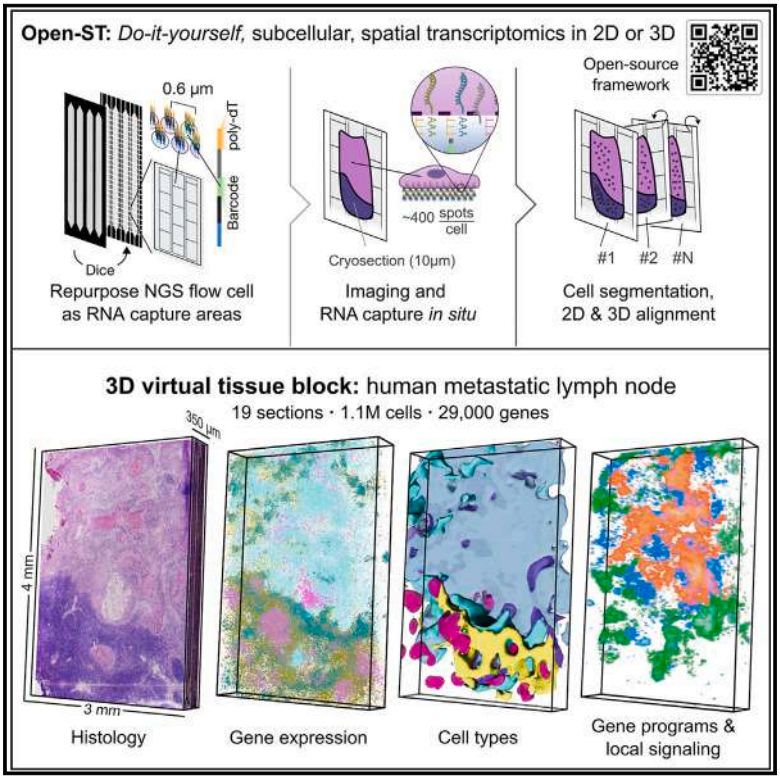
Cell

Open-ST: High-resolution spatial transcriptomics in 3D

Graphical abstract



Highlights

- Open-ST is an end-to-end, open-source framework applicable to any species
- Open-ST is easy to use, high resolution, cost efficient, and 3D scalable
- Open-ST dissects tissue heterogeneity in human clinical samples
- Open-ST generates 3D virtual tissue blocks to identify potential biomarkers in 3D

Authors

Marie Schott, Daniel León-Periñán, Elena Splendiani, ..., Giuseppe Macino, Nikos Karaiskos, Nikolaus Rajewsky

Correspondence

macino@bce.uniroma1.it (G.M.), nikolaos.karaiskos@mdc-berlin.de (N.K.), rajewsky@mdc-berlin.de (N.R.)

In brief

Open-ST is an end-to-end experimental and computational workflow for do-it-yourself subcellular spatial transcriptomics in 2D or 3D at low cost.

Schott et al., 2024, Cell *187*, 3953–3972
July 25, 2024
https://doi.org/10.1016/j.cell.2024.05.055

CellPress

Cell

CellPress
OPEN ACCESS

Article

Open-ST: High-resolution spatial transcriptomics in 3D

Marie Schott,[1,13] Daniel León-Periñán,[1,13] Elena Splendiani,[1,2,13] Leon Strenger,[1] Jan Robin Licha,[1] Tancredi Massimo Pentimalli,[1] Simon Schallenberg,[3] Jonathan Alles,[1] Sarah Samut Tagliaferro,[1] Anastasiya Boltengagen,[1] Sebastian Ehrig,[1] Stefano Abbiati,[1,4] Steffen Dommerich,[5] Massimiliano Pagani,[4,6] Elisabetta Ferretti,[2] Giuseppe Macino,[1,7,*] Nikos Karaiskos,[1,*] and Nikolaus Rajewsky[1,8,9,10,11,12,14,*]

[1]Laboratory for Systems Biology of Regulatory Elements, Berlin Institute for Medical Systems Biology (BIMSB), Max-Delbrück-Centrum for Molecular Medicine in the Helmholtz Association (MDC), Hannoversche Str. 28, 10115 Berlin, Germany
[2]Department of Experimental Medicine, Sapienza University, Rome, Italy
[3]Institute of Pathology, Charité - Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt-Universität Berlin, 10117 Berlin, Germany
[4]IFOM ETS - The AIRC Institute of Molecular Oncology, Milan, Italy
[5]Department of Otorhinolaryngology, Charité - Universitätsmedizin Berlin, Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Charitéplatz 1, Berlin 13353, Germany
[6]Department of Medical Biotechnology and Translational Medicine, Università degli Studi, Milan, Italy
[7]Department of Cellular Biotechnologies and Hematology, La Sapienza University of Rome, 00161 Rome, Italy
[8]Charité - Universitätsmedizin, Charitéplatz 1, 10117 Berlin, Germany
[9]German Center for Cardiovascular Research (DZHK), Site Berlin, Berlin, Germany
[10]NeuroCure Cluster of Excellence, Berlin, Germany
[11]German Cancer Consortium (DKTK), Heidelberg, Germany
[12]National Center for Tumor Diseases (NCT), Site Berlin, Berlin, Germany
[13]These authors contributed equally
[14]Lead contact
*Correspondence: macino@bce.uniroma1.it (G.M.), nikolaos.karaiskos@mdc-berlin.de (N.K.), rajewsky@mdc-berlin.de (N.R.)
https://doi.org/10.1016/j.cell.2024.05.055

SUMMARY

Spatial transcriptomics (ST) methods unlock molecular mechanisms underlying tissue development, homeostasis, or disease. However, there is a need for easy-to-use, high-resolution, cost-efficient, and 3D-scalable methods. Here, we report Open-ST, a sequencing-based, open-source experimental and computational resource to address these challenges and to study the molecular organization of tissues in 2D and 3D. In mouse brain, Open-ST captured transcripts at subcellular resolution and reconstructed cell types. In primary head-and-neck tumors and patient-matched healthy/metastatic lymph nodes, Open-ST captured the diversity of immune, stromal, and tumor populations in space, validated by imaging-based ST. Distinct cell states were organized around cell-cell communication hotspots in the tumor but not the metastasis. Strikingly, the 3D reconstruction and multimodal analysis of the metastatic lymph node revealed spatially contiguous structures not visible in 2D and potential biomarkers precisely at the 3D tumor/lymph node boundary. All protocols and software are available at https://rajewsky-lab.github.io/openst.

INTRODUCTION

Recent years have witnessed a massive increase in the development and application of spatial transcriptomics (ST) methods.[1] Unlike standard single-cell methods, ST retains the spatial context of the captured transcriptome and thus allows the direct observation of the arrangements of cells and their interactions in tissue space. These data are fundamental for understanding molecular mechanisms in health and critical for identifying and targeting the molecular origins of diseases.[2–4] For example, tumor microenvironment interactions or the spatial structure of lymph nodes are critical to understand function.[5,6] Moreover, ST avoids biases introduced by single-cell dissociation, which depletes certain cell types and activates stress pathways.[7–9]

Commercially available ST technologies that provide non-targeted capture of transcriptomes are limited by their relatively high costs and/or limited resolution—these include Visium (10× Genomics), CurioSeeker (Curio Bioscience), and Stereo-seq (STOmics).[10] Probe-based methods, including CosMx Spatial Molecular Imager (Nanostring), Molecular Cartography (Resolve Biosciences), and Xenium *In Situ* (10× Genomics), target a predesigned panel of genes and are therefore not suited for unbiased discovery or spatial genotyping.[11–13] Non-commercial spatial technologies are limited by inefficient capture, low resolution, or laborious/challenging experimental setups.[14–20]

Finally, although cells operate and communicate in 3D, building up functional tissues and organs, no end-to-end platform currently exists to generate and computationally analyze ST in 3D. Open-ST was conceived as a cost-effective method combining high-resolution and efficient whole-transcriptome capture, with extensive experimental resources and open-source software for seamless data processing and analysis in 2D and 3D.

Open-ST operates by converting Illumina flow cells into ST capture areas, an approach previously implemented in Seq-Scope.[16] Our method encompasses several key enhancements. First, we use patterned flow cell technology to create densely barcoded areas that capture polyadenylated RNA from a tissue section at a capture spot resolution of ~0.6 μm. To control the fragmentation of the flow cell into distinct capture areas, we provide a 3D-printable cutting guide. Our simplified library preparation only requires standard lab equipment and comes with a total cost of <€130 per 12 mm^2 capture area.

Compared with other sequencing-based technologies, Open-ST required the least sequencing depth to obtain equivalent transcriptomic information, with a standard sample (3 × 4 mm, 400 million [M] sequencing reads, ~50,000 cells) at ~1,000 unique molecular identifiers (UMIs) per cell, making it cost-effective. Open-ST is scalable, as a single researcher can prepare 10–15 libraries in 3 days starting from prepared capture areas, and versatile, as capture area size is adjustable within the limitations of the flow cell size. Our hematoxylin and eosin (H&E) imaging pipeline, which we optimized for fresh-frozen samples, produces high-resolution images from the same section, which Open-ST uses for cell segmentation and integration with transcriptomic data. Open-ST 2D data are robust enough to be computationally integrated into 3D ("virtual tissue blocks"). The single-cell segmentation, subcellular resolution, and 3D tissue reconstruction and interrogation capabilities are powered by a stack of modular and open-source computational tools designed for Open-ST data. This yields interactive objects agnostic to the original slicing direction, enabling the discovery of 3D molecular patterns and potential biomarkers.

Open-ST recapitulates cell types and marker genes of several tissues (mouse and human) with subcellular precision. Embryonic mouse head and adult mouse hippocampus were used to benchmark the precision, sensitivity, and spatial resolution of RNA capture against published gene expression data.

Furthermore, Open-ST can be used to study clinically relevant tissues encompassing drastically different morphologies and cell sizes (small immune cells to ~100× larger adipocytes). As a proof-of-principle, we profiled patient-matched samples from human head and neck squamous cell carcinoma (HNSCC). This cancer type has been shown to have a high diversity of transcriptional profiles and cell type composition, whose relative abundance, spatial organization, and interaction have implications in survival and therapy response.[21–24] Across all 21 sections from human samples processed with Open-ST, we reproducibly captured a median of ~600–2,000 spatially mapped transcripts per cell, covering more than 25,000 genes. We also show that our libraries were sequenced far from saturation.

For the human metastatic lymph node, we applied Open-ST to serial sections spanning 350 μm. We constructed a 3D virtual tissue block with over a million cells and 851 M transcripts embedded in the H&E stainings. In the patient-matched primary tumor, Open-ST recapitulated the transcriptomic identity of stromal, immune, and tumor cell types from the metastatic tissue, independently validated with a Xenium run. In particular, subclustering identified 10 tumor cell subtypes present in both the primary and the metastatic tumor. These subtypes, spatially patterned in the primary tumor but not in the metastasis, included proliferative, inflammatory, keratinizing, and invasive phenotypes and correlated strongly to the spatial localization of cell-cell communication hotspots, computationally predicted by ligand-receptor analyses. With the 3D virtual tissue block of the metastatic lymph node, we identified a spatially organized cholesterol biosynthesis signature and a population of macrophages at the 3D boundary between tumor and lymphoid tissue, as a potential 3D biomarker.

Due to its ease of use, cost-effectiveness, and wide applicability, we envision Open-ST to become a valuable method for spatial omics studies. To aid researchers in implementing Open-ST, we have set up an online resource with detailed experimental and computational protocols/software (https://rajewsky-lab.github.io/openst, Figure 1A).

RESULTS

The Open-ST workflow

To generate the mRNA capture areas, we leverage Illumina's sequencing-by-synthesis technology, using a custom sequencing recipe (Data S1). In brief, we register spatial barcode sequences and their associated (x and y) coordinates on the flow cell by sequencing oligos, which comprise 32-nt barcodes, adapters, and a poly(dT) region (Figure 1B; STAR Methods). Bridge amplification generates densely packed spots, each containing thousands of clonal oligonucleotides with a unique barcode sequence. Open-ST employs NovaSeq6000 S4 flow cells for spot generation, which contain regularly spaced nanowells with a center-to-center distance of ~0.6 μm. Patterning reduces the probability of mixed-barcode signals within a single spot and yields higher spot density than non-patterned flow cells.[25]

Following barcode sequencing, we process the oligonucleotides to allow poly(A) transcript capture and open the flow cell (STAR Methods). Our custom, 3D-printable tool facilitates the cutting of the opened flow cell into small capture areas while preventing surface scratches (Figure 1C; STAR Methods). The dimensions of the capture area can be chosen based on the experimental design, maximally 6.3 × 89 mm, limited by the sequenced area of each flow cell lane (Figure 1C). At 3 × 4 mm, ~360 capture areas can be made from one flow cell, each at around €35. These showed high spatial regularity of capture spots with few artifacts (Figures S1B and S1C) and barcode sequences with the expected structure (Figure S1D).

Open-ST allows the analysis of tissue morphology (H&E) and ST from the same cryosection (Figure 1D). Pepsin and hybridization buffer (2× saline-sodium citrate [SSC]) were combined in one solution to promote the simultaneous tissue permeabilization and RNA capture by reducing electrostatic repulsion of the single-stranded DNA and RNA molecules. A qPCR assay is

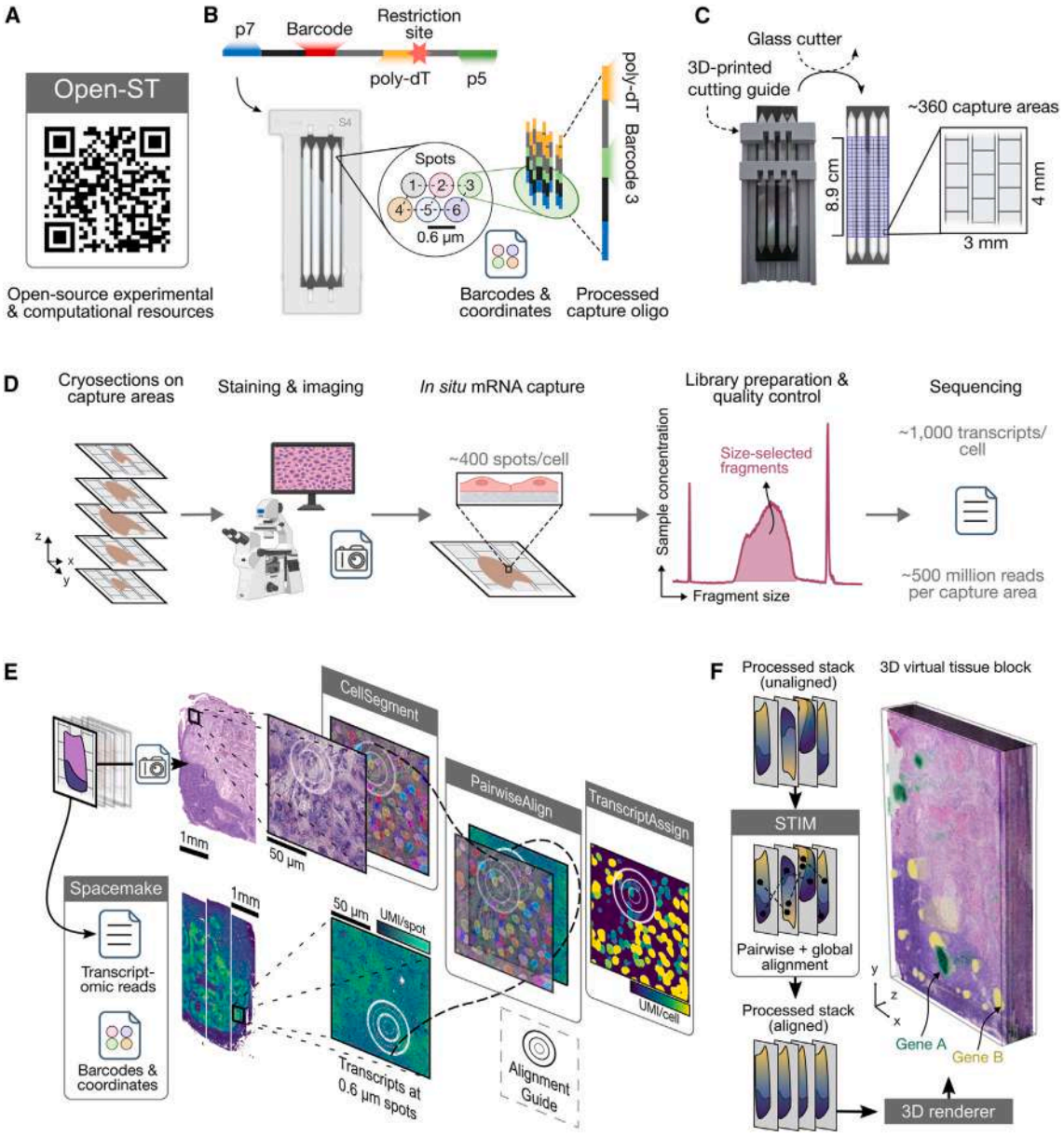


Figure 1. Open-ST workflow for high-resolution spatial transcriptomics of segmented single cells in 2D or 3D

(A) All experimental and computational resources are open-source and available under https://rajewsky-lab.github.io/openst (QR code).

(B) Sequencing designed oligos in patterned Illumina flow cells (here, NovaSeq6000 S4) allows barcode registration in regularly spaced spots (left). Oligonucleotides are processed to allow capture of polyadenylated RNA (right). Flow cell image courtesy of Illumina. p5/p7: Illumina adapters.

(C) Our custom 3D-printable device guides cutting into capture areas of desired size (STAR Methods). The shaded grid indicates the imaged area of the flow cell used for Open-ST composed of four 7 mm wide lanes (6.3 mm imaged).

(D) Transcriptomic and H&E imaging data are generated from the same fresh-frozen tissue section. Optimized RNA capture conditions and a single-amplification library preparation result in high library complexity (STAR Methods).

(E) Tissue morphological information (imaging) is integrated with ST data from single sections with our open-source openst package, including automatic cell segmentation, pairwise alignment of modalities, and quantification of transcripts in segmented cells (STAR Methods).

(F) Serial sections can be used for three-dimensional reconstruction of tissue histology and transcriptome, using STIM (left). Imaging and transcriptomics data can be visualized and interrogated as a 3D virtual tissue block using any 3D rendering engine (right) (STAR Methods). The smoothed, volumetric rendering of two genes (gene A: *S100A7*; gene B: *FDCSP*) is shown for illustration purposes.

See also Figure S1.

used to determine optimal permeabilization conditions for maximum mRNA capture. Implementation of a qPCR-based quality control during library preparation allows to determine the optimal number of PCR cycles to avoid over- or under-amplification of the library. Additionally, the one-step library amplification reduces amplification bias or sample losses due to bead purifications between PCRs (Figures S1E and S1F; STAR Methods).

The availability of morphological features from H&E staining (e.g., nuclear size and aspect ratio, cellular density, or distance between cell centroids) and molecular readouts from the same tissue section enables cell segmentation (Figure 1E). Raw H&E images are automatically preprocessed (Figure S1G) and segmented into single cells using a fine-tuned Cellpose model (Figure S1H; STAR Methods).[26] Following nuclei segmentation, radial extension of nuclei boundaries adds cytoplasmic context (Figures S1G and S1H; STAR Methods). Using the barcode coordinates from the first sequencing run, the transcriptomic reads are processed and mapped in tissue space using Spacemake.[27] Subsequently, circular marks, visible in both the imaging and spatial transcriptome modalities, are automatically detected for unsupervised pairwise alignment, resulting in registration at an accuracy of ~1 μm (Figure S1I; STAR Methods). Our segmentation and alignment protocol unbiasedly adapts to tissues with heterogeneous cell sizes and densities and automatically excludes areas without cells from downstream analyses (Figure S1H; STAR Methods). Image preprocessing and fine-tuning of the segmentation model increased the precision of the segmentation, evidenced by benchmarking against a manual segmentation (Figure S1J). This pipeline is robust to the choice of radial extension distance, alignment accuracy, and other parameters; the defaults offer a balance between the number of segmented cells, number of UMIs and genes per cell, and the accumulation of mitochondrial transcripts upon radial extension (Figures S1K–S1M). In summary, Open-ST data can be analyzed at the single-cell level and integrated with the imaging data.

Given the efficient transcript capture and integrated computational pipeline, Open-ST is suitable for 3D spatial reconstruction of any tissue. Here, we reconstruct a metastatic lymph node, integrating H&E staining and gene expression (Figure 1F). Serial sections were aligned using the Spatial Transcriptomics Imaging Framework (STIM), resulting in a 3D representation of the tissue.[28] Both imaging and transcriptomics modalities can be interrogated at once as a 3D virtual tissue block using existing software for scientific visualization.[29]

Open-ST robustly captures transcripts with high efficiency

We successfully applied our method in diverse tissues: embryonic mouse head, adult mouse hippocampus, human primary tumor (HNSCC), and the patient-matched healthy and metastatic lymph nodes.

All samples exhibited a high percentage of transcripts mapping uniquely to the genome (65%–78%, Table S1), with only a small percentage of reads consisting of ribosomal RNA (2.5%–15.3%, Table S1). On average, ~55% of reads mapping to genic regions were assigned to a spatial barcode from the first sequencing run (Table S1). Open-ST retains comparable or more spatially mapping, deduplicated reads when benchmarked against alternative solutions, consistently across samples (Figure S2A). The percentage of barcode reads from first sequencing detected in the transcriptome library was related to tissue coverage and cell density (Figure S2B). Outliers may represent segmentation errors, capture of ambient RNA in regions without cells, or areas with cells but no spatial barcodes due to capture area irregularities. Figure S2C illustrates the breakdown of the total read numbers for a representative sample after alignment. Reads were mostly assigned as coding sequences (CDSs) or 3′ UTRs and, to a lesser extent, as introns, transcription end and start sites (TES/TSS) within 10 kb of genes, and 5′ UTRs.

In a sagittal E13 mouse head section (~12 mm^2), we segmented 58,881 cells from the H&E image. Sequenced at a depth of 478 M reads, 56,627 cells (96%) contained at least one transcript, and 49,048 (83%) were high-quality (≥250 UMIs and <10% mitochondrial counts), capturing a total of 21,609 genes (Figure 2A). Open-ST's efficient RNA capture resulted in a median of 621 genes and 880 UMIs (Figure 2B; Table S2), with 42% of the high-quality cells (32% of all segmented cells) containing over 1,000 transcripts. From all captured transcripts, 82% were found inside segmented cells. Across the entire dataset, 10,000 genes account for 95% of the captured transcriptomic information. Median contributions of background transcripts were 15.4% for ribosomal proteins and 3.5% for mitochondria-encoded transcripts (Figure S2D). At the same time, Open-ST resulted in rich quantification of all genes, ranging from highly abundant transcripts (e.g., *mt-Cytb*) to lowly expressed but highly specific gene markers (Figure S2E). Thus, a notable portion of detected genes is well represented across all cells.

We assessed the library complexity of Open-ST data in comparison with existing sequencing-based ST methods by calculating UMI counts per pseudo-cell (100 μm^2) relative to sequencing depth (Figure 2C; STAR Methods). We selected prototype datasets generated by the respective developers of the methods and processed all raw data with the same computational pipeline (STAR Methods). We further selected similar tissue types for comparison wherever feasible; however, we acknowledge that the analysis may still be influenced by the specific choice of tissue analyzed. Open-ST consistently outperformed alternative solutions in capture efficiency for the same sequencing depth with the exception of 10× Visium, which demonstrated comparable performance. Capture efficiency was similar across Open-ST processed samples, despite their diverse cellular composition. Biological reproducibility of Open-ST is supported by the comparable capture efficiency (UMIs/100 μm^2) of two independent E13 mouse head libraries (Figure S2G). To further evaluate library complexity across technologies, we benchmarked the reads-to-UMIs ratio as a function of genic reads (Figure S2F, STAR Methods). Open-ST consistently showed the lowest ratio and scaled best with increasing read depth, together with Slide-seqV2. A low reads-to-UMIs ratio results from high initial transcript capture and efficient library amplification and is vital for extracting new information through deeper sequencing. Since sequencing is the largest cost per sample, Open-ST emerges as an affordable and highly efficient solution.

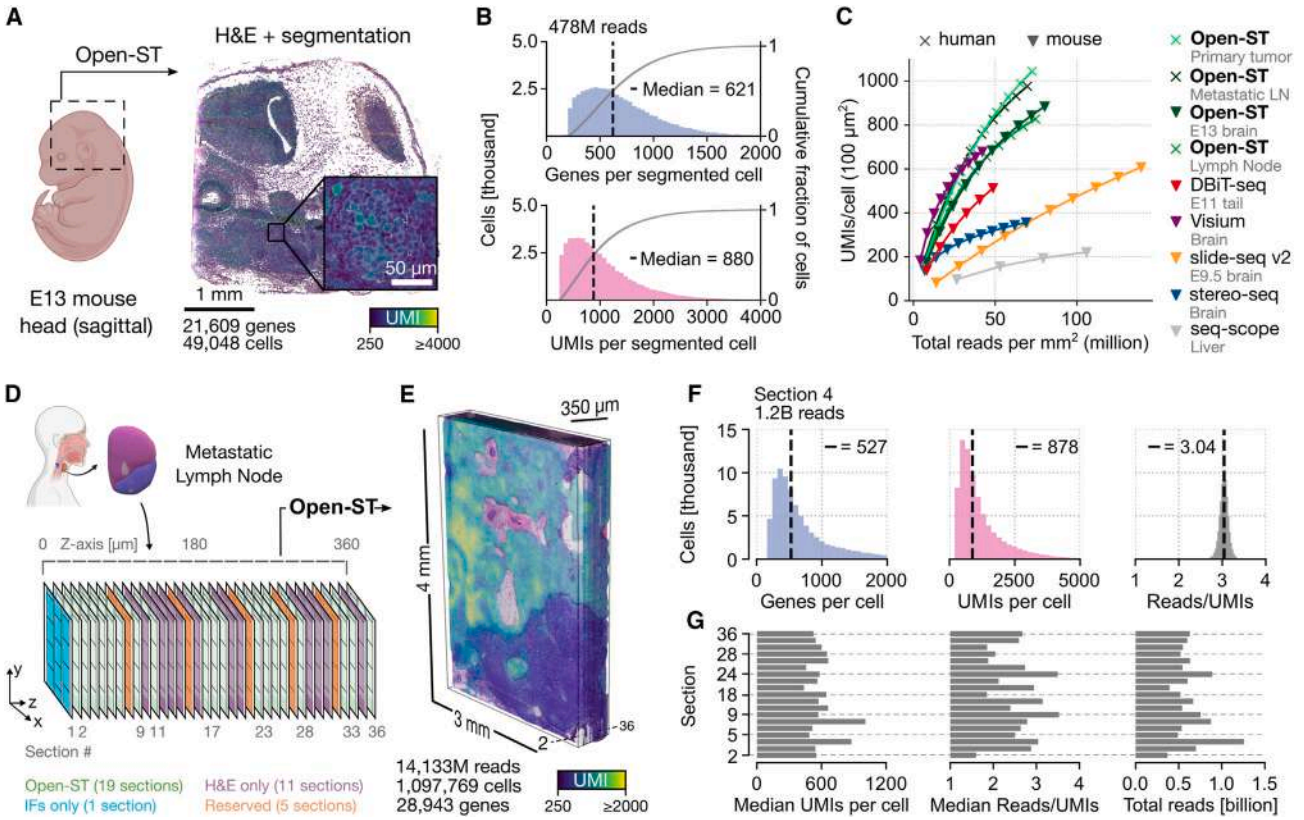


Figure 2. Open-ST robustly captures transcripts with high efficiency

(A) Spatial distribution of UMI counts per segmented cell in an Open-ST processed E13 mouse head sagittal section, overlaid on H&E staining of the same tissue.

(B) Distribution of genes and UMIs per segmented cell, and sample as in (A).

(C) Open-ST produces the highest transcript capture per pseudo-cell (UMIs/100 μm^2) for the same number of sequencing reads compared with alternative sequencing-based ST methods (STAR Methods).

(D) Experimental design for processing a human metastatic lymph node. 10 μm sections were profiled with Open-ST, stained with H&E or immunofluorescence (IF) only, or reserved for validations.

(E) Aligned transcriptomic and imaging data, and rendering of transcript capture colored by UMI counts is overlaid on a 3D virtual tissue block (STAR Methods).

(F) Distributions of UMIs and gene counts per segmented cell for section 4. Median values are indicated.

(G) Total reads, median UMIs, and median reads/UMIs per segmented cell across all 19 sections.

See also Figure S2.

Open-ST generates 3D virtual tissue blocks

To demonstrate robustness and reproducibility of Open-ST across multiple sections in a clinically relevant complex tissue, we processed a human metastatic lymph node. In this proof-of-principle, large-scale experiment, we obtained 10 μm sections spanning 350 μm of tissue depth. Our experimental design consisted of profiling sections with Open-ST (19), H&E staining only (11), immunofluorescence (1), and reserving for validation (5) (Figure 2D). In total, we obtained gene expression profiles of over 1 M cells across 19 sections, with high median capture of genes (313–624) and UMIs (438–1,008) per segmented cell (Figure S2I; Table S2). The samples demonstrated a similar percentage of reads aligning to rRNA (<15%) and mapping uniquely to the genome (>65%) (Figure S2J; Table S1).

By leveraging STIM and the Visualization Toolkit, we aligned the serial sections processed with Open-ST to reconstruct and visualize a 3D virtual tissue block, a computational object combining transcriptomics and histological modalities (Figure S2K).[28,30] A smooth isosurface rendering of the spatially mapped UMI counts indicates reproducible capture efficiency with visible transcript enrichment in the tumor compartment (Figure 2E). Section 4 was deeply sequenced (1.2B reads), exhibiting low reads/UMIs (median = 3.04) and resulting in a median of 527 (878) genes (UMIs) per segmented cell (Figure 2F). The reads-to-UMIs ratio per 100 μm^2 was consistently low and comparable across the 19 sections, with the exception of a single outlier, demonstrating the technical reproducibility and cost-effectiveness of Open-ST (Figures 2G and S2H).

Taken together, Open-ST robustly captures transcripts across serial sections, making it well-suited for high-throughput studies and 3D-transcriptome reconstruction of tissues.

Open-ST locally captures marker genes with high accuracy

We corroborated localized transcript capture of Open-ST by exploring the spatial organization of transcriptomic clusters

and marker genes in the E13 mouse head. Clusters were annotated using a developmental mouse atlas as a reference (Table S3; see STAR Methods for what follows in this section).[31] Coarse clustering of segmented Open-ST data reflects the major anatomical regions in the embryonic mouse head (Figures 3A and S3A). Spatial expression of selected marker genes per cluster co-localizes with the respective clusters (Figures 3B and S3B).

We identified three neuronal clusters, each characterized by *Tbr1*, *Lhx9*, or *Nefl* expression. A radial glia cluster, with cells mainly located in the dorsal forebrain, was identified by *Dmrta2*. The eye's retina is distinguished by localized expression of *Atoh7*, and the choroid plexus is characterized by *Ttr* expression. Further clusters include fibroblasts (expressing collagens), chondrocytes (*Col11a1*, *Sox9*, and *Hapln1*), myocytes (*Actc1* and *Mylpf*), mesenchymal (*Tsc22d1*), and endothelial cells (*Cd34* and *Kdr*). Coarse clustering revealed a group of cells with no identifiable markers. On these, we performed subclustering and identified groups of cells that we termed neural crest (*Mdk*), craniofacial development (*Barx1*, *Sp7*), musculoskeletal development (*Kctd12*, *Igf12*), immune (*Selenop*, *Tyrobp*, *C1qc*, *Lyz2*, *C1qb*), and early vascular (*Akap12*, *Gja1*, *Cldn11*). "Blood" was identified based on high expression of hemoglobin genes (e.g., *Hbb-y*). A cluster highly expressing the cell cycle gene *Ccnd2*, with transcriptomic distance closest to the neuron clusters, was termed "cycling." This cluster was in spatial proximity to the blood cluster, which could explain the contaminating expression of hemoglobin genes (Figure S3B).

Open-ST's capability to capture rich transcriptomic information allowed us to achieve more refined clustering of the segmented cells of the fore-, mid-, and hindbrain (Figures 3C, S4B, S4D, S4E, S4G, and S4H). Cell type diversity and spatial distribution of cells in the forebrain resembled that obtained by integrating and transferring cell type labels from a single-cell reference atlas (STAR Methods) (Figures 3C and S4A). This congruence extends to marker genes identified from each unbiased cluster in our Open-ST data, consistent with the respective populations in the reference (Figure S4A). Clustering of segmented Open-ST data provides labels akin to those obtained through clustering of single-cell RNA sequencing (scRNA-seq) data. Furthermore, separate clustering of an independent E13 mouse head replicate resulted in similar transcriptomic identities (Figure S4J), further confirmed by joint embedding and integration of both E13 datasets (Figure S4K). Most cell type clusters were present in both replicates and showed similar mixing before and after integration. Other clusters, such as "retinal progenitors," were not present in the biological replicate as this structure fell outside the capture area.

To investigate capture localization, transcripts were plotted on the H&E as a virtual *in situ*, with intensity relating to gene expression, resulting in a comparable stain with chromogenic *in situ* hybridization (ISH) staining from the Allen Developing Mouse Brain Atlas.[32,33] The expression of select marker genes showed high accordance with ISH (Figures 3D–3F, S4C, S4F, and S4I). For example, regionalized expression of the transcription factors *Neurod6* and *Pbx3* could be detected in the forebrain (Figure 3D).[34] As expected, the pan-neuronal marker *Tubb3* was absent in the choroid plexus while being detected in the hindbrain (Figure 3E).

Quantifying the transcript density of *Ttr*, in the choroid plexus, and *Atoh7*, in the retina, we observed a corresponding sharp rise and fall of signal intensity (Figures 3F and S3G). The majority of these transcripts were captured in regions covered by tissue. Transcripts observed in regions without tissue may be due to inaccuracies in manually defining the tissue boundary, or local lateral diffusion. For *Ttr*, the transition between minimum to maximum occurred within the diameter of a cell. *Atoh7* exhibited a less sharp boundary, which may be an effect of a wider region being sampled due to overall sparser expression and/or detection.

We further performed a crosstalk analysis to assess sources of spatial biases that may impact our interpretation (STAR Methods; Figure S3D). Focusing on blood and "chondrocyte" clusters, this analysis highlighted increased gene marker mixing at cell pairs within 20 μm proximity, underscoring the impact of local biases in clustering (Figures S3E and S3F).

Local capture of Open-ST reflects nuclear-cytoplasmic cell architecture

To assess the theoretical subcellular resolution of Open-ST based on the spots' density, we processed a coronal section of an adult mouse hippocampus hemisphere (Figure 3G).

Compared with other Open-ST datasets, this tissue exhibited lower reads-to-UMIs ratio (Figure S3H; Table S2), likely due to overall lower cell density resulting in a reduced RNA input across the capture area. The tissue's variety of cell densities, however, rendered it suitable for assessing nuclear/cytoplasmic enrichment. Visualization of nuclear-retained *Malat1* and mitochondria-encoded transcript counts within the nuclear and cytoplasmic regions of the cell segmentation mask showed accurate transcript localization (Figure 3G). Average distribution of these transcripts in nuclear and cytoplasmic compartments was calculated for different cell densities across 1,791 regions (5,000 μm^2 each) and projected onto a standardized cell (Figures 3H and S3I; STAR Methods).

Malat1 was significantly enriched in the nucleus ($\log_2$(odds ratio [OR] cytoplasmic/nuclear) = −1.24, $p_{Fisher} < 0.01$), whereas mitochondrial transcripts were enriched in the cytoplasm ($\log_2$(OR cytoplasmic/nuclear) = 0.20, $p_{Fisher} < 0.01$). This was true across regions with variable cellular densities, while no significant enrichment was detected when offsetting the segmentation mask by 5 μm (Figures S3J and S3K). Nuclear enrichment remained consistent when assessing the distribution of UMIs relative to their distance to the segmented nuclear edge but was depleted after offsetting the pairwise-aligned coordinates. In summary, localized RNA capture on the high-density capture array reflected nuclear-cytoplasmic cell architecture.

Open-ST captures spatial cell-type complexity in human primary tissues

To showcase the ability of Open-ST to dissect transcriptomic diversity in structurally complex tissues, we spatially sequenced a primary HNSCC tumor and a healthy and metastatic lymph node from a single patient.

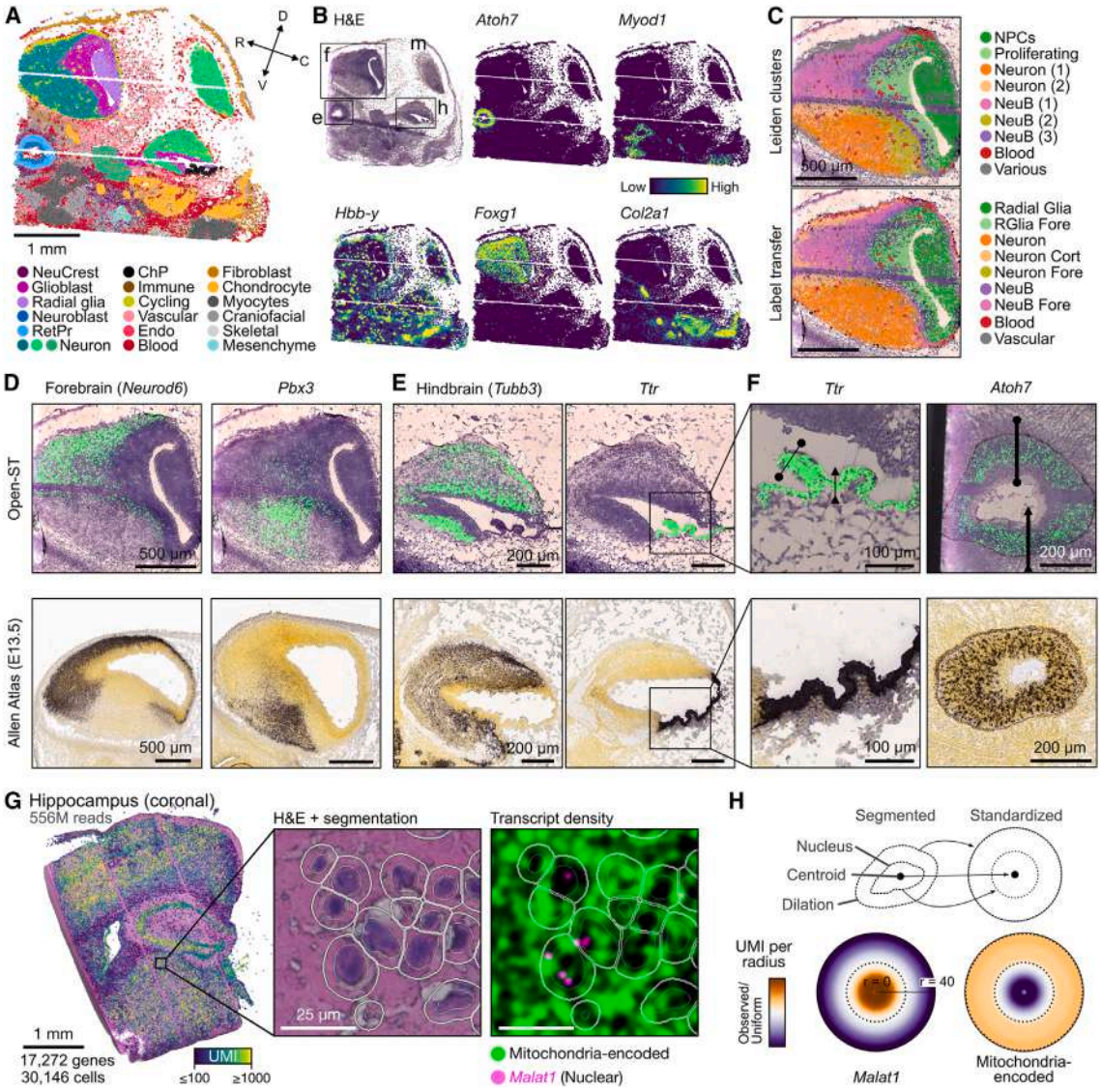


Figure 3. Open-ST captures transcripts at subcellular resolution

(A) Spatial distribution of cell types of an E13 mouse head sagittal section. D, dorsal; V, ventral; R, rostral; and C, caudal.

(B) Annotation of forebrain (f), midbrain (m), hindbrain (h), and the developing eye (e) on the H&E, and localization of selected marker genes colored by normalized expression.

(C) Spatial distribution of E13 mouse forebrain cells colored by annotated Leiden cluster (top) or by label transfer from an E13.5 single-cell reference atlas (bottom) (STAR Methods).[31]

(D) Localized gene capture of transcription factors *Neurod6* and *Pbx3* in the E13 mouse forebrain profiled with Open-ST (top) compared with *in situ* hybridization (ISH) images from an E13.5 mouse from the Allen Atlas (bottom). High expression is colored in green (black) for Open-ST (Allen Atlas).

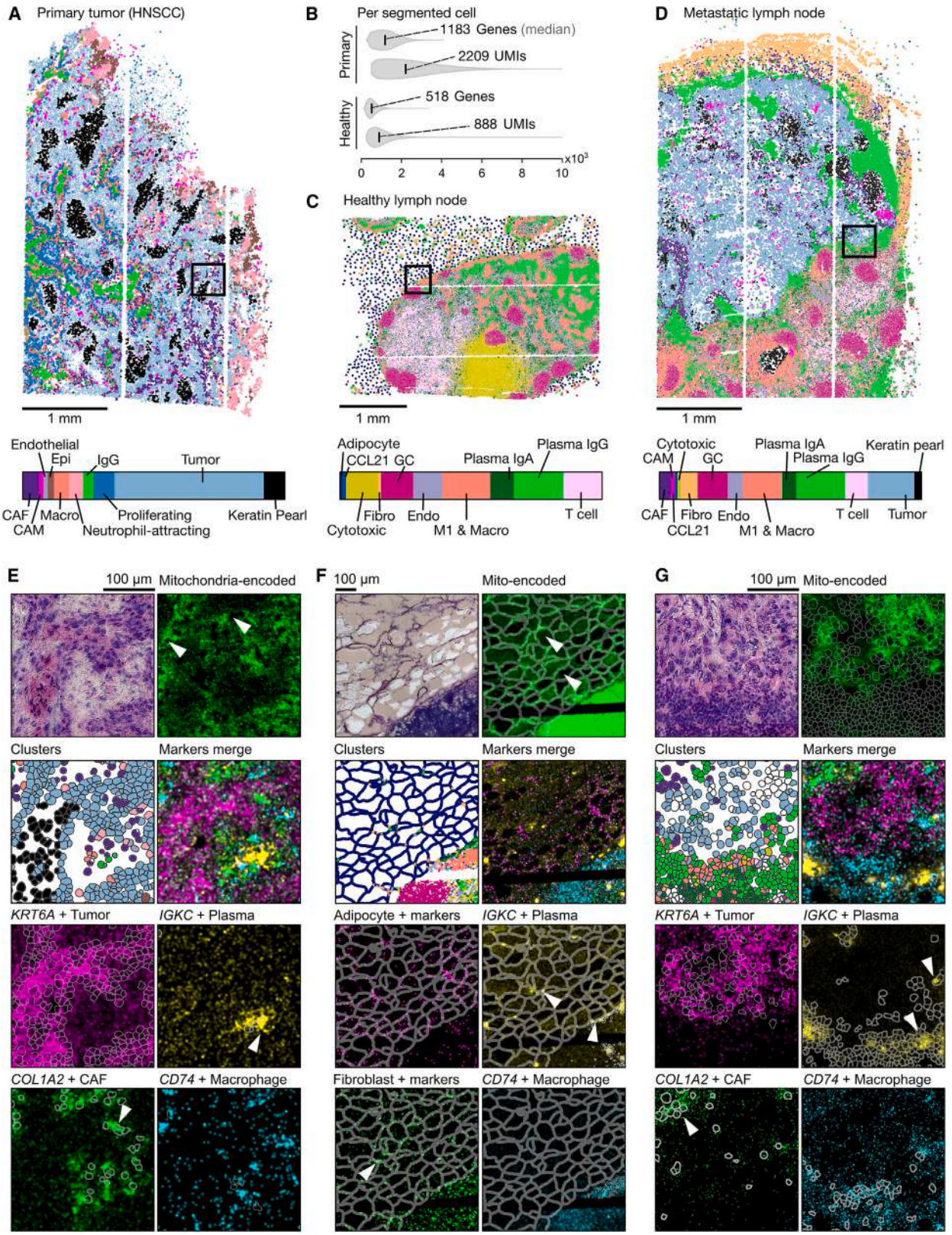
(E) Localized gene capture in the E13 mouse hindbrain (*Tubb3*) and choroid plexus (*Ttr*).

(F) Transcript density for *Ttr* in the choroid plexus and *Atoh7* in the developing eye; top: transcript density in space is visualized as a 2D Kernel density estimate of the spot coordinates weighted by transcript counts; bottom: ISH images (E13.5, sagittal) from the Allen Atlas.

(G) Subcellular transcript capture precision in an adult mouse coronal hippocampus section. Left: spatial distribution of UMIs per segmented cell reflects tissue morphology; center: close up of H&E segmentation with thin (thick) white lines indicating nuclear (cell) boundaries; right: nuclear-enriched *Malat1* and mitochondria-encoded transcript density plotted onto the segmentation mask.

(H) *Malat1* and mitochondria-encoded transcripts are enriched in the nucleus and cytoplasm, respectively. Observed/uniform UMI per radius (n_{radii} = 40), visualized on a standardized cell, quantified over 22,376 cells sampled from regions where 15%–30% area is covered by nuclei. Color mapping of expression is clipped between the 5th and 95th percentile of values (STAR Methods).

See also Figures S3 and S4.



(legend on next page)

Annotation of cell types from unbiased clusters was performed based on literature-informed marker genes,[24] and refined by jointly clustering the lymph node tissues of the metastatic and healthy samples (STAR Methods). In all three datasets, we identified signatures corresponding to the transcriptional diversity of tumor, stroma, and immune populations (Figures 4A, 4C, 4D, and S5A–S5C; Tables S4, S5, and S6), whose spatial neighborhood enrichments matched the annotations independently outlined in H&E images by a pathologist (Figures S5A–S5D, S5F, and S5H). Median number of genes and UMIs per cell in the tumor-enriched primary sample was higher than in the healthy lymph node (Figure 4B). Annotation of segmented cells matched the abundance of marker genes in space (white arrowheads, Figures 4E–4G).

The expression profile of tumor cells reflected its epithelial origin, with high expression of keratins (*KRT6A*, *KRT5*), cell adhesion (*ITGB6*), and desmosomal (*DSG3*) markers (Figures S5A and S5C; Tables S4 and S6). Additionally, "keratin pearls" (KPs), which characterize squamous cell carcinomas, were found in both the primary and metastatic tumor tissues. Neighborhood analysis showed the vicinity of these cells, in line with their known formation of compact and distinct keratinized structures in the tumor bed (Figures 4A, 4D, S5A, and S5C). These expressed the same epithelial markers found in tumor cells, but also high levels of cornification-related molecules (*S100A10*) and desmosome markers (*DSP*), compatible with the activation of keratinocyte differentiation pathways (Figures S5A and S5C; Tables S4 and S6). In the primary tumor, a cluster transcriptionally similar to the tumor and KP was defined as "proliferating," with marker genes (*RRM2*, *HMGB2*, *CDCA5*) implicated in cell cycle regulation and/or DNA repair (Figures S5A and S5E). Furthermore, tumor cells and cancer-associated macrophages (CAMs) from both primary and metastatic tissues expressed *FABP5*, suggesting a dysregulation in the lipid metabolism of the cancer (Figures S5A and S5C).

Cancer-associated fibroblasts (CAFs) (*COL1A1+*) were identified in the microenvironment of both tumor tissues, with a subset in the metastatic tumor expressing high levels of *ACTA2*, suggesting a myofibroblast phenotype.[35] Endothelial cells (*PECAM1*, *ACTA2*) were identified in the primary tumor, while endothelial cells of the lymph nodes expressed *ACKR1*, a marker for post-capillary venules, *CCL21*, a chemokine expressed in lymphatic endothelial cells, and *PECAM1* in the metastatic lymph node (Figures S5A–S5C).[36]

Macrophages (*LYZ+*) were detected in all tissues. Pro-inflammatory M1 macrophages were identified in the metastatic lymph node, with high *TIMP1* expression in a subset of the cells. CAMs exhibited strong *SPP1* expression, coding for a secreted phosphoprotein implicated in macrophage polarity previously associated with negative human papillomavirus status and poor prognosis in HNSCC.[37] This cluster frequently colocalized with tumor cells, highlighting their interactive potential in the tissue microenvironment. A cluster defined as neutrophil-recruiting, due to high expression of *CXCL8*, was detected in the primary tumor, with *CXCL8* expression also present in the epithelial and CAM clusters (Figure S5A).[38] Within the metastatic lymph node, *CXCL8* was expressed in the CAM population, colocalizing with the tumor and KP cluster (Table S6; Figures 4D, S5C, S5H, and S5I). *CCL21*, with known expression in lymphatic endothelial cells directing the migration of T cells and dendritic cells, was broadly detected in the lymphatic tissues (Figures S5B and S5C). Detection of *CCL21* in the T cell, macrophage, adipocyte, and plasma cell clusters may represent crosstalk resulting from close physical proximity of these cell types in every dimension. We acknowledge that the annotation may relate to domains enriched in specific cell types in the case of "double-positive" artifacts rather than pure cell type labels.

Within both lymph nodes, we identified T cells (*CD3E+*) and a cluster additionally expressing *IRF8* as interferon-responsive cytotoxic T cells. We also detected multiple germinal centers. These exhibited localized expression of *FDSCP*, a protein secreted by follicular dendritic cells, and *IGHM*, suggesting an early germinal center status (Figures 4C, 4D, S5B, S5C, S5G, and S5H).[39,40] Furthermore, we classified two clusters with strong *IGHG3* or *IGHA1* expression as immunoglobulin IgG or IgA plasma cells, respectively.

White adipose tissue visibly surrounded the healthy lymph node in the annotated H&E images (Figure S5F). In this sample, our segmentation strategy was adapted to retain the transcriptomic profiles of large unilocular adipocytes and smaller cells of the stromal vascular fraction (STAR Methods). Unbiased cell type markers characterization identified immune (e.g., *IGKC*-positive plasma cells) and stromal cells (e.g., collagen-expressing fibroblasts) within the adipocyte population (*FABP4*, *ADIPOQ*, *PLIN1*, *SORBS1*, *CIDEC*, and *SIK2*) (Figures 4F, S5B, and S5G).[41,42] High expression of stromal markers within adipocytes is likely an artifact of library-size normalization, which can be mitigated using cell area/volume normalization (Table S5).

Figure 4. Open-ST accurately dissects cell-type complexity in primary human tissues

(A) Spatial distribution of cell types (top) and cluster proportions (bottom) in a primary head and neck squamous cell carcinoma (HNSCC) section.

(B) Distributions of UMI and gene counts per segmented cell in the primary HNSCC and healthy lymph node sections.

(C) As in (A), but for the healthy lymph node.

(D) As in (A), but for the metastatic lymph node. Spatial plots (A, C, and D) visualize cell centers.

(E–G) Expression of selected marker genes in space for the primary HNSCC (E), the healthy lymph node (F), and the metastatic lymph node (G). Top left: H&E staining; top right: expression of mitochondria-encoded transcripts; second row: segmentation mask colored by cluster (left) and gene expression of a marker from every cluster (right); third and fourth rows: gene expression of selected markers for different clusters, together with the corresponding segmentation masks. Pseudoimages show the smoothened expression of the indicated marker genes at spot resolution for the respective cell types (STAR Methods). In (F) expression of all collagen genes is shown in fibroblasts and for 11 specific markers in "adipocyte+," i.e., *ADIPOQ*, *LEP*, and *CIDEC*, among others. In the colored segmentation masks, white cells represent those not passing filters (STAR Methods) or not annotated. CAF, cancer-associated fibroblast; CAM, cancer-associated macrophage; Epi, epithelial; IFNr, interferon production regulator; M1, M1 macrophage; Macro, macrophage; Fibro, fibroblast; Endo, endothelial; GC, germinal center.

See also Figure S5.

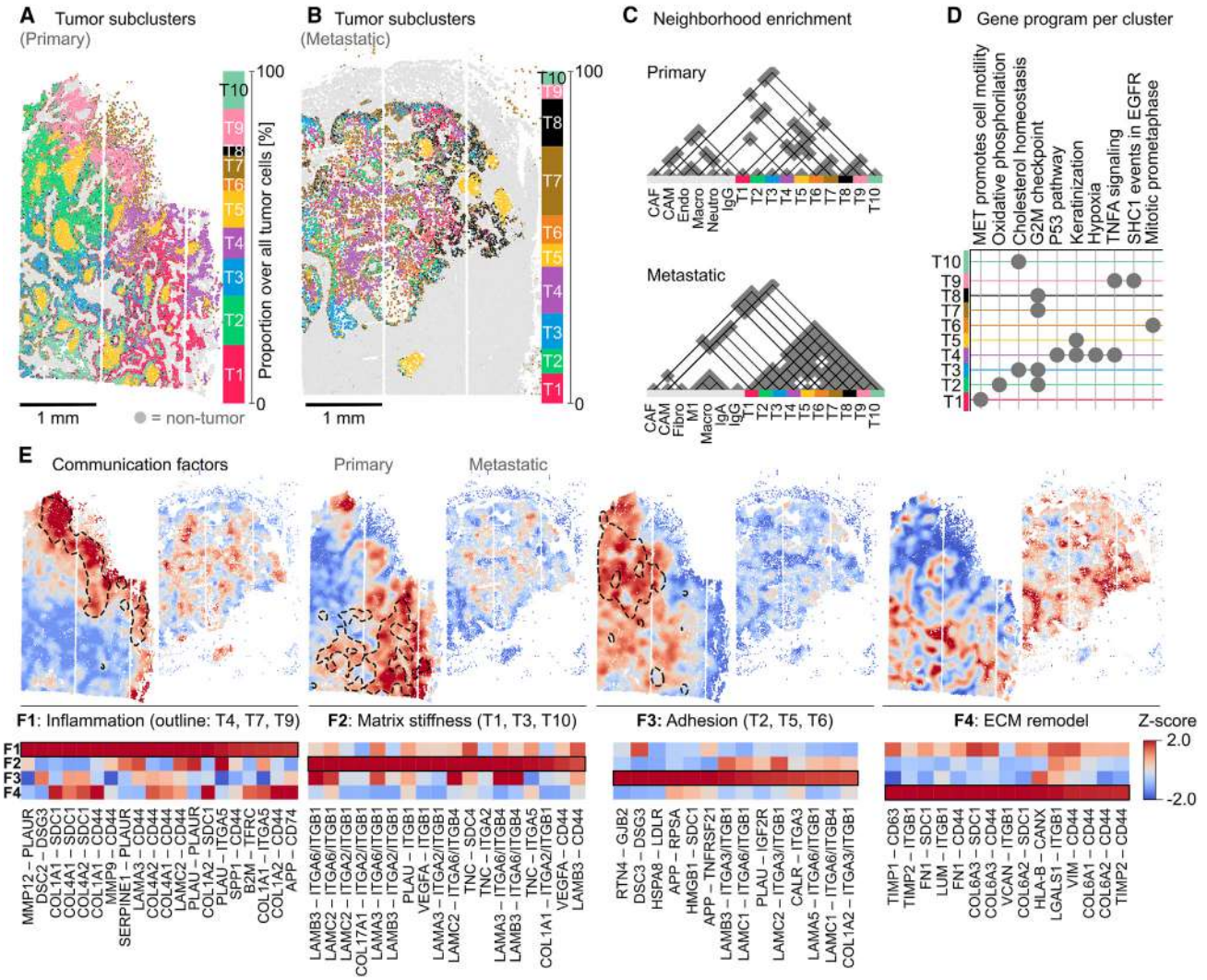


Figure 5. Transcriptomic tumor heterogeneity is organized into spatial domains with distinct communication signatures

(A) Spatial and transcriptomic heterogeneity of tumor cells across the primary tumor tissue.

(B) As in (A), but in the metastatic tissue.

(C) Spatial neighborhood enrichment of tumor and stroma populations. Significant spatial interactions (permutation test p value < 0.05) between cells belonging to a pair of clusters are depicted by gray triangles and connected by solid lines (STAR Methods).

(D) Normalized enrichment score (NES) in the tumor subclusters for the primary tumor tissue highlights differentially active gene programs. Dots are shown for cases where NES > 1 and false discovery rate (FDR)-adjusted p value < 0.05 (STAR Methods).

(E) Cell-cell communication is organized as spatial motifs discovered with non-negative matrix factorization, and their localization corresponds with the observed tumor cell heterogeneity in the primary and metastatic samples. Top: spatial visualization of factor contributions after z-normalization; bottom: min-max normalized communication score for top representative ligand-receptor pairs for each factor (STAR Methods).

See also Figure S6.

Taken together, Open-ST accurately dissected the complex primary human tissue composition.

Open-ST reveals spatially constrained heterogeneity in primary and metastatic HNSCC

Building on our exploration of coarse cell types, we investigated the transcriptomic heterogeneity of tumor cells in the primary HNSCC and metastatic lymph node. By integrating and clustering 42,132 tumor cells from both samples, we identified 10 distinct transcriptomic states (T1–10), providing a detailed map of their heterogeneity, spatial distribution, and stromal interactions (Figures 5 and S6) (STAR Methods).

The tumor transcriptomic states showed distinct spatial organization across the primary and metastatic samples. In particular, tumor states in the primary tumor were organized into continuous spatial domains prevalent for a specific state (Figure 5A). By contrast, in the metastatic tissue, tumor states appear more intermixed without clearly defined boundaries (Figure 5B), validated via neighborhood enrichment analysis (Figure 5C; STAR Methods).

We further assessed the transcriptomic heterogeneity of these spatially organized tumor states via differential gene expression and gene set enrichment analysis (STAR Methods). We detected state-enriched gene programs, with non-random spatial organization across the two tissues (Figures 5D and S6B–S6E; STAR Methods). For instance, in the metastatic lymph node, we detected increased cell cycle activity at the tumor-lymphoid boundary (T2, T3, T6, and T8) and around a region linked to hypoxic response (T4 and T7). Notably, the cholesterol biosynthesis program (*HMGCR*, *HMGCS1*, *DHCR7*, and *DHCR24*) was spatially restricted to the tumor side of the tumor-lymphoid boundary (Figures S6E and S6F; STAR Methods). On the other hand, the lymphoid (peritumoral) side of the boundary was enriched in cells expressing markers of immune activation (*GZMA*, *GZMB*, *LYZ*, *CXCL9*, and *CXCL10*) (Figure S6G).

To characterize the interplay between communication and transcriptomic identity, we performed unsupervised discovery of spatial hotspots for receptor-ligand interaction, based on their expression levels (STAR Methods). This resulted in four non-negative factors across primary tumor and metastasis (Figures 5E and 5F), each with unique interaction signatures: inflammation (F1), matrix stiffness (F2), adhesion (F3), and extracellular matrix (ECM) remodeling (F4). These factors displayed higher degree of spatial organization in the primary tumor compared with the metastatic, following a more intermixed organization of tumor states.

Different tumor states were restricted to specific cell-cell interaction hotspots. In the primary tissue, T1 was found in a CAF-rich stroma and showed cell motility phenotype (*LAMC2*^{hi}/*PTHLH*^{hi}/*ITGB1*^{hi}) (Figure S6A), as well as spatially restricted communication events related to matrix stiffness and migration, e.g., integrin and syndecan-mediated communication (Figure 5E), similar to a previously reported partial epithelial-mesenchymal transition (pEMT) subcluster of HNSCC *LAMC2*^{hi}/*TGFBI*^{hi}/*SNAI2*^{hi} tumor cells.[24] This set of tumor cells was the only one with a spatial enrichment with CAFs in the primary tumor, which subsequently surrounded endothelial cells. T2 showed a predominant proliferation phenotype. It was enriched within a communication hotspot for cell adhesion mediated via desmoglein-desmocollin interactions, in the primary tumor only showing spatial interactions with tumor cells (T6, T8). T3 also had a marked proliferation signature, additionally showing metabolic phenotypes, particularly cholesterol biosynthesis, as also enriched in T10. In the primary tumor, T3 localized to the communication hotspot for matrix stiffness, surrounded by a microenvironment only composed of other tumor cells, remarkably T6 with similar proliferation phenotype, which neighbored macrophages. T4 and T9 showed phenotypes involving tumor necrosis factor alpha (TNF-α) signaling, hypoxia, and other stress-induced pathways and localized near receptor-ligand interaction hotspots involving metalloproteases (*MMP9-CD44* and *MMP12-PLAUR*) and osteopontin (*SPP1-CD44*), processes that collectively can contribute to inflammation (Figures 5D, 5E, and S6C).

While T9 localized with CAM in both primary and metastatic lymph nodes, T4 was in close proximity to the *CXCL8*$^{+}$ neutrophil-attracting populations in the primary tumor and to the *SPP1*$^{+}$ *CXCL8*$^{+}$ CAM population in the metastatic lymph node (Figure 5C). *SPP1*$^{+}$ CAMs were previously identified to be enriched in areas of hypoxia in HNSCC, with *in vitro* studies indicating that hypoxia promotes *SPP1* expression.[43]

T5 displayed significant keratinization signatures and activity of the desmosome pathway, a known signaling mode for cell adhesion. T7 and T8 were sparse in the primary tissue and more prevalent in the metastasis. T10 was abundant and spatially organized in the primary tissue, in close proximity to cornified cells (T5) and CAMs, showing a cholesterol biosynthesis phenotype similar to T3. Lastly, we globally detected differentially expressed genes between the two samples across all tumor cells: a strikingly strong *AMTN* expression was identified exclusively in the metastatic lymph node, while *FGFBP1* and *PRR9* were expressed in the primary tissue with minimal detection in the metastasis (Table S7) (STAR Methods).

In summary, Open-ST reveals diverse tumor cell states coexisting in close proximity within the primary tumor in spatially restricted domains whose organization closely follows that of cell-cell communication hotspots.

Benchmarking Open-ST against imaging-based spatial transcriptomics

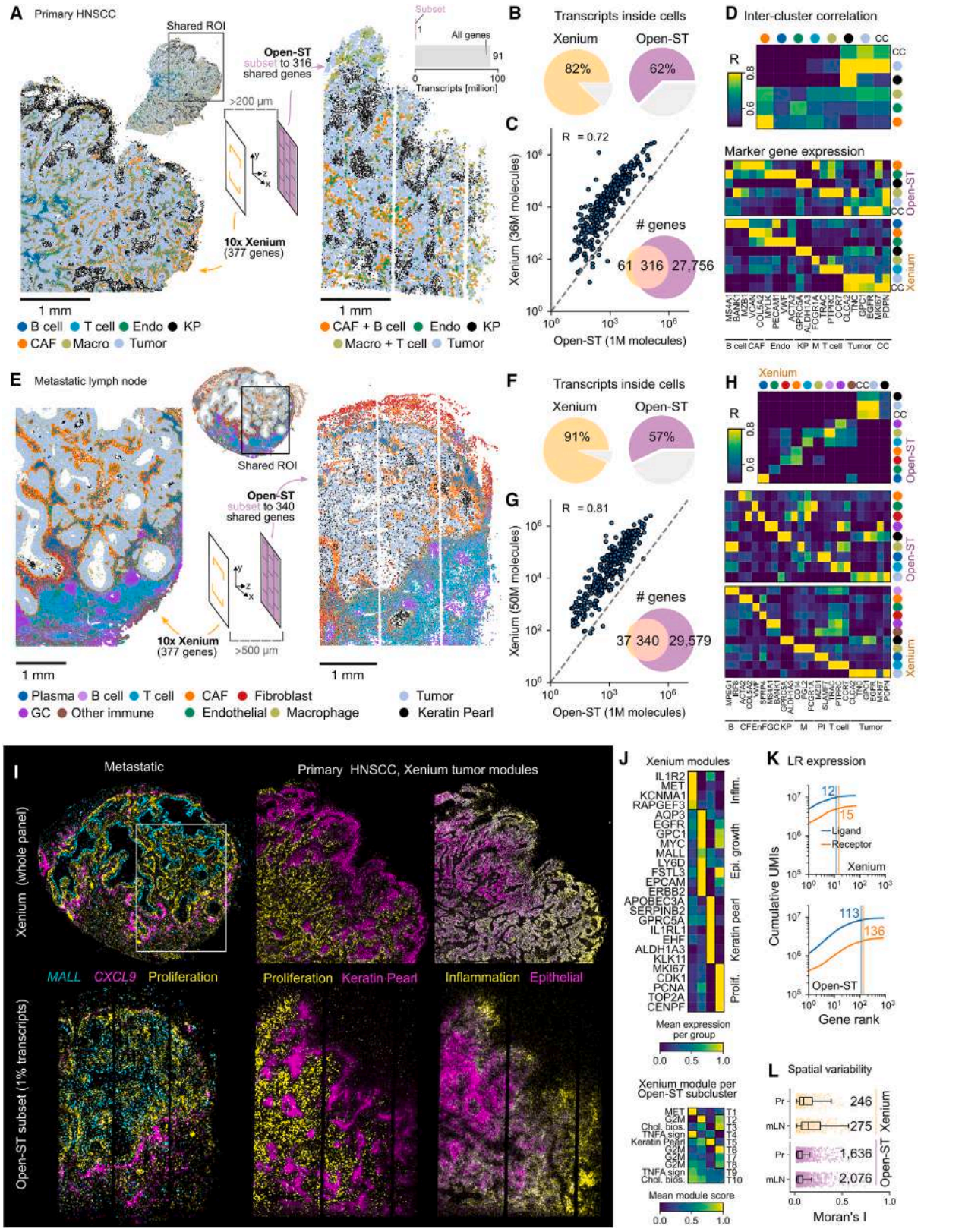
We benchmarked Open-ST against an imaging-based ST method (Xenium *In Situ*, 10× Genomics) on the human primary HNSCC and patient-matched metastatic lymph node.

We used the predesigned human multi-tissue and cancer probe panel (377 genes), mimicking a scenario with little prior knowledge of the tissue's molecular composition. For comparison, we subset the primary HNSCC and metastatic lymph node Open-ST datasets from >20,000 genes to the 316 and 340 genes shared with the Xenium panel, respectively, reducing from ~100 to ~1 M transcripts per sample. By contrast, Xenium provided 36 and 50 M transcripts for the same gene set, a result of the targeted and sensitive nature of imaging-based assays.

After subsetting the Open-ST data, we performed clustering and annotation based on marker genes separately for both technologies (STAR Methods). Both resolved cell types of similar diversity and spatial distribution across both samples, albeit being more than 100 μm apart in the sectioning axis (Figures 6A and 6E). Clusters independently annotated with the same identity across modalities showed the gene expression largely agreed (Figures 6D and 6H). However, in the (subset) Open-ST primary HNSCC dataset, CAFs could not be clustered separately from B cells, nor macrophages from T cells, likely a consequence of the low capture of B and T cell markers shared with the Xenium panel.

With a similar segmentation approach (nuclear segmentation + 3.45 μm radial extension) both methods yielded ~5,200–5,500 cells per mm^2. The proportion of transcripts detected inside cells was ~20% less in Open-ST compared with Xenium (Figures 6B and 6F), regardless of subsetting. Despite these differences in signal-to-noise, the total expression of all genes across the whole sample was highly correlated (Figures 6C and 6G). Furthermore, Open-ST with and without subsetting yielded similar classification and spatial distribution of coarse cell types.

We validated the spatial localization of individual genes or gene sets across tissues and technologies (Figure 6I). In the metastatic lymph node, we confirmed the layering of *CXCL9* expression in proximity of the tumor-lymphoid boundary, proliferation



(legend on next page)

markers (*MKI67*, *TOP2A*, *CENPF*, *UBE2C*, and *CDK1*) at the edge of the tumor mass, and we detected *MALL* expression surrounding areas devoid of cells within the tissue, i.e., necrotic areas, similar to the hypoxia spatial signature (Figure S6E).

Lastly, we compared the extent to which Xenium and Open-ST can identify finer transcriptomic identities of tumor cells in this patient's samples. Subclustering of the tumor cells from both Xenium datasets identified three major classes, in addition to the already separate KPs: (1) inflammatory response (*IL1R2*, *MET*, and *KCNMA1*), (2) epithelial growth (*EGFR*, *ERBB2*…), and (3) proliferating (*CENPF*, *TOP2A*, *MKI67*, *UBE2C*, and *CDK1*) (Figures 6I and 6J). Some of the ten Open-ST tumor subclusters, previously identified on the full data (Figure 5), were separable with respect to these marker genes (Figures 6I and 6J). Having the additional information of >20,000 genes, Open-ST identified more marker genes, enabling downstream tasks such as gene set enrichment analysis (Figure 5D), which covered more diverse functionality, i.e., genes that can be used as proxies for receptor-ligand activity (Figure 6K), and a higher number of genes showing significant spatial variation, assessed by Moran's I values (Figure 6L), with their ratio (in respect to total genes) being lower in Open-ST compared with Xenium, as the latter uses a probe panel optimized for capturing genes showing spatial variability in human tissues (Figure 6L).

In summary, Open-ST yields spatial gene expression and transcriptomic clusters comparable to an imaging-based, targeted approach.

Exploring 3D virtual tissue blocks: Cell types, gene programs, and receptor-ligand interactions in 3D

Open-ST enables the interrogation of gene expression in 3D via volumetric renderings and smoothed representations for enhanced clarity (STAR Methods). As an example, we visualize the lymphoid, metastatic tumor, and stroma tissue compartments via the expression of selected marker genes in 3D (Figure 7A).

These representations of gene expression coincide with the surface renderings generated from transcriptomic clusters, obtained after transferring the annotations from section 4 to the remaining 18 (STAR Methods). We classified over 1 M cells into 14 major types, spanning tumor, immune, and stromal compartments. The spatial distributions and proportions of cell types in 3D were consistent across consecutive sections (Figures S7A and S7B). Our transcriptome annotation was also in agreement with the manual pathologist annotation of the most distant section 36 (Figure S7C), which showed high correlation to the gene expression from section 4, displaying reproducible RNA capture throughout the tissue block (Figure S7D). All sections displayed a high degree of mixing in the latent space when performing dimensionality reduction on all sections or after projection of principal components (PCs) from section 4 (Figure S7E), supporting our clustering approach (STAR Methods).

Our reconstructed tissue can be utilized for comprehensive exploratory 3D analysis, querying both the transcriptomic and imaging modalities. Rendering techniques such as volume cropping and plane clipping can be leveraged to explore gene expression in any direction regardless of the sectioning plane (Figure 7B). To illustrate, we visualized the spatial gene expression of *ACKR1*, an endothelial marker restricted to post-capillary venules.[45–47] Querying the tissue with a plane clip, which mimics cutting the tissue using a different sectioning plane, allowed us to follow a post-capillary venule (*ACKR1*) across the z-dimension and reveals how its expression is confined to the immune regions of the metastatic lymph node, forming a network-like structure (Figure 7B, bottom right).

Furthermore, we examined a KP, which appeared as an isolated structure within the lymph node in the initial sections (Figure 4D). Our 3D reconstruction revealed its connection to the metastatic mass in the later sections, demonstrating continuity across imaging and transcriptomic modalities (Figure 7C). Similarly, we could follow the continuity of the invading tumor and its stroma within the lymph node context via the markers *SPP1* (CAMs), *ACTA2* (CAFs surrounding the tumor, proximal to lymph node populations), and *LYZ* (macrophages), as a transcriptomic feature located at the lymph sinus, as well on invaded tissue.[48] This structure is in general found in close proximity to afferent lymphatic vessels, and sinus macrophages are the first to be exposed to the metastasis.[49]

The power of 3D reconstruction extends well beyond the study of cell types and tissue structures; for instance, inspecting gene program activities in 3D. We validated the three-dimensional organization of cells with high cholesterol biosynthesis gene set activity

Figure 6. Comparison of Open-ST to imaging-based spatial transcriptomics

(A) Spatial visualization of annotated Leiden clusters of human primary HNSCC, profiled with the Xenium *in situ* platform (10× Genomics) (left) and Open-ST (right), subset from whole transcriptome to the genes shared with Xenium (316/377).

(B) Proportion of total detected transcripts that fall within segmented cells (subset for Open-ST).

(C) Correlation of pseudo-bulk gene expression for overlapping genes between Xenium and Open-ST, for the sections shown in (A); each point indicates the sum of transcripts assigned to cells, for both modalities.

(D) Top: Pearson correlation coefficients, computed for the mean normalized gene expression for all pairs of clusters across Xenium and Open-ST (subset). Values are clipped from 0.5 to 0.8. Bottom: average expression of marker genes for populations represented in the primary HNSCC section across Xenium and Open-ST data, normalized per gene.

(E–H) As (A)–(D), but for the metastatic lymph node analyzed with Xenium and Open-ST (previously shown, section 4) subset to the same shared genes (340/377).

(I) Spatial distribution of gene expression (for *MALL* and *CXCL9*), proliferation (*MKI67*, *TOP2A*, *CENPF*, *UBE2C*, and *CDK1*), keratin pearl (*APOBEC3A*, *SERPINB2*, *GPRC5A*, *IL1RL1*, *EHF*, *ALDH1A3*, *KLK11*), epithelial (*AQP3*, *EGFR*, *GPC1*, *MYC*, *MALL*, *LY6D*, *FSTL3*, *EPCAM*, *ERBB2*), and inflammation (*IL1R2*, *MET*, *KCNMA1*, *RAPGEF3*) gene signatures in the primary HNSCC section (right) and the metastatic lymph node (left), for Xenium (top) and Open-ST (bottom).

(J) Gene expression per tumor subcluster in the Xenium primary tissue (top) and mean Xenium module score (aggregate normalized expression of module genes) per tumor subcluster from the Open-ST data with >20,000 genes (bottom).

(K) Expression of ligand-receptor genes from the liana consensus database, for the Xenium (top) and Open-ST data (bottom).

(L) Spatially variable genes in Xenium and Open-ST. CC, cell cycle; KP, keratin pearl; Endo/En, endothelial; CAF/CF, cancer-associated fibroblast; Macro/M, macrophages; GC, germinal center; ROI, region of interest; Fi, fibroblasts; Im, immune.

See also Figure S6.

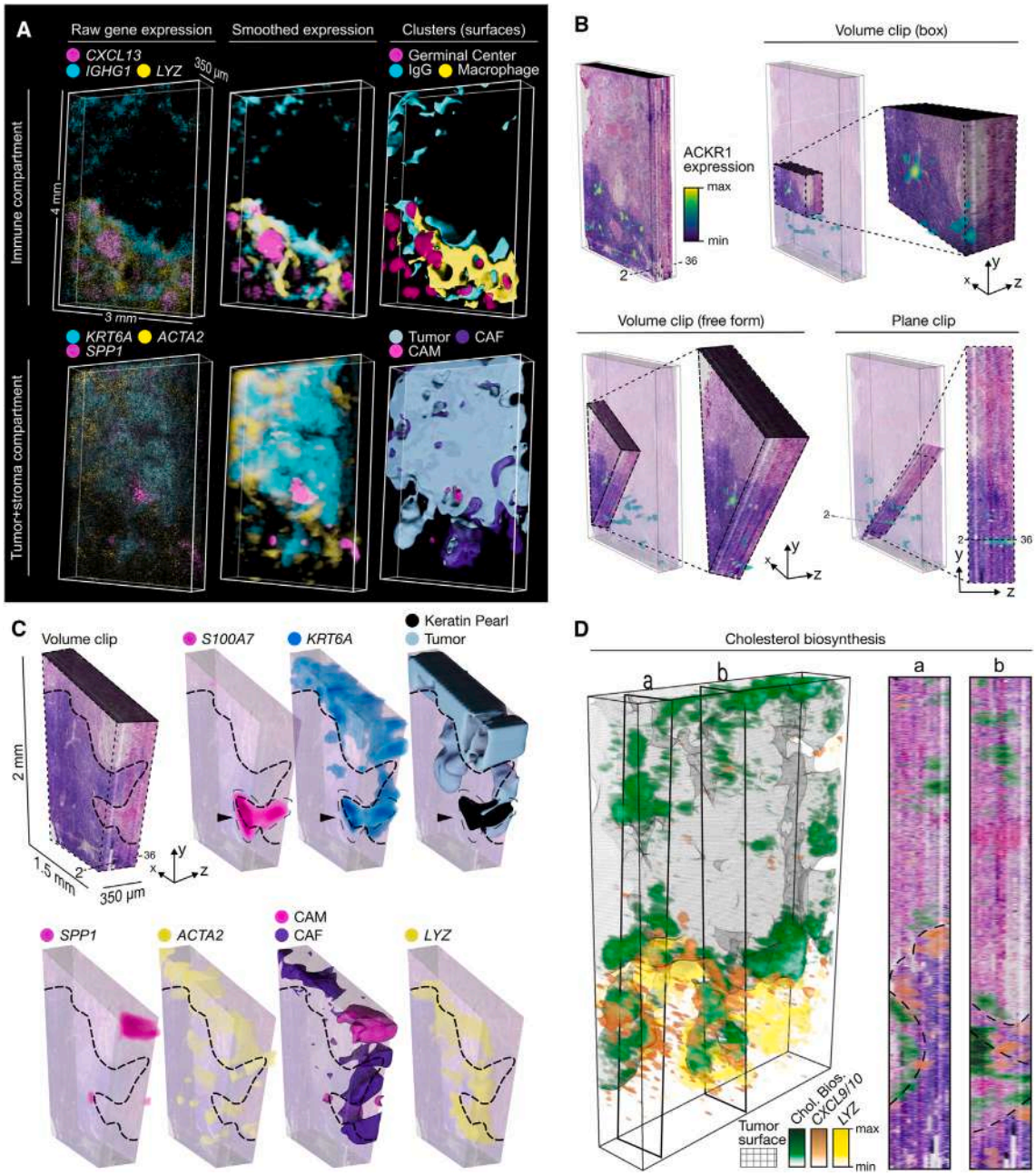


Figure 7. 3D virtual tissue block: the metastatic lymph node

(A) 3D visualizations highlight the spatial compartmentalization and continuity of gene expression. From left to right: gene expression levels of the markers rendered at each cell, smoothed expression, and 3D rendering of clusters as smooth surfaces. Top: immune compartment represented by the germinal center (*CXCL13*), IgG plasma cells (*IGHG1*), and macrophages (*LYZ*); bottom: tumor and stroma compartments represented by epithelial cells (*KRT6A*), CAMs (*SPP1*), and CAFs (*ACTA2*).

(B) Rendering techniques allow exploration of the continuity of gene expression (e.g., for venular endothelial marker *ACKR1*) in 3D, from the otherwise obstructed view from a fully opaque 3D virtual block. Volume clips (box or free-form) and plane clipping showing expression across the z axis.

(C) From left to right, a volume clip of the H&E tissue block shows the continuity of a keratin pearl structure that appears disconnected from the main tumor mass in early sections. Gene expression supports this continuity along the z axis, reflected by *S100A7* (specific to keratin pearls) and *KRT6A* (all tumor cells) over the semi-transparent tissue block. The stroma, demarcated by *SPP1* (CAMs) and *ACTA2* (CAFs) expression, also follows this tumor structure penetrating the lymph node, whose boundary is rich in *LYZ*. Dashed lines indicate the tumor/immune boundary as determined from the H&E.

(legend continued on next page)

at the boundary between tumor and lymphoid tissue, in close proximity to the sinus, hinted by the presence of *LYZ*+, *CXCL9*+, and *CXCL10*+ cells (Figure 7D; STAR Methods). The spatial activity of the G2M checkpoint around areas with hypoxic phenotype at the core of the tumor mass, previously hinted in a single 2D section (Figure S6E), was confirmed in 3D (Figures S7F and S7G).

Virtual tissue blocks support additional modalities beyond H&E and transcriptomics: we aligned immunofluorescence stainings performed on a section adjacent to section 2 from the metastatic lymph node (Figures S7H and S7I; STAR Methods). This proof-of-principle multimodal integration serves as an external validation of the Open-ST capture quality (Figures S7H and S7I). Tumor cells, expressing cytokeratins, and endothelial venules, expressing vimentin (*VIM*), displayed good agreement of the spatial distribution of their respective mRNA/protein. Specifically, vimentin was found at the outermost layer of cells at endothelial venules in both modalities, also supported by the *PECAM1* and *ACKR1* markers, with similar spatial organization (Figure S7I). On the other hand, we observed a partial discrepancy in the levels of keratin mRNA and protein (panCK) (arrowheads in Figure S7H), possibly due to the necrotic nature of these areas.

In summary, Open-ST provides a versatile and powerful framework for comprehensive analysis of gene expression in 2D and 3D, including the unveiling of molecular and/or cellular spatial structures obscured in traditional 2D representations.

DISCUSSION

We introduced Open-ST, a framework of experimental and computational tools for the efficient capture and interrogation of transcripts in tissue space at subcellular resolution.

We illustrated the quality of our data in the stereotypical developing E13 mouse head, where transcript localization showed high accordance with ISHs from the Allen Atlas. We further demonstrated the wide applicability of Open-ST by spatially sequencing structurally complex primary human tissues of clinical interest. Benchmarked against comparable spatial technologies (array-based), Open-ST outperforms in terms of UMI counts per cell at similar sequencing depth while retaining a low reads-to-UMIs ratio.

We compared Open-ST with an imaging-based technology (Xenium), using the same human tumor samples. We subset Open-ST to the 316 (or 344) genes common to both technologies, thus downscaling Open-ST to ~1% of the captured transcripts. The cellular arrangement and spatial patterns captured by Open-ST closely align with those by Xenium, although the signal from individual genes was sparser and thus less precise for structural assessment. The additional information from the >20,000 genes in Open-ST allows finer, unbiased characterization of molecular phenotypes, a challenge for probe-based technologies: these require customized gene panels for refined analyses, increasing costs and possibly the noisiness of the data when increasing probe number.[50–52]

Open-ST recapitulates tumor heterogeneity at single-cell resolution from primary and metastatic lymph node samples of the same patient. In the primary tissue, we discovered heterogeneous states of tumor and stromal cells, restricted to specific spatial domains with enrichment of specific gene programs and receptor-ligand interaction hotspots. These states involved inflammatory response, epithelial differentiation and keratinization, proliferation, and matrix remodeling with migratory potential. The latter population is situated in close proximity to CAF/endothelial-rich stroma, suggesting a potential relationship between the metastasis-initiating cells and the stromal compartment.[24,53] Moreover, tumor cell and stromal heterogeneity were maintained from the primary to metastatic tissue. The differences in spatial arrangement of tumor subclusters in the metastasis suggest that transcriptomic states are established around specific communication hotspots in the primary tissue and are mostly maintained upon new microenvironments. While transcriptomic states were generally similar for the detected subclusters, we found unique features in the metastatic tissue, such as the upregulation of *AMTN*, potentially caused by the differences in signaling at the new microenvironment.[54,55] Moreover, 3D visualization confirmed the spatial patterns of gene program activity in the metastatic lymph node. At the tumor/lymphoid boundary, we observed the upregulation of a cholesterol biosynthesis signature and gene expression patterns (i.e., *LYZ*+ *CXCL9*+ *CXCL10*+ macrophages) that hint the structural integrity of the subcapsular sinus, a structure located at the outer layer of the lymph node. We hypothesize that the growth of the tumor mass along the boundary pushes the lymphoid tissue rather than breaking it. This provides mechanical cues linked to the regulation of lipid metabolism.[56,57] The upregulation of cholesterol metabolism is an immune-evasive mechanism: tumor cells with cholesterol-enriched membranes have been shown to impair T cell-mediated cytotoxicity.[58] Moreover, previous studies report dysregulation of cholesterol metabolism in HNSCC as a potential therapeutic target, in particular, the enzyme HMGCR (3-hydroxy-3-methylglutaryl-coenzyme A [CoA] reductase).[59,60] Similarly, a fatty acid metabolism signature at the tumor/peritumor boundary was reported in liver cancer.[4]

All our analyses were obtained by using the Open-ST computational toolkit. We anticipate our toolkit can be applied to other experimental setups and processed samples, requiring minimal user intervention. Additionally, the modular approach enables the use of alternative tools and algorithms for any (pre)processing step, when necessary. Altogether, we provide detailed experimental resources and open-source computational tools with potential for rendering Open-ST as a standard for community-driven generation and analysis of ST.

Limitations and future developments

Open-ST is currently limited to polyadenylated transcripts and carries a 3′ capture bias. The library preparation can be adapted to retain entire transcript length, enabling the spatial analysis of

(D) Spatial expression of *LYZ* and the chemokines *CXCL9*/*CXCL10*, and spatial activity of the cholesterol biosynthesis gene sets (Reactome), per segmented cell (AUCell), and visualized as smoothed volumetric renderings (left).[44] Tumor surface is shown as a skeletal representation (wireframe) in gray. Gene expression and program activity are additionally shown in two cross sections (right, a and b).
See also Figure S7.

isoform diversity. Similarly, the capture oligos can be modified and designed to capture specific targets, such as splice variants or bacterial 16S transcripts.

Moreover, we anticipate to render Open-ST compatible with formalin-fixed paraffin-embedded (FFPE) tissues, widening its applicability. FFPE tissues are well-suited for long-term storage and better preserve tissue morphology. FFPE tissue sections require additional steps of deparaffinization and decrosslinking to allow RNA release. Also, FFPE processing leads to fragmentation of RNA, reducing intact polyadenylated RNA available for poly(dT) priming. To overcome limitations in poly(A) capture due to RNA fragmentation, a probe-based approach could be adopted, as in Visium FFPE, or *in situ* polyadenylation could be performed to capture fragmented and non-polyadenylated RNA.[61,62]

Open-ST requires an initial financial investment to generate the capture areas and is therefore not well-suited for low-throughput experiments. Collaborations or centralized capture area generation by institutes can overcome this hurdle.

Cell segmentation is currently guided by nuclear staining, limiting its precision and contributing to the cross-contamination of cellular transcriptomes. This can be addressed by introducing membrane staining, additionally enabling the detection of multinucleated cells. Incorporating immunofluorescence or immunohistochemical stainings may negatively affect RNA quality and/or localization of RNA capture and, thus, require separate optimization. In this regard, a comprehensive benchmarking of segmentation tools for sequencing-based technologies is also required. The discrepancy in z- versus x-y resolution (10 μm vs. 0.6 μm) presents a complementary limitation, as transcripts deriving from different cells layered along the z-dimension may be captured, contributing to the presence of double-positive signal and confounding single-cell level analyses. Future development of computational methods will allow deconvolution of this data.

Currently, no standardized data analysis pipeline exists to account for artifactual local biases of sequencing-based ST methods: anisotropic lateral diffusion, spatially autocorrelated background signal, and misassigned transcripts due to inaccurate two-dimensional segmentation or the lack of resolution in the z axis. Here, we leveraged standard workflows tailored for scRNA-seq that rely on normalization and manifold learning from the cells-by-genes matrix without considering spatial autocorrelation as a source of covariance.[63–65] We anticipate that spatial autocorrelation might break the assumptions underlying normalization and manifold learning, leading to spurious cell types and inaccurate differential gene expression and activity of programs. Future developments in spatially aware normalization, dimensionality reduction, and clustering will address questions that rely on modeling count data in space, such as differential expression analysis, as well as analyses relying on manifold structure, such as cell typing, pseudotime, and RNA velocity.

STAR★METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.05.055.

ACKNOWLEDGMENTS

We thank all members of the Rajewsky lab for critical and helpful discussions. We thank Alexandra Tschernycheff and Margareta Herzog for supporting the organization of the lab. Thank you to Cledi Alicia Cerda-Jara and Gwendolin Thomas for providing mouse samples and Lieke van de Haar for discussions on mouse brain anatomy. We thank Sonja Fritzsche for providing the antibodies and her and Kamil Lisek for advice on the immunofluorescence stainings. We also thank Salah Ayoub, Marvin Jens, and our former lab member Asija Diag for their input to the spatial transcriptomics method development. We thank the Genomics Technology Platform at the MDC/BIH@Charité, particularly Tatiana Borodina and Jeannine Wilde, for performing sequencing runs and providing us with custom flow cells and Izabela Plumbom for performing the Xenium *in situ* run. We thank the anonymous reviewers for their helpful and constructive feedback. We thank Susanne Mandrup (University of Southern Denmark) for helpful discussions about Open-ST adipocyte data.

Illustrations in Figures 1D, 2A, and 2D were created with BioRender. M.S. was financially supported by the Berlin Institute of Health at Charité (BIH) between December 2020 and February 2022, D.L.-P. by the Helmholtz Einstein

International Berlin Research School in Data Science (HEIBRiDS) program of the Helmholtz Association, and E.S. received an Erasmus fellowship. L.S. received financial support from the International Research Training Group (IRTG) 2403 program of the Humboldt-Universität zu Berlin. T.M.P. was financially supported by the Berlin School of Integrative Oncology through the GSSP program of the German Academy of Exchange Service (DAAD) and by the Add-on Fellowship of the Joachim Herz Foundation. J.A. received funding from the MDC/NYU PhD Exchange Program, S.S.T. from the Berlin Institute of Health at Charité (BIH) between December 2020 and July 2022, and A.B. from the Leibnizpreis DFG RA 838/5-1. N.K. was financially supported by the Deutsche Forschungsgemeinschaft (DFG), grant number KA 5006/1-1. E.S., E.F., and G.M. acknowledge funding from the European Union: The National Recovery Plan, Mission 4 Component 2 Investment 1.4, Next Generation EU project CN3, code CN_00000041. G.M. received financial support via a BIH Visiting Professorship "Single Cell Genetics and Epigenetics of Patient Derived Tumor and Brain Organoids" by the Stiftung Charité from 2018 to 2022. N.R. thanks the Leibniz prize of the Deutsche Forschungsgemeinschaft (DFG) (grant number RA 838/5-1), the DFG Exzellenzcluster 2049 NeuroCure, the Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK) grant numbers 81Z0100105 and 81X2100155, and the Chan Zuckerberg Foundation (CZI)/Seed Network.

AUTHOR CONTRIBUTIONS

G.M. and N.R. conceived the project. J.A., S.A., A.B., S.E., J.R.L., G.M., N.R., S.S.T., M.S., and E.S. contributed to the experimental method development. N.K., G.M., D.L.-P., T.M.P., N.R., M.S., and E.S. designed the experiments. S.D. and S.S. collected the human samples and performed the pathological analysis. M.S. and E.S. processed all samples except for the mouse head, which was processed jointly with J.R.L. N.K. and D.L.-P. performed the computational method development. D.L.-P. led the computational analysis, supported by N.K. and L.S. N.K., G.M., and N.R. supervised the project. S.A. and E.S. were supported by M.P. and E.F., respectively. N.K., D.L.-P., N.R., M.S., E.S., and L.S. wrote the original draft. All authors revised the manuscript and approved the final version.

DECLARATION OF INTERESTS

N.K., J.R.L., G.M., N.R., M.S., and E.S. are listed as inventors of a patent application relating to the work. The patent application was submitted through the Technology Transfer Office of the Max-Delbrueck Center, with the MDC being the patent applicant.

Received: January 4, 2024
Revised: April 5, 2024
Accepted: May 30, 2024
Published: June 24, 2024

REFERENCES

1. Bressan, D., Battistoni, G., and Hannon, G.J. (2023). The dawn of spatial omics. Science *381*, eabq4964. https://doi.org/10.1126/science.abq4964.
2. Rajewsky, N., Almouzni, G., Gorski, S.A., Aerts, S., Amit, I., Bertero, M.G., Bock, C., Bredenoord, A.L., Cavalli, G., Chiocca, S., et al. (2020). LifeTime and improving European healthcare through cell-based interceptive medicine. Nature *587*, 377–386. https://doi.org/10.1038/s41586-020-2715-9.
3. Wang, X.Q., Danenberg, E., Huang, C.-S., Egle, D., Callari, M., Bermejo, B., Dugo, M., Zamagni, C., Thill, M., Anton, A., et al. (2023). Spatial predictors of immunotherapy response in triple-negative breast cancer. Nature *621*, 868–876. https://doi.org/10.1038/s41586-023-06498-3.
4. Wu, L., Yan, J., Bai, Y., Chen, F., Zou, X., Xu, J., Huang, A., Hou, L., Zhong, Y., Jing, Z., et al. (2023). An invasive zone in human liver cancer identified by Stereo-seq promotes hepatocyte-tumor cell crosstalk, local immunosuppression and tumor progression. Cell Res. *33*, 585–603. https://doi.org/10.1038/s41422-023-00831-1.
5. Grant, S.M., Lou, M., Yao, L., Germain, R.N., and Radtke, A.J. (2020). The lymph node at a glance - how spatial organization optimizes the immune response. J. Cell Sci. *133*, jcs241828. https://doi.org/10.1242/jcs.241828.
6. Elhanani, O., Ben-Uri, R., and Keren, L. (2023). Spatial profiling technologies illuminate the tumor microenvironment. Cancer Cell *41*, 404–420. https://doi.org/10.1016/j.ccell.2023.01.010.
7. Moses, L., and Pachter, L. (2022). Museum of spatial transcriptomics. Nat. Methods *19*, 534–546. https://doi.org/10.1038/s41592-022-01409-2.
8. Neuschulz, A., Bakina, O., Badillo-Lisakowski, V., Olivares-Chauvet, P., Conrad, T., Gotthardt, M., Kettenmann, H., and Junker, J.P. (2023). A single-cell RNA labeling strategy for measuring stress response upon tissue dissociation. Mol. Syst. Biol. *19*, e11147. https://doi.org/10.15252/msb.202211147.
9. Denisenko, E., Guo, B.B., Jones, M., Hou, R., de Kock, L., Lassmann, T., Poppe, D., Clément, O., Simmons, R.K., Lister, R., et al. (2020). Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. Genome Biol. *21*, 130. https://doi.org/10.1186/s13059-020-02048-6.
10. 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.
11. He, S., Bhatt, R., Birditt, B., Brown, C., Brown, E., Chantranuvatana, K., Danaher, P., Dunaway, D., Filanoski, B., Garrison, R.G., et al. (2022). High-Plex Multiomic Analysis in FFPE Tissue at Single-Cellular and Subcellular Resolution by Spatial Molecular Imaging. Nat. Biotechnol. *40*, 1794–1806. https://doi.org/10.1038/s41587-022-01483-z.
12. Merritt, C.R., Ong, G.T., Church, S.E., Barker, K., Danaher, P., Geiss, G., Hoang, M., Jung, J., Liang, Y., McKay-Fleisch, J., et al. (2020). Multiplex digital spatial profiling of proteins and RNA in fixed tissue. Nat. Biotechnol. *38*, 586–599. https://doi.org/10.1038/s41587-020-0472-9.
13. Janesick, A., Shelansky, R., Gottscho, A., Wagner, F., Rouault, M., Beliakoff, G., Faria de Oliveira, M., Kohlway, A., Abousoud, J., Morrison, C., et al. (2023). High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue. Nat. Commun *14*, 8353. https://doi.org/10.1101/2022.10.06.510405.
14. Lee, Y., Bogdanoff, D., Wang, Y., Hartoularos, G.C., Woo, J.M., Mowery, C.T., Nisonoff, H.M., Lee, D.S., Sun, Y., Lee, J., et al. (2021). XYZeq: Spatially resolved single-cell RNA sequencing reveals expression heterogeneity in the tumor microenvironment. Sci. Adv. *7*, eabg4755. https://doi.org/10.1126/sciadv.abg4755.
15. 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.
16. 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.
17. Srivatsan, S.R., Regier, M.C., Barkan, E., Franks, J.M., Packer, J.S., Grosjean, P., Duran, M., Saxton, S., Ladd, J.J., Spielmann, M., et al. (2021). Embryo-scale, single-cell spatial transcriptomics. Science *373*, 111–117. https://doi.org/10.1126/science.abb9536.
18. 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.
19. Marshall, J.L., Noel, T., Wang, Q.S., Chen, H., Murray, E., Subramanian, A., Vernon, K.A., Bazua-Valenti, S., Liguori, K., Keller, K., et al. (2022). High-resolution Slide-seqV2 spatial transcriptomics enables discovery of disease-specific cell neighborhoods and pathways. iScience *25*, 104097. https://doi.org/10.1016/j.isci.2022.104097.

20. 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.

21. Arora, R., Cao, C., Kumar, M., Sinha, S., Chanda, A., McNeil, R., Samuel, D., Arora, R.K., Matthews, T.W., Chandarana, S., et al. (2023). Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. Nat. Commun. *14*, 5029. https://doi.org/10.1038/s41467-023-40271-4.

22. Choi, J.-H., Lee, B.-S., Jang, J.Y., Lee, Y.S., Kim, H.J., Roh, J., Shin, Y.S., Woo, H.G., and Kim, C.-H. (2023). Single-cell transcriptome profiling of the stepwise progression of head and neck cancer. Nat. Commun. *14*, 1055. https://doi.org/10.1038/s41467-023-36691-x.

23. Quah, H.S., Cao, E.Y., Suteja, L., Li, C.H., Leong, H.S., Chong, F.T., Gupta, S., Arcinas, C., Ouyang, J.F., Ang, V., et al. (2023). Single cell analysis in head and neck cancer reveals potential immune evasion mechanisms during early metastasis. Nat. Commun. *14*, 1680. https://doi.org/10.1038/s41467-023-37379-y.

24. Puram, S.V., Tirosh, I., Parikh, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C.L., Mroz, E.A., Emerick, K.S., et al. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell *171*, 1611–1624.e24. https://doi.org/10.1016/j.cell.2017.10.044.

25. Shen, M.-J.R., Boutell, J.M., Stephens, K.M., Ronaghi, M., Gunderson, K.L., Venkatesan, B.M., Bowen, M.S., and Vijayan, K. (2013). Kinetic exclusion amplification of nucleic acid libraries. US patent US2013/045491.

26. Pachitariu, M., and Stringer, C. (2022). Cellpose 2.0: how to train your own model. Nat. Methods *19*, 1634–1641. https://doi.org/10.1038/s41592-022-01663-4.

27. Sztanka-Toth, T.R., Jens, M., Karaiskos, N., and Rajewsky, N. (2022). Spacemake: processing and analysis of large-scale spatial transcriptomics data. Gigascience *11*, giac064. https://doi.org/10.1093/gigascience/giac064.

28. Preibisch, S., Karaiskos, N., and Rajewsky, N. (2021). Image-based representation of massive spatial transcriptomics datasets. Preprint at bioRxiv. https://doi.org/10.1101/2021.12.07.471629.

29. Ahrens, J., Geveci, B., and Law, C. (2005). ParaView: An End-User Tool for Large-Data Visualization. In Visualization Handbook (Elsevier), pp. 717–731. https://doi.org/10.1016/B978-012387582-2/50038-1.

30. Schroeder, W., Martin, K.M., and Lorensen, W. (1998). The Visualization Toolkit: An Object-Oriented approach to 3D Graphics, Second Edition (Prentice Hall PTR).

31. La Manno, G., Siletti, K., Furlan, A., Gyllborg, D., Vinsland, E., Mossi Albiach, A., Mattsson Langseth, C., Khven, I., Lederer, A.R., Dratva, L.M., et al. (2021). Molecular architecture of the developing mouse brain. Nature *596*, 92–96. https://doi.org/10.1038/s41586-021-03775-x.

32. Allen Brain Atlas (2024). Allen Institute for Brain Science Allen Developing Mouse Brain Atlas. http://developingmouse.brain-map.org/.

33. Thompson, C.L., Ng, L., Menon, V., Martinez, S., Lee, C.-K., Glattfelder, K., Sunkin, S.M., Henry, A., Lau, C., Dang, C., et al. (2014). A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. Neuron *83*, 309–323. https://doi.org/10.1016/j.neuron.2014.05.033.

34. Ypsilanti, A.R., Pattabiraman, K., Catta-Preta, R., Golonzhka, O., Lindtner, S., Tang, K., Jones, I.R., Abnousi, A., Juric, I., Hu, M., et al. (2021). Transcriptional network orchestrating regional patterning of cortical progenitors. Proc. Natl. Acad. Sci. USA *118*, e2024795118. https://doi.org/10.1073/pnas.2024795118.

35. Dwivedi, N., Shukla, N., Prathima, K.M., Das, M., and Dhar, S.K. (2023). Novel CAF-identifiers via transcriptomic and protein level analysis in HNSC patients. Sci. Rep. *13*, 13899. https://doi.org/10.1038/s41598-023-40908-w.

36. Xiang, M., Grosso, R.A., Takeda, A., Pan, J., Bekkhus, T., Brulois, K., Dermadi, D., Nordling, S., Vanlandewijck, M., Jalkanen, S., et al. (2020). A Single-Cell Transcriptional Roadmap of the Mouse and Human Lymph Node Lymphatic Vasculature. Front. Cardiovasc. Med. *7*, 52. https://doi.org/10.3389/fcvm.2020.00052.

37. Cai, X., Zhang, H., and Li, T. (2022). The role of SPP1 as a prognostic biomarker and therapeutic target in head and neck squamous cell carcinoma. Int. J. Oral Maxillofac. Surg. *51*, 732–741. https://doi.org/10.1016/j.ijom.2021.07.022.

38. Cambier, S., Gouwy, M., and Proost, P. (2023). The chemokines CXCL8 and CXCL12: molecular and functional properties, role in disease and efforts towards pharmacological intervention. Cell. Mol. Immunol. *20*, 217–251. https://doi.org/10.1038/s41423-023-00974-6.

39. Marshall, A.J., Du, Q., Draves, K.E., Shikishima, Y., HayGlass, K.T., and Clark, E.A. (2002). FDC-SP, a novel secreted protein expressed by follicular dendritic cells. J. Immunol. *169*, 2381–2389. https://doi.org/10.4049/jimmunol.169.5.2381.

40. Sundling, C., Lau, A.W.Y., Bourne, K., Young, C., Laurianto, C., Hermes, J.R., Menzies, R.J., Butt, D., Kräutler, N.J., Zahra, D., et al. (2021). Positive selection of IgG+ over IgM+ B cells in the germinal center reaction. Immunity *54*, 988–1001.e5. https://doi.org/10.1016/j.immuni.2021.03.013.

41. Emont, M.P., Jacobs, C., Essene, A.L., Pant, D., Tenen, D., Colleluori, G., Di Vincenzo, A., Jørgensen, A.M., Dashti, H., Stefek, A., et al. (2022). A single-cell atlas of human and mouse white adipose tissue. Nature *603*, 926–933. https://doi.org/10.1038/s41586-022-04518-2.

42. Maniyadath, B., Zhang, Q., Gupta, R.K., and Mandrup, S. (2023). Adipose tissue at single-cell resolution. Cell Metab. *35*, 386–413. https://doi.org/10.1016/j.cmet.2023.02.002.

43. Bill, R., Wirapati, P., Messemaker, M., Roh, W., Zitti, B., Duval, F., Kiss, M., Park, J.C., Saal, T.M., Hoelzl, J., et al. (2023). CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. Science *381*, 515–524. https://doi.org/10.1126/science.ade2292.

44. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.

45. Hadley, T.J., Lu, Z.H., Wasniowska, K., Martin, A.W., Peiper, S.C., Hesselgesser, J., and Horuk, R. (1994). Postcapillary venule endothelial cells in kidney express a multispecific chemokine receptor that is structurally and functionally identical to the erythroid isoform, which is the Duffy blood group antigen. J. Clin. Invest. *94*, 985–991. https://doi.org/10.1172/JCI117465.

46. Peiper, S.C., Wang, Z.X., Neote, K., Martin, A.W., Showell, H.J., Conklyn, M.J., Ogborne, K., Hadley, T.J., Lu, Z.H., Hesselgesser, J., et al. (1995). The Duffy antigen/receptor for chemokines (DARC) is expressed in endothelial cells of Duffy negative individuals who lack the erythrocyte receptor. J. Exp. Med. *181*, 1311–1317. https://doi.org/10.1084/jem.181.4.1311.

47. Kashiwazaki, M., Tanaka, T., Kanda, H., Ebisuno, Y., Izawa, D., Fukuma, N., Akimitsu, N., Sekimizu, K., Monden, M., and Miyasaka, M. (2003). A high endothelial venule-expressing promiscuous chemokine receptor DARC can bind inflammatory, but not lymphoid, chemokines and is dispensable for lymphocyte homing under physiological conditions. Int. Immunol. *15*, 1219–1227. https://doi.org/10.1093/intimm/dxg121.

48. Pellin, D., Claudio, N., Guo, Z., Ziglari, T., and Pucci, F. (2021). Gene Expression Profiling of Lymph Node Sub-Capsular Sinus Macrophages in Cancer. Front. Immunol. *12*, 672123. https://doi.org/10.3389/fimmu.2021.672123.

49. Louie, D.A.P., and Liao, S. (2019). Lymph node subcapsular sinus macrophages as the frontline of lymphatic immune defense. Front. Immunol. *10*, 347. https://doi.org/10.3389/fimmu.2019.00347.

50. Wang, H., Huang, R., Nelson, J., Gao, C., Tran, M., Yeaton, A., Felt, K., Pfaff, K.L., Bowman, T., Rodig, S.J., et al. (2023). Systematic benchmarking of imaging spatial transcriptomics platforms in FFPE tissues. Preprint at bioRxiv. https://doi.org/10.1101/2023.12.07.570603.

51. Cook, D.P., Jensen, K.B., Wise, K., Roache, M.J., Segato Dezem, F., Ryan, N.K., Zamojski, M., Vlachos, I., Knott, S.R., Butler, L.M., et al. (2023). A Comparative Analysis of Imaging-Based Spatial Transcriptomics Platforms. Preprint at bioRxiv. https://doi.org/10.1101/2023.12.13.571385.

52. Hartman, A., and Satija, R. (2024). Comparative analysis of multiplexed in situ gene expression profiling technologies. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.11.575135.

53. Lawson, D.A., Kessenbrock, K., Davis, R.T., Pervolarakis, N., and Werb, Z. (2018). Tumour heterogeneity and metastasis at single-cell resolution. Nat. Cell Biol. *20*, 1349–1360. https://doi.org/10.1038/s41556-018-0236-7.

54. Nakayama, Y., Tsuruya, Y., Noda, K., Yamazaki-Takai, M., Iwai, Y., Ganss, B., and Ogata, Y. (2019). Negative feedback by SNAI2 regulates TGFβ1-induced amelotin gene transcription in epithelial-mesenchymal transition. J. Cell. Physiol. *234*, 11474–11489. https://doi.org/10.1002/jcp.27804.

55. Schinke, H., Pan, M., Akyol, M., Zhou, J., Shi, E., Kranz, G., Libl, D., Quadt, T., Simon, F., Canis, M., et al. (2022). SLUG-related partial epithelial-to-mesenchymal transition is a transcriptomic prognosticator of head and neck cancer survival. Mol. Oncol. *16*, 347–367. https://doi.org/10.1002/1878-0261.13075.

56. Romani, P., Brian, I., Santinon, G., Pocaterra, A., Audano, M., Pedretti, S., Mathieu, S., Forcato, M., Bicciato, S., Manneville, J.-B., et al. (2019). Extracellular matrix mechanical cues regulate lipid metabolism through Lipin-1 and SREBP. Nat. Cell Biol. *21*, 338–347. https://doi.org/10.1038/s41556-018-0270-5.

57. Bertolio, R., Napoletano, F., Mano, M., Maurer-Stroh, S., Fantuz, M., Zannini, A., Bicciato, S., Sorrentino, G., and Del Sal, G. (2019). Sterol regulatory element binding protein 1 couples mechanical cues and lipid metabolism. Nat. Commun. *10*, 1326. https://doi.org/10.1038/s41467-019-09152-7.

58. Lei, K., Kurum, A., Kaynak, M., Bonati, L., Han, Y., Cencen, V., Gao, M., Xie, Y.-Q., Guo, Y., Hannebelle, M.T.M., et al. (2021). Cancer-cell stiffening via cholesterol depletion enhances adoptive T-cell immunotherapy. Nat. Biomed. Eng. *5*, 1411–1425. https://doi.org/10.1038/s41551-021-00826-6.

59. Ahmadi, M., Amiri, S., Pecic, S., Machaj, F., Rosik, J., Łos, M.J., Alizadeh, J., Mahdian, R., da Silva Rosa, S.C., Schaafsma, D., et al. (2020). Pleiotropic effects of statins: A focus on cancer. Biochim. Biophys. Acta Mol. Basis Dis. *1866*, 165968. https://doi.org/10.1016/j.bbadis.2020.165968.

60. Kwon, M., Nam, G.-H., Jung, H., Kim, S.A., Kim, S., Choi, Y., Lee, Y.S., Cho, H.J., and Kim, I.-S. (2021). Statin in combination with cisplatin makes favorable tumor-immune microenvironment for immunotherapy of head and neck squamous cell carcinoma. Cancer Lett. *522*, 198–210. https://doi.org/10.1016/j.canlet.2021.09.029.

61. 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.

62. Bai, Z., Zhang, D., Gao, Y., Tao, B., Bao, S., Enninful, A., Zhang, D., Su, G., Tian, X., Zhang, N., et al. (2024). Spatially Exploring RNA Biology in Archival Formalin-Fixed Paraffin-Embedded Tissues. Preprint at bioRxiv. https://doi.org/10.1101/2024.02.06.579143.

63. Moses, L., Einarsson, P.H., Jackson, K., Luebbert, L., Booeshaghi, A.S., Antonsson, S., Bray, N., Melsted, P., and Pachter, L. (2023). Voyager: exploratory single-cell genomics data analysis with geospatial statistics. Preprint at bioRxiv. https://doi.org/10.1101/2023.07.20.549945.

64. Shinn, M. (2023). Phantom oscillations in principal component analysis. Proc. Natl. Acad. Sci. USA *120*, e2311420120. https://doi.org/10.1073/pnas.2311420120.

65. Novembre, J., and Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. Nat. Genet. *40*, 646–649. https://doi.org/10.1038/ng.139.

66. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

67. 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.

68. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

69. Broad Institute. Drop-seq: Java tools for analyzing Drop-seq data. GitHub. https://github.com/broadinstitute/Drop-seq.

70. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

71. van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., and Yu, T.; scikit-image contributors (2014). scikit-image: image processing in Python. PeerJ *2*, e453. https://doi.org/10.7717/peerj.453.

72. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

73. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

74. Riba, E., Mishkin, D., Shi, J., Ponsa, D., Moreno-Noguer, F., and Bradski, G. (2020). A survey on Kornia: an Open Source Differentiable Computer Vision Library for PyTorch. Preprint at arXiv.

75. Palla, G., Spitzer, H., Klein, M., Fischer, D., Schaar, A.C., Kuemmerle, L.B., Rybakov, S., Ibarra, I.L., Holmberg, O., Virshup, I., et al. (2022). Squidpy: a scalable framework for spatial omics analysis. Nat. Methods *19*, 171–178. https://doi.org/10.1038/s41592-021-01358-2.

76. 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.

77. Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Valiollah Pour Amiri, V., Hong, J., Wu, K., Jayasuriya, M., Mehlman, E., Langevin, M., et al. (2022). A Python library for probabilistic analysis of single-cell omics data. Nat. Biotechnol. *40*, 163–166. https://doi.org/10.1038/s41587-021-01206-w.

78. Dimitrov, D., Türei, D., Garrido-Rodriguez, M., Burmedi, P.L., Nagai, J.S., Boys, C., Ramirez Flores, R.O., Kim, H., Szalai, B., Costa, I.G., et al. (2022). Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. Nat. Commun. *13*, 3224. https://doi.org/10.1038/s41467-022-30755-0.

79. Badia-I-Mompel, P., Vélez Santiago, J., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S., Taus, P., Dugourd, A., Holland, C.H., Ramirez Flores, R.O., et al. (2022). decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinform. Adv. *2*, vbac016. https://doi.org/10.1093/bioadv/vbac016.

80. Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185. https://doi.org/10.1093/bioinformatics/bts356.

81. Chiu, C.L., and Clack, N. (2022). napari: a Python Multi-Dimensional Image Viewer Platform for the Research Community. Microsc. Microanal. *28*, 1576–1577. https://doi.org/10.1017/S1431927622006328.

82. Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82. https://doi.org/10.1126/science.aaf2403.

83. Park, T., Efros, A.A., Zhang, R., and Zhu, J.-Y. (2020). Contrastive Learning for Unpaired Image-to-Image Translation. In Computer Vision – ECCV 2020: 16th European Conference, Proceedings, Part IX Lecture Notes in Computer Science, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, eds. (Springer International Publishing), pp. 319–345. https://doi.org/10.1007/978-3-030-58545-7_19.

84. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods *18*, 100–106. https://doi.org/10.1038/s41592-020-01018-x.

85. Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). Cell Detection with Star-Convex Polygons. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2018: 21st International Conference, Proceedings, Part II Lecture notes in Computer Science, A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds. (Springer International Publishing), pp. 265–273. https://doi.org/10.1007/978-3-030-00934-2_30.

86. Muzellec, B., Teleńczuk, M., Cabeli, V., and Andreux, M. (2023). PyDESeq2: a python package for bulk RNA-seq differential expression analysis. Bioinformatics *39*, btad547. https://doi.org/10.1093/bioinformatics/btad547.

87. Varrone, M., Tavernari, D., Santamaria-Martínez, A., Walsh, L.A., and Ciriello, G. (2024). CellCharter reveals spatial cell niches associated with tissue remodeling and cell plasticity. Nat. Genet. *56*, 74–84. https://doi.org/10.1038/s41588-023-01588-4.

88. Stickels, R.R., Murray, E., Kumar, P., Li, J., Marshall, J.L., Di Bella, D.J., Arlotta, P., Macosko, E.Z., and Chen, F. (2021). Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. Nat. Biotechnol. *39*, 313–319. https://doi.org/10.1038/s41587-020-0739-1.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Pan Cytokeratin (panCK), mouse mAb, clones AE1 and AE3, eFluor™ 570 conjugate	ThermoFisher	Cat# 41-9003-82; RRID: AB_11218704
Vimentin (mouse mAb, clone V9, Alexa Fluor® 750 conjugate	Bio-Techne	Cat# NBP1-97670AF750; RRID: AB_10013743
Biological samples		
Frozen sagittal E13 wild-type mouse head sections (C57BL/6N)	This study	RRID:MGI:2159965
Frozen coronal brain hemisphere sections from male p60 wild-type mouse (C57BL/6N)	This study	RRID:MGI:2159965
Human head and neck squamous cell carcinoma tissue	This manuscript	N/A
Human metastatic lymph node tissue	This manuscript	N/A
Human healthy lymph node tissue	This manuscript	N/A
Chemicals, peptides, and recombinant proteins		
Dra I enzyme	NEB	Cat#R0129
Alkaline Phosphatase Calf Intestinal (CIAP) enzyme	Promega	Cat#M1821
Exonuclease I enzyme	NEB	Cat#M0293
NaOH solution, for molecular biology, 10 M in H2O	Sigma	Cat#72068
UltraPure™1M Tris-HCl pH7.5	Invitrogen™	Cat#15567027
Tissue-Tek OCT	Sakura	Cat#4583
Methanol (min. 99.8%)	Th. Geyer	Cat#1437
2-propanol (min 99.9%)	Th. Geyer	Cat#1197
Haematoxylin, Mayer's	Agilent Dako	Cat#S3309
Bluing buffer	Agilent Dako	Cat#CS702
Eosin Y, aqueous	Sigma	Cat#HT110216
Pepsin from porcine gastric mucosa	Sigma	Cat#P7000
20x SSC	Sigma	Cat#S6639-1L
Hydrochloric Acid (HCl), 10N	AppliChem	Cat#187051
BSA Molecular Biology Grade (conc. 20 mg/ml)	NEB	Cat#B9000S
dNTP SET 100mM 4X1Ml	Life Technologies	Cat#R0182
SuperScript IV Reverse Transcriptase	Life Technologies	Cat#18090010
RiboLock RNase Inhibitor	Thermo Scientific	Cat#EO0381
Tris-HCl Buffer, pH 8.0, 1M	Life Technologies	Cat#AM9855G
Sodium chloride NaCl (5M), RNase-free	Invitrogen	Cat#AM9760G
Roti®-Stock 20 % SDS ready-to-use, sterile filtered	Roth	Cat#1057.1
UltraPure 0.5M EDTA, pH 8.0-4 x 100 Ml	Life Technologies	Cat#15575020
Proteinase K (800 mU/μL)	NEB	Cat#P8107S
DNA Polymerase Large Fragment exo- Klenow Fragment (3'-5' exo-)	NEB	Cat#M0212
Ampure XP beads	Beckman coulter	Cat#A63881
Kapa HiFi Hotstart Readymix, KK2612	Roche	Cat#7958960001
1.5% Agarose Cassettes, dye-free, int. Stds BluePippin, 250bp-1.5kb, Marker R2 or 1.5% Agarose, PippinHT, 300-1500 bp	Biozym	342BDF1550 or HTC1510
Qubit dsDNA HS Assay Kit	Invitrogen	Cat#Q32854
High sensitivity DNA kit	Agilent	Cat#5067-4626

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HS RNA tapestation	Agilent	5067-5579/ - 5580/ -5581
Blue S'Green qPCR mix	Biozym	Cat#331416
KAPA LQ Primer + Mastermix (Illumina/ LC480)	Roche	Cat#7960573001
KAPA Library Quantification DNA Standards (Illumina)	Roche	Cat#7960387001
Phosphate Buffered Saline (10X), pH 7.4, RNase-free	Invitrogen	Cat#10055154
Tween™ 20 Surfact-Amps™ detergent solution	Thermo Scientific	Cat#85113
Ethanol denatured 96%	Serva Electrophoresis	Cat#11096.02
Dimethyl Sulfoxide (DMSO)	Sigma	Cat#D8418-1L
KCl (2M), RNase-free	Life Technologies	Cat#AM9640G
Formaldehyde, for molecular biology, 36.5-38% in H2O	Sigma-Aldrich	Cat#F8775
Triton-X-100	Sigma-Aldrich	Cat#T8787
Normal Donkey Serum	Biozol Diagnostica	Cat#SBA-0030-01
DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride)	Bio-Trend	Cat##40011
Glycerol, ROTIPURAN® ≥99,5 %, p.a, anhydrous	Roth	Cat#3783.2
DPBS, no calcium, no magnesium	Gibco™	Cat#14190169
Critical commercial assays		
NovaSeq 6000 S4 reagent kit v1.5 (35 cycles)	Illumina	Cat#20044417
Xenium Human Multi-Tissue and Cancer Panel	10X Genomics	Panel number 1000626
Xenium Slides & sample Prep reagents	10X Genomics	1000460
Xenium Decoding Reagent Module A	10X Genomics	1000624
Xenium Decoding Reagent Module B	10X Genomics	1000625
Xenium Decoding Consumables	10X Genomics	1000487
Deposited data		
Open-ST (this publication)	GEO	GSE251926
10x Xenium HNSCC and metastatic lymph node (this publication)	GEO	GSE263498
Immunofluorescence images of metastatic lymph node section #1(this publication)	Zenodo	https://doi.org/10.5281/zenodo.11395256
Slide-seqV2 E9.5 mouse brain data	GEO	GSE197353
Seq-Scope mouse liver data	GEO	GSE169706
Stereo-seq	CNGB Sequence Archive	CNX0422301
10X Visium	10X Genomics	https://www.10xgenomics.com/resources/datasets/mouse-brain-serial-section-2-sagittal-anterior-1-standard
DBiT-seq E11 mouse embryo	GEO	GSE137986
Single-cell RNA seq of primary and metastatic tumor of HNSCC	GEO	GSE103322
Allen Developing Mouse Brain Atlas ISH data	Allen Institute	http://developingmouse.brain-map.org/; RRID: SCR_002990
Oligonucleotides		
HDMI32-DraI: CAAGCAGAAGACGGCATACGAGAT TCTTTCCCTACACGACGCTCTTCCGATCTNNVNBV NNVNNVNNVNNVNNVNNVNNNNNNTCTTGTGA CTACAGCACCCTCGACTCTCGCTTTTTTTTTTTTTT TTTTTTTTTTTTTTTAAAGACTTTCACCAGTCCATG ATGTGTAGATCTCGGTGGTCGCCGTATCATT	Cho et al.[16]	N/A
Randomer: TCAGACGTGTGCTCTTCCGATCT NNNNNNNNN	Cho et al.[16]	N/A
Read1-DraI: ATCATGGACTGGTGAAAGTCTT TAAAAAAAAAAAAAAAAAAAAAAAAAAAAA GCGAGAGTCGAGGGTGCTGTAGTCACAAGA	Cho et al.[16]	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
P5 Fwd: AATGATACGGCGACCACCGAGATCTACACT CTTTCCCTACACGACGCTCT*T*C	Cho et al.[16]	N/A
P7 Rev indexing: CAAGCAGAAGACGGCATACGAGAT [8-mer index sequence] GTGACTGGA GTTCAGACG TGTGCTCTTCC*G*A	Cho et al.[16]	N/A
Software and algorithms		
IdeaMaker v4.3.2	Raise3D	https://www.raise3d.com/download/
Fiji v1.53t	Schindelin et al.[66]	https://imagej.net/software/fiji/
BZ-X Viewer 1.3.1.1	Keyence	N/A
BZ-X Analyzer 1.4.1.1	Keyence	N/A
Paraview v5.11.0	Ahrens et al.[29]	https://www.paraview.org/
Cellpose v2.2	Pachitariu and Stringer[26]	https://github.com/MouseLand/cellpose
STAR v2.7.10b	Dobin et al.[67]	https://github.com/alexdobin/STAR
Bowtie2 v2.5.1	Langmead et al.[68]	https://github.com/BenLangmead/bowtie2
Drop-seq tools v2.5.1	Broad Institute[69]	https://github.com/broadinstitute/Drop-seq
spacemake v0.7.3	Sztanka-Toth et al.[27]	https://github.com/rajewsky-lab/spacemake
scanpy v1.9.3	Wolf et al.[70]	https://github.com/scverse/scanpy
scikit-image v0.19.3	van der Walt et al.[71]	https://github.com/scikit-image/scikit-image
STIM v0.2.0	Preibisch et al.[28]	https://github.com/PreibischLab/STIM
samtools v1.17	Li et al.[72]	https://github.com/samtools/samtools
microfilm v0.2.1	N/A	https://github.com/guiwitz/microfilm
scipy v1.10.0	Virtanen et al.[73]	https://github.com/scipy/scipy
Kornia v0.7.0	Riba et al.[74]	https://github.com/kornia/kornia
openst v0.1.0	This publication	https://github.com/rajewsky-lab/openst
squidpy v1.3.0	Palla et al.[75]	https://github.com/scverse/squidpy
QuPath v0.4.0	Bankhead et al.[76]	https://qupath.github.io/
scvi-tools v1.0.2	Gayoso et al.[77]	https://github.com/scverse/scvi-tools
liana-py 1.0.1	Dimitrov et al.[78]	https://github.com/saezlab/liana-py
decoupler-py 1.4.0	Badia-I-Mompel et al.[79]	https://github.com/saezlab/decoupler-py
Xenium Ranger	10× Genomics	https://www.10xgenomics.com/support/software/xenium-ranger/latest
Kneed 0.8.5	N/A	https://github.com/arvkevi/kneed
RSeQc 5.0.3	Wang et al.[80]	https://rseqc.sourceforge.net
napari 0.4.19.post1	Chiu et al.[81]	https://napari.org/stable/
Leica Application Suite X (LAS X) with Leica Thunder widefield deconvolution software	Leica	v.3.9.0.28093
Other		
16-Well ProChamber Microarray System	Grace Bio-Labs	Cat#645508
Tungsten Carbide Tip Scriber	For example, BioTrend	Cat#MD9-29
Glass cutter, 138°	Toyo	Cat#TC17B
Cutting guide 3D-print.*stl* file for NovaSeq S4 flow cell (Illumina)	This publication	https://rajewsky-lab.github.io/openst

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents may be directed to Nikolaus Rajewsky (rajewsky@mdc-berlin.de).

Materials availability

The design of the 3D-printed cutting guide is provided in our online resource (https://rajewsky-lab.github.io/openst). All other materials used are commercially available.

Data and code availability

Open-ST RNA-seq and microscopy data, and 10x Xenium data generated in this study, have been deposited at GEO (accession numbers GSE251926 and GSE263498) and Zenodo (https://doi.org/10.5281/zenodo.11395256) and are publicly available as of the date of publication. This paper also analyzes existing, publicly available data. Accession numbers for all datasets are listed in the key resources table. All original code has been deposited at https://github.com/rajewsky-lab/openst (accessed 3 November 2023) and is publicly available as of the date of publication. DOIs are listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Mouse tissue sample

Mouse adult brain was collected from a postnatal-day 60 C57BL/6N wild-type male mouse (RRID:MGI:2159965). Embryonic stage 13 (E13) mouse heads were collected from C57BL/6N wild-type mice (RRID:MGI:2159965). Tissue was immediately embedded in Tissue-Tek® O.C.T.™ Compound (Sakura, 4583) on powdered dry ice and stored at -80°C. The left hemisphere of the adult brain was embedded and sectioned coronally. Mouse E13 head was embedded and sectioned sagittally.

Animal care and mouse work were conducted according to the guidelines of the Institutional Animal Care and Use Committee of the Max Delbrück Center for Molecular Medicine and the Landesamt für Gesundheit und Soziales of the federal state of Berlin (Berlin, Germany).

Human patient characteristics and sample collection

The resected head and neck tumor, as well as a normal and a metastatic lymph node, were collected from a 56 year-old male patient with a moderately differentiated, keratinizing squamous cell carcinoma (SCC) in the larynx.

Staging workup by cranial computed tomography (CCT) and Positron emission tomography–computed tomography (PET-CT) showed bilateral cervical lymph node metastases, but no distant metastases. A laryngectomy and bilateral neck dissection was performed and the patient received an ambulant adjuvant radiotherapy.

The pathological examination revealed tumor infiltration of the supraglottis, glottis, subglottis regions, pre-epiglottic soft tissue, and two cervical lymph node metastases with a maximum diameter of 12 mm and without extracapsular invasion.

A part of the primary tumor, as well as half of resected level III cervical lymph nodes with and without metastasis, were snap-frozen after surgery and embedded in pre-cooled Tissue-Tek® O.C.T.™ Compound (Sakura, 4583) on powdered dry ice. Tissue was stored at -80°C before and after embedding and used for Open-ST.

The remaining resected specimen was formalin-fixed, decalcified, cut and paraffin-embedded (FFPE) for histological examination. The pathological tumor classification was as follows: pT3 pN2c (2/29) L0 V0 Pn0 G2 R0 (according to the 8th edition of the TNM classification (AJCC)). The FFPE tissue was stored at room temperature at the archive of the Institute of Pathology at the Charité University Hospital, Campus Mitte. The study was performed according to the ethical principles for medical research of the Declaration of Helsinki and approval was approved by the Ethics Committee of the Charité University Medical Department in Berlin (EA4/082/22).

METHOD DETAILS

Capture array generation and disassembly

The capture array was generated as in Cho et al. with several protocol adaptations.[16] The synthetic HDMI32-DraI library was produced using the Ultramer service from IDT. The library was sequenced on an Illumina® NovaSeq 6000 S4 flow cell (35 cycles), at a loading concentration of 200 pM. A single-end 37 cycle read was sequenced, using Read1-DraI (IDT) as a custom primer. A custom sequencing recipe was used, stopping the run immediately after read 1 prior to on-instrument washes (Data S1). The custom sequencing recipe was used in a sequencing run with the following versions and may require adjustment for use with different consumable or software versions: RTA v3.4.4, Flow Cell Consumable v1, Sbs Consumable v3, NovaSeq control Software v1.7.5.

After sequencing, enzymatic reactions were performed by pipetting reaction mixes into the flow cell lanes, covering the inlets/outlets with tape for the incubations. First, the flow cell was washed by flowing through 500 uL ultrapure water, then incubated overnight with DraI mix (2 U/μL DraI enzyme (NEB, R0129) in 1X CutSmart buffer) at 37°C. A repeat 5h incubation at 37°C with fresh mix was done, to ensure efficient digestion in all areas. The DraI cuts the double stranded DNA of the clusters after the poly-d(T) tail. After washes with 80% ethanol, then ultrapure water, the flow cell was incubated with Exonuclease I mix (1 U/μL Exo I (NEB, M0293L), 0.05 U/μL CIAP (Cat#M1821) in 1x Exo I buffer) for 45 min at 37°C. Subsequently, the flow cell was washed three times with ultrapure water before separating the two glass surfaces by scoring along the sides with a scalpel (Video S1).

The second strand was denatured by washing the opened flow cell in a beaker of 0.1 N NaOH for 5 min. After denaturation three washes each with 0.1M Tris HCl (pH 7.5) and ultrapure water were performed.

Using a glass cutter and a 3D-printed tool (available at https://rajewsky-lab.github.io/openst), scores were made along the flow cell top or bottom pieces at the desired breaking points. The cutting tool allows clean scores on the surface without oligos, without damaging the capture area surface. By applying even pressure at the scored sites, the flow cell was broken into approximately 3x4 mm capture area pieces. The first and last 1.5 cm of the S4 flow cell was removed, as these areas are not images by the sequencer and thus do not contain registered barcodes. The opening, scoring and breaking of the flow cell into capture areas is demonstrated in Video S1. Capture area pieces were stuck (with oligos facing up) on a sticky tape to facilitate the handling (Video S2).

The cutting guide was printed with 1.75 mm polylactic acid (PLA) filament on the Pro 3 Dual Extruder 3D printer from Raise 3D. The*.stl* file was prepared for printing using IdeaMaker v.4.3.2 as the 3D slicing software. Capture areas from three different flow cells were used for experiments: fc_1 (human samples, embryonic mouse head) and fc_2 (mouse hippocampus), fc_3 (embryonic mouse head replicate).

RNA quality control

To assess RNA quality, total RNA was extracted from cryosections lysed in Trizol using the Direct-zol RNA Miniprep kit (Zymo Research, R2050) according to the manufacturer's instructions. Concentrations were measured using the Nanodrop-1000 spectrophotometer (Thermo Fisher) and RIN values assessed using the Agilent 2200 TapeStation (High sensitivity RNA kit, Agilent). Tissues used in this study had RIN values ranging from 6.7 - 8.9 (Table S1); the E13 mouse head replicate 2 was not assayed for RNA integrity.

Cryosectioning and fixation

OCT-embedded fresh frozen tissue was sectioned at 10 μm thickness using a CryoStar NX70 cryostat (Epredia™). Tissue sections were placed and melted onto the capture area (Video S3) and fixed in methanol at -20°C for 30 min.

H&E staining and imaging

The section was dried before H&E staining by 1 min incubation with isopropanol and air-drying at room temperature. For H&E staining, Mayer's haematoxylin (Agilent, S3309) was applied for 5 min, the section was washed ten times in water and incubated with bluing buffer (Agilent, CS702) for 2 min. After washing in dH20, the tissue was treated with a 1:1 dilution of eosin Y (Sigma, HT110216) and 0.45 M tris-acetic acid pH 6 for 1 min. The sections were washed in water and left to air-dry completely before imaging. The tissues were imaged in brightfield with a 20x objective on the Keyence BZ-X710 inverted fluorescence phase contrast microscope, placed dry onto a # 1.5 coverslip. Handling of capture areas during staining and imaging is demonstrated in Videos S4 and S5, respectively.

Permeabilization time optimization

To select a suitable permeabilization condition we firstly tried the tissue optimization assay.[82] Fluorescently labeled nucleotides are integrated into the cDNA during RNA retrotranscription; the signal intensity should correspond to the amount of RNA captured, with localization of signal indicating possible diffusion. In our experience this assay was not reliable to quantify capture efficiency, so instead we define RNA capture efficiency using qPCR, where low amplification cycling numbers correspond to a higher concentration of starting material.

We tested different concentrations of pepsin (0.7- 1.4 U/μL) for different incubation times (0, 15, 30, 45, 60 min) at 37°C. Library preparation was done as for a regular Open-ST assay, but only until second strand synthesis. A qPCR was performed as described in the methods section "PCR cycling number assessment". If multiple samples amplified earliest together, we chose the shorter time or lower pepsin concentration for permeabilization.

mRNA release and capture

Reactions on the capture array were performed in a 16-well ProChamber microarray gasket (GraceBiolabs, 645508). Each flow-cell piece was placed in one gasket well with the tissue on the capture area facing up. To cover the entire surface a volume of 100 μl was used for all subsequently described reactions. mRNA was released by treating the tissue with pepsin (P7000, Sigma) in 2x SSC (pH 2.5) at 37°C, with time and concentration differing between the tissue. The primary human samples were permeabilized for 45 min with 1.4 U/μL pepsin. The mouse hippocampus and E13 head were permeabilized with 0.7 U/μL pepsin for 30 min (or 15 min for the E13 mouse head replicate).

The tissue was washed with a reverse transcription buffer (1x SuperScript IV buffer (Thermo Fisher), 1 U/μL Ribolock RNase inhibitor (Thermo Fisher)). Next, an overnight incubation was done at 42°C with reverse transcription mix (6.67 U/μL SuperScript IV (ThermoFisher), 1x SuperScript IV buffer, 5 mM DTT 0.187 mg/mL BSA, 1 mM dNTPs, 1 U/μL Ribolock (ThermoFisher). The wells of the gaskets were sealed with strips of plate sealing tape to prevent evaporation.

Tissue removal

Tissue removal was done as in Cho et al., but with 100 μL mix or wash per sample to completely cover the flow cell piece.[16] Tissue digestion was added after the reverse transcription reaction (100 mM Tris-HCl pH 8.0, 200 mM NaCl, 2% SDS, 5 mM EDTA, 16 mU/μL proteinase K (NEB, P8107S)). An incubation for 40 min at 37°C was followed by three ultrapure water washes, three 5 min 0.1 N NaOH

incubations to denature the hybridized mRNA, three 0.1 M Tris (pH 7.5) washes for neutralization and finally three ultrapure water washes. The capture area was visually inspected to confirm tissue removal before proceeding with 2nd strand synthesis.

Second strand synthesis and purification

Following the permeabilization step, the tissues were incubated with a Second Strand Synthesis mix for 2h at 37°C. The mix contains 1x NEBuffer-2 (NEB, #B7002S), 10 μM of Randomer (custom DNA oligo, IDT), 1 mM of dNTPs mix, 0.5 U/μl Klenow exo (-) Fragment (NEB, #M0212L). The random priming site serves as a UMI in downstream analysis.

The slides were washed three times with ultrapure water and incubated twice for 5 min with 100 uL 0.1 N NaOH to elute the strand. To neutralize the product Tris-HCl pH 7.5 was added to reach 0.125 M. AMPure XP magnetic beads (BeckmanCoulter, #A63881) were used to clean up and concentrate the second strand eluate to 82.5 μL, following the manufacturer's instructions.

PCR cycling number assessment

qPCR was performed on the StepOne™ Real-Time PCR System (Applied Biosystems) to assess the number of amplification cycles needed (Figure S1F).

Around 3% of the purified Secondary Strand (2.5 μL) was used as input, in a mix of 1x Blue S'Green qPCR mix plus ROX (Biozym, #331416) and 1 μM p5 Fwd and p7 Rev indexing primers (Custom DNA oligo, IDT). Cycling conditions were as follows: 95°C for 2 min, followed by 40 cycles of 95°C for 5 sec and 60°C for 30 sec.

The threshold was set at 50% of the peak ΔRn. For each sample the cycle number at the intersection of the threshold and amplification curve was determined. Five cycles less were considered for the Secondary Strand amplification, to account for the input. For all the samples 12 or 13 cycles were selected for amplification.

Library construction

The libraries were prepared for amplification, combining 80 μL of purified second strand with a final 1x KAPA HiFi Library Amp kit (Roche, #07958960001), 1μM of p5 Fwd primer and p7 indexing Rev primer in a volume of 200 μL, split into 4 PCR tubes of 50 μL each. The following cycling conditions were used: initial denaturation for 3 min at 95°C, x cycles of 95°C for 30 sec, 60°C for 1 min and 72°C for 1 min, a 2 min final elongation at 72°C and hold at 4°C.

The AMPure XP magnetic beads (BeckmanCoulter, #A63881) were used at a 1:1 ratio to clean and concentrate the library into a final volume of 20 μl following the manufacturer's instructions. To completely remove short artifacts, such as primer dimers, libraries were selected to have a length from 300-400 bp to 1100 bp, according to library composition and shape (Figure S1E). Size selection was performed using 1.5% agarose gel cassettes on the BluePippin or PippinHT (Sage science, BDF1510 or HTC1510) following the manufacturer's instructions. The product was quantified using the Qubit™ dsDNA HS Kit (Invitrogen, Q32851) and the Bioanalyzer Agilent High Sensitivity DNA Kit (Agilent, 5067-4626).

Spatial transcriptome sequencing

Libraries were quantified for sequencing using the KAPA Library Quantification Kit (optimized for Roche® LightCycler 480, 07960298001). Sequencing was performed on the Illumina® NovaSeq 6000 and NextSeq 2000 sequencing systems, with 130 pM and 650 pM loading, respectively, and 1-5% PhiX spike-in. A minimum of 28 cycles for read 1 and 90 cycles for read 2, as well as an 8-cycle index 1 read, were used.

Xenium *In Situ* Workflow

Xenium *in situ* analysis (10x Genomics) was performed on the fresh-frozen primary HNSCC and patient-matched metastatic lymph node samples, previously processed with Open-ST. The sections analyzed with the two methods were over 100 μm apart. A 10 μm section of each tissue was sectioned and placed onto one Xenium slide according to specifications delineated in the 10x Genomics protocol CG000579 (Rev C). Fixation and permeabilization was performed as described in protocol CG000581 (Rev C, 10x Genomics). Probe hybridization, ligation, amplification, as well as autofluorescence quenching and nuclei staining was performed as specified in CG000582 (Rev E, 10x Genomics) using the pre-designed Xenium Human Multi-Tissue and Cancer Panel (10x Genomics, 1000626). The slide was loaded onto the Xenium analyzer and the regions of interests (ROIs) selected according to the 10x Genomics protocol CG000584 (Rev E). All user-supplied and 10x Genomics reagents used in the run are listed in the key resources table.

Immunofluorescence staining

Immunofluorescent (IF) staining was performed on the first cryosection of the metastatic lymph node reserved for validations, as shown in the experimental setup in Figure 2D. This slide was reserved at -80°C for ~15 months, before proceeding to IF staining. Steps were performed at room temperature unless stated otherwise. Upon drying the slide, the OCT was removed in a 10-minute PBS wash. Next, the section was fixed with 4% formaldehyde (Sigma-Aldrich, F8775) for 15 minutes and then washed with DPBS (no calcium, no magnesium, Gibco™, 14190169) three times. Blocking and permeabilization was done by incubating in 0.25% Triton-X (Sigma-Aldrich, T8787) and 5% normal donkey serum (Biozol Diagnostica, SBA-0030-01) in DPBS for 1 hour.

The section was incubated overnight at 4°C in the dark, with primary-conjugated antibodies diluted in a DPBS buffer with 0.1% Triton-X and 5% normal donkey serum, as follows: 1:100 for Pan Cytokeratin (mouse mAb, clones AE1and AE3, eFluor™ 570

conjugate, ThermoFisher, Cat# 41-9003-82), 1:50 for Vimentin (mouse mAb, clone V9, Alexa Fluor® 750 conjugate, Bio-Techne, Cat# NBP1-97670AF750). Following three 5-minute DPBS washes, DAPI staining was performed with 1 ug/mL DAPI (Bio-Trend, #40011) in DPBS for 10 min in the dark. After three DPBS rinses, the section was dried and then mounted with 85% glycerol.

Images were acquired on the Leica Thunder DMi8 imager using a Leica DFC 9000GT sCMOS fluorescence camera, a 20X objective and the Leica Application Suite (LAS) X software (v.3.9.0.28093). Thunder instant computational clearing was performed on the image. Images were converted to pyramid ome.tiff files using QuPath v0.4.0, and aligned to the H&E staining image of the Open-ST metastatic lymph node (section 2) via manual selection of landmarks with napari.[81] The IF stack was warped into the Open-ST coordinates via an affine transformation matrix.

QUANTIFICATION AND STATISTICAL ANALYSIS

Spatial reconstruction of barcode sequencing

The barcode sequencing for capture area generation enables the mapping of spatial barcodes to their corresponding spatial coordinates within the flow cell. Per flow cell, we processed each of the 3,744 tiles leveraging bcl2fastq v2.20.0.422. Subsequently, the barcode_preprocessing function from our openst tools was used to trim barcodes, compute their reverse complement ("cell_bc"), and supplement these with spatial coordinates ("xcoord" and "ycoord"), ultimately generating a single coordinate file per tile. Notably, the spatial coordinates acquired with bcl2fastq are in a tile-specific coordinate system. Consequently, mapping to a global coordinate system becomes necessary for samples spanning multiple tiles. This was done with the puck_collection functionality from spacemake v0.7.3, using the provided NovaSeq S4 coordinate system.

Processing of transcriptomic reads

The transformation of raw sequencing data into spatially-mapped expression matrices was carried out utilizing spacemake,[27] an automated pipeline designed for the preprocessing, alignment, and quantification of single-cell and spatial transcriptomics data.

Preprocessing

Raw basecalls were demultiplexed with bcl2fastq v2.20.0.422, yielding two FASTQ files per sample: the Read 1 (R1) file, containing the spatial barcode, and the Read 2 (R2) file, containing the unique molecular identifier (UMI) and transcriptomic data. The R1 and R2 FASTQ pairs were merged into a single bam file per sample. Finally, the consolidated bam files underwent trimming of p5 adapter sequences and polyA tails using *TrimStartingSequence* and *PolyATrimmer* from Drop-seq tools v2.5.1, respectively. In spacemake, this preprocessing strategy is represented by the "open-st" barcode_flavor and run_mode.

Mapping

Initially, preprocessed and trimmed reads were aligned to the PhiX reference sequence, with aligned reads subsequently discarded. Remaining reads were subjected to alignment against the corresponding rRNA reference, dependent on the sample's origin (mouse or human). Aligned rRNA reads were discarded. These two steps were performed with bowtie2 v2.5.1 for fast alignment. The final step involved mapping the remaining unmapped reads to the species genome using STAR v2.7.10b, adhering to default parameters within spacemake v0.7.3. For mouse-origin reads, GRCm39 primary assembly with the Gencode vM30 annotation was utilized. Likewise, the GRCh38 genome reference with the Gencode v43 basic annotation was utilized for alignment of human-origin reads. The above mapping strategy can be reproduced in spacemake >=v0.7 by specifying a map_strategy as follows: "bowtie2:phiX->bowtie2:rRNA->STAR:genome:final".

Quantification

Uniquely mapped reads were spatially matched against a library of spatial barcodes derived from the initial flow cell sequencing. Multi-mapping reads were discarded, except for the cases where a read mapped to one genic and one intergenic locus only; in these cases the genic reads were retained. Spatial tiles were designated as part of a specific sample, if >=10% of spatial barcodes of the sample match to the tile. The aligned reads were subsequently distributed across their corresponding spatial tiles per sample. Only zero Hamming distance matches between the known barcodes and the sequenced library barcodes were considered. The gene expression of tile-specific bam files was quantified with *DigitalExpression* from Drop-seq Tools (v2.5.1). The resulting quantified gene expression data was then transformed into a single h5ad file per sample, via outer merge of features (genes).

The percentage of spatially mapping reads was computed as follows. First, only uniquely mapped reads were retained, containing both genic and non-genic sequences. Quantifying gene expression via the *DigitalExpression* tool for all barcodes present in the bam file provided the total number of genic reads. Calculating gene expression in the same manner, but only for the sample barcodes, provided the total number of spatially mapping genic reads, and therefore the spatially mapping percentage. Additionally, the number of tags (3'UTR, 5'UTR, CDS, intron, and distance from TSS/TES) found for each uniquely mapping read was computed using RSeQc 5.0.3.[80]

H&E image preprocessing and cell segmentation

Tile-scan images of H&E-stained tissue sections were stitched together using the Grid/Collection stitching plugin included in Fiji 1.53t, generating a composite image of the entire section.[66] RGB color tile-scans were converted to a HSV (Hue/Saturation/Value) image; then, the saturation channel was blurred with a Gaussian filter, and binarized with Otsu thresholding. This delivers a mask used to isolate the tissue from the background, mostly consisting of the flow cell piece. Optionally, tile-scans were further

preprocessed by style transfer with a fine-tuned Contrastive Unpaired Translation (CUT) model, trained with default parameters on unpaired 512x512 pixel patches from tile-scans imaged with high noise/variance (source style), and low noise/variance (destination style).[83] This model was specifically applied to the metastatic lymph node H&E imaging data to equalize the style between sections and remove artifacts due to low aperture size during acquisition.

Then, nuclei were segmented with cellpose 2.2, upon fine-tuning the cyto2 model on pairs of H&E images from the metastatic lymph node sample (section 4) and manually refined segmentation masks from the same pretrained model.[26,84] Segmentation of all images was performed with –diameter 20 (~7 μm), –flow_threshold 2, and –cellprob_threshold –1, unless otherwise stated. Segmented nuclei were extended 10 pixels (~3.45 μm), radially and in a non-overlapping manner, using the function segmentation.expand_labels from scikit-image (v0.19.3).[71] In samples with adipocytes (e.g., healthy human lymph node sample), a second round of segmentation with our fine-tuned cellpose model was carried out, with –diameter parameter set to 100 pixels (~34.5 μm). Segmentation masks were similarly extended 50 pixels (~17.25 μm) radially. To keep all small and large segmented cells, two rounds of segmentation were performed in the same image. Then, the two segmentation masks (one with and other without adipocytes) were merged by removing segmented cells from the mask with small diameter using the segmentation with large diameter as a negative binary mask. Then, the filtered small-nuclei mask and adipocyte mask were combined via AND operation.

We benchmarked the segmentation model on three randomly-selected 512x512 px regions of interest from the metastatic lymph node (section 4) tile scan. Manual nuclei segmentations, carried out with QuPath, were used as ground truth masks.[76] Average precision at different Intersection over Union (IoU) thresholds were quantified as previously described.[85]

Pairwise alignment of data modalities

A two-step protocol was designed to align spatial transcriptomics data to tile scans of tissue staining (H&E in this study) from the same section, such that capture spot data can be aggregated into (segmented) single cells, instead of using arbitrary grids. This protocol relies on the generation of *pseudoimages* from the spatial transcriptomics data (see "generation of Open-ST pseudoimages"), using libraries from scikit-image. First, rescaled H&E images (~7 μm/pixel) were coarsely aligned to low-resolution (~7 μm/pixel) *pseudoimages* of ST data via the pre-trained Detector-Free Local Feature Matching with Transformers (LoFTR) *outdoor* model, from kornia (v0.7.0). A robust transformation model was estimated with Random Sample Consensus (RANSAC). For a more precise alignment (~1 μm error, spot resolution), fine registration was performed, leveraging feature matching on H&E images and pseudoimages with higher resolution (~1.5 μm/pixel). Fiducial markers visible on both modalities were detected on the full-resolution images (0.345 and 0.5 μm/pixel) and appended to the features used for fine registration. The detection of fiducial marks was carried out automatically with custom YOLO-based object detection models. Likewise, the LoFTR-detected features and YOLO-detected marks were matched across modalities via brute-force and nearest neighbors matching, followed by RANSAC estimation of the transformation model. Since pairwise alignment is performed between an H&E image and ST data from the same section, rigid models were used to transform ST coordinates into H&E image space, as no distortions were expected. Finally, we used the GUI provided by our openst package (manual_pairwise_aligner_gui) to visually assess and manually refine the results of the automatic alignment pipeline. The code for alignment, GUI for visual validation, feature detection models and examples for Open-ST pairwise registration are publicly available at https://github.com/rajewsky-lab/openst.

Finally, spatial cell-by-gene expression matrices were created by aggregating the initial NxG matrix (N, capture spots; G genes) into a MxG matrix (M, segmented cells; G, genes), where the mapping of N to M takes place via the segmentation mask. That is, for all segmented regions (see "H&E image preproccesing and cell segmentation"), spots falling within the spatial coordinates of a segmented region are aggregated into a single "segmented cell" with identifier equal to the cell mask label.

We assessed the accuracy of the pairwise alignment based on fiducial marks by calculating the reproducibility of two manual alignments performed on the same randomly chosen sample, (section #9 of the metastatic lymph node). Two independent annotators (L.M.S. and D.L-P.) selected at least two pairs of corresponding points per tile, using the GUI tool (manual_pairwise_aligner_gui) starting from the same automatic coarse alignment, that was performed with pairwise_aligner. Similarity matrices were computed for each tile from both lists of keypoints and applied to the coarse-aligned barcode coordinates. Lastly, the distribution of euclidean distances across annotators was calculated.

Generation of Open-ST pseudoimages

The generation of *pseudoimages* was tailored to distinct visualization needs, encompassing two methodologies. On the one hand, for aggregated cell-level depictions in 2D, the pl.spatial function from the scanpy v1.9.3 package was used.

On the other hand, higher-resolution visualizations focusing on individual transcripts (typically, at ~100x100 μm zoom) relied on local 2D Kernel Density Estimation (KDE) of UMI-scaled spatial coordinates, for any chosen transcript. The KDE was performed with a bandwidth of ~1 μm. For visualization, the KDE is treated as image data visualized with the microshow function from microfilm v0.2.1 package. Intensity limits were adjusted from the 5th to the 95th percentile. These renderings are useful to visualize and quantify transcript density in space, rather than the sparser raw counts.

Alternatively, raw UMI counts can be displayed as an image by producing an empty canvas defined by re-centered coordinate bounds, mapping coordinates onto the canvas, and aggregating UMIs within pixel bins. These resultant images resemble conventional pixel-based grids and can be readily subjected to standard image processing techniques.

3D reconstruction

The Spatial Transcriptomics ImgLib2/Imaging Project (STIM, v0.2.0) was leveraged for the alignment of the 19 Open-ST sections of the metastatic lymph node dataset. For alignment purposes, we treated sections as sequential. First, the coordinates of each pairwise-aligned and segmented h5ad file were rescaled by a factor 1:2. The coordinate and gene expression information of these datasets were converted into the n5 format, optimized for efficient image processing, via the st-resave function. The st-align-pairs function was subsequently used for pairwise alignment of three sections below and above each section (r=3), by creating image channels of the expression of prespecified genes, aggregated per cell as a Gauss rendering around centroids, parametrized with a smoothness factor. We qualitatively assessed *KRT6A*, *KRT6B*, *S100A2*, *LYZ*, *CD74*, *IGKC*, *IGHG1*, *IGHA1*, *JCHAIN*, *CD74* and *AMTN* to have high expression and spatial variation within sections. This function results in a set of feature matches between pairs of sections, filtered via affine model; st-align-pairs was parameterized with—minNumInliers 15, –scale 0.03, and -sf 4.0 (smoothness factor). Finally, we ran st-align-global to find affine transformations that globally minimize the distances of feature matches between sections, across the entire stack. The n5 container was then converted back to the h5ad format, for subsequent downstream analyses. We additionally transferred the resultant affine transformation matrices onto preprocessed and background-removed H&E images, rescaled to an equivalent 1:4 factor. An aligned imaging volume, used for subsequent 3D visualization, was created via the transform.warp function from scikit-image. This can be reproduced with the from_3d_registration program from our openst package.

3D renderings

Three-dimensional rendering was used to display the two *raw* data modalities comprising Open-ST data (tissue staining images and spatial transcriptomics data), and results from downstream analyses (i.e., cell clustering).

Tissue staining renderings were generated with ParaView v5.11, from the rescaled image stack resulting from the 3D alignment (see STAR Methods section "3D reconstruction"). The x-y axes were scaled 1:4 with respect to the physical dimensions, and the z axis was scaled 7:1. The x-y coordinates of the volume were clipped into a box of 4x3x0.36 mm, to remove empty areas outside of the tissue block, as well as areas with low section coverage along the z axis.

Raw or normalized gene expression of individual genes was visualized upon downscaling the expression levels per cell by a factor 1:40. Then, volumes of ~500x500x19 voxels were generated by concatenating ~500x500 pixel *pseudoimages* of the selected gene at each section. Upon concatenation, values were interpolated by summing the pixel-wise values from section *n-1* to section *n*, removing spatial irregularities due to uneven coverage along the z-axis. The voxelized volume data was smoothed with a gaussian filter (sigma=2), and exported as a TIFF file. Volume renders of gene expression were visualized with ParaView v5.11.

Surfaces of unsupervised, annotated clusters were generated from similarly constructed *pseudoimages* of categorical variables, with gaussian filtering and linear interpolation. This resulted in volumes of ~500x500x19 voxels, one per annotated cluster. Within ParaView v5.11, voxel data was thresholded for isosurface extraction. Surfaces were smoothed for 1,000 iterations.

Clustering analysis & cell typing

We performed cell type annotation of the E13 embryonic mouse brain, human primary head and neck squamous cell carcinoma, and matched healthy and metastatic lymph node samples, by following standard practices for single-cell analysis and by using scanpy. In the case of the metastatic lymph node, the following steps were performed on one section (#4), used to build a reference annotation that was subsequently transferred to the remaining 18 sections. Data preprocessing involved applying unique molecular identifier (UMI) thresholds per segmented cell (at least 250, at most 10,000), mitochondrial count filtering (at most 10% of the counts for the mouse sample, 20% of the counts for the human sample), and retaining genes expressed in a minimum of 10 cells. In the human primary tumor sample, we adjusted the cell filtering thresholds to at least 500, at most 10,000 UMIs per cell, and at most 15% of mitochondrial counts per cell. Then, we normalized the raw counts per segmented cell employing the scanpy functions sc.pp.normalize_total and sc.pp.log1p. Concurrently, we identified the top 2,000 highly variable genes through sc.tl.highly_variable_genes using the 'seurat' method. Following normalization, we performed dimensionality reduction via Principal Component Analysis (PCA) on the normalized counts of the selected highly variable genes. The subsequent construction of a nearest neighbor graph in the PC space, encompassing the first 30 principal components, was accomplished using sc.tl.neighbors. Then, we leveraged the Leiden algorithm for community clustering (sc.tl.leiden). We mapped Leiden clusters to cell types by identifying marker genes via sc.tl.rank_genes_groups, using the 'Wilcoxon' method, with a filter for $\log_2$FC (fold change) $\geq$ 1 and Bonferroni-adjusted p-value < 0.05. Subsequently, we visualized marker genes with sc.pl.rank_genes_groups_dotplot, normalizing expression per gene feature (standard_scale='var'). We annotated cell types from Leiden clusters by incorporating these unbiased marker genes and literature-informed markers.

For the E13 embryonic mouse head, we further clustered the cells at the forebrain, midbrain and hindbrain, separately, using the same dimensionality reduction and community clustering approach. For this, we manually generated three sets of cells based on their brain location, by filtering their x-y coordinates, and excluded cells that were annotated as chondrocytes, fibroblast, endothelial, mesenchyme, myocytes and blood. Then, clusters were annotated using literature markers. We compared our unsupervised clustering and annotation to the annotation from a reference atlas of the E13 mouse brain.[31] We used the scVI and scANVI models from scvi-tools v1.0.2 to perform label transfer of the reference atlas to the segmented cells from our Open-ST dataset, using default settings.[77] The degree of mixing before and after integration was measured as the fraction of k-nearest neighbors (k-NN) in the embedding space (scVI latent space) with different batch labels, via sklearn.neighbors.kneighbors_graph.

For the human primary HNSCC, metastatic and healthy lymph node samples, in a secondary pipeline we refined cell type labels by aggregating subsets of cells from different samples and performing joint annotation. This allowed us to validate the robustness of our annotation regardless of the sample of origin. For the metastatic and lymph node samples, we aggregated cells not previously annotated as tumor and excluding adipocytes, which we integrated with scvi-tools v1.0.2 on raw counts of the 2,000 most variable genes. We used the default scVI model trained for 100 epochs, with 2 hidden layers and 30 latent dimensions. Similarly, we approached tumor cell sub-clustering by consolidating tumor cells from primary and metastatic samples, followed by scVI integration for batch correction on the 2,000 highly variable genes. Tumor labels were identified through Leiden cluster numbers. Lastly, cell type annotations from the metastatic lymph node section 4 were transferred across the whole dataset (rest of sections) by projecting the PCs fitted on section 4 and using a *Nearest Neighbors* classifier to map cluster labels, via sc.tl.ingest with default parameters.

Gene set and communication analysis

We explored gene programs and cell-cell communication within the primary HNSCC, healthy, and metastatic lymph nodes starting from a Differential Gene Expression (DGE) analysis, by merging the samples following a pseudobulk approach. We iteratively computed the DGE between tumor subclusters (after reclustering of tumor cells annotated as 'tumor', 'keratin pearl' or 'proliferating') versus rest, using the sample of origin and binary indicator of cluster as design matrix, with pydeseq2.[86] We discerned differentially expressed genes per cell type (one versus rest) within each sample by including features with an absolute $\log_2$FC > 0.5 and an adjusted p-value < 0.05.

Subsequently, the identified differentially expressed genes per cluster underwent Gene Set Enrichment Analysis (GSEA) using the dc.get_gsea_df from the decoupler-py package. The Reactome pathway gene set signatures were employed, filtering for a FDR p-value < 0.05 and retaining programs with a NES (Normalized Enrichment Score) > 1. This analysis was executed on the various tumor cell subclusters from metastatic and primary samples, shedding light on the active pathways within unbiased clusters. We additionally validated the localization of these programs in space through AUCell scores (implemented in decoupler-py) for the respective signature gene sets.

Moving to spatial receptor-ligand analysis, we utilized the liana-py package, specifying a distance of 50 μm on a manually-curated consensus database of interactions.[78] Spatial hotspots were approached as non-negative factors, computed from normalized communication scores via the decomposition.NMF function from scikit-learn, on tumor and stromal cells. The optimal number of factors was automatically selected upon the elbow method, with the kneed package, from the reconstruction error of the inferred factors and the training data. Factors were literature-informed annotated based on the top ligand-receptor pairs contributing to the non-negative factors.

Resolution analysis

Pseudoimages were generated for the E13 mouse head dataset through the 2D KDE protocol at specific ROIs. The detectable scale of changes of transcript density was analyzed along manually selected lines, drawn to pass from tissue regions with *baseline* (off) to high levels (on) of expression of a specific marker. This assesses the scale at which changes from *off* to *on* can be detected, as a proxy for the effective resolution of the method (upper bound). Spatial density changes along the lines were measured via the measure.profile_line function from the scikit-image v0.19.3 library. Notably, for transcripts like *Ttr* (choroid plexus region), a line width of 10 pixels was employed, while for *Atoh7* (eye marker), a line width of 50 pixels was utilized. The findings of this analysis were subjected to qualitative comparison with analogous regions from a sagittal E13.5 brain, as cataloged via in-situ hybridization in the Allen Developing Mouse Brain Atlas as of 20 December 2023, for *Neurod6* (Image 2), Pbx3 (Image 2), *Tubb3* (image 7), *Atoh7* (image 1), *Ttr* (image 7), *Cnr1* (Image 3), *Dbi* (Images 1 and 4), *Eomes* (Image 4), *Tbr1* (Image 5), *Pax6* (Image 2), *Shox2* (Image 8), *Npy* (Image 8), *Six3* (Image 10), *Enc1* (Images 3 and 5), *Nes* (Image 6), *Htr2c* (Image 5), *Cyp26b1* (Image 6), and *Foxp2* (Image 5).

Subcellular localization analysis

To study the average subcellular localization patterns of transcripts, the adult mouse hippocampus dataset, consisting of transcript locations at ~0.6 μm spots and the pairwise-aligned segmentation mask, was partitioned into a regular grid of 70x70 μm squares. The cellular density of all patches was assessed by computing the proportion of nuclear segmentation mask coverage over total patch area. Patches were further categorized into three tiers: 10-15%, 15-30%, and 30-50% mask coverage, or nuclei density.

The same process was applied within each density group, for three transcript or transcript families: *Malat1* (nuclear), *mt-** (all mitochondrially-encoded transcripts), and *mt-Tt* (a transcript with similar total UMIs to *Malat1*). A 3.5 μm extension surrounding the nuclear segmentation mask was enacted, and for distinct cells, distance transforms were calculated utilizing the ndimage.distance_transform_edt function from scipy v1.10.0.[73] Through a linear rescaling and inversion, these distance transform values were mapped to the continuous [0, 2] interval, wherein 0-1 corresponds to the nuclear core and 1-2 pertains to the extended boundary. This allows for a standardized comparison across cells. For both *Malat1* and mitochondrial transcripts, the coordinates were projected onto the rescaled distance transform, producing the ''observed'' distributions that encapsulate the spatial behavior of the transcripts within cells.

An "uniform" distribution was generated by shuffling the location of transcripts prior to projection to the distance transform, for later statistical comparisons. Kernel density estimation was performed, utilizing 40 bins within the [0, 2] interval, to approximate a smooth distribution shape. Measurements from all cells were aggregated, and an observed-to-uniform ratio was computed for each

transcript or transcript family. These ratios were visualized within standardized cells, via polar projection of the density estimates at each rescaled distance value, equivalent to radii of a standardized (circular) cell. As an additional validation, the histogram and kernel density estimate of *Malat1* and *mt-** counts respect to the distance to nuclear edges was computed. To this end, we projected the points into a distance-transformed mask, where zero values correspond to the nuclear edge (pixels delimiting the boundary between nuclear segmentation and background), negative values point towards the centroid of the nucleus, and positive values diverge from the edge to the outside of nuclei.

Crosstalk and spatial bias analysis

To evaluate the impact of spatial biases of Open-ST data on the transcriptomic profiles of identified cell types, especially in spatial locations with high transcriptomic diversity, we quantified crosstalk in the E13 mouse head. More specifically, we examined the biases in transcriptomic profiles of cells by the spatial biases in transcript capture by defining crosstalk as how the expression from a cell of population *a* has components of cell of population *b*, depending on their pairwise spatial distance. We define a crosstalk coefficient k_{ab} as the expression of markers from cell type *b* measured at cells *a*:

$$k_{a,b} = \int_{\alpha_1}^{\alpha_2} H_b(\alpha) d\alpha$$

where α_1 and α_2 are the expression (or gene set score) limits for population *a*, and H_b is a function that returns the amount of expression (or gene set score) for population *b*, measured at level α in population *a*.

We subset the E13 mouse head data to the cells of the clusters labeled as 'Blood' and 'Fibroblast' due to their widespread presence across the sample and their distinct marker expression, determined through unbiased clustering (detailed in the 'Clustering' section). We performed the following analysis on specific cell pairs using either a segmentation mask or a hexagonal grid with a 7 μm radius: first, we calculated marker gene set scores for these clusters using the scanpy function scanpy.tl.score_genes; then, we calculated pairwise spatial distances between 'Blood' and 'Fibroblast' cells. Employing a spatial filtering criterion of a center-to-center distance under 20 μm, we generated 2D contour plots of kernel density estimates, portraying gene set scores of 'Fibroblast' and 'Blood' markers in the x and y axes grouping the density clouds based on the label of the cells of origin. In practice, the crosstalk coefficient k_{ab} measures the overlap of the two KDE clouds. Finally, to extend the analysis, we performed this analysis on a dataset with simulated lateral diffusion for individual transcript coordinates, with an equivalent diffusion coefficient D= 1 $\mu m/s^2$ for 100 seconds, via random walk (dt = 1 s).

We start from the assumption that (segmented) cells should be unique sources of transcriptomic information in the dataset, whose location in the transcriptomic manifold can be clustered into distinct populations or types. Within a proximity of 20 μm or less, we observed a higher degree of mixing of cell signatures than at greater distances, suggesting the existence of local biases in transcript capture, which may lead to unexpected gene expression patterns. Specifically, cells labeled as 'Chondrocyte', characterized by markers *Col2a1*, *Col9a1*, *Col11a1*, *Sox9* and *Col9a3*, often exhibited expression of markers *Hbb-y*, *Hbb-bt*, *Hbb-bs*, *Hba-a1* and *Hba-a2* (associated with cluster 'Blood') when in close spatial proximity to cells annotated as 'Blood' (Figure S3D). We validated our quantification of crosstalk in two ways. Simulated diffusion of transcripts resulted in the merging of distinct cell populations, even at larger distances. Moreover, using a regular hexagonal grid as a segmentation mask led to the emergence of a mixed state corresponding to the merging of two cells under the same spatial location (Figure S3E). In datasets with perfect segmentation and no prior local biases in transcript capture, we expect to find no signal of the program *A* at cell *b*, and vice versa; that is, a scenario comparable to scRNA-seq data. These findings emphasize that our method may encounter limitations when dealing with closely spaced cell populations, where segmentation is more challenging and lateral diffusion can have a stronger effect. Importantly, spatial bias can impact current dimensionality reduction and normalization approaches, as borrowed from single-cell analyses toolboxes.[63–65]

Neighborhood analysis

We quantified the spatial relationships between distinct cell types in primary HNSCC and metastatic samples by studying local cellular neighborhoods. Utilizing a Delaunay triangulation, we computed interaction graphs where two cells are connected if they share an edge. This provided a quantitative measure of the frequency of cell type neighborhoods, that further allows to assess the enrichment of neighborhoods via permutation tests. For this purpose, we employed the squidpy and CellCharter frameworks[75,87] and binarized the asymmetric neighborhood enrichment matrix to keep values with p-value < 0.05 and relative enrichment > 0, i.e., we ignore neighborhood *depletion* events.

Tumor boundary analysis

To define the boundary between the tumor mass and lymph node in the metastatic sample, we start from the radius neighborhood graph and define edges for pairs of cells with center-to-center distance smaller than 50 μm. The value for a node is then computed as the number of edges connecting to a cell type from the lymph node, if the node is a tumor cell, and vice versa. The set of these values

along all nodes is the boundary strength *S*. To calculate the correlation of two signals in space, e.g. gene expression to a boundary, we compute a value (*I*) inspired by the bivariate counterpart of Moran's I:

$$I = \sum_i \sum_j w_{ij} S_i X_j^+$$

where *w* is the weighted adjacency matrix from the spatial graph, X^+ is the query signal, and *S* is the boundary strength. X^+ results from setting negative values from *X*, the normalized and z-scored gene expression matrix, to zeroes. We analyzed the 9,000 genes with highest mean expression, and reported the ranking of top-50 genes with highest I^* values, computed as $I^* = 1^+(I)\,/1^-(I)$, where 1^c is an indicator function setting ones where the values fulfill the comparison term *c*. We computed the score of the set of top-50 genes sorted by highest I^* values with scanpy.tl.score_genes and visualized it in space to validate the enrichments.

Flow cell imperfection analysis

To quantify macroscopic imperfections exceeding 2 μm, an iterative approach was employed, separately for each tile at the the top and bottom parts of the S4 flow cell. First, a *pseudoimage* of the flow cell was generated by rescaling the coordinates of individual capture areas sourced from spatial barcodes (see STAR Methods section "Spatial reconstruction of barcode sequencing") by a factor of 0.1. This rescaling aligns consecutive spatial points to an interval of one unit length (1, 2, 3... corresponding to sequential positions), resulting in a 4430x4600 pixel image. Gaussian blurring, with a sigma of 5, was applied to the *pseudoimage*, followed by Otsu thresholding. The outcome is a segmentation of space filled by barcoded spots, wherein pixel values are either 1 (presence of spatial barcodes) or 0 (absence). By implementing this procedure, high-frequency fluctuations were suppressed, isolating larger-scale spatial voids.

This sequence of operations was replicated for all tiles within the flow cell. Subsequently, the proportion of pixels with a value of 0 within the segmentation mask is computed against the total pixel count. This value, termed the 'percentage of irregularities,' characterizes regions with deviations in the flow cell surface. Such irregularities may stem from diverse factors, including fiducial marks whose contrast is enhanced following blur and thresholding.

Comparison to Xenium

We ran the segmentation and transcript aggregation with Xenium Ranger 1.7, with nuclear masks + 3.45 μm extension, equivalent to the settings used by the Open-ST segmentation pipeline – otherwise, by default, Xenium output consists of 15 μm radial extensions from nuclear masks. Cells with less than 50 transcripts, and genes appearing in less than 10 cells, were removed from processing. Open-ST datasets were subset to the overlap of genes with the Xenium data (i.e., 316 for the primary HNSCC and 340 for the metastatic lymph node). Similarly, cells with less than 15 transcripts and 10 genes, and genes appearing in less than 5 cells were removed from further processing. The threshold for these filters were lower than for Xenium, as the median amount of transcripts and genes per cell was similarly lower (∼225 in Xenium, vs ∼30 in subset Open-ST). Then, both Xenium and the subset Open-ST data underwent the same normalization, dimensionality reduction, clustering and cell type annotation workflow as previously described (see STAR Methods section "clustering analysis & cell typing"). These were applied independently, without integration – specifically, annotation of transcriptomic clusters was performed using literature informed marker genes, on a per-sample and per-technology basis. Subclustering and marker gene analysis of tumor cells from Xenium data was run independently (i.e., without integration). Correlation of cell types was computed as the pairwise Pearson correlation coefficients of bulk profiles for all annotated clusters. Correspondence of ROIs between Xenium and Open-ST was defined manually, based on visual features and the knowledge from tissue sectioning. For the ligand-receptor (LR) comparison, we computed total counts (sum across all cells) for the genes from the consensus database available in liana-py v1.0.1, across the Xenium and Open-ST primary HNSCC samples. The amount of LR genes with *non-background* expression was estimated via the knee point of the cumulative counts of ranked LR total counts, using the kneed python package. For the comparison of spatial patterns, Moran's I values for all genes were computed for Xenium and Open-ST data using squidpy v1.3.0.

Benchmarking analysis

Slide-seq, Seq-Scope and DBiT-seq data were downloaded from the Gene Expression Omnibus (accession codes: GSE197353 (sample GSM5915059), GSE169706 (sample GSM5212844, runs SRR14082753-SRR14082757), and GSE137986 (sample GSM4364242), respectively. Visium data was downloaded from the 10x Genomics website and Stereo-seq data from the CNGB Nucleotide Sequence Archive (experiment ID: CNX0422301).[10,16,88] All barcode files were formatted to: bc \t x_pos \t y_pos. To reduce computational costs for the Stereo-seq data the barcode file was split into chunks of 10 M barcodes each and 1.5 billion of the 5.2 billion total reads were randomly sampled. The Seq-Scope data were downsampled to a total of 570 M total reads. All datasets were processed using Spacemake (v0.7.3).[27] Seq-Scope, Stereo-seq and Open-ST data were binned into hexagons with 7 μm edge length, resulting in an area of 127.3 μm^2 per spatial unit. To remove background signal and low quality beads cutoffs of 20, 180, 800, 5000, 200, 700, 700, 350, 600 UMIs per spatial unit were applied to Slide-seq, Seq-Scope, DBiT-seq, Visium, Stereo-seq, and Open-ST mouse brain, healthy lymph node, metastatic lymph node, primary tumor data, respectively. A field of view with high tissue coverage, high total UMI counts and, if possible, a moderate variance in UMI count distribution per spatial unit to account for

differences in e.g. cell type composition was chosen (Methods S1). Assuming sizes of 78.5 μm^2, 625 μm^2, 2,376 μm^2 and 127.3 μm^2 for Slide-seq, DBiT-seq, Visium, and Seq-Scope, Stereo-seq, Open-ST, respectively, spatial units within the field of view were randomly sampled such that the total covered area corresponds to 1 mm^2, resulting in 12,738, 1,600, 420 and 7,855 spatial units, respectively. For each dataset, reads belonging to the sampled spatial units were extracted from the output.bam file created by spacemake with samtools (v1.17) view -D and afterwards shuffled using samtools collate. The shuffled reads were downsampled in steps of 4 M reads up to a maximum of 40 M. All the downsampling files were processed using the DigitalExpression function from Drop-seq tools (v2.5.1) to obtain the genic reads and transcript numbers per barcode from the resulting summary files. These were summed to get the total number of genic reads and transcripts in the 1 mm^2 area. The total number of transcripts was divided by 10,000 to get the average number on a 100 μm^2 area. The total number of reads was computed as: total reads = total genic reads / (1 - PhiX mapping%) / (1 - rRNA mapping%) / spatial mapping%, where PhiX, rRNA and spatial mapping% where obtained as specified in *Quantification*. Code to reproduce this analysis and the spacemake configuration file are provided in our repository (https://github.com/rajewsky-lab/openst/).

Additional resources

We have set up an online resource with detailed descriptions of all experimental and computational steps: https://rajewsky-lab.github.io/openst.

CellPress
OPEN ACCESS

Supplemental figures

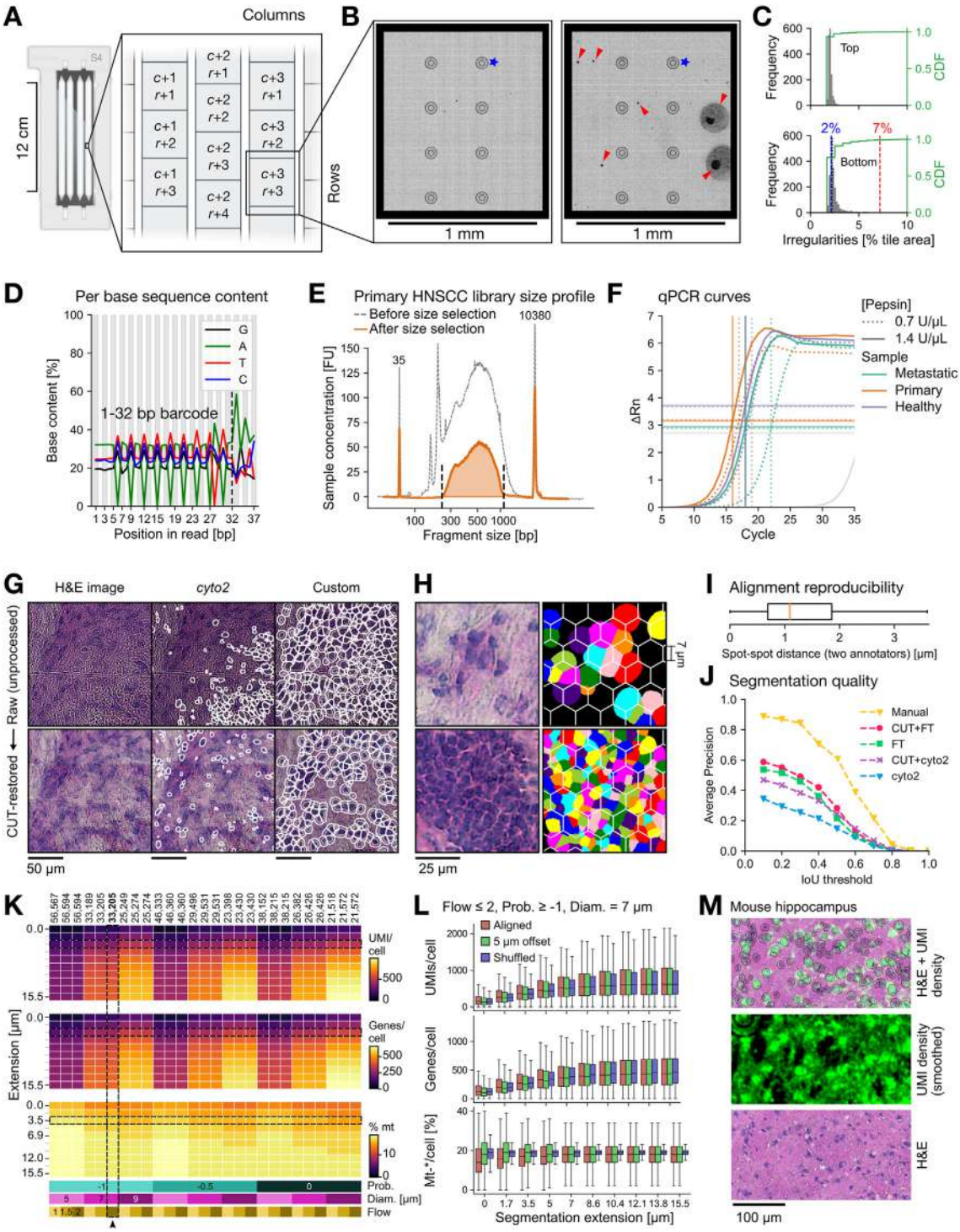


Figure S1. Open-ST: quality control and image processing, related to Figure 1

(A) A NovaSeq S4 flow cell consists of 4 lanes with a top and bottom surface, each with 6 columns (c) and 78 rows (r) per lane, totaling 3,744 tiles. Tiles are discrete sections of the flow cell imaged during sequencing. Distance between tiles is 55.5 µm in the x and 5.3 µm in the y axis.

(legend continued on next page)

(B) Representative example of tiles with average number of irregularities (left) and higher number of irregularities (right). Blue star: space without spots due to fiducial markers; red arrowhead: imperfections that may result from flow cell manufacturing errors (small black dots), bubbles during sequencing (medium black areas), or dust obstructing the imaging (large black areas).
(C) Frequency of irregularities (>2 μm) on the flow cell fc_1 from the 1st sequencing. The top flow cell surface contains fewer irregularities (CDF, cumulative density function).
(D) Per base sequence content of 1st sequencing of fc_1. Bases 1–32 correspond to the barcodes, with drops in "A" or "T" at expected sites in sequence (B or V in IUPAC nucleotide code).
(E) Automated electrophoresis profiles of the primary HNSCC library before and after size selection. Fluorescent units (FUs) relate to bioanalyzer input and not total sample concentration. Peaks at 35 and 10,380 bp are the upper and lower markers.
(F) Comparison of tissue permeabilization conditions via qPCR assay (STAR Methods). Example data with two pepsin concentrations tested per tissue type: permeabilization with higher concentration captured more mRNA.
(G) Image restoration and segmentation pipeline: raw images after stitching (upper) and restored images after applying a contrastive unpaired translation (CUT) model (STAR Methods). Left: H&E images used as input; middle: segmentation masks generated by Cellpose 2.0; right: segmentation masks via our fine-tuned version trained on manually annotated raw and CUT images (STAR Methods).
(H) Robustness of segmentation for different cell densities; two regions with ~65% (upper) and ~99% density (bottom) shown for the metastatic lymph node. CUT-restored images (left) were segmented with our fine-tuned (FT) model (right). Each segmented cell (nucleus with a 3.5 μm extension) is depicted with a different color. A regular lattice of hexagons (7 μm side) is overlaid onto the segmentation masks, illustrating the difference between data meshing and segmentation.
(I) Accuracy of the spatial, pairwise alignment between staining image and barcoded spots, measured as the pairwise Euclidean distance of barcoded spot coordinates after two independent manual selections of correspondences (STAR Methods).
(J) Benchmark of the segmentation models, measured as the average precision of the restoration + segmentation models at different intersection over union (IoU) thresholds (STAR Methods).
(K) Effect of user-defined *cellpose* parameters (Prob., cell probability threshold; Diam., expected cell diameter; Flow, flow threshold) on the median number of UMIs and genes per cell and % of mitochondrially encoded transcripts (mt). Parameters are indicated by the color-coded x axes (bottom). The number of segmented cells for each parameter combination is given above the top heatmap. Different non-overlapping expansion radii (in μm) are applied after nuclear segmentation and visualized in the y axis. The default parameters for the *openst* pipeline are highlighted using dotted lines (Prob., Diam., Flow) = (−1, 7, 2).
(L) Distributions of UMIs, genes, and % mitochondrial transcripts per cell for the default parameters used in the *openst* package. Three conditions are displayed: after pairwise alignment of segmentation masks and spatially mapped transcripts (aligned), after applying an offset to the pairwise alignment (offset), and after uniform shuffling of the location of all transcripts (shuffled).
(M) Qualitative assessment for the overlap of captured transcripts with respect to the cells segmented from H&E images (as nuclei + radial extension, with default parameters). (K), (L), and (M) show data from the mouse hippocampus.

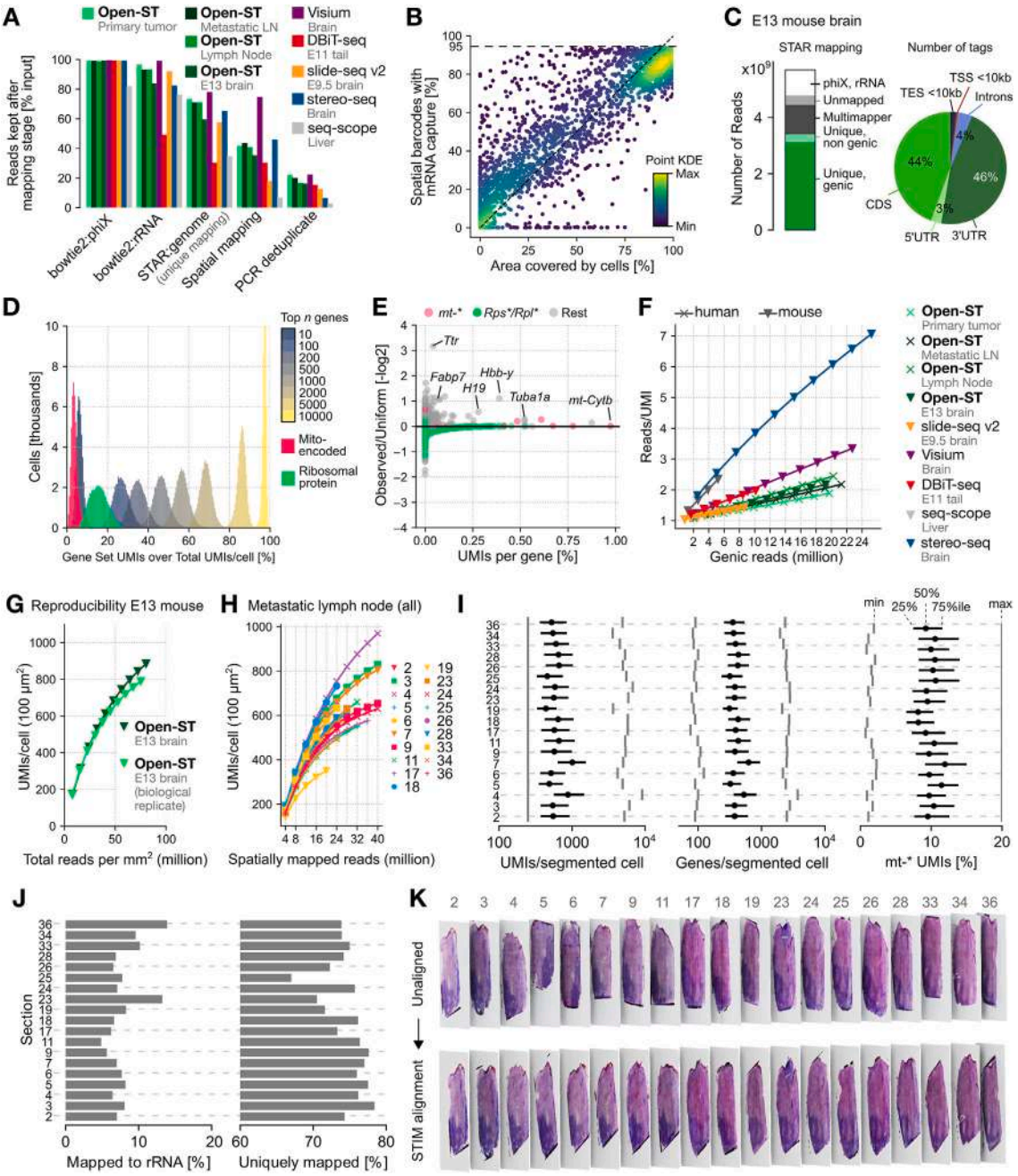


Figure S2. Reproducibility and performance of Open-ST, related to Figure 2

(A) Reads retained across different technology datasets after discarding those aligning to PhiX and rRNA, multi-mappers, reads not associated with a spatial barcode, and removing PCR duplicates. Coloring as in (F).

(B) Percentage of spatial barcodes with captured transcripts relative to the area covered by cells at non-overlapping ~5,000 μm^2 square regions. Diagonal dashed line: y = x; horizontal dashed line: y_{max} across the dataset. Color gradient shows point density.

(C) Left: stacked barplot of input reads filtered out during two-stage alignment against PhiX and human rRNA, and summarized STAR mapping statistics from the remaining reads, for the E13 mouse brain dataset (replicate 1); right: distribution of tags assigned to uniquely mapped reads—a single read may contain more than one tag (these are later quantified as ambiguous) (STAR Methods).

(D) Percentage of total unique molecular identifiers (UMIs) per segmented cell accounting for several gene sets: top *n* genes by total UMIs, mitochondrially encoded transcripts (*mt-**), and ribosomal protein transcripts (*Rps*/Rpl**).

(E) Transcriptomic information measured as the observed/uniform ratio of cells expressing a gene, over the total UMIs of that gene across cells.

(F) Library complexity measured as reads/UMIs ratio over number of genic reads across samples.

(legend continued on next page)

(G and H) Distribution of UMIs/pseudo-cell (100 μm^2) over total reads per mm^2 for the two biological replicates of the E13 head sample (G) and over spatially mapped reads for all 19 sections of the metastatic lymph node sample (H).
(I) Distribution of UMIs and genes captured per segmented cell, as well as % UMIs mapping to the mitochondrial genome, across all 19 sections of the metastatic lymph node.
(J) Left: % reads mapping to rRNA after excluding PhiX-mapping reads; right: % reads uniquely mapping to the genome after excluding rRNA-mapping reads. Data shown across all 19 sections of the metastatic lymph node.
(K) Stacked 3D rendering of H&E images before (top) and after global alignment with STIM (bottom).

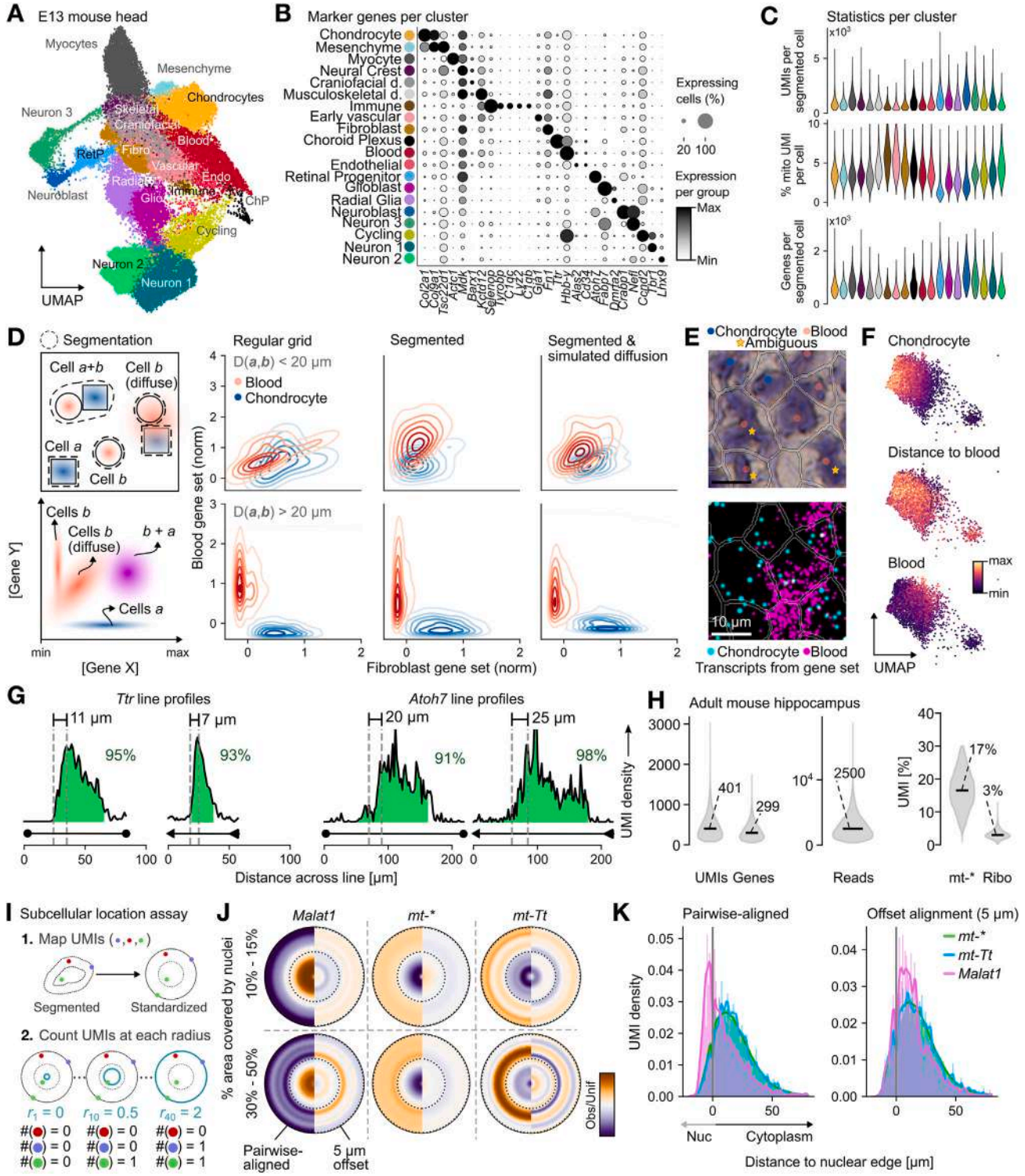


Figure S3. Localized transcript capture yields accurate annotation of mouse head cell types, related to Figure 3

(A) Uniform manifold approximation and projection (UMAP) visualization of the E13 mouse head clusters.

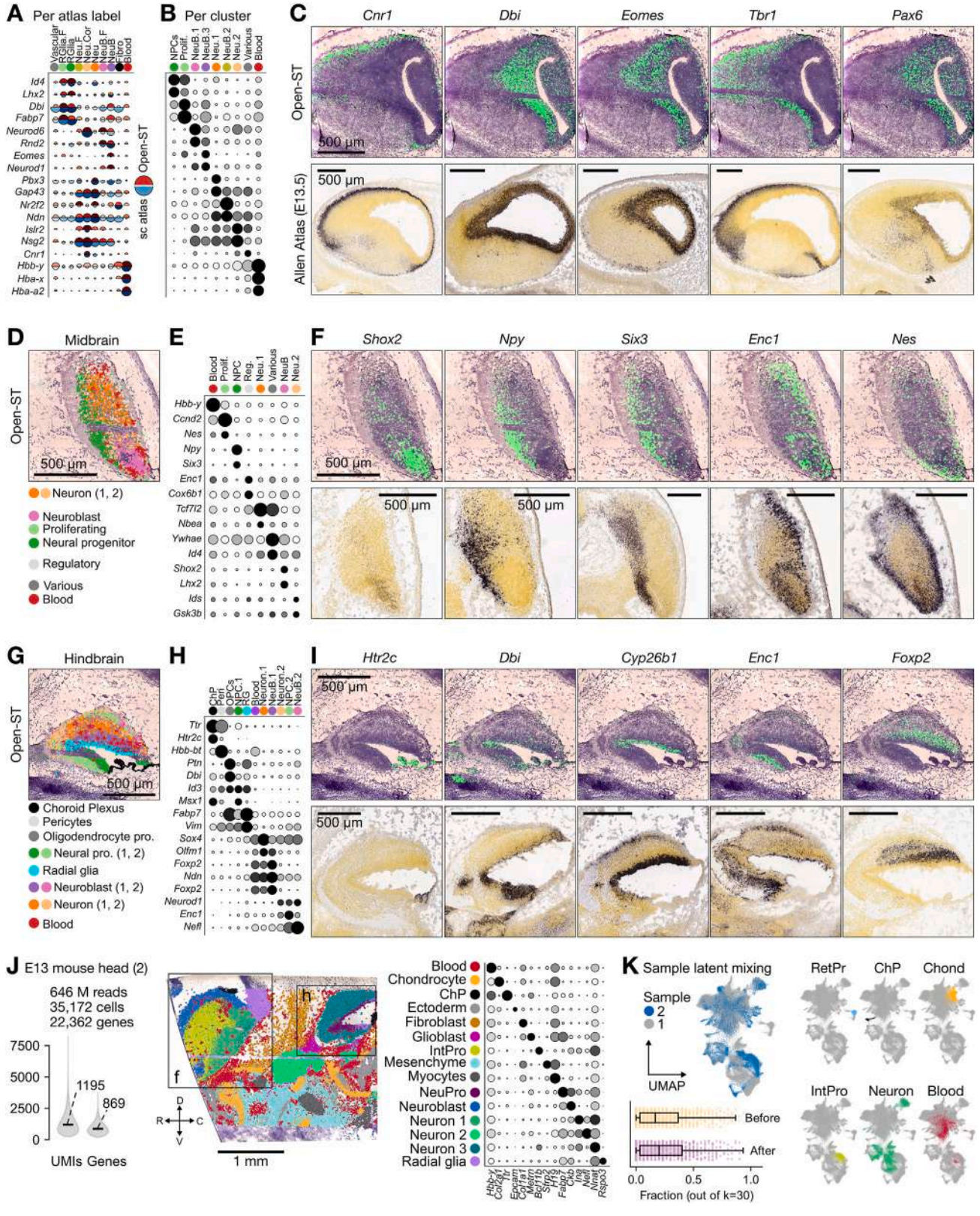
(B) Normalized expression of selected marker genes across the annotated Leiden clusters of the mouse E13 head.

(C) Distributions of UMIs, % of mitochondria-encoded transcripts, and genes per segmented cell for the mouse E13 head.

(D) Crosstalk analysis between blood and fibroblast clusters. Left, top: sketch of two cell types (*a*: circle, *b*: square) with their expression of marker genes in space (gene Y: orange, gene X: blue), alongside dashed outlines indicating cell segmentation. Left, bottom: measurement of gene expression of X and Y, in cells of types

(legend continued on next page)

a and *b*. Right: quantifications of the blood and fibroblast marker gene sets at cells. From left to right, cells are delimited via a regular grid (hexagons of 7 μm side) or via nuclear segmentation masks with radial extension. Additionally, lateral diffusion of transcripts is simulated and grouped into the same segmentation masks. Measurements focus on cells with neighboring blood or fibroblast cells at less (top) or more (bottom) than 20 μm center-to-center distance (STAR Methods).
(E) A region of the E13 mouse head with blood and "chondrocyte" cells in close proximity on the H&E staining with the segmentation mask outline (top) and the expression of marker gene sets in space (bottom). Cells clustered as blood but with low expression of blood and high expression of chondrocyte markers are indicated as "ambiguous."
(F) UMAP for the subset of fibroblast cells. From top to bottom, gene set scores for fibroblast markers, distance of cells to the nearest cell classified as blood in space, and gene set score of blood markers.
(G) Linear intensity profile of *Ttr* and *Atoh7* transcript density across selected tissue regions indicated by lines in Figure 3F Local maxima and minima used for distance measurement are depicted by the vertical dotted lines. Proportion of area under the curve within tissue boundaries (green) is given (STAR Methods).
(H) Distributions of UMI, gene, and read counts per segmented cell in a coronal section of the adult mouse hippocampus hemisphere (left) and the relative distribution of UMIs corresponding to mitochondria-encoded or ribosomal proteins (right).
(I) Outline of the method for mapping and counting spatial UMIs at a standardized (circular) cell.
(J) Observed over uniform UMI per radius (n_{radii} = 40) at two cellular densities, for *Malat1*, all mitochondrially encoded transcripts, and *mt-Tt* (STAR Methods).
(K) Distribution of UMI counts with respect to the nuclear edge. Left: spatial distribution of UMIs for *Malat1*, *mt-Tt*, and all mitochondria-encoded transcripts. Right: distribution profile after applying a two-dimensional offset of 5 μm to the pairwise-aligned spatial coordinates. Negative distances correspond to space within the nucleus (STAR Methods).



(legend on next page)

Figure S4. Marker gene localization in E13 murine brain regions and reproducibility on independent biological replicate, related to Figure 3

(A) Normalized expression of selected marker genes of the E13 mouse forebrain in the segmented cells of the Open-ST data (red, top) and in the cells from the E13 reference atlas (blue, bottom), grouped by the labels in the published atlas (STAR Methods).[31]

(B) Normalized expression of selected marker genes of the E13 mouse forebrain subclusters, grouped by the cluster annotation labels. Genes shown in (C) are indicated in bold.

(C) Localized capture of selected marker genes in the E13 mouse forebrain profiled with Open-ST (top) compared with *in situ* hybridization images of the E13.5 mouse from the Allen Developing Mouse Brain Atlas (bottom). High expression is colored in green (black) for Open-ST (Allen Atlas).

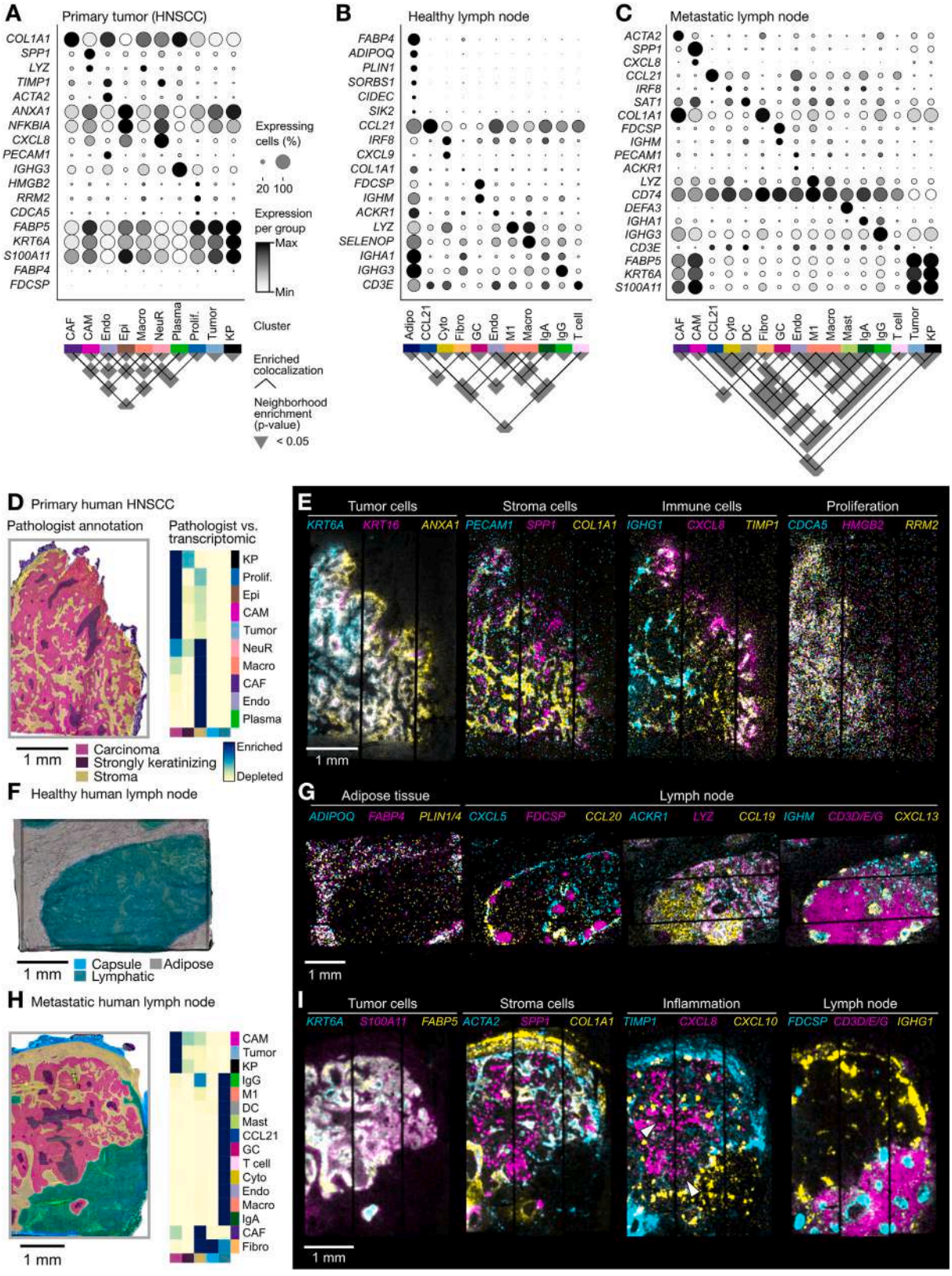
(D) Spatial distribution of E13 mouse midbrain subclusters (top) and corresponding region with annotation of morphological regions from the Allen Developing Mouse Brain Atlas (bottom).

(E and F) As in (B) and (C), but for the midbrain.

(G–I) As in (D)–(F), but for the hindbrain.

(J) Clustering and annotation of segmented cells of an independently Open-ST processed biological replicate of the E13 mouse head (sample 2; sample 1 was introduced in Figures 2 and 3). Left: distributions of the number of UMIs and genes per segmented cell. Center: spatial distribution of clusters overlaid on the H&E of the tissue. Forebrain (f) and hindbrain (h) locations are indicated by the boxes. Right: top-1 marker per cell type.

(K) Integration of the two E13 mouse head samples with scvi, using the sample of origin as batch label (see STAR Methods). Left top: uniform manifold approximation and projection (UMAP) computed from scvi embeddings, colored by sample of origin. Left bottom: sample mixing in latent space (before: 30D principal components [PCs] of both samples; after: 30D scvi embeddings). Right: localization of specific transcriptomic clusters in the joint embedding. RetPr, retinal progenitors; ChP, choroid plexus; IntPro, intermediate progenitors; NeuPro, neural progenitors.



(legend on next page)

Figure S5. Transcriptomic clusters are consistent with pathologist's annotation, related to Figure 4

(A–C) Normalized expression of selected marker genes (top) and neighborhood enrichment (bottom) across the annotated Leiden clusters of the primary HNSCC (A), the healthy lymph node (B), and the metastatic lymph node (C).

(D) Left: pathologist's manual annotation of tissue domains superimposed on the H&E image of the HNSCC section. Right: relative abundance of Open-ST transcriptomic clusters at manually annotated areas. Values are normalized per column.

(E) Spatial distribution of gene expression in the primary HNSCC section. Representative gene markers depict the molecular landscape of distinct tissue regions. Visualized as a merged representation of intensity channels, the coexpression of genes manifests as a "sum" of different colors (white) (STAR Methods). Pseudoimages show the smoothened expression of the indicated marker genes at cellular resolution (STAR Methods).

(F) As in (D), but in the healthy lymph node.

(G) As in (E), but in the healthy lymph node.

(H) As in (D), but in the metastatic lymph node section 4.

(I) As in (E), but in the metastatic lymph node section 4. Prolif, proliferating; KP, keratin pearl; NeuR, neutrophil-recruiting; Epi, epithelial; CAM, cancer-associated macrophage; Macro, macrophage; Endo, endothelial; CAF, cancer-associated fibroblast; GC, germinal center; Fibro, fibroblast; Cyto, cytotoxic T cell; M1, M1 macrophage; Fibro, fibroblast; Endo, endothelial; Adipo, adipocyte.

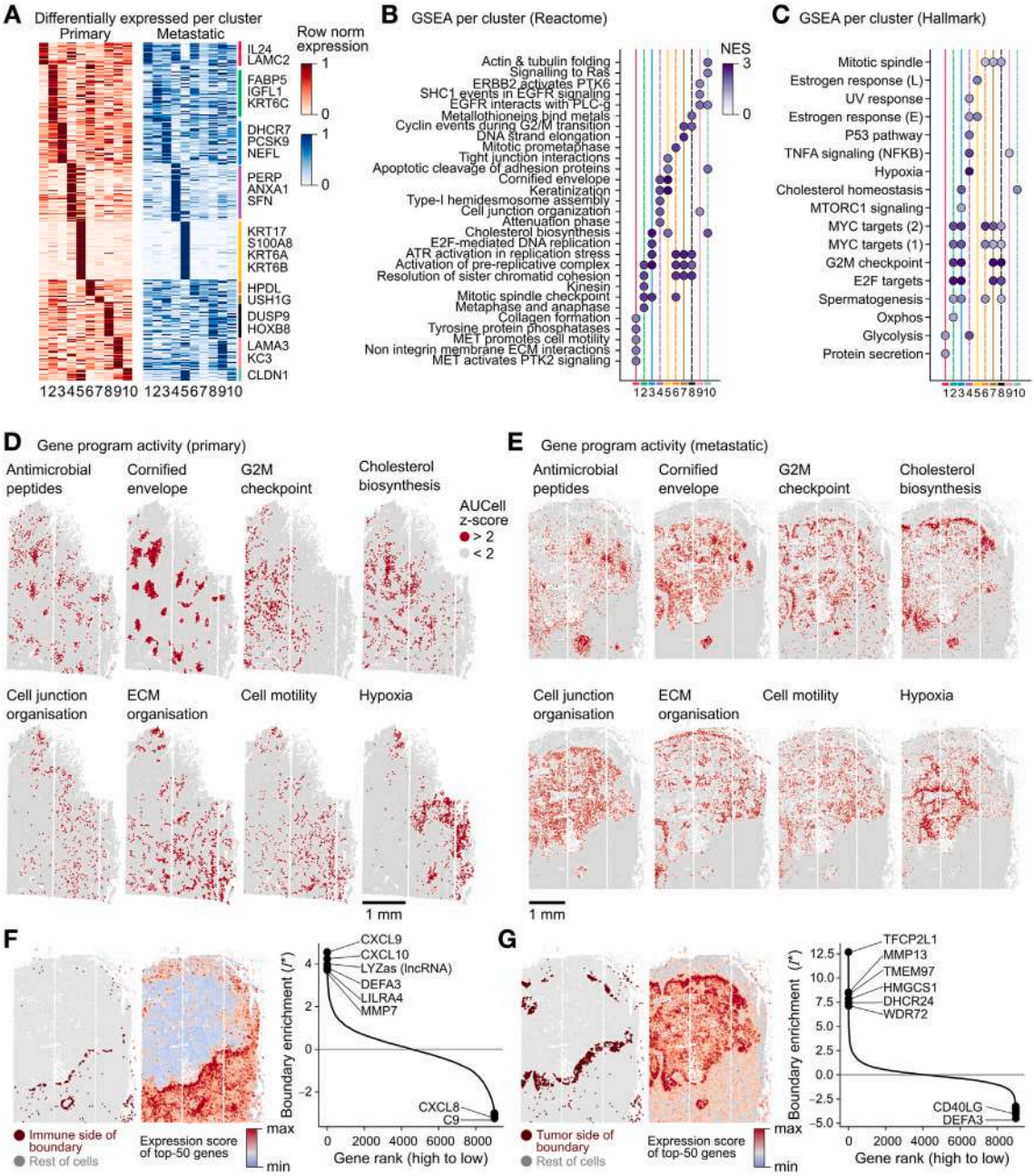


Figure S6. Tumor subclusters are linked to distinct spatial activity of gene programs, related to Figure 5

(A) Normalized expression of selected marker genes across the annotated Leiden clusters of tumor cells in the primary HNSCC (red) and the metastatic lymph node (blue).

(B) Top five reactome pathways sorted by normalized enrichment score per subclusters of the primary and metastatic HNSCC tumor cells, with NES > 1 and FDR-adjusted p value < 0.05 (STAR Methods).

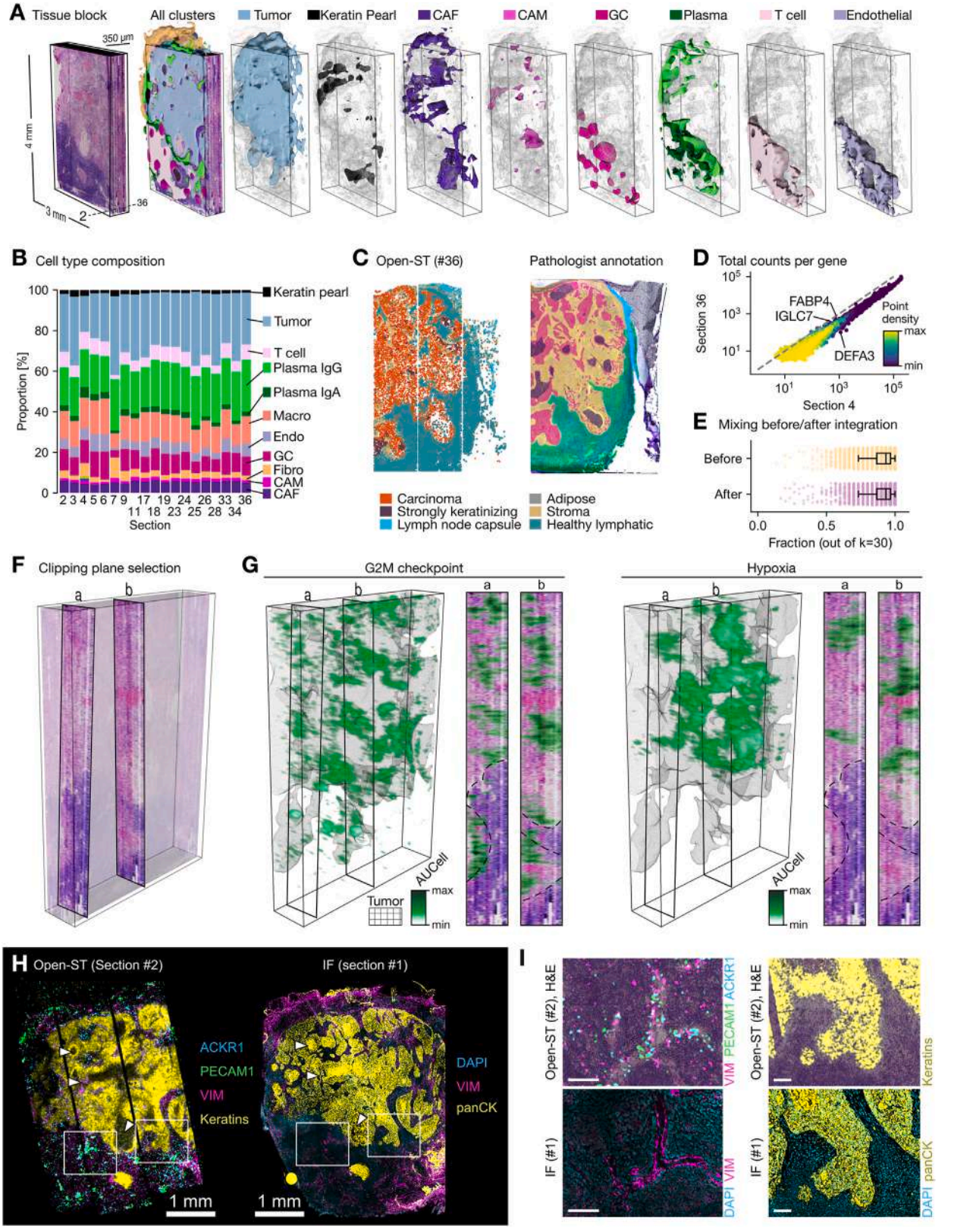
(C) Top hallmark pathways sorted by normalized enrichment score per subclusters of the primary and metastatic HNSCC tumor cells (STAR Methods).

(D) Spatial activity of reactome programs in the primary tissue, depicted as AUCell Z scores higher than 2 (STAR Methods).[44]

(E) As in (D), but for the metastatic lymph node tissue.

(F) Gene expression enrichment in cells located at the tumor-immune boundary of metastatic lymph node section 4. Left: the defined boundary at the lymph node side (STAR Methods); center: aggregated expression score of the top-50 genes, sorted by boundary enrichment values (l^*); right: sorted rank of boundary enrichment values, across the 9,000 genes with the highest mean expression.

(G) As in (F), but for cells residing in the tumor side of the tumor-immune boundary.



(legend on next page)

Figure S7. Dissecting the multimodal 3D virtual tissue block, related to Figure 7

(A) 3D rendering of tumor, stromal, and immune clusters as smooth surfaces overlaid on the virtual tissue block. Remaining clusters are shown as skeletal representation (wireframe) in the background.

(B) Cell type composition across the 19 sections of the metastatic lymph node.

(C) Tissue domains of a section as identified by Open-ST (left) and manually annotated by the pathologist, superimposed on H&E image (right). Open-ST clusters were merged for comparing to pathologist-annotated domains: carcinoma (tumor), strongly keratinizing (keratin pearl), stroma (CAF, CAM), healthy lymphatic (T cells, IgA/IgG plasma cells, macrophages, endothelial, germinal center, Mast cells, cytotoxic T cell, and CCL21-expressing), lymph node capsule (fibroblasts), adipose (adipocytes).

(D) Correlation of gene expression between section 4 and section 35 for cell type markers, illustrating the relationship of mean gene expression levels (depth-corrected and log-normalized) across cells of spatially distant sections. Diagonal dashed line: y = x.

(E) Sample mixing in latent space, expressed as the fraction of neighbors in the k-NN graph (of PCs) that have different section identifiers—values closer to 1 indicate better mixing. Before: 50D PCA computed jointly for all samples; after: 50D PCA computed on section 4, then projected to the rest of sections (STAR Methods).

(F) Two clipping planes orthogonal to the cutting direction (*a*, *b*) taken from the staining channel, used to project gene set activity and gene expression.

(G) Spatial activity of the G2M checkpoint and hypoxia hallmark gene sets, quantified per segmented cell with AUCell and visualized as smoothed volumetric renderings.[44] Tumor surface is shown as a skeletal representation (wireframe) in gray.

(H) Comparison of Open-ST to immunofluorescence staining of a consecutive, 10 μm-thick section (IF section #1, aligned onto Open-ST section #2). Arrowheads indicate necrotic areas in the tissue showing discordant mRNA and protein levels.

(I) Close-up views of two ROIs, indicated in (H) with white rectangles, for a region with an endothelial venule (left) and a region containing tumor and lymph node tissues (right). Scale bars: 100 μm. Macro, macrophage; Endo, endothelial; CAM, cancer-associated macrophage; CAF, cancer-associated fibroblast; GC, germinal center; Fibro, fibroblast; Cyto, cytotoxic; M1, M1 macrophage; Fibro, fibroblast; Endo, endothelial. Keratins: combined expression of KRT1/2/3/4/5/6A/6B/7/8/10/14/15/16/19.