RESEARCH ARTICLE

Same-Slide Spatial Multiomics Integration with IN-DEPTH Reveals Tumor Virus–Linked Spatial Reorganization of the Tumor Microenvironment

Stephanie Pei Tung Yiu[1], Yuzhou Chang[1,2,3], Yao Yu Yeo[1,4], Huaying Qiu[1], Wenrui Wu[1], Hendrik A. Michel[1,4], Xiaojie Jin[2,3], Rongting Huang[5], Shoko Kure[6], Lindsay Parmelee[1], Shuli Luo[1], Precious Cramer[1], Jia Le Lee[1], Yang Wang[1], Zhangxin Zhao[1], Jason Yeung[1], Nourhan El Ahmar[7], Berkay Simsek[7], Razan Mohanna[7], McKayla Van Orden[8], Wesley S. Lu[8], Kenneth J. Livak[8], Shuqiang Li[8], Ce Gao[9], Melinda Burgess[10], Colm Keane[10], Jahanbanoo Shahryari[5], Leandra G. Kingsley[5], Reem N. Al-Humadi[5], Sahar Nasr[5], Dingani Nkosi[7], Sam Sadigh[7], Philip Rock[11], Leonie Frauenfeld[12], Louisa Kaufmann[12], Bokai Zhu[13], Ankit Basak[13], Nagendra Dhanikonda[1], Chi Ngai Chan[1], Jordan Krull[2,3], Ye Won Cho[14], Chia-Yu Chen[14], Jonathan Brown[15], Hongbo Wang[1,16], Bo Zhao[16], Jia-Ying Joey Lee[17], Lit-Hsin Loo[17], David M. Kim[14], Vassiliki A. Boussiotis[18], Baochun Zhang[6], Kevin Wei[9], Alex K. Shalek[13], Brooke E. Howitt[5], Sabina Signoretti[7], Christian M. Schürch[12,19], F. Stephen Hodi[6], W. Richard Burack[11], Scott J. Rodig[7], Qin Ma[2,3], and Sizun Jiang[1,4,7,13,20]



ABSTRACT Spatial transcriptomics and proteomics have enabled profound insights into tissue organization, yet these technologies remain largely disparate, and emerging same-slide multiomics approaches are limited in plex, spatial resolution, signal retention, and integrative analytics. We introduce IN-situ DEtailed Phenotyping To High-resolution transcriptomics (IN-DEPTH), a streamlined, resource-efficient, commercially compatible workflow using single-cell spatial proteomics-derived imaging to guide transcriptomic capture on the same slide without RNA signal loss. To integrate modalities beyond niche-level mapping, we developed Spectral Graph Cross-Correlation (SGCC), a proteomic–transcriptomic framework resolving spatially coordinated functional state changes across interacting cell populations. Applied to diffuse large B-cell lymphoma (DLBCL), IN-DEPTH and SGCC enabled stepwise discovery from Epstein–Barr virus (EBV)–positive and EBV-negative tumor comparisons with single-cell resolution, revealing coordinated tumor–macrophage–CD4 T-cell remodeling, immunosuppressive C1Q macrophage enrichment, CD4 T-cell dysfunction, and a candidate IL27–STAT3 signaling axis. Collectively, IN-DEPTH enables scalable spatial multiomics to uncover clinically relevant microenvironmental mechanisms and toward robust spatial multimodal AI models.

SIGNIFICANCE: IN-DEPTH enables same-slide spatial multiomics across commercial platforms via a protein-first strategy preserving protein epitopes, RNA quality, and tissue integrity. Coupled with SGCC, it resolves coordinated spatial immune remodeling, revealing EBV/LMP1-driven C1Q macrophage polarization and CD4 T-cell dysfunction in DLBCL, with broad applicability to other diseases.

INTRODUCTION

Spatial transcriptomics and spatial proteomics are recent technological breakthroughs that have enabled investigations of complex biological systems at unprecedented detail within native tissue contexts (1–4). Effective combination of both approaches on the same tissue section is currently the rate-limiting step for novel biological insights, particularly given the complementary strengths of assessing both RNA and proteins. Although spatial transcriptomics offers higher feature coverage and pathway-level insights, the technology faces inherent biological limitations in predicting functional outcomes due to posttranscriptional regulation and variable RNA-to-protein correlations (5–7), whereas spatial proteomics directly captures functional molecular phenotypes and functional states with high signal-to-noise ratios and data acquisition speeds, albeit with lower multiplexing capacity. Spatial multiomics methods that can simultaneously profile both transcripts and proteins from the same tissue section would enable insights into regulatory mechanisms while preserving spatial context to bridge the gap between gene expression and functional protein dynamics in complex biological systems and archival clinical specimens.

[1]Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts. [2]Pelotonia Institute for Immuno-Oncology, The Ohio State University, Columbus, Ohio. [3]Department of Biomedical Informatics, College of Medicine, Ohio State University, Columbus, Ohio. [4]Program in Virology, Division of Medical Sciences, Harvard Medical School, Boston, Massachusetts. [5]Department of Pathology, Stanford University, Stanford, California. [6]Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts. [7]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts. [8]Translational Immunogenomics Lab, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts. [9]Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts. [10]Frazer Institute, Translational Research Institute, University of Queensland, Brisbane, Australia. [11]Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, Rochester, New York. [12]Department of Pathology and Neuropathology, University Hospital and Comprehensive Cancer Center Tübingen, Tübingen, Germany. [13]Broad Institute of Harvard and MIT, Cambridge, Massachusetts. [14]Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Boston, Massachusetts. [15]Department of Mathematics, University of Dayton, Dayton, Ohio. [16]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts. [17]Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. [18]Department of Hematology Oncology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts. [19]Cluster of Excellence iFIT (EXC 2180) "Image-Guided and Functionally Instructed Tumor Therapies", University of Tübingen, Tübingen, Germany. [20]Department of Pathology, Dana Farber Cancer Institute, Boston, Massachusetts.

S.P.T. Yiu, Y. Chang, Y.Y. Yeo, H. Qiu, and W. Wu are co-first authors.

Q. Ma and S. Jiang are senior authors.

Corresponding Author: Sizun Jiang, Center for Life Science, 3 Blackfan Cir 1034, Boston, MA 02115. E-mail: sjiang3@bidmc.harvard.edu

Cancer Discov 2026;16:1611–48

doi: 10.1158/2159-8290.CD-25-0775

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2026 The Authors; Published by the American Association for Cancer Research

Several innovative approaches have successfully demonstrated the potential of integrating spatial protein and RNA imaging on the same tissue sample (bioRxiv 2023.10.27.564191; refs. 8–13). Although these pioneering methods have provided valuable insights, current technical constraints, such as multiplexing capacity (bioRxiv 2023.10.27.564191; refs. 8, 10, 11, 14) and spatial resolution in grid/spot-based approaches (8, 9, 12, 13), suggest opportunities for further advancements. Spatial transcriptomics approaches also often incorporate protease treatment of tissue sections for efficient RNA detection, which will compromise protein epitope integrity and affect downstream protein analysis (https://dx.doi.org/10.17504/protocols.io.q26g71rwqgwz/v1; refs. 10, 14). An additional key limitation for broad clinical application and adoption is the compatibility with formalin-fixed, paraffin-embedded (FFPE) tissues, the standard preservation method in clinical pathology (15). There is also significant potential to expand and enhance complementary computational approaches to fully empower multimodal analysis for meaningful biological insights (16).

We herein present IN-situ DEtailed Phenotyping To High-resolution transcriptomics (IN-DEPTH), a cost-efficient, commercial platform-compatible, and robust spatial multiomics approach that utilizes single-cell spatial proteomics to guide subsequent targeted or genome-wide spatial transcriptomics capture on the same slide without compromising protein or RNA signals. IN-DEPTH advances our conceptual approach of spatial multiomics data generation by linking spatially resolved proteomic characterization of cellular context and tissue architecture with transcriptomic interrogation of functional pathways in a biologically relevant manner. To quantify tissue spatially linked transcriptomic pathways revealed by IN-DEPTH, we developed Spectral Graph Cross-Correlation (SGCC) to determine spatial covarying relationships between cell pairs using an unbiased graph signal representation method (17, 18). Here, the spatial arrangement and pattern of each cell phenotype is a graph signal in which cells serve as nodes, spatial patterns are node attributes, and spatial distances are edges. This allows an unbiased representation of spatial patterns of each cell population on tissues through spectral graph signals to resolve underlying spatial relationships between cell types and their coordinated functional states.

We demonstrate the broad applicability of IN-DEPTH across multiple commercially available spatial platforms and highlight the combination of IN-DEPTH and SGCC to accurately identify human tonsil multimodal features at global and local scales. We further demonstrate the synergistic potential of IN-DEPTH and SGCC to uncover novel biological insights into the impact of the prototypic tumor virus Epstein–Barr virus (EBV) on the diffuse large B-cell lymphoma (DLBCL) tumor microenvironment (TME) and immune dysregulation. Leveraging proteomics-guided, same-slide spatial multiomics integration, our stepwise analyses progress from multicellular comparisons of EBV-positive and EBV-negative tumors to single-cell resolution, revealing coordinated spatial remodeling of tumor, macrophage, and CD4 T-cell functional states. This approach identifies enrichment of immunosuppressive C1Q-associated macrophage polarization coupled with CD4 T-cell dysfunction and further localizes these programs to tumor cells exhibiting active viral oncoprotein LMP1 signaling. Integration of spatial context with transcriptomic pathway analysis additionally implicates the IL27–STAT3 signaling axis as a potential mediator of this immunosuppressive niche. Together, these findings illustrate how viral-driven tumor-intrinsic programs can spatially rewire immune interactions within the TME and highlight the value of IN-DEPTH and SGCC for revealing mechanistically interpretable interaction networks that may inform targeted therapeutic strategies in virus-associated malignancies.

RESULTS

IN-DEPTH Combines Antibody Staining and RNA Probe Hybridization on the Same Slide While Retaining Protein and RNA Quality

IN-DEPTH leverages high-dimensional spatial proteomics to provide initial cellular and architectural context that guides subsequent spatial transcriptomics capture, either through phenotype-informed region of interest (ROI) selection or direct image-guided targeting across other platforms (Fig. 1A). This streamlined approach ensures the biological relevance of spatial transcriptomics by tying it to spatial proteomics-guided identification of tissue ROI, thus reducing the resource-intense cost and time barriers associated with spatial transcriptomics of whole slides, while retaining high sensitivity (Supplementary Fig. S1A and S1B). Given the impact of the protease digestion step during spatial transcriptomics on subsequent protein imaging (https://dx.doi.org/10.17504/protocols.io.q26g71rwqgwz/v1; refs. 10, 14), we postulated that performing spatial proteomics first before transcriptomics will circumvent this challenge. As various spatial proteomics platforms also differ in recommended tissue retrieval conditions, we first implemented a standardized heat-induced epitope retrieval step at 97°C for 20 minutes using a pH 9.0 retrieval buffer followed by a 1-hour photobleaching step, optimized across our prior experiments (bioRxiv 2024.03.05.583586; refs. 10, 11, 19).

To systematically evaluate the feasibility of integrating spatial proteomics with transcriptomics within a generalizable framework, we focused on four multiplexed immunofluorescence-based spatial proteomics platforms [CODEX (20), SignalStar (21), Polaris (22), and Orion (23)] due to their demonstrated utility in clinical applications, preservation of tissue integrity, rapid whole-slide imaging capabilities, and complementary protein labeling strategies.

These platforms represent diverse methodologies including cyclic immunofluorescence, signal amplification, and spectral deconvolution, providing a diverse foundation for method development. We also selected representative spatial transcriptomics platforms [GeoMx (8), VisiumHD (24), CosMx (25), and Xenium (26)], encompassing complementary transcriptomic capture strategies, including pseudobulk region-based profiling, high-resolution spot-based capture, and single-cell imaging-based transcript detection using either direct probe hybridization or signal amplification strategies. These platforms were selected based on their broad availability both within and beyond our laboratories, using stringently optimized protocols to ensure experimental compatibility across modalities (see "Methods").

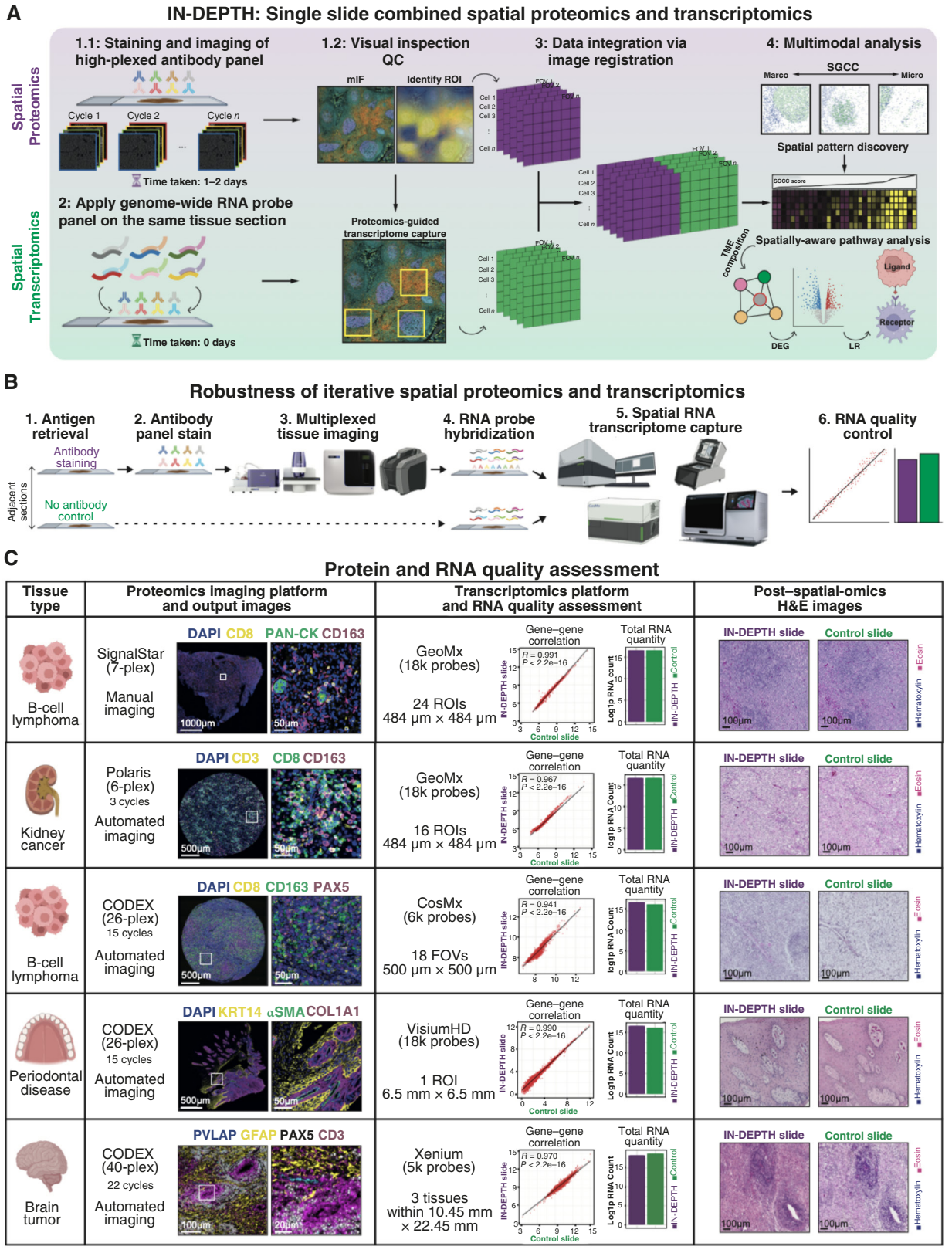


Figure 1. IN-DEPTH combines spatial proteomics and transcriptomics on the same slide without loss of protein or RNA quality. **A,** Schematic overview of IN-DEPTH, in which spatial proteomics was used to guide cell type–specific, genome-wide transcriptomic capture on the same slide. **B,** Experimental outline to assess the impact of spatial proteomics workflow on RNA capture, with an adjacent tissue section processed for RNA capture only as a control. **C,** Assessment of tissue imaging and RNA capture quality after IN-DEPTH. (*continued on following page*)

To determine if prior spatial proteomics on tissue samples affects downstream RNA signal recovery, we first compared the spatial transcriptome signal of adjacent tissue slides, in which one slide was subjected to IN-DEPTH (spatial proteomics followed by spatial transcriptomics), whereas the other slide was only subjected to the corresponding spatial transcriptomics platform as a control (Fig. 1B). Both slides subsequently underwent hematoxylin and eosin (H&E) staining to assess the retention of tissue morphology. In our initial proof of concept, we applied CODEX–GeoMx IN-DEPTH to FFPE tonsil tissues and observed robust antibody staining. Relative to the control slide, IN-DEPTH samples demonstrated high gene-to-gene concordance (R = 0.938) and comparable total RNA capture [Supplementary Fig. S1C (row 1); Supplementary Table S1]. We next demonstrated the platform flexibility of the CODEX workflow by performing manual stripping and detection by oligo hybridization (20, 27), followed by whole-slide imaging using the GeoMx slide scanner functionality and subsequent RNA recovery, achieving highly concordant RNA signal preservation [R = 0.952; Supplementary Fig. S1C (row 2)].

We next expanded upon these initial IN-DEPTH results across multiple combinations of spatial proteomics and spatial transcriptomics platforms using a variety of FFPE tissue samples, including lymphoma, kidney cancer, periodontal disease, brain cancer, and uterine cancer. Across combinations, we observed consistently high gene-to-gene concordance (R >0.94), comparable total transcript recovery, and uniformly strong sequencing and transcriptomic quality metrics, including Q30 scores, mean transcripts per cell, and reads mapped to probe sets, between IN-DEPTH and control slides (Fig. 1C; Supplementary Fig. S1C; Supplementary Table S2), except the Orion–GeoMx combination, which showed lower gene-level concordance [R = 0.692; Supplementary Fig. S1C (row 5)]. H&E staining quality was preserved across all conditions, with IN-DEPTH slides demonstrating equal or improved staining performance relative to control slides in several platform combinations.

To further evaluate the robustness of IN-DEPTH and assess tissue integrity during extended imaging, we determined the maximum number of imaging cycles that could be performed without compromising RNA quality. CODEX experiments were conducted on brain tissue sections using the PhenoCycler-Fusion platform and a 40-plex antibody panel, with each marker imaged twice using either the Cy3 or Cy5 channels. Plasmalemma vesicle–associated protein (PLVAP), a vascular permeability marker, was included in every cycle and imaged in the Cy7 channel to monitor staining consistency. The experiment was carried out to the instrument's maximum physical limit of 60 imaging cycles. Following CODEX imaging, the same slide was processed with VisiumHD, whereas an adjacent tissue section was processed in parallel with VisiumHD alone as a control [Supplementary Fig. S1C (row 6)]. Protein staining remained stable throughout the experiment, with minimal detectable changes in signal patterns by both visual and quantitative assessment. This stability was supported by consistent relative target registration error (rTRE) scores, reflecting preserved image alignment accuracy, and stable Pearson correlation coefficients (PCC), indicating sustained marker intensity similarity (Supplementary Fig. S2A–S2C). Mean PLVAP signal intensity also remained consistent across imaging cycles (Supplementary Fig. S2D), further supporting preserved tissue integrity during prolonged imaging. Consistent H&E morphology between IN-DEPTH and control slides further confirmed the absence of tissue degradation [Supplementary Fig. S1C (row 6)]. For RNA quality, although a slight reduction in total RNA counts was observed relative to the VisiumHD-only control, we nonetheless observed that a strong gene-to-gene correlation (R = 0.966) was maintained [Supplementary Fig. S1C (row 6)], demonstrating that extended multiplexed imaging minimally affects transcriptomic integrity.

Having established that the sequential order of spatial proteomics followed by spatial transcriptomics maintained RNA quality, we next evaluated the impact of reversing this workflow order on protein signal preservation. We compared staining quality and quantified protein marker intensities between tissues subjected to CODEX followed by Xenium (CODEX-Xenium) and those performed in the reverse order [Xenium-CODEX; Fig. 1C (row 5); Supplementary Fig. S1C (row 7)]. Consistent with other IN-DEPTH combinations, we observed comparable total RNA quantities and robust gene-to-gene correlations (R = 0.970) between the two approaches. However, the Xenium-CODEX samples exhibited a marked reduction in mean staining intensities across most markers relative to the CODEX-Xenium slide (Supplementary Fig. S2E and S2F), confirming that performing spatial proteomics first preserves both protein signal quality and RNA integrity, establishing IN-DEPTH as a viable framework for integrated same-slide spatial multiomics.

These data collectively demonstrate the robustness of spatial protein and RNA signals with IN-DEPTH while allowing user flexibility for cross-platform and region-specific RNA capture. We also validated the segmentation module (see "Methods") across diverse tissue types, including immune-rich (tonsil, B-cell lymphoma) and structurally complex tissues (periodontal and brain), ensuring reliable performance beyond circular, immune-based morphologies (Supplementary Fig. S2G). Among the validated platform combinations, we selected CODEX–GeoMx for subsequent IN-DEPTH development due to its compatibility with FFPE tissues, rapid whole-slide imaging, reproducible multiplexed protein detection (20, 27, 28), and automated whole-transcriptome profiling with high regional selectivity and efficient processing (Supplementary Fig. S1B; ref. 8). This combination provided a robust platform for further methodologic and biological validation.

Figure 1. ***(Continued)*** Each row represents a different combination of spatial platforms evaluated, along with the corresponding tissue type, number of imaging markers and cycles, probe plex, and number of capture regions. Each column (from left to right) shows multiplexed protein images, RNA quality assessment results, and H&E images. Key RNA and sequencing-related metrics for each ROI across all combinations are summarized in Supplementary Table S2. Detailed experimental procedures are described in "Methods," and step-by-step protocols for all IN-DEPTH spatial proteomics and transcriptomics combinations presented in this study are available at https://sizunjianglab.github.io/IN-DEPTH/. mIF, Multiplexed Immunofluorescence

IN-DEPTH Enables Reproducible and Robust Spatial Multiomics Profiling and Reveals Functional Cell States within the Native Tissue Architecture

We next performed IN-DEPTH (CODEX–GeoMx) on two adjacent FFPE tonsil sections, with each section undergoing RNA capture on two independent GeoMx instruments to assess for technical reproducibility. We applied a 12-plex antibody panel consisting of cell phenotyping markers on both slides (Supplementary Table S1) and imaged them in parallel on the PhenoCycler-Fusion system (Fig. 2A; Supplementary Fig. S3A). We then performed cell segmentation and phenotyping for 11 cell populations using the background-subtracted images acquired from the PhenoCycler-Fusion (Fig. 2B).

To capture cell type–specific transcriptomes, we imported these cell type–specific masks onto the GeoMx for custom spatial transcriptome capture using the human whole-transcriptome atlas library consisting of >18,000 targets in the human genome. We selected 16 paired and continuous 660 × 760 μm rectangular ROIs on each adjacent tissue replicate slide that include B follicles and T-cell zones (Supplementary Fig. S3B; Supplementary Table S2). We first confirmed the specificity of our antibody panel and accuracy of spatial proteomics cell type annotation for both tissues (replicate 1, Fig. 2B; replicate 2, Supplementary Fig. S3C). All multiplex staining patterns and marker-based cell type annotations were reviewed by board-certified pathologists, with same-slide H&E histology used to verify tissue context and morphology as needed (replicate 1, Fig. 2B; replicate 2, Supplementary Fig. S3C).

We further evaluated the specificity and accuracy of cell phenotyping by confirming the expected enrichment of antibody marker expression across the 11 annotated cell populations [replicate 1, Fig. 2C (left); replicate 2, Supplementary Fig. S3D (left)]. We then orthogonally verified the spatial transcriptomics capture specificity by quantifying the enrichment of cell type–specific gene signatures for each cell population against a single-cell tonsil atlas [replicate 1, Fig. 2C (right); replicate 2, Supplementary Fig. S3D (right); refer to Supplementary Table S3 and "Methods" for details in gene signature curation; ref. 29]. We showed that these two technical replicates displayed the expected cell composition of tonsil tissues (Fig. 2B; Supplementary Fig. S3C), high consistency between the protein and transcriptome signatures (Supplementary Fig. S3D), gene-to-gene correlation (Supplementary Fig. S3E), total RNA capture (Supplementary Fig. S3F), and low signals from nontargeting negative control probes (Supplementary Fig. S3G). These results highlight the robust technical reproducibility of IN-DEPTH across different instruments.

We recognize that spatial proteomics-guided transcriptomes with IN-DEPTH are well suited to address the challenge of accurate real-world reference standards currently missing for deconvolution approaches (30–33). We demonstrate this application by systematically benchmarking the performances of common deconvolution algorithms CIBERSORT (34), dtangle (35), MuSic (36), and SpatialDecon (37) on our reference gene signatures curated from the single-cell tonsil atlas (Supplementary Table S3; see "Methods"; ref. 29). We observed that the results from CIBERSORT (34), dtangle (35), and MuSiC (36) were relatively consistent for the top three cell type components, that is, BCL6-positive B cells, BCL6-negative B cells, and CD4 T cells (Fig. 2D; Supplementary Fig. S3H). Ranking the tonsil ROIs by cell type proportion complexity, as estimated by the Gini–Simpson index applied to the cell type proportions acquired from the IN-DEPTH dataset (Fig. 2E), revealed that all four methods achieved high correlation (>0.9) with the IN-DEPTH dataset for ROIs of low complexity (e.g., ROIs 1, 2, 3, 4, 5, 9). Together, these results validate IN-DEPTH as a robust approach for generating high-quality, ROI-level spatial reference data suitable for evaluating deconvolution algorithms. Moreover, they provide practical guidance for selecting and optimizing computational methods tailored to specific tissue contexts and biological questions.

To demonstrate the utility of paired spatial proteomics and transcriptomics data from IN-DEPTH, we next examined the functional and spatial dynamics of lymphocytes in the tonsillar tissue. We focused on CD4 T follicular helper (Tfh) cells, which are known to migrate into B follicles (i.e., B cell–dense zone based on protein imaging) during the activation and maturation process [Fig. 2F (top); ref. 29]. Although Tfh cells can be easily identified from our CD4 T-cell population as spatially residing within B follicles, they cannot be confidently identified using the limited spatial proteomics markers alone, as our antibody panel did not include classic Tfh markers such as PD-1 or CXCR5 (Supplementary Table S1). To overcome this, we tested whether transcriptomic data could identify Tfh-like cells by evaluating Tfh gene signature enrichment (Supplementary Table S3) in CD4 T cells located inside versus outside follicles. Our results revealed significantly higher Tfh transcriptomic signatures for CD4 T cells in ROIs also containing dense B follicles [Fig. 2F (bottom left)]. Moreover, Tfh gene set variation analysis (GSVA) scores positively correlated with the proportion of B cells across all ROIs in both tissues replicates [$R = 0.75$; Fig. 2F (bottom right)], consistent with the known Tfh cell trafficking and maturation processes in the tonsil (29).

To further characterize tissue-wide, cell type–specific transcriptional programs, we applied consensus nonnegative matrix factorization (cNMF; ref. 38) and identified 10 predominant gene expression programs (GEP; Supplementary Table S4). These programs were annotated using Gene Ontology Biological Process (GOBP) signatures (Supplementary Table S4) and exhibited distributions consistent with known cellular functions (29). For example, $BCL6^+$ B cells were enriched for programs related to DNA modification and somatic hypermutation; T cells showed activation-associated signatures; endothelial cells were enriched for vascularization programs; $CD68^+$ macrophages showed ER stress responses; and epithelial-like cells from tonsillar crypts exhibited epithelial differentiation signatures (Fig. 2G).

Together, these results demonstrate that IN-DEPTH faithfully recapitulates canonical, spatially restricted biology while providing added value through precise cell type identification, inference of functional states beyond available protein markers, and elucidation of coordinated transcriptional programs across tissue compartments. Beyond descriptive biology, these data establish a high-quality spatial reference that supports and enhances downstream computational analyses, including cell deconvolution.

Coordinated Spatial Transitions in Cellular States and Tissue Organization

To investigate how spatial organization relates to cellular function and to maximize the utility of IN-DEPTH multiomics data, we developed SGCC, a mathematical formulation built upon graph signal processing approaches to analyze pairwise coordinated spatial patterns. SGCC leverages the unbiased representation and interpretability of the spectral graph wavelet transform (SGWT) to explore the distributional relationships between cell pairs. In our previous study (17), any spatial omics feature (e.g., cell phenotype labels) can be treated as a graph signal, in which the underlying graph can be lattice (a pixel graph with nodes representing pixels and edges defined by pixel-to-pixel distance) or irregular (a cell graph with nodes representing cells and edges defined by cell-to-cell distance). Subsequently, graph Fourier transform (GFT) is applied to project vertex-domain graph signals onto the frequency domain via Fourier modes (FM), yielding a set of interpretable Fourier coefficients (FC) across whole spectral domain. As low-frequency FMs capture spatially organized components of the graph signal (39, 40) and other non–low-frequency FMs capture local variation components, it lays the foundation of correlating pairwise cell phenotype in frequency domain by computing the similarity of these band-passed FCs using SGWT theory (see "Methods" and SGCC Supplementary Notes for more details).

SGCC quantitatively measures the spatial distributional relationships and underlying patterns between two cell phenotypes via the following three steps (Fig. 3A). First, by binning cell phenotypes from the cell graph into a pixel graph, all ROIs' FCs are placed within the same linear space, ensuring subsequent cross-correlation calculations that can be compared across samples. Second, the binned cell phenotype signals are processed using SGWT, which decomposes each graph signal into a low-pass (scaling) component and multiple band-pass (wavelet) components defined on the underlying spatial graph. The low-pass SGWT component captures broad-scale spatial organization, whereas the band-pass components encode finer-scale and localized spatial structure. Third, pairwise cosine similarities between cell phenotype signals are computed separately for the low- and band-pass SGWT coefficients and combined using energy-based weights, resulting in c(m, 2) pairwise comparisons, where m represents the number of cell phenotypes. These SGCC scores reflect the spatial distribution patterns between two cell types.

When multiple samples are available, SGCC can be treated as a continuous or ordinal variable serving as a spatial factor. At multicellular level, a negative SGCC value indicates reduced spatial co-occurrence, whereas a positive value indicates increased spatial co-occurrence between cell phenotypes. Consequently, SGCC can be used to predict genes covarying with spatial factors. For example, one can apply the ImpulseDE2 model (41) to treat SGCC as a continuous spatial variable or employ edgeR (42) to treat it as an ordinal spatial variable, thereby enabling the identification of spatially dynamic genes (Fig. 3B).

To examine graph signal representation by SGWT and benchmark SGCC with other methods, we simulated two datasets, each representing a 60 × 60-pixel graph, to create ring-like distributions of two cell phenotypes. These datasets systematically varied in area and complementarity, spanning global-to-local and distal-to-proximal relationships [Supplementary Fig. S4A (pattern 1 and pattern 2), respectively]. Our results demonstrate that the pattern variation of low-frequency FM is related to the underlying graph structure. With increasing neighborhood connectivity, the low-frequency FMs exhibit increased spectral smoothness (Supplementary Fig. S4B and S4C), and a relatively small number of low-frequency FMs suffices to capture the dominant large-scale spatial variation of the graph signal (Supplementary Fig. S4D and S4E). These observations are consistent with theoretical expectations that low-frequency FMs encode global spatial organization determined by graph structure (18). Furthermore, our results showed that introducing band-pass FM can improve the representation of spatial patterns. In the ablation setting with a constrained low-frequency basis ($k = 10$), reconstruction using SGWT band-pass components significantly outperformed reconstruction using low-frequency components ($k = 400$) alone (Supplementary Fig. S4F; $P = 4.32 \times 10^{-12}$). Together, these results indicate that effective spatial pattern representation requires both low-frequency structure and band-pass details and that SGWT provides a theoretical multiscale framework for jointly extracting global and local features in graph-based spatial data.

We then computed SGCC scores and compared them with classical spatial statistical methods including cross-variogram (43), Pearson correlation, bivariate Moran I (44), and local spatial cross-correlation index (45). SGCC uniquely and robustly discriminated spatial patterns, increasing under locally complementary patterns and decreasing under globally complementary patterns (Supplementary Fig. S4G). The same trend held in the spatial proximal–distal simulations, with SGCC scores decreasing in spatially distal and increasing in spatially proximal patterns (Supplementary Fig. S4H). By contrast, the other spatial statistical methods showed only subtle shifts with changing patterns, indicating that SGCC can more sensitively discriminate fine spatial differences.

We next demonstrated the applicability of SGCC to real-world IN-DEPTH data (CODEX–GeoMx, FFPE tonsil tissue acquired in Fig. 2) to resolve fine-grained state transitions between CD4 T cells and BCL6$^+$ B cells, which are key modulators of germinal center (GC) reactions (Fig. 3C). SGCC between these T and B cells uncovered coordinated shifts in tissue organization and transcriptional programs (Fig. 3C; Supplementary Table S4). B cells in regions with lower SGCC scores showed enrichment in pathways for DNA replication and DNA topologic change [Fig. 3C (left)]. In contrast, regions with higher SGCC scores were characterized by enrichment of cytokine production and TCR signaling in CD4 T cells and antigen processing and presentation pathways in B cells [Fig. 3C (right)]. These patterns recapitulate canonical dark-zone (DZ)/light-zone (LZ) polarization, with low-SGCC regions reflecting T- and B-cell self-aggregation consistent with DZ somatic hypermutation [Fig. 3C (left)], whereas high-SGCC regions exhibit increased T–B cross-talk characteristics of LZ selection [Fig. 3C (right)]. Extending beyond broad immune-active/immunosuppressive macrophage labels, MoMacVERSE (46) curated macrophage programs (Supplementary Table S3) showed that low-SGCC (DZ-like)

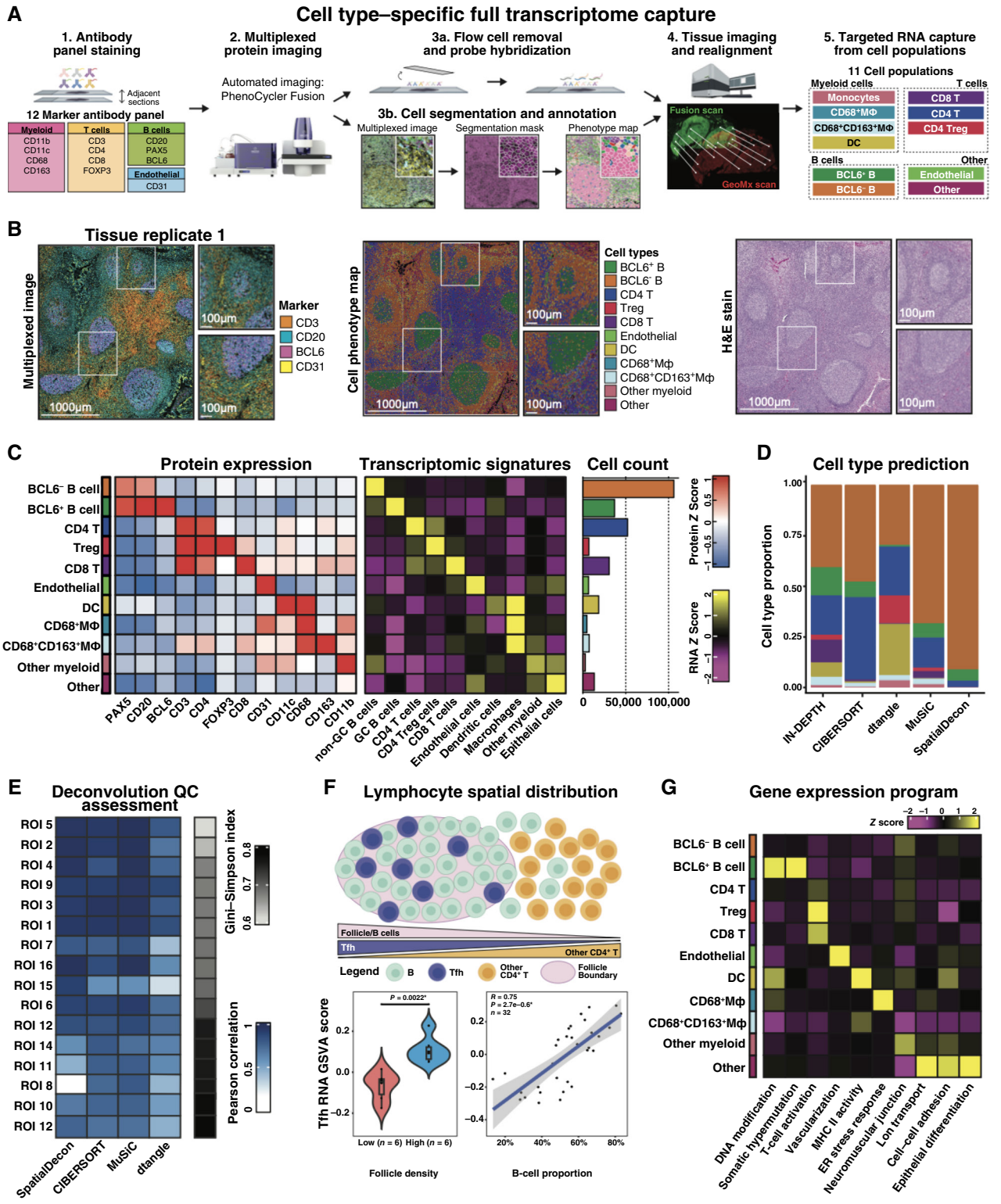


Figure 2. IN-DEPTH enables reproducible and systematic characterization of tonsillar tissue architecture through integrated spatial proteomics and transcriptomics. **A,** Schematic workflow of IN-DEPTH, illustrating the 12-marker antibody imaging, cell segmentation and phenotyping, cross-platform tissue image registration, and targeted RNA capture from identified cell populations on the same slide. **B,** Visualization of key cellular features in tonsillar tissues using CODEX multiplexed imaging (left) showing T cells (CD3), B cells (CD20 and BCL6), and endothelial cells (CD31), with the corresponding cell phenotype map (middle) and H&E image (right) of tissue replicate 1, as part of the IN-DEPTH workflow. Sixteen 660 × 760 μm rectangular ROIs were selected on each adjacent tissue replicates. **C,** Cell type–specific protein expression levels (left), gene signatures (middle), and cell counts (right) for the annotated cell types. Data shown are generated from tissue replicate 1. (*continued on following page*)

regions are enriched for HES1$^+$FOLR2$^+$, C1Q$^+$, and FTL$^+$ macrophage states (consistent with tingible-body/efferocytic and iron-handling functions), whereas high-SGCC (LZ-like) regions are enriched for IL1B$^+$ and CD16$^-$ monocytes (compatible with cytokine-driven activation and immune-complex handling), aligning with GC zonation and prior reports on macrophage polarization and monocyte spatial patterning in tonsil (47).

Notably, SGCC scores between these cell populations yielded concordant spatial patterns and corresponding state changes across tissue replicates (Supplementary Fig. S4I and S4J), underscoring the robustness of SGCC for resolving coordinated transitions. Together, these data highlight the unique insight gained by integrating IN-DEPTH spatial multiomics with SGCC analysis to reveal spatially orchestrated shifts in cell state and function, beyond the reach of either modality alone (Fig. 3D).

IN-DEPTH Reveals an EBV-Linked Macrophage Immunosuppression and Associated CD4 T-cell Dysfunction in the DLBCL TME

To investigate the complex tumor–immune interactions in the viral-linked TME, we next applied IN-DEPTH to dissect the poorly understood TME of EBV-positive and EBV-negative DLBCL. Using a multi-institutional cohort of FFPE tissues from 17 EBV-positive and 13 EBV-negative patients, we performed IN-DEPTH (CODEX–GeoMx) with a 30-marker antibody panel for cell phenotyping and functional analysis (Fig. 4A; Supplementary Fig. S5A; Supplementary Table S1). We identified eight distinct cell populations (Fig. 4B; Supplementary Fig. S5B) and performed scale-invariant feature transform (SIFT) to warp the CODEX-derived cell type–specific masks onto the GeoMx coordinate space (Supplementary Fig. S6A–S6C, see "Methods"), ultimately capturing genome-wide transcriptomes across 38 ROIs. GeoMx data were batch effects corrected (see "Methods"; Supplementary Fig. S6D; Supplementary Table S5), and all multiplexed staining patterns and marker-based cell type annotations were reviewed by board-certified pathologists, with same-slide H&E histology used to verify tissue context and morphology (see "Methods"; Fig. 4B and C; Supplementary Fig. S6E).

Building upon our prior findings of increased T-cell dysfunction in EBV-positive classical Hodgkin lymphoma TME (bioRxiv 2024.03.05.583586), we hypothesized there to be distinctive immune composition and organization within the EBV-stratified DLBCL TME. Our initial analysis across the entire tissue microarray (TMA) at core level revealed striking differences in TME composition, with EBV-positive DLBCL consisting of higher immune infiltrates compared with the tumor-heavy EBV-negative cases (Fig. 4D). Further dissection of the immune populations at the tissue level demonstrated an EBV-associated increase in regulatory T cells (Treg), and a distinctive macrophage polarization marked by elevated immunosuppressive CD68$^+$CD163$^+$ macrophages and diminished immune-active CD68$^+$ macrophages in the EBV-positive DLBCL (Fig. 4E and F).

At the tissue level, the EBV-positive DLBCL TME exhibited a significant reduction in MHC class II expression, whereas MHC class I expression did not differ significantly compared with EBV-negative DLBCL (Fig. 4G), suggesting a preferential impairment of CD4 T cell–mediated immunity consistent with reduced MHC class II antigen presentation. Using CD4 and CD8 T-cell dysfunction signatures assessed at both the protein (core-level) and transcript (ROI-level) scales (see "Methods"; Supplementary Table S3; refs. 48–50), we observed that CD4 T cells exhibited significantly higher dysfunction scores in EBV-positive compared with EBV-negative DLBCL at both the protein and RNA levels, whereas CD8 T-cell dysfunction scores did not differ between groups (Fig. 4H). Notably, CD4 T-cell dysfunction scores derived from protein and transcriptomic data were significantly positively correlated, whereas no such correlation was observed for CD8 T cells (Supplementary Fig. S7A). Together, these results indicate that EBV-associated immune dysfunction preferentially affects CD4 T cells and is consistently captured across modalities, in line with the selective reduction of MHC class II antigen presentation observed in EBV-positive tumors (51). Moreover, the orthogonal confirmation of T-cell dysfunction across modalities underscores the value of same-slide multiomics via IN-DEPTH for integrated biological discovery and validation.

To further explore the spatial context potentially contributing to CD4 T-cell dysfunction, we characterized the immediate cellular neighborhoods surrounding CD4 T cells using a network-based 1-hop adjacency analysis that integrated spatial proteomics and transcriptomics data at the ROI level. *K*-means clustering identified five distinct CD4 T cell–associated neighborhood motifs (Fig. 4I; Supplementary Fig. S7B–S7D), of which immune-rich Motif 1 [enriched in macrophages, Tregs, dendritic cells (DC), and endothelial cells] and Motif 4 (enriched in CD8 T cells) were significantly more prevalent in EBV-positive cases (Fig. 4J). Within this spatial and immunologic context, CD4 T cells exhibited a graded increase in dysfunction, with the highest dysfunction observed in EBV-positive immune-rich motifs (Motif 1) and the lowest in EBV-negative immune-deficient motifs (Motifs 2–5; Fig. 4K).

Figure 2. ***(Continued)*** Results related to tissue replicate 2 can be found in Supplementary Fig. S3. Refer to Supplementary Table S3 and "Methods" for details in gene signature curation. **D,** Systematic evaluation of four computational deconvolution algorithms using IN-DEPTH data as the reference. **E,** Assessment of deconvolution accuracy was performed by calculating the Pearson correlation between the computed cell proportions from each deconvolution algorithm and the IN-DEPTH–derived ground truth measurements across 11 cell types for each ROI. Cell type proportion complexity of each ROI is calculated using Gini–Simpson index. All ROIs were ranked by the Gini–Simpson index from low to high (top to bottom), indicating cell type proportion complexity from low to high. **F,** Spatial multimodal analysis of Tfh cells showing their distribution relative to B-cell follicles (top schematic) and quantitative validation through differential Tfh gene signature enrichment between follicle-high and follicle-low regions (bottom left, six ROIs chosen each) and correlation with B-cell density (bottom right). A two-sided Wilcoxon rank-sum test was performed, with the null hypothesis that there is no difference in the Tfh signature between follicle-low and follicle-high regions (bottom left), and a Spearman's correlation was used for the correlation test (bottom right). Refer to Supplementary Table S3 for Tfh signatures. **G,** Top cell type–specific GEPs identified, and their relative enrichment across the 12 annotated cell populations. * indicates statistical significance.

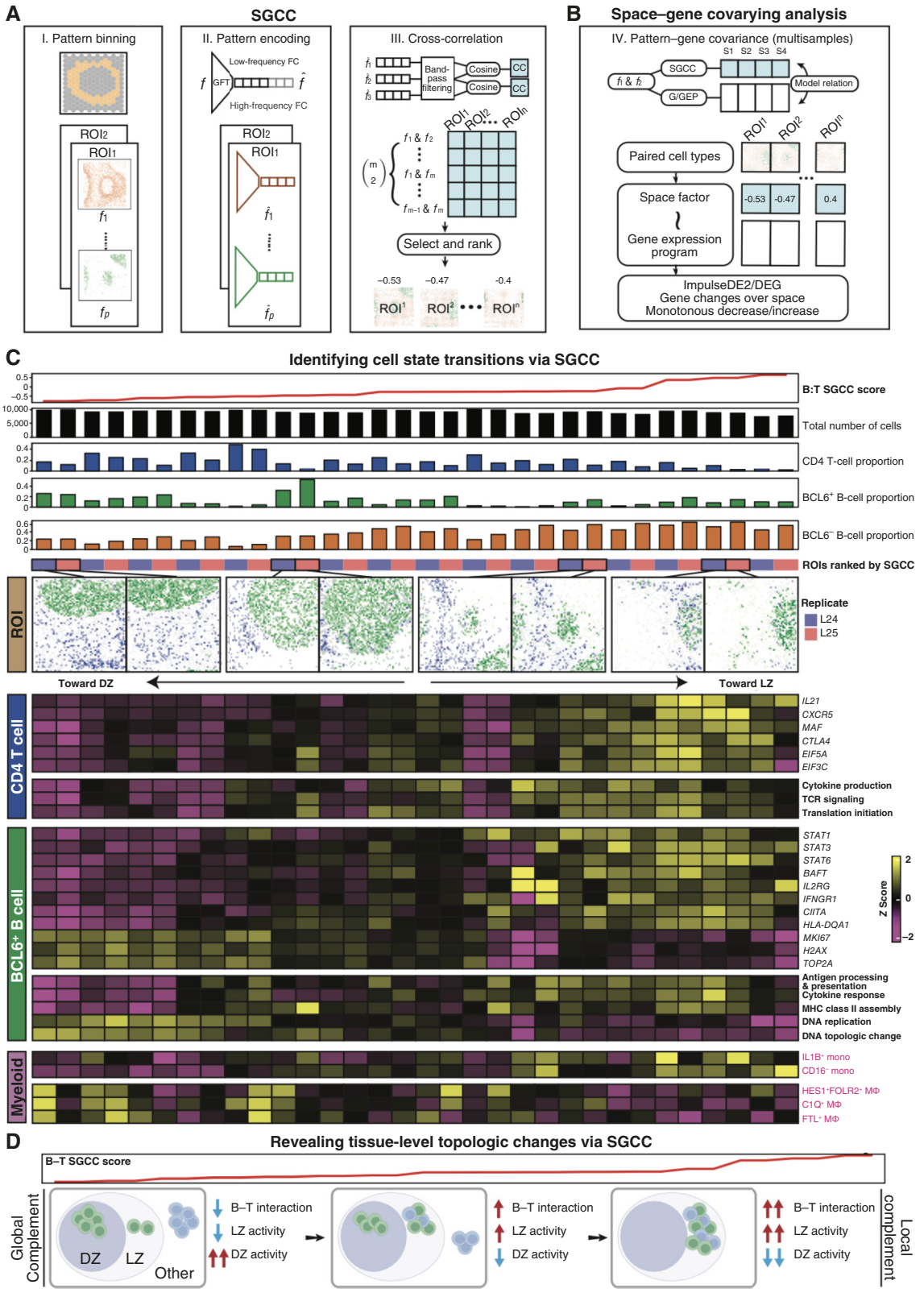


Figure 3. SGCC reveals coordinated spatial transitions in cellular states and tissue architecture. **A,** Schematic overview of the SGCC methodology showing (I) pattern binning of single cells in spatial proteomics data, followed by (II) pattern encoding through GFT to generate low-frequency FCs and (III) cross-correlation analysis to identified coordinated spatial patterns for downstream integration with transcriptomics. **B,** Integration framework for identifying genes covarying with spatial pattern across the tissue, linking spatial factors to gene expression for functional analysis. (*continued on following page*)

To directly assess whether macrophage composition contributed to these EBV-associated motif differences, we applied negative binomial regression to quantify macrophage subsets across motifs (Supplementary Fig. S7E; Supplementary Table S6). EBV-positive samples exhibited a marked skewing toward immunosuppressive CD68$^+$CD163$^+$ macrophages, with a 1.91-fold higher expected count relative to EBV-negative samples [$P < 0.05$; 95% confidence interval (CI), 1.64–2.25], accompanied by a concomitant reduction in immune-active CD68$^+$ macrophages (0.86-fold relative to EBV-negative; $P < 0.05$; 95% CI, 0.74–0.99). Consistent with this polarization, macrophages associated with EBV-positive tumors displayed increased PD-L1 and decreased HLA-DR expression with increasing tumor density, whereas macrophages associated with EBV-negative tumors exhibited the opposite pattern (Supplementary Fig. S7F), indicating a shift toward an immunosuppressive phenotype in EBV-positive cases. Collectively, these findings support a model in which EBV reshapes the DLBCL microenvironment through coordinated reduction in MHC class II expression and elevation of PD-L1 expression, thereby conditioning an immunosuppressive macrophage-enriched microenvironment that promote CD4 T-cell dysfunction (Fig. 4L).

SGCC Analysis Reveals a Spatially Coordinated Tumor–Macrophage–CD4 T Cell Axis Driving Immune Dysfunction in EBV-Linked DLBCL

To further dissect the molecular mechanisms underlying our proposed model of EBV-associated CD4 T-cell dysfunction (Fig. 4L), we extended SGCC to interrogate the spatial relationships among tumor cells, macrophages, and CD4 T cells and elucidate coordinated molecular mechanisms driving this biological process.

As EBV is primarily present in tumor cells (52, 53), we first examined how EBV-positive versus EBV-negative tumor cells modulate macrophage functional states. Consistent with clinical annotation, EBV-positive tumor cells exhibited elevated expression of the LMP1 viral oncoprotein and viral transcripts relative to EBV-negative tumor cells [Fig. 5A (top); Supplementary Fig. S8A]. SGCC analysis revealed distinct EBV-associated immunomodulatory programs in tumor cells. EBV-positive tumor cells were enriched for pathways linked to macrophage proliferation and chemotaxis, whereas EBV-negative tumor cells preferentially activated programs associated with innate immune signaling and positive regulation of leukocyte cytotoxicity (Fig. 5A; Supplementary Fig. S8A; Supplementary Table S7). These tumor-associated programs align with the immunosuppressive and immune-active TME states previously shown in EBV-positive and EBV-negative DLBCLs, respectively (Fig. 4).

Using the MoMacVERSE macrophage signatures (Supplementary Table S3; ref. 46), we further resolved macrophage polarization states associated with these tumor programs. Macrophages in EBV-positive TMEs exhibited increased expression of transcriptional profiles corresponding to FTL$^+$, C1Q$^+$ and HES1$^+$FOLR2$^+$ macrophage states, whereas macrophages in EBV-negative TMEs showed profiles consistent with TREM2$^+$ and IL4I1$^+$ macrophage states (Fig. 5A; Supplementary Table S7). These macrophage states were associated with divergent functional programs, with macrophages in EBV-positive TMEs showed enrichment of pathways related to negative regulation of inflammation and TNF signaling, whereas macrophages in EBV-negative TMEs demonstrated increased activation of mTOR and NF-κB signaling pathways (Fig. 5A; Supplementary Fig. S8A; Supplementary Table S7). Importantly, these patterns correlated with increasing SGCC scores (from global to local organizations), linking EBV-associated tumor cell programs to macrophage functional polarization in a spatially structured manner.

Given these EBV-dependent differences in macrophage states and functional programs observed across the TMEs, we next examined how macrophages influence CD4 T-cell functional states. In EBV-positive DLBCL, increasing SGCC score was associated with reduced macrophage HLA-DR protein expression and diminished activation of gene programs related to MHC class II, regulation of T-cell activation, and T-cell differentiation. In contrast, these macrophage-associated immune activation signatures were progressively enhanced with increasing SGCC score in EBV-negative DLBCL (Fig. 5B; Supplementary Fig. S8B; Supplementary Table S7). Consistent with these macrophage-associated trends, CD4 T cells in EBV-positive DLBCL exhibited suppression of gene programs linked to T-cell immunity, proliferation, and antigen receptor signaling, indicative of a dysfunctional state. Notably, these transcriptional features aligned with the elevated CD4 T-cell dysfunction scores shown earlier (Fig. 4; Supplementary Fig. S7A). Conversely, CD4 T cells in EBV-negative DLBCL showed increased activation of these pathways from global to local organizations, that is, increasing SGCC score (Fig. 5B; Supplementary Fig. S8B; Supplementary Table S7). Together, these results indicate that macrophage–CD4 T cell spatial associations are linked to divergent immunomodulatory states that are strongly conditioned by EBV status.

To further resolve the complexity of this tripartite spatial interaction, we visualized tumor–macrophage–CD4 T cell relationships using ternary analysis of SGCC scores (Fig. 5C). Although SGCC scores were broadly distributed across the three cell populations, EBV-positive TMEs exhibited enrichment of CD4 T cell–centric SGCC scores, whereas EBV-negative TMEs showed enrichment of macrophage-centric scores. CD4 T-cell dysfunction peaked in regions in which all three cell types co-localized (Fig. 5D; Supplementary Table S8), supporting a tripartite spatial interaction axis that promotes CD4 T-cell dysfunction. Adjacency enrichment statistic (AES) analysis further revealed preferential tumor–macrophage interactions in EBV-positive DLBCL, compared with macrophage–CD4 T cell

Figure 3. ***(Continued)*** **C,** Analysis of CD4 T cells and BCL6$^+$ B cells via IN-DEPTH proteomics and transcriptomics analysis, showing SGCC scores, total number of cells, CD4 T-cell proportion, BCL6$^+$ and BCL6$^-$ B-cell proportion, their associated spatial distribution of cells in bins, and the coordinated GEPs reflecting intrinsic cell programs and T cell–B cell cross-talk (bottom). Enriched genes are italicized, pathways are bolded, and myeloid subtypes are in purple. The full gene pathway names can be found in Supplementary Table S4. **D,** A schematic illustrating tissue-level organization derived from SGCC analysis depicting the transitions in T cell–B cell interactions across the DZ and LZ.

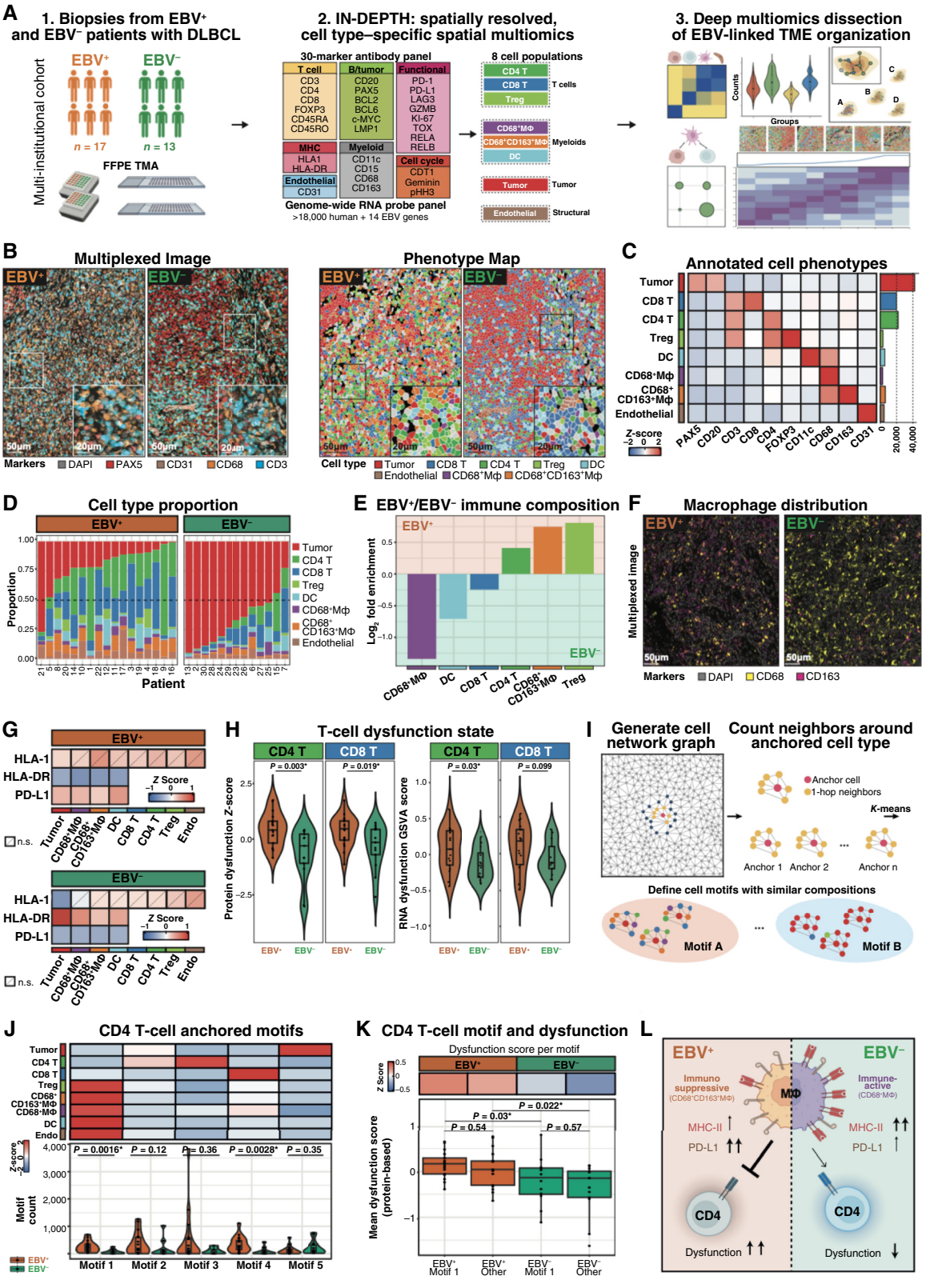


Figure 4. Iterative spatial multiomics dissection of EBV-positive and EBV-negative DLBCL via IN-DEPTH reveals a macrophage-linked CD4 T-cell dysfunction interaction axis. **A,** IN-DEPTH workflow on EBV-positive ($n = 17$) and EBV-negative ($n = 13$) DLBCL biopsy samples (Supplementary Table S11), using a 30-marker antibody panel (Supplementary Table S1) and a genome-wide RNA probe panel spiked in with custom-designed probes targeting 14 EBV genes. One 660 × 785 μm rectangular ROI was drawn for each patient core with emphasis on tumor-enriched regions. (*continued on following page*)

interactions in EBV-negative cases (Fig. 5D; Supplementary Table S8). These findings support a model in which tumor–macrophage cross-talk and immunosuppression dominate in EBV-positive DLBCL TMEs, constraining CD4 T-cell activation and promoting dysfunction. Notably, LMP1 abundance was enriched at the center of the ternary plot (Fig. 5D), and LMP1-expressing tumor cells showed positive correlations with immunosuppressive $CD68^+CD163^+$ macrophage polarization and CD4 T-cell dysfunction [Fig. 5E (right column); Supplementary Fig. S8C]. In contrast, EBV-negative TMEs exhibited increased macrophage–CD4 T cell engagement (Fig. 5D), consistent with a more immune-active microenvironment characterized by enhanced CD4 T-cell activation and functional immune responses.

Collectively, integrating IN-DEPTH with SGCC extended our proposed mechanism (Fig. 4L) and revealed two spatially orchestrated cellular circuits within the DLBCL TME. In EBV-positive DLBCL, tumor cells preferentially associate with macrophages to establish an immunosuppressive niche that limits CD4 T-cell activation and promotes dysfunction. Conversely, in EBV-negative TMEs, enhanced macrophage-CD4 T-cell interactions foster an immune-active microenvironment with preserved CD4 T-cell function (Fig. 5F; refs. 54–61).

SGCC Reveals Finer Details of EBV-Driven Tumor–Macrophage–CD4 T Cell Coordinated Functional Changes at Single-Cell Resolution

To extend our multicellular-level analyses and capture coordinated functional changes at higher spatial resolution, we performed additional sections of the DLBCL cohort using an expanded IN-DEPTH workflow that combined multiplexed spatial proteomics (26 protein markers, including additional macrophage markers, Supplementary Table S1) with high-plex single-cell spatial transcriptomics (about 6,000 human and EBV genes; Fig. 6A). The multiomics data were aligned, and transcripts were assigned at single-cell level (Supplementary Fig. S9A–S9C; see "Methods"). In parallel, we conducted CosMx- and CODEX-only validations using the same marker panel in two independent DLBCL cohorts (Fig. 6A), enabling cross-platform and cross-cohort validation of our findings.

Across a total of 55 EBV-positive and 41 EBV-negative fields of view (FOV) derived from four independent cohorts spanning three clinical collection sites in the validation cohort, we consistently observed higher $C1Q^+$ macrophage signature scores in EBV-positive DLBCL compared with EBV-negative DLBCL (Supplementary Fig. S9D; Supplementary Table S3). Moreover, $C1Q^+$ macrophage scores showed significant positive correlations with CD4 T-cell dysfunction scores in the single-cell IN-DEPTH dataset as well as in the independent CosMx-only validation cohorts [Fig. 6B (left)]. A similar trend was observed at the core level in the IN-DEPTH dataset [Supplementary Fig. S9E (bottom)], whereas no significant association was detected between $C1Q^+$ macrophage signature scores and CD8 T-cell dysfunction [Supplementary Fig. S9E (top)]. Together, these findings support a selective association between C1Q-associated macrophage polarization and CD4 T-cell dysfunction, consistent with prior reports linking $C1Q^+$ macrophages to immunosuppressive TMEs and adverse clinical outcomes in DLBCL (62, 63).

Given the established association between EBV positivity and poorer prognosis (58), we next examined whether EBV burden was directly linked to $C1Q^+$ macrophage polarization. Indeed, EBV burden positively correlated with $C1Q^+$ macrophage subtype signatures in the validation cohorts and showed a similar trend in the IN-DEPTH cohort, albeit at borderline significance [Fig. 6B (right)], suggesting a coordinated relationship among EBV status, $C1Q^+$ macrophage polarization, and CD4 T-cell dysfunction in EBV-positive DLBCL.

To systematically capture spatially coordinated functional changes among tumor cells, macrophages, and T cells at single-cell resolution, we next applied SGCC to the IN-DEPTH

Figure 4. ***(Continued)*** A total of 30 ROI were drawn for this experiment. The number of cells per annotated cell type (i.e., nuclei count), read counts, and Q30 values are in Supplementary Table S2. **B,** Representative CODEX multiplexed images (left) with markers for nuclei (DAPI), B/tumor cells (PAX5), endothelial cells (CD31), macrophages (CD68), and T cells (CD3) shown, as well as the corresponding phenotype maps (right) of EBV-positive and EBV-negative DLBCL tissues. Individual marker images are in Supplementary Fig. S5A. Phenotype maps and H&E images for each tissue sample core are in Supplementary Figs. S5B and S6E, respectively. **C,** Relative protein expression levels (left) and cell counts (right) at full slide level across the entire TMA for the annotated cell types from this DLBCL cohort. **D,** Relative proportions of annotated cell types across EBV-positive and EBV-negative (left) tissues at full slide level. **E,** Log_2 fold enrichment plot of immune cell proportions between EBV-positive and EBV-negative DLBCL tissues in this patient cohort. **F,** Representative multiplexed images showing macrophage distribution in EBV-positive and EBV-negative tissues, using markers nuclei (DAPI, gray), CD68 (yellow), and CD163 (magenta). **G,** Relative protein expression of MHC class I (HLA1), MHC class II (HLA-DR), and PD-L1, on the corresponding cell types that express these molecules across EBV-positive (top) and EBV-negative (bottom) DLBCL tissues in this patient cohort. Diagonal hatching indicates lack of statistical significance. **H,** Left, Comparison of CD4 and CD8 T-cell dysfunction scores calculated by summing the expression intensities of LAG3, TOX1/2, and CD45RO and subtracting the intensities of CD45RA, KI67, and GZMB protein markers between EBV-positive and EBV-negative DLBCL tissues at core level. Right, Comparison of CD4 and CD8 T-cell dysfunction scores calculated based on GSVA scoring of RNA gene set consists of *CTLA4*, *LAG3*, *HAVCR2*, *PDCD1*, *BTLA*, *TIGIT*, *CD160*, *CD244*, *ENTPD1*, *VSIR*, *NT5E*, *ADORA2A*, *PVRIG*, *SIGLEC7*, and *SIGLEC9* in EBV-positive and EBV-negative DLBCL tissues at GeoMx ROI level. A one-sided Wilcoxon rank-sum test was performed, with the alternative hypothesis that the T-cell dysfunction signature was greater in the EBV-positive tissues. The selection of protein markers and RNA gene signatures, as well as the calculation methods for the dysfunction scores, are detailed in "Methods" and Supplementary Table S3. **I,** Schematic representation of identifying different cellular motifs through *n*-hop neighborhood analysis anchored on a cell type of interest using integrated proteomics and transcriptomics information of each ROI (*n*-hop neighborhood analysis and the selection of the optimal number of motifs are detailed in the Supplementary Fig. S7B–S7D and "Methods"). **J,** Top, Cell type enrichment from each identified cellular motif, with CD4 T cells set as the anchor cell. Bottom, Comparison of motif abundance between EBV-positive and EBV-negative DLBCL. A two-sided Wilcoxon rank-sum test was performed, with the null hypothesis that there is no difference between motif abundance in EBV-positive and EBV-negative tissues. **K,** Heatmap (top) and boxplot (bottom) of the protein-level CD4 T-cell dysfunction score between immune-enriched motif 1 and all other motifs combined. The "Other" category was obtained by combining Motifs 2–5. Analyses were performed using integrated spatial proteomics and transcriptomics information at ROI level. **L,** Cartoon model depicting key differences in macrophage and CD4 T-cell dysfunction states between EBV-positive and EBV-negative DLBCL. * indicates statistical significance.

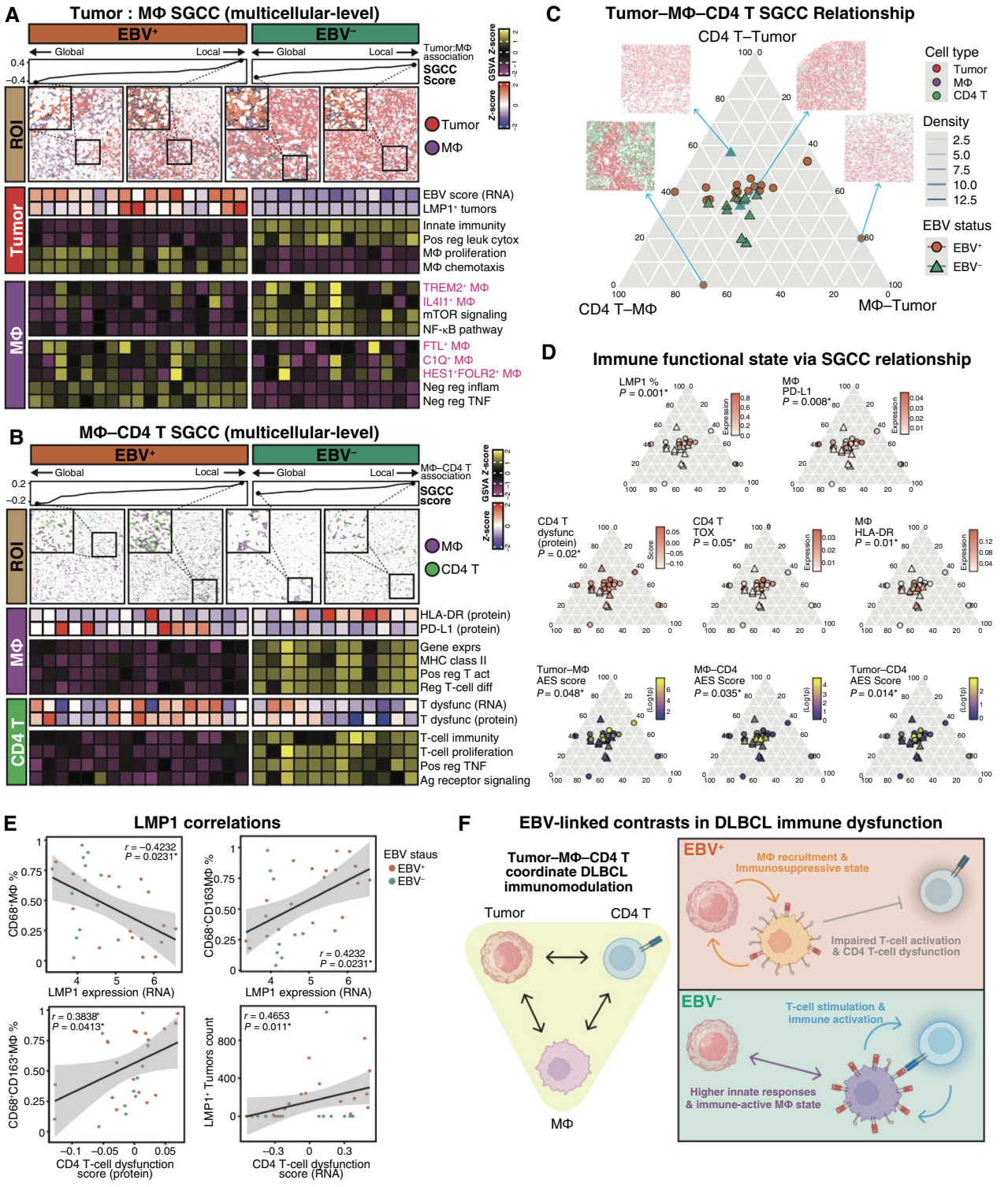


Figure 5. SGCC reveals coordinated spatial multimodal interactions and EBV-linked cell states in the tumor–macrophage–CD4 T cell axis. **A,** Analysis of tumor–macrophage spatial relationships. Top, SGCC-ranked spatial distributions and representative images. Middle, EBV score (transcript levels), LMP1+ tumor cells, and tumor-associated signaling pathways across SGCC scores. Bottom, Changes in macrophage polarization states (based on MoMacVERSE signature; ref. 49) and associated pathway signatures with increasing SGCC scores. Each column shows an individual ROI. Full pathway names are in Supplementary Table S7. **B,** Analysis of macrophage–CD4 T cell spatial relationship. Top, SGCC-ranked spatial distributions and representative images. Middle, Changes in HLA-DR and PD-L1 expression in macrophage and functional programs across SGCC scores. Bottom, Changes in T-cell dysfunction signatures (RNA and protein) and immune signaling pathways across SGCC scores. Each column shows an individual ROI. Full pathway names are in Supplementary Table S7. **C,** Ternary plot depicting a three-way SGCC relationship between CD4 T cells and tumor (top vertex), CD4 T cells and macrophages (bottom left vertex), and macrophages and tumor (bottom right vertex). (*continued on following page*)

dataset. Prior to downstream analyses, we assessed the robustness of SGCC across spatial resolutions and platforms. Cell type specificity and major functional programs identified at the multicellular level (Fig. 5) were well preserved at single-cell resolution across cohorts and platforms (Supplementary Fig. S9F). In particular, we observed concordant trends for pathways related to macrophage proliferation, chemotaxis, and leukocyte-mediated cytotoxicity in tumor cells; regulation of inflammatory responses and T-cell activation in macrophages; and antigen receptor signaling, T cell–mediated immunity, and dysfunction-related programs in CD4 T cells (Supplementary Fig. S9F). These results confirmed the reproducibility of SGCC across platforms and spatial resolutions and validated its use for dissecting fine-scale spatial coordination among cellular compartments.

In this framework, lower SGCC scores reflect spatial contexts in which coordinated functional changes are dominated by tumor–macrophage association, whereas higher SGCC scores indicate contexts in which such coordination is less specifically attributable to tumor–macrophage proximity and increasingly shaped by interactions with additional cell types within the 250 μm × 250 μm spatial window. Using this definition, we examined tumor–macrophage coordination at single-cell resolution. We found that in EBV-negative DLBCL, innate immune pathways were enriched when SGCC scores were low, indicating strong functional coordination between tumor cells and macrophages, and progressively downregulated as coordination decreased [high SGCC scores; Fig. 6C (middle); Supplementary Table S8; Supplementary Fig. S9G]. In contrast, innate immune activity was generally dampened in EBV-positive DLBCL, consistent with the more immunosuppressive TME observed at the multicellular level. Pathways related to cell-cycle progression and negative regulation of apoptosis followed similar proximity-dependent trends in EBV-negative tumors, indicating enhanced tumor proliferative and survival capacity when functionally coordinated with macrophages. These programs were markedly attenuated in EBV-positive tumors [Fig. 6C (middle)]. Notably, the elevated cell-cycle and antiapoptotic activity in EBV-negative tumor B cells may contribute to the higher tumor cellularity observed across independent cohorts (Fig. 4D).

Given these distinct tumor functional states, we next assessed how tumor–macrophage coordination shapes macrophage functional programs. Building on the macrophage subtypes identified using MoMacVERSE (Fig. 5A; Supplementary Table S7; ref. 46), single-cell SGCC enabled further refinement of macrophage functional states. In EBV-positive tumors, macrophages exhibited enrichment of IL4/IL13 signaling, cytokine and IL signaling, wound healing–associated programs, and Fcγ receptor–mediated signaling when tumor–macrophage coordination was dominating [low SGCC scores; Fig. 6C (bottom); Supplementary Fig. S9G]. Together, these pathways are characteristic of an immunosuppressive environment and are consistent with polarization toward the C1Q$^+$ macrophage subtype identified at the multicellular level. These macrophages also showed reduced activation of complement cascade pathways, which progressively increased as tumor–macrophage coordination weakened (high SGCC scores), suggesting selective tumor-mediated decoupling of C1Q-associated phagocytic functions from complement-driven inflammatory effector activity. In contrast, macrophages in EBV-negative DLBCL displayed distinct functional programs. TLR4 signaling and lipid metabolic pathways were enriched when tumor–macrophage coordination was strong (low SGCC scores) and progressively declined as tumor influence diminished [high SGCC scores; Fig. 6C (bottom); Supplementary Fig. S9G]. These features align with the immune-active microenvironment and lipid-sensing properties of TREM2$^+$ macrophages identified earlier, indicating that innate danger-sensing programs in macrophages are tightly regulated by tumor-derived cues in EBV-negative disease.

We next examined how these macrophage functional states influence CD4 T-cell responses. In EBV-positive DLBCL, CD4 T cells located in regions with dominant macrophage influence (low SGCC scores) exhibited enrichment of cytokine and IL signaling, oxidative phosphorylation, cellular responses to oxidative stress, and apoptotic programs [Fig. 6D (bottom); Supplementary Fig. S9H]. This constellation of features supports a chronically stimulated, stress-associated T-cell state, consistent with the CD4 T-cell dysfunction reported previously (Figs. 4 and 5). In contrast, these pathways were generally dampened in EBV-negative CD4 T cells indicating limited dependence on macrophage [Fig. 6D (bottom); Supplementary Fig. S9H].

Together, SGCC analysis at single-cell resolution revealed coordinated, association-dependent functional relationships among tumor cells, macrophages, and CD4 T cells that are not resolvable at multicellular-level resolution. In EBV-positive DLBCL, macrophages exhibited enrichment of immunoregulatory programs consistent with C1Q-associated polarization under conditions of dominant tumor–macrophage association, whereas complement pathway activity and C1Q protein expression were concurrently reduced and progressively increased as tumor–macrophage association weakened (Fig. 6C). This divergence between macrophage polarization and complement effector output indicates that tumor-associated niches selectively constrain complement-mediated macrophage functions in an EBV-specific manner, linking EBV-associated tumor states to spatially organized immune regulation. Collectively, these findings establish SGCC as a powerful framework for resolving association-dependent immune regulation within the TME.

Figure 5. ***(Continued)*** Points located near the vertices indicate colocalization between two specific cell types while forming a complementary structure with the third cell type (e.g., the ROI from Rochester 4 at the left bottom end of the triangle demonstrates colocalization between CD4 T cells and macrophages while complementing the tumor). In contrast, points near the center of the triangle may signify colocalization among all three cell types. EBV-positive and EBV-negative samples are indicated in circle and triangle, respectively). **D,** Ternary plots across the tumor–macrophage–CD4 T cell axis colored by their expression of key immune dysfunction features (top two rows) or AES (bottom row; Supplementary Table S8). **E,** Correlation of LMP1-expressing tumors with CD4 T-cell dysfunction signatures and macrophage polarization states across protein and RNA levels. **F,** Cartoon model depicting contrasting immune state differences in the tumor–macrophage–CD4 T cell interaction axis between EBV-positive (immunosuppressive) and EBV-negative (immune-active) DLBCL TMEs. * indicates statistical significance.

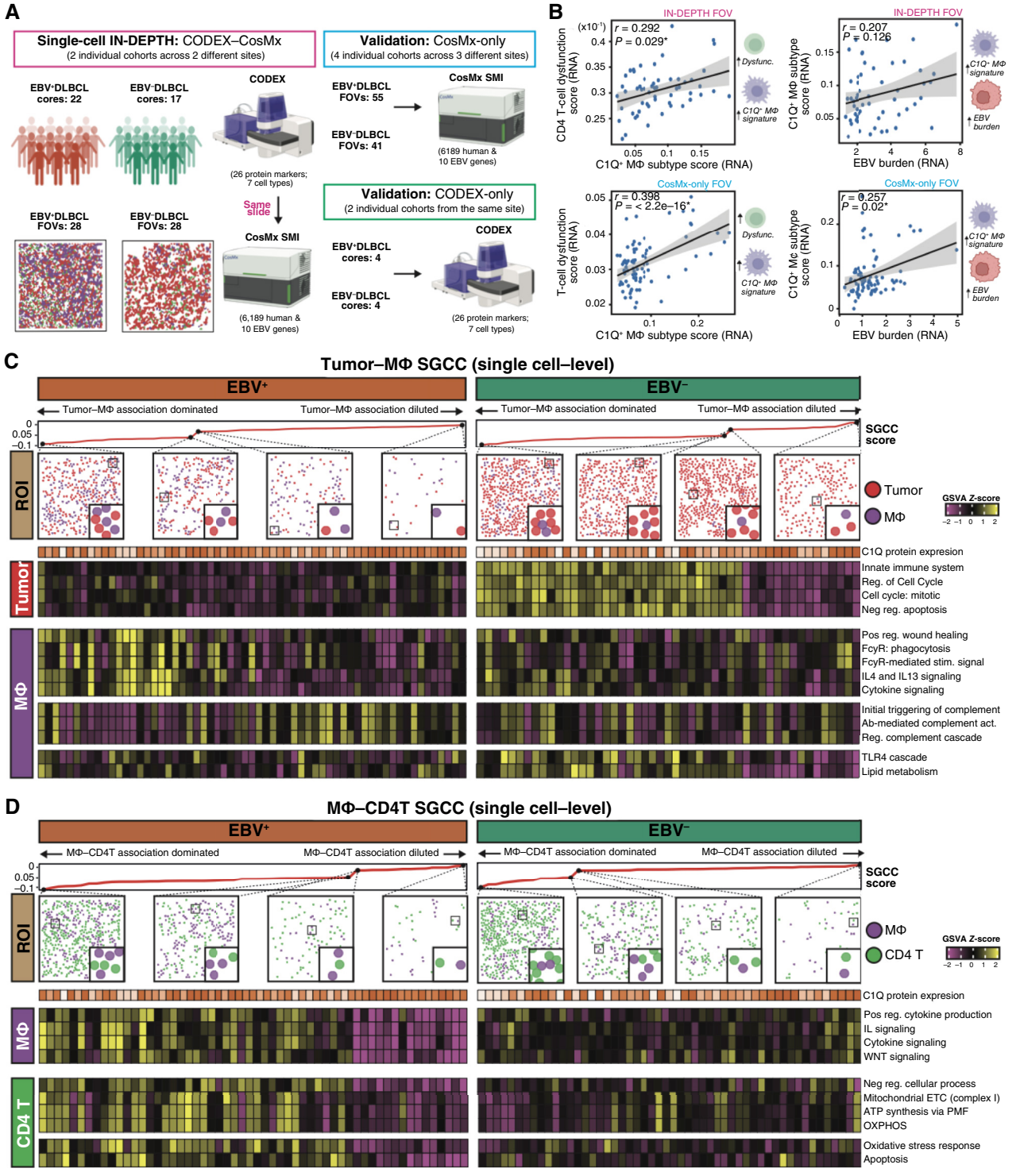


Figure 6. Single-cell SGCC reveals finer details of EBV-driven tumor–macrophage–CD4 T cell coordinated functional changes. **A,** Schematic illustration of additional cohorts for single-cell SGCC using CODEX–CosMx (22 EBV$^+$ and 17 EBV$^-$ cores; 28 EBV$^+$ and 28 EBV$^-$ FOVs) with the IN-DEPTH workflow and validation cohorts with CosMx-only (55 EBV$^+$ and 41 EBV$^-$ FOVs) and CODEX-only (4 EBV$^+$ and 4 EBV$^-$ cores) experiments (Supplementary Table S11). **B,** Top, Scatter plots showing the correlations between C1Q$^+$ macrophage signature with CD4 T-cell dysfunction (left) and C1Q$^+$ macrophage signature with EBV burden (right) using the single-cell IN-DEPTH dataset. Bottom, Scatter plots showing the correlations between C1Q$^+$ macrophage signature with T-cell dysfunction (left) and C1Q$^+$ macrophage signature with EBV burden (right) using the CosMx-only validation cohort. Gene signatures are in Supplementary Table S3. **C,** Single-cell SGCC between tumor and macrophage association. Low SGCC score corresponds to spatial organization in which tumor-macrophage association is dominated, whereas high SGCC score corresponds to spatial organization in which tumor-macrophage association is diluted by effects imposed by other interacting cell types. Representative spatial organizations of tumor and macrophage at different SGCC scores are shown (top). Each column represents an individual FOV. Full pathway names are in Supplementary Table S7. (*continued on following page*)

EBV LMP1 Marks Tumor Neighborhoods Enriched for C1Q-Associated Macrophages and IL27–STAT3 Signaling in EBV-Positive DLBCL

Given the apparent EBV specificity of these association-dependent macrophage features and our previous observation that EBV burden was positively associated with C1Q-associated macrophage signatures (Fig. 6B; Supplementary Fig. S9E), which in turn correlated with CD4 T-cell dysfunction (Supplementary Fig. S9E), we next considered potential viral drivers of this spatial immune remodeling. CD4 T-cell dysfunction was also previously shown to be positively associated with expression of the EBV-encoded oncogene LMP1 (Fig. 5E). Together, these linked associations suggest that EBV-derived tumor-intrinsic signals, particularly LMP1, may actively shape macrophage-enriched spatial niches.

We therefore examined whether tumor cells with active LMP1 signaling (i.e., LMP1$^+$ tumor, based on protein expression) are preferentially embedded within C1Q-associated macrophage neighborhoods (1-hop, within 20 μm; Fig. 7A). We found that the proportion of C1Q$^+$ macrophages (based on protein expression) surrounding EBV-positive tumor cells exhibited an estimated $\log_{10}$ fold change of 2 (95% CI, 1.91–2.10) relative to EBV-negative controls. Furthermore, within the EBV-positive subset, LMP1 positivity was associated with a $\log_{10}$ fold change of 1.22 (95% CI, 1.13–1.30) compared with LMP1-negative counterparts [Fig. 7B (left)]. A similar trend was also observed in the independent validation cohort [Fig. 7B (right)]. These data suggest that LMP1 expression marks tumor regions embedded within C1Q$^+$ macrophage subtype–rich niches.

To further investigate the role of LMP1 in shaping these niches, we anchored each macrophage to its nearest tumor cell at 20 μm and classified macrophages as proximal to either LMP1-positive ("near LMP1$^+$ tumor") or LMP1-negative ("near LMP1$^-$ tumor") tumor cells. This spatial classification captures local clusters of active versus inactive LMP1 protein signaling within EBV-positive DLBCL.

We first compared the functional states of tumor cells with and without active LMP1 signaling. Consistent with the established role of LMP1 as a constitutively active CD40 mimic (64, 65), tumors with active LMP1 signaling exhibited differential expression of genes, for example, *CD40*, *TRAF1*, *EBI3*, *NFKBIA*, and *STAT5A* (Fig. 7C). Gene set enrichment analysis (GSEA) revealed enrichment of signaling pathways downstream of the CD40 axis, including STAT5 activation, TNF, IL27 and NFκB signaling, and positive regulation of telomerase, in LMP1-positive tumor cells relative to LMP1-negative tumor cells (Fig. 7C; Supplementary Table S9), which aligned with the reported roles of LMP1 and its effect on telomerase as one of the mechanism in driving tumor development (66, 67). In contrast, tumors lacking active LMP1 signaling were enriched for pathways related to metabolism, calcium transport, and basement membrane organization (Fig. 7C; Supplementary Table S9). These results indicate that even within EBV-positive disease, tumor cells are exhibiting distinct functional states depending on the status of LMP1 signaling.

Given the functional differences observed between these tumor groups, we next examined how LMP1 status influences the transcriptional programs of spatially associated macrophages. Differential expression analysis revealed that macrophages proximal to LMP1-positive tumors upregulated immunoregulatory genes, for example, *IDO1*, *LGALS9*, *GPNMB*, and *CSTB*, whereas macrophages near LMP1-negative tumors preferentially expressed genes such as *COX1*, *COX2*, and *PTEN* (Fig. 7D). GSEA further showed that macrophages adjacent to LMP1-positive tumors were enriched for pathways associated with negative regulation of T-cell proliferation and activation, complement signaling, and ferroptosis (Fig. 7D; Supplementary Table S9). In contrast, macrophages near LMP1-negative tumors exhibited enrichment for pathways related to oxidative phosphorylation, negative regulation of focal adhesion, and PI3K signaling (Fig. 7D; Supplementary Table S9). Together, these findings suggest that macrophages in proximity to LMP1-positive tumors adopt an immunoregulatory state consistent with the C1Q-associated macrophage phenotype, whereas those near LMP1-negative tumors display alternative, metabolically remodeled programs.

To identify potential tumor–macrophage interaction axes underlying these spatially distinct phenotypes, we inferred cell–cell interactions using Squidpy (Fig. 7E; Supplementary Table S10; ref. 68). This analysis revealed significantly enriched tumor (ligand)–to–macrophage (receptor) signaling along the IL27–STAT3 axis in LMP1-positive tumors and their neighboring macrophages, including EBI3–STAT3, EBI3–IL27RA, and EBI3–IL6ST interactions. The coordinated enrichment of ligand subunits (EBI3), receptor components (IL27RA and IL6ST), and downstream signaling mediators (STAT3) supports activation of this signaling pathway (69). IL27–STAT3 signaling has been reported to promote immunosuppressive macrophage polarization and upregulate PD-L1 expression (70, 71), consistent with the elevated PD-L1 levels observed in EBV-positive DLBCL in our cohort (Fig. 4G; Supplementary Fig. S7F). In addition, enrichment of the CCL22–CCR4 interaction suggests enhanced macrophage recruitment and stabilization of immunosuppressive niches (Fig. 7E). In contrast, tumor–macrophage interactions enriched in LMP1-negative tumors included both immune-activating and immune-regulatory axes, such as CCN1–TLR4 and ICAM3–CD209, indicating the presence of comparatively more immune-reactive features (Fig. 7E; Supplementary Table S10). These inferred interactions were supported by ligand–receptor (L–R) expression patterns, with significantly higher expression of EBI3–IL27RA, EBI3–IL6ST, EBI3–STAT3, CCL22–CCR4, and

Figure 6. ***(Continued)*** **D,** SGCC between macrophage and CD4 T-cell association at single-cell resolution. Low SGCC score corresponds to a spatial organization in which macrophage–CD4 T cell association is dominated, whereas a high SGCC score corresponds to a spatial organization in which macrophage–CD4 T cell association is diluted by effects imposed by other interacting cell types. Representative spatial organization of macrophage–CD4 T cell at different SGCC scores are shown (top). Functional pathways for macrophage (middle) and CD4 T cells (bottom) are shown per FOV. Each column represents an individual FOV. Full pathway names are in Supplementary Table S7. ETC, Electron Transport Chain; PMF, Proton Motive Force; OXPHOS, Oxidative Phosphorylation. * indicates statistical significance.

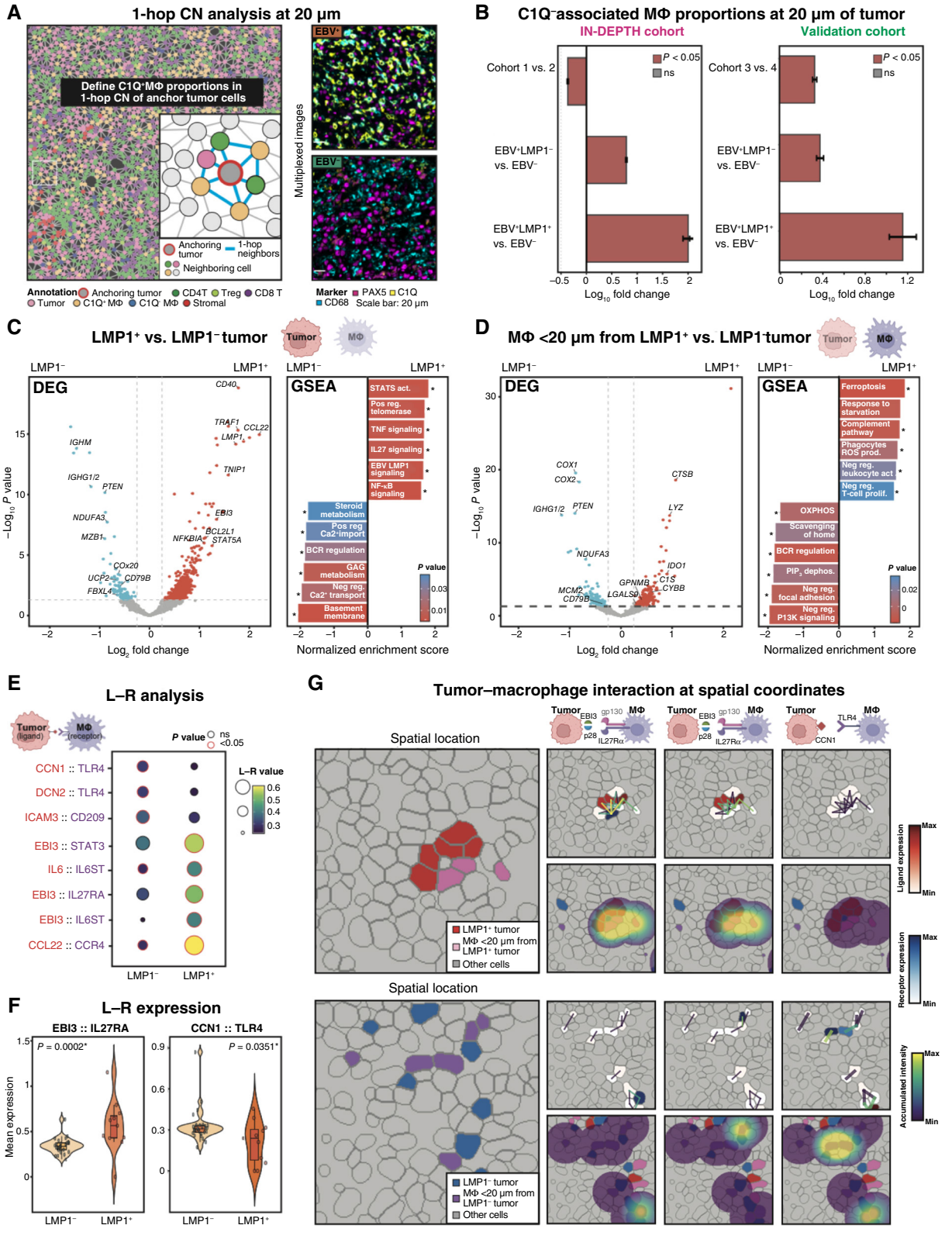


Figure 7. EBV LMP1 marks tumor neighborhoods enriched for C1Q-associated macrophages and IL27–STAT3 signaling in EBV-positive DLBCL. **A,** Left, Schematic of the 1-hop analysis used to quantify the proportion of C1Q-associated macrophages within 20 μm of anchor tumor cells. Right, Representative multiplexed images showing PAX5 (magenta), C1Q (yellow), and CD68 (cyan) in EBV-positive and EBV-negative samples. Scale bar, 20 μm. **B,** Estimated $\log_{10}$ fold change in C1Q-associated macrophage proportions for EBV-positive (LMP1-negative) vs. EBV-negative and EBV-positive (LMP1-positive) vs. EBV-negative comparisons in the IN-DEPTH (left) and validation (right) cohorts. Error bars indicate 95% CIs. (*continued on following page*)

IL6–IL6ST interactions in LMP1-positive tumor–macrophage pairs, whereas CCN1–TLR4, DCN2–TLR4, and ICAM3–CD209 interactions were elevated in LMP1-negative tumor–macrophage pairs (Fig. 7F; Supplementary Fig. S10A).

Finally, to directly link inferred tumor–macrophage interactions to their spatial tissue context, we mapped LMP1-positive and LMP1-negative tumor–macrophage interactions onto single-cell spatial coordinates. Spatial visualization demonstrated increased L–R signaling activity for EBI3–IL27RA and EBI3–IL6ST in LMP1-positive tumor–macrophage cell pairs, characterized by higher mean ligand and receptor expression as well as increased accumulated L–R interaction intensity. In contrast, CCN1–TLR4 signaling activity was higher in spatially associated LMP1-negative tumor–macrophage cell pairs, indicating distinct L–R interaction programs between LMP1-positive and LMP1-negative tumor regions (Fig. 7G). Together, these multiscale spatial analyses demonstrate that tumor-intrinsic LMP1 status shapes coherent tumor–macrophage interaction programs that are consistently reflected across single-cell transcriptional states, pairwise cell–cell signaling, and higher-order spatial organization.

Collectively, these same-slide spatial multiomics data support a model in which active LMP1 signaling acts as a key tumor-intrinsic determinant for local immune niche organization, marking tumor regions enriched for C1Q-associated immunosuppressive macrophages and driving macrophage polarization through the IL27–STAT3 signaling axis.

DISCUSSION

IN-DEPTH addresses key limitations in current spatial multiomics approaches by enabling same-slide protein and RNA profiling, thereby substantially increasing the number of simultaneously measurable biomolecules without proportional increases in cost and experimental time (Supplementary Fig. S1A). By performing protein and transcripts measurements on the same tissue section, IN-DEPTH eliminates the need for computational integration of adjacent slides, a process that is often technically challenging and prone to registration artifacts.

A central feature of the IN-DEPTH workflow is the protein-first strategy, which allows targeted spatial transcriptome interrogation guided by prior biological context derived from high-quality protein imaging. This strategy provides a resource-effective alternative to whole-slide transcriptome profiling in a platform agnostic manner. Importantly, it preserves transcriptomic performance while maintaining robust protein imaging quality, in contrast to RNA-first workflows that may compromise epitope integrity. Although IN-DEPTH is compatible with multiple commercially available spatial platforms, platform-specific considerations remain essential. For instance, tyramide signal amplification used in the Polaris platform results in covalent deposition of Opal fluorophores (22), which may require extensive photobleaching or alternative strategies to ensure compatibility with fluorescence-based *in situ* hybridization–based spatial transcriptomics platforms such as CosMx, which rely on RNA fluorescent imaging. Notably, IN-DEPTH is carefully optimized to preserve tissue integrity while supporting robust protein epitope staining, RNA signal retention, and subsequent H&E staining for histopathologic evaluation, thereby enabling the potential integration of additional spatial modalities beyond protein and RNA on the same slide (72).

SGCC, computed at both single-cell and multicellular levels, is derived from IN-DEPTH same-slide multimodal data using graph signal processing and SGWT-based mathematical principles (Fig. 3). The same-slide design is critical for SGCC, as it avoids the need for cross-section cell or spot matching required by sequential slide integration methods such as SpatialGlue (73), MaxFuse (74), and MARIO (75). Even in consecutive tissue sections, local discrepancies frequently arise due to missing cells, tissue distortion, microenvironmental or architectural variation, or imperfect registration. These artifacts can lead to incomplete spatial components and introduce additional topologic noise, particularly in graph-based integration frameworks, thereby reducing sensitivity to fine-scale local spatial features. By leveraging coregistered protein and RNA measurements from the same tissue section, SGCC operates on spatially faithful ground truth data, enabling robust quantification of cell–cell spatial coupling. Specifically, SGCC measures the relative spatial positioning of any two cellular phenotypes in the low-frequency domain, providing an unbiased and interpretable metric of spatial association. Owing to the mathematically grounded and modality-agnostic nature of SGWT, SGCC can be broadly applied to assess spatial relationships across diverse tissue contexts, whereas the IN-DEPTH framework additionally provides a benchmark dataset for evaluating and validating computational spatial integration methods.

When multiple samples or ROIs are available, SGCC can be treated as a continuous or ordinal spatial factor, enabling direct integration with transcriptomic data to identify genes and pathways that covary with spatial organization. In this capacity, SGCC integrates protein and RNA measurements into a unified spatial coupling axis that captures coordinated changes in cellular functional states across interacting cell populations. Rather than functioning as a predictive classifier,

Figure 7. ***(Continued)*** **C,** Differential gene expression and GSEA comparing tumor cells with active LMP1 signaling (LMP1-positive tumors) vs. tumors lacking LMP1 signaling (LMP1-negative tumors) in EBV-positive DLBCL. Enriched pathways are listed in Supplementary Table S9. **D,** DEG and GSEA comparing macrophages within 20 μm of LMP1-positive tumors vs. macrophages within 20 μm of LMP1-negative tumors. Enriched pathways are listed in Supplementary Table S9. **E,** Squidpy-inferred L–R interactions between (i) LMP1-positive tumors and nearby macrophages (within 20 μm) and (ii) LMP1-negative tumors and nearby macrophages (within 20 μm). Significant interaction pairs are provided in Supplementary Table S10. **F,** Mean expression of L–R pair identified from LMP1-positive and LMP1-negative tumor with their respective macrophages. **G,** Representative FOV showing the spatial mapping of LMP1-positive and LMP1-negative tumor–macrophage interactions at single-cell resolution. Column 1, Spatial localization of LMP1-positive and LMP1-negative tumor cells with their associated macrophages. Columns 2–4, EBI3–IL27RA, EBI3–IL6ST (gp130), and CCN1–TLR4 interactions, respectively. For each L–R pair, the top show mean ligand expression (red scale) and receptor expression (blue scale), with lines indicating mean L–R expression, and the bottom show accumulated L–R interaction intensity. * indicates statistical significance.

SGCC facilitates mechanistically interpretable analyses of how molecular programs are organized in space and coordinated across diseased and healthy compartments. Although demonstrated here using protein and transcriptomic data, the SGCC framework is inherently extensible and provides a generalizable foundation for integrating additional spatial molecular modalities, which is expected to further enhance biological interpretation beyond what is achievable with any single modality alone. Importantly, multiple resolutions of SGCC are introduced in this study, providing analytic flexibility to tailor spatial interrogation to distinct biological scales and disease contexts.

We demonstrate the utility of IN-DEPTH in dissecting EBV-associated immune modulation in the DLBCL TME, revealing distinct spatially organized immunoregulatory states. EBV-positive DLBCL is characterized by enrichment of immunosuppressive macrophages with reduced HLA-DR and elevated PD-L1 expression, coinciding with unique increased in CD4 T-cell dysfunction (Fig. 4), consistent with prior studies linking EBV positivity to poor prognosis and impaired antitumor immunity (54–61, 76). Notably, EBV-positive and EBV-negative DLBCL exhibit distinct T-cell immune states; these EBV-stratified differences in T-cell dysfunction may, in part, explain the variable responses to immune checkpoint blockade observed in DLBCL (77–79). By integrating macrophage subtype signatures with SGCC, we resolve functional macrophage diversity *in situ* and identify enrichment of C1Q$^+$ macrophage programs in EBV-positive TMEs, enabling spatially informed analysis of a tumor–macrophage–CD4 T cell interaction axis (Fig. 5). Extending these analyses to single-cell resolution, single cell Spectral Graph Cross-Correlation (scSGCC) revealed finer, association-dependent coordination that was not apparent at the multicellular level, including selective uncoupling of macrophage polarization from effector complement function within tumor-associated niches. Subsequent hop analyses further associated this spatial reorganization with expression of the EBV oncoprotein LMP1 (Fig. 7A), a constitutively active CD40 mimic known to activate NF-κB signaling in EBV-associated cancer cells (64, 65). Although LMP1 has been extensively studied for its tumor-intrinsic roles (66, 67), our data extend these observations by linking active LMP1 signaling to enrichment of pathways consistent with CD40-driven signaling programs, including NF-κB activation and telomerase-associated processes, alongside coordinated remodeling of the surrounding immune microenvironment.

Subsequent L–R interaction analysis further identified a potential IL27–STAT3 signaling axis that may contribute to this immunosuppressive macrophage phenotype (Fig. 7E), which has previously been implicated in promoting macrophage-mediated immune suppression and PD-L1 expression in other tumor contexts (70, 71). Importantly, C1Q$^+$ macrophages have been previously shown to associate with T-cell dysfunction and poorer prognosis in DLBCL and other malignancies (62, 63), situating our findings within an established biological and clinical framework.

Together, these findings connect EBV/LMP1-driven tumor-intrinsic signaling with the spatial organization of immunosuppressive macrophage niches and CD4 T-cell dysfunction in DLBCL. More broadly, they illustrate the ability of IN-DEPTH and SGCC to integrate combined protein and RNA measurements into spatially constrained, mechanistically interpretable interaction networks, enabling identification of coordinated, multicompartment functional state transitions that would not be detectable using either modality alone.

Several exciting opportunities exist for further development of the IN-DEPTH framework. Although we demonstrated strong gene–gene correlations (at least R >0.938) across 11 different combinations of leading and widely adopted spatial proteomics and transcriptomics platforms, highlighting the broad compatibility among current spatial multiomics technologies, the Orion–GeoMx correlation was comparatively lower. We are actively working to optimize this specific integration and plan to address it more comprehensively in future studies. Importantly, this isolated observation does not diminish the overall potential for robust cross-platform integration enabled by IN-DEPTH. In addition, we evaluated RNA quality in brain tumor tissue subjected to the physical upper limit of iterative imaging on the PhenoCycler Fusion. Although this resulted in a modest reduction in total RNA yield relative to the control slide (≈2.5-fold), gene-to-gene correlations remained robust (R = 0.966) and comparable with other platform combinations tested. Together, these results demonstrate that the IN-DEPTH protocol is highly robust, with RNA integrity largely preserved even under extensive iterative immunofluorescence imaging. In parallel, emerging technologies such as DBiTplus (80), which integrate sequencing-based spatial transcriptomics with image-based proteomics within a unified workflow and support both FFPE and fresh-frozen tissues, provide powerful and complementary capabilities for spatial multiomics profiling. IN-DEPTH is designed to address distinct and complementary objectives by enabling (i) robust protein-based cell type annotation to guide targeted RNA profiling, (ii) broad flexibility across commercially available proteomic and transcriptomic platforms, and (iii) demonstrated preservation of tissue integrity across extended iterative imaging cycles and experimental procedures (Supplementary Fig. S10B). Together, these features position IN-DEPTH as a versatile and extensible framework for future expansion. By preserving tissue integrity during iterative imaging and profiling, IN-DEPTH establishes a foundation for same-slide integration of additional spatial platforms, including protein-guided RNA profiling coupled with mask-based laser-capture mass spectrometry and other emerging biomolecular assays. Moreover, IN-DEPTH datasets provide a high-confidence reference resource for downstream computational developments, including bulk deconvolution, multimodal data integration, and related analytic applications.

Together, the experimental and computational advances presented herein demonstrate the potential for comprehensive tissue analysis and the generation of new biological insights through same-slide integrated spatial multiomics. We anticipate this approach will be broadly applicable across spatial platforms and will accelerate discovery and mechanistic research across diverse disease contexts and toward the generation of multimodal AI models for discovery and clinical impact.

METHODS

All reagents and resources used in this study are listed in the Key Resource Table in the Supplementary Notes file.

Human Tissue Acquisition and Patient Consent

All FFPE tissues used in this study were sectioned 5 μm thick on SuperFrost glass slides (VWR, 48311-703) and obtained from the following sources. The tonsil tissues in Figs. 1 and 2 and Supplementary Fig. S1B were generously provided by S.J. Rodig from the Brigham and Women's hospital (IRB# 2016P002769 and 2014P001026), the DLBCL tissue for SignalStar–GeoMx [Fig. 1C (row 1)] was purchased from Amsbio (AMS-31010), the kidney cancer [Fig. 1C (row 2)] and lymph node tissues [Supplementary Fig. S1B (row 3)] were generously provided by S. Signoretti from the Dana-Farber Cancer Institute (DFCI; IRB# DFCI 13-425), the DLBCL tissue for CODEX–CosMx [Fig. 1C (row 3)] was obtained from W.R. Burack from University of Rochester Medical Center (IRB# STUDY159), the periodontal disease tissue for CODEX–VisiumHD [Fig. 1C (row 4)] was generously provided by D.M. Kim from Harvard Dental School (IRB# 22-0587), the primary central nervous system lymphoma (PCNSL) tissue for CODEX–VisiumHD and CODEX–Xenium [Fig. 1C (rows 5 and 6)] were generously provided by C. Keane from Frazer Institute, and the uterine cancer tissues [Supplementary Fig. S1B (row 5)] were generously provided by B.E. Howitt from Stanford University Medical School.

For comparing EBV-positive and EBV-negative DLBCL using the CODEX–GeoMx workflow (Figs. 4 and 5), 30 patient samples (17 EBV-positive, 13 EBV-negative) were sectioned from two TMA. The DFCI TMA, constructed by S. Signoretti and S.J. Rodig (IRB# 2016P002769 and 2014P001026), includes 1 core from each patient (6 EBV-positive, 8 EBV-negative) and 1 tonsil control core, with each core measuring 1.5 mm in diameter. The University of Rochester Medical Center (Rochester) TMA, constructed by D. Nkosi, P. Rock, and W.R. Burack (IRB# STUDY159), includes 1 core from each patient (11 EBV-positive, 5 EBV-negative) and 1 tonsil control core, with each core measuring 2 mm in diameter. For dissecting EBV-positive versus EBV-negative signatures at single-cell level using the CODEX–CosMx workflow (Figs. 6 and 7), 39 patient samples (22 EBV-positive, 17 EBV-negative) were sectioned from the same DFCI and Rochester TMA as mentioned above. For the validation cohort (Fig. 6), one 1.5 mm diameter core from each of 18 patient samples (4 EBV-positive, 4 EBV-negative) were sectioned from two TMAs from University Hospital and Comprehensive Cancer Center Tübingen that was constructed by L. Frauenfeld, L. Kaufmann, and C.M. Schürch. EBV status for all DLBCL biopsies was verified using *in situ* hybridization for EBER as part of the routine clinical pathology process. Detailed deidentified information for the patients with DLBCL are in Supplementary Table S11.

Antibody Panel Selection, Conjugation, and Titration

Antibodies used in the CODEX experiments were conjugated in-house and include previously validated antibody clones (bioRxiv 2024.03.05.583586; refs. 10, 81). In brief, the specificity of antibody candidates was first validated via immunohistochemistry (IHC) on FFPE cell pellets or FFPE lymphoid tissues to ensure robustness of staining. The selected antibody clones were then conjugated by maleimide, lysine, or biotinylation chemistries, and each conjugated antibody was titrated and validated via immunofluorescence on FFPE lymphoid tissues. Interested readers are referred to the following publications for a more detailed guide on antibody target selection and optimization (19, 82). Antibodies used for the SignalStar, Polaris, and Orion experiments were obtained from their respective commercial sources. Details about the antibody clones, vendors, conjugated channels, titers, exposure times, and assigned channels throughout the study are in Supplementary Table S1.

Maleimide-based conjugations were performed with minor modifications from a previously published protocol (27). Briefly, 50 or 100 μg of carrier-free antibody was concentrated using a Phosphate Buffered Saline with Tween-20 (PBS-T) pre-wet 50 kDa filter (Sigma Millipore, UFC5050BK) and then incubated with 0.9 μmol/L Tris (2-carboxy-ethyl)-phosphin-HCl (Sigma, C4706-10G) for 10 to 30 minutes in a 37°C water bath to reduce the thiol groups for conjugation. Reduction was quenched by two washes with buffer C [1 mmol/L Tris (pH 7.5), 1 mmol/L Tris (pH 7.0), 150 mmol/L NaCl, 1 mmol/L EDTA] supplemented with 0.02% NaN_3. Maleimide oligos were resuspended in buffer C supplemented with NaCl (buffer C, 250 mmol/L NaCl). The reduced antibody was next incubated with 100 or 200 μg (for 50 or 100 μg of antibody, respectively) of maleimide oligos (Biomers, 5′-maleimide) in a 37°C water bath for 2 hours. The resulting conjugated antibody was purified by washing for 3 to 5 times with the 50 kDa filter with high-salt PBS (1× DPBS, 0.9 mol/L NaCl, 0.02% NaN_3). The conjugated antibody was quantified in IgG mode at A280 using a NanoDrop (Thermo Fisher Scientific, ND-2000). The final concentration was adjusted by adding >30% v/v Candor Antibody Stabilizer (Fisher Scientific, NC0414486) supplemented with 0.2% NaN_3, and the antibody was stored at 4°C.

Lysine-based conjugations were performed according to the official Alexa Fluor 532/594/647 Labeling Kit protocols (Thermo Fisher Scientific; A20182, A20185, and A20186). Briefly, 100 μg of carrier-free antibody was adjusted to a concentration of 1 mg/mL and mixed with 10 μL of 1 mol/L sodium bicarbonate buffer with gentle agitation for 5 minutes. The basic pH antibody was then transferred into the Alexa Fluor reactive dye with gentle pipetting to dissolve the dye. The labeling reaction proceeded in the dark for 1 hour at room temperature (RT), and the vial was gently inverted five times every 15 minutes. A purification resin bed was prepared by thoroughly resuspending the resin by violent agitation and then centrifuging the resin through the provided filters at 1,200 × *g* for 8 minutes until there was minimal residual buffer remaining in the resin bed. The conjugated antibody was then pipetted into the resin bed and allowed to absorb into the bed for 1 minute. The antibody was collected by centrifuging at 1,200 × *g* for 5 minutes and then stored at 4°C.

Biotinylation was performed using a commercial rapid biotinylation kit (Biotium, 92244) according to manufacturer's instructions. Briefly, 75 μg of carrier-free antibody was biotinylated, with a conjugation time of 15 minutes. The conjugated antibody was diluted in 300 μL provided storage buffer and then stored at 4°C.

Spatial Proteomics: Antibody Staining and Imaging

The tissue antigen retrieval and photobleaching steps were standardized across all spatial proteomics assays accordingly. Briefly, FFPE tissue slides were baked in an oven (VWR, 10055-006) at 70°C for 1 hour and then thoroughly deparaffinized by immersing in xylenes for 2 × 5 minutes. The slides were then subject to a series of graded solutions for rehydration using a linear stainer (Leica Biosystems, ST4020), with each step proceeding for 3 minutes: 3× xylene, 2× 100% EtOH, 2× 95% EtOH, 1× 80% EtOH, 1× 70% EtOH, 3× UltraPure water (Invitrogen, 10977-023), and finally left in UltraPure water (Invitrogen, 10977-023). Antigen retrieval was then performed at 97°C for 20 minutes with pH 9 Target Retrieval Solution (Agilent, S236784-2) using a PT Module (Thermo Fisher Scientific, A80400012), after which the slides were cooled to RT on the benchtop and washed in 1× PBS for 5 minutes. Tissue regions were circled with a hydrophobic barrier pen (Vector Laboratories, H-4000), rinsed in 1× PBS to remove residual ink, and then washed in 1× Tris-Buffered Saline with Tween-20 (TBS-T) prior to photobleaching and antibody

blocking. For assays that include staining with a biotinylated antibody, an extra biotin blocking step was included at this point with a commercial Biotin Blocking kit (BioLegend, 927301). Briefly, slides were first incubated with the avidin solution for 30 minutes at RT followed by two quick rinses 1× TBS-T and one 2-minute wash with 1× TBS-T and next incubated with the biotin solution for 30 minutes at RT followed by two quick rinses 1× TBS-T and one 2-minute wash with 1× TBS-T. Photobleaching and antibody blocking was then performed by first washing the slides in S2 Buffer (2.5 mmol/L EDTA, 0.5× DPBS, 0.25% BSA, 0.02% NaN_3, 250 mmol/L NaCl, 61 mmol/L Na_2HPO_4, 39 mmol/L NaH_2PO_4) for 20 minutes and then blocking using blocking buffer with donkey serum (BBDG) [5% normal donkey serum, 0.05% NaN_3 in 1× TBS-T wash buffer (Sigma, 935B-09)] supplemented with 50 μg/mL mouse IgG [diluted from 1 mg/mL stock (Sigma, I5381-10 mg) in S2], 50 μg/mL rat IgG [diluted from 1 mg/mL stock (Sigma, I4141-10 mg) in S2], 500 μg/mL sheared salmon sperm DNA (Thermo Fisher Scientific, AM9680), and 50 nmol/L oligo block (diluted from stock with 500 nmol/L of each oligo in 1× Tris-EDTA (TE) buffer (pH 8.0; Invitrogen, AM9849). Blocking was performed in a humidity chamber on ice during photobleaching, with the duration adjusted based on the tissue's collagen content, which is known to contribute to autofluorescence (83, 84). Specifically, tissue such as tonsil, B-cell lymphoma, kidney cancer, and lymph node were photobleached for 1 hour, whereas periodontal disease tissue, which contains higher collagen levels, underwent photobleaching for 2 hours. Photobleaching was performed using HappyLights (Verilux, VT22), with the temperature continuously monitored to ensure that it was kept below 40°C. After photobleaching and antibody blocking, tissues were stained and imaged accordingly based on the respective assay, as described below. Note that the photobleaching and blocking setup was different for the Orion (more details below).

CODEX. Tissues were stained for 1 hour at RT in a humidity chamber and then washed in S2 Buffer twice for 2 minutes each at RT. The slides were first fixed in 1.6% paraformaldehyde [PFA; diluted from 16% stock (EMS Diasum, 15740-04) in S4 Buffer (4.5 mmol/L EDTA, 0.9× DPBS, 0.45% BSA, 0.02% NaN_3, 500 mmol/L NaCl)] twice for 5 minutes each at RT, after which the slides were rinsed twice in 1× PBS followed by a 2 minutes wash in 1× PBS. The slides were next fixed with ice-cold methanol for 5 minutes on ice (while intermittently lifted to scrape off the hydrophobic barrier using a cotton-tipped applicator starting from the 3-minute time point), after which the slides immediately rinsed twice in 1× PBS followed by a 2 minutes wash in 1× PBS. The slides were finally fixed in 4 μg/μL of BS3 Final Fixative [diluted from 200 μg/μL stock (Thermo Fisher Scientific, 21580) in 1× PBS] twice for 10 minutes each in the dark at RT, after which the slides were rinsed twice in 1× PBS followed by a 2 minutes wash in 1× PBS.

To prepare the slides for imaging in the automated PhenoCycler Fusion platform, flow cells (Akoya Biosciences, 240205) were mounted by securely pressing them on each tissue slide for 30 seconds, followed by 10 minutes of incubation in 1× CODEX buffer [10 mmol/L Tris (pH 7.5), 0.02% NaN_3, 0.1% Triton X-100, 10 mmol/L $MgCl_2$-$6H_2O$, 150 mmol/L NaCl]. A reporter plate was also prepared for each tissue slide such that each well corresponds to each imaging cycle. Briefly, a 96-well black reporter plate (BRANDTECH, 781607) was prepared by filling each well with plate buffer (500 μg/mL sheared salmon sperm DNA in 1× CODEX buffer) supplemented with 1:300 (54.11 mmol/L) of Hoechst 33342 (Thermo Fisher Scientific, H3570) and adding complementary reporter oligos conjugated with ATTO550 or Alexa Fluor 647 [GenScript, high-performance liquid chromatography (HPLC)–purified] to a final concentration of 100 nmol/L each. The wells were then sealed using aluminum plate seal (Thermo Fisher Scientific, AB0626) and mixed by inverting the plate several times. Low DMSO (80% 1× CODEX buffer, 20% DMSO) and high DMSO (10% 1× CODEX buffer, 90% DMSO) buffers were also prepared fresh each run by mixing 1× CODEX buffer in DMSO (Sigma, 472301-4L), which was used by the PhenoCycler Fusion to strip and hybridize the reporter oligos. After imaging, the flow cell was removed prior to RNA probe hybridization by using a razor blade to pry the flow cell and gently scrape off any adhesive while repeatedly dipping in 1× PBS (video tutorials can be found in our study webpage: https://sizunjianglab.github.io/IN-DEPTH/). Personal protective equipment was worn at all times at this step. After the flow cell and adhesive were removed, slides were washed twice in 1× PBS.

For the data acquired by manual cycling imaging [Supplementary Fig. S1B (row 2)], the slides were first rinsed in 1× CODEX buffer followed by an initial stripping cycle in stripping buffer (25% 10× CODEX buffer, 75% DMSO) twice for 5 minutes each. The slides were subsequently washed twice in 1× CODEX buffer for 5 minutes each, incubated for 10 minutes with plate buffer supplemented with 100 nmol/L SYTO13 (Thermo Fisher Scientific, S7575), and then washed twice again for 5 minutes each in 1× CODEX buffer. The slides were then loaded into the GeoMx and scanned as the initial blank cycle. Subsequent cycles were carried out as follows: 2 × 5-minute incubation in stripping buffer, washing twice in 1× CODEX for 5 minutes each, 10-minute incubation in plate buffer supplemented with 100 nmol/L SYTO13 and three 100 nmol/L reporter oligos conjugated to Alexa Fluor 532, 594, or 647 (GenScript, HPLC-purified), and finally washing in 1× CODEX buffer twice for 5 minutes each. After all marker cycles, a final blank cycle stained with only 100 nmol/L SYTO13 was also included to ensure clearance of signal. All steps were performed at RT on the benchtop, all stripping and washing steps were performed in polypropylene Coplin jars (Ted Pella, 21038), whereas all reporter oligo incubations were performed in a humidity chamber. For all imaging, slides were loaded into the provided slide holder in the GeoMx and hydrated with 3 mL of Buffer S prior to operating the instrument. After imaging, slides were washed twice in 1× PBS.

SignalStar. The SignalStar reaction occurs in two rounds with four antibodies imaged per round and was performed using the commercial buffers (Cell Signaling Technology, 63043S) unless otherwise mentioned. Briefly, during each round, tissues were first incubated with SignalStar Amplification Solution 1 (1:100 of each SignalStar complementary oligo diluted in amplification buffer) for 2 hours (round 1 that includes 1:100 of each antibody) or 40 minutes (round 2 that does not contain antibodies) at 4°C and and then rinsed in 1× TBS-T for 30 seconds. Tissues were then fixed in 4% PFA [diluted from 16% stock (EMS Diasum, 15740-04) in 1× PBS] for 5 minutes at RT. After washing using UltraPure water (Invitrogen, 10977-023), eight rounds of amplification was performed accordingly using the corresponding amplification solution (1:50 of each amplification oligo diluted in amplification buffer), with a 30-second UltraPure water (Invitrogen 10977-023) rinse between each round of amplification. A 20-minute ligation step was performed accordingly using SignalStar Ligation Solution [50% Ligation Buffer, 2% T4 ligase (from a stock "5 U per mL"), and 1 mmol/L ATP prepared using UltraPure water (Invitrogen 10977-023)], followed by another 30 seconds Ultrapure water (Invitrogen 10977-023) rinse. Tissues were then stained with 1:300 of Hoechst 33342 (Thermo Fisher Scientific, H3570) for 5 minutes at RT, rinsed with 1× TBS-T, and coverslipped with ProLong Gold Antifade Mountant (P36930). Tissues were then imaged on the corresponding four-color channels using the PhenoCycler Fusion platform. After imaging, the coverslip was removed by dipping in 1× TBS-T followed by incubation with the SignalStar Fluorescent Removal Solution for 2 hours at 37°C and rinsed with UltraPure water (Invitrogen, 10977-023) for 30 seconds. To ensure complete removal of signal, tissues were stained with 1:300 of Hoechst 33342 (Thermo Fisher Scientific, H3570) for 5 minutes at RT

and then imaged again. The coverslip was similarly removed by dipping in 1× TBS-T. After both SignalStar reactions, slides were finally washed 5 times in 1× PBS to ensure complete removal of glycerol.

Polaris. An optimized tissue staining assay was performed on a BOND RX Autostainer (Leica Biosystems) using the Akoya Biosciences Opal tyramide signal system. The antibody/fluorophore pairings are CD8 on Opal Polaris 480 (1:50), PD-1 on Opal Polaris 690 (1:100), TIM-3 on Opal Polaris 620 (1:150), LAG3 on Opal Polaris 570 (1:50), CD20 on Opal Polaris 520 (1:150), and CD163 on Opal Polaris 780 (1:25)/TSA-DIG (1:100). Prior to imaging, slides were mounted using 1× PBS and sealed with nail polish. Whole-slide multispectral images were acquired at 20× magnification using the PhenoImager HT automated quantitative pathology imaging system (Akoya Biosciences), whereas the Inform 3.0 software was then used to deconvolute the multispectral images. After imaging, a cotton swab dipped with xylenes was used to remove the nail polish and unmount the coverslip, and slides were then washed twice in 1× PBS.

Orion. After antigen retrieval, the autofluorescence quenching, blocking, and antibody staining steps were instead performed according to the manufacturer's protocol. After antibody staining, tissues were coverslipped using 1× PBS and sealed with nail polish. Whole-slide images were acquired using the Orion (Rarecyte). After imaging, a cotton swab dipped with xylenes was similarly used to unmount the tissue, followed by washing twice in 1× PBS.

Spatial Transcriptomics: Probe Hybridization and Transcriptome Capture

At this point, all tissues were equilibrated in 1× PBS, including the control slides that were paused after antigen retrieval. Tissues were then hybridized for transcriptome capture accordingly based on the respective assay, as described below.

GeoMx. The RNA probe staining cocktail was prepared using the RNA Slide Prep kit (NanoString, 121300313) using the Human Whole Transcriptome Atlas detection probe set (NanoString, 121401102). The RNA probe cocktail was then applied to the tissue slides, sealed with a hybridization cover slip (EMS Diasum, 70329-40), and incubated overnight (around 18 hours) at 37°C. After RNA probe hybridization, tissue slides were first washed twice in Stringent Wash Buffer [2× saline sodium citrate (SSC); MilliporeSigma, S6639] in 50% formamide (MilliporeSigma, 344206-1L-M) for 5 minutes each at 37°C and subsequently washed twice with 2× SSC for 5 minutes each at RT on a Belly Dancer. Tissues were then stained with SYTO13 (100 nmol/L) for 10 minutes at RT and washed twice in 2× SSC for 2 minutes each at RT to visualize nuclear morphology. Slides were then scanned on the GeoMx for ROI selection while ensuring that the IN-DEPTH stained and control slides were always scanned in parallel. Square 484 × 484 μm ROIs were drawn for each experiment: 24 in DLBCL tissue [Fig. 1C (row 1)], 16 in kidney cancer tissue [Fig. 1C (row 2)], 18 in tonsil tissue [Supplementary Fig. S1C (row 2)], 8 in lymph node tissue [Supplementary Fig. S1C (row 3)], 18 in tonsil tissue [Supplementary Fig. S1C (row 4)], and 25 in uterine cancer tissue [Supplementary Fig. S1C (row 5)].

For the tonsil biological validation component (Fig. 2), sixteen 660 × 760 μm rectangular ROIs were selected on each adjacent tissue section with emphasis on lymphoid nodules (Fig. 2B; Supplementary Fig. S2B). The location of each ROI on the GeoMx was then recorded by their four vertices, and these coordinates were used to crop out one subregion for each ROI from the CODEX-to-GeoMx registered full-tissue segmentation mask. Within each subregion for each ROI, a segmentation mask for each annotated cell population was iteratively generated to enable cell type–specific RNA collection. Each cell type–specific segmentation mask was then converted into a binary mask by setting the pixel value of all the cell areas to 255 and pixel value for all background areas to 0. These masks were then reuploaded onto the GeoMx instrument to guide cell type–specific RNA genome-wide transcriptome extraction, ranked from the lowest to highest cell proportion within each ROI, such that transcript collection would proceed in this order.

For the EBV-positive versus EBV-negative DLBCL component (Figs. 4 and 5), the NanoString Human Whole Transcriptome Atlas detection probe was combined with a custom spike-in panel of probes against 14 targeted EBV genes (*EBER1*, *EBER2*, *EBNA1*, *EBNA2*, *EBNALP*, *LMP1*, *RPMS1*, *BALF1 BCRF1*, *BHRF1*, *BNLF2A*, *BNLF2B*, *BNRF1*, *BZLF1*). After 2× SSC and formamide washing, slides were stained with antibodies against TOX1/2, c-MYC for 1 hour at RT, followed by SYTO13 (100 nmol/L) streptavidin (used to visualize the biotinylated PD-L1 antibody) for 10 minutes at RT. The stained slides were then washed twice in 2× SSC for 2 minutes each at RT prior to GeoMx scanning. One 660 × 785 μm rectangular ROI was drawn for each patient core with emphasis on tumor-enriched regions. The location of each ROI on the GeoMx was similarly recorded by their four vertices and used to crop out the corresponding subregions, from binary 0/255 segmentation masks for each annotated cell population were iteratively generated, ranked, and uploaded onto the GeoMx for transcriptome extraction.

After transcriptome capture, unique molecular barcodes for the RNA probes were aspirated from each cell population to 96-well collection plates (NanoString, 100473), except for the first aspirate for each plate, which is the default negative control. Collection plates that were fully filled were dried according to official NanoString protocol and stored at −20°C until transcript collection for all other collection plates within each experiment was completed. Sequencing library preparation was then performed starting from the dried collection plates. Each aspirate was first resuspended in 10 μL of UltraPure water (Invitrogen, 10977-023) and then uniquely indexed using the Illumina i5 × i7 dual indexing system as part of the NGS library preparation kits (NanoString; 121400201, 121400202, 121400203, and 121400204). The PCR was prepared in 96-well PCR plates (Thermo Fisher Scientific, 4306737), in which each well contained 4 μL of aspirate, 1 μmol/L of each i5 and i7 primers, and 1× library preparation PCR Master Mix, adding up to 10 μL per well. The PCR conditions were 37°C for 30 minutes, 50°C for 10 minutes, 95°C for 3 minutes, followed by 18 cycles of 95°C for 15 seconds, 65°C for 60 seconds, 68°C for 30 seconds, followed by a final extension of 68°C for 5 minutes before holding indefinitely at 12°C. Next, 4 μL of PCR product from each well was pooled into DNA LoBind tubes (Eppendorf, 022431021) for purification, with 1 LoBind tube used per collection plate. For the first round of purification, 1.2× volume of AMPure XP beads (Beckman Coulter, A63881) were first added to the pooled PCR products and incubated at RT for 5 minutes. Beads were then pelleted on a magnetic stand (Thermo Fisher Scientific, 12321D), washed twice with 1 mL of 80% ethanol, and eluted with 54 μL of elution buffer [10 mmol/L Tris-HCl (pH 8.0), 0.05% Tween-20]. The second round of purification was performed using 50 μL of eluted DNA from the first round, incubated with 1.2× volume of AMPure XP beads, and washed twice in 1 mL of 80% ethanol. A final elution was done at 2:1 ratio of aspirate (number of wells) to elution buffer (volume in μL), and 0.5 μL of the final eluate was diluted in 4.5 μL of UltraPure water (Invitrogen, 10977-023; 1:10 dilution) to confirm library purity and concentration on the Agilent TapeStation.

For each experiment, the same concentration of each sublibrary (eluted in individual DNA LoBind tubes) was pooled into one LoBind tube to be sent for next-generation sequencing. PhiX sequencing control (Illumina, FC-110-3002) was added into the library, with amount adjusted based on the percentage of total reads allocated

for PhiX as per the sequencing platform used (5% on the NovaSeq X Plus, 20% on the NextSeq 2000). Paired-end sequencing was then performed on the NovaSeq X Plus (tonsil tissue experiments, Figs. 1 and 2) or NextSeq 2000 (DLBCL experiment, Fig. 4), with a total sequencing depth calculated as follows:

$$1.2 \times 100 \times Total\ ROI\ Area\ (\mu m^2) \times \frac{1}{100 - (PhiX\%)}$$

VisiumHD. Slides were first subjected to H&E staining and imaging as described in the next section. Afterward, tissues were dried at 37°C for 3 minutes using a thermal cycler. Tissues were then destained with 0.1 mol/L HCl at 42°C for 15 minutes, followed by 3× washes and incubations with TE buffer, and finally submerged in 1× PBS.

As the default VisiumHD workflow has a de-cross-linking step prior to probe hybridization, the control VisiumHD-only slide was subjected to de-cross-linking at 80°C for 30 minutes using the de-cross-linking mix provided by the manufacturer followed by probe hybridization at 50°C overnight following the manufacturer's protocols (10x Genomics, #1000668 and #1000466). For the CODEX–VisiumHD slide, tissues were incubated with 2 μg/mL Proteinase K (Thermo Fisher Scientific, AM2546) prepared with 1× PBS at 40°C for 20 minutes, followed by three washes in UltraPure water (Invitrogen, 10977-023). Tissues were then fixed in 10% Neutral Buffered Formalin (NBF; EMS Diasum, 15740-04) at RT for 1 minute, and the fixation process was stopped by incubating the tissue twice in NBF stop buffer (0.1 mol/L Tris and 0.1 mol/L glycine) for 5 minutes each at RT, followed by a 1× PBS wash for 5 minutes at RT. The tissues were then similarly subjected to probe hybridization (10x Genomics, #1000466) at 50°C overnight following the manufacturer's protocols.

Following posthybridization wash, the tissues were subjected to probe ligation at 37°C for 1 hour, washed with postligation wash (10x Genomics, #1000668) at 57°C for 5 minutes, and finally with 2× SSC buffer. The tissues were then stained with 10% eosin at RT for 1 minute and washed with 1× PBS. The tissues were loaded into the Visium CytAssist, adjusted to align with the slide subjected to Visium HD, followed by probe release. One square 6.5 × 6.5 mm ROIs were drawn for experiment presented in Fig. 1C (row 4), Supplementary Fig. S1C (row 6), and Fig. 6 due to the inherent size of each cassette (10x Genomics, #1000669 and #1000670). Probes were then extended with a thermal cycler and eluted with 0.08 mol/L KOH. Probes from each of the tissue samples were amplified with individual Dual Index TS Set A (10x Genomics, #PN-1000251) in a thermal cycler followed by PCR clean-up with SPRIselect Reagent (Beckman Coulter, #B23317). The libraries were quality controlled (QC) through High Sensitivity DNA Assay (Agilent Technologies) and sequenced paired-end on a HiSeq 2000 (Illumina).

CosMx. An incubation frame was first applied on each slide to ensure that liquid remains on the tissue surface. Tissues were then digested with 2 μg/mL Proteinase K (Thermo Fisher Scientific, AM2546) prepared with 1× PBS for 20 minutes at 40°C, followed by three washes in UltraPure water (Invitrogen, 10977-023). Fiducial solution [0.001% of fiducials in 2× Saline-Sodium citrate with Tween-20 buffer (SSC-T)] was applied afterward for 5 minutes at RT, which is immediately followed by tissue fixation in 10% NBF (EMS Diasum, 15740-04) for 1 minute at RT. The fixation process was quenched twice in NBF stop buffer (0.1 mol/L Tris and 0.1 mol/L glycine) for 5 minutes each at RT, followed by a 1× PBS wash for 5 minutes at RT. To block nonspecific probe and antibody binding, a 100 mmol/L NHS–acetate mixture was prepared immediately prior to application and incubated for 15 minutes at RT in a humidified chamber. Slides were then washed twice in 2× SSC for 5 minutes each at RT.

The RNA detection probes were prepared by denaturing at 95°C for 2 minutes using a preheated thermal cycler and then immediately chilled in an ice bucket for 1 minute. Note that different detection probe panels were used, with a 1k panel for Supplementary Fig. S1C (row 4) and a 6k panel for Fig. 1C (row 3). Afterward, the RNA probe cocktail was prepared according to the manufacturer's guidelines. The upper layer of the incubation frame was carefully removed to apply the probe cocktail while ensuring the liquid remains within the incubation frame boundary without any bubbles introduced, after which an incubation frame cover was used to seal the RNA probe cocktail within. Probes were allowed to hybridize at 37°C for 16 hours. After RNA probe hybridization, tissue slides were first washed twice in Stringent Wash Buffer (2× SSC MilliporeSigma, S6639) in 50% formamide (MilliporeSigma, 344206-1L-M)] for 25 minutes each at 37°C and subsequently washed twice with 2× SSC for 5 minutes each at RT on a belly dancer shaker. Tissues were then stained with SYTO13 (100 nmol/L) buffered in blocking buffer for 15 minutes at RT and washed in 1× PBS for 5 minutes, followed by staining with a designated antibody cocktail for 1 hour at RT to demarcate cell boundaries. After antibody staining, slides were washed 3 times in 1× PBS followed by another round of incubation using freshly prepared NHS–acetate mixture for 15 minutes at RT. Slides were then washed twice in 2× SSC for 5 minutes each at RT. Slides were then scanned on the CosMx for ROI selection while ensuring that the IN-DEPTH stained and control slides were always scanned in parallel. Square 500 × 500 μm ROIs were drawn for each experiment: 18 in DLBCL tissue [Fig. 1C (row 3)], 36 in tonsil tissue [Supplementary Fig. S1C (row 1)], 56 in DLBCL scSGCC cohort (Figs. 6 and 7), and 96 in DLBCL validation cohort (Fig. 6).

Xenium. Tissues were processed with minor modifications to the manufacturer's protocol (10x Genomics). Briefly, tissues were first digested with 2 μg/mL Protease K (Thermo Fisher Scientific, AM2546) prepared with 1× PBS for 20 minutes at 40°C, followed by three washes in UltraPure water (Invitrogen, 10977-023). Tissues were then immediately fixed with 4% PFA for 1 minute at RT. The fixation process was quenched twice in NBF stop buffer (0.1 mol/L Tris and 0.1 mol/L glycine) for 5 minutes each at RT, followed by a 1× PBS wash for 5 minutes at RT. Tissue slides were then assembled in the Xenium Cassette v2 following instructions from 10x Genomics. Tissues were then hybridized with priming oligos (10x Genomics, 2001224), which were prepared by denaturing at 95°C for 2 minutes and immediately in ice for at least 1 minute. Priming hybridization was performed at 50°C for 90 minutes, followed by two washes with PBS-T for 1 minute at RT, and subsequently with postpriming wash buffer (10x Genomics, 2001228) at 50°C for 30 minutes. Tissues were then washed 3 times with PBS-T at RT. After that, they were treated with RNase (10x Genomics, 3000593) at 37°C for 20 minutes, washed with 0.5x SSC-T 3 times for 1 minute at RT, and proceeded with polishing (10x Genomic, 20001230) at 37°C for 1 hour. Tissues were washed 3 times with PBS-T for 1 minute at RT and hybridized with the Xenium 5K Human PTP Panel Probes (10x Genomics, 2001225), which were prepared by denaturing at 95°C for 2 minutes and immediately in ice for at least 1 minute. Probe hybridization was performed at 50°C for 16 to 24 hours.

After probe hybridized for at least 16 hours, tissues were first washed twice with PBS-T for 1 minute at RT, followed with posthybridization wash (10x Genomics, 2000395) at 35°C for 15 minutes. After that, tissues were washed 3 times with PBS-T for 1 minute at RT. Probe ligation reaction (10x Genomics, 2000397 and 2000398) was then performed at 42°C for 30 minutes. Following that, tissues were washed 3 times with PBS-T for 1 minute at RT and immediately proceed to Amplification Enhancement (10x Genomics, 2001235) at 4°C for 2 hours. Tissues were then washed with Amplification Enhancer Wash Buffer (10x Genomics, 2001236) for 1 minute at RT and immediately proceeded to amplification (10x Genomics, 2000392) at 30°C for 90 minutes. Tissues were subsequently washed 3 times with TE buffer for 1 minute at RT.

To allow cell segmentation, tissues were first washed once with 70% ethanol for 2 minutes at RT, followed by twice with 100% ethanol for 2 minutes at RT and lastly with 70% ethanol for 2 minutes at RT for one time. Tissues were taken extra care not to dry up during this process. Tissues were blocked with the diluted Xenium Block and Stain Buffer (10x Genomics, 2002083) for 1 hour at RT. After that, tissues were stained with Xenium Multi-Tissue Stain Mix (10x Genomics, 2000991) at 4°C for 16 to 24 hours. After staining is completed, the tissues were first washed 3 times with PBS-T for 1 minute at RT, followed by incubation with the Xenium Stain Enhancer solution (10x Genomics, 2000992) for 20 minutes at RT. Tissues were then washed twice with PBS-T for 1 minute at RT, followed by incubating with Nuclei Staining Buffer (10x Genomics, 200762) in dark, for 1 minute at RT. Finally, tissues were washed 3 times with PBS-T for 1 minute at RT. The tissue slides were then scanned with the Xenium Analyzer. Each tissue, PCNSL [Fig. 1C (row 5)] and tonsil [Supplementary Fig. S1C (row 7)], was selected as a single FOV.

Detailed information and key metrics for each captured region across all transcriptomics datasets, including the number of cells profiled, mean transcripts per cell, and reads mapped to probe sets, are summarized in Supplementary Table S2.

H&E Staining and Imaging

VisiumHD. H&E staining was part of the VisiumHD protocol. Slides were first immersed twice in UltraPure water (Invitrogen, 10977-023) for 20 seconds each. H&E staining was performed a serial incubation in hematoxylin (StatLab, HXMMHPT), blueing buffer (StatLab, HXB00588E), and eosin (StatLab, STE0243) for 1 minute each at RT, with three UltraPure water (Invitrogen, 10977-023) washes between each incubation. Next, glycerol was used to coverslip the VisiumHD only slide, whereas UltraPure water (Invitrogen, 10977-023) was used to coverslip the CODEX–VisiumHD slide. Slides were then scanned using the Ocus40 slide scanner (Grundium, MGU-00003). After scanning, the coverslip was removed by immersing the slides in UltraPure water (Invitrogen, 10977-023) and continued with drying and destaining and detailed in the previous section.

GeoMx, CosMx, and Xenium. All slides were stored in 2× SSC at 4°C after transcriptome capture for H&E staining to visualize and confirm tissue morphology immediately after completing quality control evaluation of the captured transcripts. Slides were first equilibrated in UltraPure water (Invitrogen, 10977-023) at RT prior to staining with Modified Mayer's Hematoxylin (StatLab, HXMMHPT) for 5 minutes at RT, followed by rinsing 3 times with UltraPure water (Invitrogen, 10977-023). Slides were then treated with Bluing Solution (StatLab, HXB00588E) to develop the blue coloration and subsequently rinsed 3 times with UltraPure water (Invitrogen, 10977-023) at RT. The slides were then equilibrated in 95% ethanol for 1 minute prior to staining with a solution of eosin Y and phloxine B (StatLab, STE0243) for 1 minute, followed by rinsing by dipping 12 times each in three changes of fresh 95% ethanol. Finally, the slides underwent graded dehydration by dipping once in 70% ethanol, once in 100% ethanol, and once in two changes of xylenes. Excess xylenes was gently dabbed off and glass coverslips (Creative Waste Solutions, CSM-2450) were mounted with xylene-based mounting medium (OptiClear xylene, SSN Solutions, CSM1112). The slides were left to dry overnight at RT, after which they were scanned using the Ocus40 slide scanner (Grundium, MGU-00003).

Step-by-step protocol of each of the IN-DEPTH combinations presented in this study can be found in our study webpage: https://sizunjianglab.github.io/IN-DEPTH/. To support experimental planning, we have provided an interactive cost calculator in Supplementary Table S12, outlining estimated expenses for implementing IN-DEPTH across various combinations of spatial proteomics and transcriptomics platforms.

Histopathologic Review in This Study

All H&E-stained sections, multiplexed antibody-stained images generated from the spatial proteomics platforms, and the resulting computational phenotype maps were reviewed by board-certified pathologists S. Kure and S.J. Rodig. Pathology review was performed using same-slide H&E histology to verify tissue context and cellular morphology alongside multiplex staining patterns and marker-based cell type annotations. This review focused on multiple criteria, including (i) expected subcellular localization of markers (membranous, cytoplasmic, or nuclear), (ii) appropriate compartmental distribution (e.g., tumor, stroma, or vasculature), (iii) cellular morphology consistent with the assigned phenotype (such as lymphocyte size and shape or macrophage morphology), and (iv) identification of nonspecific or off-target staining patterns, including diffuse background haze, edge effects, autofluorescence-prone structures, or unexpected marker coexpression (e.g., T-cell markers on tumor cells). This rigorous histopathologic quality control ensured that multiplex staining patterns and computational annotations were biologically consistent and provided a reliable foundation for downstream spatial and molecular analyses.

Image Registration and Integration across Spatial Proteomics and Transcriptomics Platforms

Image Registration between CODEX and GeoMx. SIFT algorithm was used (85) for feature detection and feature description of the Fusion DAPI image and the GeoMx SYTO13 image. Then, a brute force matcher was used to match the features between the two images. A ratio test was used to determine if a specific match should be considered as a "good match." The source point (the CODEX image) and the destination point (the GeoMx image) of the "good matches" were used to calculate the affine transformation matrix that would register the CODEX image's coordinates into the GeoMx image's coordinate system. The software used and the specific hyperparameters for the algorithm and ratio test are in Supplementary Table S13. A step-by-step tutorial of this registration method can be found at https://github.com/SizunJiangLab/IN-DEPTH.

Single-Cell Integration between CODEX and CosMx. DAPI images from CODEX and CosMx were registered using the VALIS (v1.1.0; ref. 86) registration framework. The rTREs were calculated to assess alignment accuracy, with rTRE <0.01 as indicative of good alignment. Nonrigid warping transformations computed by the registrar were applied to CosMx transcript coordinates to transform them from the native CosMx coordinate system into the CODEX coordinate space. Cell segmentation masks generated from CODEX imaging were then used to assign transcripts to individual cells based on their warped spatial locations, thereby constructing CODEX-segmented single-cell CosMx RNA expression matrices. Cell type annotations derived from CODEX protein expression data were transferred to the corresponding RNA expression profiles. These CODEX-segmented and annotated single-cell transcriptomic datasets were subsequently used for downstream integrative analyses.

Spatial Proteomics: Data Processing

Cell Segmentation. Segmentation for all tissues were performed only with the CODEX images using the Mesmer model of DeepCell (v0.12.2; refs. 87, 88), with maxima_threshold set to 0.075 and interior_threshold set to 0.05. The nuclear channel input of MESMER was DAPI for all datasets. The membrane channel input of Mesmer for the tonsil dataset (Fig. 2) was a summation of CD11b, CD68, CD20, CD163, CD31, and CD3. Those for the EBV-positive versus EBV-negative DLBCL datasets (Figs. 4–7) were a summation of HLA1, HLA-DR, and CD31 and HLA-DR, CD3, and CD68, respectively. To further demonstrate the adaptability and robustness of

the MESMER segmentation workflow across diverse tissue architectures, the same parameters were applied to representative nonimmune tissues, including periodontal and PCNSL samples that contain stromal and structured cell populations (Fig. 1C; Supplementary Figs. S1C and S2G). The membrane channel inputs for these datasets included KRT14, VIM, α-SMA, CD31, CD3, CD68, and HLA-I (periodontal) and PVLAP, GFAP, HLA-I, and CD31 (brain).

Single-Cell Feature Extraction. For each marker, the pixel value within the area of each cell (determined by the segmentation mask) was summed and then divided by the area of each cell, and the resulting cell-size scaled sum was set as the expression value for a given marker. For the DLBCL dataset (Fig. 4) in which 3 markers were acquired on the GeoMx, the segmentation mask generated from the CODEX image was applied to the GeoMx image to ensure that the same cell imaged between the two instruments contained the same cell label, from which the cell features were similarly extracted and scaled to cell size. Finally, the scaled single-cell features extracted from the Fusion and GeoMx images were joined together by cell label and tissue core ID.

Cell Clustering and Annotation. The extracted features were first scaled to a standardized range of [0,1], and cell phenotyping was then performed through an iterative clustering and annotating process with PhenoGraph (k = 45, seed = 23; ref. 89). For the tonsil dataset (Fig. 2), the 12 phenotyping markers used were CD20, PAX5, BCL6, CD3, CD8, CD4, FOXP3, CD11c, CD31, CD68, CD163, and CD11B, which allowed the annotation of $BCL6^+$ B cells, $BCL6^-$ B cells, CD4 T cells, CD8 T cells, endothelial cells, Tregs, DCs, immune-active macrophages, immunosuppressive macrophages, and other myeloids. For the EBV-positive versus EBV-negative DLBCL dataset (Fig. 4), the phenotyping markers used were CD20, PAX5, CD3, CD8, CD4, FOXP3, CD11C, CD31, CD68, and CD163, which allowed the annotation of CD4 T cells, CD8 T cells, endothelial cells, Tregs, DCs, immune-active macrophages, immunosuppressive macrophages, and tumor cells. For the EBV-positive versus EBV-negative DLBCL dataset (Fig. 6), the phenotyping markers used were PAX5, BCL6, CD3, CD8, CD4, FOXP3, CD68, CD163, and α-SMA, which allowed the annotation of CD4 T cells, CD8 T cells, Tregs, immune-active macrophages, immunosuppressive macrophages, tumor, and stromal cells. Cells that showed unclear marker enrichment patterns were annotated as "Other" cells.

During the annotation process, clustering results were first visualized using a heatmap showing the *Z*-score of each marker within each cluster. This was used as a basis to annotate each cluster based on their marker *Z*-score combinations while visually inspecting the original images to confirm annotation accuracy. After an initial round of clustering with PhenoGraph was performed, clusters with clear enrichment patterns were annotated, whereas clusters with mixed patterns underwent additional rounds of clustering and annotation using a targeted set of phenotyping markers. This process was iterated until all identifiable cells were annotated. To visualize and confirm the assigned annotations, Mantis Viewer (v1.2.0-beta.1; Zenodo; available from https://doi.org/10.5281/zenodo.4909620) was utilized to overlay the annotation onto the segmentation mask and the marker image for visual inspection. Final cell type annotations were subsequently reviewed and validated by board-certified pathologists S. Kure and S.J. Rodig using multiplexed marker expression patterns and corresponding H&E-stained sections, as described in the "Histopathologic Review" section.

For the tonsil experiment (Fig. 2), we annotated one tissue section using the above-described procedure. Leveraging upon the advantage of adjacent tissue sections and the reproducible high-quality tissue staining, annotation of the adjacent section was guided by Machine learning for Analysis of Proteomics in Spatial biology (MAPS) (90), followed by further refinement using the same procedures as described above.

Image Processing. For functional markers included in the analysis in Fig. 4 (HLA-1, HLA-DR, CD45RO, CD45RA, KI-67, PD-1, LAG3, Granzyme B), the 16-bit intermediate QPTIFFs, generated by the PhenoCycler-Fusion, were used to ensure optimal dynamic range of data. The QPTIFFs were processed firstly by subtracting the last blank cycle scaled by the ratio between current channel cycle and total cycle number, that is,

$$X'_{i,j} = X_{i,j,0} - \left(\frac{i}{N}\right) \times X_{\varepsilon}$$

where $X'_{i,j}$ is the blank-subtracted image of marker j in cycle i; $X_{i,j,0}$ is 16-bit intermediate image of marker j in cycle i; and X_{ε} is the last blank cycle. Then, the last blank-subtracted image was processed in ImageJ using the "Math" and "Subtract Background" functionalities under "Process":

1. Subtract the mean pixel value of the image to get rid of most of the "salt and pepper" noise.
2. Subtract the background generated by the sliding paraboloid algorithm with a 5-pixel radius.

As GeoMx images were outputted as 16-bit images by default and were already fully processed internally by the instrument, TOX and PD-L1 were not processed by the abovementioned pipeline. Finally, for each core and each marker, a lower bound and an optional upper bound (in case of high pixel intensity artifacts) were applied to remove the remaining unspecific staining, noise, and artifacts. The lower bound and upper bound were determined by visual inspection of the images in QuPath, and the values can be found in Supplementary Table S14.

Note that cell phenotyping was performed based on the final 8-bit QPTIFF generated by the PhenoCycler-Fusion. As the 8-bit QPTIFF was processed completely by the PhenoCycler-Fusion's software, the blank subtraction and the ImageJ processing were not applied. However, similar to the 16-bit images, lower bounds were set for each core and each marker in order to get rid of as much of unspecific staining (e.g., nuclear signal of a supposedly membrane marker) as possible. The lower bound values can be found in Supplementary Table S14.

To improve the signal-to-noise characteristics of the EBV-positive and EBV-negative DLBCL data in Fig. 6, the 16-bit raw QPTIFF images acquired by the PhenoCycler-Fusion were processed using a custom background correction pipeline. Briefly, the Yen thresholding method (91) was applied to the blank channel to identify regions exhibiting unusually high fluorescence intensity, which are commonly associated with red blood cell–related artifacts. These regions were masked and excluded from subsequent analyses across all channels. The resulting filtered blank channels were then used to estimate background-related effects via a nuisance regression framework, and the inferred effects were subtracted from each marker image. In addition to autofluorescence captured by the blank channel, nonspecific signal was further reduced in a marker-dependent manner by estimating a background threshold from pixels outside the marker's expected subcellular localization (e.g., nonmembrane regions for membrane markers). Finally, the autofluorescence-corrected images were adjusted by subtracting the estimated background threshold, yielding images with reduced contributions from autofluorescence and nonspecific staining.

Marker Preprocessing. Functional markers (HLA-1, HLA-DR, CD45RO, CD45RA, KI-67, PD-1, LAG3, Granzyme B, TOX, PD-L1, C1Q, LMP1) were scaled by the respective median nuclear signal (DAPI for markers captured on Fusion and SYTO13 for markers captured on GeoMx) of each tissue sample in order to adjust for different binding efficiency of markers. Then, a global min–max scaling was applied to scale the marker expression levels to be within [0,1].

For phenotyping markers (PAX5, CD20, CD3, CD8, CD4, FOXP3, CD11c, CD68, CD163, CD31), the same median nuclear signal scaling was applied. Then, the markers were further scaled within each tissue sample by a (0.001, 0.999) quantile scaling and then truncated at 0 and 1. Unlike the functional markers, the phenotyping markers were scaled at a local level to compensate for tissue samples with an overall weaker pixel intensity.

Spatial Proteomics: Analysis

Marker Enrichment Heatmap. The marker enrichment heatmap showed the *Z*-score of a given (marker, cell type, EBV status) tuple. In other words, it showed how many standard deviations away is the mean of marker A expression of cell type B given an EBV condition from the population mean of marker A expression:

$$Z_{i,j,k} = \frac{\left(\mu_{i,j,k} - \mu_i\right)}{\sigma_i}$$

where $Z_{i,j,k}$ stands for the *Z*-score for marker i, cell type j, and EBV status k; $\mu_{i,j,k}$ stands for the mean expression for marker i, cell type j, and EBV status k; μ_i stands for the population mean of marker i; and σ_i stands for the population standard deviation of marker i.

Cell Type Proportion and Enrichment. Cell type enrichment was presented as $\log_2$ of the ratio between the proportion of cell types in EBV-positive and EBV-negative DLBCL samples:

$$log_2 \frac{P_{i,EBV+}}{P_{i,EBV-}}$$

where $P_{i,EBV+}$ is the proportion of cell type i in EBV-positive and $P_{i,EBV-}$ is the proportion of cell type i in EBV-negative.

Dysfunction Score. The T-cell dysfunction score was calculated by summing the expression intensities of LAG3, TOX1/2, and CD45RO and subtracting the intensities from CD45RA, KI-67, and GZMB. PD-1 was not included due to its lower staining quality in this tissue cohort, as well as its additional biological function as an activation marker (92).

$$S = \sum_{i \in \mathcal{M}+} X_i - \sum_{j \in \mathcal{M}-} X_j$$

where S stands for the dysfunction score; X_i and X_j stands for the expression level of marker i or marker j of a cell; $\mathcal{M}+$ stands for a set of markers that signify contributive effects to cell dysfunction, $\mathcal{M}+ = \{\text{LAG3, CD45RO, TOX}\}$; and $\mathcal{M}-$ stands for a set of markers that signify counteractive effects to cell dysfunction, $\mathcal{M}- = \{\text{CD45RA, KI-67, GZMB}\}$.

Cell Motif Analysis. For a tissue sample, the spatial coordinates of individual cells were defined by the centroids of their segmentation masks. Using these centroids, a Delaunaay triangulation was first performed; then a graph was constructed using the simplices. Two nodes (cells) were connected if their Euclidean distance was less than or equal to 20 μm. For each node of interest, for example, CD4 T-cell nodes, the one-hop neighborhood, which comprised all directly connected cells, were identified. The composition of a given one-hop neighborhood was summarized into a vector representing the count of each cell type. For example, if four cell types were present in the dataset and a one-hop neighborhood consists of two CD4 T cells and one CD8 T cell, the summary vector would be (2, 1, 0, 0). These vectors were then clustered using *k*-means clustering to identify recurrent spatial motifs.

To determine the optimal number of clusters (k) for CD4 T-anchored motifs, the elbow method was applied to assess within-cluster variance reduction (Supplementary Fig. S7B). A distinct inflection point was observed between $k = 5$ and 7, indicating this range as optimal. *k*-means clustering was subsequently performed for $k = 5$, 6, and 7, revealing consistent motif patterns across values, with only minor subdivision in motif 3 at higher k (Supplementary Fig. S7C). As the subclusters were highly similar in their immediate cellular composition and spatial context, $k = 5$ was selected as the final clustering solution, balancing resolution, and interpretability (Fig. 4J; Supplementary Fig. S7D).

Binarization of LMP1 and C1Q. Visual inspection of CODEX multiplexed images for LMP1 and C1Q was first performed to identify TMA cores with reliable LMP1 staining patterns, defined by bright ring-like staining in tumor B cells. Notably, even in cores with appropriate LMP1 staining, most LMP1-negative cells exhibited faint nuclear signal. Therefore, comparisons based on absolute LMP1 expression values could confound true biological signal with technical background noise. To address this, the Yen thresholding method (91) was applied to [0,1]-transformed LMP1 expression values (see "Marker Preprocessing") on a per-core basis. Cells with LMP1 expression levels exceeding the threshold were classified as LMP1-positive.

A similar strategy was applied to binarize C1Q expression. The log-transformed distribution of C1Q expression exhibited clear bimodality after trimming values of 0 and 1. In addition to the Yen threshold, the position of the second peak in the distribution was determined. The final threshold was defined as follows:

$$T_{final} = \max\left\{T_{Yen}, T_{second\ peak}\right\}$$

Cells with C1Q expression levels exceeding this final threshold were classified as C1Q-positive.

Generalized Linear Model. Two negative binomial regression models were fitted to explore the effect of EBV status, membership of motif, and their interaction on immune-active and immunosuppressive macrophage counts within the one-hop neighborhood anchoring on CD4 T cells (Supplementary Fig. S7A and S7E; Supplementary Table S6). The proposed model is as follows:

$$\ln E[Y_i] = \beta_0 + \beta_1 I_{EBV} + \sum_{i=2}^{5} \beta_i I_i + \sum_{i=1}^{4} \gamma_i J_{EBV,i}$$

where

$$I_{EBV} = \begin{cases} 1, EBV_+ \\ 0, EBV_- \end{cases},$$

$$I_i = \begin{cases} 1, \text{Motif } i \\ 0, \text{Not Motif } i \end{cases},$$

$$J_{EBV,i} = \begin{cases} 1, EBV_+, \text{Motif } i+1 \\ 0, \text{Not } EBV_+, \text{Motif } i+1 \end{cases}$$

An ordinary least square model was fitted to explore the effect of EBV and LMP1 status on the proportion of $C1Q^+$ macrophages in the 1-hop neighborhood of tumors, accounting for the confounding effect of TMA membership. A graph was first built for each FOV using Delaunay triangulation.

For each tumor, the proportion of each cell type in its 1-hop neighborhood was calculated and transformed using $\log_{10}$ (x + 1e−6). Then, a linear model

$$\log_{10}\left(C1Q^+ \text{macrophage proportion} + 10^{-6}\right) \sim \text{TMA} + \text{Viral status}$$

was fitted. Viral status (EBV^-, EBV^+ $LMP1^+$, EBV^+ $LMP1^-$) was defined to avoid multicollinearity between EBV status and LMP1 status as all cells in an EBV^- tissue must be $LMP1^-$.

Tumor Density Score. Tumors were first classified into three categories:

- EBV-positive, LMP1-high: if a tumor is in an EBV-positive sample and its LMP1 expression is greater than the median LMP1 expression of all tumors
- EBV-positive, LMP1-low: if a tumor is in an EBV-positive sample and its LMP1 expression is less than or equal to the median LMP1 expression of all tumors
- EBV-negative: if a tumor is in an EBV-negative sample

Tumor density score was then calculated as described in bioRxiv 2024.03.05.583586. Briefly, within each of these categories, for each nontumor cell, three tumor scores were calculated, one for each tumor class. The score was calculated based on a cell's distance to tumors within a closed neighborhood of radius r. Let $J = \{1,\ldots,m\}$ denote the indices of all the tumors in the dataset and $d_{i,j}$ denotes the distance from the cell i to tumor j. Then, the tumor score is calculated as follows:

$$S_i = \sum_{j \in \{k | d_{i,j} \le r\}} \frac{1}{d_{i,j}}$$

Then, the score was transformed into

$$S_i' = \exp(-S_i)$$

Spatial Transcriptomics: Data Processing

Batch Correction: GeoMx Data. The demultiplexed FASTQ output files from next-generation sequencing were used to map and quantify the human probes (and EBV probes for DLBCL data) through the GeoMx Data Analysis software pipeline (8). The .dcc files produced were then uploaded onto the GeoMx to generate gene counts tables using the default "QC" and "Biological probe QC" settings without filtering out any genes.

Cell type annotations distinguished multiple T cells (CD4T, CD8T), macrophage, endothelial, and tumor, as shown in Supplementary Fig. S5. Gene expression data from both cohorts were then combined into a single, unified count matrix with genes as rows and spatial segments (ROI × cell type) as columns. Segments matched with fully annotated metadata were retained. Raw gene counts were then normalized, and for the EBV-positive versus EBV-negative DLBCL dataset (Fig. 4), additional rigorous batch correction steps were adopted as described below.

Rationale for Batch Correction. Overall, GeoMx datasets often involve samples from multiple cohorts and experimental batches, each potentially introducing technical artifacts that can obscure true biological variation. In the context of our cohort of patients with DLBCL, in which samples are derived from diverse sources, correcting for batch effects is critical to ensure that the observed differences in gene expression reflect underlying biology rather than technical or sample processing discrepancies. Batch correction methods help remove these unwanted sources of variation while preserving genuine differences arising from biological conditions and cell types. This step is important for downstream analyses such as differentially expressed gene (DEG) analysis and gene signature validation, as it ensures that identified biomarkers and signatures are robust and not confounded by technical and other unwanted factors.

Normalization Methods, Negative Control Genes, and Unwanted Covariant Factor Preparation for Batch Correction. The standR (v1.9.3; ref. 93) pipeline was used for normalization and reducing patient-level batch effects using the RUV4 method. Two normalization methods were adopted, including log counts-per-million (CPM) reads via the logNormCounts function of scater package (v1.28.0) and quantile normalization via GeoMxNorm function of standR. Batch effect correction was implemented via a grid searching strategy to optimize parameter combinations for minimizing individual patient-level variations (e.g., tissue sources) while retaining biological variations due to EBV condition and cell types. Five grids of the number of negative control genes (NCG) were selected: 1,000; 2,000; 3,000; 4,000; and 5,000 via, findNCGs function. The three grids of the number of unwanted factors (i.e., *k*-values) for the RUV4 method (94) were set to 1, 2, and 3 using the GeoMxBatchCorrection function. The result of each batch correction run was a normalized and adjusted expression matrix for DEG.

DEG Parameter Settings. Following batch correction, a two-step approach was employed to evaluate and refine DEG parameters. First, the suitability and effectiveness of batch correction strategies were assessed by examining their ability to produce biologically interpretable DEGs. To do this, pairwise comparisons were conducted between key cell populations of interest (e.g., tumor, $CD4^+$ T, $CD8^+$ T, and macrophage compared with endothelial cells) across different EBV status subsets (EBV-positive, EBV-negative, and combined). These contrasts aimed to reveal condition-dependent DEGs that are biologically meaningful.

Second, the DEG model parameters were optimized to recover cell type–specific gene signatures robustly. DEG analyses were performed using a pipeline that integrated edgeR (v3.42.4; ref. 42) and limma (v3.56.2; ref. 95). The modeling framework allowed for the inclusion of weight matrices from RUV4 in the design matrix of the linear model as covariates. Four confounder sets were tested:

1. No confounders
2. One confounder if the *k*-value is equal to or greater than 1: one weight matrix from RUV4
3. Two confounders if the *k*-value is equal to or greater than 2: two weight matrices from RUV4
4. Three confounders if the *k*-value is equal to 3: three weight matrices from RUV4

Additionally, each confounder set was tested with two scenarios: with and without controlling for cell-type abundance (i.e., including or excluding cell counts as a covariate). DEGs were then identified using moderated linear modeling (limma) and empirical Bayes shrinkage. Significance thresholds included an adjusted *P* value threshold of 0.01. *P* values were adjusted for multiple testing using the false discovery rate (FDR) method.

Benchmarking and Signature Validation. To systematically assess the combined influence of batch correction and DEG model parameters, all combinations ($N = 540$) of number NCGs, *k*-values for unwanted variation, EBV status subsets, confounder sets, and cell type abundance adjustments were evaluated. The DEGs identified under each parameter setting were then evaluated against known cell type–specific signatures. Signatures (Supplementary Table S3) included well-established lineage and function markers for CD4 T cells (96), CD8 T cells (96), macrophages (97, 98), and DLBCL tumor cells (99). Enrichment of known markers within each DEG list was assessed via hypergeometric tests, confirming whether the parameters chosen successfully recovered expected biological signatures.

CosMx Data. The acquired data were automatically uploaded onto the AtoMx spatial informatics platform, with the normalized transcript counts of each FOV generated in the platform, as well as image preprocessing and feature extraction, To identify single-cell features, a pretrained neural network model Cellpose was used for segmentation (100). Single-cell RNA expression profiles were generated by counting transcripts of each gene falling within different segmented areas. Cells with fewer than 20 total transcripts were removed from downstream data analysis.

VisiumHD Data. The demultiplexed FASTQ output files from next-generation sequencing were used to map and quantify the human probes through the 10x Genomics Space Ranger v3.1.1 count pipeline. Manual alignment and tissue detection was first performed with 10x Genomics Loupe Browser v8.0.0 using the CytAssist image and the H&E-stained microscope image. These images, together with the human transcriptome reference GRCh38, Visium probe set v2.0, and the FASTQ files, were input into the Space Ranger's count pipeline. Due to varying ROI sizes in the tissue samples, unique molecular identifier (UMI) counts were normalized by the number of bins

within each ROI, with a scaling factor of 10,000. Note that batch effect correction was similarly not performed for the analysis in Fig. 1C and Supplementary Fig. S1C.

SGCC Development Rationale

To quantify spatial relationships between two cell phenotype patterns defined on a tissue graph, we developed SGCC based on the SGWT. Existing approaches for spatial cross-correlation, such as Pearson correlation, bivariate Moran I, local spatial cross-correlation indices, and cross-variograms, ignore spatial topology, rely on predefined spatial weights or distance assumptions, or provide single-scale or purely local summaries, limiting their ability to capture heterogeneous spatial organization on irregular graphs. A detailed comparison between SGCC and existing cross-correlation approaches is provided in the Supplementary Note, together with their underlying assumptions, limitations, and interpretability for graph-structured spatial data. In contrast, SGWT provides a principled spectral decomposition of graph signals into low-frequency components that encode smooth, global tissue organization determined by graph structure, and band-pass components that capture localized, fine-scale spatial variations. This multiscale representation is particularly important for spatial omics data, in which biologically meaningful patterns may arise simultaneously at tissue-domain and cell–cell interaction scales. Building on SGWT, SGCC compares two graph signals directly in the spectral domain by computing cosine similarities between their low-frequency and band-pass coefficients and combining them using energy-based weights, yielding a bounded and interpretable metric that captures multiscale spatial co-occurrence or segregation on irregular spatial graphs.

SGCC addresses these limitations by leveraging SGWT to analyze graph signals in the frequency domain. The rationale for SGCC lies in its ability to

1. *Extend beyond single-signal analysis*: Although spatial autocorrelation measures, such as Moran I, evaluate the spatial coherence of a single signal, SGCC quantifies the cross-correlation between two graph signals using a spectral graph method, capturing their spatial relationship in terms of complementarity or co-occurrence.
2. *Incorporate graph structure*: SGCC operates directly on graph-structured data, integrating spatial adjacency information into the analysis. This allows it to adapt to both regular (e.g., pixel grids) and irregular (e.g., cell–cell adjacency) spatial graphs, ensuring an accurate representation of spatial relationships.
3. *Use band-pass signals to study spatially organized structures*: Referring to our previous study, SpaGFT (17), we demonstrated that a k-bandlimited signal is a smooth and slow graph signal, which can be biologically defined as a spatially organized structure (e.g., the GC pattern of the tonsil). Such a signal can be effectively captured by the first k FMs, which are eigenvectors of the graph Laplacian that represent broad, large-scale patterns in the graph data, such as gradual and organized distributions. This low-frequency representation is well suited for describing global variables and overarching tissue organization. However, k-bandlimited signals alone cannot adequately capture all biologically meaningful organization, as they do not focus on the localized, cellular-level features in an observation window. In contrast, band-pass signals are not restricted to the first k FMs and can include other frequency components (e.g., high frequency) that encode fine-scale, localized, or even abrupt variations. Although some high-frequency content reflects technical noise, others capture biologically relevant microstructures, such as sharp tissue boundaries and cellular interactions that would be lost if only low-frequency signals were considered. By integrating both approximately k-bandlimited and band-pass signals, SGCC transcends the limitations of a merely low-frequency representation and achieves a balanced, multiscale view of tissue architecture, capturing both smooth global organization and fine-grained spatial complexity.
4. *Address multiscale FOV in multiomics profiling*: In multiomics profiling, such as IN-DEPTH, a multiscale structure arises naturally from the flexibility of defining different FOVs. A larger FOV captures broad organizational domains and global tissue patterns, which are well described by approximately k-bandlimited (low-frequency) signals. Conversely, a smaller FOV highlights localized cell–cell interactions and niche-specific features, which manifest as band-pass signals with higher-frequency content. Therefore, a robust method must capture both signal types simultaneously, ensuring that multiscale features, from smooth global organization to fine-grained local interactions, are faithfully represented.
5. *Provide a quantitative and interpretable metric*: SGCC is implemented as a weighted cosine similarity between the FCs of two graph signals after SGWT. The similarity is computed separately for low-frequency (scaling) and high-frequency (wavelet) components and then combined using energy-based weights that reflect the relative contribution of each band. In this way, SGCC provides a balanced measure that captures both smooth, large-scale spatial organization and fine-scale, localized interactions.
6. *Enable cross-sample comparisons*: To allow consistent measurement across multiple samples, SGCC represents spatial data in the form of a spot graph, in which each node corresponds to a bin containing a small group of cells rather than individual cells. This approach provides two key advantages. First, by binning cells into spots, spatial signals from heterogeneous cell populations are aggregated, reducing single-cell noise and enhancing biological interpretability, as spots can be viewed as functional tissue units rather than isolated cells. Second, constructing spot graphs ensures that all FOVs are projected into a shared eigenvector space derived by the graph Laplacian. This unified spectral basis makes cross-sample comparison feasible: Fourier or wavelet coefficients from different samples can be directly aligned, enabling quantitative evaluation of spatial similarity across conditions. In this way, SGCC leverages spot graphs to balance biological meaning with mathematical comparability, ensuring both local tissue structure and global tissue organization are faithfully captured across datasets.
7. *Link spatial patterns to functional insights*: As IN-DEPTH generates matched spatial omics datasets on the same tissue slides, SGCC can first be applied to rank FOVs by their spatial cross-correlation patterns (i.e., spatial factors). When these spatial factors are established, they can be used in a *post hoc* analysis to connect phenotype-level spatial organization with molecular programs. Specifically, SGCC enables a spatial analysis that identifies covarying genes or gene programs associated with the observed phenotypic arrangements. This integration bridges spatial phenotypes (e.g., paired immune–tumor patterns) with functional interpretation, providing mechanistic insights into how spatial organization shapes gene expression and tissue function.

SGCC Development

All implementation details, including graph construction, parameter selection, SGWT filter design, and energy normalization, are provided in the SGCC Supplementary Note. The key steps are described below.

Let $G=(V,E)$ denote a spatial graph constructed from binned tissue spots, with $n=|V|$ nodes. For each cell phenotype p, a graph signal is defined as follows:

$$\mathrm{f}_p=\left(f_{p1},f_{p2},\ldots,f_{pn}\right)^{\mathrm{T}}\in\mathbb{R}^n, \quad \text{(A)}$$

where f_{pi} represents the abundance of phenotype p at node i. Given the graph Laplacian L, spectral decomposition yields the following:

$$\mathrm{L}=\mathrm{U}\Lambda\mathrm{U}^{\mathrm{T}} \quad \text{(B)}$$

where columns of $U = [\mu_1, \ldots, \mu_n]$ are graph FMs ordered by increasing eigenvalues λ_k, corresponding to increasing spatial frequencies. The GFT of f_p is as follows:

$$\hat{f}_p = U^T f_p \quad \text{(C)}$$

To capture spatial structure across scales, a SGWT is applied using a low-pass (scaling) filter and a set of band-pass (wavelet) filters. This yields low-frequency coefficients $\hat{f}_p^{(\text{low})}$, which encode smooth, global tissue organization, and band-pass coefficients $\hat{f}_p^{(\text{band})}$, which encode localized and fine-scale spatial variations.

Given two graph signals f_1 and f_2, cosine similarities are computed separately for the low-frequency and band-pass components:

$$c_{\text{low}} = \frac{\left\langle \hat{f}_1^{(\text{low})}, \hat{f}_2^{(\text{low})} \right\rangle}{\left\| \hat{f}_1^{(\text{low})} \right\| \left\| \hat{f}_2^{(\text{low})} \right\|}, \quad c_{\text{band}} = \frac{\left\langle \hat{f}_1^{(\text{band})}, \hat{f}_2^{(\text{band})} \right\rangle}{\left\| \hat{f}_1^{(\text{band})} \right\| \left\| \hat{f}_2^{(\text{band})} \right\|} \quad \text{(D)}$$

To balance contributions from global and local spatial structure, energy-based weights w_{low} and $w_{\text{band}} = 1 - w_{\text{low}}$ are defined from the corresponding spectral energies of the two signals. The SGCC is then computed as follows:

$$\text{SGCC} = w_{\text{low}} c_{\text{low}} + w_{\text{band}} c_{\text{band}} \quad \text{(E)}$$

which is bounded in $[-1,1]$. Positive SGCC values indicate multiscale spatial co-occurrence of the two cell phenotypes, whereas negative values indicate systematic spatial segregation. Values near zero may arise from cancellation between low-frequency and band-pass similarities rather than absence of spatial association.

SGCC Validation Analysis

Simulation 1 (Ring Pattern). The simulation process begins by defining a regular 60 by 60 grid to represent the spatial domain, with each cell having x and y coordinates. An inner circle is generated with a fixed radius from a predefined range (2.5, 5, 7.5, 10, 12.5, 15, 17.5, and 20), centered in the middle of the grid ($x = 30, y = 30$). To simulate the dynamic behavior of an outer ring shrinking toward the inner circle, a sequence of radii is defined for the outer ring in 10 incremental steps, starting from a large initial radius and progressively decreasing to slightly larger than the inner circle's radius. For each step, the grid is analyzed to classify points as either inside the inner circle, within the outer ring (defined as the area between the shrinking outer radius and the inner circle), or outside both regions. The spatial distribution of these classifications is aggregated for all steps, resulting in a set of data that capture the interaction between the inner circle and the shrinking outer ring at different stages of the simulation. This process enables the generation of 80 datasets to demonstrate local and global complementary patterns.

Simulation 2 (Moving Pattern). The simulation method generates data to model the spatial interactions between two dynamically moving circular regions on a 60 by 60 grid. For each simulation, the radius of the first circle is varied within a specified range (6, 7, 8, 9, 10, 11, 12, 13, and 14), whereas the radius of the second circle is set to be 1.5 times the radius of the first circle. Initially, the centers of the two circles are positioned symmetrically at a distance of 30 U from the centerline of the grid. Over 10 incremental steps, the centers of the circles move inward toward the grid's center. At each movement step, the Euclidean distance from every grid point to the centers of the circles is calculated to determine whether a point lies within the first circle, the second circle, both circles, or outside both. This classification is updated at each step to reflect the movement of the two circles. The resulting data for each simulation step include the binary indicators for points being within each circle and the overlap between the two. This process enables the generation of 80 datasets to demonstrate moving pattern of two cell types.

Benchmarking and Parameter Sensitivity Analysis of SGCC. We benchmarked SGCC on simulated multiscale spatial patterns by treating the two simulated signals as graph signals on a shared two-dimensional grid and computing SGCC after SGWT in a common spectral basis. A k-nearest neighbor (kNN) spatial graph was constructed with $k = 20$ using a normalized graph Laplacian, and the number of retained eigenvalues was determined by a k-bandlimited knee-point diagnostic, with a full-spectrum fallback when bandlimitedness was not supported. SGWT was implemented using the BioGSP package, employing a heat-kernel spectral filter with $J = 4$ and a scaling factor of 2 for the main analysis, after which SGCC was computed between the two signals for each pattern, optionally decomposed into low-frequency and non-low-frequency components.

For comparison, classical spatial correlation methods were evaluated on the same simulated patterns using standard implementations. Bivariate Moran I was computed using the spdep package (moran_bv), with spatial weights constructed from the same kNN graph (row-standardized weights, style "W") and $n_{sim} = 10$ Monte Carlo permutations. Pearson correlation was computed directly from paired signals using the base stats package (cor). Local spatial cross-correlation index (LSCI) was computed using the spatialEco package (crossCorrelation) with inverse-power distance weighting and coordinate scaling. Cross-variogram similarity was computed using the gstat package (variogram, cross-variogram mode) with lag width = 1 and cutoff = 10, and summarized as the negative mean cross-variogram value to ensure consistency in similarity direction. Parameter robustness was assessed using pattern-level coefficients of variation across method-specific parameter grids: SGWT with $J \in \{2,4,6\}$ and scaling factor $\in \{1,2,2.5\}$; bivariate Moran I with $k \in \{10,20,30\}$ and weight style $\in \{W, B\}$; LSCI with $k \in \{10,20,30\}$; and cross-variogram with cutoff $\in \{2,4,\ldots,20\}$ (with width fixed at 1), whereas Pearson correlation has no tunable spatial parameters.

Selection of the KNN Parameter. The KNN parameter was selected based on eigenspectral stability analyses across a range of k values. The eigenvalue spectrum exhibited a plateau between $k \approx 250$ and 600, indicating limited sensitivity of the graph structure to k within this interval (Supplementary Fig. S4B). Pairwise correlations of eigenvalues across different k values were consistently high for $k \geq 400$ (Supplementary Fig. S4C), and the variability of eigenvalue ranks was reduced in this range (Supplementary Fig. S4D), reflecting increased stability of the graph representation. At smaller k values, the graph was more sensitive to local variability, whereas larger k values better captured global structure. Based on these observations, we selected $k = 400$ as a representative value within the stable regime. For the comparison between SGWT with combined low-pass and band-pass filtering and GFT with a low-pass filter, we selected the ring-shaped configuration of pattern 1 across 80 simulated patterns for evaluation. In the GFT setting, a KNN graph with $k = 400$ was used, and the number of eigenvectors (i.e., FMs) was determined using the *kneedle* algorithm to detect the knee point of the eigenvalue spectrum. For SGWT, we fixed $k = 10$ and evaluated multiple choices of retained FMs (200, 400, 800, and 1,600). A Wilcoxon test was used to assess differences between GFT ($k = 400$) and SGWT (k = 10, 200 retained eigenvectors). Downstream analyses were robust to the specific choice of k within this range.

Space-Gene Covarying Analysis. To investigate spatially covarying gene expression in relation to cell–cell spatial pattern dynamics across multiple samples, SGCC scores are leveraged as spatial factors and treated as time series variables within the ImpulseDE2 framework (v0.99.10; ref. 41). ImpulseDE2 is a statistical tool designed for differential expression analysis, employing a sigmoid-based impulse model to represent continuous trends across time. By utilizing SGCC

scores as a continuous spatial variable, this approach facilitates the identification of genes in which expression systematically correlates with spatially defined paired cell phenotype patterns, enabling the exploration of underlying molecular mechanisms associated with changed spatial organization across multiple samples or ROIs.

The workflow begins by addressing batch effects using previously established batch correction methods (as detailed above and also in bioRxiv 2024.03.05.583586). Following this, the input consists of a gene expression matrix, sample metadata, and SGCC scores, which represent the spatial relationships between paired cell phenotypes. The dataset is preprocessed by subsetting to include relevant cell phenotypes and experimental conditions while correcting for batch factors using default ImpulseDE2 settings. In Fig. 3C, CD4T cells and BCL6-positive B cells were selected. When metadata is available, it is constructed for each sample, incorporating binary conditions (e.g., case vs. control), SGCC scores as continuous spatial factors, and batch information. SGCC scores are then discretized into time bins to represent progression along the spatial factor for time series modeling. Using ImpulseDE2, a sigmoid-based impulse model is applied to capture nonlinear gene expression dynamics across SGCC-defined time bins. Genes are ranked based on their temporal expression trends and categorized into patterns such as increasing, decreasing, or transient, and significant genes are identified using an adjusted *P* value threshold based on the Benjamini–Hochberg (BH) method. The output consists of a ranked list of genes that covary with the spatial factor, classified patterns of gene expression, and insights into spatially regulated molecular mechanisms linked to changes in paired cell phenotypical patterns.

Spatial Transcriptomics: Analysis

RNA Quantity Comparison. The non-batch-corrected CPM counts (GeoMx data), UMI counts (VisiumHD data), and transcript counts (CosMx data) were used as gene expression measurements after log1p transformation. PCCs were calculated for each adjacent IN-DEPTH and control slide pairs, with each datapoint being 1 unique gene. Total RNA quantity, as well as total control RNA quantity, were generated by first summing all the respective gene counts across the ROIs and then visualized on a log1p scale. Genes labeled as "NegProbe" or "Neg" in the GeoMx and CosMx probe kits were used to determine the control probe counts; note that the VisiumHD probe panel did not include any internal negative controls.

Gene Signature Curation and Scoring. All gene signatures used in this study (101), apart from those that were manually curated, were obtained using the R package "msigdbr" (v7.5.1), and the enrichment of gene signatures within cell populations were calculated using GSVA (102) through the R package "gsva" (v1.52.3) with the default parameters.

The gene signatures used to validate the transcriptomic signature of annotated cell populations [Fig. 2C (middle)] were derived from a tonsil single-cell RNA sequencing (scRNA-seq) atlas comprising more than 556,000 cells (29). They were used to (i) calculate cell type associated DEGs for enrichment analysis of IN-DEPTH captured transcriptomics data and (ii) provide scRNA-seq reference for deconvolution analyses. The processing workflow began by loading Seurat objects (v4.4.0; ref. 103). Cells were subsampled and refined to merge to reduce dataset complexity based on the annotation with 135 cell types. Specifically, "SELENOP FUCA1 PTGDS macrophages," "C1Q HLA macrophages," "ITGAX ZEB2 macrophages," and "IL7R MMP12 macrophages" were assigned as immunosuppressive macrophages, "Mono/Macro" and "cycling myeloid" were assigned as myeloid cells. Cell types unrelated to this study, such as "cycling FDC," "cycling T," "granulocytes," "DN," "Granulocytes," "ILC," "Mast," "NK," and "preB/T," were excluded from the analysis. The major B-cell populations, including naive B cells (NBC), memory B cells (MBC), and GC B cells (GCBC), were refined by removing corresponding cell subsets with fewer than 100 cells. Overall, NBC, MBC, GCBC, CD4 T cell, CD8 T cell, Treg, immunosuppressive macrophages, immune-active macrophages, myeloid, DCs, and epithelial cells were refined and extracted for enrichment and deconvolution analyses. Note that endothelial signatures were collected separately (104). Additionally, the Tfh signature used in Fig. 2F was curated using all unique genes from four annotated Tfh populations ("Tfh TB border," "Tfh-LZ-GC," "GC-Tfh-SAP," "GC-Tfh-OX40") in the same atlas resource (29).

DEG analysis was subsequently performed using Seurat (v4.4.0; ref. 103) to identify gene signatures associated with specific cell types. Followed by the log-CPM (LogCPM) normalization method, the "FindMarkers" function was applied with default parameters, including a log fold change threshold ($\log_2$ fold change > 0.25) and an adjusted *P* value threshold ($P_{adj} < 0.05$). For each cell type, DEGs were calculated by comparing the target cell population with all other cell types. Specifically, DEGs of NBC, MBC, GCBC, CD8 T cells, DC, and epithelial cells were identified by comparing each cell type with other cell types. DEGs of CD4 T cells and Tregs were compared directly against each other. DEGs of immunosuppressive macrophage were compared with immune-active macrophage. GSVA (v.1.52.23; ref. 102) was used to determine enrichment of each gene signature (Fig. 2C). All gene signatures used in Fig. 2C and D, for tonsil cell types and Tfh cells, are in Supplementary Table S3.

The RNA gene signature for T-cell dysfunction [Fig. 4H (right)] was curated based on genes previously described as markers of dysfunctional or exhausted CD4 and CD8 T cells (49, 50, 105–107), including *CTLA4*, *HAVCR2*, *LAG3*, *PDCD1*, *BTLA*, *TIGIT*, *CD160*, *CD244*, *ENTPD1*, and *VSIR*. The EBV RNA gene signature shown in Fig. 5A was generated by averaging normalized expression counts of detected EBV transcripts in the GeoMx platform, including *EBER1*, *EBER2*, *EBNA1*, *EBNA2*, *EBNALP*, *LMP1*, *RPMS1*, *BALF1*, *BCRF1*, *BHRF1*, *BNLF2A*, *BNLF2B*, *BNRF1*, and *BZLF1*. Expression levels of individual EBV transcripts are shown in Supplementary Fig. S8. For the CosMx platform (Figs. 6 and 7), the EBV signature was constructed from detected transcripts, including *BCRF1*, *BLLF1*, *BNLF2A*, *BZLF1*, *EBNA2*, *EBNA3A*, *EBNA3BC*, *LMP1*, *LMP2*, and *RPMS1*. Macrophage subtype gene signatures were curated from the MoMacVERSE dataset (46). Complete gene signatures used across Figs. 3–7 are provided in Supplementary Table S3.

Lymphocyte Spatial Distribution. The follicle-high and follicle-low regions were visually identified, with ROIs 3, 5, and 16 from both tissues used for the former and ROIs 1, 7, and 14 from both tissues used for the latter (Supplementary Fig. S3B) to generate six data points for each follicle regions, after which the CD4 T-cell Tfh GSVA scores were compared between these two follicle regions. Tfh correlation was determined by performing a Spearman correlation across all ROIs between each ROI's B-cell proportion and CD4 T-cell Tfh GSVA score.

GEP Identification. GEPs were identified using cNMF (38). The number of highly variable genes to use for cNMF was determined by setting a minimum threshold of 10% of all genes (at least 1,800 genes in this case). The variance for all genes was then determined using the "FindVariableFeatures" function in Seurat (v4.4.0; ref. 103), followed by *k*-means clustering with nine centers with the random seed 1, to identify the cluster with the optimal cutoff for the number of highly variable genes. The number of genes chosen was then rounded up to the nearest 100 and used for cNMF. A range of 25 to 30 components (also known as GEPs) was tested for cNMF, an empirically determined optimum based on prior experience. The number of components with highest stability, in which the stability is larger than the error, was chosen; in this case, it was 26. The R package "enrichR" (v3.2; ref. 108) was then used to infer the biological function of each

GEP by referencing the top 5 enriched GOBP gene signatures (Supplementary Table S4). GEPs with at least 1 statistically significant (P_{adj} < 0.05) GOBP signature were determined to be distinctly enriched and were annotated based on their significant GOBP terms. The annotatable GEPs were then used to determine their relative enrichments across all the tonsil cell subpopulations in Fig. 2G.

CosMx Cell Phenotyping. Seurat (v4.4.0; ref. 103) was used to perform unsupervised clustering and annotation of single cells. Harmony (v1.2.0; ref. 109) was used for batch effect correction across different FOVs. Afterward, the read count for each gene was divided by the total gene counts within each cell, multiplied by a scale factor of 100,000, and natural log-transformed. Principal component (PC) analysis was performed on the normalized expression matrix using 2,000 highly variable genes. The top 15 PCs were selected with a resolution parameter equal to 1. The clustering results were visualized using Uniform Manifold Approximation and Projection for Dimension Reduction (110). We annotated cells into five major types according to their marker genes: *CD3D*, *CD4*, and *CD8A* for T cells; *CD79A*, *MS4A1*, *MZB1*, and *JCHAIN* for B/plasma cells, which were reannotated as tumor cells; *LYZ*, *CD68*, and *C1Q* for myeloid cells; *COL1A1* and *ACTA2* for fibroblasts; and *VWF*, *PECAM1*, and *ENG* for endothelial cells. Note that batch correction was only performed for the analysis in Fig. 5E.

Area under the Curve–Based Gene Set Enrichment at the Single-Cell Level Scoring and Correlation Analysis. Area under the curve–based gene set enrichment at the single-cell level (AUCell), implemented in rapids_singlecell (v0.13.4; Zenodo. Available from: https://doi.org/10.5281/zenodo.683900), was used to compute gene set enrichment scores for individual cells. The T-cell dysfunction signature comprised 15 immune checkpoint and exhaustion-associated genes (*CTLA4*, *HAVCR2*, *LAG3*, *PDCD1*, *TIGIT*, *ENTPD1*, *BTLA*, *CD244*, *VSIR*, *CD160*, *NT5E*, *ADORA2A*, *PVRIG*, *SIGLEC7*, *SIGLEC9*). The C1QC$^+$ macrophage signature included 17 marker genes (*C1QA*, *C1QB*, *C1QC*, *ITM2B*, *HLA-DMB*, *MS4A6A*, *CTSC*, *TBXAS1*, *TMEM176B*, *SYNGR2*, *ARHGDIB*, *TMEM176A*, *UCP2*, *CAPZB*, *MAF*, *TREM2*, *MSR1*). EBV burden was quantified as the summed expression of viral transcripts (*BCRF1*, *BLLF1*, *BNLF2A*, *BZLF1*, *EBNA2*, *EBNA3A*, *EBNA3BC*, *LMP1*, *LMP2*, *RPMS1*) across all cells. Mean AUCell scores for T-cell dysfunction, the C1QC$^+$ macrophage signature, and EBV burden were calculated separately for T cells (CD4 and CD8), macrophages, and tumor cells within each spatial region (core or FOV). Spearman rank correlation coefficients were computed using R (v4.4.1) to assess associations between these scores across regions, and a *P* value < 0.05 was considered statistically significant.

Differential Gene Expression and Pathway Analysis of LMP1-Positive and LMP1-Negative Tumors and Their Neighboring Macrophages. Tumor cells were annotated as LMP1$^+$ or LMP1$^-$ based on LMP1 protein expression detected in CODEX imaging data. Because the CODEX and CosMx datasets were spatially registered, LMP1 protein expression measured by CODEX was used to assign LMP1 status and match corresponding cell identities in the CosMx transcriptomic dataset. The neighborhood composition within a 20 μm radius of each tumor cell was then quantified. Macrophages surrounded exclusively by LMP1$^-$ tumor cells were classified as *macrophages_near_LMP1−_tumor*, whereas macrophages with at least one LMP1$^+$ tumor cell within their vicinity were classified as *macrophages_near_LMP1+_tumor*. These tumor and macrophage populations were subsequently subjected to differential gene expression, pathway enrichment, and cell–cell communication inference analyses.

The Scanpy package (v1.11.5; ref. 111) was used to identify DEGs between LMP1$^+$ and LMP1$^-$ tumor cells, as well as between *macrophages_near_LMP1−_tumor* and *macrophages_near_LMP1+_tumor*. BH correction was applied for multiple testing. Genes with a *P* value < 0.05 and $|\log_2$ fold change$|$ >0.25 were considered differentially expressed. Functional pathway enrichment analysis was performed using the GSEApy package (v1.1.11; ref. 112). FDR correction was applied for multiple testing. A *P* value < 0.05 was considered statistically significant.

Ligand–Receptor Communication Analysis. Cells were divided into two groups based on LMP1 status: the LMP1$^+$ group (comprising LMP1$^+$ tumor cells and *macrophages_near_LMP1+_tumor*) and the LMP1$^-$ group (comprising LMP1$^-$ tumor cells and *macrophages_near_LMP1−_tumor*). For population-level interaction analysis, cell–cell communication was inferred using the Squidpy package (v1.7.0; ref. 68). Analyses were performed separately for each group to identify significant L–R pairs, with $P < 0.05$ considered statistically significant. The $\log_2$ fold change of mean interaction strength for each L–R pair was calculated between groups (LMP1$^+$ vs. LMP1$^-$).

Cell–cell communication was further evaluated at the single-cell pair level. Tumor cells and macrophages located within 20 μm of each other were paired within each group to generate tumor–macrophage pairs. For each pair, interaction strength was quantified as the mean expression of ligand and receptor genes:

$$\frac{1}{2}(\text{ligand expression} + \text{receptor expression})$$

The mean interaction strength of each L–R pair was then calculated across all tumor–macrophage pairs within each FOV. Differences in mean interaction strength between the LMP1$^+$ and LMP1$^-$ groups were assessed using the Wilcoxon rank-sum test, with $P < 0.05$ considered statistically significant.

Benchmarking of Deconvolution Softwares

CIBERSORT. CIBERSORT (34) is a computational method designed for cell type deconvolution from bulk tissue gene expression data using a reference-based approach. It employs a support vector regression framework (nu-SVR) to estimate cell proportions within a mixed tissue sample. The input includes a gene expression reference matrix, derived from the create_profile_matrix function of SpatialDecon, and a bulk tissue expression matrix in raw count format, created by combining and merging data across ROIs. The method is executed using the cibersort function, with parameters specifying the reference matrix and bulk expression data, enabling a robust deconvolution process that accurately quantifies cell type proportions.

dtangle. dtangle (v2.0.9; ref. 35) is another method based on single-cell reference data that use a linear scoring approach to estimate cell type proportions in bulk tissue samples. The input consists of a bulk tissue expression matrix and a single-cell dataset, both preprocessed to retain the most informative genes and cell types. The function dtangle facilitates the deconvolution by specifying parameters such as the combined dataset, the number of markers to use, and the data type. This ensures precise estimation of cell type proportions while maintaining compatibility with bulk and single-cell data formats.

MuSiC. MuSiC (36) leverages single-cell reference data for cell type deconvolution in bulk gene expression profiles. It employs weighted nonnegative least squares to estimate the contributions of distinct cell types within bulk samples. The input includes the same bulk expression matrix used in CIBERSORT and a single-cell expression dataset formatted as a SingleCellExperiment object. This dataset is preprocessed to include cell types of interest and DEGs to enhance deconvolution accuracy. The deconvolution process is implemented through the music_prop function, in which users specify key parameters, including cell type annotations and sample identifiers, ensuring the alignment of single-cell and bulk datasets.

SpatialDecon. SpatialDeconn (v1.13.2; ref. 37) utilizes a log-normal regression model to perform gene expression deconvolution. Unlike other tools, it can integrate normalized bulk expression data and single-cell reference matrices. The method aligns genes across datasets to ensure consistency during deconvolution. The spatialdecon function allows users to specify the normalized bulk expression data, background adjustment parameters, and the reference matrix. This method is particularly effective in leveraging both single-cell and bulk datasets to provide accurate cell type proportion estimates, whereas the alignment step enhances consistency across data sources.

Reference scRNA-Seq Dataset Preparation for Deconvolution Methods. CIBERSORT (34), dtangle (35), MuSiC (36), and SpatialDecon (37) require a reference transcriptome for deconvolution. Therefore, a published tonsil single-cell transcriptomic atlas comprising more than 556,000 cells was used for preparing the reference (29). To improve computational efficiency while preserving biological diversity, we downsampled this dataset prior to analysis. Specifically, we selected the nine immune cell types present in the GeoMx dataset ("Immune-active Macrophages," "DC," "myeloid," "CD8 T," "CD4 T," "MBC," "GCBC," "Tregs," and "Immune-suppressive Macrophages") and randomly sampled 5% of the 460,000 cells assigned to these categories. This produced a reduced single-cell reference matrix with preserved representation of the cell types of interest. Except for adapting input formats to meet the requirements of each tool (bulk GeoMx ROI expression matrix, single-cell reference expression matrix, and single-cell cell type annotations), we did not perform additional preprocessing or modifications.

Pseudobulk RNA-Seq Creation. All deconvolution methods (CIBERSORT, dtangle, MuSiC, and SpatialDecon) were applied at the ROI level. For each IN-DEPTH ROI, expression profiles from all annotated segments (i.e., cell type masks) were first aggregated into a single ROI-level profile, and deconvolution was then performed on each cell type aggregated ROI.

Benchmarking Aggregated Cell Type Proportions. For each deconvolution method, cell type proportion estimates across 16 ROIs were first reshaped into a long format containing three variables: method, cell type, and predicted cell proportion. To obtain a representative value per method, proportions were averaged across all ROIs for each tool and cell type combination. Within each method, these mean values were then normalized by dividing by the sum of all cell types, resulting in relative contributions that sum to 1. Stacked bar plots were generated using ggplot2 with geom_bar (stat = "identity," position = "stack"), thereby displaying the mean relative cellular composition predicted by each method. A fixed color palette was applied across figures to ensure consistent mapping of cell types.

IN-DEPTH Dataset Reference Visualization. Cell type compositions from the IN-DEPTH dataset were aggregated across ROIs to compute the mean abundance of each cell type. These mean values were visualized as a reference distribution using stacked bar plots, with proportions normalized to sum to 1. This visualization provided a direct comparison baseline against which method-level estimates could be interpreted.

Diversity Quantification. To assess intra-ROI heterogeneity, we calculated the Gini–Simpson index for each ROI. Let $Value_i$ denote the raw abundance (score) of cell type i within a given ROI, and let $Value_j$ denote the raw abundances of all cell types j present in the same ROI. The relative proportion of cell type i was defined as follows:

$$p_i = \frac{Value_i}{\sum_j Value_j}$$

where p_i denotes the relative abundance of cell type i within the ROI. The Gini–Simpson index was then computed as follows:

$$GS = 1 - \sum_i p_i^2$$

These indices were computed independently for each ROI. ROIs were ranked according to their Gini–Simpson index, from lowest to highest diversity, and this ordering was subsequently used to arrange rows in the benchmarking heatmap.

Heatmap Generation. Method performance was evaluated by computing PCCs between predicted and ground truth compositions across ROIs. The PCC matrix was visualized using the ComplexHeatmap package. Rows (ROIs) were ordered according to the diversity ranking described above, and columns (methods) were kept in a fixed order without clustering. PCC values were mapped onto a white–blue gradient, whereas missing values were displayed in gray. To highlight diversity alongside performance, a continuous annotation bar was added to the left of the heatmap, encoding Gini–Simpson scores as a gradient from gray to black.

Application of SGCC on DLBCL Dataset

To analyze DLBCL GeoMx data, we first calculated SGCC scores to capture spatial relationships between the cell phenotypes. Samples were merged and discretized into a uniform 60 by 60 bin grid. Pairwise SGCC scores were computed for all cell types, reflecting their large-scale spatial distributions.

DEG Analysis. For DEG analysis between EBV-positive and EBV-negative conditions, we applied edgeR (42) and limma (95) frameworks with batch corrected data (batch correction performed as described in the Batch Correction section). Batch-corrected data were fitted to a linear model using the "mFit" function, incorporating a predefined design matrix. Empirical Bayes moderation was applied using the "eBayes" function to stabilize variance estimates, followed by DEG identification with the "topTable" function, ranked by adjusted P values. Specific normalization strategies and batch correction parameters were applied based on cell types:

- CD4 T cells: LogCPM normalization, top 5,000 NCGs, $k = 2$, using two weight matrices from RUV4 batch correction, with cell type number included as a covariate in the design model
- Macrophages: LogCPM normalization, top 1,000 NCGs, $k = 3$, using three weight matrices from RUV4 batch correction as covariates
- Tumor cells: LogCPM normalization, top 1,000 NCGs, $k = 3$, using one weight matrix from RUV4 batch correction as a covariate

DEGs between EBV-positive and EBV-negative conditions for CD4 T cells, macrophages, and tumor cells were filtered based on adjusted P value thresholds ($P_{adj} < 0.01$, BH method). Enrichment analysis was performed for each DEG set using the enrichR (v3.2; ref. 108) database, focusing on "Reactome_2022," "GO_Biological_Process_2023," and "KEGG_2021_Human." Genes enriched in biologically meaningful pathways (Figs. 5 and 6; Supplementary Figs. S8 and S9; Supplementary Table S7) were selected for GSVA to refine functional insights. Heatmap visualization was subsequently generated to highlight pathway activity across conditions based on ComplexHeatmap (v2.16.0).

Ternary Analysis. ggtern (v3.5.0) was used for visualizing CD4 T cell, tumor, and macrophage ternary plots using SGCC scores from CD4 T cell–tumor, macrophage–tumor, and CD4 T cell–macrophage (Fig. 5; Supplementary Table S8). The AES for each cell pair was determined as described in ref. (113), in which the expected number of edges between cell types was computed based on the frequencies

of the cell types and the total number of edges in the graph. Specifically, AES was then calculated by comparing the observed number of edges connecting the two cell types to the expected number of edges. An AES of 0 indicates no enrichment over expectation, whereas positive and negative values indicate enrichment and depletion, respectively. Additionally, the density transparency was mapped to contour levels and color-coded by EBV status (i.e., "EBV$^+$" and "EBV$^-$").

Ternary plots were solely used for visualization, and all statistical testing were performed independently of the ternary representation. For each score, differences between EBV-positive and EBV-negative samples were evaluated using linear regression models of the form value EBV status + cohort, in which EBV status was treated as the primary variable of interest and cohort was included as a covariate to account for batch effects across datasets. Model coefficients and corresponding *P* values for the EBV effect were obtained from the fitted linear models, and statistical significance was determined based on the corresponding Wald test *P* values.

Single Cell–Level SGCC Analysis. Cell centroids and curated cell type annotations were used to partition each FOV into overlapping spatial neighborhoods using a sliding-window strategy by setting window size as 250 μm × 250 μm, defining local microenvironments that capture cell organization. To ensure robust estimation of spatial relationships, only windows containing sufficient numbers of both cell types were retained. Within each valid window, single cells of a given type were converted into a spatial signal by binning cell locations onto a regular two-dimensional lattice, yielding one spatial signal per cell type per window that encodes local cell density and spatial arrangement (50 × 50 bins). Each lattice was treated as a graph, and spatial signals were analyzed in the graph spectral domain to capture multiscale spatial organization while respecting neighborhood connectivity. SGCC produces a single scalar score per window that reflects whether two cell types exhibit spatial organization within that neighborhood. Window-level SGCC scores were subsequently propagated back to the single-cell metadata by assigning each cell the SGCC score of its corresponding sliding window. To define discrete SGCC categories, groups were defined using percentile-based thresholds. Windows falling within the lowest percentiles of the SGCC score distribution were classified as SGCC-low, whereas windows within the highest percentiles were classified as SGCC-high, with all intermediate windows excluded from this categorization. In practice, this corresponded to selecting the bottom 25% and top 25% of windows to represent low and high interaction states, respectively. To link spatial interaction states to transcriptional programs, gene expression was aggregated within each window for a selected cell type to generate window-level pseudobulk gene expression profiles and applied RUV4 batch correction in the same manner as the previous section. Differential expression analysis was performed by comparing SGCC-high and SGCC-low windows, with analyses conducted separately for EBV-positive and EBV-negative samples to control for viral status. Batch effects were accounted for during normalization and modeling. Genes differentially expressed between SGCC states were further analyzed using pathway enrichment and GSVA to identify and visualize biological processes associated with distinct spatial interaction patterns.

Data Availability

All data generated in this study are publicly available at: https://doi.org/10.5281/zenodo.14530077 and https://doi.org/10.5281/zenodo.18379155. All analysis codes used in this study, along with tutorials for key workflows (including SGCC), are available on our study Github: https://github.com/SizunJiangLab/IN-DEPTH. Detailed experimental protocols for each IN-DEPTH platform combination presented in this study, as well as accompanying video tutorials, are available at our study webpage: https://sizunjianglab.github.io/IN-DEPTH/. The SGCC R package is publicly available through CRAN by using install.packages ("BioGSP").

Authors' Disclosures

S. Li reports grants from the NIH/NCI during the conduct of the study. C. Keane reports grants from Medical Research Future Fund (MRFF) and National Health and Medical Research Council (NHMRC) during the conduct of the study, as well as personal fees from Merck, Roche, and Gilead and nonfinancial support from BeOne and Miltenyi outside the submitted work. B.E. Howitt reports grants and personal fees from Cartography, grants from Santa Ana and Link Cell Sciences, and personal fees from AbbVie, Tempus, Leica, ACDBio, Roche, and J&J outside the submitted work. S. Signoretti reports grants and personal fees from AstraZeneca, Bristol Myers Squibb, Exelixis, NiKang Therapeutics, and Arsenal Biosciences; grants and personal fees from Merck; and personal fees from NextPoint Therapeutics, American Association for Cancer Research, and the NCI outside the submitted work, as well as a patent for BioGenex with royalties paid, and that S. Signoretti has mentored several non-US citizens on research projects with potential funding (in part) from non-US sources/foreign components. C.M. Schürch reports personal fees and other support from Vicinity Bio GmbH and nonfinancial support and other support from Enable Medicine Inc. outside the submitted work. F.S. Hodi reports grants and personal fees from Bristol Myers Squibb and personal fees from Merck, Novartis, Compass Therapeutics, Apricity, Bicara, Checkpoint Therapeutics, Bioentre, Gossamer, Iovance, Catalym, Immunocore, Kairos, Rheos, Bayer, Zumutor, Corner Therapeutics, Puretech, Curis, AstraZeneca, Pliant, Solu Therapeutics, and Vir Biotechnology outside the submitted work, as well as a patent for Methods for Treating MICA-Related Disorders (#20100111973) pending; a patent for Tumor Antigens and Uses Thereof (#7250291) issued; a patent for Angiopoiten-2 Biomarkers Predictive of Anti-immune Checkpoint Response (#20170248603) pending; a patent for Compositions and Methods for Identification, Assessment, Prevention, and Treatment of Melanoma using PD-L1 Isoforms (#20160340407) pending; a patent for Therapeutic Peptides Patent Number: 9402905 issued; a patent for Methods of Using Pembrolizumab and Trebananib pending; a patent for Vaccine Compositions and Methods for Restoring NKG2D Pathway Function against Cancers Patent Number: 10279021 issued and with royalties paid; a patent for Antibodies That Bind to MHC Class I Polypeptide-Related Sequence A Patent Number: 10106611 issued; a patent for Anti-Galectin Antibody Biomarkers Predictive of Anti-Immune Checkpoint and Anti-Angiogenesis Responses Publication Number: 20170343552 pending; and a patent for Antibodies against EDIL3 and Methods of Use Thereof pending. S.J. Rodig reports research support from KITE/Gilead, Bristol Myers Squibb, and Coherus Therapeutics and that S.J. Rodig is a member of Immunitas Therapaeutics. S. Jiang reports grants from the NIH, Gilead Sciences, Bill & Melinda Gates Foundation, Broad Institute, Dye Family Foundation, and Koch Institute for Integrative Cancer Research at MIT and the Dana-Farber/Harvard Cancer Center Bridge Project during the conduct of the study, as well as other support from Elucidate Bio Inc. and Saterra AI Inc., grants from Roche and Novartis, personal fees and nonfinancial support from Cell Signaling Technology, and personal fees from Quanterix and AstraZeneca outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

S.P.T. Yiu: Conceptualization, resources, data curation, formal analysis, validation, investigation, visualization, methodology, writing–original draft, writing–review and editing. **Y. Chang:** Conceptualization, resources, data curation, software, formal analysis,

validation, visualization, methodology, writing–original draft, writing–review and editing. **Y.Y. Yeo:** Conceptualization, formal analysis, investigation, visualization, writing–original draft, writing–review and editing. **H. Qiu:** Resources, data curation, software, formal analysis, validation, visualization, methodology, writing–original draft, writing–review and editing. **W. Wu:** Resources, data curation, software, formal analysis, validation, visualization, methodology, writing–original draft, writing–review and editing. **H.A. Michel:** Resources, validation, investigation, visualization. **X. Jin:** Resources, formal analysis. **R. Huang:** Resources, data curation, software, formal analysis, validation, visualization. **S. Kure:** Resources, data curation. **L. Parmelee:** Resources. **S. Luo:** Resources, investigation. **P. Cramer:** Resources, investigation. **J.L. Lee:** Resources, software, formal analysis, visualization. **Y. Wang:** Resources, software, formal analysis, visualization. **Z. Zhao:** Resources, data curation. **J. Yeung:** Resources, formal analysis. **N. El Ahmar:** Resources, investigation. **B. Simsek:** Resources, investigation. **R. Mohanna:** Resources, investigation. **M. Van Orden:** Resources, investigation. **W.S. Lu:** Resources, investigation. **K.J. Livak:** Resources, investigation. **S. Li:** Resources, investigation. **C. Gao:** Resources, investigation. **M. Burgess:** Resources, data curation, investigation. **C. Keane:** Resources, data curation. **J. Shahryari:** Resources, investigation. **L.G. Kingsley:** Resources, investigation. **R.N. Al-Humadi:** Resources, investigation. **S. Nasr:** Resources, investigation. **D. Nkosi:** Resources, data curation. **S. Sadigh:** Resources, data curation. **P. Rock:** Resources, data curation, investigation. **L. Frauenfeld:** Resources, data curation, investigation. **L. Kaufmann:** Resources, data curation, investigation. **B. Zhu:** Resources, investigation. **A. Basak:** Resources, investigation. **N. Dhanikonda:** Resources, investigation. **C.N. Chan:** Resources. **J. Krull:** Resources. **Y.W. Cho:** Resources, data curation, investigation. **C.-Y. Chen:** Resources, data curation. **J. Brown:** Resources. **H. Wang:** Resources. **B. Zhao:** Resources. **J.-Y.J. Lee:** Resources, formal analysis. **L.-H. Loo:** Resources, software, formal analysis. **D.M. Kim:** Resources, data curation. **V.A. Boussiotis:** Resources. **B. Zhang:** Resources. **K. Wei:** Resources. **A.K. Shalek:** Resources. **B.E. Howitt:** Resources, data curation. **S. Signoretti:** Resources, data curation. **C.M. Schürch:** Resources, data curation. **F.S. Hodi:** Resources. **W.R. Burack:** Resources, data curation. **S.J. Rodig:** Resources, data curation. **Q. Ma:** Conceptualization, resources, supervision, funding acquisition, methodology, project administration, writing–review and editing. **S. Jiang:** Conceptualization, resources, supervision, funding acquisition, methodology, writing–original draft, project administration, writing–review and editing.

Acknowledgments

We thank Craig Lassy and Michael Hair from Akoya for PhenoCycler-Fusion technical support and Marvin Nayan, Adam Limb, Mike Chen, Brendan Collins, Nicholas Merino, Clement David, Sarah Miseirvitch, Prajan Divakar, Ozge Getkin, Tim Riordan, and Sarah Weigel from NanoString for technical support. We also thank Jixin Liu, Jim DeCaprio, and other members of the Jiang and Ma labs for insightful discussions. ChatGPT (OpenAI) was used as a tool to assist with code development and grammatical editing. The author reviewed, verified, and edited all generated content to ensure accuracy and correctness. S. Jiang is supported in part by NIH DP2AI171139, P01AI177687, R01AI149672, R01GM152585, U24CA224331, and P50CA272390; a Gilead's Research Scholars Program in Hematologic Malignancies; a Sanofi iAward; the Dye Family Foundation, the Broad Next Generation Award; and the Bridge Project, a partnership between the Koch Institute for Integrative Cancer Research at MIT and the Dana-Farber/Harvard Cancer Center. Q. Ma is supported in part by NIH R01GM152585, P01CA278732, P01AI177687, U54AG075931, and R01DK138504, and the Pelotonia Institute of Immuno-Oncology. S.J. Rodig is supported by a Blood Cancer Discoveries Grant Program from the Leukemia Lymphoma Society, and the Paul G. Allen Frontiers Group. Y.Y. Yeo is a recipient of the Albert J Ryan Fellowship. S.P.T. Yiu is a MacMillan Family Foundation Awardee of the Life Sciences Research Foundation. R. Huang is supported by a Cancer Research Institute Immuno-Informatics Postdoctoral Fellowship (CRI Award #14614). This article reflects the views of the authors and should not be construed as representing the views or policies of the institutions that provided funding.

Note

Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

Received May 5, 2025; revised February 9, 2026; accepted March 20, 2026; posted first March 24, 2026.

REFERENCES

1. Marx V. Method of the year: spatially resolved transcriptomics. Nat Methods 2021;18:9–14.
2. Method of the year 2024: spatial proteomics. Nat Methods 2024;21:2195–6.
3. Vandereyken K, Sifrim A, Thienpont B, Voet T. Methods and applications for single-cell and spatial multi-omics. Nat Rev Genet 2023;24:494–515.
4. Bodenmiller B. Highly multiplexed imaging in the omics era: understanding tissue structures in health and disease. Nat Methods 2024;21:2209–11.
5. de Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C. Global signatures of protein and mRNA expression levels. Mol Biosyst 2009;5:1512–26.
6. Frei AP, Bava FA, Zunder ER, Hsieh EWY, Chen SY, Nolan GP, et al. Highly multiplexed simultaneous detection of RNAs and proteins in single cells. Nat Methods 2016;13:269–75.
7. Woo J, Williams SM, Markillie LM, Feng S, Tsai CF, Aguilera-Vazquez V, et al. High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. Nat Commun 2021;12:6246.
8. Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G, et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. Nat Biotechnol 2020;38:586–99.
9. Liu Y, Yang M, Deng Y, Su G, Enninful A, Guo CC, et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. Cell 2020;183:1665–81.e18.
10. Jiang S, Chan CN, Rovira-Clavé X, Chen H, Bai Y, Zhu B, et al. Combined protein and nucleic acid imaging reveals virus-dependent B cell and macrophage immunosuppression of tissue microenvironments. Immunity 2022;55:1118–34.e8.
11. Deisher A, Yeo YY, Jiang S. Combined protein and nucleic acid staining in tissues with PANINI. STAR Protoc 2022;3:101663.
12. Ben-Chetrit N, Niu X, Swett AD, Sotelo J, Jiao MS, Stewart CM, et al. Integration of whole transcriptome spatial profiling with protein markers. Nat Biotechnol 2023;41:788–93.
13. Liu Y, DiStasio M, Su G, Asashima H, Enninful A, Qin X, et al. High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. Nat Biotechnol 2023;41:1405–9.
14. Schulz D, Zanotelli VRT, Fischer JR, Schapiro D, Engler S, Lun XK, et al. Simultaneous multiplexed imaging of mRNA and proteins with subcellular resolution in breast cancer tissue samples by mass cytometry. Cell Syst 2018;6:25–36.e5.
15. Blow N. Tissue preparation: tissue issues. Nature 2007;448:959–63.
16. Bussi Y, Keren L. Multiplexed image analysis: what have we achieved and where are we headed? Nat Methods 2024;21:2212–15.
17. Chang Y, Liu J, Jiang Y, Ma A, Yeo YY, Guo Q, et al. Graph Fourier transform for spatial omics representation and analyses of complex organs. Nat Commun 2024;15:7467.

18. Shuman DI, Narang SK, Frossard P, Ortega A, Vandergheynst P. The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Process Mag 2013;30:83–98.
19. Yeo YY, Cramer P, Deisher A, Bai Y, Zhu B, Yeo WJ, et al. A Hitchhiker's guide to high-dimensional tissue imaging with multiplexed ion beam imaging. Methods Cell Biol 2024;186:213–31.
20. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, et al. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. Cell 2018;174:968–81.e15.
21. Papalegis D, Tkachev S, Vu L, Klein S. 114 SignalStar is a novel multiplex IHC technology that demonstrates flexibility and reproducibility. J Immunother Cancer 2023;11(Suppl 1):129.
22. Stack EC, Wang C, Roman KA, Hoyt CC. Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of tyramide signal amplification, multispectral imaging and multiplex analysis. Methods 2014;70:46–58.
23. Lin JR, Chen YA, Campton D, Cooper J, Coy S, Yapp C, et al. High-plex immunofluorescence imaging and traditional histology of the same tissue section for discovering image-based biomarkers. Nat Cancer 2023;4:1036–52.
24. Oliveira MF, Romero JP, Chung M, Williams S, Gottscho AD, Gupta A, et al. High-definition spatial transcriptomic profiling of immune cell populations in colorectal cancer. Nat Genet 2025;57:1512–23.
25. He S, Bhatt R, Brown C, Brown EA, Buhr DL, Chantranuvatana K, et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. Nat Biotechnol 2022;40:1794–806.
26. Janesick A, Shelansky R, Gottscho AD, Wagner F, Williams SR, Rouault M, et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. Nat Commun 2023;14:8353.
27. Black S, Phillips D, Hickey JW, Kennedy-Darling J, Venkataraaman VG, Samusik N, et al. CODEX multiplexed tissue imaging with DNA-conjugated antibodies. Nat Protoc 2021;16:3802–35.
28. Schürch CM, Bhate SS, Barlow GL, Phillips DJ, Noti L, Zlobec I, et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. Cell 2020;182:1341–59.e19.
29. Massoni-Badosa R, Aguilar-Fernández S, Nieto JC, Soler-Vila P, Elosua-Bayes M, Marchese D, et al. An atlas of cells in the human tonsil. Immunity 2024;57:379–99.e18.
30. Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P, De Preter K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat Commun 2020;11:5650.
31. Jin H, Liu Z. A benchmark for RNA-seq deconvolution analysis under dynamic testing environments. Genome Biol 2021;22:102.
32. Sutton GJ, Poppe D, Simmons RK, Walsh K, Nawaz U, Lister R, et al. Comprehensive evaluation of deconvolution methods for human brain gene expression. Nat Commun 2022;13:1358.
33. Li H, Zhou J, Li Z, Chen S, Liao X, Zhang B, et al. A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. Nat Commun 2023;14:1548.
34. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453–7.
35. Hunt GJ, Freytag S, Bahlo M, Gagnon-Bartsch JA. dtangle: accurate and robust cell type deconvolution. Bioinformatics 2019;35:2093–9.
36. Wang X, Park J, Susztak K, Zhang NR, Li M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun 2019;10:380.
37. Danaher P, Kim Y, Nelson B, Griswold M, Yang Z, Piazza E, et al. Advances in mixed cell deconvolution enable quantification of cell types in spatial transcriptomic data. Nat Commun 2022;13:385.
38. Kotliar D, Veres A, Nagy MA, Tabrizi S, Hodis E, Melton DA, et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-seq. Elife 2019;8:e43803.
39. Dong X, Thanou D, Frossard P, Vandergheynst P. Learning Laplacian matrix in smooth graph signal representations. IEEE Trans Signal Process 2016;64:6160–73.
40. Dong X, Thanou D, Rabbat M, Frossard P. Learning graphs from data: a signal representation perspective. IEEE Signal Process Mag 2019;36:44–63.
41. Fischer DS, Theis FJ, Yosef N. Impulse model-based differential expression analysis of time course sequencing data. Nucleic Acids Res 2018;46:e119.
42. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–40.
43. Gräler B, Pebesma E, Heuvelink G. Spatio-temporal interpolation using gstat. R J 2016;8:204–18.
44. Lee SI. Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I. J Geogr Syst 2001;3:369–85.
45. Chen Y. A new methodology of spatial cross-correlation analysis. PLoS One 2015;10:e0126158.
46. Mulder K, Patel AA, Kong WT, Piot C, Halitzki E, Dunsmore G, et al. Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. Immunity 2021;54:1883–900.e5.
47. Liu M, Bertolazzi G, Sridhar S, Lee RX, Jaynes P, Mulder K, et al. Spatially-resolved transcriptomics reveal macrophage heterogeneity and prognostic significance in diffuse large B-cell lymphoma. Nat Commun 2024;15:2113.
48. Wherry EJ. T cell exhaustion. Nat Immunol 2011;12:492–9.
49. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol 2015;15:486–99.
50. Baessler A, Vignali DAA. T cell exhaustion. Annu Rev Immunol 2024;42:179–206.
51. Jiang XN, Yu BH, Yan WH, Lee J, Zhou XY, Li XQ. Epstein–Barr virus-positive diffuse large B-cell lymphoma features disrupted antigen capture/presentation and hijacked T-cell suppression. Oncoimmunology 2019;9:1683346.
52. Young LS, Murray PG. Epstein–Barr virus and oncogenesis: from latent genes to tumours. Oncogene 2003;22:5108–21.
53. Pattle SB, Farrell PJ. The role of Epstein–Barr virus in cancer. Expert Opin Biol Ther 2006;6:1193–205.
54. Park S, Lee J, Ko YH, Han A, Jun HJ, Lee SC, et al. The impact of Epstein-Barr virus status on clinical outcome in diffuse large B-cell lymphoma. Blood 2007;110:972–8.
55. Oyama T, Yamamoto K, Asano N, Oshiro A, Suzuki R, Kagami Y, et al. Age-related EBV-associated B-cell lymphoproliferative disorders constitute a distinct clinicopathologic group: a study of 96 patients. Clin Cancer Res 2007;13:5124–32.
56. Montes-Moreno S, Odqvist L, Diaz-Perez JA, Batlle Lopez A, Gonzalez De Villambrosía S, Mazorra F, et al. EBV-positive diffuse large B-cell lymphoma of the elderly is an aggressive post-germinal center B-cell neoplasm characterized by prominent nuclear factor-κB activation. Mod Pathol 2012;25:968–82.
57. Okamoto A, Yanada M, Inaguma Y, Tokuda M, Morishima S, Kanie T, et al. The prognostic significance of EBV DNA load and EBER status in diagnostic specimens from diffuse large B-cell lymphoma patients. Hematol Oncol 2017;35:87–93.
58. Lu TX, Liang JH, Miao Y, Fan L, Wang L, Qu XY, et al. Epstein-Barr virus positive diffuse large B-cell lymphoma predict poor outcome, regardless of the age. Sci Rep 2015;5:12168.
59. Stuhlmann-Laeisz C, Borchert A, Quintanilla-Martinez L, Hoeller S, Tzankov A, Oschlies I, et al. In Europe expression of EBNA2 is associated with poor survival in EBV-positive diffuse large B-cell lymphoma of the elderly. Leuk Lymphoma 2016;57:39–44.
60. Bourbon E, Maucort-Boulch D, Fontaine J, Mauduit C, Sesques P, Safar V, et al. Clinico-pathological features and survival in EBV-positive diffuse large B-cell lymphoma not otherwise specified. Blood Adv 2021;5:3227–39.
61. Malpica L, Marques-Piubelli ML, Beltran BE, Chavez JC, Miranda RN, Castillo JJ. EBV-positive diffuse large B-cell lymphoma, not otherwise specified: 2024 update on the diagnosis, risk-stratification, and management. Am J Hematol 2024;99:2002–15.

62. Gao G, Sun N, Zhang Y, Li J, Jiang Y, Chen N, et al. Single-cell sequencing in diffuse large B-cell lymphoma: C1qc is a potential tumor-promoting factor. Int Immunopharmacol 2024;143:113319.
63. Revel M, Sautès-Fridman C, Fridman WH, Roumenina LT. C1q+ macrophages: passengers or drivers of cancer progression. Trends Cancer 2022;8:517–26.
64. Uchida J, Yasui T, Takaoka-Shichijo Y, Muraoka M, Kulwichit W, Raab-Traub N, et al. Mimicry of CD40 signals by Epstein-Barr virus LMP1 in B lymphocyte responses. Science 1999;286:300–3.
65. Giehler F, Ostertag MS, Sommermann T, Weidl D, Sterz KR, Kutz H, et al. Epstein-Barr virus-driven B cell lymphoma mediated by a direct LMP1-TRAF6 complex. Nat Commun 2024;15:414.
66. Lajoie V, Lemieux B, Sawan B, Lichtensztejn D, Lichtensztejn Z, Wellinger R, et al. LMP1 mediates multinuclearity through downregulation of shelterin proteins and formation of telomeric aggregates. Blood 2015;125:2101–10.
67. Ding L, Li L, Yang J, Zhou S, Li W, Tang M, et al. Latent membrane protein 1 encoded by Epstein–Barr virus induces telomerase activity via p16INK4a/Rb/E2F1 and JNK signaling pathways. J Med Virol 2007;79:1153–63.
68. Palla G, Spitzer H, Klein M, Fischer D, Schaar AC, Kuemmerle LB, et al. Squidpy: a scalable framework for spatial omics analysis. Nat Methods 2022;19:171–8.
69. Caveney NA, Glassman CR, Jude KM, Tsutsumi N, Garcia KC. Structure of the IL-27 quaternary receptor signaling complex. eLife 2022;11:e78463.
70. Horlad H, Ma C, Yano H, Pan C, Ohnishi K, Fujiwara Y, et al. An IL-27/STAT3 axis induces expression of programmed cell death 1 ligands (PD-L1/2) on infiltrating macrophages in lymphoma. Cancer Sci 2016;107:1696–704.
71. Xia T, Zhang M, Lei W, Yang R, Fu S, Fan Z, et al. Advances in the role of STAT3 in macrophage polarization. Front Immunol 2023;14:1160719.
72. Nordmann TM, Mund A, Mann M. A new understanding of tissue biology from MS-based proteomics at single-cell resolution. Nat Methods 2024;21:2220–2.
73. Long Y, Ang KS, Sethi R, Liao S, Heng Y, van Olst L, et al. Deciphering spatial domains from spatial multi-omics with SpatialGlue. Nat Methods 2024;21:1658–67.
74. Chen S, Zhu B, Huang S, Hickey JW, Lin KZ, Snyder M, et al. Integration of spatial and single-cell data across modalities with weakly linked features. Nat Biotechnol 2024;42:1096–106.
75. Zhu B, Chen S, Bai Y, Chen H, Liao G, Mukherjee N, et al. Robust single-cell matching and multimodal analysis using shared and distinct features. Nat Methods 2023;20:304–15.
76. Dai Y, Kizhakeyil A, Chihara D, Li X, Liu Y, Sainz Zuniga TP, et al. Multi-modal spatial characterization of tumor immune microenvironments identifies targetable inflammatory niches in diffuse large B-cell lymphoma. Nat Genet 2025;57:2715–27.
77. Rodig SJ, Gusenleitner D, Jackson DG, Gjini E, Giobbie-Hurder A, Jin C, et al. MHC proteins confer differential sensitivity to CTLA-4 and PD-1 blockade in untreated metastatic melanoma. Sci Transl Med 2018;10:eaar3342.
78. Chen BJ, Dashnamoorthy R, Galera P, Makarenko V, Chang H, Ghosh S, et al. The immune checkpoint molecules PD-1, PD-L1, TIM-3 and LAG-3 in diffuse large B-cell lymphoma. Oncotarget 2019;10:2030–40.
79. Bednarska K, Nath K, Nicol W, Gandhi MK. Immunity reloaded: deconstruction of the PD-1 axis in B cell lymphomas. Blood Rev 2021;50:100832.
80. Enninful A, Zhang Z, Klymyshyn D, Ingalls M, Yang M, Zong H, et al. Integration of imaging-based and sequencing-based spatial omics mapping on the same tissue section via DBiTplus. Nat Methods 2026 Jan 15 [Epub ahead of print].
81. Phillips D, Schürch CM, Khodadoust MS, Kim YH, Nolan GP, Jiang S. Highly multiplexed phenotyping of immunoregulatory proteins in the tumor microenvironment by CODEX tissue imaging. Front Immunol 2021;12:687673.
82. Hickey JW, Neumann EK, Radtke AJ, Camarillo JM, Beuschel RT, Albanese A, et al. Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging. Nat Methods 2022;19:284–95.
83. Hoell T, Huschak G, Beier A, Hüttmann G, Minkus Y, Holzhausen HJ, et al. Autofluorescence of intervertebral disc tissue: a new diagnostic tool. Eur Spine J 2006;15 Suppl 3(Suppl 3):345–53.
84. Gill EM, Malpica A, Alford RE, Nath AR, Follen M, Richards-Kortum RR, et al. Relationship between collagen autofluorescence of the human cervix and menopausal status. Photochem Photobiol 2003;77:653–8.
85. Lowe DG. Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. Kerkyra (Greece): IEEE; 1999. Vol. 2. p. 1150–7.
86. Gatenbee CD, Baker AM, Prabhakaran S, Swinyard O, Slebos RJC, Mandal G, et al. Virtual alignment of pathology image series for multi-gigapixel whole slide images. Nat Commun 2023;14:4502.
87. Van Valen DA, Kudo T, Lane KM, Macklin DN, Quach NT, DeFelice MM, et al. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. PLoS Comput Biol 2016;12:e1005177.
88. Greenwald NF, Miller G, Moen E, Kong A, Kagel A, Dougherty T, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nat Biotechnol 2022;40:555–65.
89. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir EA, Tadmor MD, et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. Cell 2015;162:184–97.
90. Shaban M, Bai Y, Qiu H, Mao S, Yeung J, Yeo YY, et al. MAPS: pathologist-level cell type annotation from tissue images through machine learning. Nat Commun 2024;15:28.
91. Yen JC, Chang FJ, Chang S. A new criterion for automatic multilevel thresholding. IEEE Trans Image Process 1995;4:370–8.
92. Simon S, Labarriere N. PD-1 expression on tumor-specific T cells: friend or foe for immunotherapy? Oncoimmunology 2018;7:e1364828.
93. Liu N, Bhuva DD, Mohamed A, Bokelund M, Kulasinghe A, Tan CW, et al. standR: spatial transcriptomic analysis for GeoMx DSP data. Nucleic Acids Res 2024;52:e2.
94. Gagnon-Bartsch JA, Jacob L, Speed TP. Removing unwanted variation from high dimensional data with negative controls. Berkeley (CA): Tech Reports from Department of Statistics, University of California; 2013. p. 1–112. Report No.: 820.
95. Ritchie ME, Phipson B, Wu DI, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
96. Chu Y, Dai E, Li Y, Han G, Pei G, Ingram DR, et al. Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. Nat Med 2023;29:1550–62.
97. Cheng S, Li Z, Gao R, Xing B, Gao Y, Yang Y, et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell 2021;184:792–809.e23.
98. Ma RY, Black A, Qian BZ. Macrophage diversity in cancer revisited in the era of single-cell omics. Trends Immunol 2022;43:546–63.
99. Ye X, Wang L, Nie M, Wang Y, Dong S, Ren W, et al. A single-cell atlas of diffuse large B cell lymphoma. Cell Rep 2022;39:110713.
100. Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: a generalist algorithm for cellular segmentation. Nat Methods 2021;18:100–6.
101. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
102. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 2013;14:7.
103. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol 2015;33:495–502.
104. Goncharov NV, Popova PI, Avdonin PP, Kudryavtsev IV, Serebryakova MK, Korf EA, et al. Markers of endothelial cells

in normal and pathological conditions. Biochem (Mosc) Suppl Ser A Membr Cell Biol 2020;14:167–83.
105. Crawford A, Angelosanto JM, Kao C, Doering TA, Odorizzi PM, Barnett BE, et al. Molecular and transcriptional basis of CD4+ T cell dysfunction during chronic infection. Immunity 2014;40:289–302.
106. Lines JL, Pantazi E, Mak J, Sempere LF, Wang L, O'Connell S, et al. VISTA is an immune checkpoint molecule for human T cells. Cancer Res 2014;74:1924–32.
107. Gupta PK, Godec J, Wolski D, Adland E, Yates K, Pauken KE, et al. CD39 expression identifies terminally exhausted CD8+ T cells. PLoS Pathog 2015;11:e1005177.
108. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Vaz Meirelles G, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 2013;14:128.
109. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 2019;16:1289–96.
110. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol 2019;37:38–44.
111. Wolf FA, Angerer P, Theis FJ. Scanpy: large-scale single-cell gene expression data analysis. Genome Biol 2018;19:15.
112. Fang Z, Liu X, Peltz G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics 2023;39:btac757.
113. Baker EAG, Schapiro D, Dumitrascu B, Vickovic S, Regev A. In silico tissue generation and power analysis for spatial omics. Nat Methods 2023;20:424–31.