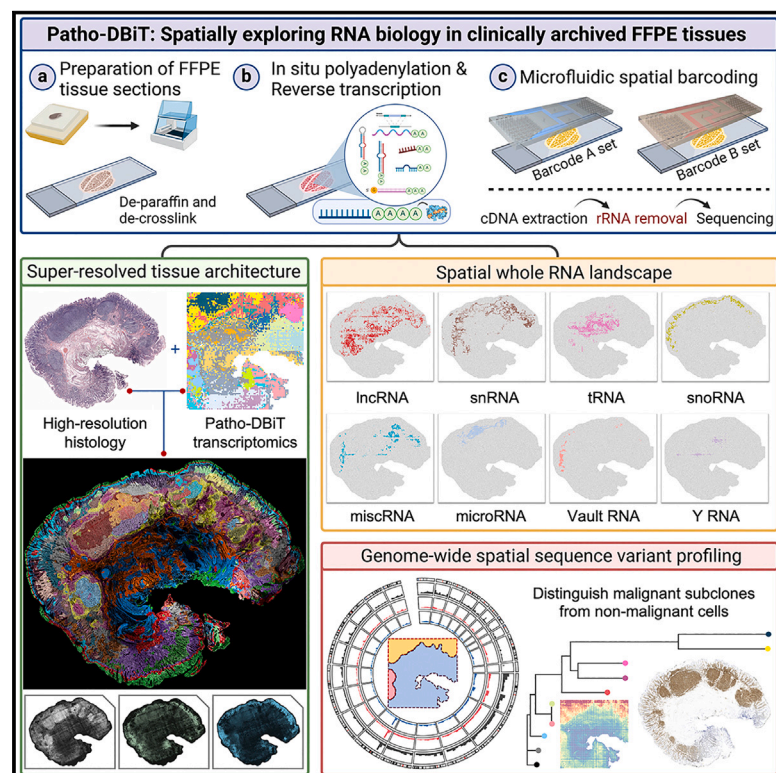# Spatially exploring RNA biology in archival formalin-fixed paraffin-embedded tissues

Graphical abstract

Authors

Zhiliang Bai, Dingyao Zhang, Yan Gao, ..., Jun Lu, Mina L. Xu, Rong Fan

Correspondence

zhiliang.bai@yale.edu (Z.B.), mingyao@pennmedicine.upenn.edu (M.L.), xingyi@chop.edu (Y.X.), jun.lu@yale.edu (J.L.), mina.xu@yale.edu (M.L.X.), rong.fan@yale.edu (R.F.)

In brief

Patho-DBiT enables spatial profiling of a range of RNA species (i.e., mRNA, microRNA, tRNA, splicing isoform, etc.) all at once in clinical formalin-fixed paraffin-embedded tissue with high sensitivity and coverage. It further allows for genome-wide mapping of single-nucleotide RNA variants to distinguish subclones and dissect spatiotemporal tumorigenesis at the cellular level.

Highlights

- Patho-DBiT enables spatial whole transcriptome sequencing in clinical FFPE tissues

- It permits spatial co-profiling of gene expression, splicing, and non-coding RNAs

- Genome-wide detection of spatial SNV distribution distinguishes malignant subclones

- It dissects spatiotemporal dynamics of lymphomagenesis at the single-cell level

CellPress

Cell

CellPress
OPEN ACCESS

Resource

Spatially exploring RNA biology in archival formalin-fixed paraffin-embedded tissues

Zhiliang Bai,[1,11,*] Dingyao Zhang,[2,3,11] Yan Gao,[4,11] Bo Tao,[5,11] Daiwei Zhang,[6,11] Shuozhen Bao,[1,11] Archibald Enninful,[1] Yadong Wang,[2] Haikuo Li,[1] Graham Su,[1] Xiaolong Tian,[1] Ningning Zhang,[2] Yang Xiao,[7] Yang Liu,[5] Mark Gerstein,[3] Mingyao Li,[6,8,*] Yi Xing,[4,8,*] Jun Lu,[2,9,*] Mina L. Xu,[5,*] and Rong Fan[1,5,9,10,12,*]

[1]Department of Biomedical Engineering, Yale University, New Haven, CT 06520, USA
[2]Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA
[3]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA
[4]Center for Computational and Genomic Medicine, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[5]Department of Pathology, Yale University School of Medicine, New Haven, CT 06520, USA
[6]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[7]Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA
[8]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[9]Yale Stem Cell Center and Yale Cancer Center, Yale University School of Medicine, New Haven, CT 06520, USA
[10]Human and Translational Immunology, Yale University School of Medicine, New Haven, CT 06520, USA
[11]These authors contributed equally
[12]Lead contact
*Correspondence: zhiliang.bai@yale.edu (Z.B.), mingyao@pennmedicine.upenn.edu (M.L.), xingyi@chop.edu (Y.X.), jun.lu@yale.edu (J.L.), mina.xu@yale.edu (M.L.X.), rong.fan@yale.edu (R.F.)
https://doi.org/10.1016/j.cell.2024.09.001

SUMMARY

The capability to spatially explore RNA biology in formalin-fixed paraffin-embedded (FFPE) tissues holds transformative potential for histopathology research. Here, we present pathology-compatible deterministic barcoding in tissue (Patho-DBiT) by combining *in situ* polyadenylation and computational innovation for spatial whole transcriptome sequencing, tailored to probe the diverse RNA species in clinically archived FFPE samples. It permits spatial co-profiling of gene expression and RNA processing, unveiling region-specific splicing isoforms, and high-sensitivity transcriptomic mapping of clinical tumor FFPE tissues stored for 5 years. Furthermore, genome-wide single-nucleotide RNA variants can be captured to distinguish malignant subclones from non-malignant cells in human lymphomas. Patho-DBiT also maps microRNA regulatory networks and RNA splicing dynamics, decoding their roles in spatial tumorigenesis. Single-cell level Patho-DBiT dissects the spatiotemporal cellular dynamics driving tumor clonal architecture and progression. Patho-DBiT stands poised as a valuable platform to unravel rich RNA biology in FFPE tissues to aid in clinical pathology evaluation.

INTRODUCTION

Spatial transcriptomics is revolutionizing our understanding of developmental biology, oncology, and disease pathology, mapping intricate gene expression patterns within their native tissue context.[1–4] However, to date, the scope of the field primarily revolves around the analysis of messenger RNA (mRNA) expression. The transcriptome in eukaryotic cells is a myriad of all dynamic and diverse RNA molecules, encompassing not only mature mRNAs encoding protein synthesis but also splicing variants, small RNAs, and other non-coding RNAs with regulatory functions.[5,6] Thus, spatial profiling of different RNA species throughout their life cycle[7,8] is imperative for dissecting the full picture of RNA biology in complex tissues.

Formalin-fixed paraffin-embedded (FFPE) tissues are essential in clinical practice, being the backbone of human histopathological diagnoses.[9] Pathology departments have accrued vast collections of FFPE blocks over time, creating a rich yet underutilized compendium of materials that, accompanied by clinical data, stands as a treasure trove for human biology and translational research.[10] Nevertheless, FFPE specimens pose certain challenges. The RNA within these samples is susceptible to fragmentation during the paraffin-embedding process and may further experience heightened degradation under suboptimal storage. Additionally, RNA may undergo chemical modifications, resulting in fragmentation or resistance to the enzymatic reactions required for sequencing. The loss of poly-A tails introduces another layer of complexity, restricting the utility of oligo-dT

Figure 1. Patho-DBiT workflow and spatial whole transcriptome mapping of mouse embryo

(A) Schematic workflow, molecular underpinnings, and technological spectrum of Patho-DBiT.

(B) Patho-DBiT's performance on an E13 mouse embryo section. Top left: H&E of an adjacent section. Top right: tissue scanning. Bottom: unsupervised clustering.

(C) Spatial pan-mRNA and UMI count maps.

(D) Correlation analysis between replicates, with the Pearson coefficient indicated.

(E) Read coverage along the gene body and the percentage of reads mapped to the 5′ UTR.

(F) Proportion of reads mapped to different RNA categories.

(G) Comparison of non-coding RNA read mapping ratios between Patho-DBiT and normal DBiT. Data shown are mean ± SD ($n = 2$).

(H) Spatial count maps of different non-coding RNAs.

(I) Left: cell patterning and embedding. Right: relative miR-142 expression in KO and WT cells. Data shown are mean ± SD ($n = 3$).

(J) Spatial miR-142 expression in patterned cells (I) and the evaluation of detection sensitivity and specificity.

(legend continued on next page)

primed reverse transcription. Consequently, options for spatially profiling RNA molecules in this tissue type are limited.

In this evolving landscape, we present pathology-compatible deterministic barcoding in tissue (Patho-DBiT) not only enabling spatial full-coverage base-by-base whole transcriptome sequencing but also crafted to address the challenges of clinically archived FFPE tissues (Figure 1A). Patho-DBiT integrates *in situ* polyadenylation,[11,12] microfluidic in tissue barcoding,[13] and computational innovations to decode rich RNA biology inherent in FFPE samples. The platform capitalizes on RNA fragmentation naturally occurring in FFPE specimens and appends poly(A) tails to a broad spectrum of RNA species, thereby overcoming traditional barriers associated with FFPE samples and even outperforming the assays conducted with frozen tissues. By spatially barcoding these RNA molecules, we now can gain a deeper appreciation of high-sensitivity transcriptomics, alternative splicing, genetic variation profile, microRNA regulation, and RNA dynamics within clinical tissues.

RESULTS

Patho-DBiT design and spatial whole transcriptome mapping of mouse embryo

Patho-DBiT initiates with tissue section deparaffinization and heat-induced crosslink reversal (Figure 1A). After tissue permeabilization, enzymatic *in situ* polyadenylation enables detection of various RNAs, followed by complementary DNA (cDNA) strand synthesis. Spatial barcoding is then achieved using microfluidic chips featuring 50 or 100 microchannels.[13,14] Post-imaging, the tissue undergoes digestion to extract barcoded cDNA to perform the downstream procedures. Polyadenylation adds poly(A) tails to all RNAs, including the predominant ribosomal RNAs (rRNAs), which constitute 80%–90% of cellular RNAs but provide limited information on the target transcriptome. To circumvent the loss of rare or low-abundance transcripts, cDNAs originating from rRNAs were selectively removed from amplicons, significantly reducing these fragments before sequencing (Figure S1A).

We applied Patho-DBiT to E13 mouse embryo FFPE section using a 50 μm pixel-sized device. Unsupervised clustering revealed 20 transcriptomic clusters, and the spatial UMAP closely aligned with the histology of an adjacent section stained with H&E (Figure 1B). Cell type-specific marker genes were identified, characterizing their expression within each individual cluster (Figure S1B). The distribution of these clusters exhibited conspicuous and distinctive spatial patterns, underscoring the high accuracy of this assay (Figure S1C). Patho-DBiT detected an average of 5,480 genes and 15,381 unique molecular identifiers (UMIs) per pixel, with the genome-wide pan-mRNA and UMI maps displaying a strong concordance with tissue morphology and density (Figure 1C). Reproducibility among replicates conducted on adjacent slices was notably high (Figures 1D and S1D). To discern the cell identities within each

cluster, we integrated our datasets with single-cell RNA sequencing (scRNA-seq) reference data from E13.5 mouse organogenesis.[15] This integration yielded a cohesive pattern, aligning our spatial pixels with the scRNA-seq dataset in a well-conformed manner (Figure S1E).

To assess read coverage across gene bodies, we generated replicate datasets on adjacent E13 sections using normal DBiT-seq without polyadenylation. While still exhibiting a 3′ bias, Patho-DBiT showcased an approximate 2-fold increase in coverage across the entire gene body, and the percentage of reads mapped to 5′ untranslated region (UTR) more than doubled (Figure 1E). In both replicates, ∼47% of reads were mapped to protein-coding RNA, while the remaining reads were mapped to long non-coding RNA (lncRNA), microRNA, miscellaneous RNA (miscRNA), rRNA, small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), small Cajal body-specific RNA (scaRNA), transfer RNA (tRNA), vault RNA, Y RNA, or other uncategorized RNAs (Figure 1F). Among these, tRNA, snRNA, and lncRNA constituted a relatively high proportion of the reads, consistent with their known high abundance in the cell.[16] Notably, Patho-DBiT demonstrated a 10- to 100-fold higher capture rate for most non-coding RNAs compared with the previous DBiT-seq protocol (Figure 1G). It maintained a low level of rRNA reads (∼3%), reaffirming the efficacy of removing this undesired category. Spatial patterns of these non-coding RNAs revealed general enrichment in the brain and liver regions, mirroring the distribution trend observed for coding mRNAs (Figure 1H). We examined the spatial expression of tRNA isotypes by grouping reads encoding the same amino acid, revealing high levels of aspartic acid and asparagine tRNAs in the heart and isoleucine and threonine tRNAs in the brain (Figure S1F). These findings align with published data showing expression level in adult mouse tissues measured using demethylase-tRNA sequencing.[17]

To benchmark Patho-DBiT's accuracy in detecting small RNA, we used miR-142 knockout (KO) Ba/F3 cells and assembled them into a pattern of the letter "Y," surrounded by wild-type control cells (Figure 1I). The assembled cell aggregate is embedded and sectioned for Patho-DBiT sequencing. The detected miR-142 distribution negatively correlates with the letter "Y" region, which was filled with KO cells, demonstrating a detection sensitivity of 86.95% and specificity of 84.19% (Figure 1J). We next examined the microRNA profile in the E13 section. In total, Patho-DBiT detected 1,063 microRNAs, peaking at 22 nt in the count of mapped reads (Figure 1K), consistent with the typical length of microRNAs.[18] To evaluate the mapping accuracy, we focused on miR-122, one of the earliest examples of a tissue-specific microRNA, constituting ∼70% of the total microRNA pool in the liver.[19] The miR-122 reads are aligned to miR-122-5p, the known dominant mature strand produced from the pre-miR-122 hairpin (Figure 1L). Additionally, miR-122 is enriched within the liver area in the embryo section.

We applied Patho-DBiT across diverse tissue types and spatial pixel sizes, yielding remarkable results (Figure 1M). At a

(K) Detection of microRNAs in the E13 section.

(L) Left: read coverage of miR-122 mapped to the genome location. Right: spatial distribution of miR-122.

(M) Distribution of gene and UMI counts in different tissue types at varying spatial pixel sizes.

See also Figure S1.

50 μm pixel size, our approach exhibited superior performance compared with the probe-based Visium for FFPE at a 55 μm feature size, even surpassing its fresh frozen counterpart. At a 20 μm pixel size, we identified over 4,000 genes per pixel in lymphoma sections and more than 3,000 genes in samples from mouse lymph nodes. Notably, employing device with 10 μm channels, we identified 2,292 genes and 6,021 UMIs from a lymphoma section at this near-cellular level.

Spatial co-profiling of region-specific alternative splicing and A-to-I RNA editing in the mouse brain

To demonstrate Patho-DBiT's capability to simultaneously map gene expression and post-transcriptional RNA processing, we profiled an FFPE mouse coronal brain section. An average of 6,786 genes and 31,063 UMIs were detected per pixel, exhibiting a spatial distribution strongly correlated with the tissue histology (Figure 2A). Clustering analysis unveiled 15 anatomical clusters characterized by specific gene markers, and the spatial arrangements broadly aligned with the region annotations on a similar section from the Allen Brain Atlas (Figures 2B and S2A). The isocortex area was precisely deconstructed into 3 layers, assigning cluster 7 to layer 1–2, cluster 4 to layer 4–5, and cluster 10 to layer 6a–b.[20] Spatial pattern of the primary defining gene in each cluster closely mirrored the in situ hybridization (ISH) results for the same genes (Figure S2B), underscoring its capacity to faithfully reflect fine tissue structures. Integration and co-embedding the Patho-DBiT data with scRNA-seq atlas from cells in the mouse cortex and hippocampus validated the identity of these clusters[21] (Figure 2C). Particularly, cells in clusters 0, 3, 9, 11, and 12, located in the midbrain or hindbrain areas, remained largely unmapped due to the absence of corresponding cells in the reference scRNA-seq dataset.

Alternatively spliced transcripts play a crucial role in neurogenesis and brain development.[22] While long-read sequencing is ideal for capturing the full exonic structure of a transcript, short-read RNA-seq can still identify splicing events, offering advantages like greater read depth and cost efficiency.[23–25] Through poly(A) tailing to the RNA transcript fragments (Figure S2C), Patho-DBiT exhibited remarkably broader coverage across the gene body than the Visium spatial transcriptomics on a frozen mouse brain section (Figure S2D). We detected a total of 3,879 distinct alternative splicing events in 2,368 genes when at least 2 splice-junction-spanning read counts were required (Figure S2E), covering all the major event types, including skipped exon (SE), retained intron (RI), alternative 3′ splice site (A3SS), alternative 5′ splice site (A5SS), and mutually exclusive exon (MXE). This number substantially decreased when the cutoff threshold was increased to 10 read counts (Figure S2E). At the 2-read threshold, each spatial spot yielded an average of 105 alternative splicing events in 85 genes, with a mean coverage level of 43 UMIs per event when aggregating all the pixels (Figure S2F).

Although current read coverage may not reliably infer splicing at the pixel level, we explored significant regional differences by aggregating reads from all pixels within an anatomical region (Figure 2D). At the regional level, alternative splicing events were detected in 393 to 1,179 genes, with a median of hundreds of read counts covered, significantly surpassing those of genes without alternative splicing events detected (Figure S2G). Utilizing criteria

set by the computational program rMATS,[23] we identified dozens to hundreds of differential alternative splicing events between two brain regions, depending on the read count thresholds applied (Figure 2E). We further examined the top-ranked genes involved in brain functions that show pronounced regional isoform switching (Figure 2F). For example, Myl6, a gene widely involved in neuronal migration and synaptic remodeling[26] with a uniform distribution of gene expression across the entire section, exhibited an enriched inclusion isoform (Myl6-206) in the fiber tracts and hindbrain, in contrast to the skipping isoform (Myl6-201) enriched in CA1 (Figure 2G). Ppp3ca has been identified as a leading modulator of genetic risk in Alzheimer's disease.[27] Our data unveiled distinct isoform usage patterns, with the inclusion isoform (Ppp3ca-201) prevailing in the isocortex and CA1, while the skipping isoform (Ppp3ca-202) was exclusively expressed in the DG (Figure 2H). Patho-DBiT also identified spatial isoform distribution of Stxbp1, a functional gene regulating neurotransmitter release[28] (Figure S2H). The spatial expression patterns of both genes and their isoforms are consistent with an independent dataset from frozen coronal sections, where isoform data were obtained using long-read nanopore sequencing[29] (Figure S2I).

Transcriptomic diversity expands through adenosine-to-inosine (A-to-I) RNA editing, an essential process vital for proper neuronal function.[30] With superior read coverage, Patho-DBiT spatially mapped A-to-I editing in situ on FFPE brain sections, unveiling a distinctive editing ratio landscape across different regions (Figure 2I). In line with prior findings,[31] thalamus exhibited a notably elevated editing ratio (mean 27.9%), while fiber tracts displayed a lower ratio (mean 12.7%). This pattern closely corresponds to the expression levels and spatial frequencies of genes encoding A-to-I editing enzymes, known as adenosine deaminases acting on RNA (ADAR).[32] Particularly, the spatial distribution of Adarb1 closely mirrored the editing ratio (Figure 2J). Intriguingly, Adar showed ubiquitous expression without correlation to RNA editing (Figure S2J). We next cross-referenced the editing ratio with the abovementioned published dataset.[29] In addition to the robust correlations in regional Adarb1 expression and editing ratio between the two technologies (Figures S2K and S2L), we observed a high correlation across 259 commonly detected editing sites showing at least 10 UMIs in both datasets (Figures 2K and S2M; Table S1). Therefore, Patho-DBiT provides a spatial delineation of gene expression, region-specific splicing isoforms, and A-to-I editing in the FFPE mouse brain section.

Patho-DBiT recapitulates lymphomagenesis biology in the clinical archival AITL sample

We next extended the spatial profiling to clinically archived FFPE tissues. Patho-DBiT was employed to barcode a tissue section obtained from the subcutaneous nodule of a patient diagnosed with angioimmunoblastic T cell lymphoma (AITL) (Figure S3A). This block had been stored at room temperature for over 5 years prior to our assay. Unsupervised clustering revealed 10 spatially organized clusters aligned with histological structures (Figure 3A). The UMAP, with an average of 5,364 genes and 11,989 UMIs per pixel, delineated distinct cell types defined by canonical markers (Figures 3B and S3B). Notably, cluster 5 exhibited high expression of connective tissue cell-related genes,
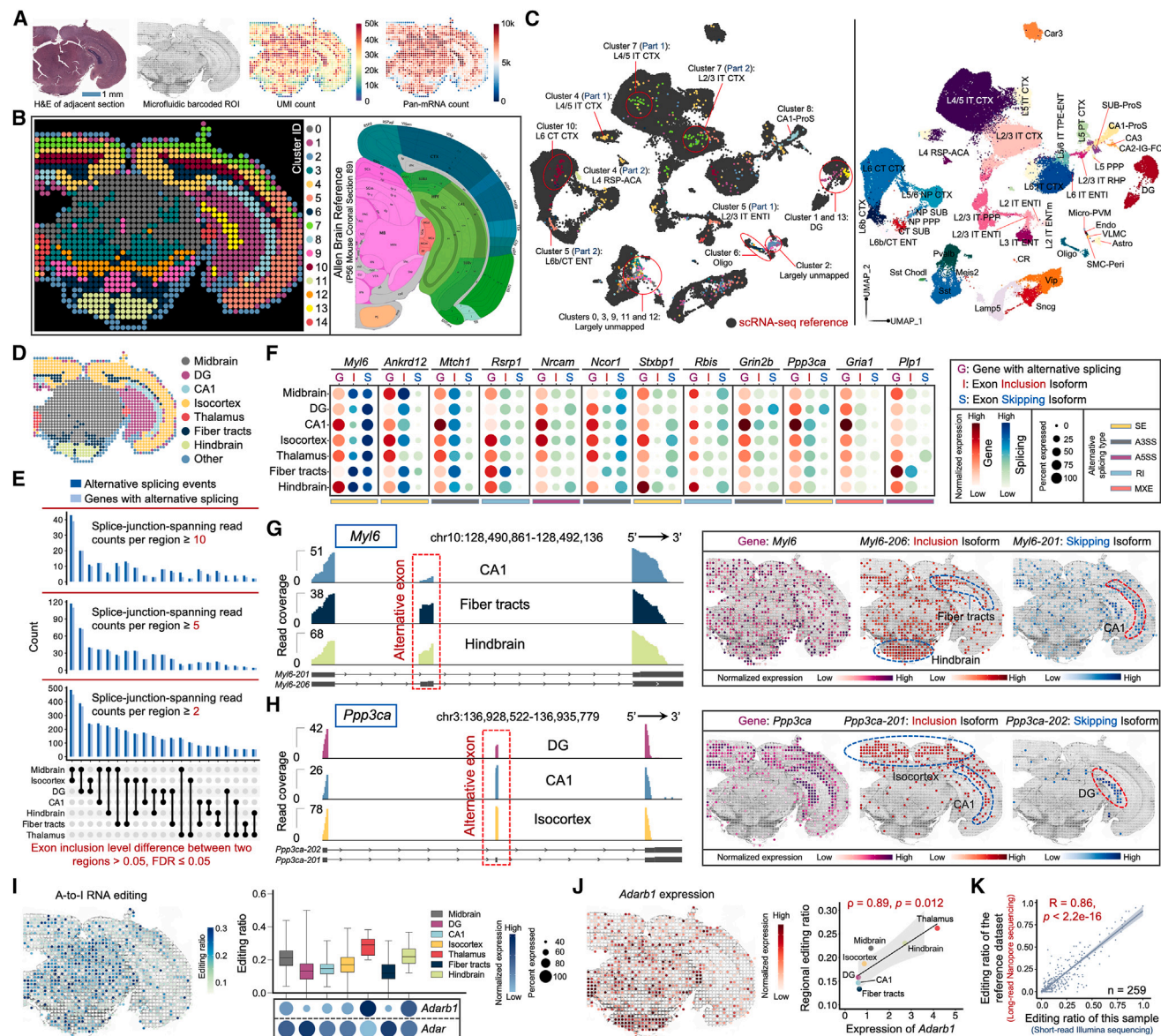
Figure 2. Spatial co-mapping of gene expression and RNA processing in the mouse brain

(A) Patho-DBiT profiling of a mouse brain section. Left: H&E of an adjacent section. Middle: tissue scanning. Right: spatial pan-mRNA and UMI count maps.

(B) Unsupervised clustering. The distribution aligned with the region annotation from the Allen Brain Atlas.

(C) Integration of spatial data with scRNA-seq dataset.

(D) Anatomical brain region labeling in the Patho-DBiT dataset.

(E) Number of significant differentially spliced events and parental genes between brain region pairs under different read count thresholds.

(F) Top-ranked 12 genes exhibiting significant regional differences in exon inclusion levels.

(G and H) Junction read coverage of *Myl6* (G) and *Ppp3ca* (H) splicing event in specific brain regions.

(I) Left: spatial variations in A-to-I RNA editing. Right: distribution of editing ratio across all editing sites and the expression level of ADAR-encoding genes in different brain regions.

(J) Left: spatial *Adarb1* expression. Right: correlation between the *Adarb1* expression and the average regional editing ratio across various brain regions, with the Spearman coefficient indicated.

(K) Editing ratio correlation between 259 sites commonly detected by Patho-DBiT and long-read nanopore sequencing, as reported in the reference literature, with the Pearson coefficient indicated.

See also Figure S2 and Table S1.

Figure 3. High-sensitivity spatial transcriptomics of an AITL sample stored for 5 years

(A) Spatial transcriptome mapping of an FFPE AITL tissue stored for 5 years. Left top: H&E of an adjacent section. Left bottom: tissue scanning. Right: unsupervised clustering.

(B) Top-ranked DEGs defining each cluster.

(C) Spatial phenotyping of an adjacent section using CODEX technology.

(D) Spatial distributions of B cells, T cells, and macrophages revealed by Patho-DBiT, exhibiting a strong Pearson correlation with the proteomic data from CODEX.

(E) Top: CODEX data from the yellow square indicated area in (C). Bottom: DEGs in cluster 0 corresponding to the indicated region.

(F) Ligand-receptor interactions within cluster 0.

(G) Signaling pathways regulated by the DEGs in cluster 0. z score is computed and used to reflect the predicted activation level ($z > 0$, activated; $z < 0$, inhibited; $z \geq 2$ or $z \leq -2$ can be considered significant).

(H) Graphical network of canonical pathways, upstream regulators, and biological functions regulated by DEGs identified in cluster 0.

See also Figure S3.

exclusively within the region displaying consistent morphology in the H&E staining.

To assess capture accuracy in this long-term stored FFPE section, we conducted high-plex spatial cell typing on an adjacent tissue section using the co-detection by indexing (CODEX) assay.[33] The proteomic data faithfully depicted the histological profile, revealing a population of malignant T follicular helper (Tfh) cells with strong CD4 expression and minimal CD8 and granzyme B (Figures 3C, S3C, and S3D). We traced the expression patterns of B cells, T cells, and macrophages and compared them to corresponding surface markers from CODEX data, revealing notable spatial correlations between the two modalities (Figure 3D). The proliferation markers *MKI67* and *PDCD1*, frequently expressed on malignant AITL cells, exhibited a dense nodular distribution in this section, consistently identified by CODEX (Figure S3E).

Analysis of differentially expressed genes (DEGs) in cluster 0 revealed significant upregulation of B-cell markers, malignant Tfh cell markers, and markers associated with the abnormally proliferative follicular dendritic cells in AITL.[34] The profile from high-power CODEX imaging of this region corroborated the active expression of CD20, CD4, and CD21 (Figure 3E). This intricate composition, evocative of AITL's distinctive tumor microenvironment, warrants further investigation. Cell-cell interaction analysis highlighted that the most prevalent interaction pair in this region was between *CXCL13* and its receptor genes (Figure 3F). Given its high expression in nearly all cases, CXCL13 serves as a specific diagnostic marker for AITL, and its interaction with CXCR5 is deeply implicated in tumorigenesis.[35] Patho-DBiT accurately unveiled this regulatory mechanism, complemented by the spatial distribution profile (Figure S3F).

We then delved into signaling pathways governed by DEGs within cluster 0. Alongside the activation of fundamental T cell immune functions in this cluster (Figure S3G), our analysis pinpointed the upregulation of a cascade of AITL-specific pathways integral to the pathogenesis process,[36] including PI3K/AKT, CD40, JAK/STAT, NF-κB, ICOS-ICOSL, VEGF, mTOR, and ERK/MAPK (Figure 3G). Conversely, general cancer suppression signals like p53 and PTEN were significantly inhibited within this cluster. Moreover, Rho GTPases signaling, known for their functional involvement in the initiation and progression of AITL,[37] showed significant activation, coinciding with upregulated pathways associated with cell cycle regulation and metabolism. Together, Patho-DBiT recapitulates the underlying biology of lymphomagenesis in an AITL biopsy sample archived at room temperature for over 5 years. This comprehensive understanding is summarized through a graphical network encompassing canonical pathways, upstream regulators, and biological functions (Figure 3H).

Super-resolved tissue architecture by integrating expanded Patho-DBiT with high-resolution histology

We further used a 100 × 100 microfluidic device to barcode 10,000 spots at a 20 μm pixel size to spatially profile a clinical tumor biopsy from a patient diagnosed with extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue (MALT), a low-grade non-Hodgkin B-cell lymphoma (BCL).[38] The section was prepared with an FFPE block stored for 3 years at room temperature. Biopsy from the gastric antrum of the stomach showed

a dense nodular infiltrate of lymphocytes primarily in the mucosa, as detected by endoscopy (Figure S4A).

In all the aforementioned samples, Patho-DBiT primarily includes reads mapped to exonic regions. While these exonic reads yield an average of 2,190 genes and 4,202 UMIs per pixel in the MALT section, we detected a substantial number of intronic reads, corresponding to a mean per pixel count of 3,749 genes and 6,535 UMIs (Figure S4B). By aggregating both expression matrices while preserving their individual identities, we identified 20 clusters that spatially mirrored the histological structures (Figure 4A). The localization of these clusters revealed clear and distinctive spatial patterns (Figure S4C). Representative cell types such as B cells, macrophages, plasma cells, and mucus-secreting cells were delineated based on canonical marker gene expression (Figure 4B). We noticed a faint expression of the plasma cell score and macrophage cell score in the designated regions P and M in Figure 4B. To validate that this signal reflects actual cellular presence rather than background noise, we conducted immunofluorescence (IF) assays targeting the immunophenotypic markers CD138 for plasma cells and CD68 for macrophages on adjacent sections (Figure 4C). The results confirmed the cell identities, providing further support for Patho-DBiT's capability to detect relatively low-abundance cell types in specific regions. Furthermore, with this improved resolution, Patho-DBiT was able to identify vasculature among the surrounding BCL cells, distinguish lymphatics from neighboring smooth muscle cells, and locate plasma cells present in the lamina propria underneath the gastric epithelium (Figure 4D). These transcriptomic neighborhoods are closely aligned with the tissue histological features annotated by an experienced pathologist using H&E staining.

Deep data fusion of spatial transcriptome and histological image has been demonstrated to generate transcriptome-wide super-resolved expression maps and further evolved to derive whole tissue slide-scale transcriptomic atlas.[39–41] Here, we utilized the newly developed computational pipeline iStar[41] to improve spatial gene expression mapping to the single-cell level. The iStar unsupervised segmentation produced intricate molecular maps of cells, also enabling gene expression prediction across the whole tissue slide (Figure 4E). Overall, the spatial distributions of gene markers for major cell types were accurately aligned with the presence of these cells as confirmed by the pathologist, with reference to original histological sections and accompanying immunohistochemical studies. Within each cell type, cellular heterogeneities were revealed at a refined level. For example, tumor B cells were classified into nine subtypes based on their cellular activities and molecular signatures, while epithelial cell heterogeneities were defined by their functional and metabolic states. The underlying biological relevance of each identified cluster was confirmed by the top-ranked expressed genes within the cluster (Figure S4D). We also calculated correlation matrices depicting the relationships among iStar clusters. Similar cell types grouped together, while divergent cell types exhibited significant negative correlations, further supporting the utility of this analysis (Figure S4E).

Spatial RNA splicing dynamics reveals the trajectory of tumor cell development

RNA velocity reconstructs dynamic gene expression changes by leveraging the ratio of newly transcribed, unspliced pre-mRNAs
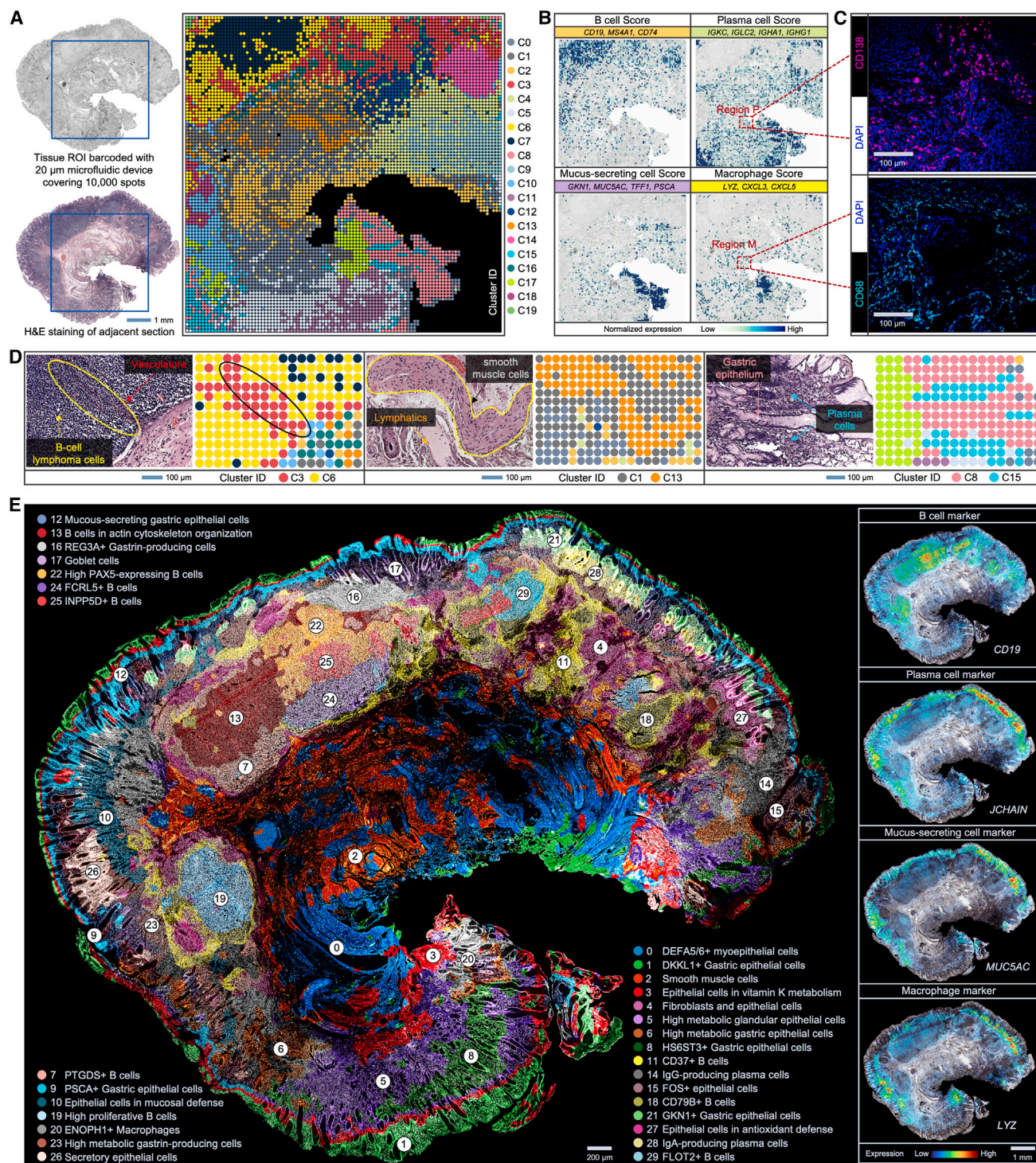
Figure 4. High-resolution tissue architecture resolved by integrating Patho-DBiT with histology

(A) Spatial mapping of the MALT section using 20 μm Patho-DBiT device covering 100,000 spots. Left top: tissue scanning. Left bottom: H&E of an adjacent section. Right: unsupervised clustering.

(B) Spatial identification of representative cell types.

(C) IF staining of plasma cell marker (CD138) and macrophage marker (CD68) in the selected regions P and M in (B).

(D) Representative transcriptomic neighborhoods revealed by Patho-DBiT, closely aligned with H&E tissue morphology.

(E) Super-resolved high-resolution spatial clustering and representative gene expression enhanced with iStar.

See also Figures S4 and S5.
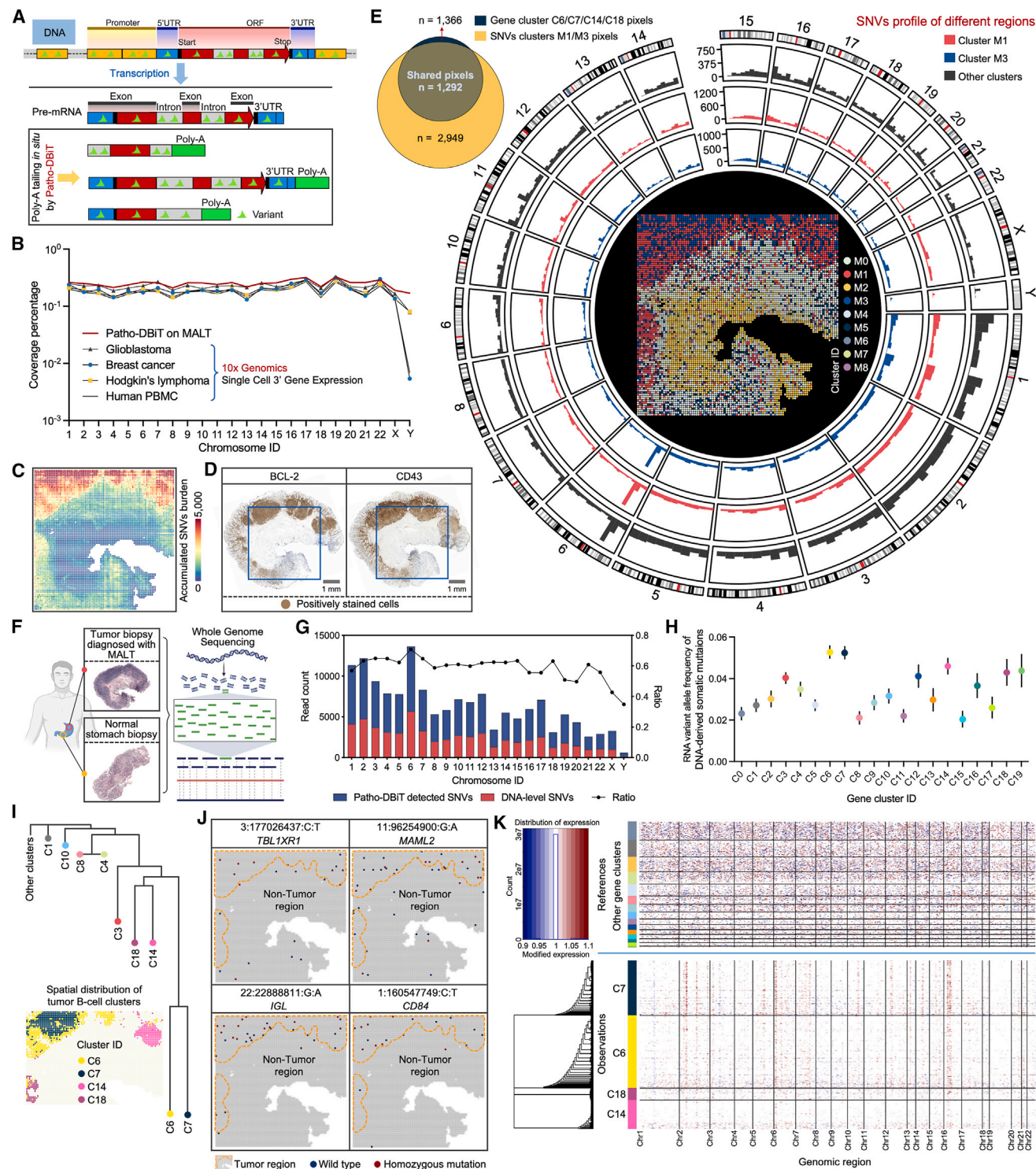
Figure 5. Genome-wide spatial variant profiling for differentiating malignant subclones

(A) Molecular underpinnings of detecting variations printed in pre-mRNA transcripts.

(B) Comparison of genomic location coverage bandwidth between Patho-DBiT and scRNA-seq datasets.

(C) Spatial expression map of accumulated SNVs burden.

(D) IHC staining of BCL-2 and CD43 commonly used for immunophenotyping MALT tumor cells.

(legend continued on next page)

to mature, spliced mRNAs, with the former identifiable through the presence of introns.[6,42] Typically, bulk sequencing or scRNA-seq datasets comprise only 15%–25% intronic sequences, primarily from secondary priming positions within intronic regions.[6] Patho-DBiT generated a higher proportion of intronic reads compared with exonic ones, enhancing RNA velocity analysis. Employing scVelo,[42] we delineated transient cellular states of all the identified clusters. While such velocity direction analysis could be incorrect for cell types that are not supposed to have any cell state transition or differentiation relationship,[43] we focused our analysis exclusively on the tumor B-cell clusters. We observed a cellular dynamics trajectory originating from C14 to C18, progressing to C6, and then extending toward C7 (Figure S5A), potentially unveiling their differentiation stages as defined by splicing rates. The 4 clusters exhibited no prominent variations in cell cycle stages, also evidenced through sparse immunohistochemistry (IHC) staining of Ki67 (Figure S5B). Furthermore, we identified prominent genes that drove the primary processes of this dynamic behavior, with the top-ranked genes clearly displaying higher splicing activities in C7 compared with C6, C14, and C18 (Figure S5C).

Next, we inferred a universal pseudotime shared among genes that represents the cell's internal clock. While all the clusters outside the tumor region displayed a static cell fate, changes began to manifest from C14 and progressively intensified toward C7 (Figure S5D). This spatial directional transition, coupled with the velocity flow from C14 to C7, further implies that tumor cells in C7 pixels were produced at a later stage. Putative driver genes contributing to this pseudotime trajectory were identified, among which the top-ranked gene, *BCL2*, is a key antiapoptotic regulator critical for lymphoma pathogenesis[44] (Figure S5E). Together, with a superior capture efficiency for intronic reads, Patho-DBiT spatially mapped RNA splicing dynamics associated with the trajectory of malignant B-cell development.

Genome-scale spatial sequence variant profiling for tumor discrimination

Sequence variations were common in RNA transcripts, largely reflecting underlying genetic mutations or RNA editing, a process frequently deregulated in cancer.[45,46] By *in situ* poly(A) tailing (Figure 5A), we hypothesized that Patho-DBiT could effectively capture sequence variations printed in pre-mRNA transcripts. Higher genomic coverage will enable more reliable detection of sequence variants. We therefore first conducted a comparative analysis of the genomic location coverage bandwidth between Patho-DBiT and scRNA-seq datasets. Here, the coverage percentage for each chromosome is calculated by dividing the cumulative length of sequenced regions in anno-

tated GENCODE[47] transcripts by the total transcript length on a chromosome. Our spatial FFPE MALT data showed higher coverage capability than scRNA-seq datasets from both fresh human cancer samples and healthy donor peripheral blood mononuclear cells (PBMCs), allowing Patho-DBiT to capture genome-wide sequence variations (Figure 5B).

To delineate the RNA variant profile of this MALT section, we implemented a variant calling pipeline to identify all potential single-nucleotide variations (SNVs). Each SNV site in a spatial pixel was cataloged and classified based on the reference and variant reads, overall yielding a matrix of detected SNVs (STAR Methods). The spatial heatmap of accumulated SNVs revealed a notably higher SNV burden in the B-cell region (Figure 5C). The tumor signature in these B cells was validated through IHC staining of canonical markers commonly used for immunophenotyping MALT tumor cells, namely BCL-2 and CD43,[48] on adjacent sections (Figure 5D). The expression of these markers exhibited a strong correlation with the spatial SNV burden profile.

We explored the potential of leveraging RNA-level variants for unsupervised tumor discrimination. Spatial clustering of the expressed SNV matrix alone revealed 9 subpopulations (Figure 5E). Clusters M1 and M3 exhibited notable overlap with the tumor region (Figures 5E and S6A). The principal-component analysis (PCA) demonstrated a distinct variation profile of these 2 clusters (Figure S6B), and the counts of genomic regions with SNV frequency >0.01% were significantly higher (Figure S6C), confirming their elevated variant levels. To validate this analysis pipeline, we performed Patho-DBiT on an FFPE lymph node section from a healthy donor using a device with the same pixel size (Figure S6D). This revealed low SNV accumulation across the entire section, with a relatively higher burden in some of the B-cell regions, presumably due to somatic hypermutation occurring in germinal center B cells (Figure S6E). Unsupervised clustering of the SNV matrix revealed no spatial pattern in this donor sample (Figure S6F). To depict the genome-wide distribution of SNVs, somatic variant calling was conducted within the M1 and M3 using pixels from the other 7 clusters as controls, revealing a chromatin- and region-specific pattern of SNVs (Figure 5E).

RNA-level SNVs may arise from several sources, including DNA-level SNVs, RNA modifications, FFPE procedure, suboptimal storage, or sequencing errors.[49,50] Although it is difficult to accurately infer DNA-level SNVs from the RNA data, we reasoned that by adding DNA sequencing of the same tissue, we could determine spatial patterns of DNA-level SNVs. Therefore, we performed bulk whole-genome sequencing (WGS) on an adjacent MALT section paired with a reference normal stomach sample from the same patient (Figure 5F). Comparing the

(E) Unsupervised clustering of the spatial SNVs matrix and genome-wide distribution of somatic variations in clusters M1 and M3. Veen plot showing the pixel overlap between gene and SNVs tumor clusters.

(F) WGS of the MALT section matched with a reference normal stomach sample from the same patient.

(G) Read count of DNA-level SNVs and Patho-DBiT detected RNA-level SNVs and their ratios.

(H) Comparison of RNA variant allele frequencies in DNA-derived somatic mutation sites across gene clusters.

(I) Evolutionary relationships among the identified gene clusters.

(J) Spatial distribution of top-ranked SNV sites upregulated in tumor B-cell clusters.

(K) InferCNV analysis of gene expression intensity across tumor genome positions in tumor B-cell clusters.

See also Figure S6.

SNVs identified in Patho-DBiT with those found in the WGS data, we verified that an average of 58.6% of RNA-level SNVs across chromosomes reflect DNA-level SNVs, most of which were germline variants (Figure 5G). We narrowed down the analysis to only RNA-level SNV sites corresponding to somatic DNA mutations and compared the RNA variant allele frequencies across all transcriptomic clusters. Consistently, higher variant frequencies were observed in the tumor B-cell-enriched clusters, particularly in C6 and C7 (Figure 5H). We also inferred the clonal relationship of these clusters, revealing a distinct separation between tumor clusters and normal tissue clusters (Figure 5I). Compared with C14 and C18, the pixels in C6 and C7 likely represent subclones with different developmental stages. Additionally, cluster C3 exhibited a close relationship with the 4 tumor clusters, possibly reflecting its identity as a mixture of tumor B cells and epithelial cells (Figures 4A and 5D). We then conducted a differential SNV analysis to identify defining variant sites in clusters C6 and C7 compared with normal tissue clusters (Figure S6G). Among the tumor region-enriched sites (Figures 5J and S6H), *TBL1XR1* emerged as a well-documented mutation known to drive extranodal lymphoma by inducing a pro-tumorigenic memory fate,[51] *MAML2* is a coactivator of NOTCH frequently upregulated in lymphomas and leukemias,[52] and *IGL* alteration is a potent enhancer relevant to lymphomagenesis.[53] *CD84*, while it may not directly contribute to tumor formation, regulates the immunosuppressive microenvironment in hematological malignancies.[54] Another notable observation is that no sites were significantly (p value < 0.05) downregulated in these two tumor B-cell clusters.

We further explored the use of Patho-DBiT RNA expression data to identify evidence of somatic large-scale chromosomal copy number variation (CNV) using inferCNV.[55] Normalizing the CNV profiles with pixels from the normal clusters revealed gains in chromosomes 2, 6, and 16 in the tumor cluster C7 (Figure 5K). These copy number gains were similarly observed in our WGS data (Figures S6I and S6J), reaffirming the accuracy of Patho-DBiT detection. Interestingly, cluster C6, as a spatial neighbor of C7, displayed less pronounced gains in these regions. This inconsistency with SNV evolutionary dynamics might suggest differing contributions of CNV and SNV to the tumor clonal architecture. Collectively, these findings suggest that Patho-DBiT possesses the capability to distinguish malignant from non-malignant cell regions in an unbiased manner without the need for known markers. It can dissect spatial clonal architectures based on both spatial SNV and CNV profiles.

## Spatial regulatory network of microRNAs in tumorigenesis

We next sought to assess Patho-DBiT's capacity for co-mapping large and small RNAs in clinical samples. Approximately 35% of the reads were mapped to non-coding RNAs, capturing tens to thousands of RNA molecules along with their spatial distribution (Figures S7A and S7B). A differential expression analysis comparing tumor B-cell clusters to non-tumor clusters uncovered distinct profiles of these RNA species that may contribute to tumorigenesis (Figure S7C; Table S2). Next, we concentrated on the in-depth analysis of microRNAs due to their roles in various pathologies, including cancer.[56] Out of 2,300 true

human mature microRNAs,[57] Patho-DBiT detected 1,352 in the MALT section, with the count of mapped reads accurately peaking at 21 nt (Figure 6A). Regarding the UMI count per pixel, 71.9% had fewer than 10 UMIs, 20.2% had 10–100 UMIs, and the remaining 7.9% had more than 100 UMIs.

We assessed detection accuracy by examining tissue-specific microRNAs. Based on both tissue morphology and the enriched expression of marker genes, cells within clusters C0, C13, and C16 were classified as smooth muscle cells[58] (Figures 6B and S4A). Two smooth muscle cell-specific microRNAs, miR-143 and miR-145,[59] exhibited markedly elevated expression in these 3 clusters, with over 2,000 reads precisely mapped to the genome location, reflecting the dominant mature strands (miR-143-3p and miR-145-5p) from their respective pre-microRNAs (Figure 6C). Their spatial distributions were prominently evident in the smooth muscle cell region. MiR-142 is necessary for the development of marginal zone B cells,[60] while both miR-146a and miR-150 are upregulated in marginal zone lymphomas.[61] Consistently, a notably high expression pattern of these three microRNAs was observed in the tumor B-cell region of the MALT section (Figure S7D).

To further elucidate the regulatory role of microRNAs in tumorigenesis, we conducted a differential microRNA expression analysis between the tumor and non-tumor regions. Among the upregulated microRNAs in the tumor region is miR-21, a well-characterized cancer-promoting "oncomiR,"[62] along with the abovementioned lymphoma-enriched microRNAs (Figure 6D). By contrast, miR-134 and miR-149, two microRNAs known to suppress the proliferation and metastasis of multiple cancer cells,[63,64] were significantly downregulated in the tumor region. Regulatory network analysis revealed positive correlations between the top upregulated microRNAs and multiple genes implicated in lymphomagenesis (Figure 6E), including *NCL* encoding a BCL-2 mRNA binding protein,[65] oncogenic lncRNA SHG7,[66] and *EEF1A1*, potentially contributing to tumor initiation and progression.[67] Conversely, a broad array of genes exhibited negative correlations with these microRNAs, especially miR-21 and miR-4472-2. While miR-4472 has been reported for its role in fostering tumor aggressiveness in breast cancer,[68] our data suggest its potential involvement in this unrelated tumor type. Likewise, interaction analysis was conducted for the top downregulated microRNAs in the tumor region, elucidating how these regulations influence the transcriptomic signatures of tumor B cells (Figure S7E).

We also performed detailed regulatory analysis with miR-21 as the initial example. This microRNA was inferred to significantly regulate 536 genes in the tumor region, with 82.46% identified as cancer-related according to the ingenuity pathway analysis (IPA) database[69] (Figure 6F). As an oncomiR known for its pivotal role in the initiation and development of various B-cell malignancies,[70] miR-155 exhibited significantly higher expression in the tumor region, accompanied by precise genome location mapping and spatial distribution (Figure 6G). This microRNA, being a transactivational target of NF-κB,[71] contributes to the promotion of PI3K-AKT signaling pathway in BCL.[72] Of 200 NF-κB signaling genes defined by gene set enrichment analysis (GSEA), 136 positively correlated with miR-155 (Figure S7F). Similarly, 65 out of 105 genes linked to PI3K-AKT signaling
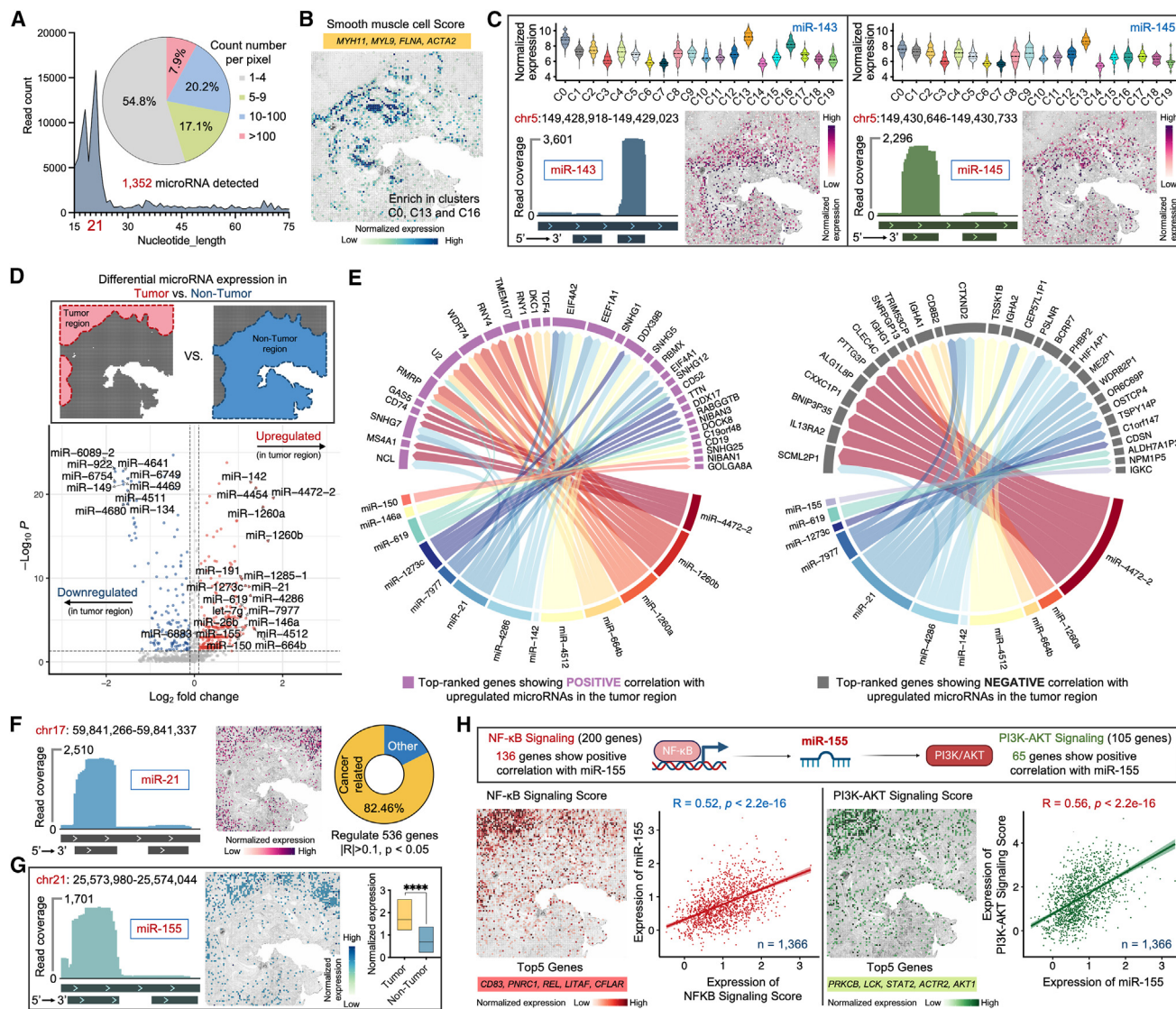
**Figure 6. Spatial microRNA regulations in the MALT section**

(A) MicroRNAs detected by Patho-DBiT in the MALT section. The pie chart illustrates the percentage distribution of the detected count number per spatial pixel.

(B) Spatial distribution of the smooth muscle cell score.

(C) Spatial mapping of smooth muscle cell-specific miR-143 and miR-145.

(D) Differentially expressed microRNAs between the tumor and non-tumor regions.

(E) Regulatory network between the top upregulated microRNAs and the gene expression in the tumor region.

(F) Spatial expression, read coverage, and regulatory relationship of miR-21.

(G) Spatial expression, read coverage, and expression comparison between tumor and non-tumor regions of miR-155. Significance level was calculated with two-tailed Mann-Whitney test, **** $p < 0.0001$.

(H) Spatial interactions involving miR-155 and its upstream and downstream signaling pathways. The Pearson correlation was calculated across 1,366 spatial pixels within the tumor B-cell region.

See also Figure S7 and Table S2.

showed a positive correlation with miR-155 (Figure S7G). The spatial correlation between miR-155 and both signaling pathway scores across 1,366 tumor cluster pixels yielded Pearson correlation coefficients of 0.52 and 0.56, respectively (Figure 6H). Taken together, Patho-DBiT enables spatially resolved co-profiling of large and small RNAs, facilitating the analysis of a microRNA regulatory network in the clinical biopsy.

Mapping the spatial evolution of tumor progression at the cellular level

We extended our analysis by spatially mapping gastric fundus nodule biopsy sections using cellular-level 10 μm pixel-sized devices. Collected from the same patient depicted in Figure 4 at the same time, this biopsy showed diffuse large BCL (DLBCL), likely arising from the indolent MALT lymphoma precursor. A diffuse
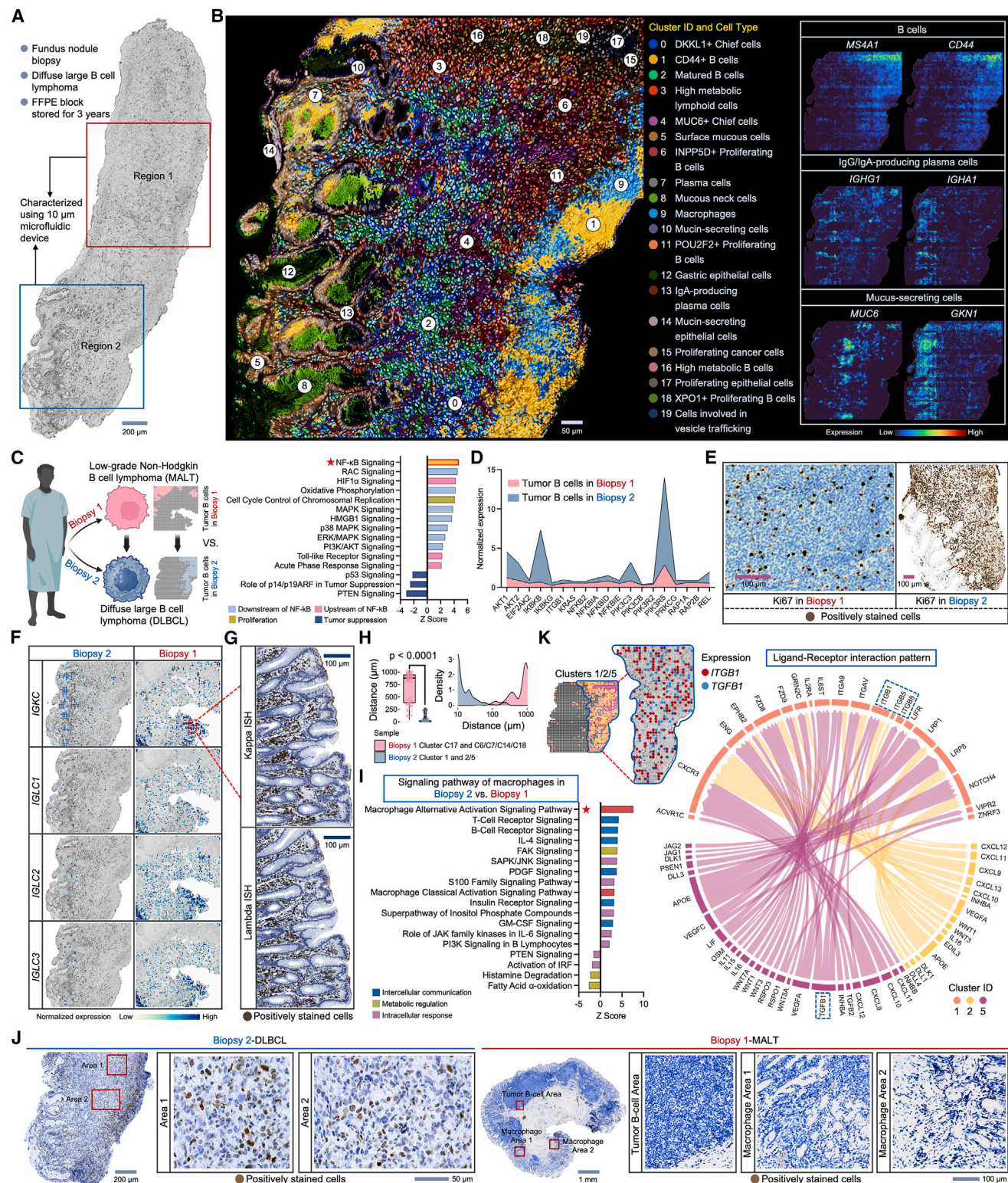
Cell
Resource

Figure 7. Cellular-level spatial mapping of a DLBCL section elucidates tumor progression

(A) Spatial transcriptome mapping of the DLBCL biopsy. Sections from two different regions underwent 10 μm microfluidic barcoding.

(B) Super-resolved high-resolution spatial clustering and representative gene expression enhanced with iStar.

(C) Left: schematic illustration showing comparative analysis. Right: signaling pathways regulated by DEGs between tumor B cells in DLBCL vs. MALT biopsy.

(legend continued on next page)

sheet of atypically large lymphocytes, extending from superficial glandular structures to the deep margin, was observed (Figure S8A). Predominantly large and pleomorphic, the tumor B cells confirmed by IHC staining (Figure S8B) exhibited irregular nuclear contours, dispersed chromatin, and a moderate amount of cytoplasm, with frequent mitoses.

Tissue sections from two distinct regions were selected for spatial barcoding, detecting an exonic average pixel count of 2,292 genes and 6,021 UMIs in region 1, and 1,507 genes and 3,466 UMIs in region 2 (Figures 7A and S8C). Unsupervised clustering of region 1 identified two similar B-cell clusters, distinguished by differential gene expression patterns (Figure S8D). In region 2, a more intricate spatial organization of diverse cell types, including B cells, macrophages, and mucus-secreting cells, was identified (Figures S8E and S8F). This cellular-level mapping revealed a neighborhood comprising various subtypes of gastric mucous cells (clusters 4, 7, and 8) infiltrated with plasma cells (cluster 6), closely mirroring the H&E-defined tissue morphology of an adjacent section (Figure S8E). The spatial clustering of region 2 was further refined through computational enhancement using iStar (Figures 7B and S8G). This analysis notably identified intrinsic heterogeneities among mucus-secreting cells, identifying cluster 4 as MUC6+ chief cells, cluster 5 as surface mucous cells, and cluster 8 as mucous neck cells, fully verified by histopathology.

Thoughtfully selected low-grade MALT and high-grade DLBCL biopsies collected from different anatomical sites at the same time point enabled us to elucidate the spatial molecular dynamics propelling tumor progression. Comparative analysis of gene expression profiles revealed a significant upregulation of NF-κB signaling and its associated upstream and downstream pathways, while tumor suppression pathways, including p53, p14/p19ARF, and PTEN, exhibited notable inhibition (Figure 7C). We observed elevated expression of a range of key genes involved in the NF-κB signaling pathway, particularly *AKT1*, *IKBKB*, *PIK3C3*, and *PIK3R5* (Figure 7D). The considerable NF-κB activation, a hallmark of aggressive DLBCL,[73] could be associated with their significantly active c-Myc (MYC) signaling, spatially concentrated in the tumor B-cell zone (Figure S8H). This was confirmed by pronounced c-Myc IHC staining (Figure S8I). We also found a strong positive correlation between MYC signaling and miR-21 expression (Figure S8J), aligning with the documented role of c-Myc in activating miR-21 by directly binding to its promoter.[74] The coordinated upregulation of these signaling pathways, functional molecules, and their networks contributes to the high proliferative index in the DLBCL

section, indicated by its high Ki67 activity (~70%) compared with the MALT section (<10%) (Figure 7E).

The detection of B-cell clonality has proven valuable for diagnosing BCLs.[75] In the MALT biopsy, a comparable level of the kappa chain constant domain gene *IGKC* and lambda chain genes (*IGLC1*, *IGLC2*, *IGLC3*) was observed and confirmed by ISH stain for kappa and lambda mRNA (Figures 7F and 7G). The finding of polytypic light chains in this biopsy is likely due to the indolent stage, where no clone dominates yet and relatively less plasmacytic differentiation. However, in the progressed DLBCL biopsy, the plasma cells were mostly kappa type, represented by a dominant expression of *IGKC* (Figure 7F), which could be associated with greater differentiation status. Consistently, we observed a significant upregulation of pathways linking inflammation to cancer and involved in tumor cell survival and malignant progression, alongside a downregulation of principal tumor suppression pathways in the DLBCL plasma cells compared with those in the MALT counterpart (Figures S8K and S8L).

Finally, we investigated the molecular profile of macrophages and their spatial interaction with tumor cells. In the MALT section, macrophages in cluster C17 remained distantly positioned from the tumor zone (Figure 4A). By contrast, a profound infiltration of macrophages (cluster 1) into the tumor region (clusters 2/5) was observed in the DLBCL section (Figure S8E), resulting in a significantly reduced macrophage-tumor distance in the latter (Figure 7H). The DEGs between macrophages in DLBCL and MALT regulated a significant activation of the macrophage alternative activation signaling pathway (Figures 7I and S8M). Associated pathways involved in intercellular communications, metabolic modulations, and intracellular responses were also affected. To validate this finding, we performed IHC staining for CD206, a widely used marker for identifying alternative macrophages,[76] and observed positively stained cells with a distribution largely consistent with the transcriptomic data (Figure 7J). By contrast, no signal was detected in either the macrophage-enriched regions or the tumor B-cell region in the MALT biopsy. The deeper exploration of ligand-receptor interactions shed light on the communication network between macrophages and B cells. This analysis unveiled a plethora of crucial pairs activated within the DLBCL microenvironment (Figure 7K), highlighting the potential role of *APOE-LRP1* pair in metabolic shaping of the conversion from classic M1 to alternative M2 macrophage,[77] *CXCL9-CXCR3* axis in orchestrating the recruitment of effector T cells,[78] and *VEGFC-IL6ST* pair in fostering tumor lymphangiogenesis.[79] A noteworthy interaction was the crosstalk between

(D) Expression comparison of key genes involved in the NF-κB signaling between DLBCL vs. MALT biopsy.

(E) IHC staining for Ki67 on adjacent sections from the two biopsies.

(F) Spatial expression mapping of genes encoding plasma cell kappa and lambda chains in the two biopsies.

(G) ISH staining for kappa and lambda chain mRNA in the designated area in (F).

(H) Distance distribution between macrophages and tumor B cells in the two biopsies. Significance level was calculated with two-tailed Mann-Whitney test, **** $p < 0.0001$.

(I) Signaling pathways regulated by DEGs between macrophages in DLBCL vs. MALT biopsy.

(J) IHC staining for CD206 on an adjacent section from the DLBCL and MALT biopsy.

(K) Ligand-receptor interactions between macrophage cluster 1 and tumor B-cell clusters 2 and 5. The distinctive communication pattern between *TGFB1* and the integrin family genes is indicated and spatially visualized.

In (C) and (I), Z score is computed and used to reflect the predicted activation level ($z > 0$, activated; $z < 0$, inhibited; $z \geq 2$ or $z \leq -2$ can be considered significant). See also Figure S8.

TGFB1 and the integrin family genes, potentially instigating the formation of pro-tumorigenic M2-macrophages.[80] The spatial expression pattern of this interaction can be distinctly visualized. Collectively, Patho-DBiT empowers us to spatially map the molecular evolution from low-grade to high-grade tumor at single-cell level, deepening our understanding of the complex interplays shaping the tumor microenvironment in transformed large BCLs.

DISCUSSION

Herein, we developed Patho-DBiT, a powerful technology enabling spatial exploration of rich RNA biology in FFPE tissues. Combining polyadenylation with spatial deterministic barcoding empowered high-yield, genome-wide detection of various RNA species, notably enhancing the capture of small non-coding RNAs lacking poly(A) tails. We demonstrated this ability by spatially profiling the expression of microRNAs. Patho-DBiT not only aligns reads to pre-microRNA regions but also accurately demarcates the boundaries of known dominant mature microRNA strands produced from the pre-microRNA hairpins. Interestingly, our own head-to-head comparative evaluation revealed that microRNA profiling performance was less satisfactory in frozen embryos. The reason why small RNAs such as microRNAs are well preserved in aged FFPE blocks is yet to be further investigated, but it may be due to the formalin fixing step that protects RNA-binding proteins, such as the Argonaute family, which bind to mature microRNAs. Beyond profiling gene expression, Patho-DBiT enables spatial profiling of post-transcriptional RNA processing, including alternative splicing and A-to-I RNA editing. The spatial patterns were cross-validated with a dataset from frozen mouse brain sections using long-read sequencing,[29] confirming the accurate molecular detection by Patho-DBiT rather than modifications arising from the FFPE preparation process.

Pathology has evolved significantly with technological advances, moving from microscopic observation to sophisticated genomic analysis. Patho-DBiT enabled high-sensitivity spatial transcriptomics profiling of a 5 year archival AITL sample, facilitating a faithful reconstruction of spatial cellular compositions and the identification of AITL-specific molecular pathways. In the characterization of a 3 year archival MALT tissue, Patho-DBiT showcased its prowess in extracting a wealth of information from a single run, ranging from exonic mRNA expression, genome-scale SNVs, microRNA regulatory networks, and splicing dynamics, with rigorous validation of accuracies using traditional staining methods including H&E, IHC, IF, ISH, and DNA data from WGS. The result is a finely crafted spatial molecular landscape that serves as a powerful resource for pathologists to disentangle the intricate process underlying MALT tumorigenesis.

Patho-DBiT holds the potential to revolutionize the analysis of human pathology by revealing vulnerabilities in the structural supports of tumors. As demonstrated in this study of indolent MALT and aggressive DLBCL cases, the same patient can harbor similar B-cell tumor cells that take advantage of different microenvironments to survive and proliferate. Current clinical diagnostic practice incorporates only bulk sequencing, if at all, coupled with histopathology, IHC, and a few routine fluorescence *in situ* hybridization studies. *In situ* RNA profiling gives insight into cellular neighborhoods and their interactions. New tools such as Patho-DBiT may allow for a deeper genomic dive that can potentially help clinicians locate the specific cancer cell genotypes or clonotypes implicated in disease evolution and drug resistance. Further inclusion of spatial RNA profiling will enable us to dissect biological pathways, predict which clones may evade drug treatment, identify patient-specific combination therapies, and strategically deploy different immune cells for immunotherapy.[81] For example, Figure S8K shows that *ROCK1* is the most upregulated gene in aggressive lymphoma compared with the indolent form in the same patient's plasma cells. While drugs like belumosudil,[82] a ROCK1/2 inhibitor, are being tested in plasma cell neoplasms, our data may suggest the off-label use of belumosudil as a potential precision medicine approach to treat this patient.

Limitations of the study

Compared with state-of-the-art technologies in sequencing-based spatial transcriptomics,[83–86] Patho-DBiT alone lacks single-cell resolution. Recognizing the value of histological images in identifying spatial patterns and the latest advances in machine learning for deep data fusion, we applied iStar to derive super-resolved single-cell-level transcriptome maps. Future advancements in device design and fabrication may further improve mapping resolution without compromising mapping area. Additionally, our alternative splicing analysis faces low coverage at the pixel level. Recently, Spatial VDJ technology has mapped full-length B-cell and T-cell receptor sequences in frozen human tissues.[87] To enhance spatial quantification of antigen receptor sequences, hybridization capture probes were directed at the targeted regions to enrich transcripts of interest. We anticipate that incorporating this strategy into Patho-DBiT could improve the detection and quantification of transcript isoforms. To harness emerging genomic knowledge and bring it to bear on current pathology diagnostics, newly developed tools like Patho-DBiT must undergo thorough testing and validation on patient samples. FFPE specimens stored for up to 32 years have proven to be a promising source of RNA for gene expression quantification.[88] Given Patho-DBiT's success in appending poly(A) tails to small microRNAs and yielding accurate spatial profiles, we envision its potential for spatial analysis of FFPE patient samples with ultralong storage times.

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Rong Fan (rong.fan@yale.edu).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- Raw and processed sequencing data for this study can be accessed in the NCBI Gene Expression Omnibus (GEO) database under the accession number GSE274641.
- All original code has been deposited at our GitHub space (https://github.com/Zhiliang-Bai/Patho-DBiT).

CellPress
OPEN ACCESS

Cell
Resource

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

AUTHOR CONTRIBUTIONS

DECLARATION OF INTERESTS

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Patient specimens
  - Surgical pathology report of the lymphoma biopsies
  - Mouse paraffin tissues
  - Sample handling and section preparation
- METHOD DETAILS
  - Fabrication of microfluidic device
  - DNA barcodes annealing
  - Tissue deparaffinization and decrosslinking
  - Permeabilization, in situ polyadenylation, and reverse transcription
  - Spatial barcoding with microfluidic devices
  - Tissue lysis and cDNA extraction
  - cDNA purification, template switch, and PCR amplification
  - rRNA removal, library preparation, and sequencing
  - Generation of miR-142 knockout cells and spatial patterning
  - CODEX spatial phenotyping using PhenoCycler-Fusion
  - H&E, immunohistochemistry (IHC) and in situ hybridization (ISH)
  - Immunofluorescence staining (IF)
  - Genomic DNA extraction from FFPE sections for whole genome sequencing (WGS)
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Sequence alignment and generation of mRNA expression matrix

- Read mapping of non-coding RNA species
- Gene data normalization and unsupervised clustering analysis
- Integration with scRNA-seq datasets
- qPCR analysis of rRNA removal efficiency
- Gene body coverage calculation
- Spatial microRNA alignment and analysis
- Spatial alternative splicing analysis
- Spatial adenosine-to-inosine (A-to-I) RNA editing analysis
- Inferring high-resolution tissue architecture using iStar
- Spatial RNA splicing dynamics
- Coverage comparison with 10x Genomics datasets
- Spatial single nucleotide variant (SNV) analysis
- WGS data alignment and analysis
- Evolutionary relationship construction
- Copy number variation (CNV) analysis
- Ligand-receptor interaction analysis
- Ingenuity Pathway Analysis
- Statistical analysis

SUPPLEMENTAL INFORMATION

REFERENCES

1. Baysoy, A., Bai, Z., Satija, R., and Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. Nat. Rev. Mol. Cell Biol. *24*, 695–713. https://doi.org/10.1038/s41580-023-00615-w.

2. Bressan, D., Battistoni, G., and Hannon, G.J. (2023). The dawn of spatial omics. Science *381*, eabq4964. https://doi.org/10.1126/science.abq4964.

3. Deng, Y., Bai, Z., and Fan, R. (2023). Microtechnologies for single-cell and spatial multi-omics. Nat. Rev. Bioeng. *1*, 769–784. https://doi.org/10.1038/s44222-023-00084-y.

4. Chen, J., Larsson, L., Swarbrick, A., and Lundeberg, J. (2024). Spatial landscapes of cancers: insights and opportunities. Nat. Rev. Clin. Oncol. *21*, 660–674. https://doi.org/10.1038/s41571-024-00926-7.

5. Harries, L.W. (2019). RNA Biology Provides New Therapeutic Targets for Human Disease. Front. Genet. *10*, 205. https://doi.org/10.3389/fgene.2019.00205.

6. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.

7. Zeng, H., Huang, J., Ren, J., Wang, C.K., Tang, Z., Zhou, H., Zhou, Y., Shi, H., Aditham, A., Sui, X., et al. (2023). Spatially resolved single-cell translatomics at molecular resolution. Science *380*, eadd3067. https://doi.org/10.1126/science.add3067.

8. Fan, R. (2023). Mapping RNA translation. Science *380*, 1321–1322. https://doi.org/10.1126/science.adi6844.

9. Blow, N. (2007). Tissue preparation: Tissue issues. Nature *448*, 959–963. https://doi.org/10.1038/448959a.

10. Chin, L., and Gray, J.W. (2008). Translating insights from the cancer genome into clinical practice. Nature *452*, 553–563. https://doi.org/10.1038/nature06914.

11. Jang, M.J., Coughlin, G.M., Jackson, C.R., Chen, X., Chuapoco, M.R., Vendemiatti, J.L., Wang, A.Z., and Gradinaru, V. (2023). Spatial transcriptomics for profiling the tropism of viral vectors in tissues. Nat. Biotechnol. *41*, 1272–1286. https://doi.org/10.1038/s41587-022-01648-w.

Cell
Resource

12. McKellar, D.W., Mantri, M., Hinchman, M.M., Parker, J.S.L., Sethupathy, P., Cosgrove, B.D., and De Vlaminck, I. (2023). Spatial mapping of the total transcriptome by in situ polyadenylation. Nat. Biotechnol. *41*, 513–520. https://doi.org/10.1038/s41587-022-01517-6.

13. Liu, Y., Yang, M., Deng, Y., Su, G., Enninful, A., Guo, C.C., Tebaldi, T., Zhang, D., Kim, D., Bai, Z., et al. (2020). High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. Cell *183*, 1665–1681.e18. https://doi.org/10.1016/j.cell.2020.10.026.

14. Su, G., Qin, X., Enninful, A., Bai, Z., Deng, Y., Liu, Y., and Fan, R. (2021). Spatial multi-omics sequencing for fixed tissue via DBiT-seq. Star Protoc. *2*, 100532. https://doi.org/10.1016/j.xpro.2021.100532.

15. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.

16. Fafard-Couture, É., Bergeron, D., Couture, S., Abou-Elela, S., and Scott, M.S. (2021). Annotation of snoRNA abundance across human tissues reveals complex snoRNA-host gene relationships. Genome Biol. *22*, 172. https://doi.org/10.1186/s13059-021-02391-2.

17. Yu, P., Zhou, S., Gao, Y., Liang, Y., Guo, W., Wang, D.O., Ding, S., Lin, S., Wang, J., and Cun, Y. (2023). Dynamic Landscapes of tRNA Transcriptomes and Translatomes in Diverse Mouse Tissues. Genomics Proteomics Bioinformatics *21*, 834–849. https://doi.org/10.1016/j.gpb.2022.07.006.

18. O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front. Endocrinol. *9*, 402. https://doi.org/10.3389/fendo.2018.00402.

19. Jopling, C. (2012). Liver-specific microRNA-122: Biogenesis and function. RNA Biol. *9*, 137–142. https://doi.org/10.4161/rna.18827.

20. Muñoz-Castañeda, R., Zingg, B., Matho, K.S., Chen, X., Wang, Q., Foster, N.N., Li, A., Narasimhan, A., Hirokawa, K.E., Huo, B., et al. (2021). Cellular anatomy of the mouse primary motor cortex. Nature *598*, 159–166. https://doi.org/10.1038/s41586-021-03970-w.

21. Yao, Z., van Velthoven, C.T.J., Nguyen, T.N., Goldy, J., Sedeno-Cortes, A.E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., et al. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell *184*, 3222–3241.e26. https://doi.org/10.1016/j.cell.2021.04.021.

22. Raj, B., and Blencowe, B.J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. Neuron *87*, 14–27. https://doi.org/10.1016/j.neuron.2015.05.004.

23. Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. USA *111*, E5593–E5601. https://doi.org/10.1073/pnas.1419161111.

24. Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The Expanding Landscape of Alternative Splicing Variation in Human Populations. Am. J. Hum. Genet. *102*, 11–26. https://doi.org/10.1016/j.ajhg.2017.11.002.

25. Gao, Y., Wang, F., Wang, R., Kutschera, E., Xu, Y., Xie, S., Wang, Y., Kadash-Edmondson, K.E., Lin, L., and Xing, Y. (2023). ESPRESSO: Robust discovery and quantification of transcript isoforms from error-prone long-read RNA-seq data. Sci. Adv. *9*, eabq5072. https://doi.org/10.1126/sciadv.abq5072.

26. Kneussel, M., and Wagner, W. (2013). Myosin motors at neuronal synapses: drivers of membrane transport and actin dynamics. Nat. Rev. Neurosci. *14*, 233–247. https://doi.org/10.1038/nrn3445.

27. Pandey, R.S., Kotredes, K.P., Sasner, M., Howell, G.R., and Carter, G.W. (2023). Differential splicing of neuronal genes in a Trem2*R47H mouse model mimics alterations associated with Alzheimer's disease. BMC Genomics *24*, 172. https://doi.org/10.1186/s12864-023-09280-x.

28. Weston, M. (2018). Trading up to a New Model of STXBP-Encephalopathy. Epilepsy Curr. *18*, 257–259. https://doi.org/10.5698/1535-7597.18.4.257.

29. Lebrigand, K., Bergenstråhle, J., Thrane, K., Mollbrink, A., Meletis, K., Barbry, P., Waldmann, R., and Lundeberg, J. (2023). The spatial landscape of gene expression isoforms in tissue sections. Nucleic Acids Res. *51*, e47. https://doi.org/10.1093/nar/gkad169.

30. Licht, K., Kapoor, U., Amman, F., Picardi, E., Martin, D., Bajad, P., and Jantsch, M.F. (2019). A high resolution A-to-I editing map in the mouse identifies editing events controlled by pre-mRNA splicing. Genome Res. *29*, 1453–1463. https://doi.org/10.1101/gr.242636.118.

31. Waterhouse, R.M., Aganezov, S., Anselmetti, Y., Lee, J., Ruzzante, L., Reijnders, M.J.M.F., Feron, R., Bérard, S., George, P., Hahn, M.W., et al. (2020). Evolutionary superscaffolding and chromosome anchoring to improve Anopheles genome assemblies. BMC Biol. *18*, 1. https://doi.org/10.1186/s12915-019-0728-3.

32. Yang, Y., Okada, S., and Sakurai, M. (2021). Adenosine-to-inosine RNA editing in neurological development and disease. RNA Biol. *18*, 999–1013. https://doi.org/10.1080/15476286.2020.1867797.

33. Black, S., Phillips, D., Hickey, J.W., Kennedy-Darling, J., Venkataraaman, V.G., Samusik, N., Goltsev, Y., Schürch, C.M., and Nolan, G.P. (2021). CODEX multiplexed tissue imaging with DNA-conjugated antibodies. Nat. Protoc. *16*, 3802–3835. https://doi.org/10.1038/s41596-021-00556-8.

34. Cortes, J.R., Ambesi-Impiombato, A., Couronné, L., Quinn, S.A., Kim, C.S., da Silva Almeida, A.C., West, Z., Belver, L., Martin, M.S., Scourzic, L., et al. (2018). RHOA G17V Induces T Follicular Helper Cell Specification and Promotes Lymphomagenesis. Cancer Cell *33*, 259–273.e7. https://doi.org/10.1016/j.ccell.2018.01.001.

35. Lunning, M.A., and Vose, J.M. (2017). Angioimmunoblastic T-cell lymphoma: the many-faced lymphoma. Blood *129*, 1095–1102. https://doi.org/10.1182/blood-2016-09-692541.

36. Jain, S., Mallick, S., Ramteke, P., and Gogia, A. (2022). Neoplasms of follicular helper T-cells: an insight into the pathobiology. Am. J. Blood Res. *12*, 64–81.

37. Voena, C., and Chiarle, R. (2019). RHO Family GTPases in the Biology of Lymphoma. Cells *8*, 646. https://doi.org/10.3390/cells8070646.

38. Zucca, E., and Bertoni, F. (2016). The spectrum of MALT lymphoma at different sites: biological and therapeutic relevance. Blood *127*, 2082–2092. https://doi.org/10.1182/blood-2015-12-624304.

39. Bergenstråhle, L., He, B., Bergenstråhle, J., Abalo, X., Mirzazadeh, R., Thrane, K., Ji, A.L., Andersson, A., Larsson, L., Stakenborg, N., et al. (2022). Super-resolved spatial transcriptomics by deep data fusion. Nat. Biotechnol. *40*, 476–479. https://doi.org/10.1038/s41587-021-01075-3.

40. Hu, J., Coleman, K., Zhang, D., Lee, E.B., Kadara, H., Wang, L., and Li, M. (2023). Deciphering tumor ecosystems at super resolution from spatial transcriptomics with TESLA. Cell Syst. *14*, 404–417.e4. https://doi.org/10.1016/j.cels.2023.03.008.

41. Zhang, D., Schroeder, A., Yan, H., Yang, H., Hu, J., Lee, M.Y.Y., Cho, K.S., Susztak, K., Xu, G.X., Feldman, M.D., et al. (2024). Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. Nat. Biotechnol. https://doi.org/10.1038/s41587-023-02019-9.

42. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414. https://doi.org/10.1038/s41587-020-0591-3.

43. Gao, M., Qiao, C., and Huang, Y. (2022). UniTVelo: temporally unified RNA velocity reinforces single-cell trajectory inference. Nat. Commun. *13*, 6586. https://doi.org/10.1038/s41467-022-34188-7.

44. Correia, C., Schneider, P.A., Dai, H., Dogan, A., Maurer, M.J., Church, A.K., Novak, A.J., Feldman, A.L., Wu, X., Ding, H., et al. (2015). BCL2 mutations are associated with increased risk of transformation and shortened survival in follicular lymphoma. Blood *125*, 658–667. https://doi.org/10.1182/blood-2014-04-571786.

45. Jiang, Q., Crews, L.A., Holm, F., and Jamieson, C.H.M. (2017). RNA editing-dependent epitranscriptome diversity in cancer stem cells. Nat. Rev. Cancer 17, 381–392. https://doi.org/10.1038/nrc.2017.23.

46. Baysal, B.E., Sharma, S., Hashemikhabir, S., and Janga, S.C. (2017). RNA Editing in Pathogenesis of Cancer. Cancer Res. 77, 3733–3739. https://doi.org/10.1158/0008-5472.Can-17-0520.

47. Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Armstrong, J., Barnes, I., et al. (2021). GENCODE 2021. Nucleic Acids Res. 49, D916–D923. https://doi.org/10.1093/nar/gkaa1087.

48. Bacon, C.M., Du, M.Q., and Dogan, A. (2007). Mucosa-associated lymphoid tissue (MALT) lymphoma: a practical guide for pathologists. J. Clin. Pathol. 60, 361–372. https://doi.org/10.1136/jcp.2005.031146.

49. Guo, Y., Zhao, S., Sheng, Q., Samuels, D.C., and Shyr, Y. (2017). The discrepancy among single nucleotide variants detected by DNA and RNA high throughput sequencing data. BMC Genomics 18 (Suppl 6), 690. https://doi.org/10.1186/s12864-017-4022-x.

50. Robbe, P., Popitsch, N., Knight, S.J.L., Antoniou, P., Becq, J., He, M., Kanapin, A., Samsonova, A., Vavoulis, D.V., Ross, M.T., et al. (2018). Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. Genet. Med. 20, 1196–1205. https://doi.org/10.1038/gim.2017.241.

51. Venturutti, L., Teater, M., Zhai, A., Chadburn, A., Babiker, L., Kim, D., Béguelin, W., Lee, T.C., Kim, Y., Chin, C.R., et al. (2020). TBL1XR1 Mutations Drive Extranodal Lymphoma by Inducing a Pro-tumorigenic Memory Fate. Cell 182, 297–316.e27. https://doi.org/10.1016/j.cell.2020.05.049.

52. Köchert, K., Ullrich, K., Kreher, S., Aster, J.C., Kitagawa, M., Jöhrens, K., Anagnostopoulos, I., Jundt, F., Lamprecht, B., Zimber-Strobl, U., et al. (2011). High-level expression of Mastermind-like 2 contributes to aberrant activation of the NOTCH signaling pathway in human lymphomas. Oncogene 30, 1831–1840. https://doi.org/10.1038/onc.2010.544.

53. Ott, G., Rosenwald, A., and Campo, E. (2013). Understanding MYC-driven aggressive B-cell lymphomas: pathogenesis and classification. Blood 122, 3884–3891. https://doi.org/10.1182/blood-2013-05-498329.

54. Lewinsky, H., Gunes, E.G., David, K., Radomir, L., Kramer, M.P., Pellegrino, B., Perpinial, M., Chen, J., He, T.-f., Mansour, A.G., et al. (2021). CD84 is a regulator of the immunosuppressive microenvironment in multiple myeloma. JCI Insight 6, e141683. https://doi.org/10.1172/jci.insight.141683.

55. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 344, 1396–1401. https://doi.org/10.1126/science.1254257.

56. Bracken, C.P., Scott, H.S., and Goodall, G.J. (2016). A network-biology perspective of microRNA function and dysfunction in cancer. Nat. Rev. Genet. 17, 719–732. https://doi.org/10.1038/nrg.2016.134.

57. Alles, J., Fehlmann, T., Fischer, U., Backes, C., Galata, V., Minet, M., Hart, M., Abu-Halima, M., Grässer, F.A., Lenhof, H.-P., et al. (2019). An estimate of the total number of true human miRNAs. Nucleic Acids Res. 47, 3353–3364. https://doi.org/10.1093/nar/gkz097.

58. Muhl, L., Mocci, G., Pietilä, R., Liu, J., He, L., Genové, G., Leptidis, S., Gustafsson, S., Buyandelger, B., Raschperger, E., et al. (2022). A single-cell transcriptomic inventory of murine smooth muscle cells. Dev. Cell 57, 2426–2443.e6. https://doi.org/10.1016/j.devcel.2022.09.015.

59. Cordes, K.R., Sheehy, N.T., White, M.P., Berry, E.C., Morton, S.U., Muth, A.N., Lee, T.-H., Miano, J.M., Ivey, K.N., and Srivastava, D. (2009). miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. Nature 460, 705–710. https://doi.org/10.1038/nature08195.

60. Kramer, N.J., Wang, W.-L., Reyes, E.Y., Kumar, B., Chen, C.-C., Ramakrishna, C., Cantin, E.M., Vonderfecht, S.L., Taganov, K.D., Chau, N., and Boldin, M.P. (2015). Altered lymphopoiesis and immunodeficiency in miR-142 null mice. Blood 125, 3720–3730. https://doi.org/10.1182/blood-2014-10-603951.

61. Di Lisio, L., Martinez, N., Montes-Moreno, S., Piris-Villaespesa, M., Sanchez-Beato, M., and Piris, M.A. (2012). The role of miRNAs in the pathogenesis and diagnosis of B-cell lymphomas. Blood 120, 1782–1790. https://doi.org/10.1182/blood-2012-05-402784.

62. Bautista-Sánchez, D., Arriaga-Canon, C., Pedroza-Torres, A., De La Rosa-Velázquez, I.A., González-Barrios, R., Contreras-Espinosa, L., Montiel-Manríquez, R., Castro-Hernández, C., Fragoso-Ontiveros, V., Álvarez-Gómez, R.M., and Herrera, L.A. (2020). The Promising Role of miR-21 as a Cancer Biomarker and Its Importance in RNA-Based Therapeutics. Mol. Ther. Nucleic Acids 20, 409–420. https://doi.org/10.1016/j.omtn.2020.03.003.

63. Pan, J.Y., Zhang, F., Sun, C.C., Li, S.J., Li, G., Gong, F.Y., Bo, T., He, J., Hua, R.X., Hu, W.D., et al. (2017). miR-134: A Human Cancer Suppressor? Mol. Ther. Nucleic Acids 6, 140–149. https://doi.org/10.1016/j.omtn.2016.11.003.

64. He, Y., Yu, D., Zhu, L., Zhong, S., Zhao, J., and Tang, J. (2018). miR-149 in Human Cancer: A Systemic Review. J. Cancer 9, 375–388. https://doi.org/10.7150/jca.21044.

65. Abdelmohsen, K., and Gorospe, M. (2012). RNA-binding protein nucleolin in disease. RNA Biol. 9, 799–808. https://doi.org/10.4161/rna.19718.

66. Najafi, S., Ghafouri-Fard, S., Hussen, B.M., Jamal, H.H., Taheri, M., and Hallajnejad, M. (2021). Oncogenic Roles of Small Nucleolar RNA Host Gene 7 (SNHG7) Long Noncoding RNA in Human Cancers and Potentials. Front. Cell Dev. Biol. 9, 809345. https://doi.org/10.3389/fcell.2021.809345.

67. Abbas, W., Kumar, A., and Herbein, G. (2015). The eEF1A Proteins: At the Crossroads of Oncogenesis, Apoptosis, and Viral Infections. Front. Oncol. 5, 75. https://doi.org/10.3389/fonc.2015.00075.

68. Li, Y., Wang, Y.W., Chen, X., Ma, R.R., Guo, X.Y., Liu, H.T., Jiang, S.J., Wei, J.M., and Gao, P. (2020). MicroRNA-4472 Promotes Tumor Proliferation and Aggressiveness in Breast Cancer by Targeting RGMA and Inducing EMT. Clin. Breast Cancer 20, e113–e126. https://doi.org/10.1016/j.clbc.2019.08.010.

69. Krämer, A., Green, J., Pollard, J., Jr., and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 30, 523–530. https://doi.org/10.1093/bioinformatics/btt703.

70. Due, H., Svendsen, P., Bødker, J.S., Schmitz, A., Bøgsted, M., Johnsen, H.E., El-Galaly, T.C., Roug, A.S., and Dybkær, K. (2016). miR-155 as a Biomarker in B-Cell Malignancies. BioMed Res. Int. 2016, 9513037. https://doi.org/10.1155/2016/9513037.

71. Ma, X., Becker Buscaglia, L.E., Barker, J.R., and Li, Y. (2011). MicroRNAs in NF-κB signaling. J. Mol. Cell Biol. 3, 159–166. https://doi.org/10.1093/jmcb/mjr007.

72. Due, H., Schönherz, A.A., Ryø, L., Primo, M.N., Jespersen, D.S., Thomsen, E.A., Roug, A.S., Xiao, M., Tan, X., Pang, Y., et al. (2019). MicroRNA-155 controls vincristine sensitivity and predicts superior clinical outcome in diffuse large B-cell lymphoma. Blood Adv. 3, 1185–1196. https://doi.org/10.1182/bloodadvances.2018029660.

73. David, A., Arnaud, N., Fradet, M., Lascaux, H., Ouk-Martin, C., Gachard, N., Zimber-Strobl, U., Feuillard, J., and Faumont, N. (2017). c-Myc dysregulation is a co-transforming event for nuclear factor-κB activated B cells. Haematologica 102, 883–894. https://doi.org/10.3324/haematol.2016.156281.

74. Li, S., He, X., Gan, Y., Zhang, J., Gao, F., Lin, L., Qiu, X., Yu, T., Zhang, X., Chen, P., et al. (2021). Targeting miR-21 with NL101 blocks c-Myc/Mxd1 loop and inhibits the growth of B cell lymphoma. Theranostics 11, 3439–3451. https://doi.org/10.7150/thno.53561.

75. Rimsza, L.M., Day, W.A., McGinn, S., Pedata, A., Natkunam, Y., Warnke, R., Cook, J.R., Marafioti, T., and Grogan, T.M. (2014). Kappa and lambda light chain mRNA in situ hybridization compared to flow cytometry and

Cell
Resource

CellPress
OPEN ACCESS

immunohistochemistry in B cell lymphomas. Diagn. Pathol. *9*, 144. https://doi.org/10.1186/1746-1596-9-144.

76. Modak, M., Mattes, A.K., Reiss, D., Skronska-Wasek, W., Langlois, R., Sabarth, N., Konopitzky, R., Ramirez, F., Lehr, K., Mayr, T., et al. (2022). CD206+ tumor-associated macrophages cross-present tumor antigen and drive antitumor immunity. JCI Insight *7*, e155022. https://doi.org/10.1172/jci.insight.155022.

77. May, P., Bock, H.H., and Nofer, J.R. (2013). Low density receptor-related protein 1 (LRP1) promotes anti-inflammatory phenotype in murine macrophages. Cell Tissue Res. *354*, 887–889. https://doi.org/10.1007/s00441-013-1699-2.

78. Marcovecchio, P.M., Thomas, G., and Salek-Ardakani, S. (2021). CXCL9-expressing tumor-associated macrophages: new players in the fight against cancer. J. Immunother. Cancer *9*, e002045. https://doi.org/10.1136/jitc-2020-002045.

79. Huang, Y.H., Yang, H.Y., Huang, S.W., Ou, G., Hsu, Y.F., and Hsu, M.J. (2016). Interleukin-6 Induces Vascular Endothelial Growth Factor-C Expression via Src-FAK-STAT3 Signaling in Lymphatic Endothelial Cells. PLoS One *11*, e0158839. https://doi.org/10.1371/journal.pone.0158839.

80. Brown, N.F., and Marshall, J.F. (2019). Integrin-Mediated TGFβ Activation Modulates the Tumour Microenvironment. Cancers *11*, 1221. https://doi.org/10.3390/cancers11091221.

81. Bai, Z., Su, G., and Fan, R. (2021). Single-Cell Analysis Technologies for Immuno-Oncology Research: From Mechanistic Delineation to Biomarker Discovery. Genomics Proteomics Bioinformatics *19*, 191–207. https://doi.org/10.1016/j.gpb.2021.02.004.

82. Barcelo, J., Samain, R., and Sanz-Moreno, V. (2023). Preclinical to clinical utility of ROCK inhibitors in cancer. Trends Cancer *9*, 250–263. https://doi.org/10.1016/j.trecan.2022.12.001.

83. Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Äijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J.F., et al. (2019). High-definition spatial transcriptomics for in situ tissue profiling. Nat. Methods *16*, 987–990. https://doi.org/10.1038/s41592-019-0548-y.

84. Cho, C.-S., Xi, J., Si, Y., Park, S.-R., Hsu, J.-E., Kim, M., Jun, G., Kang, H.M., and Lee, J.H. (2021). Microscopic examination of spatial transcriptome using Seq-Scope. Cell *184*, 3559–3572.e22. https://doi.org/10.1016/j.cell.2021.05.010.

85. Fu, X., Sun, L., Dong, R., Chen, J.Y., Silakit, R., Condon, L.F., Lin, Y., Lin, S., Palmiter, R.D., and Gu, L. (2022). Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. Cell *185*, 4621–4633.e17. https://doi.org/10.1016/j.cell.2022.10.021.

86. Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., et al. (2022). Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell *185*, 1777–1792.e21. https://doi.org/10.1016/j.cell.2022.04.003.

87. Engblom, C., Thrane, K., Lin, Q., Andersson, A., Toosi, H., Chen, X., Steiner, E., Lu, C., Mantovani, G., Hagemann-Jensen, M., et al. (2023). Spatial transcriptomics of B cell and T cell receptors reveals lymphocyte clonal dynamics. Science *382*, eadf8486. https://doi.org/10.1126/science.adf8486.

88. Zhao, Y., Mehta, M., Walton, A., Talsania, K., Levin, Y., Shetty, J., Gillanders, E.M., Tran, B., and Carrick, D.M. (2019). Robustness of RNA sequencing on older formalin-fixed paraffin-embedded tissue from high-grade ovarian serous adenocarcinomas. PLoS One *14*, e0216050. https://doi.org/10.1371/journal.pone.0216050.

89. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open source software for digital pathology image analysis. Sci. Rep. *7*, 16878. https://doi.org/10.1038/s41598-017-17204-5.

90. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. *17*, 3. https://doi.org/10.14806/ej.17.1.200.

91. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

92. Navarro, J.F., Sjöstrand, J., Salmén, F., Lundeberg, J., and Ståhl, P.L. (2017). ST Pipeline: an automated pipeline for spatial mapping of unique transcripts. Bioinformatics *33*, 2591–2593. https://doi.org/10.1093/bioinformatics/btx211.

93. Chan, P.P., and Lowe, T.M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res. *44*, D184–D189. https://doi.org/10.1093/nar/gkv1309.

94. The RNAcentral Consortium, Petrov, A.I., Kay, S.J.E., Kalvari, I., Howe, K.L., Gray, K.A., Bruford, E.A., Kersey, P.J., Cochrane, G., Finn, R.D., et al. (2017). RNAcentral: a comprehensive database of non-coding RNA sequences. Nucleic Acids Res. *45*, D128–D134. https://doi.org/10.1093/nar/gkw1008.

95. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

96. Wang, L., Nie, J., Sicotte, H., Li, Y., Eckel-Passow, J.E., Dasari, S., Vedell, P.T., Barman, P., Wang, L., Weinshiboum, R., et al. (2016). Measure transcript integrity using RNA-seq data. BMC Bioinformatics *17*, 58. https://doi.org/10.1186/s12859-016-0922-z.

97. Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Res. *47*, D155–D162. https://doi.org/10.1093/nar/gky1141.

98. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

99. Picardi, E., D'Erchia, A.M., Lo Giudice, C., and Pesole, G. (2017). REDIportal: a comprehensive database of A-to-I RNA editing events in humans. Nucleic Acids Res. *45*, D750–D757. https://doi.org/10.1093/nar/gkw767.

100. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

101. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

102. Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., and Saunders, C.T. (2018). Strelka2: fast and accurate calling of germline and somatic variants. Nat. Methods *15*, 591–594. https://doi.org/10.1038/s41592-018-0051-x.

103. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

104. Xu, S., Li, L., Luo, X., Chen, M., Tang, W., Zhan, L., Dai, Z., Lam, T.T., Guan, Y., and Yu, G. (2022). Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. Imeta *1*, e56. https://doi.org/10.1002/imt2.56.

105. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comp. Biol. *12*, e1004873. https://doi.org/10.1371/journal.pcbi.1004873.

106. Raredon, M.S.B., Adams, T.S., Suhail, Y., Schupp, J.C., Poli, S., Neumark, N., Leiby, K.L., Greaney, A.M., Yuan, Y., Horien, C., et al. (2019).

Single-cell connectomic analysis of adult mammalian lungs. Sci. Adv. *5*, eaaw3851. https://doi.org/10.1126/sciadv.aaw3851.

107. Cheng, J., Roden, C.A., Pan, W., Zhu, S., Baccei, A., Pan, X., Jiang, T., Kluger, Y., Weissman, S.M., Guo, S., et al. (2016). A Molecular Chipper technology for CRISPR sgRNA library generation and functional mapping of noncoding regions. Nat. Commun. *7*, 11178. https://doi.org/10.1038/ncomms11178.

108. Adams, B.D., Guo, S., Bai, H., Guo, Y., Megyola, C.M., Cheng, J., Heydari, K., Xiao, C., Reddy, E.P., and Lu, J. (2012). An in vivo functional screen uncovers miR-150-mediated regulation of hematopoietic injury response. Cell Rep. *2*, 1048–1060. https://doi.org/10.1016/j.celrep.2012.09.014.

109. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. *20*, 296. https://doi.org/10.1186/s13059-019-1874-1.

110. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

111. Tao, B., Kraehling, J.R., Ghaffari, S., Ramirez, C.M., Lee, S., Fowler, J.W., Lee, W.L., Fernandez-Hernando, C., Eichmann, A., and Sessa, W.C. (2020). BMP-9 and LDL crosstalk regulates ALK-1 endocytosis and LDL transcytosis in endothelial cells. J. Biol. Chem. *295*, 18179–18188. https://doi.org/10.1074/jbc.RA120.015680.

112. Robert, C., and Watson, M. (2015). Errors in RNA-Seq quantification affect genes of relevance to human disease. Genome Biol. *16*, 177. https://doi.org/10.1186/s13059-015-0734-x.

Cell
Resource

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Purified anti-human CD68 Recombinant Antibody	Biolegend	Cat#118212; RRID: AB_2922613
Alexa Fluor 594 anti-human CD138 (Syndecan-1) Antibody	Biolegend	Cat#352323; RRID: AB_2728336
Alexa Fluor 647 anti-human CD20 Antibody	Biolegend	Cat#382803; RRID: AB_3068232
CODEX antibodies	Akoya Biosciences	See Table S4
Biological samples		
Mouse C57 Embryo E13 Paraffin Tissue Sections	Zyagen	MP-104-13-C57
Mouse C57 Brain Reg.9 Paraffin Tissue Sections	Zyagen	MP-201-09-C57
Mouse C57 Brain Reg.9 Paraffin Tissue Sections	Zyagen	MP-201-09-C57
Mouse C57 Lymph nodes Paraffin Sections	Zyagen	MP-703-C57
Human lymphoma FFPE tissue blocks	Yale Pathology Tissue Services	https://medicine.yale.edu/pathology/research/ypts/
Chemicals, peptides, and recombinant proteins		
Acetone	Sigma-Aldrich	179124-500ML
SU-8 photoresist	MicroChem	SU-8 2010 or 2025
Chlorotrimethylsilane	Sigma-Aldrich	89595-10X1ML
Polydimethylsiloxane (PDMS)	Dow Silicones Corporation	GMID: 02065622
Tris-HCl pH 8.0	Sigma-Aldrich	T2694-100ML
5M NaCl	Thermo Fisher Scientific	AM9759
0.5M EDTA	Life technologies	15575-038
Xylene	Sigma-Aldrich	247642-4L-CB
Ethanol	Sigma-Aldrich	E7023-500ML
Antigen retrieval buffer	abcam	ab93684
Triton X-100	Sigma-Aldrich	T8787-50ML
DPBS	Gibco	14190-144
Nuclease-free water	Thermo Fisher Scientific	AM9932
RNase Inhibitor	Qiagen	Y9240L
E. coli Poly(A) Polymerase	New England Biolabs	M0276L
SUPERase·In RNase Inhibitor	Thermo Fisher Scientific	AM2694
Maxima H Minus Reverse Transcriptase	Thermo Fisher Scientific	EP0753
dNTP Mix	Thermo Fisher Scientific	R0192
NEBuffer 3.1	New England Biolabs	B6003S
T4 DNA ligase	New England Biolabs	M0202L
SDS	Sigma-Aldrich	71736-500ML
Proteinase K, recombinant, PCR grade	Thermo Fisher Scientific	EO0491
Phenylmethylsulfonyl fluoride (PMSF)	Sigma-Aldrich	10837091001
Dynabead MyOne Streptavidin C1	Thermo Fisher Scientific	65001
Tris-HCl pH 7.5	Sigma-Aldrich	T2694-100ML
Tween-20	AmericanBio	AB02038-00500
20% Ficoll PM-400	Sigma-Aldrich	F5415-25ML
2X KAPA HiFi HotStart ReadyMix	KAPA Biosystems	KK2602
20X EvaGreen Plus Dye	VWR International	76423-470
SPRIselect beads	Beckman Coulter	B23318
Formaldehyde	Thermo Fisher Scientific	BP53125

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bovine serum albumin	Sigma-Aldrich	A7030
DAPI Solution	Thermo Fisher Scientific	62248
Mm_Gapdh_3_SG QuantiTect Primer Assay	Qiagen	QT01658692
Mm_Actb_2_SG QuantiTect Primer Assay	Qiagen	QT01136772
Fetal Bovine Serum	Thermo Fisher Scientific	10438026
Penicillin-Streptomycin-Glutamine (100X)	Thermo Fisher Scientific	10378016
Recombinant Mouse IL-3	Biolegend	575504
Critical commercial assays		
D5000 ScreenTape Assay	Agilent Technologies	50675588 and 5589
SEQuoia RiboDepletion Kit	BIO-RAD	17006487
DNeasy Blood & Tissue Kit	Qiagen	69504
Deposited data		
Raw and processed sequencing data	This paper	GSE274641
Original code	This paper	https://github.com/Zhiliang-Bai/Patho-DBiT
Oligonucleotides		
Primers, ligation linkers, and DNA barcodes	IDT	See Table S3
Primers targeting Rrna	IDT	See Table S5
Software and algorithms		
QuPath V0.5.0	Bankhead et al.[89]	https://qupath.github.io/
Cutadapt V3.4	Martin[90]	https://cutadapt.readthedocs.io/en/v3.4/
STAR V2.7.7a	Dobin et al.[91]	https://github.com/alexdobin/STAR/releases/tag/2.7.7a
ST_Pipeline V1.7.6	Navarro et al.[92]	https://github.com/SpatialTranscriptomicsResearch/st_pipeline
GENCODE	Frankish et al.[47]	https://www.gencodegenes.org/
GtRNAdb	Chan et al.[93]	https://gtrnadb.ucsc.edu/
RNAcentral	Petrov et al.[94]	https://rnacentral.org/
RStudio V2023.06.2+561	RStudio	https://posit.co/download/rstudio-desktop/
R V4.3.1	CRAN	https://www.r-project.org/
Seurat V4	Hao et al.[95]	https://satijalab.org/seurat/
RSeQC package V5.0.1	Wang et al.[96]	https://rseqc.sourceforge.net/
miRBase	Kozomara et al.[97]	https://www.mirbase.org/
Integrative Genomics Viewer (IGV)	Robinson et al.[98]	https://igv.org/
rMATS-turbo pipeline V4.1.2	Shen et al.[23]	https://github.com/Xinglab/rmats-turbo
REDIportal V2.0	Picardi et al.[99]	http://srv00.recas.ba.infn.it/atlas/
samtools V1.16.1	Danecek et al.[100]	https://github.com/samtools/samtools
iStar	Zhang et al.[41]	https://github.com/daviddaiweizhang/istar
bedtools V 2.31.0	Quinlan et al.[101]	https://github.com/arq5x/bedtools2
scVelo	Bergen et al.[42]	https://scvelo.readthedocs.io/en/stable/
Strelka V2.9.10	Kim et al.[102]	https://github.com/Illumina/strelka
BWA V0.7.17	Li et al.[103]	https://github.com/lh3/bwa
GATK V4.2.0.0	Data Sciences Platform at the Broad Institute	https://gatk.broadinstitute.org/hc/en-us/sections/4405443482011-4-2-2-0
ggtree V3.8.2	Xu et al.[104]	https://www.bioconductor.org/packages/release/bioc/html/ggtree.html

Cell
Resource

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
inferCNV V1.16.0	inferCNV of the Trinity CTAT Project	https://github.com/broadinstitute/inferCNV/wiki
CNVkit V0.9.11	Talevich et al.[105]	https://cnvkit.readthedocs.io/en/stable/
Connectome V1.0.0	Raredon et al.[106]	https://msraredon.github.io/Connectome/
Ingenuity Pathway Analysis (IPA)	Krämer et al.[69]	https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/
Prism V9	GraphPad	https://www.graphpad.com/
Other		
Poly-L-Lysine Slide	Electron Microscopy Sciences	63478-AS
Chrome photomask	Front Range Photomasks	N/A
Silicon wafer	WaferPro	C04004
EVOS M7000 Imaging System	Thermo Fisher Scientific	AMF7000
DNA low-bind tube	Eppendorf SE	022431021
TapeStation 4150	Agilent Technologies	G2992AA
NovaSeq 6000 Sequencing System	Illumina	N/A
CODEX PhenoCycler-Fusion	Akoya Biosciences	N/A
Leica TCS SP5 Confocal microscope	Leica Microsystems	N/A
CFX Connect Real-Time System	BIO-RAD	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Patient specimens

De-identified archival formalin-fixed paraffin-embedded (FFPE) human lymphoma tissue blocks, originally collected by physicians for diagnostic purposes, were sourced from the Yale Pathology Tissue Services (YPTS), a Pathology-based Central Tissue Resource Lab that provides comprehensive tissue-related services and materials in de-identified format for investigators at Yale University. The tissue collection was conducted with Yale University Institutional Review Board approval with oversight by Tissue Resource Oversight Committee. Written informed consent for participation, including cases where identification was collected alongside the specimen, was obtained from patients or their guardians, adhering to the principles of the Declaration of Helsinki. Each sample was handled in strict compliance with HIPAA regulations, University Research Policies, Pathology Department diagnostic requirements, and Hospital by-laws. The excisional biopsy from the left upper arm subcutaneous nodule was collected and embedded in 2018 from a patient presenting with angioimmunoblastic T-cell lymphoma (AITL) in multiple lymph nodes and subcutaneous sites. Biopsies from the gastric antrum revealing marginal zone lymphoma of mucosa-associated lymphoid tissue (MALT) and the fundus nodule indicating diffuse large B-cell lymphoma (DLBCL) were collected and embedded in 2020. These biopsies were obtained from a patient who incidentally presented with retroperitoneal lymphadenopathy during imaging originally performed for an orthopedic visit. Upper endoscopy revealed multiple areas of erosion in the stomach, and a breath test for H. pylori was positive.

Surgical pathology report of the lymphoma biopsies

The AITL sections show a sheet of lymphocytes, some with atypical morphology. There are thick and thin bands of fibrosis and interspersed blood vessels. The atypical cells have irregular to round nuclei, speckled chromatin, variable small nucleoli, and a small amount of cytoplasm. There is infiltration into the adjacent fat. Significant mitotic figures, apoptotic figures, or necrosis is not identified. The atypical cells are CD3-positive T cells that are positive for CD4, CD2, CD5, CD10, CXCL13, and PD-1. They are negative for CD25 and CD8 with partial loss of CD7. There are abundant background CD20-postive B cells. The Ki-67 proliferation index is overall approximately 20-30%. T cell gene rearrangement was positive, showing same peaks as other sites of involvement. Flow cytometric analysis reveals CD4+ T cells are increased in the specimen, representing about 36% of total lymphocytes with few CD8+ elements detected. In addition, CD4+ T cells possess an abnormal immunophenotype.

The MALT sections reveal gastric antral mucosa with numerous lymphoid follicles showing monotonous small lymphocytes that demonstrate ovoid nuclei, condensed chromatin, and indistinct nucleoli. No large cell component is seen in this part. The tumor cells are CD20 positive B cells that co-express BCL-2 and CD43, are negative for CD5, CD10, BCL-6, CD23, LEF1, and cyclinD1. Ki-67 is low at <10%. CD3 highlights scattered small T cells. H.pylori immunostaining is negative.

The DLBCL sections reveal sheets of large pleomorphic lymphocytes, some with horseshoe shaped nuclei, dispersed chromatin, prominent nucleoli, and moderate amount of eosinophilic cytoplasm. There are numerous eosinophils in the background and no

substantial small cell lymphoma. The tumor cells are positive for CD20, CD43, and MUM1 and negative for CD10, cyclinD1, and CD30. BCL-6 is faintly expressed in <20% of cells. C-myc is expressed in >80% of tumor cells and BCL-2 is expressed in >70% of cells. Ki-67 proliferation index is approximately 70%. CD3 positive small T cells are scattered. Para-aortic lymph node biopsy performed simultaneously shows involvement by metastatic DLBCL.

Mouse paraffin tissues

The mouse E13 embryo, caudal hippocampus coronal brain/Region.9, and lymph node sections were purchased from Zyagen (San Diego, CA). The in-house Institutional Animal Care and Use Committee of Zyagen review and approve all protocols. Tissues were freshly harvested from C57BL/6 mice fixed in 10% Neutral Buffered formalin and processed for embedding in low temperature melting paraffin. All tissue preparation steps from harvesting to embedding in paraffin were done in RNase-, DNase-, and protease-free conditions. Tissue sections were hematoxylin and eosin (H&E) stained and examined by histologists with extensive experience to be sure of excellent morphology and high quality.

Sample handling and section preparation

For both human and mouse samples, paraffin blocks were sectioned at a thickness of 7-10 μm and mounted on the center of Poly-L-Lysine coated 1 x 3" glass slides. Serial tissue sections were collected simultaneously for Patho-DBiT and other staining. The sectioning of lymphoma patient samples was carried out at YPTS, while mouse sectioning was performed by Zyagen technicians. Paraffin sections were shipped in tightly closed slide boxes or slide mailers at room temperature and stored at $-80°C$ upon receipt until use.

METHOD DETAILS

Fabrication of microfluidic device

The comprehensive fabrication process, employing standard soft lithography, has been detailed in our previous publication.[14] Briefly, high-resolution chrome photomasks with a customized pattern were printed and ordered from Front Range Photomasks (Lake Havasu City, AZ). Upon receipt, the masks underwent cleaning with acetone to remove any dirt or dust. Master wafers were then produced using SU-8 negative photoresist (SU-2010 or SU-2025) on silicon wafers following the manufacturer's guidelines, with feature width of 50 μm, 20 μm, or 10 μm. The newly fabricated wafers were treated with chlorotrimethylsilane for 20 minutes to develop high-fidelity hydrophobic surfaces. Subsequently, polydimethylsiloxane (PDMS) microfluidic chips were fabricated through a replication molding process. The base and curing agents were mixed thoroughly with a 10:1 ratio following the manufacturer's guidelines and poured over the master wafers. After degassing in the vacuum for 30 min, the PDMS was cured at 70°C for at least 2 hours. The solidified PDMS slab was cut out, and the inlets and outlets were punched for further use.

DNA barcodes annealing

DNA oligos used in this study were procured from Integrated DNA Technologies (IDT, Coralville, IA) and the sequences were listed in Table S3. Barcode (100 μM) and ligation linker (100 μM) were annealed at a 1:1 ratio in 2X annealing buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM EDTA) with the following PCR program: 95°C for 5 min, slow cooling to 20°C at a rate of $-0.1°C/s$, followed by 12°C for 3 min. The annealed barcodes can be stored at $-20°C$ until use.

Tissue deparaffinization and decrosslinking

Tissue section was retrieved from the $-80°C$ freezer and equilibrated to room temperature for 10 minutes until all moisture dissipated. Following this, the tissue slide underwent a 1-hour baking process at 60°C to facilitate softening and melting of the paraffin. Removal of paraffin was achieved by immersing slides in Xylene for two changes, followed by rehydration in a series of ethanol dilutions, including two rounds of 100% ethanol and once each of 90%, 70%, and 50% ethanol, culminating in a final wash with distilled water. Each step was performed for a duration of 5 min. Subsequently, the tissue slide was submerged in 1X antigen retrieval buffer and subjected to steaming using boiling water for 30 minutes, followed by a 30-minute cooldown to room temperature. After a brief dip in distilled water, intact tissue scan was captured using a 10X objective on the EVOS M7000 Imaging System.

Permeabilization, *in situ* polyadenylation, and reverse transcription

The tissue was permeabilized for 20 min at room temperature with 1% Triton X-100 in DPBS, followed by 0.5X DPBS-RI (1X DPBS diluted with nuclease-free water, 0.05 U/μL RNase Inhibitor) wash to halt permeabilization. The tissue slide was then air-dried and equipped with a PDMS reservoir covering the region of interest (ROI). *In situ* polyadenylation was performed using *E. coli* Poly(A) Polymerase. Initially, samples were equilibrated by adding 100 μL wash buffer (88 μL nuclease-free water, 10 μL 10X Poly(A) Reaction Buffer, 2 μL 40 U/μL RNase Inhibitor) and incubating at room temperature for 5 min. Following wash buffer removal, 60 μL of the Poly(A) enzymatic mix (38.4 μL nuclease-free water, 6 μL 10X Poly(A) Reaction Buffer, 6 μL 5U/μL Poly(A) Polymerase, 6 μL 10mM ATP, 2.4 μL 20 U/μL SUPERase·In RNase Inhibitor, 1.2 μL 40 U/μL RNase Inhibitor) was added to the reaction chamber and incubated in a humidified box at 37°C for 30 min. To remove excessive reagents, the slide was dipped in 50 mL DPBS and shake-washed for 5 min after the reaction. Subsequently, 60 μL of the reverse transcription mix (20 μL 25 μM RT Primer, 16.3 μL

0.5X DPBS-RI, 12 µL 5X RT Buffer, 6 µL 200U/µL Maxima H Minus Reverse Transcriptase, 4.5 µL 10mM dNTPs, 0.8 µL 20 U/µL SUPERase·In RNase Inhibitor, 0.4 µL 40 U/µL RNase Inhibitor) was loaded into the PDMS reservoir and sealed with parafilm. The sample was incubated at room temperature for 30 min and then at 42°C for 90 min, followed by a 50 mL DPBS wash as described before.

Spatial barcoding with microfluidic devices
To ligate barcode A *in situ*, the first PDMS device was meticulously positioned atop the tissue slide, aligning the center channels over the ROI. The chip was imaged to record the positions for downstream alignment and analysis. Afterwards, an acrylic clamp was applied to firmly secure the PDMS to the slide, preventing any inter-channel leakage. The ligation mix, comprising 100 µL 1X NEBuffer 3.1, 61.3 µL nuclease-free water, 26 µL 10X T4 ligase buffer, 15 µL T4 DNA ligase, 5 µL 5% Triton X-100, 2 µL 40 U/µL RNase Inhibitor, and 0.7 µL 20 U/µL SUPERase·In RNase Inhibitor, was then prepared. For the barcoding reaction, 5 µL of the ligation solution, containing 4 µL ligation mix and 1 µL 25 µM DNA barcode A, was introduced into each of the 50 or 100 inlets. The solution was withdrawn to flow through the entire channel using a delicately adjusted vacuum. After a 30-minute incubation at 37°C, the PDMS chip was removed, and the slide was washed with 50 mL DPBS. Subsequently, the second PDMS device, featuring 50 or 100 channels perpendicular to the first PDMS, was attached to the ROI on the air-dried slide. A bright-field image was captured, and the ligation of barcode B set was performed similarly. Finally, after five flow-washes with 1 mL nuclease-free water to remove residual salt, the final scan was conducted to record the microchannel marks imprinted onto the tissue ROI.

Tissue lysis and cDNA extraction
The barcoded tissue ROI was enclosed with a clean PDMS reservoir and securely clamped using acrylic chips. A 2X lysis buffer was prepared in advance, consisting of 20 mM Tris-HCl pH 8.0, 400 mM NaCl, 100 mM EDTA, and 4.4% SDS. For tissue digestion, 70 µL of the lysis mix (30 µL 1X DPBS, 30 µL 2X lysis buffer, 10 µL 20 µg/µL Proteinase K solution) was loaded into the PDMS reservoir, sealed with parafilm, and incubated in a humidified box at 55°C for 2 hours. After the reaction, the parafilm was removed, and all the liquid containing cDNA was collected into a 1.5mL DNA low-bind tube. Additionally, 40 µL of fresh lysis mix was loaded into the reservoir to collect any remaining cDNA material. The tissue lysate was incubated overnight at 55°C to completely reverse cross-links, after which it could be stored at − 80°C until the subsequent steps.

cDNA purification, template switch, and PCR amplification
To inhibit Proteinase K activity, 5 µL of 100 µM phenylmethylsulfonyl fluoride (PMSF) in ethanol was introduced into the lysate and incubated at room temperature for 10 min with rotation. Following this, ∼35 µL of nuclease-free water was added to adjust the total volume to 150 µL. The cDNA was purified using 40 µL of Dynabeads MyOne Streptavidin C1 beads resuspended in 150 µL of 2X B&W buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl). The mixture was incubated at room temperature for 60 min with rotation to ensure sufficient binding, followed by magnetic separation and two washes with 1X B&W buffer with 0.05% Tween-20, and an additional two washes with 10 mM Tris-HCl pH 7.5 containing 0.1% Tween-20. Streptavidin beads bound with cDNA molecules were then resuspended in 200 µL of TSO Mix (75 µL nuclease-free water, 40 µL 5X RT buffer, 40 µL 20% Ficoll PM-400, 20 µL 10mM dNTPs, 10 µL 200U/µL Maxima H Minus Reverse Transcriptase, 5 µL 40 U/µL RNase Inhibitor, 10 µL 100 µM TSO Primer). The template switch reaction was conducted at room temperature for 30 min and then at 42°C for 90 min with gentle rotation. After a single wash with 10 mM Tris-HCl pH 7.5 containing 0.1% Tween-20 and another wash with nuclease-free water, the beads were resuspended in 200 µL of PCR Mix (100 µL 2X KAPA HiFi HotStart ReadyMix, 84 µL nuclease-free water, 8 µL 10 µM PCR Primer 1, 8 µL 10 µM PCR Primer 2). This suspension was then distributed into PCR stripe tubes. An initial amplification was conducted with the following PCR program: 95°C for 3 min, cycling five times at 98C for 20 s, 63C for 45 s, 72C for 3 min, followed by an extension at 72°C for 3 min and 4°C hold. Following magnetic removal of the beads, 19 µL of the PCR solution was combined with 1 µL 20X EvaGreen for quantitative real-time PCR (qPCR) analysis using the same program. The remaining samples underwent further amplification, with the cycle numbers determined by 1/2 of the saturated signal observed in qPCR results. The PCR product was then purified using SPRIselect beads at a 0.8X ratio, adhering to the standard manufacturer's instructions. The resulting cDNA amplicon underwent analysis using a TapeStation system with D5000 DNA ScreenTape and reagents. This stage provides a secure stopping point, allowing the sample to be stored at − 20°C until the next steps.

rRNA removal, library preparation, and sequencing
The SEQuoia RiboDepletion Kit was employed to eliminate fragments derived from rRNA and mitochondrial rRNA from the amplified cDNA product, following the manufacturer's guidelines. Based on the TapeStation readout profile, 20 ng of cDNA was used as the input amount, and three rounds of depletion were performed. Subsequently, 7 cycles of the aforementioned PCR program were executed to directly ligate sequencing primers, using a 100 µL system consisting of 50 µL 2X KAPA HiFi HotStart ReadyMix, ∼42 µL solution from the rRNA removal step, 4 µL 10 µM P5 Primer, and 4 µL 10 µM P7 Primer. The resulting library underwent purification using SPRIselect beads at a 0.8X ratio, quality control checked using TapeStation, and was then sequenced on an Illumina NovaSeq 6000 Sequencing System with a paired-end 150bp read length.

Generation of miR-142 knockout cells and spatial patterning

Murine Ba/F3 cells from our laboratory stock were maintained in RPMI 1640 medium supplemented with 10% heat-inactivated fetal bovine serum, 1% of penicillin-streptomycin-glutamine (100X), and 3 ng/ml recombinant murine IL-3. Lentiviral CRISPR targeting miR-142 was performed on Ba/F3 cells containing a miR-142-3p reporter construct, which was generated based on a bidirectional EGFP/mCherry reporter with four copies of miR-142-3p complementary sequences inserted in the 3' UTR of EGFP.[107] Cells were infected with a Cas9-sgRNA (cggagaccacgccacgccg) vector (Addgene, 49535) against miR-142-3p via spin-infection, following our established methods.[108] Selection was based on EGFP expression levels, and successful miR-142 knockout cells were confirmed by qPCR and Sanger sequencing. Finally, wild-type Ba/F3 control cells and miR-142 knockout cells were patterned into specific shapes, embedded and sectioned for the Patho-DBiT assay.

CODEX spatial phenotyping using PhenoCycler-Fusion

Spatial high-plex phenotyping of the adjacent FFPE section was performed following the CODEX PhenoCycler-Fusion user guide (https://www.akoyabio.com/wp-content/uploads/2021/01/CODEX-User-Manual.pdf). Briefly, the tissue section underwent deparaffinization, hydration, antigen retrieval and equilibration in staining buffer, followed by antibody cocktail staining incubated at room temperature for 3 hours in a humidity chamber. After the completion of the incubation, a series of sequential steps, including post-fixation, ice-cold methanol incubation, and a final fixative step, were performed. The tissue section, attached to the flow cell, was then incubated in 1X PhenoCycler buffer with additive for a minimum of 10 minutes to enhance adhesion. Afterwards, the CODEX cycles were configured, the reporter plate was prepared and loaded, and the imaging process commenced. Upon completion of the imaging cycles, a final QPTIFF file was generated, which could be visualized using QuPath V0.5.0.[89] Information about PhenoCycler antibody panels, experimental cycle design, and reporter plate volumes can be found in Table S4.

H&E, immunohistochemistry (IHC) and *in situ* hybridization (ISH)

Histological H&E staining and clinical-level IHC and ISH on adjacent FFPE sections were conducted at Yale University School of Medicine, Department of Pathology and at YPTS. These procedures adhered to Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory protocols as well as YPTS's rigorous standard protocols, ensuring precision and accuracy in the analysis of tissue samples.

Immunofluorescence staining (IF)

The adjacent FFPE sections underwent a standard IF procedure. After deparaffinization and antigen retrieval, the tissue sections were fixed in 4% formaldehyde for 10 minutes and subsequently blocked with DPBS containing 5% bovine serum albumin for 1 hour at room temperature. CD68 antibodies, diluted at 1:100 in the blocking buffer, were applied and left to incubate overnight at 4°C. Secondary antibodies for CD68, Alexa-594 labeled CD138, and Alexa-647 labeled CD20 were then introduced following a standard IF protocol, with a 30-minute incubation at room temperature. The nuclei were counterstained with DAPI at a 1:4000 dilution. Imaging was conducted using a Leica TCS SP5 Confocal microscope.

Genomic DNA extraction from FFPE sections for whole genome sequencing (WGS)

FFPE tissue sections were deparaffinized and decrosslinked following the procedures described above. Total DNA was extracted and purified using the Qiagen DNeasy Blood & Tissue Kit according to the manufacturer's protocol. Samples underwent quality control via agarose gel electrophoresis. Genomic DNA was randomly sheared into shorter fragments, which were end-repaired, A-tailed, and ligated with Illumina adapters. The adapter-ligated fragments were size-selected, PCR amplified, purified, quantified, and sequenced at 50X depth on an Illumina NovaSeqXPlus system with a 150bp paired-end read length.

QUANTIFICATION AND STATISTICAL ANALYSIS

Sequence alignment and generation of mRNA expression matrix

To decode sequencing data, the FASTQ file Read 2 underwent processing, involving the extraction of unique molecular identifiers (UMIs) and spatial Barcode A and Barcode B. The Read 1 containing cDNA sequences was trimmed using Cutadapt V3.4[90] and then aligned to either the mouse GRCm38-mm10 or human GRCh38 reference genome using STAR V2.7.7a.[91] Utilizing ST_Pipeline V1.7.6,[92] spatial barcode sequences were demultiplexed based on the predefined coordinates of the microfluidic channels and ENSEMBL IDs were converted to gene names, generating the gene-by-pixel expression matrix for downstream analysis. Matrix entries corresponding to pixel positions devoid of tissues were excluded. Missing pixels are inferred from nearby data to facilitate clustering analysis across the entire mapped area.

Read mapping of non-coding RNA species

Genomic data of various non-coding RNA species for mouse GRCm38-mm39 or human GRCh38 were collected from multiple databases, including GENCODE,[47] GtRNAdb,[93] and RNAcentral.[94] For RNA fragments mapping to multiple overlapping annotations, an 'overlap score' was calculated using the formula $(L_o - L_{no})/L_a$, where L_o denotes the overlapped length between the query RNA

fragment and an annotation, L_{no} represents the non-overlapped length of the query RNA fragment, and L_a is the total length of the genomic annotation. The annotation with the highest overlap score was considered the true annotation for the RNA fragment.

Gene data normalization and unsupervised clustering analysis

Spatial gene expression analysis was conducted through the Seurat V4 pipeline.[95] First, SCTransform, designed for normalization and variance stabilization in single-cell RNA sequencing (scRNA-seq) datasets,[109] was employed to normalize gene expression within each pixel. Linear dimensional reduction was performed using the "RunPCA" function, and the optimal number of principal components for subsequent analysis was determined through a heuristic method, generating an 'Elbow plot' that ranks PCA components based on their percentage of variances. Second, "FindNeighbors" function was utilized to embed pixels in a K-nearest neighbor graph structure based on the Euclidean distance in PCA space, and "FindClusters" was implemented using a modularity optimization technique to cluster the pixels. Finally, the non-linear dimensional reduction function "RunUMAP" was applied to visually explore spatial heterogeneities using the Uniform Manifold Approximation and Projection (UMAP) algorithm, and the identification of differentially expressed genes (DEGs) defining each cluster was accomplished through the "FindMarkers" function for pairwise comparison between groups of pixels.

Integration with scRNA-seq datasets

At a pixel size of 50 μm, Patho-DBiT assay pixels capture the expression profiles of multiple cells. We employed the 'anchor'-based integration workflow included into Seurat V4 to deconvolute each spatial voxel, predicting the underlying composition of cell types.[110] This facilitated the probabilistic transfer of annotations from a reference to a query set. After standard "SCTransform" normalization of both Patho-DBiT and reference scRNA-seq data, the "FindTransferAnchors" function identified anchors between the reference scRNA-seq and our query Patho-DBiT object. Subsequently, the "TransferData" function was applied for label transfer, providing a probabilistic classification for each spatial pixel based on well-annotated scRNA-seq identities. These predictions were added as a new assay to the Patho-DBiT object. Unsupervised clustering was then performed on the combined Patho-DBiT and reference dataset, resulting in an integrated UMAP where Patho-DBiT pixels were projected onto the scRNA-seq cluster landscape. The mouse organogenesis reference dataset was obtained from GSE119945,[15] and the mouse brain cortex and hippocampus reference dataset was downloaded from the Allen Mouse Brain Atlas (https://portal.brain-map.org/atlases-and-data/rnaseq).[21]

qPCR analysis of rRNA removal efficiency

To assess the rRNA removal efficiency of Patho-DBiT, qPCR analysis was performed on cDNA amplicons obtained from three independent FFPE mouse E13 embryos before and after rRNA removal. Each sample, with an input amount of 2.5 ng cDNA, underwent a total volume of 25 μL in the KAPA HiFi HotStart ReadyMix reaction system. Forward and reverse primers targeting cytoplasmic (5S, 5.8S, 18S, and 28S) and mitochondrial (12S and 16S) rRNA were custom-designed and ordered from IDT (sequences provided in Table S5). QuantiTect Primer Assays for mouse GAPDH and β-actin genes served as internal controls. The qPCR reactions were conducted on a CFX Connect Real-Time System, and fold changes were determined using the comparative CT method.[111]

Gene body coverage calculation

For each sample, we computed the percentile coverage along the gene body from 5' to 3' using the "geneBody_coverage.py" module from the RSeQC package V5.0.1 with default settings.[96]

Spatial microRNA alignment and analysis

The transcriptome output function of STAR was used to generate the microRNA transcriptome BAM file using annotations obtained from miRBase.[97] Only primary alignment of each read mapped to microRNA was preserved, and microRNAs with detected UMI count ≥ 1 were included in the downstream analysis. The nucleotide length of each mapped microRNA read was calculated and the count distribution across all identified microRNAs was generated. To visualize read coverage across the reference genomic region, the BAM file of specific microRNAs was directly imported into the Integrative Genomics Viewer (IGV),[98] focusing on the precursor microRNA region, including the mature 5p-strand and 3p-strand, for detailed visualization. The spatial microRNA-by-pixel expression matrix was generated by decoding barcode sequences, and standard functions integrated into Seurat V4 were utilized for normalization and spatial visualization.

Spatial alternative splicing analysis

For both our Patho-DBiT FFPE mouse brain and 10x Genomics Visium fresh-frozen mouse brain samples, we evaluated five types of alternative splicing events (SE, RI, A3SS, A5SS, MXE) and their respective splice-junction-spanning read counts from the Binary Alignment Map (BAM) file of each sample. The rMATS-turbo pipeline V4.1.2[23] with parameters "-t single –allow-clipping –variable-read-length" and the GRCm38-mm10 mouse gene annotation were employed for this analysis. Candidate events were considered for further analysis if their inclusion and skipping isoform read counts were both ≥ 2 when aggregated from all pixels within the sample. Within each spatial pixel, a gene was deemed to have alternative splicing information if at least one splice-junction-spanning read of either inclusion or skipping isoform was detected. To identify alternative splicing events showing regional differences in our Patho-DBiT data, pseudo-bulk BAM files of each brain region were generated by merging reads from all the pixels within the same

region. Pairwise regional differential alternative splicing analysis was performed by running rMATS-turbo on the generated pseudo-bulk BAM files for each pair of two regions. An alternative splicing event was considered significant if it exhibited an exon inclusion level difference of > 0.05 between two regions, with a false discovery rate (FDR) of ≤ 0.05. Exon inclusion levels and FDRs were obtained from rMATS-turbo's splice-junction-read-based outputs (*.MATS.JC.txt). The spatial locations of reads corresponding to alternative splicing events were deciphered using their barcode sequences, resulting in distinct inclusion and skipping isoform expression matrices for each event type. Seurat V4's "NormalizeData" with a "LogNormalize"-based global-scaling normalization was applied, and the "SpatialFeaturePlot" was employed to visualize the spatial distribution of selected isoforms. The download links for the 10x Genomics datasets can be found in Table S6. For cross-validation with an independent dataset,[29] spatial expression profiles of representative genes and corresponding isoforms were obtained from an online database (http://carra.ipmc.cnrs.fr:3838/SiTx/) provided by the published work.

Spatial adenosine-to-inosine (A-to-I) RNA editing analysis

A total of 107,095 reference mouse A-to-I RNA editing sites were retrieved from the REDIportal V2.0 database[99] as of the download date on 9-20-2023. The counts of edited and unedited reads for each editing site were calculated from the BAM file containing all spatial pixels using the "mpileup" subcommand of samtools V1.16.1,[100] with parameters "–no-output-ins –no-output-ins –no-output-del –no-output-del –no-output-ends -B -d 0 -Q 25 -q 25" along with the reference editing site list and the GRCm38-mm10 mouse reference genome. Reads with bases "A" and "G" at editing sites were classified as unedited and edited, respectively. Candidate A-to-I RNA editing sites for further analysis were defined as those with a total coverage of ≥ 10 and an edited read count of ≥ 1 when aggregated from all pixels within the sample. The overall editing ratio for each editing site was computed by dividing the total number of edited reads across all pixels by the total coverage of that site. Similarly, the average editing ratio for each pixel or brain region was determined by dividing the total edited reads by the total coverage of all editing sites within that specific area. The reference spatial dataset, containing editing sites, editing ratios, and total read counts from long-read Nanopore sequencing of fresh frozen mouse brain sections, was obtained from the literature.[29] For comparison with our Patho-DBiT dataset, only sites with ≥ 10 long reads were included.

Inferring high-resolution tissue architecture using iStar

The iStar algorithm comprises three components: a histology feature extractor, a high-resolution gene expression predictor, and a tissue architecture annotator, as detailed in our previous publication.[41] Briefly, H&E images were rescaled to standardize pixel size, partitioned into hierarchical image tiles, and histology features were extracted and trained using a vision transformer. A feed-forward neural network, trained through weakly supervised learning, then predicted superpixel-level gene expressions using the top 1,000 highly variable genes from the Patho-DBiT expression matrix. The tissue was segmented by clustering superpixels based on gene expression, co-registered with H&E staining images. The top 10 genes defining each cluster were used for biological interpretation, refined by a pathologist according to tissue morphology.

Spatial RNA splicing dynamics

The analysis involved extracting counts of spliced and unspliced reads independently from the aligned BAM file. Genomic regions corresponding to exons and introns were obtained from the GENCODE annotation.[47] Utilizing the "intersect" tool within bedtools V2.31.0,[101] reads overlapping with intronic regions were identified, and the associations between each read and its corresponding gene were documented. The remaining reads that overlapped with exonic regions were selected, and their connections to the overlapped genes were documented as well. After demultiplexing their spatial coordinates, reads containing region records were processed to generate spliced and unspliced count matrices, respectively. Following this, the two matrices were imported into the scVelo pipeline,[42] where RNA velocity, pseudotime analysis, and visualization were implemented using default settings. Pixel annotations, featuring assigned cluster identities, were transferred from the Seurat clustering analysis conducted on the combined exonic and intronic expression matrices.

Coverage comparison with 10x Genomics datasets

To assess and compare the genomic location coverage bandwidth between Patho-DBiT and 10x Genomics 3' scRNA-seq datasets, we obtained the aligned BAM files from the respective website. The sequencing depth was normalized by randomly selecting an equivalent number of reads in each 10x Genomics file and our Patho-DBiT data. Genomic regions with at least one detected read were considered covered. The download links for the 10x Genomics datasets can be found in Table S6.

Spatial single nucleotide variant (SNV) analysis

The germline mode of Strelka V2.9.10[102] with "–rna" flag was utilized to identify potential SNVs from the mapped BAM file. Only high-confidence variant loci marked as "PASS" in Strelka, along with SNV sites having sequencing counts ≥ 10, were retained for further analysis. Each pixel and SNV site were assigned values: 0 for no read coverage, 1 for wild type, 2 for heterozygous mutation, or 3 for homozygous mutation. Positions with no detected mutated nucleotides were labeled as wild type, those with both mutated and wild-type nucleotides were classified as heterozygous mutation, and sites with only mutated nucleotides were categorized as homozygous mutation. For each pixel, only SNV sites identified by Strelka were incorporated into the profile, considering that RNA-seq data

may not cover the entire genome.[112] By combining spatial coordinates defined by barcode sequences, a mutation-by-pixel matrix was generated, and the cumulative number of SNVs within each pixel was calculated to delineate spatial mutational burden. Subsequently, this mutational matrix was input into the Seurat V4 pipeline to perform unsupervised clustering analysis using standard normalization, dimensional reduction, and spatial visualization methods.

WGS data alignment and analysis
Raw FASTQ files were mapped to the human GRCh38 reference genome using BWA V0.7.17.[103] Duplicate reads were marked by GATK V4.2.0.0 and removed with samtools V1.16.1.[100] De-duplicated BAM files were then used to call somatic mutations with Strelka V2.9.10[102] in somatic mode.

Evolutionary relationship construction
To construct evolutionary relationships among the identified gene clusters, the DNA mutation list was refined using strict criteria: (1) mutational frequency <10% in the normal sample and (2) significantly lower mutation rate in the normal sample compared to the tumor sample (Fisher test p-value < 0.001). The refined DNA mutation list was cross-referenced with the SNV list to generate a DNA-verified SNV list. A variation site in a cluster was considered valid if the read coverage was >2. Each valid SNV site was annotated as 'mutated' if the mutational frequency was >0.25, otherwise as 'wild-type'. The distance between clusters was calculated using the formula N_d/N_v, where N_d is the number of sites with different genotypes between two clusters, and N_v is the number of valid sites in both clusters. The phylogenetic relationship was constructed using the UPGMA method with the 'hclust' function in R V4.3.1 and visualized using ggtree V3.8.2.[104]

Copy number variation (CNV) analysis
CNV inference analysis was conducted using inferCNV V1.16.0 with the Patho-DBiT gene expression matrix as input. The resulting output matrix was utilized to identify genomic regions exhibiting significant CNV events. CNV analysis based on WGS data was performed using CNVkit V0.9.11.[105]

Ligand-receptor interaction analysis
The R toolkit Connectome V1.0.0[106] was employed to investigate cell-cell connectivity patterns using ligand and receptor expressions from our Patho-DBiT datasets. The normalized Seurat object served as input, and cluster identities were utilized to define nodes in the interaction networks, resulting in an edge list connecting pairs of nodes through specific ligand-receptor mechanisms. We selected top-ranked interaction pairs, prioritizing those more likely to be biologically and statistically significant based on the scaled weights of each pair. The "sources.include" and "targets.include" parameters were applied to specify the source cluster emitting ligand signals and the target cluster expressing receptor genes that sense the ligands.

Ingenuity Pathway Analysis
Ingenuity Pathway Analysis (IPA, QIAGEN)[69] was employed to uncover the underlying signaling pathways regulated by the DEGs characterizing each identified cluster or two groups. The DEG list, along with the corresponding fold change value, p-value, and adjusted p-value of each gene, was imported into the software. The Ingenuity Knowledge Base (genes only) served as the reference set for performing Core Expression Analysis. The z-score was utilized to assess the activation or inhibition level of specific pathways. Conceptually, the z-score is a statistical measure gauging how closely the actual expression pattern of molecules in our DEG dataset aligns with the expected pattern based on the literature for a particular annotation. A z-score >0 signifies activation or upregulation, while a z-score <0 indicates inhibition or downregulation. A z-score ≥ 2 or ≤ -2 is considered significant. The p-value for each identified signaling pathway is calculated using the right-tailed Fisher's Exact Test. This significance reflects the probability of the association of molecules from our Patho-DBiT dataset with the canonical pathway reference dataset. Additionally, a graphical summary (Figures 3H and S8H) was generated to provide an overview of the major biological themes in our IPA Core Analysis and illustrate how these concepts interrelate. A machine learning algorithm, relying entirely on prior knowledge, was deployed to score inferred relationships between molecules, functions, and pathways. Networks were constructed from the IPA analysis results using a heuristic graph algorithm.

Statistical analysis
Statistical analyses were performed using Prism V9 (GraphPad), with the specific tests employed indicated.

Supplemental figures
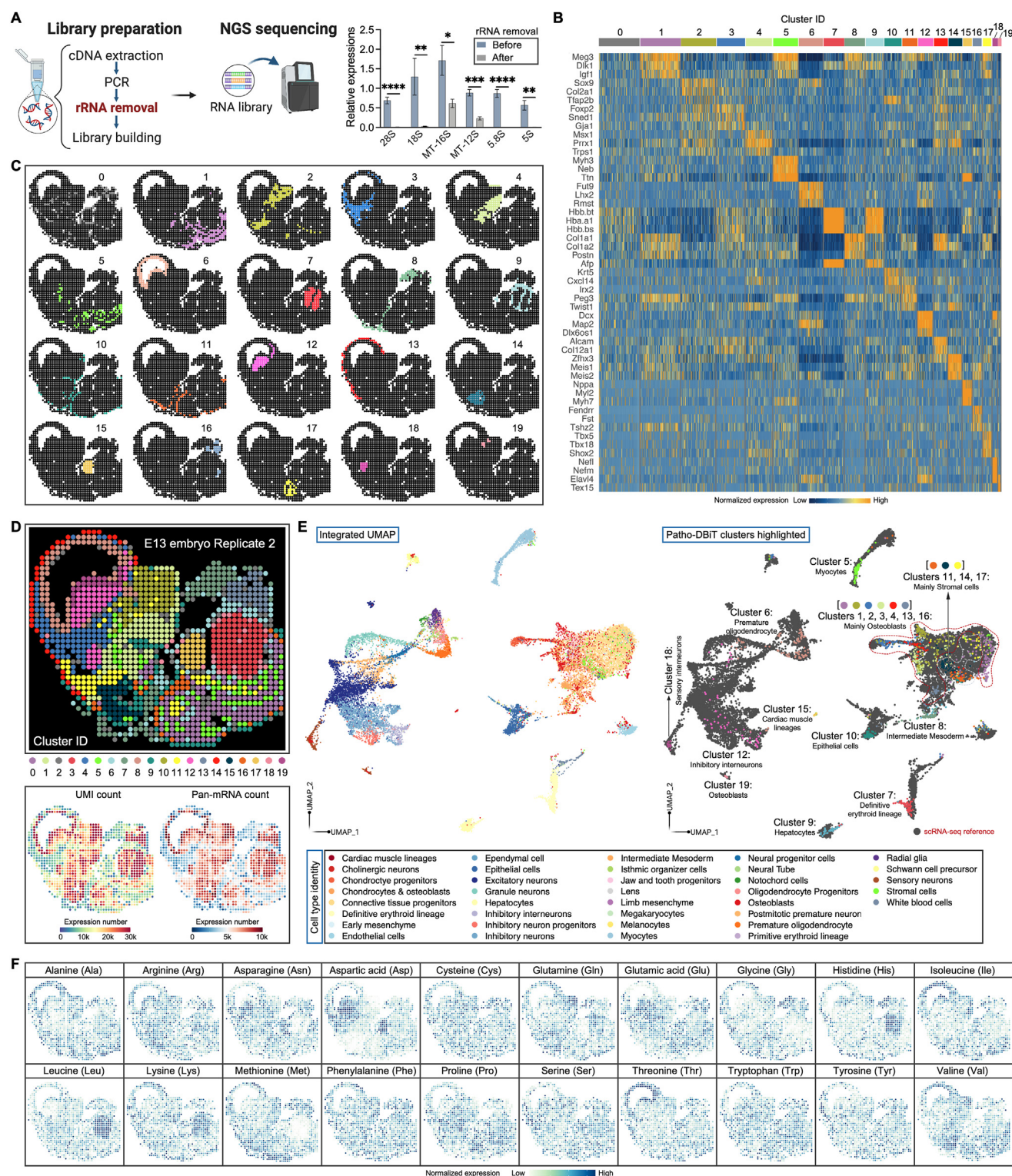
(legend on next page)

Figure S1. Patho-DBiT performance and spatial tRNA profiling in E13 mouse embryo sections, related to Figure 1

(A) Left: Patho-DBiT steps post tissue barcoding. Right: qPCR analysis of the rRNA removal efficiency.

(B) Top-ranked DEGs defining each cluster.

(C) Spatial distribution of identified clusters.

(D) Top: unsupervised clustering of the replicate E13 mouse embryo section. Bottom: spatial pan-mRNA and UMI count maps of the replicate section.

(E) Integration of spatial RNA data with scRNA-seq mouse organogenesis data (Cao et al.[15]).

(F) Spatial expression of tRNA isotypes in the E13 mouse embryo section.
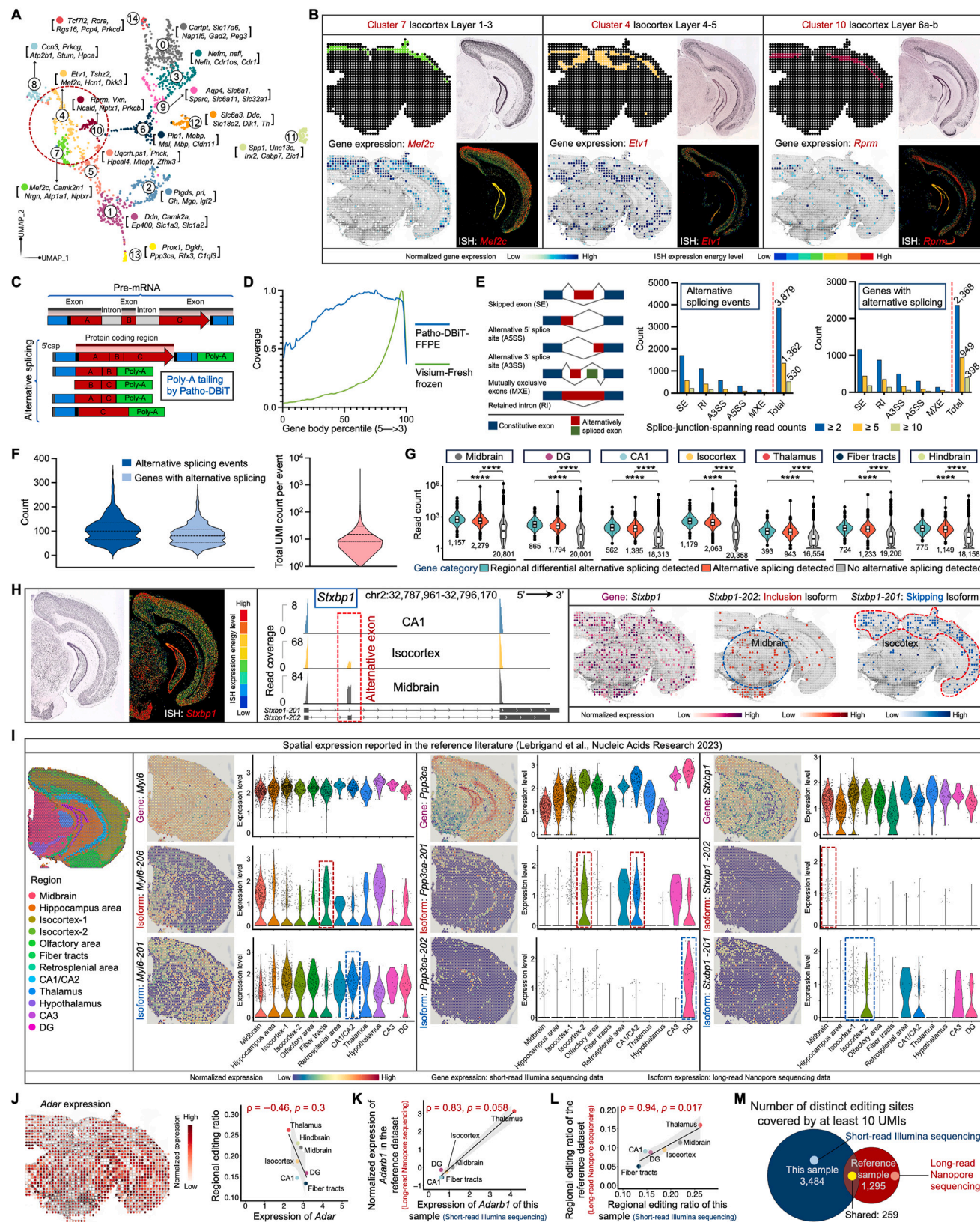
(legend on next page)

Figure S2. Spatial co-profiling of gene expression, regional alternative splicing, and A-to-I RNA editing in the mouse brain, related to Figure 2

(A) UMAP showing the clustering analysis of Patho-DBiT data in FFPE mouse brain section. Top 5 genes defining each cluster are indicated.

(B) Spatial distribution of clusters 4, 7, and 10 in the isocortex region alongside their principle-defining genes. ISH staining and expression images are from the Allen Mouse Brain Atlas.

(C) Molecular underpinnings of alternative splicing detection by Patho-DBiT.

(D) Read coverage along the gene body from 5′ to 3′. Comparison involves Patho-DBiT data and 10× Genomics Visium dataset on a fresh-frozen brain section.

(E) Number of detected spliced events and corresponding parental genes under different read count thresholds.

(F) Distribution of detected splicing events and their corresponding parental genes across spatial pixels, along with the UMI count distribution for each splicing event.

(G) Regional read count coverage of genes with or without alternative splicing events detected. The number of genes within each violin plot is indicated. Significance level was calculated with two-tailed unpaired Student's t test, **** $p < 0.0001$.

(H) Junction read coverage of *Stxbp1* splicing event in specific brain regions. Spatial expression patterns of the gene, exon inclusion isoform, and exon skipping isoform are shown. ISH staining and expression images are from the Allen Mouse Brain Atlas.

(I) Spatial expression pattern of representative genes and their corresponding isoforms, as obtained from a published database (http://carra.ipmc.cnrs.fr:3838/SiTx/) using long-read nanopore sequencing.

(J) Left: spatial *Adar* expression. Right: correlation between the *Adar* expression and the average regional editing ratio across various brain regions.

(K and L) Correlation between regional *Adarb1* expression (J) or editing ratio (K) detected by short-read Illumina sequencing-based Patho-DBiT and those detected by long-read nanopore sequencing (Lebrigand et al.[29]). Note, the hindbrain was not included in the brain sample profiled in the reference dataset.

(M) Venn plot showing the count of distinct editing sites covered by a minimum of 10 UMIs in both datasets.
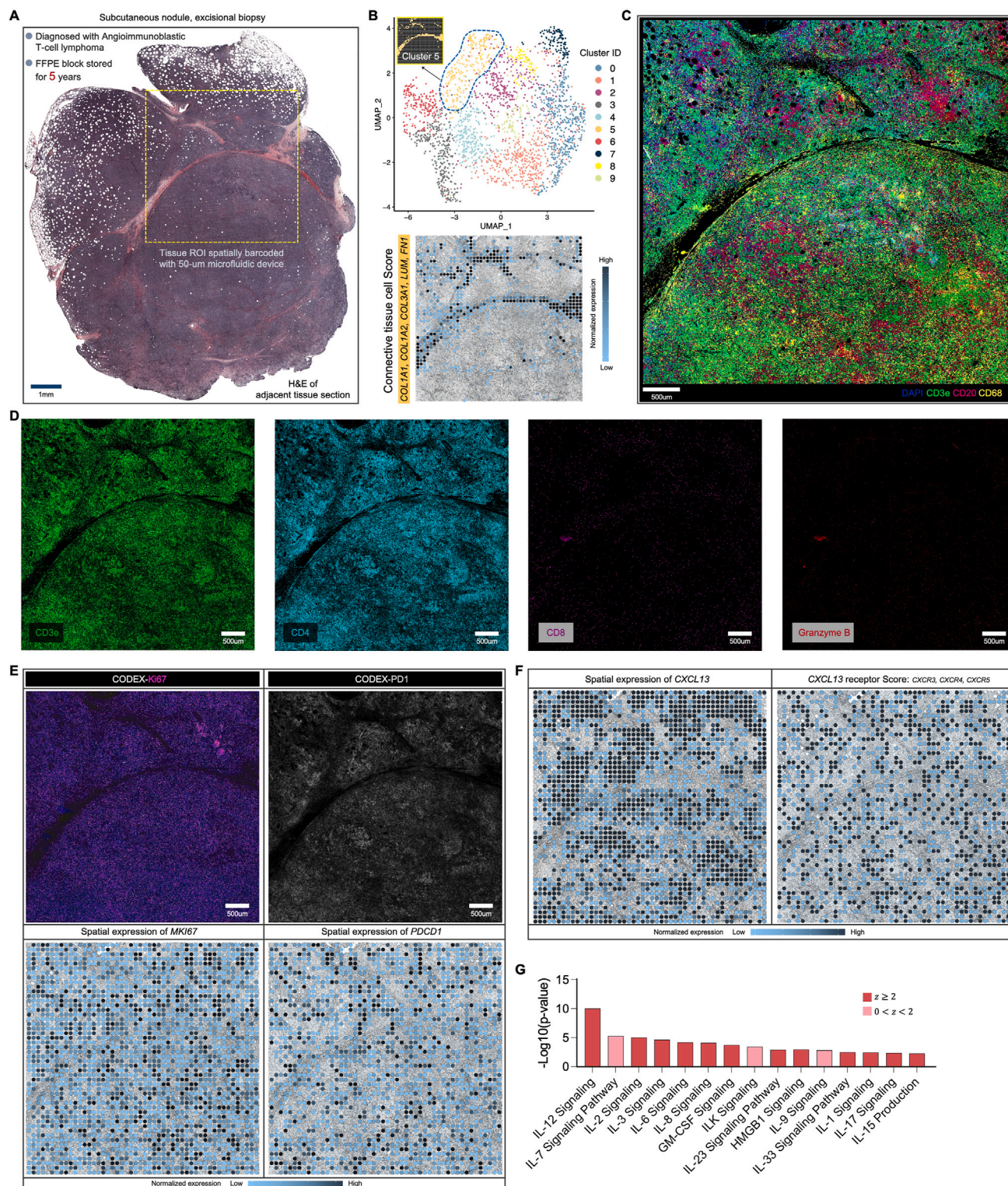
Figure S3. Detection accuracy of Patho-DBiT in the AITL section cross-validated with CODEX, related to Figure 3

(A) Full scan of the H&E staining of an adjacent section. Yellow square indicates the region of interest (ROI) spatially barcoded with the 50 μm microfluidic device.

(B) Top: UMAP showing the clustering analysis of Patho-DBiT data in the AITL section. The spatial distribution of cluster 5 is indicated. Bottom: spatial expression of the connective tissue cell score. Genes defining this module score are listed.

(C) Enlarged CODEX image corresponding to the ROI barcoded by Patho-DBiT.

(legend continued on next page)

(D) CODEX images showing expression of CD3e, CD4, CD8, and granzyme B.

(E) Spatial expressions of *MKI67* and *PDCD1* and the corresponding CODEX images of Ki67 and PD-1.

(F) Spatial expressions of *CXCL13* and the *CXCL13* receptor Score. Genes defining this module score are listed on top.

(G) Signaling pathways related to T cell functions in cluster 0, a malignant T cell population. *Z* score is computed and used to reflect the predicted activation level ($z > 0$, activated; $z < 0$, inhibited; $z \geq 2$ or $z \leq -2$ can be considered significant).
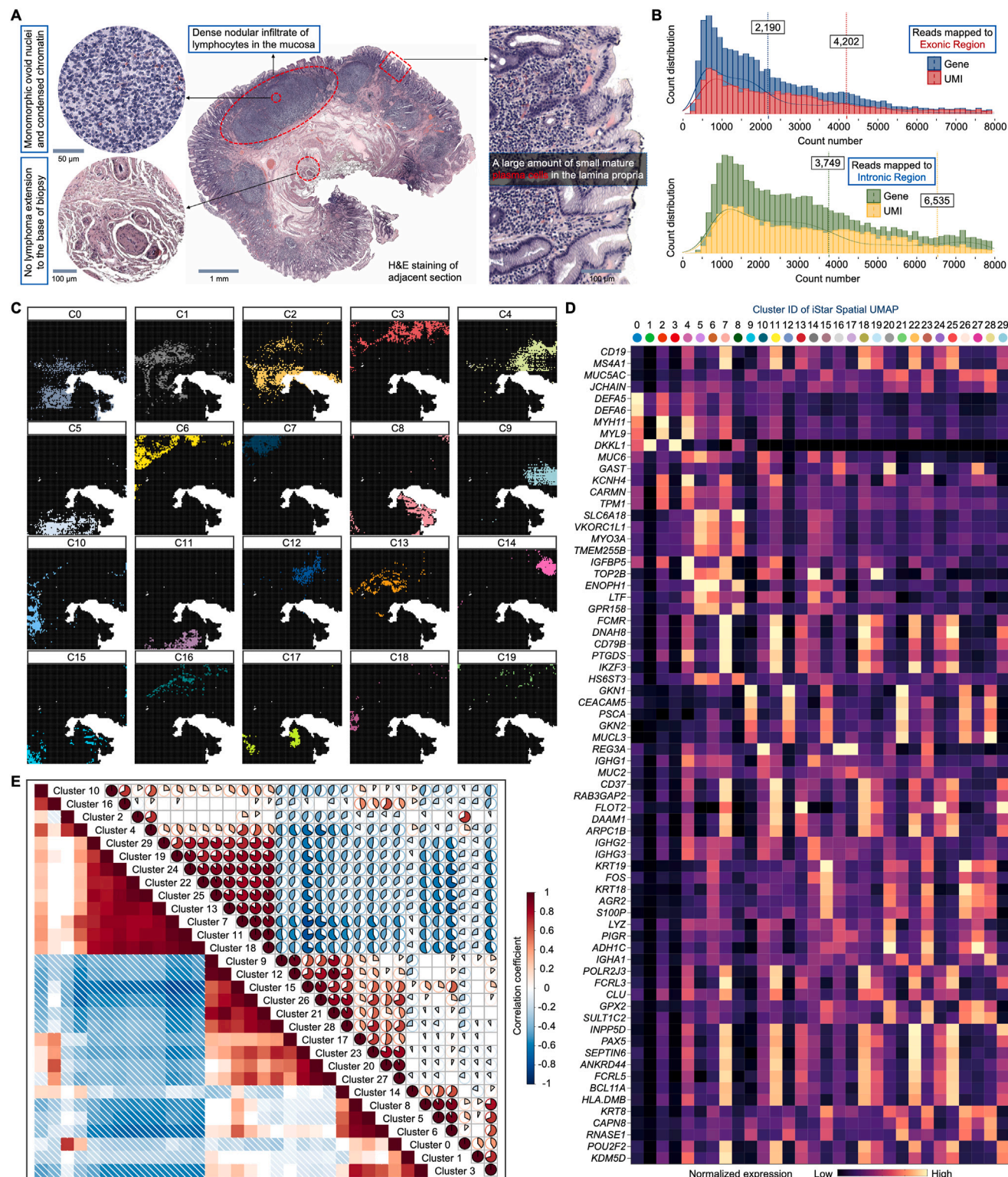
Figure S4. High-resolution spatial profiling of the MALT section, related to Figure 4

(A) H&E staining of an adjacent section with key histological information annotated by a pathologist.

(B) Distribution of detected gene/UMI counts per spatial pixel from reads mapped to exonic or intronic region. The dashed lines indicate average level of gene or UMI count.

(legend continued on next page)

(C) Spatial distribution of identified clusters.

(D) Expression of top-ranked defining genes in each iStar cluster.

(E) Correlation matrices depicting relationships among iStar clusters based on gene expression profiles. Color intensity reflects Pearson's correlation coefficients, with only statistically significant correlations ($p < 0.05$) indicated.
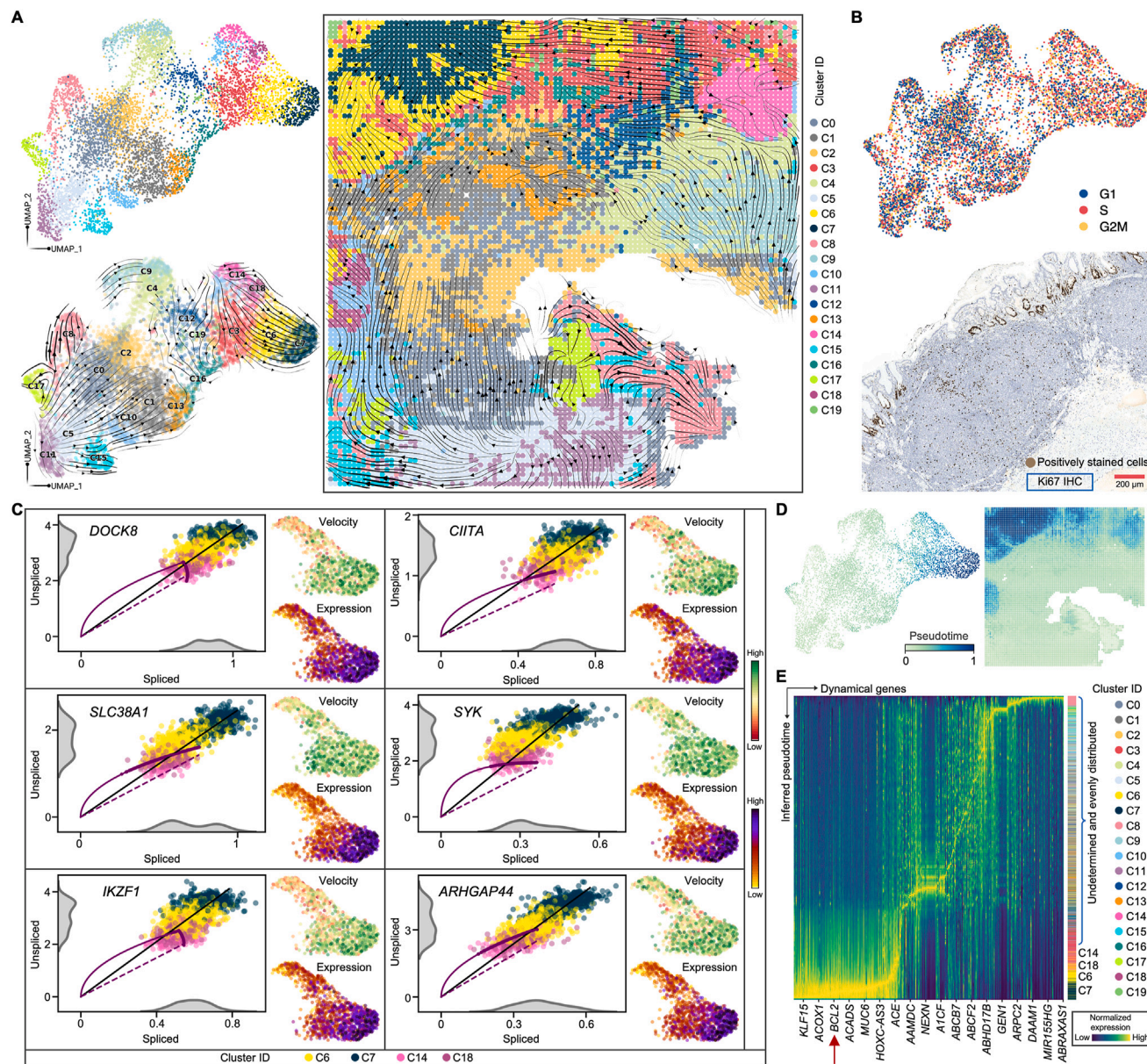
**Figure S5. Tumor differentiation trajectory revealed by spatial RNA splicing dynamics, related to Figure 4**

(A) Left top: UMAP visualization of clusters identified in the MALT sample. Left bottom: velocities derived from the dynamical RNA splicing activities are visualized as streamlines in a UMAP-based embedding. Right: spatial velocity pattern.

(B) Top: cell cycle score colored by the G1, S, or G2/M stage. Bottom: IHC staining for Ki67 in the tumor region of an adjacent section.

(C) Phase portraits showing the ratio of unspliced and spliced RNA for top-ranked genes driving the dynamic flow from cluster C14 to C7, along with their expression and velocity level within the three tumor clusters. The dashed purple line corresponds to the estimated splicing steady state.

(D) Spatial pseudotime of underlying cellular processes based on the transcriptional dynamics. A discernible change is evident exclusively within the tumor clusters, where a higher pseudotime number denotes a later differentiation stage.

(E) Gene expression dynamics resolved along the pseudotime direction showing a clear cascade of transcription of the top-ranked genes.
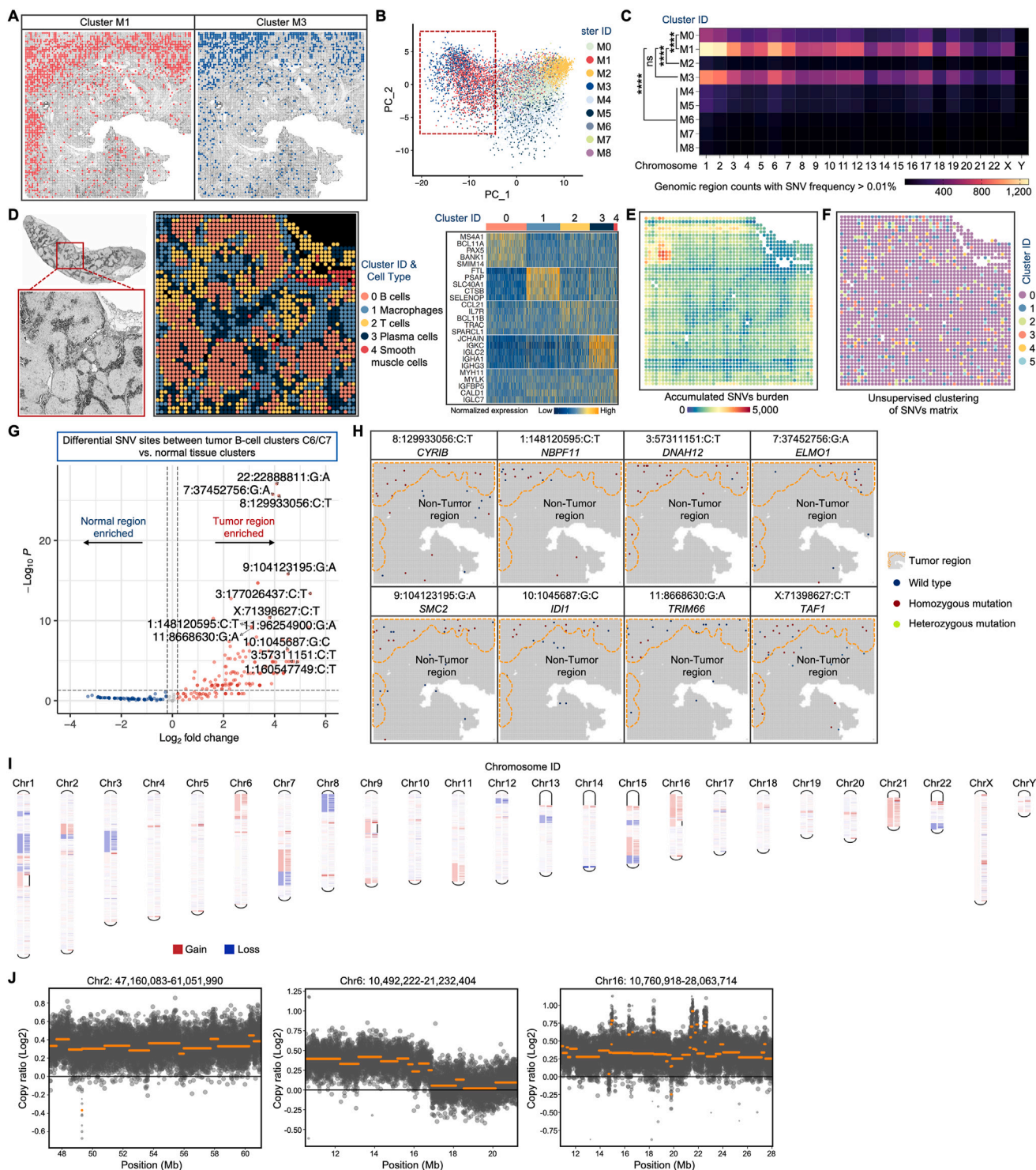
Figure S6. Validation of spatial variant analysis using a healthy donor sample and WGS, related to Figure 5

(A) Spatial distribution of single-nucleotide variant (SNV) clusters M1 and M3.

(B) Principal-component analysis (PCA) of the identified SNV clusters.

(C) Count comparison of genomic regions with SNV frequency >0.01% between SNV clusters. The SNV frequency was counted within sliding genomic regions of 10,000 bp.

(D) Patho-DBiT mapping of an FFPE lymph node section from a healthy donor using a 20 μm device. Left: tissue scanning with the enlarged region of interest. Middle: unsupervised clustering revealed 5 distinct clusters. Right: top 5 genes defining each cluster, which were used for cell type annotation.

(legend continued on next page)

(E) Spatial expression map of accumulated SNVs burden in the donor sample.

(F) Unsupervised clustering of the spatial SNV matrix in the donor sample.

(G) Differentially expressed SNV sites between the tumor and non-tumor clusters.

(H) Spatial distribution of top-ranked SNV sites upregulated in tumor B-cell clusters.

(I) Copy number variation (CNV) profile across all chromosomes based on WGS data.

(J) Scatter plot showing CNVs in selected regions.
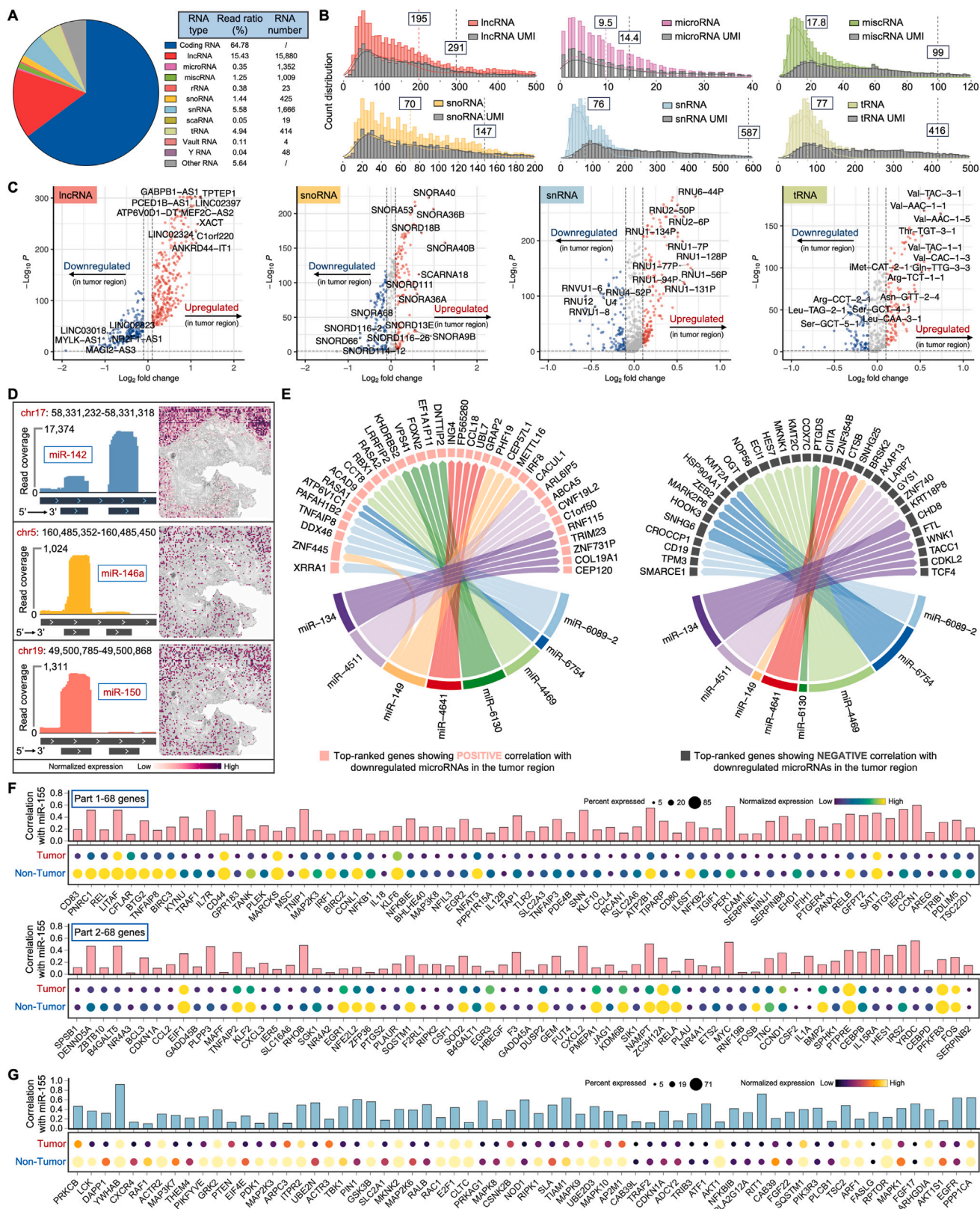
(legend on next page)

Figure S7. Spatial expression profile of non-coding RNAs and microRNA regulations in the MALT section, related to Figure 6

(A) Proportion of reads mapped to different RNA categories and the corresponding counts of each RNA type.

(B) Distribution of selected non-coding RNAs and corresponding UMI counts per spatial pixel. Dashed lines indicate the average number of RNAs or UMI counts.

(C) Differentially expressed non-coding RNAs between tumor B-cell clusters and non-tumor clusters.

(D) Spatial mapping of microRNAs enriched in marginal zone B cells or lymphoma. The read coverage mapped to the reference genome location and spatial distribution are shown.

(E) Regulatory network between the top downregulated microRNAs and the gene expression in the tumor region. Genes with the highest rankings, demonstrating positive or negative correlations with the microRNAs, were separately illustrated. Edge thickness is proportional to correlation weights.

(F and G) Correlation analysis between miR-155 and each gene in the GSEA-defined NF-κB signaling (F) or PI3K-AKT signaling (G). Enhanced expression of genes participating in both signaling pathways is observed in the tumor region compared with the non-tumor region. The dot size indicates the percentage of pixels expressing the gene, and the color shade represents normalized expression level.
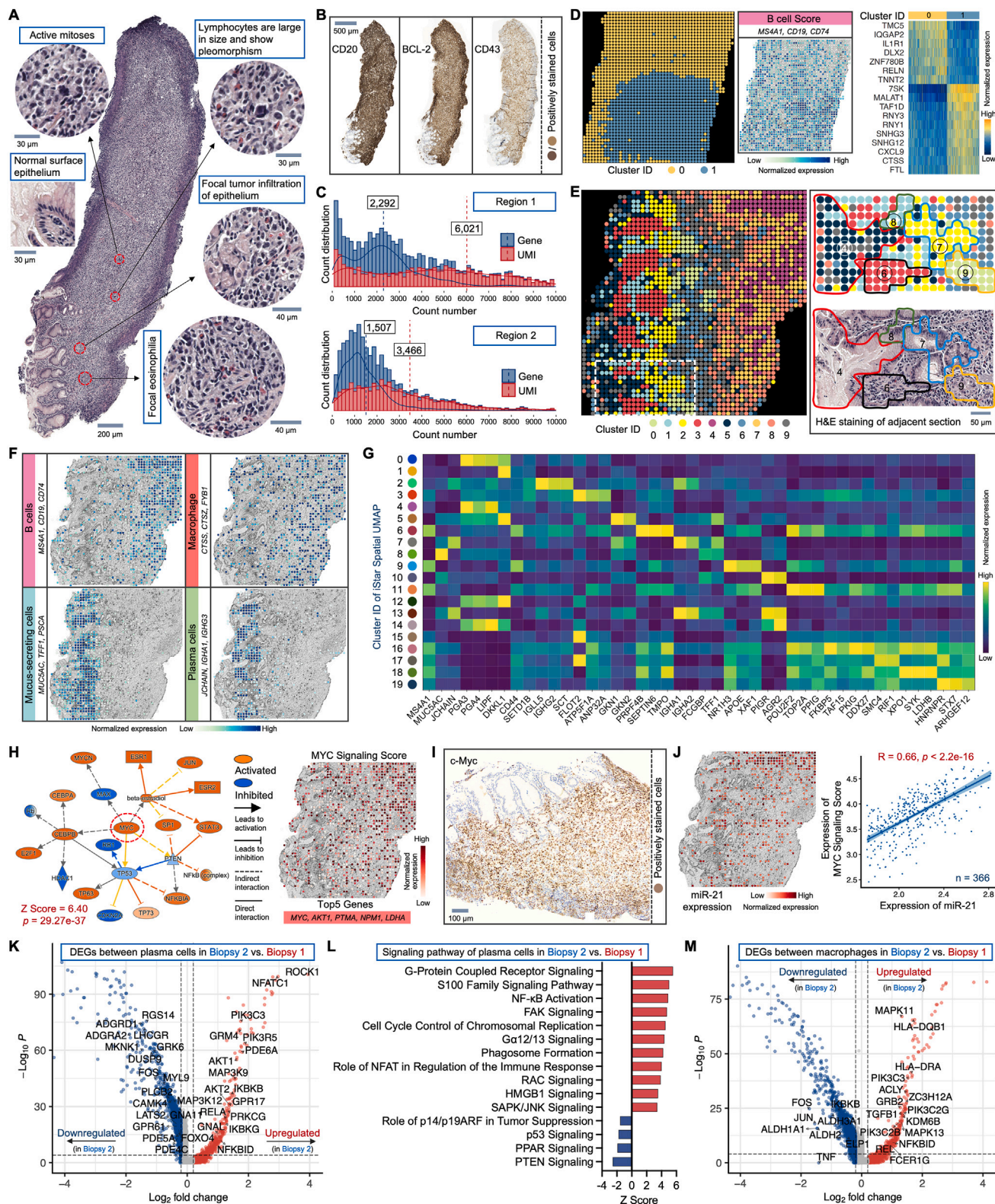
Figure S8. Spatial transcriptome and microRNA analysis of the DLBCL section, related to Figure 7

(A) H&E staining of an adjacent section with key histological information annotated by a pathologist.

(B) IHC staining of CD20 and surface markers commonly detected in non-Hodgkin B-cell lymphomas (BCL-2 and CD43) on adjacent sections.

(legend continued on next page)

(C) Distribution of detected gene/UMI counts per spatial pixel within two DLBCL regions. Only reads mapped to the exonic region were considered. Dashed lines represent the average gene or UMI count level.

(D) Spatial analysis of the region 1 section. Left: unsupervised clustering identified 2 clusters. Middle: spatial expression of the B-cell score, with defining genes listed. Right: top-ranked genes defining each cluster.

(E) Spatial analysis of the region 2 section. Left: unsupervised clustering identified 10 clusters. Right top: enlarged transcriptomic neighborhood highlighted by the white square in the left spatial UMAP. Right bottom: tissue morphology of the corresponding area defined by H&E staining of an adjacent section.

(F) Spatial expression of representative cell types in the region 2 section. Genes defining each module score are listed.

(G) Expression of top-ranked defining genes in each iStar cluster.

(H) Left: mechanistic network analysis identifying a significant activation of the master regulator MYC in tumor B cells. Right: spatial expression of the MYC signaling score. Top 5 genes defining this module score are listed.

(I) IHC staining for c-Myc on an adjacent section.

(J) Left: spatial expression map of miR-21. Right: correlation analysis between miR-21 expression and the MYC signaling score. The Pearson correlation was calculated across 366 spatial pixels within the tumor region.

(K) DEGs between plasma cells in DLBCL vs. MALT biopsy.

(L) Signaling pathways regulated by DEGs identified in (K).

(M) DEGs between macrophages in DLBCL vs. MALT biopsy.

In (H) and (L), Z score is computed and used to reflect the predicted activation level ($z > 0$, activated; $z < 0$, inhibited; $z \geq 2$ or $z \leq -2$ can be considered significant).