 human dorsolateral prefrontal cortex (DLPFC) datasets (*15*) with manual annotation of tissue structures. The accuracy of spatial domain detection was evaluated using the adjusted Rand index (*ARI*), which measures the consistency between predicted spatial domains and manual layer assignments. SpaTopic outperformed all the evaluated clustering methods, including STAGATE (*10*), SpaGCN (*9*), and BayesSpace (*8*), mclust (*16*) and Louvain (Fig. 1, D and E, and fig. S2). With a focus on the tissue slice 151676, SpaTopic successfully identified spatial domains that demonstrated a higher level of agreement with the manually annotated tissue layers compared to other methods (Fig. 1D). While SpaTopic can use various clustering methods for initial cluster detection (with STAGATE showing better performance in particular), it consistently enhances the accuracy of spatial domain identification, regardless of the specific clustering method used (fig. S2). This suggests that the integration of cell composition information holds potential value in improving spatial domain annotation across various scenarios. Overall, the above evaluation highlights the effectiveness of SpaTopic in accurately capturing the underlying spatial organization within the analyzed data.

## Annotation of tumor-associated spatial domains in pancreatic ductal adenocarcinoma

To showcase the efficacy of SpaTopic in annotating spatial domains within complex tissues, we sought to leverage its power to decode the spatial cell organization within the TME for subsequent analyses. We first applied SpaTopic to a human pancreatic ductal adenocarcinoma (PDAC) dataset (denoted as PDAC-A) (*17*), which contains annotated SRT (with manually annotated tissue regions by histologists; Fig. 2A) and matched scRNA-seq data [comprising 15 annotated cell types (*17*)] for the same patient. The analysis identified 13 distinct topics in PDAC-A, with each topic exhibiting different scores across the spatial clusters (fig. S3).

To improve the interpretability of the score distribution derived from SpaTopic, we used binary classification of topic scores to demarcate the spatial structural domains, resulting in four CellTopics effectively covering specific regions (cancer, pancreatic, ductal, and stroma regions) within the PDAC sample (Fig. 2B). These CellTopics can be concurrently interpreted by considering cellular compositions (Fig. 2C), variably expressed genes (Fig. 2D and figs. S4 to S7), or enriched functional terms (Fig. 2E). In particular, CellTopic1 represented the stromal region with an accumulation of ductal cells, mast cells, macrophages, and monocytes, while CellTopic2 represented the ductal epithelial region, predominantly composed of ductal cells (Fig. 2C). CellTopic4 delineated the normal pancreatic tissue region, primarily enriched with acinar, endocrine, and dendritic cells (DCs). CellTopic3 characterized the cancer region, featuring a high enrichment of neoplastic cells (such as cancer clone A and clone B cells) and fibroblasts (Fig. 2C). The expression of genes such as *TM4SF1* from cancer clone A, *S100A4* from cancer clone B, and cancer-associated fibroblast (CAF) genes (*COL1A1* and *COL1A2*) showed activation in the CellTopic3 spatial domain (Fig. 2D). The highly expressed genes in this spatial domain are significantly enriched in stromal and immune-related processes and endodermal cell differentiation, suggesting the potential presence of cancer stem cells in CellTopic3. Consequently, CellTopic3 may contribute to initiation of tumorigenesis, development of chemoresistance, and facilitation of metastasis in PDAC (*18*). The aforementioned results align with the findings obtained through multimodal intersection analysis (*17*), providing further support for the validity of CellTopics.

SpaTopic identifies spatially coherent regions or domains (CellTopics) characterized by consistent cellular composition and gene expression profiles. These spatial domains may represent specific functional zones within the TME. Exploring the interactions among these domains would help to understand the overall dynamics of the TME. We therefore investigated communication patterns among or within CellTopics by assessing the signaling communication probability across spatial domains (see Materials and Methods). We observed that the stroma region (CellTopic1) barely interacted with other regions (Fig. 2F). Further analysis indicated that the integrin signaling pathways were largely blocked in interactions associated with CellTopic1 (Fig. 2G). PDAC stroma is known to function as a physical barrier within the PDAC TME (*19*). While the interaction frequencies between the tumor region (CellTopic3) and the nonmalignant duct epithelium (CellTopic2) or the normal pancreatic tissue region (CellTopic4) were predicted to be notably high, the physical isolation by the stroma region (Fig. 2A) suggests that these predicted interactions are unlikely to occur over long distances. Because juxtacrine and paracrine signals typically operate within a range of 0 to 200 μm (*20*, *21*), we limited the range of cell communication to 400 μm when analyzing communication patterns between different spatial domains. The interactions between CellTopic3 and CellTopic2/CellTopic4 do not exist when the spatial distance is taken into account (fig. S8). By mapping the spatial organization of cells, analyzing gene expression patterns, and exploring interactions between different tissue domains, SpaTopic uncovers specific cell types, spatial regions, and signaling pathways that may play crucial roles in disease progression.

## Precise prediction of TLSs in primary liver cancer

TLSs are distinctive immune microenvironments that develop in nonlymphoid tissues (termed ectopic lymphoid organs) in response to chronic inflammation (*22*). These structures encompass a diverse range of essential immune response cell types, such as B cells, T cells, DCs, and stromal cells (*23*). Recently, TLSs have been extensively studied and characterized in various cancer types using SRT (*22*). However, the specific composition of cell types within TLSs can vary depending on the tissue and the underlying inflammatory condition, presenting a considerable challenge for computationally predicting TLSs. We anticipated that using the cell-type composition of spatial domains predicted by SpaTopic would offer an unbiased method for detecting TLSs in tumor SRT data.

We used SpaTopic on a SRT dataset of primary liver cancer (PLC), including the subtypes of HCC, combined hepatocellular and cholangiocarcinoma (cHC), and intrahepatic cholangiocarcinoma (ICC) (*24*) in six slices (ICC-1L/HCC-1L/HCC-2L/HCC-3L/HCC-4L/cHC-1L) (Fig. 3A). The regions corresponding to TLSs in these slices were manually annotated by pathologists based on

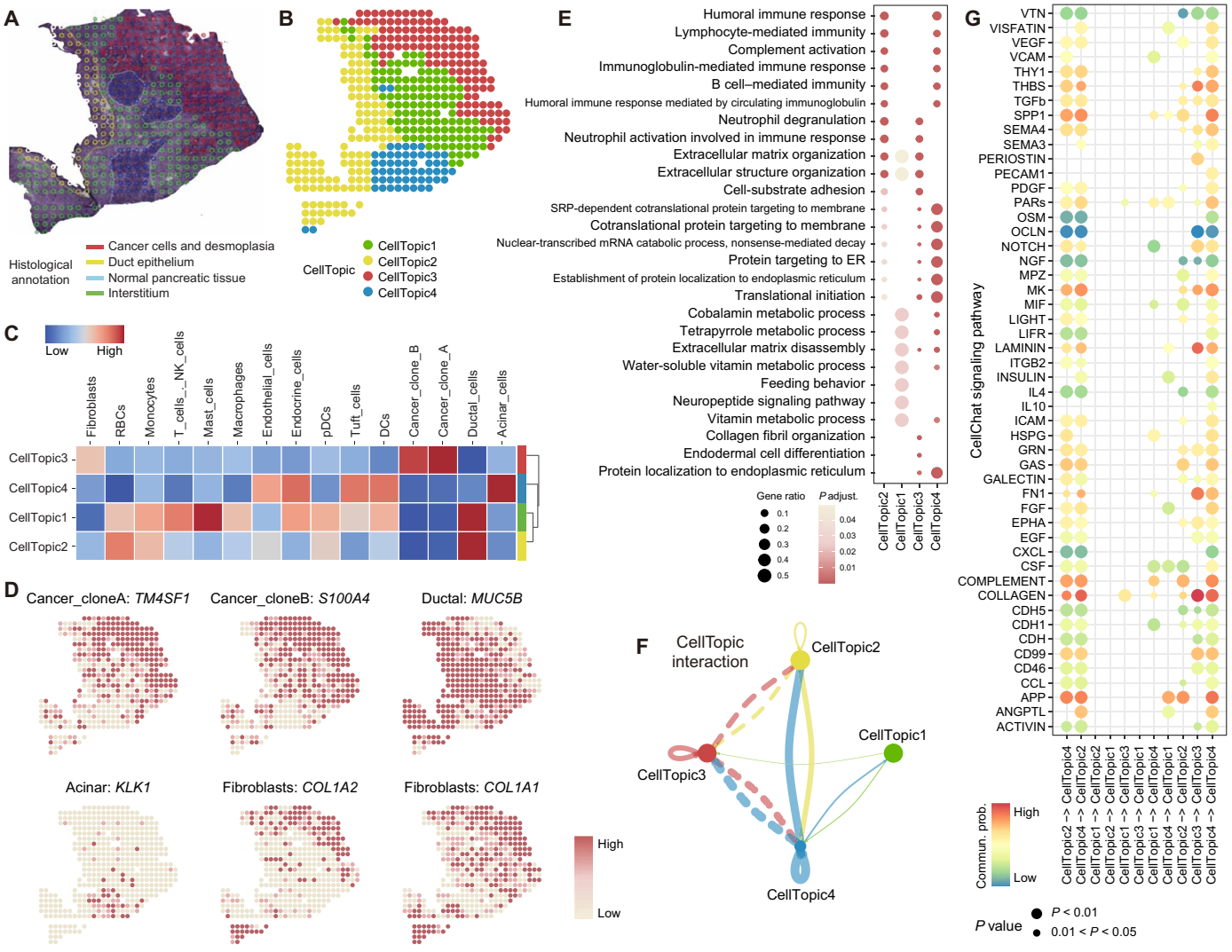


**Fig. 2. Analyzing the PDAC data by SpaTopic.** (**A**) Histologically annotated regions of the PDAC tissue from the original study (*17*). (**B**) Inferred CellTopics (spatial domains) of the PDAC data by SpaTopic. (**C**) Heatmap showing the cell-type compositions (normalized topic score) of different CellTopics. (**D**) Spatial map of representative cell type–specific marker genes. (**E**) GO enrichment analysis for each CellTopic based the maker genes (*P* value < 0.05 and avg_log2FC > 0). (**F**) Cell-cell interaction networks of different CellTopics. The thickness of the edge is proportion to the communication intensity (i.e., the number of signaling pathways). Dashed lines indicate physically isolated CellTopics; thus, the inferred interactions are unlikely to occur. (**G**) Dot plot showing the communication probability of CellChat signaling pathways among different CellTopics.

hematoxylin and eosin (H&E) staining (*24*). We then calculated the overlap between TLS regions annotated by experts and CellTopics predicted by SpaTopic using a hypergeometric test to evaluate the significance of enrichment. For instance, we observed a significant overlap between CellTopic11 in cHC-1L and the annotated TLS region of the SRT data (Fig. 3B). Expanding this analysis to all six slices allowed for the identification of TLS-related CellTopics in PLC (Fig. 3B and fig. S9). Consistently, these TLS-related CellTopics notably colocalized with B cells, T cells, and DCs, forming a network spatially intertwined with specialized fibroblasts (Fig. 3C). This composition closely resembled the typical structure of TLSs commonly found in tumors (*25*). Moreover, the highly expressed genes within these TLS-related CellTopics showed significant enrichment for various immune response pathways (Fig. 3D). Overall, the identified TLS-related CellTopics in PLC closely resembled the spatial domains of TLSs annotated by pathologists in terms of their spatial locations, cellular composition, and functional characterization.

Furthermore, we assessed the confidence of TLS-related CellTopics using "TLS-50" features, representing the top 50 genes that are highly and specifically expressed in manually annotated TLS regions of an HCC sample (*24*). CellTopics associated with TLSs exhibited a significantly higher module score (*26*) or the TLS-50 gene program compared to other regions (Fig. 3E). Because TLS-related CellTopics did not precisely overlap with manually annotated TLSs, we therefore termed these spatial domains associated with TLSs identified by SpaTopic as TLS-like domains. However, considering that the cellular composition and gene expression pattern within TLS-like domains are generally uniform, as indicated by unsupervised learning approaches, TLS-like domains may represent bona fide TLSs in tumors.

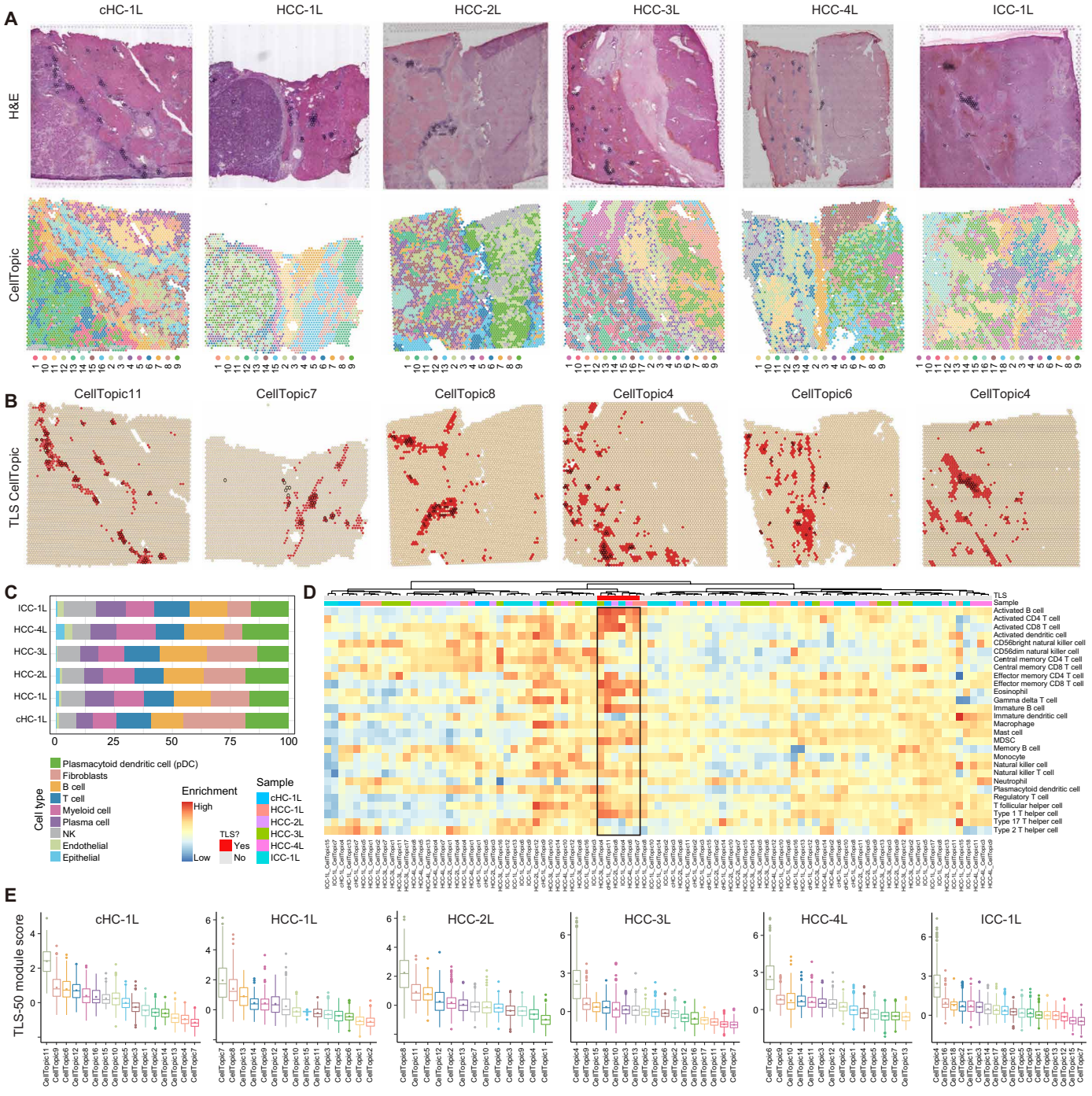


**Fig. 3. Accurate prediction of TLSs in PLC by SpaTopic.** (**A**) H&E staining of six PLC tissue sections (top) and the spatial distribution of CellTopics (or spatial domains) inferred by SpaTopic (bottom). Note that the number of CellTopic varies among different samples. (**B**) Representation of the TLS-associated CellTopic (in red domain). The expert-annotated locations of TLSs are highlighted by black circles. (**C**) Cell-type compositions (normalized topic score) of TLS-associated CellTopics. (**D**) Immuno-immersion analysis (*60*) of marker genes for different spatial domains. (**E**) Gene module score for TLS-50 marker genes (*24*) in different CellTopics. TLS-associated CellTopics consistently show a significantly higher module score than other CellTopics.

### SpaTopic-derived gene signatures improving the accuracy of TLS predictions

While the prognostic significance of TLSs in various cancer types is widely acknowledged, there exists a requirement for a standardized set of biomarkers to precisely define and to characterize TLSs (*22*). We sought to investigate if gene signatures derived from TLS-like domains could improve the prediction of TLSs using SRT data, as compared to existing signatures like TLS-50 (*24*) and cell type–specific signatures (*27*). To this end, we identified a shared set of genes ($n$ = 83) (table S1) that were consistently highly expressed in the TLS-like domains across the six SRT datasets of PLC in the above analysis (Fig. 4A and fig. S10). We observed a significant overlap between our gene set and the TLS-50 signatures, revealing 25 common genes (termed TLS-25), which encompass crucial players such as *CCL19* (*28*), *CCL21* (*29*), *TRBC2* (*24*), *LTB* (*30*), and *CXCL13* (*29*), known for their essential roles in the TLS formation and development (Fig. 4A) (*31*, *32*). Overall, the TLS-25 signatures demonstrate significant enrichment for functions related to TLSs, including neutrophil and DC chemotaxis as well as the activation of TLS cell composition (Fig. 4B and fig. S11). Moreover, the expression score of TLS-25 signatures is significantly associated with better overall survival in liver cancer (Fig. 4C).

We then evaluated the predictive performance of TLS-25 signatures in identifying TLS regions using SRT data. In the PLC dataset, the TLS-25 signatures demonstrated comparable performance to TLS-50 signatures, outperforming other methods (*5*, *32*–*34*) in predicting TLS regions annotated by pathologists (Fig. 4D). Spatial regions with high expression scores (highlighted in red) consistently and closely matched expert-annotated TLS regions (indicated by black circles) (Fig. 4E), suggesting that the TLS-25 signatures are

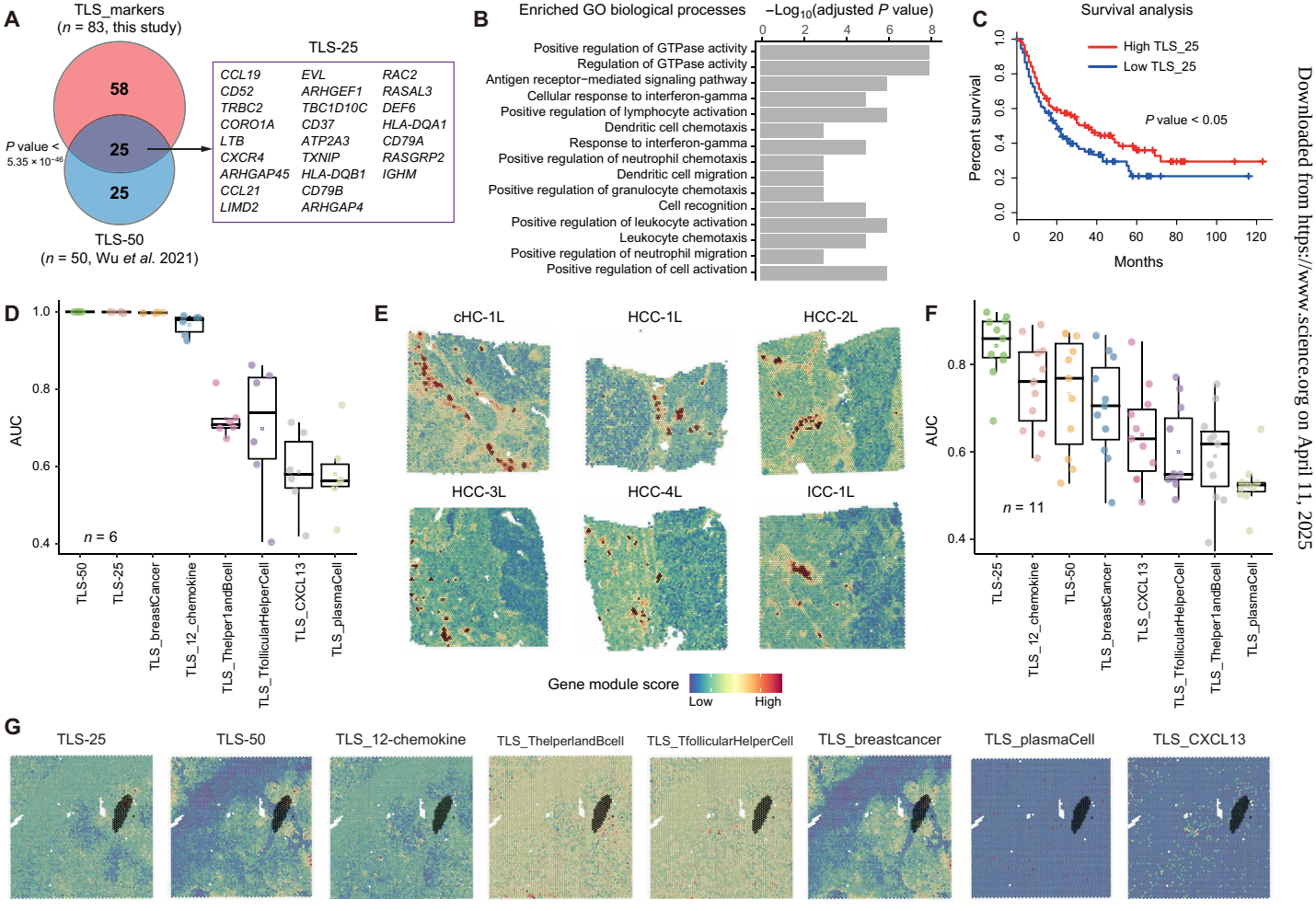


**Fig. 4. Improved prediction of TLSs by a refined marker gene set.** (**A**) Significant intersection (labeled as TLS-25) of differential genes derived from with TLS-related CellTopics in PLC and the reported TLS-50 genes (*24*). (**B**) Functional enrichment analysis of the TLS-25 genes. (**C**) Kaplan-Meier curves of patients with liver cancer from The Cancer Genome Atlas (TCGA) program ($n$ = 364) showing the survival rates grouped by the rank of TLS-25 gene module score. The $P$ value is calculated with two-sided log-rank test. (**D**) Evaluation of the prediction performance of TLSs in six PLC samples (*24*) using eight different TLSs marker gene sets. TLS-50 from ref. (*24*), TLS-25 from this study, "TLS_breastCancer" from ref. (*5*), "TLS_12_chemokine" from ref. (*33*), and "TLS_CXCL13," "TLS_plasmaCell," "TLS_TfollicularHelperCell," and "TLS_ThelperlandBcell" from ref. (*32*). AUC, area under the curve. (**E**) Spatial map representation of the TLS-25 gene module scores in the PLC data. Expert-annotated TLSs are labeled by black circles. (**F**) Evaluation of the prediction performance of TLSs in 11 ccRCC samples (*35*) using different gene sets. (**G**) Spatial patterns of gene module scores for different gene sets in the GSM5924035 sample of ccRCC.

effective in predicting TLSs in PLC (Fig. 4E). Furthermore, we investigated if the TLS-25 signatures derived from liver cancers (including cHC, HCC, and ICC) could predict TLSs in other cancer types. We applied SpaTopic to analyze SRT data from clear cell renal cell carcinoma (ccRCC) (*35*), where TLSs were identified in 11 slices based on expert manual annotations, immunostaining assays, as well as B lineage and T cell scores. SpaTopic accurately predicted TLS locations [with an area under the curve (AUC) up to 0.95] in all TLS-positive tumors based on the TLS-25 signatures and consistently outperformed existing methods (including TLS-50 signatures) in ccRCC (Fig. 4F). As exemplified in Fig. 4G, the TLS-25 signatures derived from liver cancers exhibited predominantly high expression scores (*24*) within the true TLS-positive regions (marked with black circles) in ccRCC, whereas other known signatures exhibited lower consistency. This suggests that TLSs across different cancer types share similar gene expression patterns according to the TLS-25 signatures. In summary, the analysis above emphasizes the effectiveness of TLS-25 signatures for accurate TLS prediction across tumors, highlighting their potential significance in cancer diagnosis.

### Illumination of cell-type architecture cross the TB

The TB, also known as the tumor margin or tumor edge, refers to the outer limit or physical border of a tumor mass within the TME. It marks the spatial region where the tumor tissue transitions into the surrounding normal tissue. Accurate delineation of the TB is essential for assessing the growth and stage of the tumor, evaluating its invasiveness, and monitoring its response to treatment (*36*). Traditional approaches, such as tumor histology and gene expression analysis, often fall short in capturing the dynamic interactions and chemotaxis between different cell populations in the microenvironmental region. Here, we explored if spatial domains identified by SpaTopic could provide insights into the spatial architecture of TBs, specifically in terms of cell composition and molecular expression patterns. We collected SRT data from four tumor types characterized by well-defined TBs, resulting in a total of 17 SRT datasets obtained from PLC (*24*), breast cancer (BRCA; 10x Genomics), ovarian cancer (OV; 10x Genomics), and oral squamous cell carcinoma (OSCC) (*37*) (Fig. 5A, fig. S12, and table S2). We used SpaTopic to identify CellTopics for each sample individually and subsequently assigned specific CellTopics to manually annotated TB regions in each slice (Fig. 5, A and B).

As illustrated in Fig. 5C, SpaTopic identified distinct CellTopics representing the "transition areas" (or TBs) (*37*) between normal regions and tumor regions in PLC (HCC-1L, HCC-3L, and HCC-4L) (*24*), OV, and BRCA (BRCA1 and BRCA2). Specifically, CellTopic8 marked the transitional boundary in HCC-1L, CellTopic6 in HCC-3L, CellTopic11 in HCC-4L, CellTopic1 in OV, CellTopic2 in BRCA1, and CellTopic1 in BRCA2. Of note, these spatial domains identified by SpaTopic are predominantly distributed around the boundary of tumors. The consistent colocalization of cell types within these CellTopics, including hepatic stellate cells [the main source of liver fibroblasts (*38*)], plasmacytoid DCs (pDCs), and B cells enriched in the PLC data (Fig. 5D), aligns with the previously reported accumulation of immune cells in the TB (*24*, *36*). In the OSCC dataset (11 slices in total), we observed a notable enrichment of fibroblasts in the spatial regions of TBs (fig. S12). Overall, our analysis uncovered a notable rise in fibroblast abundance localized within the boundary of diverse tumors. This conserved microenvironmental pattern underscores the crucial role that fibroblasts may play in tumor progression and metastasis.

In the PLC samples (HCC-1L, HCC-3L, and HCC-4L), the boundary regions distinctly delineated the TME into two halves. We identified a common set of spatially variably expressed genes ($n = 13$) that were consistently up-regulated in the boundary regions of all three samples (Fig. 5E). Among these genes, *MYL9*, a member of the myosin superfamily expressed predominantly in CAFs, is known to promote cancer proliferation, invasion, metastasis, and angiogenesis (*39*–*41*). The overall expression patterns of these genes were highly specific to the CellTopic regions corresponding to TBs in PLC (Fig. 5F), indicating their predictive capability for TB regions. This notion is further supported by the observation that the identified TB-associated genes in PLC were also notably highly expressed at the boundaries in OV, BRCA, and OSCC (Fig. 5F), underscoring the similarity in expression patterns of the boundary regions across different tumor types. On the other hand, marker genes highly expressed in boundary regions from other cancer types (i.e., OV or BRCA) demonstrated a significantly positive correlation in terms of gene module scores across PLC, BRCA, and OV (Fig. 5G), providing additional evidence for conserved gene expression patterns in TBs across different cancer types. Functional analysis of the boundary-associated genes unveiled significant enrichment in biological pathways related to matrix remodeling and formation, including extracellular matrix structure organization, regulation, aggregation and activation of platelets, as well as cell migration and adherence (Fig. 5H).

In TB-associated spatial domains, gene expression may be dependent on their spatial location. For example, cells at the boundary of a spatial domain may have expression profiles quite different from those at the center of the spatial domain. To test this, we applied the "FindSpatiallyVariables" function implemented in Seurat (*42*) to identify spatially variable genes within the TB-associated spatial domains in the three slices (HCC-1L, HCC-3L, and HCC-4L) of the PLC data. Among the top 50 spatially variable features from each of the three samples, 20 conserved genes were identified in common (fig. S13A). Furthermore, these 20 conserved genes and the 13 markers for the TB (Fig. 5E) showed a significant overlap (hypergeometric test, $P$ value $< 9.9 \times 10^{-7}$; fig. S13B), including *TAGLN*, *MYL9*, *ACTA2*, and *COL1A2* (fig. S13C). The analysis indicates that gene signatures for the TB-associated spatial domains displayed spatially variable expression patterns within the spatial domain.

In summary, SpaTopic accurately identifies spatial domains corresponding to the TB in various cancer types, unveiling distinctive and conserved cellular and molecular patterns that contribute to TB formation.

### Comparison of spatial domains in colorectal cancer and liver metastatic tumors

The score of CellTopics facilitates quantitative comparison of spatial domains across different SRT datasets. To validate this capability, we used SpaTopic to characterize and compare spatial cellular organization in colorectal cancer (CRC) primary and liver metastatic tumors (*43*). SpaTopic respectively identified distinct CellTopics in primary (C1 to C4) and metastatic (L1 and L2) tumors based on available cell-type annotations (Fig. 6A). Clustering analysis based on topic scores revealed seven major categories of CellTopics, referred to as MetaTopics, wherein each MetaTopic comprised CellTopics sourced from distinct tumors (Fig. 6B). Generally, CellTopics

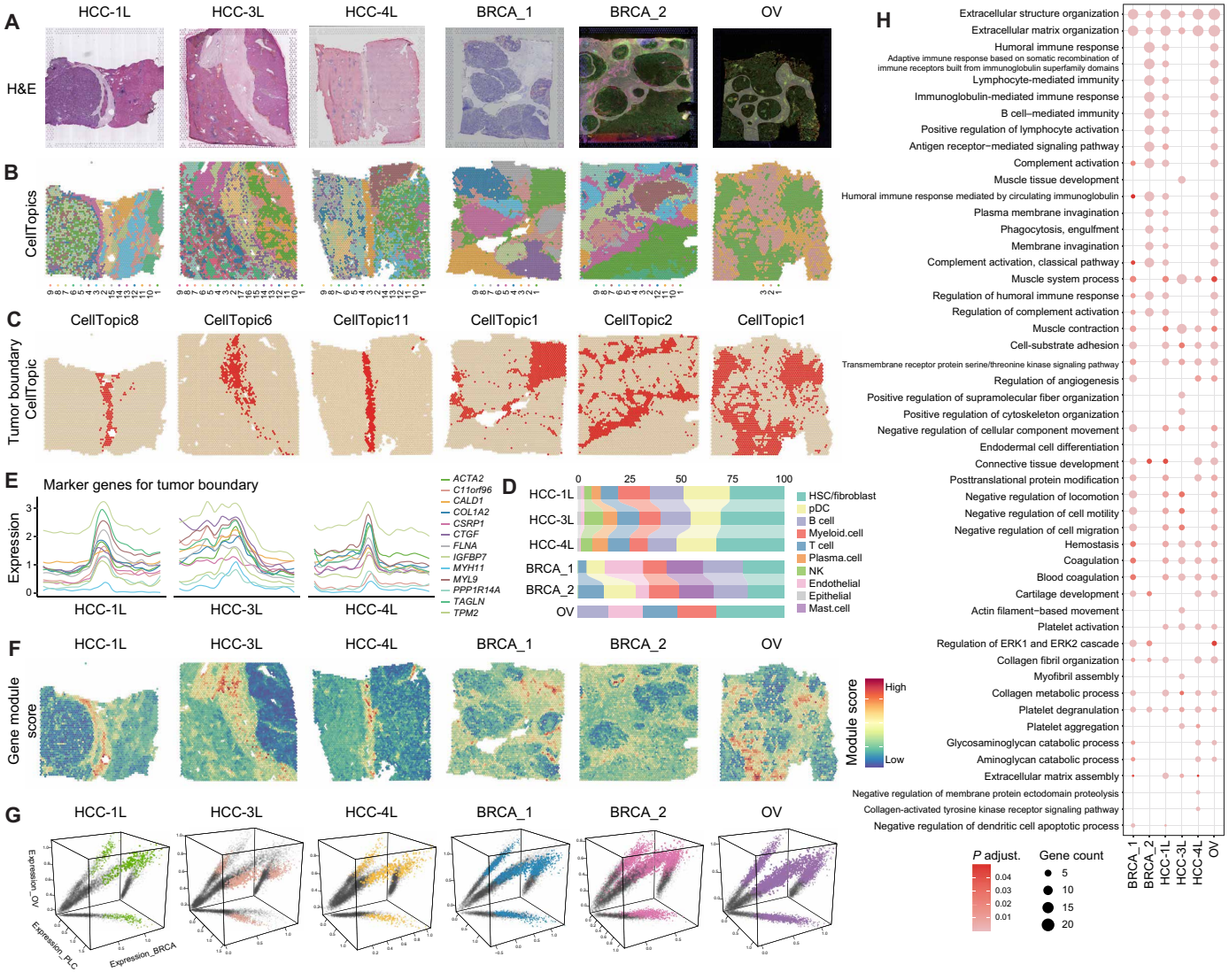


**Fig. 5. Pan-cancer prediction and annotation of TBs by SpaTopic.** (**A**) H&E staining of six tumors from three different cancer types. The white areas in the image are the potential TB regions according to manual annotation. (**B**) CellTopics identified by SpaTopic. (**C**) CellTopics associated with TBs in each tumor tissue slice (highlighted in red). (**D**) Composition of cell types within the CellTopics related to TBs in each tissue slices. (**E**) Expression of the module score of 13 common variable genes spatially distributed across TBs in the three tissue sections of PLC. (**F**) Spatial map of the module score of the 13 TB-associated marker genes derived from the PLC data. (**G**) Correlation analysis of the module score for highly expressed genes in TB regions across three different cancer types (PLC, BRCA, and OV). The highlighted spots correspond to the CellTopics associated with TBs in each slice. (**H**) GO enrichment analysis of the TB-associated genes in each tissue slices.

in the same MetaTopic showed consistent cellular organization enriched for combinations of specific cell types (Fig. 6C). For instance, the Mac_SPP1 cell subset demonstrated high enrichment in the CellTopics from MetaTopic 2 (M2), which were specifically derived from primary tumors. In contrast, the Mac_CXCL9 subset exhibited an increase in M4 and M6, where the CellTopics were sourced from metastatic tumors (Fig. 6C). These findings of cell subset enrichment in SRT data align with the results obtained through single-cell data analysis in the original study (*43*). Overall, both primary and metastatic tumors displayed a combination of shared and distinct MetaTopics. Notably, M6 and M7 were exclusive to metastatic tumors, whereas M2, M3, and M5 were specific to primary tumors. M1 and M4 were found to be shared between both primary and metastatic tumors (Fig. 6D). The analysis could provide valuable insights into the variations and similarities in the spatial architecture of tumors originating in the colon/rectum and their metastatic counterparts in the liver.

To further explore spatial gene expression programs associating with tumor metastasis, we conducted a comparative analysis of gene expression patterns among different MetaTopics (see Materials and Methods). This analysis revealed 907 genes that exhibited significant differential expression across MetaTopics [analysis of variance (ANOVA), adjusted $P$ value < 0.05; Fig. 7A]. These genes were further categorized into seven distinct gene modules (K1 to K7) based on their expression patterns. For instance, the gene module K3 exhibits notably high expression specifically in the metastatic-specific MetaTopic M7, with these genes enriched for biological processes such as fatty acid metabolic process and acute inflammatory response,

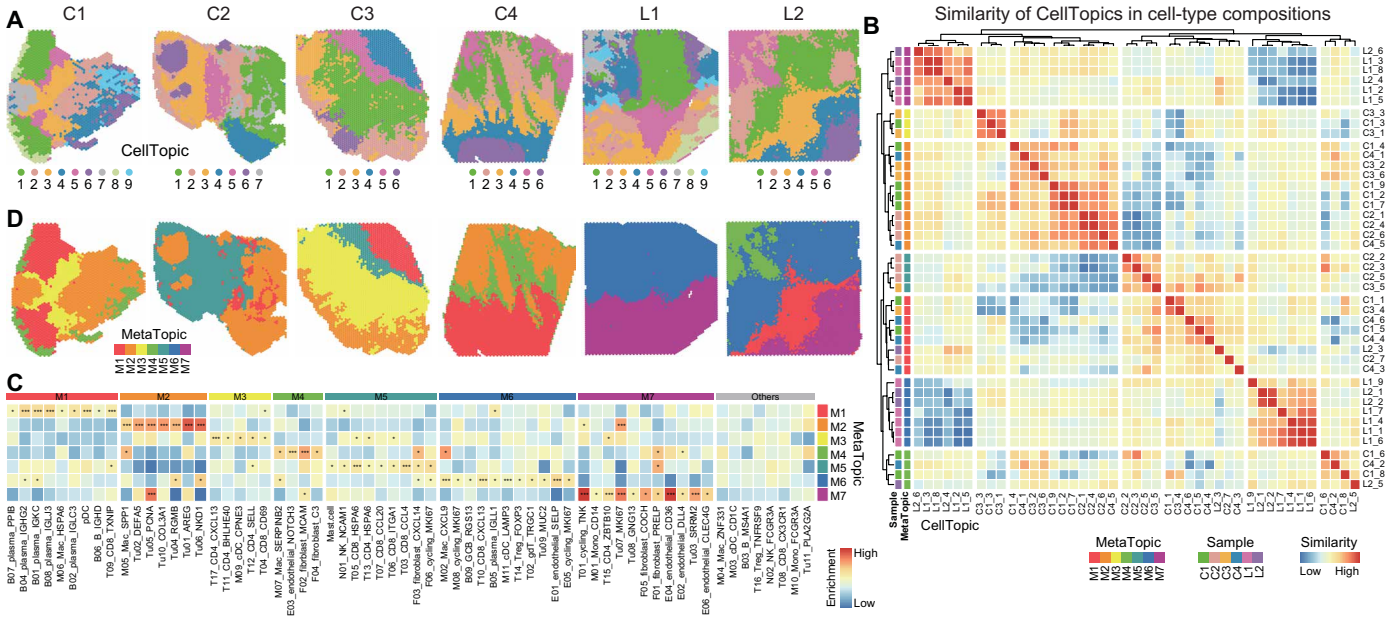


**Fig. 6. Quantitative comparison of spatial domains in CRC and liver metastatic tumors.** (**A**) SpaTopic identified CellTopics in four instances of primary CRC (C1, C2, C3, and C4) and two instances of liver cancer metastasis (L1 and L2). (**B**) Heatmap showing the similarity in cell-type compositions (i.e., normalized topic score) of CellTopics identified in the six slices. Hierarchical clustering analysis reveals seven distinct groups of CellTopics, termed MetaTopics. (**C**) Heatmap demonstrating the average cell-type compositions across the MetaTopics. "*" and "***" denote the significant level of enrichment (Wilcoxon tests). *0.01 < $P \leq 0.05$; ***$P \leq 0.01$. (**D**) Spatial map of MetaTopics across the different slices.

well-established hallmarks of tumor metastasis (Fig. 7A) (*44*–*46*). Conversely, the gene modules K4 and K5 exhibit specific activation in primary tumor–related MetaTopics (M2 and M3 for K4 and M1 to M5 for K5), highlighting biological functions including energy production, metabolic regulation, synthesis, and metabolism of reactive oxygen species, actin cytoskeleton organization, cellular morphogenesis, as well as the assembly and organization of protein-containing complexes and supramolecular fibers. The gene module K7 is shared between primary and metastasis-related MetaTopics (M1, M5, and M7), with functional enrichment of various immune-related pathways. As expected, the gene module score (*26*) showed specific enrichment in the corresponding spatial domains identified by SpaTopic (Fig. 7B), indicating the presence of distinct spatial gene expression programs with functional implications.

To validate if gene modules derived from MetaTopics capture the biological variability among different SRT data, we harmonized the six SRT datasets with the Seurat CCA (canonical correlation analysis) integration approach (*47*), without considering the spot location information (Fig. 7C). The MetaTopics were distinctly mapped into specific spot clusters with similar gene expression patterns (Fig. 7, C and D). In line with this observation, the gene module score was notably enriched in closely clustered spots on the integrated spot map (Fig. 7E). The above results collectively support the efficient use of SpaTopic for quantitative comparison and functional annotation of spatial domains.

## DISCUSSION

Tumor tissues represent intricate microenvironments shaped not only by the diverse array of cell types present but also by the spatial organization that governs their interactions. In the TME, the spatial arrangement of fibroblasts and endothelial cells can influence angiogenesis, a critical process for tumor growth and metastasis (*48*). Spatial domains are localized regions within tissues featuring distinct anatomical structures and specialized functions. Each domain is generally characterized by unique local features, including specific cell-type compositions, gene expression patterns, and cell-cell interactions (*49*). Spatial domains coordinate with each other in carrying out the overall tissue functions. TLSs and TBs exemplify such spatial domains known to affect tumor development and metastasis (*23*, *36*). Improved understanding of the cellular and molecular aspects of spatial domain development and function may pave the way for leveraging these structures to enhance cancer therapies.

Emerging technologies, such as SRT, enable the high-resolution mapping of gene expression within the context of tissue architecture (*6*). Various computational tools have been developed to identify and annotate spatial domains in SRT data, on the basis of either spatial clustering or cell-type deconvolution approaches (*8*–*10*, *49*–*53*). However, most of these tools are designed and validated using SRT data from developmental tissues, where there is a clear and distinct organization of cell types along the tissue layer (*50*). Exploring tumor spatial domains with biological functions (e.g., TLSs and TBs), it is crucial to identify both variable genes and cell-type composition that show coherent enrichment in the identified domains. In this study, we presented SpaTopic, a method that jointly considers spatial clustering and cell-type deconvolution for spatial domain identification and annotation, with a specific focus on exploring the TME. In this manner, SpaTopic addresses the challenge of spatial domain detection being predominantly influenced by gene expression, thus reducing potential discrepancies between the identified domains and the underlying anatomical structure of the tissue (especially for tumor tissues). Furthermore, SpaTopic excels in capturing

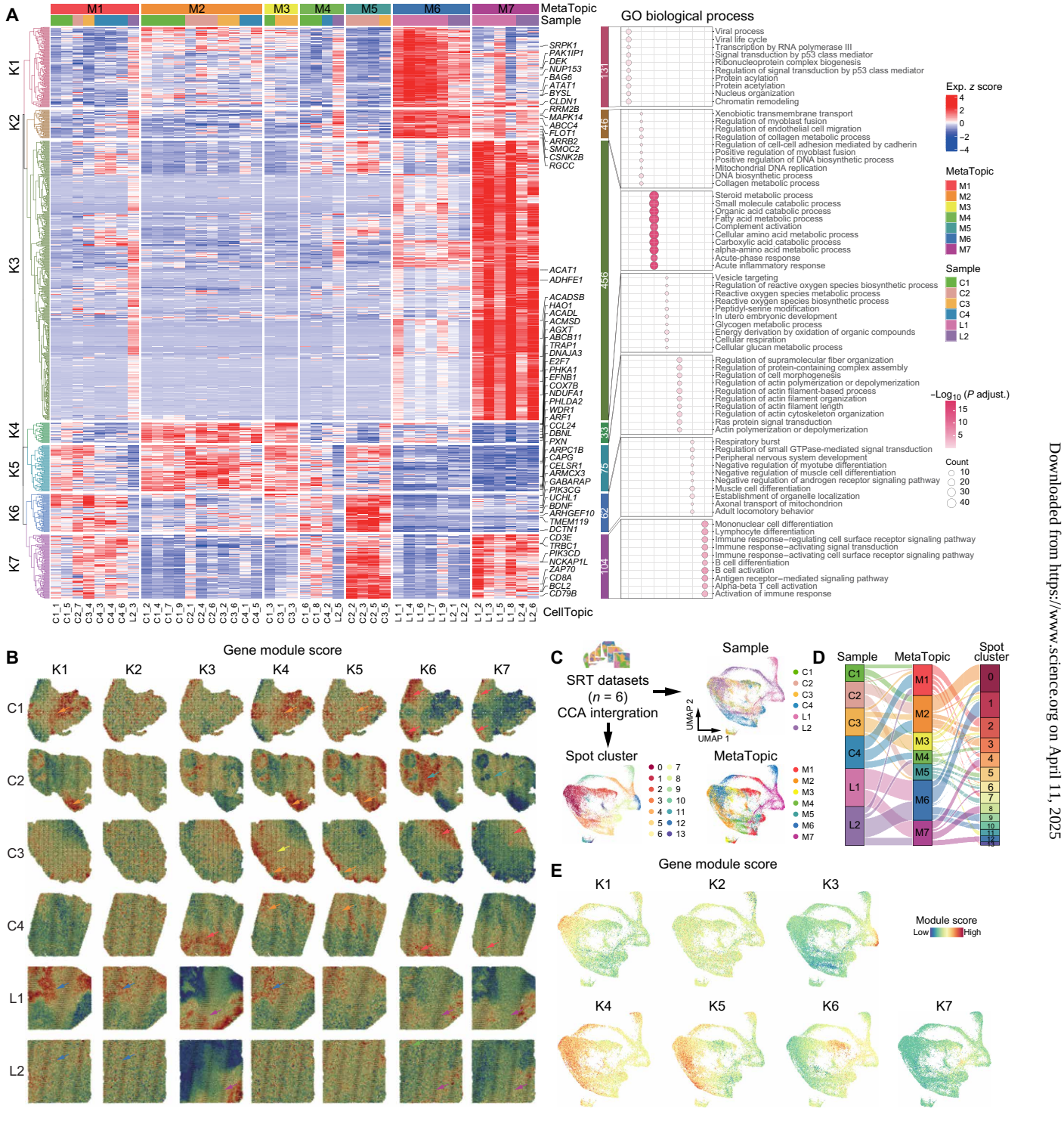


**Fig. 7. Expression analysis of MetaTopics.** (**A**) Heatmap showing the expression patterns of genes that are significantly highly expressed in specific MetaTopics as defined in Fig. 6. Seven gene modules (K1 to K7) are observed based on the expression pattern. Example genes are shown. (**B**) Spatial map of module scores of the seven gene modules across the six slices. Arrows indicate MetaTopic-matched spatial regions with high module scores. (**C**) Integration of the six SRT data by the Seurat CCA method (top left). Uniform manifold approximation and projection (UMAP) plots show the distribution of spots based on samples (top right), spot clusters (bottom left), or MetaTopics (bottom right). (**D**) Sankey diagram representing the relationship of spots among the sample origin, MetaTopic, and integrated spot clusters. (**E**) Expression pattern of the gene module score over the UMAP.

variations in both spatial and cell-type gene expression patterns for robust spatial domain detection. Through comprehensive analyses across diverse cancer types, we demonstrate the ability of SpaTopic to accurately predict TLSs and TBs. Our results consistently showed that SpaTopic can identify spatial domains with both coherent gene expression and cell-type composition and a core set of genes that have much clearer spatial expression patterns and biological interpretations. In addition, marker genes for TLSs and TBs derived from SpaTopic demonstrate transferability and could be used for potential targets for cancer treatment.

The detailed insights into cell-cell colocalization provided by SpaTopic offer valuable clues about crucial regulatory mechanisms in disease contexts. In the context of TLSs, the tight spatial coupling between lymphocytes and nonhematopoietic stromal cells, particularly fibroblasts, suggests the possibility of their organization into an immunofibroblast network (*54*) potentially playing a role in the establishment and maintenance of TLSs. In contrast, in the context of TBs, CAFs are the pivotal components of boundary structures, yet the specific colocalized cellular partners can vary significantly among different tumor types. Knowledge of such cell-cell spatial relationships in the TME can support the development and implementation of CAF-targeting strategies for more effective tumor treatment (*55*).

Another key advantage of SpaTopic is its ability to prioritize the contribution of cell types to spatial domains. This feature not only facilitates quantitative comparison of spatial domains across different SRT datasets based on their cell-type compositions but also empowers the identification of previously unidentified spatial domains for subsequent functional characterization and experimental validation. Moreover, SpaTopic identifies gene modules with distinct spatial gene expression patterns, facilitating the functional characterization of spatial domains. We anticipate that SpaTopic will be instrumental in method development and the exploration of SRT data.

As single-cell resolution spatial transcriptomics technologies continue to evolve, SpaTopic can be applied directly to single-cell resolution SRT data, eliminating the need for deconvolution analysis. In this context, the deconvoluted spot-cell type matrix can be replaced by the cell-cell type matrix provided by SRT data. To test this possibility, we applied SpaTopic to a single-cell resolution dataset from the NanoString CosMx platform on non–small cell lung cancer (NSCLC) (*56*). SpaTopic successfully identified and annotated spatial domains supported by known cell-type compositions and marker gene expression patterns (fig. S14). This demonstrates the adaptability and potential of SpaTopic for long-term utility in the spatial research area, regardless of the resolution of the data.

However, it should be noted that the outcome of SpaTopic might be influenced by the cell-type resolution used in the analysis. In SpaTopic, we use CARD for deconvolution, and the accuracy of CARD's deconvolution is affected by the cell-type resolution (*12*). If too many sub-cell types are included, then the specificity of the gene expression profiles may be diluted, leading to less accurate deconvolution. Conversely, if too few cell types are considered, then important distinctions between different cell types may be missed, also affecting the accuracy. In general, the choice of cell-type resolution level in SpaTopic should align with the biological question being investigated. For example, when analyzing spatial domains like TLSs, a resolution at the major cell-type level might be necessary to accurately identify and characterize the involved cell types. For de novo detection of spatial domains, we recommend a balanced approach—using a sufficient number of cell types to capture the diversity within the tissue while ensuring that each cell type has a distinct gene expression profile to avoid overlap. Although different cell-type resolutions lead to distinct cell-type combinations within spatial domains, we found that the top representative cell-type compositions in a given spatial domain are generally consistent between major and minor cell types (fig. S15), indicating that SpaTopic maintains considerable accuracy and consistency across different cell-type resolutions.

## MATERIALS AND METHODS

### Overview of the SpaTopic framework

SpaTopic uses both single-cell data, annotated with cell-type information, and SRT data, annotated with location information, as the input. First, unsupervised clustering is applied to the SRT data to generate preclustering of spots. Second, deconvolution analysis is performed on the combined single-cell and SRT data to estimate cell-type proportions for each spatial spot. Third, the results of deconvolution and clustering analysis is used to compute the specific score of cell types for each spot cluster using the KS test, which aims to enhance the identification of specific cell types in specific spatial clusters. Fourth, the generative probability model LDA is used to model the specific score matrix and derive intermediate layer topics. LDA produces two distribution matrices: **C1** and **C2**. The **C1** distribution matrix signifies that each topic comprises multiple cell types, describing a combination of cell-type distributions. The **C2** distribution matrix signifies that each domain comprises a mixture of multiple topics, describing a combination of spatial domain distributions. Last, by binarizing distribution **C2**, we amalgamate informative topics as CellTopic to represent distinct spatial domains and predict the proportion of cell types in the corresponding CellTopic.

### Preclustering of SRT data

Preclustering of spots can be achieved using either nonspatial clustering methods based solely on expression information [such as mclust (*16*) and Louvain] or by integrating spatial coordinates and expression information via spatial clustering approaches [such as STAGATE (*10*), SpaGCN (*9*), and BayesSpace (*8*)]. While both nonspatial and spatial clustering methods are applicable for preclustering spots in SpaTopic, we use the spatial clustering method STAGATE as the default approach due to its robust performance. The analysis results in a preclustering of spots, where each spot corresponds to an initial spatial domain. However, it is important to note that this initial preclustering of spots will undergo further refinement and improvement by SpaTopic in subsequent steps.

### Deconvolution analysis

Deconvolution analysis is conducted to infer the spatial distribution and abundance of different cell types within the tissue sample based on the integrated single-cell and SRT data. In SpaTopic, the gene expression signatures of known cell types obtained from the single-cell transcriptomic data are used to deconvolve the spatially resolved gene expression profiles from SRT data by CARD (*12*). Specifically, we denote **D** (spot × celltype) as the cell-type composition matrix, representing the proportion of cell types at each spatial location. The spatial transcriptome data matrix is denoted as **X** (gene × spot), and the cell type–specific score matrix is denoted as **C** (gene × cell types),

which can usually be estimated from single-cell data. Three matrices can be harmonized into a non-negative matrix factorization model (Eq. 1), where **E** is the residual matrix

$$\mathbf{X} = \mathbf{C} \times \mathbf{D}^T + \mathbf{E} \tag{1}$$

In CARD (*12*), spatial constraints are added in the model to accurately estimate **D** (*57*) because cell types in neighboring locations tend to be similar.

### Calculation of the cell type–specific score

The above deconvoluted cell-type abundance matrix **D** and the preclustering of spots are used to calculate the specific score for cell types across different spot clusters. To do so, we use the KS test, a nonparametric usually used to compare the significance of differences between two empirical distributions. For each cell type, we assess the specific score of estimated cellular abundance in a given spot cluster by comparing the cumulative distribution functions (CDFs) between the specified cluster and all clusters. We denote $F_m(x)$ and $F_0(x)$ as the CDFs for cluster $m$ and all clusters, respectively. Given that the domain $m$ containing $n_m$ spots and all the domains $n_0$ spots, the test statistic $Z$ (Eq. 2) follows an approximate normal distribution

$$Z = max\left|F_m(x) - F_0(x)\right| \sqrt{\frac{n_m n_0}{n_m + n_0}} \tag{2}$$

where *max* is the maximum absolute difference between the two CDFs over all possible values of $x$. On the basis of the statistical metric $Z$, we obtained the specific score for individual cell types in a given spatial cluster. The cell type–specific score matrix **S** represents the specific score for each spatial cluster (row) and each cell type (column).

### Topic modeling

LDA is a probabilistic generative model extensively used in natural language processing and machine learning for topic modeling. LDA is designed to uncover the underlying topics or themes within a collection of documents. In LDA, each document is represented as a mixture of topics, and each topic is represented as a distribution over a fixed vocabulary of terms. The model assumes that documents exhibit multiple topics, and each word in a document is attributable to one of the document's topics. Here, we use topic modeling to describe the characteristics of each spatial domain. In this context, we liken each spatial domain to a document and each cell type to a word to reveal the spatial distribution patterns of different cell types.

In theory, consider the topic distribution of topics $z$ in spatial domain $d$, denoted as $v_d$, and the distribution of cell type $c$ under topic $z_m$, denoted as $\varphi_m$

$$v_d : p(z \mid d, \alpha) \sim \text{Dirichlet}(\alpha) = \frac{1}{B(\alpha)} \prod_{k=1}^{K} z_k^{\alpha_k - 1} \tag{3}$$

$$\varphi_m : p(c \mid z_m, \gamma) \sim \text{Dirichlet}(\gamma) = \frac{1}{B(\gamma)} \prod_{k=1}^{K} c_k^{\gamma_k - 1} \tag{4}$$

where $\alpha$ and $\gamma$ serve as hyperparameters for the Dirichlet distribution, and $B(.)$ is the multivariate beta function. For each spatial domain $d$, the number of occurrences for each topic is denoted as $n_d$, and $N_d$ represents the number of cell types in $d$. For each spatial domain $d$ and each cell type $c$, latent variables $z_{d,c}$ are introduced to signify their topic assignments. Considering all cell types, latent variables $z_{d,c}$ can be combined into $z_d$. These assignments are sampled from the Dirichlet distribution $p(z|d, \alpha)$. Repeat these steps in every spatial domain, then the topic to which each domain in $S$ belongs can be determined

$$n_d \sim \text{Multinomial}\left(n_d \mid v_d, N_d\right) \tag{5}$$

$$p\left(z_d \mid \alpha\right) = \frac{B\left(\alpha + z_d\right)}{B(\alpha)} \tag{6}$$

Subsequently, the distribution of cell type $c$ under topic $z$ is considered. For a specific topic $z_m$, given $z_{d,c}$, and this distribution, we obtain the number of cell types, which has been assigned to topic $z_m$, denoted as $N_m$. Then, variables $n_{m,c}$ are introduced, representing the number of occurrences for a certain cell type $c$ in topic $z_m$. Considering all cell types, variables $n_{m,c}$ can be combined into $n_m$. When the distribution of $n_m$ is determined, the distribution of cell type $c$ in topic $z_m$ can be derived from Bayesian theory

$$n_m \sim \text{Multinomial}\left(n_m \mid \varphi_m, N_m\right) \tag{7}$$

$$p\left(c_{z_m} \mid \gamma\right) = \frac{B\left(\gamma + n_m\right)}{B(\gamma)} \tag{8}$$

Through such a model, we can derive the joint distribution of cell type $c$ and topics $z$ in domain $d$

$$p(c, z \mid \alpha, \gamma) = \left[\prod_{m=1}^{K} p\left(c_{z_m} \mid \gamma\right)\right] * \left[\prod_{d=1}^{D} p\left(z_d \mid \alpha\right)\right] \tag{9}$$

where $K$ denotes the number of topics and $D$ denotes the number of domains. In this manner, we can obtain the posterior distributions of topic and cell-type distributions through Gibbs sampling, ultimately revealing the latent structures of cell types and spatial regions.

When applied the above model to the cell type–specific score matrix **S**, LDA yields two critical statistical distributions, denoted as **C1** and **C2**. Distribution **C1** conceptualizes each topic as a composition of diverse cell types, effectively illustrating mixed cell-type distributions. Conversely, distribution **C2** portrays each spatial domain as a distribution of cell-type topics.

### Cluster-topic binarization for spatial domain calling

To assign a spatial cluster to one or more specific topics, binarization methods are applied to the cluster-topic matrix. This step allows for the refinement of the initial spot clusters into spatial domains based on the binarized cell-type topics (termed CellTopics).

Our first method involves exploring a dynamic fusion of topics within spatial clusters. It commences by pinpointing clusters through topic selection, where each topic undergoes scoring within the spatial cluster, with the highest-scoring cluster retained for each topic. Subsequently, we pivot toward topic selection within spatial clusters, integrating this with the score of the top candidate topic within the cluster to orchestrate a tiered selection process. The objective is to discern an ensemble of candidate topics collectively contributing to 60% of the total significance across all spatial clusters. This ensemble is considered as the combined CellTopics.

Furthermore, a complementary approach is adopted, focusing exclusively on a single topic per cluster. Similar to the first method, the initial step involves identifying a specific cluster for each topic,

retaining those with the highest scores. However, instead of pursuing a combined contribution, we concentrate on individual significance. The highest-scoring topic within each domain is selected, establishing a one-to-one correspondence between topics and spatial clusters. In cases where a cluster remains unmatched with any topic, we use a cyclical approach, revisiting the selection process and considering the next best scores until each cluster is paired with a topic.

### Method comparisons

We compared SpaTopic with five methods for spatial domain detection: (i) STAGATE (*10*), (ii) SpaGCN (*9*), (iii) BayesSpace (*8*), (iv) mclust (*16*), and (v) Louvain. For each method, we followed the tutorial on the corresponding GitHub pages and used the recommended default parameter settings for clustering analysis. For the STAGATE algorithm, the version is strictly set to match the original paper (*10*) to prevent the version from influencing the results (i.e., Python version 3.7 and TensorFlow version 1.15.0). In SpaGCN, we used the following recommended parameter settings from the original study (*9*): random seed set to 100, number of clusters to 7, learning rate (lr) of 0.05, neighborhood expression contribution percentage of 0.5, max_epoch of 20, start of 0.7, step of 0.1, and tolerance (tol) of $5 \times 10^{-3}$, among others. For the BayesSpace mclust and Louvain algorithm, we used the default parameters in the R package and obtained results consistent with the original study (*8*, *16*).

### Adjusted Rand index

The *ARI* is a measure of the similarity between two clustering results, taking into account both the agreement and disagreement between the cluster assignments. It is commonly used to evaluate the performance of clustering algorithms. The *ARI* is calculated as shown in Eq. 10

$$ARI = \frac{2(ad - bc)}{(a + b)(b + d) + (a + c)(c + d)} \quad (10)$$

where $a$ represents the number of pairs of data points that belong to the same cluster in both the true and predicted cluster, $b$ represents the number of pairs of data points that belong to the same cluster in the true clustering but are assigned to different clusters in the predicted cluster, $c$ represents the number of pairs of data points that belong to different clusters in the true clustering but are assigned to the same cluster in the predicted clustering, and $d$ represents the number of pairs of data points that belong to different clusters in both the true and predicted clustering.

The *ARI* ranges from −1 to 1, where a value of 1 indicates a perfect match between the true and predicted cluster, a value of 0 indicates a random agreement, and a value less than 0 indicates a disagreement between the cluster.

### CellTopic interaction analysis

Spot-spot interactions among the CellTopics were estimated by CellChat (*21*).

### Gene module scoring

The average expression levels [called module score (*58*)] of gene modules were calculated through the "AddModuleScore" function implemented in the Seurat package.

### Functional enrichment analysis

Gene ontology (GO) enrichment analysis of marker genes in this study was conducted using the clusterProfiler package (*59*). Gene set variation analysis (GSVA) was performed using the GSVA package (*60*).

### Survival analysis

Survival curves were generated using the Kaplan-Meier method with the GEPIA2 (*61*). Significance between two groups was assessed using the log-rank test statistics (*P* values).

### Statistical analysis and data visualization

If not specified, all statistical analyses and data visualization were done in R (version 4.0.0). Heatmap visualization was done using the ComplexHeatmap package (*62*). SRT and single-cell data processing and visualization was mainly performed using the Seurat package (*58*).

## Supplementary Materials

**This PDF file includes:**

Figs. S1 to S15
Tables S1 and S2

## REFERENCES AND NOTES


1. X. Zhou, R. A. Franklin, M. Adler, J. B. Jacox, W. Bailis, J. A. Shyer, R. A. Flavell, A. Mayo, U. Alon, R. Medzhitov, Circuit design features of a stable two-cell system. *Cell* **172**, 744–757.e17 (2018).
2. E. Armingol, A. Officer, O. Harismendy, N. E. Lewis, Deciphering cell-cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 (2021).
3. T. Krausgruber, N. Fortelny, V. Fife-Gernedl, M. Senekowitsch, L. C. Schuster, A. Lercher, A. Nemc, C. Schmidl, A. F. Rendeiro, A. Bergthaler, C. Bock, Structural cells are key regulators of organ-specific immune responses. *Nature* **583**, 296–302 (2020).
4. M. V. Hunter, R. Moncada, J. M. Weiss, I. Yanai, R. M. White, Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. *Nat. Commun.* **12**, 6278 (2021).
5. P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Å. Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, J. Frisén, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
6. A. Rao, D. Barkley, G. S. Franca, I. Yanai, Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
7. L. Tian, F. Chen, E. Z. Macosko, The expanding vistas of spatial transcriptomics. *Nat. Biotechnol.* **41**, 773–782 (2023).
8. E. Zhao, M. R. Stone, X. Ren, J. Guenthoer, K. S. Smythe, T. Pulliam, S. R. Williams, C. R. Uytingco, S. E. B. Taylor, P. Nghiem, J. H. Bielas, R. Gottardo, Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol.* **39**, 1375–1384 (2021).
9. J. Hu, X. Li, K. Coleman, A. Schroeder, N. Ma, D. J. Irwin, E. B. Lee, R. T. Shinohara, M. Li, SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat. Methods* **18**, 1342–1351 (2021).
10. K. Dong, S. Zhang, Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat. Commun.* **13**, 1739 (2022).
11. V. Kleshchevnikov, A. Shmatko, E. Dann, A. Aivazidis, H. W. King, T. Li, R. Elmentaite, A. Lomakin, V. Kedlian, A. Gayoso, M. S. Jain, J. S. Park, L. Ramona, E. Tuck, A. Arutyunyan, R. Vento-Tormo, M. Gerstung, L. James, O. Stegle, O. A. Bayraktar, Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
12. Y. Ma, X. Zhou, Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat. Biotechnol.* **40**, 1349–1359 (2022).
13. M. Elosua-Bayes, P. Nieto, E. Mereu, I. Gut, H. Heyn, SPOTlight: Seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**, e50 (2021).
14. D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
15. K. R. Maynard, L. Collado-Torres, L. M. Weber, C. Uytingco, B. K. Barry, S. R. Williams, J. L. Catallini II, M. N. Tran, Z. Besich, M. Tippani, J. Chew, Y. Yin, J. E. Kleinman, T. M. Hyde, N. Rao, S. C. Hicks, K. Martinowich, A. E. Jaffe, Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021).
16. L. Scrucca, M. Fop, T. B. Murphy, A. E. Raftery, mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R J.* **8**, 289–317 (2016).
17. R. Moncada, D. Barkley, F. Wagner, M. Chiodin, J. C. Devlin, M. Baron, C. H. Hajdu, D. M. Simeone, I. Yanai, Integrating microarray-based spatial transcriptomics and

single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**, 333–342 (2020).
18. S. Valle, S. Alcalá, L. Martin-Hijano, P. Cabezas-Sáinz, D. Navarro, E. R. Muñoz, L. Yuste, K. Tiwary, K. Walter, L. Ruiz-Cañas, M. Alonso-Nocelo, J. A. Rubiolo, E. González-Arnay, C. Heeschen, L. Garcia-Bermejo, P. C. Hermann, L. Sánchez, P. Sancho, M. Á. Fernández-Moreno, B. Sainz Jr., Exploiting oxidative phosphorylation to promote the stem and immunoevasive properties of pancreatic cancer stem cells. *Nat. Commun.* **11**, 5265 (2020).
19. A. N. Hosein, R. A. Brekken, A. Maitra, Pancreatic cancer stroma: An update on therapeutic targeting strategies. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 487–505 (2020).
20. S. K. Longo, M. G. Guo, A. L. Ji, P. A. Khavari, Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat. Rev. Genet.* **22**, 627–644 (2021).
21. S. Jin, C. F. Guerrero-Juarez, L. Zhang, I. Chang, R. Ramos, C.-H. Kuan, P. Myung, M. V. Plikus, Q. Nie, Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
22. T. N. Schumacher, D. S. Thommen, Tertiary lymphoid structures in cancer. *Science* **375**, eabf9419 (2022).
23. W. H. Fridman, M. Meylan, G. Pupier, A. Calvez, I. Hernandez, C. Sautès-Fridman, Tertiary lymphoid structures and B cells: An intratumoral immunity cycle. *Immunity* **56**, 2254–2269 (2023).
24. R. Wu, W. Guo, X. Qiu, S. Wang, C. Sui, Q. Lian, J. Wu, Y. Shan, Z. Yang, S. Yang, T. Wu, K. Wang, Y. Zhu, S. Wang, C. Liu, Y. Zhang, B. Zheng, Z. Li, Y. Zhang, S. Shen, Y. Zhao, W. Wang, J. Bao, J. Hu, X. Wu, X. Jiang, H. Wang, J. Gu, L. Chen, Comprehensive analysis of spatial architecture in primary liver cancer. *Sci. Adv.* **7**, eabg3750 (2021).
25. F. Petitprez, A. de Reyniès, E. Z. Keung, T. W.-W. Chen, C.-M. Sun, J. Calderaro, Y.-M. Jeng, L.-P. Hsiao, L. Lacroix, A. Bougoüin, M. Moreira, G. Lacroix, I. Natario, J. Adam, C. Lucchesi, Y. H. Laizet, M. Toulmonde, M. A. Burgess, V. Bolejack, D. Reinke, K. M. Wani, W.-L. Wang, A. J. Lazar, C. L. Roland, J. A. Wargo, A. Italiano, C. Sautès-Fridman, H. A. Tawbi, W. H. Fridman, B cells are associated with survival and immunotherapy response in sarcoma. *Nature* **577**, 556–560 (2020).
26. I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth II, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J.-R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. K. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A.-C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jané-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K. Shalek, A. Regev, L. A. Garraway, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
27. T. Shang, T. Jiang, T. Lu, H. Wang, X. Cui, Y. Pan, M. Xu, M. Pei, Z. Ding, X. Feng, Y. Lin, X. Li, Y. Tan, F. Feng, H. Dong, H. Wang, L. Dong, Tertiary lymphoid structures predict the prognosis and immunotherapy response of cholangiocarcinoma. *Front. Immunol.* **14**, 1166497 (2023).
28. S.-Y. Wu, S.-W. Zhang, D. Ma, Y. Xiao, Y. Liu, L. Chen, X.-Q. Song, X.-Y. Ma, Y. Xu, W.-J. Chai, X. Jin, Z.-M. Shao, Y.-Z. Jiang, CCL19$^{+}$ dendritic cells potentiate clinical benefit of anti-PD-(L)1 immunotherapy in triple-negative breast cancer. *Med* **4**, 373–393.e8 (2023).
29. M. Liu, R. Cai, T. Wang, X. Yang, M. Wang, Z. Kuang, Y. Xie, J. Zhang, Y. Zheng, LAMC2 promotes the proliferation of cancer cells and induce infiltration of macrophages in non-small cell lung cancer. *Ann. Transl. Med.* **9**, 1392 (2021).
30. A. T. Ruffin, A. R. Cillo, T. Tabib, A. Liu, S. Onkar, S. R. Kunning, C. Lampenfeld, H. I. Atiya, I. Abecassis, C. H. L. Kürten, Z. Qi, R. Soose, U. Duvvuri, S. Kim, S. Oesterrich, R. Lafyatis, L. G. Coffman, R. L. Ferris, D. A. A. Vignali, T. C. Bruno, B cell signatures and tertiary lymphoid structures contribute to outcome in head and neck squamous cell carcinoma. *Nat. Commun.* **12**, 3349 (2021).
31. Y. Sato, K. Silina, M. van den Broek, K. Hirahara, M. Yanagita, The roles of tertiary lymphoid structures in chronic diseases. *Nat. Rev. Nephrol.* **19**, 525–537 (2023).
32. C. Sautès-Fridman, F. Petitprez, J. Calderaro, W. H. Fridman, Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat. Rev. Cancer* **19**, 307–325 (2019).
33. W. Xu, C. Ma, W. Liu, A. Anwaier, X. Tian, G. Shi, Y. Qu, S. Wei, H. Zhang, D. Ye, Prognostic value, DNA variation and immunologic features of a tertiary lymphoid structure-related chemokine signature in clear cell renal cell carcinoma. *Cancer Immunol. Immunother.* **71**, 1923–1935 (2022).
34. A. Andersson, L. Larsson, L. Stenbeck, F. Salmén, A. Ehinger, S. Z. Wu, G. Al-Eryani, D. Roden, A. Swarbrick, Å. Borg, J. Frisén, C. Engblom, J. Lundeberg, Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat. Commun.* **12**, 6012 (2021).
35. M. Meylan, F. Petitprez, E. Becht, A. Bougoüin, G. Pupier, A. Calvez, I. Giglioli, V. Verkarre, G. Lacroix, J. Verneau, C.-M. Sun, P. Laurent-Puig, Y.-A. Vano, R. Elaïdi, A. Méjean, R. Sanchez-Salas, E. Barret, X. Cathelineau, S. Oudard, C.-A. Reynaud, A. de Reyniès, C. Sautès-Fridman, W. H. Fridman, Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. *Immunity* **55**, 527–541.e5 (2022).
36. L. Wu, J. Yan, Y. Bai, F. Chen, X. Zou, J. Xu, A. Huang, L. Hou, Y. Zhong, Z. Jing, Q. Yu, X. Zhou, Z. Jiang, C. Wang, M. Cheng, Y. Ji, Y. Hou, R. Luo, Q. Li, L. Wu, J. Cheng, P. Wang, D. Guo, W. Huang, J. Lei, S. Liu, Y. Yan, Y. Chen, S. Liao, Y. Li, H. Sun, N. Yao, X. Zhang, S. Zhang, X. Chen, Y. Yu, Y. Li, F. Liu, Z. Wang, S. Zhou, H. Yang, S. Yang, X. Xu, L. Liu, Q. Gao, Z. Tang, X. Wang, J. Wang, J. Fan, S. Liu, X. Yang, A. Chen, J. Zhou, An invasive zone in human liver cancer identified by Stereo-seq promotes hepatocyte-tumor cell crosstalk, local immunosuppression and tumor progression. *Cell Res.* **33**, 585–603 (2023).
37. R. Arora, C. Cao, M. Kumar, S. Sinha, A. Chanda, R. McNeil, D. Samuel, R. K. Arora, T. W. Matthews, S. Chandarana, R. Hart, J. C. Dort, J. Biernaskie, P. Neri, M. D. Hyrcza, P. Bose, Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. *Nat. Commun.* **14**, 5029 (2023).
38. A. Filliol, Y. Saito, A. Nair, D. H. Dapito, L.-X. Yu, A. Ravichandra, S. Bhattacharjee, S. Affo, N. Fujiwara, H. Su, Q. Sun, T. M. Savage, J. R. Wilson-Kanamori, J. M. Caviglia, L. K. Chin, D. Chen, X. Wang, S. Caruso, J. K. Kang, A. D. Amin, S. Wallace, R. Dobie, D. Yin, O. M. Rodriguez-Fiallos, C. Yin, A. Mehal, B. Izar, R. A. Friedman, R. G. Wells, U. B. Pajvani, Y. Hoshida, H. E. Remotti, N. Arpaia, J. Zucman-Rossi, M. Karin, N. C. Henderson, I. Tabas, R. F. Schwabe, Opposing roles of hepatic stellate cell subpopulations in hepatocarcinogenesis. *Nature* **610**, 356–365 (2022).
39. S. Deng, D. Cheng, J. Wang, J. Gu, Y. Xue, Z. Jiang, L. Qin, F. Mao, Y. Cao, K. Cai, MYL9 expressed in cancer-associated fibroblasts regulate the immune microenvironment of colorectal cancer and promotes tumor progression in an autocrine manner. *J. Exp. Clin. Cancer Res.* **42**, 294 (2023).
40. D. Lavie, A. Ben-Shmuel, N. Erez, R. Scherz-Shouval, Cancer-associated fibroblasts in the single-cell era. *Nat. Cancer.* **3**, 793–807 (2022).
41. Y. Zhou, S. Bian, X. Zhou, Y. Cui, W. Wang, L. Wen, L. Guo, W. Fu, F. Tang, Single-cell multiomics sequencing reveals prevalent genomic alterations in tumor stromal cells of human colorectal cancer. *Cancer Cell* **38**, 818–828.e5 (2020).
42. Y. Hao, T. Stuart, M. H. Kowalski, S. Choudhary, P. Hoffman, A. Hartman, A. Srivastava, G. Molla, S. Madad, C. Fernandez-Granda, R. Satija, Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
43. F. Wang, J. Long, L. Li, Z.-X. Wu, T.-T. da, X.-Q. Wang, C. Huang, Y.-H. Jiang, X.-Q. Yao, H.-Q. Ma, Z.-X. Lian, Z.-B. Zhao, J. Cao, Single-cell and spatial transcriptome analysis reveals the cellular heterogeneity of liver metastatic colorectal cancer. *Sci. Adv.* **9**, eadf5464 (2023).
44. F. Wang, J. Long, L. Li, Z.-X. Wu, T.-T. Da, X.-Q. Wang, C. Huang, Y.-H. Jiang, X.-Q. Yao, H.-Q. Ma, Z.-X. Lian, Z.-B. Zhao, J. Cao, Emerging roles of lipid metabolism in cancer metastasis. *Mol. Cancer* **16**, 76 (2017).
45. Z. Li, H. Zhang, Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. *Cell. Mol. Life Sci.* **73**, 377–392 (2016).
46. H. Zhao, L. Wu, G. Yan, Y. Chen, M. Zhou, Y. Wu, Y. Li, Inflammation and tumor progression: Signaling pathways and targeted intervention. *Signal Transduct. Target. Ther.* **6**, 263 (2021).
47. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck III, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
48. S. M. Weis, D. A. Cheresh, Tumor angiogenesis: Molecular pathways and therapeutic targets. *Nat. Med.* **17**, 1359–1370 (2011).
49. L. Shang, X. Zhou, Spatially aware dimension reduction for spatial transcriptomics. *Nat. Commun.* **13**, 7203 (2022).
50. Z. Xun, X. Ding, Y. Zhang, B. Zhang, S. Lai, D. Zou, J. Zheng, G. Chen, B. Su, L. Han, Y. Ye, Reconstruction of the tumor spatial microenvironment along the malignant-boundary-nonmalignant axis. *Nat. Commun.* **14**, 933 (2023).
51. G. Palla, H. Spitzer, M. Klein, D. Fischer, A. C. Schaar, L. B. Kuemmerle, S. Rybakov, I. L. Ibarra, O. Holmberg, I. Virshup, M. Lotfollahi, S. Richter, F. J. Theis, Squidpy: A scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178 (2022).
52. Y. Long, K. S. Ang, M. Li, K. L. K. Chong, R. Sethi, C. Zhong, H. Xu, Z. Ong, K. Sachaphibulkij, A. Chen, L. Zeng, H. Fu, M. Wu, L. H. K. Lim, L. Liu, J. Chen, Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun.* **14**, 1155 (2023).
53. J. Qian, J. Liao, Z. Liu, Y. Chi, Y. Fang, Y. Zheng, X. Shao, B. Liu, Y. Cui, W. Guo, Y. Hu, H. Bao, P. Yang, Q. Chen, M. Li, B. Zhang, X. Fan, Reconstruction of the cell pseudo-space from single-cell RNA sequencing data with scSpace. *Nat. Commun.* **14**, 2484 (2023).
54. S. Nayar, J. Campos, C. G. Smith, V. Iannizzotto, D. H. Gardner, F. Mourcin, D. Roulois, J. Turner, M. Sylvestre, S. Asam, B. Glaysher, S. J. Bowman, D. T. Fearon, A. Filer, K. Tarte, S. A. Luther, B. A. Fisher, C. D. Buckley, M. C. Coles, F. Barone, Immunofibroblasts are pivotal drivers of tertiary lymphoid structure formation and local pathology. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 13490–13497 (2019).
55. Y. Chen, K. M. McAndrews, R. Kalluri, Clinical and therapeutic relevance of cancer-associated fibroblasts. *Nat. Rev. Clin. Oncol.* **18**, 792–804 (2021).
56. S. He, R. Bhatt, C. Brown, E. A. Brown, D. L. Buhr, K. Chantranuvatana, P. Danaher, D. Dunaway, R. G. Garrison, G. Geiss, M. T. Gregory, M. L. Hoang, R. Khafizov, E. E. Killingbeck, D. Kim, T. K. Kim, Y. Kim, A. Klock, M. Korukonda, A. Kutchma, Z. R. Lewis, Y. Liang, J. S. Nelson, G. T. Ong, E. P. Perillo, J. C. Phan, T. Phan-Everson, E. Piazza, T. Rane, Z. Reitz, M. Rhodes, A. Rosenbloom, D. Ross, H. Sato, A. W. Wardhani, C. A. Williams-Wietzikoski, L. Wu, J. M. Beechem, High-plex imaging of RNA and proteins

at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol.* **40**, 1794–1806 (2022).
57. V. De Oliveira, Bayesian analysis of conditional autoregressive models. *Ann. Inst. Stat. Math.* **64**, 107–133 (2012).
58. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
59. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
60. S. Hänzelmann, R. Castelo, J. Guinney, GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
61. Z. Tang, B. Kang, C. Li, T. Chen, Z. Zhang, GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **47**, W556–W560 (2019).
62. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
63. Y. Zhang, B. Yu, W. Ming, D. Chen, Codes for the paper "SpaTopic: A StatisticalLearning Framework for Exploring Tumor Spatial Architecture from Spatially Resolved Transcriptomic Data", version v1.0.0, Zenodo (2024); https://doi.org/10.5281/zenodo.13283829.
64. Y. Sun, L. Wu, Y. Zhong, K. Zhou, Y. Hou, Z. Wang, Z. Zhang, J. Xie, C. Wang, D. Chen, Y. Huang, X. Wei, Y. Shi, Z. Zhao, Y. Li, Z. Guo, Q. Yu, L. Xu, G. Volpe, S. Qiu, J. Fan, Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. *Cell* **184**, 404–421.e16 (2021).

**Acknowledgments:** We acknowledge the Center for Information Technology and the High Performance Computing Center of Nanjing University for providing high-performance computing resources. **Funding:** This work is supported by the National Natural Science Foundation of China (no. 32070656) to D.C. **Author contributions:** Conceptualization: D.C. Resources: D.C. and J.W. Supervision: D.C. Funding acquisition: D.C. Software: Y.Z. and B.Y. Formal analysis: Y.Z., W.M., B.Y., and X.Z. Methodology: D.C. and Y.Z. Writing—original draft: D.C., Y.Z., and X.Z. Writing—review and editing: D.C., J.W., Y.Z., W.M., and B.Y. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The source code in R for SpaTopic is freely available at the Zenodo repository (*63*) or Github (https://github.com/compbioNJU/SpaTopic). All datasets used in this study are publicly available. The DLPFC spatial and single-cell transcriptome data were downloaded from https://github.com/LieberInstitute/spatialLIBD. The BRCA data (including SRT and single-cell transcriptome data) were obtained from the 10x Genomics Visium website (10x Genomics): https://support.10xgenomics.com/spatial-gene-expression/. The OV data were downloaded from https://10xgenomics.com/resources/datasets/human-ovarian-cancer-whole-transcriptome-analysis-stains-dapi-anti-pan-ck-anti-cd-45-1-standard-1-2-0 (for SRT data) and https://lambrechtslab.sites.vib.be/en/high-grade-serous-tubo-ovarian-cancer-refined-single-cell-rna-sequencing-specific-cell-subtypes (for single-cell data). The PLC SRT data were collected from Wu *et al.* (*24*) and a matched scRNA-seq dataset (*64*) for cell-type annotation. Note that all cell-type annotation is based on the corresponding studies to better explain our analysis results. All other data (such as TLS markers and cell marker genes) were obtained from the original paper. The NSCLC (*56*) from the NanoString CosMx spatial transcriptomics platform were downloaded from https://nanostring.com/products/cosmx-spatial-molecular-imager/ffpe-dataset/nsclc-ffpe-dataset/. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 28 March 2024
Accepted 21 August 2024
Published 27 September 2024
10.1126/sciadv.adp4942