



# HHS Public Access

## Author manuscript

*Nat Methods.* Author manuscript; available in PMC 2025 April 10.

Published in final edited form as:

*Nat Methods.* 2025 January ; 22(1): 145–155. doi:10.1038/s41592-024-02501-5.

## Repurposing large-format microarrays for scalable spatial transcriptomics

**Denis Cipurko<sup>1,2</sup>, Tatsuki Ueda<sup>1</sup>, Linghan Mei<sup>1</sup>, Nicolas Chevrier<sup>1</sup>**

<sup>1</sup>Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL, USA.

<sup>2</sup>Medical Scientist Training Program, University of Chicago, Chicago, IL, USA.

### Abstract

Spatiomolecular analyses are key to study tissue functions and malfunctions. However, we lack profiling tools for spatial transcriptomics that are easy to adopt, low cost and scalable in terms of sample size and number. Here, we describe a method, Array-seq, to repurpose classical oligonucleotide microarrays for spatial transcriptomics profiling. We generate Array-seq slides from microarrays carrying custom-design probes that contain common sequences flanking unique barcodes at known coordinates. Then we perform a simple, two-step reaction that produces mRNA capture probes across all spots on the microarray. We demonstrate that Array-seq yields spatial transcriptomes with high detection sensitivity and localization specificity using histological sections from mouse tissues as test systems. Moreover, we show that the large surface area of Array-seq slides yields spatial transcriptomes (i) at high throughput by profiling multi-organ sections, (ii) in three dimensions by processing serial sections from one sample, and (iii) across whole human organs. Thus, by combining classical DNA microarrays and next-generation sequencing, we have created a simple and flexible platform for spatiomolecular studies of small-to-large specimens at scale.

---

The architecture of a tissue determines its function and malfunction in health and disease. Therefore, technologies to study the spatial organization of cells and molecules in tissues are fundamental to biomedical research and clinical pathology. Recent advances in the field of spatial transcriptomics (ST) have enabled the sequencing of the transcriptome associated with specific areas of tissue sections<sup>1–4</sup>. Transcriptome-wide ST methods most commonly

---

**Correspondence and requests for materials** should be addressed to Nicolas Chevrier. nchevrier@uchicago.edu.

Author contributions

N.C. developed the concept. D.C. and N.C. designed the experiments and analysis approaches. D.C. performed the experiments and conducted the analyses. L.M. and T.U. performed optimization experiments. D.C. and N.C. wrote the manuscript with input from all authors.

Code availability

All scripts are publicly available at Zenodo via <https://doi.org/10.5281/zenodo.10963424>.

Competing interests

D.C. and N.C. are authors on patent PCT/US23/13010 covering the described technology. The remaining authors declare no competing interests.

Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41592-024-02501-5>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41592-024-02501-5>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

Author Manuscript  
Author Manuscript  
Author Manuscript  
Author Manuscript

rely on mRNA capture via spatially barcoded oligo(dT) probes that are anchored to a solid substrate. The data obtained by ST profiling are composed of mRNA counts at each of the spots tiling the area of the substrate under the tissue section. Several technologies for ST profiling through mRNA capture on a slide have emerged from academic laboratories in recent years<sup>5–11</sup>. However, existing ST methods face limitations including small surface area, lack of compatibility with hematoxylin and eosin (H&E) staining, low throughput and high costs. In addition, the spread of these methods beyond their institutions of origin has been impeded by important barriers to adoption including the need for special instrumentation, expertise and custom reagents. One notable exception is that of the first platform developed for ST, now Visium, which has become widespread thanks to commercialization and ease of use<sup>5,12</sup>. However, this platform remains limited by its small surface area and high cost. Together, these gaps in ST technologies preclude broad adoption across fields and detailed spatiomolecular analyses across many tissue samples, small or large, for basic and clinical research.

To fill these gaps in spatiomolecular profiling methods, we hypothesized that classical oligonucleotide microarrays could be repurposed for ST. High-density oligonucleotide microarrays were a workhorse in biomedical research for more than a decade until the emergence of next-generation sequencing in the late 2000s. Microarrays have been widely used to measure changes in gene expression<sup>13</sup>, genomic DNA sequence and structure<sup>14</sup> or protein–DNA interactions<sup>15</sup>. However, the oligonucleotide length (~25-mer to 85-mer) and its 3' tethering to the glass slide precludes using off-the-shelf arrays for ST because they cannot carry 5' anchored probes with an oligo(dT) in its 3' end for mRNA capture. In addition, the chemistry for oligonucleotide synthesis on arrays is not compatible with the addition of random bases needed for quantifying mRNA molecules using unique molecular identifiers (UMIs)<sup>16</sup>. Nonetheless, finding ways to repurpose microarrays into ST-compatible slides would be attractive in theory due to several advantageous characteristics of this platform. First, the synthesis of custom-design barcodes at predetermined *x*-*y* coordinates can be done on demand using *in situ* synthesis of oligonucleotides by an ink-jet printing method based on phosphoramidite chemistry<sup>17</sup>. On the contrary, most existing ST tools rely on the random dispersion of barcoded sequences across the two-dimensional (2D) plane of a substrate<sup>6,7,9–11</sup>. Therefore, these ST methods require additional sequencing steps before the ST profiling of tissue sections to determine the spatial localization of barcodes, which compounds with other barriers to adoption and increases time and costs. Second, over two decades of technology development and refinement by many academic and industrial groups have led to the emergence of robust array printing technologies that enable the rapid, high-accuracy, and on-demand synthesis of microarrays with custom sequences at low costs. Third, microarrays use glass slides of the standard microscope format, which makes them readily compatible with commonly used instrumentation for H&E staining, imaging, and automation.

Here, we transformed custom-sequence microarrays into ST slides using a simple, two-step reaction yielding spatially barcoded probes for *in situ* mRNA capture in tissue sections. Using large-format arrays, our method produced Array-seq slides carrying 974,016 spots of 30 µm in diameter (36.65 µm center-to-center distance) across 11.31 cm<sup>2</sup> in total surface area, thereby providing a high degree of flexibility in terms of the size and number

of samples that can be profiled on a single slide. We demonstrated that Array-seq generated high-quality data using mouse and human tissues in two- and three-dimensional configurations. We thus created a highly scalable ST platform poised to propel the broad adoption and application of spatiomolecular profiling across basic and clinical fields of inquiry.

## Results

### Assembly of mRNA capture probes on custom microarrays

We designed custom oligonucleotide microarrays with spatially barcoded sequences flanked by common sequences that enable the on-slide assembly of mRNA capture probes necessary for ST profiling (Fig. 1a and Methods). Custom microarrays were synthesized with 3'-anchored oligonucleotides (52-mer or 58-mer in total length) carrying spatial barcodes (12-mer or 18-mer) unique to each spot and flanked by two sequences, anchors 1 (24-mer) and 2 (16-mer), which were common across all spots on the array (Fig. 1b and Extended Data Fig. 1a). We initially used 58-mer oligonucleotides to facilitate the identification of the various oligonucleotide species generated during the optimization of our method for on-slide probe assembly. Subsequently, arrays carrying 52-mer oligonucleotides with 12-mer spatial barcodes were used for all ST profiling experiments to decrease next-generation sequencing cycle requirements (88 cycles in total). The anchor 1 sequence, tethered to the glass slide by its 3' end, matches the first 24 bases of the Illumina read 1 sequencing primer, and serves as a primer for both cDNA amplification and Illumina sequencing (Extended Data Fig. 1a). The 16-mer anchor 2 sequence matched the commonly used M13 forward (M13F) primer, chosen because (1) it is short, which limited the number of repetitive bases that are sequenced between the spatial barcode and UMI sequences, and (2) its melting temperature ( $T_m = 50.7\text{ }^\circ\text{C}$ ) was compatible with downstream cleanup steps to remove incorrectly assembled spatial probes (Extended Data Fig. 1a). For spatial barcodes, we designed 974,016 unique 12-mer sequences (hamming distance = 2) while minimizing self-dimerization and interactions with the two anchor sequences (Methods).

Next, to assemble ST probes on the microarray, we first hybridized arrays with two oligonucleotides: (a) a partial Illumina read 1 sequence (anchor 1), and (b) a probe for mRNA capture containing the reverse complement of anchor 2, a UMI, and an oligo(dT) (Fig. 1b and Extended Data Fig. 1a). Second, we performed an on-slide, extension–ligation reaction (or ‘gap-fill’) using a DNA polymerase to synthesize the reverse complement of the spatial barcode at the 3' end of the hybridized anchor 1, and a DNA ligase to attach the newly synthesized barcode to the phosphorylated, 5' end of anchor 2 (Fig. 1b and Extended Data Fig. 1a). The conditions for the gap-fill reaction were optimized using a DNA polymerase, Phusion, shown to have minimal strand displacement activity<sup>18</sup>, which would lead to the unwanted stripping of the annealed anchor 2 before ligation (Fig. 1c). Moreover, we maximized ligation efficiency while minimizing ligation biases toward specific sequences (Extended Data Fig. 1b,c). Third, after gap-filling, we sought to remove unligated anchor 2-UMI-Oligo(dT) oligonucleotides from the array to avoid a sink effect for tissue mRNAs, which would decrease sensitivity. Unligated anchor 2-UMI-Oligo(dT) oligonucleotides are hybridized on array probes by the M13F sequence only (16-mer with

$T_m = 50.7^\circ\text{C}$ ), whereas correctly ligated spatial probes (anchor 1-spatial barcode-anchor 2-UMI-Oligo(dT)<sub>30</sub>VN) are hybridized to the full length of each probe sequence (58-mer with mean  $T_m = 71.3^\circ\text{C}$  and s.d. = 0.5 °C across all spatial barcodes). We thus washed off unligated anchor 2-UMI-Oligo(dT) oligonucleotides using a temperature of 58 °C because it is between the two melting points of the ligated and unligated capture probes, yielding fully assembled mRNA capture probes at >75% purity (Fig. 1c).

Resulting ST-compatible arrays carried 974,016 spots of 30 μm in diameter (36.65 μm center-to-center distance) arranged in 1,068 rows and 912 columns across 11.31 cm<sup>2</sup> in total surface area usable for ST profiling (Fig. 1a). Each spot on the array carried a unique spatial barcode sequence at known  $x$  and  $y$  coordinates for localization, UMIs for quantification and an oligo(dT) for capturing poly-adenylated transcripts (Fig. 1b). Thus, our custom-sequence design followed by on-slide probe assembly yielded Array-seq slides compatible with downstream in situ reverse transcription to perform ST profiling (Extended Data Fig. 1a). To estimate how many cells were captured per spot on an Array-seq slide, we counted the number of nuclei in tissue sections that overlap with each spot across ten mouse organ types (Fig. 1d,e). Array-seq spots overlapped with a median number of nuclei ranging from 2 and 3 for muscle and brain to 15 and 19 for bone marrow and spleen (Fig. 1e). All other tissue types profiled—brown fat, liver, heart, lung, colon and kidney—displayed median numbers of nuclei per spot ranging from 5 to 8 (Fig. 1e). One Array-seq slide can therefore capture between ~2 million and 20 million cells depending on the tissue type and the number and size of the tissue sections to analyze.

### Array-seq yields high-quality ST data

To test Array-seq, we first profiled the main olfactory bulb (MOB) system of the mouse brain, a tissue of choice for the benchmarking of ST methods due to its small size and well-defined morphological layering<sup>5–7,9</sup>. Since Array-seq slides are generated from a standard microscopy slide substrate, tissue sectioning, fixation and staining can be done with standard reagents and slide scanners. Following tissue permeabilization, library preparation and sequencing of the MOB, we detected 3,582.2 UMIs ± 1,553.8 s.d. and 1,971.5 genes ± 709.4 s.d. per spot on the array, whose patterns of intensity reflected the histological features of the tissue section as observed by H&E imaging (Fig. 2a–c). We observed similar numbers in an independent MOB section with 3,492.1 UMIs ± 1,437.9 s.d. and 1,968.9 genes ± 682.7 s.d. per spot and a high degree of correlation between our two MOB datasets (Pearson’s correlation coefficient 0.986; Extended Data Fig. 2a). Next, we applied unsupervised clustering to the MOB data using the Leiden algorithm<sup>19</sup>, which identified clusters corresponding to all known layers of the MOB tissue (Fig. 2d,e and Extended Data Fig. 2b,c). Using different clustering algorithms led to similar results recapitulating all the layers of the MOB from the olfactory nerve layer to the granule cell layer, although only the Leiden algorithm correctly identified the inner-most, rostral migratory system layer<sup>20,21</sup> (Extended Data Fig. 2d,e). Moreover, the spatial distribution of marker genes for MOB tissue layers agreed with in situ hybridization data published by others<sup>22</sup> (Fig. 2d), and with histological regions defined by our H&E imaging data (Fig. 2f and Extended Data Fig. 2f).

Author Manuscript  
Author Manuscript  
Author Manuscript  
Author Manuscript

Next, we examined the potential diffusion of mRNA molecules outside their physical location of origin by computing the percentage of UMIs captured under the tissue section versus outside it. We found that an average of 3,581.7 UMIs per spot were detectable under the tissue compared to 168.76 outside that area across both MOB datasets combined, suggesting that on average 4.5% of the total UMIs detected were not under the tissue and likely resulting from diffusion (Extended Data Fig. 3a–c). For comparison, a publicly available MOB dataset generated on the Visium platform showed 5.4% of UMIs being detected by mRNA capture probes localized outside the tissue section, suggesting similar diffusion rates across Array-seq and Visium platforms (Methods). Furthermore, the spatial containment of marker genes for each histological layer of the MOB was well preserved and closely followed the anatomical regions defined by H&E and unsupervised clustering (Fig. 2f).

Lastly, we sought to test the spatial assignment of cell types across the histological layers of the MOB tissue. We used conditional autoregressive-based deconvolution (CARD) to infer the cell-type composition for each Array-seq spot under the two MOB sections<sup>23</sup>. We used publicly available single-cell RNA-sequencing (RNA-seq) data from the MOB tissue as a reference for spatial cell-type assignments<sup>24</sup>. We found that the cell types annotated across spots were correctly assigned to specific MOB layers (Fig. 2g). For example, immature migrating adult-born neurons localized specifically to the rostral migratory system layer and a subset of granule cells (GC-5) in the granule cell layer region (Fig. 2g).

We also found that the results obtained for cell-type calling across each Array-seq spot were comparable across various algorithms, including CARD, robust cell-type decomposition (RCTD) and Cell-2Location<sup>23,25,26</sup> (Extended Data Fig. 3d). Together, Array-seq generates ST profiles that detect genes, cell types and histological regions at high resolution.

### Benchmarking Array-seq against the Visium platform

We benchmarked Array-seq against the well-established commercial platform Visium (10x Genomics). Array-seq slides provided an 8.1-fold increase in spot density (861.2 versus 106.2 spots per mm<sup>2</sup>; Extended Data Fig. 4a,b), a 216.8-fold increase in total number of spots on the slide (974,016 versus 4,992 spots; Extended Data Fig. 4c) and a 26.7-fold increase in total active surface area (11.31 cm<sup>2</sup> versus 0.42 cm<sup>2</sup>). We profiled four sets of two immediately adjacent mouse kidney sections by dividing each pair of sections between Array-seq and Visium slides and using comparable sequencing depth across samples and platforms. On average, we obtained a total of 193,357,436 mapped reads  $\pm$  29,929,596 s.d. ( $n=4$ ) for Array-seq and 230,739,959  $\pm$  107,06,663 s.d. ( $n=4$ ) for Visium, leading to 5.39 mapped reads per  $\mu\text{m}^2$  spots in Array-seq and 7.64 mapped reads per  $\mu\text{m}^2$  spots in Visium. Moreover, a downsampling analysis revealed similar degrees of sequencing saturation and gene detection across both platforms at the sequencing depth used in our experiments (Extended Data Fig. 4d). We obtained ST data from an average of 30,579 spots  $\pm$  3,740 s.d. ( $n=4$ ) for Array-seq and 3,447 spots  $\pm$  430 s.d. ( $n=4$ ) for Visium using the four pairs of adjacent kidney sections (Extended Data Fig. 4e). While Visium captured more UMIs (25,820.6  $\pm$  1,322.6 s.d.) and genes (4,918.8  $\pm$  93.4 s.d.) per spot than Array-seq (2,563.5 UMIs  $\pm$  390.7 s.d. and 1,292.1 genes  $\pm$  119.7 s.d.) due to the larger size spot diameter (55

μm in Visium versus 30 μm in Array-seq), normalizing those numbers per square micron revealed a similar detection sensitivity between the two methods with 3.10 UMIs per  $\mu\text{m}^2 \pm 0.06$  s.d. and 0.55 genes per  $\mu\text{m}^2 \pm 0.01$  s.d. for Visium and 2.28 UMIs per  $\mu\text{m}^2 \pm 0.28$  s.d. and 1.09 genes per  $\mu\text{m}^2 \pm 0.08$  s.d. for Array-seq (Extended Data Fig. 4f–h).

Next, unsupervised clustering identified tissue structures matching known histological regions of the kidney for both methods, albeit at a higher resolution in Array-seq due to the higher density and smaller diameter of spots (Fig. 3a,b and Extended Data Fig. 4i,j). For example, finer tissue structures such as the renal urothelium or glomeruli were more consistently detectable and well defined in Array-seq than in Visium (Fig. 3a,b and Extended Data Fig. 4i,j). To further our analysis, we tested the spatial assignment of the cell-type composition for each spot using the CARD algorithm with publicly available single-cell RNA-seq data from mouse kidneys<sup>27</sup>. In agreement with the clustering analysis, we found that Array-seq outperformed Visium in its ability to find and correctly assign cell types at high resolution, including for cell types that were inferred to be present in small numbers of spots, such as podocytes (0.025% of total spots) or connecting tubule cells (0.02%; Fig. 3c and Extended Data Fig. 4i–k).

We then performed gene expression analyses to identify and compare subregion-specific gene markers across both platforms. The spatial distribution of differentially expressed marker genes for kidney tissue regions matched the zoning expected based on the H&E analysis for both methods although at higher resolution with Array-seq (Fig. 3d and Extended Data Fig. 5a). The top differentially expressed marker genes between kidney tissue regions agreed across both methods (Extended Data Fig. 5b). Moreover, our results suggested that the spatial distributions of genes, including highly expressed genes such as *Cryab* (renal medulla) or *Aadat* (outer stripe of outer medulla), showed a high degree of specificity for their expected histological regions, suggesting that spatial patterns of gene expression showed little to no alteration due to mRNA diffusion before capture on the slide (Fig. 3d and Extended Data Fig. 5a). Lastly, we confirmed that the mRNA expression profiles obtained with both methods correlated significantly with whole-kidney RNA-seq data (Extended Data Fig. 5c).

Together, Array-seq outperforms the gold-standard platform for ST profiling, Visium, in metrics such as spot density, diameter, and total surface area, while maintaining similar levels of sensitivity (for example, genes detected per surface area) and specificity (for example, expected spatial localization of marker genes).

### Array-seq enables high-throughput profiling in 2D and 3D

We tested the performance of Array-seq for the 3D profiling of spatial transcriptomes using the mouse kidney as a test system (Fig. 4a). The surface area of an Array-seq capture area ( $11.31 \text{ cm}^2$ ) allows for the analysis of serial mouse kidney sections in parallel on the same slide. To test this, we generated eight, 10-μm-thick sections by performing serial sectioning of a mouse kidney and analyzing one section every ~80–120 μm across a total of ~800 μm of tissue depth (Fig. 4b,c). We selected an area of depth in the kidney tissue that captured the major histological regions of the organ across all sections collected as shown by H&E imaging analysis (Fig. 4b,c). The histological features of the 3D kidney volume observed

by H&E were correctly recapitulated by unsupervised clustering analysis of the Array-seq data on the same tissue sections (Fig. 4d,e and Extended Data Fig. 6a–d). The surface area covered by each tissue subregion varied according to the depth of a given section within the kidney. For example, we observed an increase in proximal convoluted tubule-enriched regions of the renal cortex and a decrease in the outer stripe of outer medulla subregions as quantified by the proportion of spot matching to these histological regions across the z-stack (Fig. 4d, e and Extended Data Fig. 6e). Furthermore, the spatial distribution of marker genes for each of the tissue subregions identified by unsupervised clustering and confirmed by histology closely aligned with their expected regional expression (Fig. 4f,g and Extended Data Fig. 6f).

Next, we sought to assess the performance of our method for measurements of multi-tissue spatial transcriptomes (Fig. 5a). To do so, we profiled eight mouse tissue sections on a single Array-seq slide, including brain ( $n=2$ ), liver ( $n=3$ ) and kidney ( $n=3$ ) tissues (Fig. 5b). Each spatial profile correlated significantly among replicates (Extended Data Fig. 7a) and with whole-tissue RNA-seq from matching tissue types (average Pearson's correlation coefficient  $0.794 \pm 0.030$  s.d. for  $n=8$  tissue sections; Extended Data Fig. 7b). Unsupervised clustering across all eight sections correctly identified known tissue subregions for each organ type (Fig. 5c and Extended Data Fig. 7c). For example, our clustering data recapitulated the major brain subregions and the expected spatial zonation of periportal and pericentral hepatocytes within liver lobules<sup>28</sup> (Fig. 5c). These observations were further corroborated by the spatial expression patterns of marker genes for tissue subregions, such as *Mobp*, *Pcp2* and *Nrep*, found in the fiber tracts, the granular layer and the molecular layer of the cerebellar cortex regions of the brain, respectively, or *Cyp2f2*, *Cyp2e1* and *Glul*, found in the periportal and pericentral zones of the liver (Fig. 5d,e and Extended Data Fig. 7d). Lastly, spatial gene-set enrichment analyses correctly assigned the biological functions of an organ to its expected histological locations, such as fluid transport, regulation of systemic arterial pressure and gluconeogenesis in the kidney renal medulla, distal convoluted tubule-enriched renal cortex and proximal convoluted tubule-enriched renal cortex, respectively (Extended Data Fig. 8a,b). Similarly, we observed that sets of genes linked to specific tissue functions were enriched in the expected spatial regions of the brain and liver as well (Extended Data Fig. 8c–f).

These results thus suggested that scaling up the number of tissues processed per Array-seq slide did not alter data quality. The large surface area of Array-seq slides enabled the generation of spatial transcriptomes at high resolution and throughput, including in three dimensions.

### Array-seq enables whole-mount, human organ profiling

While a large surface area is useful for comparing many conditions or sample types, it could also, in theory, enable the analysis of whole-mount sections from larger tissues. To test this, we obtained fresh human spleen tissue, which we processed for Array-seq profiling (Fig. 6a). We generated whole-mount, longitudinal sections using a cryomacrotome due to the large surface area covered by the human splenic tissue (~12 cm<sup>2</sup>; Fig. 6b). H&E staining followed by imaging of human spleen sections revealed the known histological structures

Author Manuscript  
Author Manuscript  
Author Manuscript  
Author Manuscript

of spleen, including the red pulp interspersed with a patchwork of white pulp nodules and trabecular vessels (Fig. 6b). The spleen tissue covered a total of 750,640 spots, covering 77% of the array surface. Clustering of the ST data obtained from the same splenic section as the one analyzed by H&E correctly identified the red pulp, white pulp and marginal zones (Fig. 6b and Extended Data Fig. 9a–c). The expression of key marker genes for splenic immune cell types, such as *CD68*, *FCGR3A*, *CSFR1* and *ITGAX* for macrophages, *CD79A*, *CD19*, *IGHD* and *CD22* for B cells, and *CD3D*, *CD3E*, *CD69* and *SELL* for T cells, was spatially restricted to the histological regions known to contain these cell types (Fig. 6c and Extended Data Fig. 9d–f).

Next, we examined the spatial expression patterns of pairs of chemokines and their matching receptors, which are key to the maintenance of splenic tissue architecture and the organization of immune cell positioning in the spleen. For example, we examined the expression of *CXCL13*, which is a ligand for the *CXCR5* receptor and helps construct B cell follicles in the human spleen<sup>28</sup>. We found that *CXCL13* and *CXCR5* mRNAs overlapped in their spatial expression at B cell follicles (Fig. 6d,e). Next, we plotted the expression of *CCL21*, which contributes to the attraction and survival of *CCR7*-expressing, naive T cells in the T cell zone of the spleen<sup>29</sup>. Our data also revealed the colocalized expression patterns of the *CCL21* and *CCR7* mRNAs in the T cell zones that are adjacent to the larger B cell follicles in the human splenic tissue (Fig. 6d,e). Lastly, our data correctly recapitulated the expected expression patterns of the CXCL12–CXCR4 axis<sup>29</sup>. We found CXCL12 to be expressed in the red pulp and contributing to the attraction of CXCR4<sup>+</sup> B cells to the red pulp, as predicted by the computational inference of spatial signaling directionality<sup>30</sup> (Fig. 6d–f). Therefore, the spatial distributions of individual genes encoding chemokines and chemokine receptors agreed with their expected localization within the splenic tissue (Fig. 6d–f).

Overall, Array-seq allows for the spatial profiling of histological sections covering the entirety of the active surface area for mRNA capture, thereby paving the way for whole-mount analyses of tissues and organisms whose sections fit within a dozen square-centimeters.

### Comparative evaluation of ST methods

We asked where Array-seq fits within the landscape of existing methods for spatiomolecular profiling. Above, we benchmarked Array-seq against Visium because it is the only other ST tool to date that also combines all the following characteristics of our method: compatible with histology analysis (H&E) on the same tissue section, and easy to adopt and access without the need for special instrumentation, expertise or custom reagents. Going further, we compared the characteristics of Array-seq to those of other existing ST methods that also rely on mRNA capture on a slide followed by cDNA sequencing. First, the total surface of active capture area for Array-seq slides was 1 to 4 logs larger than all existing methods for sequencing-based spatial profiling (Extended Data Fig. 10a–c), with perhaps the exception of a recent implementation of the Stereo-seq method on macaque brain sections<sup>31</sup>. Second, the resolution of Array-seq (30-μm spots with 36.65-μm center-to-center distance between spots) was 8.1-fold higher spots per mm<sup>2</sup> than Visium, but lower than

Author Manuscript  
Author Manuscript  
Author Manuscript  
Author Manuscript

methods with higher resolution such as DBiT-seq<sup>8</sup> (10 µm), Slide-seq<sup>6</sup> (10 µm), Visium HD (2 µm), HDST<sup>7</sup> (2 µm), Pixel-seq<sup>11</sup> (1 µm), Open-ST<sup>32</sup> (~0.6 µm), Seq-Scope<sup>10</sup> (0.63 µm) and Stereo-seq<sup>9</sup> (0.22 µm; Extended Data Fig. 10a). However, methods with a physical resolution on the substrate below 10 µm, such as Stereo-seq, Pixel-seq, Open-ST and Seq-Scope, require spatial binning of the data for most downstream analyses and visualization due to the sparsity of count data, suggesting that the effective difference in resolution between Array-seq and these methods at the spatial data level is less pronounced than what the physical properties of spots across methods suggest. Third, existing ST methods other than Visium have rarely been adopted outside their institutions of origin due to complex requirements in instrumentation, expertise and custom reagents, making these methods difficult to implement by independent labs<sup>1</sup> (Extended Data Fig. 10a). Fourth, multiple methods are not compatible with H&E staining on the same tissue section as the one used for ST profiling due to the use of substrates other than glass for the mRNA capture slide (Extended Data Fig. 10a). Array-seq is compatible with simultaneous H&E analysis, which is, in our view, a critical prerequisite for any ST platform aiming to integrate with existing clinical workflows and research requiring dual H&E and spatiomolecular profiling on the same histological section. Fifth, we found that the sensitivity of Array-seq, as measured by the number of UMIs per µm<sup>2</sup>, was on par with Visium but lower than other ST methods (Extended Data Fig. 10d and Supplementary Table 1), suggesting that further optimizations of Array-seq are needed. Lastly, the cost of Array-seq is lower than all other methods reported to date and, for example, by a decrease of nearly 50-fold when comparing to Visium (Extended Data Fig. 10e and Supplementary Table 2). We estimated that the cost of sequencing all ~1 M spots on one Array-seq slide at a depth of 5,000–10,000 raw sequencing reads per spot (that is, ~5–10 B reads total) would cost US\$3,550 to US\$7,100 using present day next-generation sequencers (that is, ~US\$314 to US\$628 per cm<sup>2</sup> using a NovaSeq X Series 10B Reagent Kit with 100 cycles).

Array-seq thus lies at a unique position when compared to existing ST platforms thanks to its large surface area, ease of adoption by any laboratory and low cost while maintaining a sensitivity similar to that of current commercially available gold standards.

## Discussion

By combining the two workhorses in gene expression from the beginning of the twenty-first century—microarrays and next-generation sequencing—we created a simple, large-format platform for ST. Our method pushes the boundaries of spatial profiling by enabling the processing of small-to-large samples ranging from murine tissues to whole-mount human organs, as shown in this work, or whole-mouse sections, as we recently demonstrated<sup>33</sup>. We also showed that Array-seq enables the 3D profiling of tissue samples by processing large numbers of serial sections from the same tissue block. Notably, Array-seq is (i) scalable thanks to its large format and low cost, (ii) easy to adopt without special expertise or instrumentation, (iii) compatible with H&E staining—the gold standard for clinical pathology diagnosis—and (iv) readily applicable to all fields of basic and clinical research. Thus, the Array-seq platform is poised to contribute to the spatial omics revolution by enabling spatiomolecular studies at scale and facilitating technology adoption across research communities.

The total surface of active area for mRNA capture of on Array-seq slide is 11.31 cm<sup>2</sup>, which is 26.7-fold larger than the only widespread method for ST profiling, Visium. Therefore, Array-seq slides provided a high degree of flexibility in the size and number of samples that can be profiled per slide, while keeping costs low thanks to cheap microarray manufacturing (nearly 50-fold less than Visium). We estimated that, at full occupancy, a single Array-seq slide could capture up to ~2 to 20 million cells depending on the tissue type and the number and size of the tissue sections being analyzed. Our method, therefore, provides high-throughput profiling capabilities, which are needed to address questions requiring large numbers of samples such as time series, functional perturbations, drug responses, animal or patient cohorts or 3D analyses of tissue structures. For example, Array-seq opens the door to functional spatiomolecular studies whereby molecular-, cellular- or tissue-level perturbations can be coupled with downstream ST profiling. Our method also enables cohort-level studies of patient samples by making the profiling of hundreds of specimens on dozens of Array-seq slides feasible and affordable, which will yield information with high clinical value.

In conclusion, Array-seq is readily deployable to address fundamental questions across all fields of basic and translational biomedical research. We foresee that the open-source nature of Array-seq, which relies on cheap, well-established microarray technology, will help democratize and enhance the reach of spatiomolecular profiling across fields of inquiry. We anticipate that further optimizations of our on-slide assembly process to generate mRNA capture probes will increase the sensitivity of our method. The main limitations of Array-seq are its resolution (30-μm spot diameter) and lack of compatibility with multimodal readouts and formalin-fixed, paraffin-embedded tissue sections. Future developments of the Array-seq platform will enable the multimodal analysis of whole or targeted transcriptomes and proteomes in both fresh-frozen and formalin-fixed, paraffin-embedded samples, as shown for other spatiomolecular profiling tools<sup>8,34,35</sup>. Moreover, we anticipate that future modifications of the technology will help increase its resolution by, for example, modifying microarray printers to decrease spot diameters or combining sample expansion protocols with current Array-seq slides<sup>36</sup>. The throughput of the method can also be readily increased using automation procedures facilitated by the format of Array-seq slides, which is that of standard microscopy slides. Array-seq will thus enable the spatiomolecular mapping of mammalian and non-mammalian organisms and organs at scale—without requiring the preselection of small regions that may bias the analysis of tissue architecture.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-024-02501-5>.

## Methods

### Mice

Female C57BL/6J mice (6–8 weeks old) were obtained from the Jackson Laboratories (stock 000664). Animals were housed in specific pathogen-free conditions at The University

of Chicago, and all experiments were performed in accordance with the US National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by The University of Chicago Institutional Animal Care and Use Committee.

### Access to human tissue and ethics oversight

The human spleen sample was obtained from a 69-year-old female with metastatic high-grade serous carcinoma of Mullerian origin. The surgical intervention involved a splenectomy. The splenic parenchyma was negative for carcinoma. The donor sample was obtained at the University of Chicago Medical Center (UCMC) with informed consent from the donor family without compensation. Deidentified tissue sample collection and subsequent experiments were approved by the University of Chicago Biological Sciences Division/UCMC Institutional Review Board (IRB no. 21–0707). The human spleen sample was collected within 1 h of resection, placed in a plastic bag, frozen in a dry-ice hexane bath for 3 min and stored at –80 °C until use.

### Custom oligonucleotide microarray design

We generated two custom array designs using the microarray platform SureSelect DNA Capture Array (1 M; Agilent, G3358A): one containing spatial barcodes of 18-mers (Agilent identifier AMADID 086624) and another with 12-mer spatial barcodes (Agilent identifier AMADID 087232). Barcodes were designed using the DNABarcodes R package<sup>37</sup>. For the 18-mer barcodes used for the initial gel-based optimizations of on-slide probe assembly, we first generated 9-mer barcodes with a minimum Hamming distance of 3, no self-complementary sequences, no consecutive base repetition 3 and with a 50% GC content. Resulting barcodes were duplicated and concatenated in all possible two-way combinations, creating a pool of over 6.5 million, 18-mer barcodes which maintained the minimum Hamming distance of 3 as in the original set. Barcode sequences were filtered to maximize sequencing performance using the following criteria: (1) does not begin with AC or CC, (2) does not end with GG, and (3) no consecutive base repetition 3. After filtering, we obtained 5,069,490 unique barcode sequences from which a random set of 974,016 barcodes was selected. For the 12-mer barcodes used for all Array-seq ST profiling experiments, we followed a similar procedure as described above for the 18-mer barcodes without the concatenation step. We first generated an initial pool of all possible 12-mer barcodes with a GC content between 40% and 60%, yielding 10,272,768 independent barcodes. Barcode sequences were then filtered to maximize sequencing performance using the following criteria: (1) no individual base present in more than half of the entire barcode sequence, (2) no base repeated consecutively four times or greater, (3) no base repeated three times with additional occurrences of the same base two bases away from the repeat in either direction, (4) no four-base stretches of As and/or Ts or Gs and/or Cs in the first four, middle four or last four bases of the 12-mer barcode. After filtering, we obtained 6,814,344 unique barcode sequences, which were further filtered to only include a pool of barcodes with a minimum Hamming distance of 2 and no self-complementarity, yielding a pool of 1,048,608 barcode sequences from which a random set of 974,016 barcodes not starting with AC or CC was selected.

## On-slide, mRNA capture probe assembly

We note that our gap-fill reaction yields three oligonucleotide species (see lane c of the gel in Fig. 1c): (i) a fully assembled capture probe (107-mer), (ii) an anchor 2-UMI-Oligo(dT) oligonucleotide and (iii) an overextended anchor 1 oligonucleotide (using the custom array design with 18-mer spatial barcodes). The inclusion of the last wash step at 58 °C for 10 min in the procedure detailed above leads to the removal of the unwanted anchor 2-UMI-Oligo(dT) oligonucleotide (see lane d of the gel in Fig. 1c). The wash step does not fully remove the overextended anchor 1 species, which is not a problem because (i) these oligonucleotides do not interfere with mRNA capture, cDNA synthesis or library construction, and (ii) they are likely to be progressively removed during wash steps including full-length cDNA cleanup with zymo columns and post-PCR cleanups with SPRI beads. We estimated that the remaining overextended anchor 1 oligonucleotides likely decrease the theoretical, maximum mRNA binding capacity of Array-seq slides by ~25%.

## PAGE analysis

On-slide probe assembly was carried out as described in the previous section using the 18-mer custom arrays (SureSelect DNA Capture Array 1 M, Agilent, G3358A, Agilent identifier AMADID = 086624) and hybridized with anchor 1 and modified anchor 2 with a 17-bp UMI sequence (5'-/5Phos/

GTAAAACGACGCCAGNNNNNNNNNNNNNNNTTTTTTTTTTTTTTTT to facilitate the identification of the various oligonucleotide species resulting from our on-slide, gap-fill reaction. Following ST probe assembly, oligonucleotide products were eluted from the microarray with 0.1 M KOH (Sigma, P4494–50ML). Oligonucleotide eluates were neutralized by adding 0.2× volume of 1 M Tris-HCl (pH = 7.4) and purified using Oligo Clean & Concentrator columns (Zymo, D4060). Purified oligonucleotide samples were mixed with an equal volume of Novex TBE-Urea Sample Buffer (2×; Thermo Fisher Scientific, LC6876) and heated at 70 °C for 3 min. Samples and the 20/100 DNA Ladder (IDT, 51–05–15–02) were loaded on a Novex 15% TBE-Urea Gel (Thermo Fisher Scientific, EC68855BOX) and ran at 200 V for 75 min using the XCell SureLock Mini-Cell gel electrophoresis system (Thermo Fisher Scientific, EI0001) in 1× Novex TBE Running Buffer (Thermo Fisher Scientific, LC6675). Gels were stained with 1× SYBR Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific, S11494) in 1× TBE Running Buffer for 10 min and imaged on the ChemiDoc MP Imaging System (Bio-Rad). Gel band intensities were quantified using ImageJ (<https://imagej.net/ij/index.html>).

### Ligation bias analysis

To quantify potential ligation biases during the extension–ligation reaction on the array for mRNA capture probe assembly, we calculated the mean and standard deviation of total UMI counts per spot across all the spots whose spatial barcodes ended with the same last single or two bases. For these analyses, we used the Array-seq data from tissue section (replicate 1) of the MOB system of the mouse brain shown in Fig. 2.

### Nuclei per spot analysis

Array-seq slides were hybridized with a fluorescently labeled anchor 1 (5'-5Cy5/CCCTACACGACGCTTCGATCT-3') as described above. After hybridization, arrays were washed with nuclease-free water and dried at room temperature. Tissue sections were placed on the prepared slides and were fixed for 30 min at –20 °C in pre-cooled 100% methanol (Sigma-Aldrich, 32213). The sections were then stained for 10 min at room temperature with 1 µg ml<sup>-1</sup> DAPI (Fisher Scientific, NC0804630), washed with nuclease-free water, dried at room temperature, and imaged at ×20 magnification using an Olympus VS2000 slide scanner. The fluorescent images capturing spots (Cy5) and nuclei (DAPI) were analyzed using Qupath-0.2.3 (<https://qupath.github.io/>). Nuclei and spots were detected using the Cell Detection tool, and the centroids for each nucleus and spot object were exported for downstream processing in R. Nuclei were assigned to spots based upon minimum Euclidean distance between nucleus and spots centroids. The distributions of the numbers of nuclei per spots for each tissue type were plotted using ggplot (<https://ggplot2.tidyverse.org/>).

### Histological sectioning

Mice were euthanized with CO<sub>2</sub> and organs were harvested, fresh-frozen in OCT (Tissue-Tek, 4583) on powdered dry ice and stored at –80 °C until further use. Frozen tissues were sectioned (10-µm thickness) using a cryomicrotome (Leica, CM3050S) and placed on Array-seq slides pre-chilled in the cryostat chamber before downstream histology analysis and ST profiling. For serial sectioning used to perform 3D ST analyses, consecutive mouse

kidney sections were placed onto Array-seq slides while skipping ~100 µm between each section that was profiled. In mouse sections intended for comparison between Array-seq and Visium, immediately adjacent kidney sections were placed onto each ST slide. The human spleen sample was taken from -80 °C storage, immediately placed in a freezing frame containing SCEM (SECTION-LAB) and frozen in a dry-ice hexane bath. The frozen human spleen was sectioned using a cryomacrotome (Leica, CM3600XP).

### Tissue section fixation, H&E staining and imaging for Array-seq

Tissue sections on Array-seq slides were thawed at 37 °C for 1 min and fixed for 30 min at -20 °C in pre-cooled 100% methanol (Sigma-Aldrich, 32213). Following fixation, slides were dried for 1 min at room temperature and incubated with 2 U µl<sup>-1</sup> Ribolock (Thermo Fisher Scientific, EO0382) in 5× saline-sodium citrate (SSC; Sigma, S6639-1L) for 5 min at room temperature. Hematoxylin (Sigma-Aldrich, MHS16-500ML) and eosin (Sigma-Aldrich, HT110216-500ML) were filtered through a 0.22-µm sterile filter (Foxx Life Sciences, 371-2215-OEM) before use. After the SSC buffer was removed from the slide surface, sections were stained for 3 min at room temperature with hematoxylin, washed in nuclease-free water, and stained for 1 min at room temperature with eosin diluted at a 1:10 ratio in 0.45 M Tris-Acetic acid buffer, pH 6.0 (Sigma-Aldrich, GE17-1321-01 and A6283). The slide was washed in nuclease-free water and kept at room temperature until dry and imaged at ×20 magnification using an Olympus VS2000 slide scanner.

### Tissue section permeabilization and in situ reverse transcription for Array-seq

Array-seq slides with H&E-stained sections were attached to a gasket with one-well or eight-well chambers (Electron Microscopy Sciences, 63484-31 or 63484-25). Tissue sections were permeabilized with 0.1% pepsin (Sigma, P7000-25G) in 0.1 M HCl (ColeParmer, EW-88011-96) at 37 °C. Permeabilization times were optimized by low-depth sequencing of Array-seq libraries using MiSeq (Illumina). Human spleen and mouse MOB were permeabilized for 20 min, and all other tissues were permeabilized for 15 min. After permeabilization, sections were washed once with 0.1× SSC buffer. After removing the wash buffer, reverse transcription mix consisting of 1× Maxima Reverse Transcriptase Buffer (Thermo Fisher Scientific, EP0753), 0.2 mg ml<sup>-1</sup> BSA (Sigma, B8667-1.25ML), 0.5 mM dNTPs (NEB, N0447L), 3 µM of template switching oligonucleotide (5'-5MeisodC//iisodG//iMe-isodC/CCCTACAC GACGCTCTTCCrGrGrG-3'), 2 U µl<sup>-1</sup> Ribolock and 10 U µl<sup>-1</sup> Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific, EP0753) was added to each well and incubated overnight at 42 °C. The array was washed once with 0.1× SSC and incubated for 1 h at 37 °C using a tissue removal buffer containing Proteinase K (Qiagen, 19131) at a 1:8 dilution in buffer PKD (Qiagen, 1034963). The array was washed once with 0.1× SSC and incubated for 1 h at 37 °C with an Exo1 mix containing 1× Exonuclease reaction buffer and 1 U µl<sup>-1</sup> Exonuclease 1 (NEB, M0293L). The array was washed with 0.1× SSC followed by buffer EB, and full-length cDNAs were eluted from the array by incubating for 10 min at room temperature in 0.1 M KOH (Sigma, P4494-50ML). Eluted cDNAs were transferred into a 1.5-ml or 15-ml tube (depending on size and number of tissues/chambers) and neutralized by adding a 0.2× volume of 1 M Tris-HCl (pH = 7.4).

### Array-seq library preparation and sequencing

We modified procedures developed by others for single-cell RNA-seq<sup>38</sup> as detailed below. cDNAs were purified using a DNA Clean & Concentrator-5 kit (Zymo Research, D4013) and amplified using single-primer PCR with the following primer: 5'-5Biosg/CCCTACACGACGCTCTTCC-3' using the KAPA HiFi HotStart ReadyMix (Roche, KK2601) in a 50-μl reaction volume and with the following cycling conditions: one cycle at 95 °C for 3 min; eight cycles at 98 °C for 20 s, 60 °C for 20 s, 72 °C for 3 min; and one cycle at 72 °C for 5 min. Amplified cDNAs were cleaned up using ×0.6 volume of magnetic SPRISelect beads (Beckman Coulter, A63880) and quantified using the Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Q32851) and a TapeStation D5000 high-sensitivity DNA Screenshot kit (Agilent). In total, 100 ng of amplified cDNA was fragmented using Tament DNA TDE1 Enzyme and buffer (Illumina, 20034197) for 5 min at 55 °C and amplified by PCR using the following forward primer: 5'-AATGATACGGCGACCACCGAGATCTACACTTTCCCTACACGACGCTCTCCG\*A\*T\*C\*T-3' and Illumina i7 reverse primers (Illumina, FC-121-1011) with the following cycling conditions: one cycle at 98 °C for 30 s; eight cycles at 98 °C for 10 s, 63 °C for 30 s and 65 °C for 30 s; and one cycle at 65 °C for 5 min. Libraries were cleaned up using 0.8× volume of magnetic SPRISelect beads followed by gel purification using 2% E-Gel EX Agarose Gels (Thermo Fisher Scientific, G402002) and MinElute columns (Qiagen, 28604) and quantified with the Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Q32851). Resulting libraries were sequenced on the NextSeq 550 or 2000 platforms (Illumina) with 38 bases for read 1, 8 bases for i7 barcodes, and 42 bases for read 2.

### Visium ST

Following sectioning, mouse kidney sections were mounted onto a Visium Spatial Gene Expression library preparation slide (10× Genomics) and samples were processed to generate ST sequencing libraries according to the manufacturer's instructions. Resulting libraries were sequenced on the NextSeq 2000 platform (Illumina), with 28 bases for read 1, 10 bases for the sample index and 50 for read 2.

### Tissue harvest for bulk RNA-seq

Mouse tissues (liver, kidney, brain) were harvested, frozen and stored as previously described<sup>39,40</sup>. Mice were anesthetized with 2,2,2-tribromoethanol (250–500 mg per kg body weight) and perfused transcardially with PBS containing 10 mM EDTA (to avoid signal contamination from blood in tissues). Immediately after perfusion, tissues were placed in RNA preserving solution (5.3 M ammonium sulfate, 25 mM sodium citrate, 20 mM EDTA) and kept at 4 °C overnight before transfer to –80 °C for storage.

### Whole-tissue RNA extraction

Whole-tissue RNA extraction was performed using procedures previously developed by us<sup>39</sup>. In brief, tissues stored in RNA-preserving solution were thawed and transferred to 2 ml tubes containing 700–1,500 μl (depending on tissue) of PureZOL (Bio-Rad, 7326890) or homemade TRIzol-like solution (38% phenol, 0.8 M guanidine thiocyanate, 0.4 M ammonium thiocyanate, 0.1 M sodium acetate, 5% glycerol). Tissues were lysed by adding

2.8-mm ceramic beads (OMNI International, 19–646) and running 1–3 cycles of 5–45 s at 3,500 rpm on the PowerLyzer 24 (Qiagen). For liver, and brain samples, tissues were lysed with 35 ml using M tubes (Miltenyi Biotec, 130–096–335) and running 1–4 cycles of the RNA\_02.01 program on the gentleMACS Octo Dissociator (Miltenyi Biotec). Next, lysates were processed in deep 96-well plates (USA Scientific, 1896–2000) by adding chloroform for phase separation by centrifugation, followed by precipitation of total RNA in the aqueous phase using magnetic beads coated with silane (Dynabeads MyOne Silane; Thermo Fisher Scientific, 37002D), buffer RLT (Qiagen 79216) and ethanol. Genomic DNA contamination was removed by on-bead DNase I (Thermo Fisher Scientific, AM2239) treatment at 37 °C for 20 min. After washing steps with 80% ethanol, RNA was eluted from beads. This RNA extraction protocol was performed on the Bravo Automated Liquid Handling Platform (Agilent). Sample concentrations were measured using a Nanodrop One (Thermo Scientific). RNA quality was confirmed using a Tapestation 4200 (Agilent Technologies).

### Whole-tissue RNA-seq

For each tissue sample, full-length cDNA was synthesized in 20 µl final reaction volume containing the following: (1) 10 µl of 10 ng µl<sup>-1</sup> RNA; (2) 1 µl containing 2 pmol of a custom RT primer biotinylated in 5' and containing sequences from 5' to 3' for the Illumina read 1 primer, a 6-bp sample barcode (up to 384), a 10-bp UMI and an anchored oligo(dT)<sub>30</sub> for priming; and (3) 9 µl of RT mix containing 4 µl of 5× RT buffer, 1 µl of 10 mM dNTPs, 2 pmol of template switching oligonucleotide and 0.25 µl of Maxima H Minus Reverse Transcriptase (Thermo Scientific, EP0753). First, barcoded RT primers were added to RNA, which were then denatured at 72 °C for 1 min and snap cooled on ice. Second, the RT mix was added, and plates were incubated at 42 °C for 120 min. For each library, double-stranded cDNA from up to 384 samples was pooled using DNA Clean & Concentrator-5 columns (Zymo Research, D4013), and residual RT primers were removed using exonuclease I (New England Biolabs, M0293). Full-length cDNAs were amplified with five to eight cycles of single-primer PCR using the Advantage 2 PCR Kit (Clontech, 639206) and cleaned up using SPRIselect magnetic beads (Beckman Coulter, B23318). cDNA was quantified with a Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, 32851) and 50 ng of cDNA per pool of samples was fragmented using the Tegment DNA Enzyme I (Illumina, 20034197) and amplified using the NEBNext Ultra II Q5 Master Mix (NEB, M0544L). Libraries were gel purified using 2% E-Gel EX Agarose Gels (Thermo Fisher Scientific, G402002), quantified with a Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Q32851) and a Tapestation 4200 (Agilent Technologies), and sequenced on the NextSeq 550 platform (Illumina).

### Sequence alignment and generation of spatial gene expression matrix

Fastq files from Array-seq and Visium sequencing experiments were generated using BaseSpace DRAGEN Analysis v1.3.0 (Illumina). For both platforms, the spatial barcodes and UMIs are contained in the read 1 sequence (Array-seq spatial barcode: 1–12 bases, Array-seq UMI: 29–38 bases; Visium spatial barcode: 1–16, Visium UMI: 17–28) while the read 2 sequence contains cDNA information. To generate spatial count matrices, fastq files were processed using STARsolo from the STAR package version 2.7.10a (<https://>

[github.com/alexdobin/STAR/blob/master/docs/STARsolo.md/](https://github.com/alexdobin/STAR/blob/master/docs/STARsolo.md/)<sup>41</sup>. Reads were aligned to the mm10 mouse or CRCh38 human reference genome, spatial barcodes were demultiplexed against the whitelist of the reverse complement sequences of the 974,016 Array-seq spatial barcodes or 4,992 Visium barcodes (using the Exact option), and UMIs were collapsed (using the 1MM\_CR option for error correction).

### Alignment of spatial data with H&E image

To filter the Array-seq data to only include spots under the tissue, we followed similar procedures described by others<sup>8</sup> with modifications described below. In brief, the STARsolo output count matrices were read into Python 3.8.5 (<http://www.python.org/>) with Scanpy<sup>42</sup> and stored in an AnnData object together with the spatial coordinates of each barcoded spot on the slide. The spot coordinates and UMI counts were used to generate a scalable vector graphic image composed of points matching the *x* and *y* positions of each barcoded spot, which were colored based on UMI counts. The scalable vector graphic image was converted to a png format using the Wand package (<https://github.com/emcconville/wand/>) and manually aligned with the H&E image of the same section using Illustrator (Adobe). The *x* and *y* positions, height and width of the aligned images were recorded. The ST AnnData object was imported and scaled to match the corresponding pixel coordinates of the H&E image. For H&E images containing multiple sections, each detected tissue sample was labeled using the scikit-image package<sup>43</sup>. The annotation of organs and replicates were mapped to the object labels, the individual section H&E images were cropped, and the spot spatial data were rescaled. Lastly, to select the spots under tissue, we kept spots overlapping with pixel coordinates containing detected tissue objects. Visium datasets were filtered for spots under tissue and aligned to H&E images by joining the tissue positions list csv, generated by the Space Ranger v1.2.0 software (10x Genomics), to the AnnData object generated from the STARsolo count matrix.

### Clustering and differential gene expression analysis

For spatial gene expression analysis using Scanpy, resulting count matrices were further filtered to keep (1) genes with at least 20 UMIs across all spots, and (2) spots with more than 120 UMIs in total. For clustering, data were normalized,  $\log_{10}$  transformed and clustered using principal component analysis, *k*-nearest neighbor identification and the Leiden clustering algorithm<sup>19</sup>. Our mouse MOB data were also clustered using the algorithms GraphST (v1.0.0)<sup>20</sup> and BayesSpace (v1.5.1)<sup>21</sup> in R-4.1.0 (<http://www.R-project.org/>). Differentially expressed genes between clusters were identified using the tl.rank\_genes\_groups() function in Scanpy. To map tissue subregions to spatial cluster obtained with the Leiden algorithm, we used the top 15 differentially expressed genes from each cluster to inform manual annotation. For plotting, spatial count or clustering data were overlaid on the grayscale H&E image using Seaborn (<https://seaborn.pydata.org/>). For plotting MOB marker spatial distribution along the *x* axis, we used the SciPy's gaussian\_kde() function to compute the smoothed spatial distribution of each gene.

### Spatial assignment of cell types

To deconvolute the cell types present at each spot, we used the following packages: CARD (v1.0.0)<sup>23</sup>, SpaceX/RCTD (v2.2.1)<sup>25</sup>, and cell2location (v0.1.3)<sup>26</sup>. As a reference, we used

publicly available single-cell RNA-seq data for the mouse MOB<sup>24</sup> and kidney tissues<sup>27</sup>. For the MOB reference data, we excluded some clusters of granule cells, transition neurons, and olfactory nerve cells, which confounded the integration of results obtained across the three algorithms listed above. For kidney datasets, cell-type proportions for each spot were calculated using the CARD deconvolution function with the ST (Array-seq or Visium) and single-cell RNA-seq count matrices as input, and results were plotted using the Seaborn package.

### mRNA diffusion analysis

We quantified the numbers of UMIs detected under the tissue section versus those detected outside the tissue section by following a similar process as others<sup>44</sup>. We manually selected a field of view surrounding MOB Array-seq replicates and labeled spots as under or adjacent to the tissue section without filtering based on minimal reads per spot. To estimate diffusion in Visium, we performed a similar analysis using publicly available mouse MOB data (<https://www.10xgenomics.com/datasets/adult-mouse-olfactory-bulb-1-standard-1/>).

### Benchmarking Array-seq and Visium sequencing statistics and analysis

For a direct comparison of Array-seq and Visium using four pairs of immediately adjacent kidney sections, we normalized sequencing depth so each tissue would have equivalent reads per mm<sup>2</sup> of tissue over the capture area. To this end, Array-seq and Visium kidney tissue section scans were manually outlined and the tissue area was quantified using QuPath v0.2.3. The reads for each of the fastq files were then proportionally downsampled with Subread v2.0.6 (<https://subread.sourceforge.net/>) before transcript mapping and counting. Sequencing saturation and detected genes were measured with STARSolo while UMI and genes per spot were calculated from the Array-seq and Visium AnnData objects.

### 3D mouse kidney Array-seq data analysis

Serial sections were aligned to the H&E scan and separated as described above. 3D alignment was performed by pairwise Jaccard index maximization of H&E tissue overlap during combinatorial adjustments of image rotations and vertical and horizontal translations. Optimal transformation values were applied to corresponding serial section ST coordinate data. After alignment, all raw count matrices were merged, normalized, clustered and analyzed for differential expression as described above.

### Spatial Gene Ontology enrichment

Gene Ontology Biological Process 2024 gene sets were obtained from GSEAp<sup>45</sup>. Enrichment of each gene set was calculated using Scanpy's tl.score\_genes() function. Gene sets with a maximum spot score greater than 1 were plotted for visualization of nonrandom enrichment score spatial distribution. For manually selected gene sets, the spot enrichment scores were added to AnnData metadata, and average scores for tissue subregion were calculated and normalized before heat map generation.

## Chemokine ligand–receptor analysis

For visualizing spatial ligand–receptor dynamics in the human spleen, we used the COMMOT package on a selected subset of the tissue<sup>29</sup>. Selected chemokine ligand–receptor interactions were inferred using Commot’s `tl.spatial_communication()` function using a distance threshold of 150. Ligand–receptor pair communication directionality was calculated with the `tl.communication_direction()` function using  $k=5$ . Vectors for interaction magnitude and directionality were plotted and overlayed onto the H&E image.

## Sensitivity analysis across existing ST methods

Array-seq MOB sequencing data sensitivity was benchmarked against publicly available ST data from MOB tissue sections, or other tissues when MOB was not found (Supplementary Table 1). To normalize the mean UMI counts per unit area, we divided the mean counts in the dataset by the number of spots per  $\mu\text{m}^2$  (based on the stated values of total spots and capture area size in the articles) or by the number of  $\mu\text{m}^2$  in the data bin size (Seq-Scope).

## Cost comparison analysis

Values for method cost were those stated in the articles of origin for each method. Array-seq costs were calculated based on the costs of individual reagents and supplies needed to implement our method (Supplementary Table 2). Visium cost was based on the list price for a four-capture area, single slide kit from a quote obtained on 7 August 2023 (Dual index and accessories not included). Curio Bioscience (Seeker) cost was based on a cost estimate provided by a sales representative on 28 February 2024. STomics cost ( $2 \times 3 \text{ cm}$  format) cost was based on a cost table provided by a sales representative on 27 March 2024.

## Whole-tissue RNA-seq data analysis

Sequencing read files were processed to generate UMI count matrices using the Python toolkit from the bcbio-nextgen project version 1.1.5 (<https://bcbio-nextgen.readthedocs.io/en/latest/>). In brief, reads were aligned to the mouse mm10 transcriptome with RapMap (<https://github.com/COMBINE-lab/RapMap/>). Quality-control metrics were compiled with a combination of FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), Qualimap and MultiQC (<https://github.com/ewels/MultiQC/>). Samples were demultiplexed using barcodes stored in read 1 (first six bases) and raw UMI count matrices were computed using UMIs stored in read 1 (bases 7 to 16; <https://github.com/vals/umis/>). Differential expression analysis was done using custom scripts in R (<http://www.R-project.org/>). Raw count matrices were filtered to keep genes with at least 20 counts per million or five UMIs in two samples and normalized across samples using the `calcNormFactor` function in edgeR (<https://github.com/StoreyLab/edge/>).

## Statistics and reproducibility

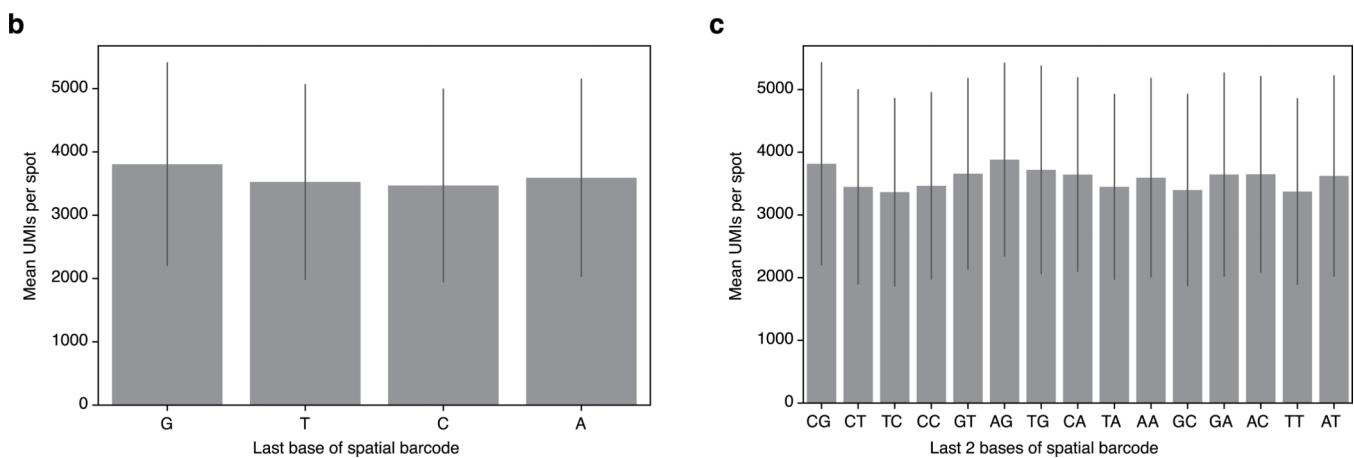
Statistical analyses were performed by Python, using the Wilcoxon rank sum test implemented by Scanpy<sup>42</sup>, or Pearson correlation coefficient when measuring correlation. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications<sup>36</sup>. No treatment groups were part of this

study, and all mice were the same sex and age. Some tissue sections were excluded from analysis if sectioning artifacts were observed. All spatial profiles of gene expression, cluster annotation and cell-type occupancy shown in this study are representative of at least two independent experiments except for the whole-mount human spleen section, which was only performed once due to limited tissue sample availability.

### Reporting summary

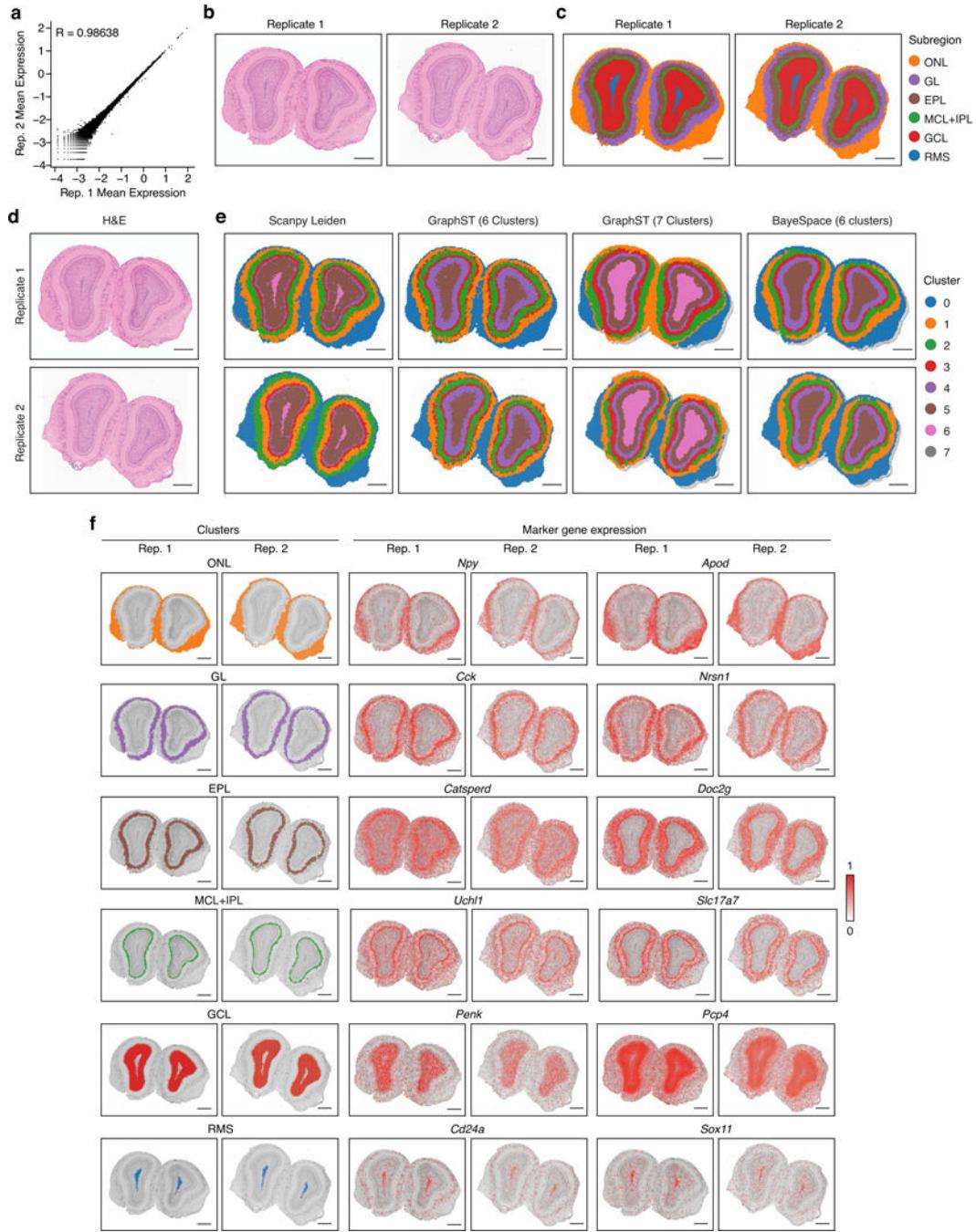
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Extended Data



**Extended Data Fig. 1 | Sequence level overview of the on-slide assembly procedure yielding mRNA capture probes.**

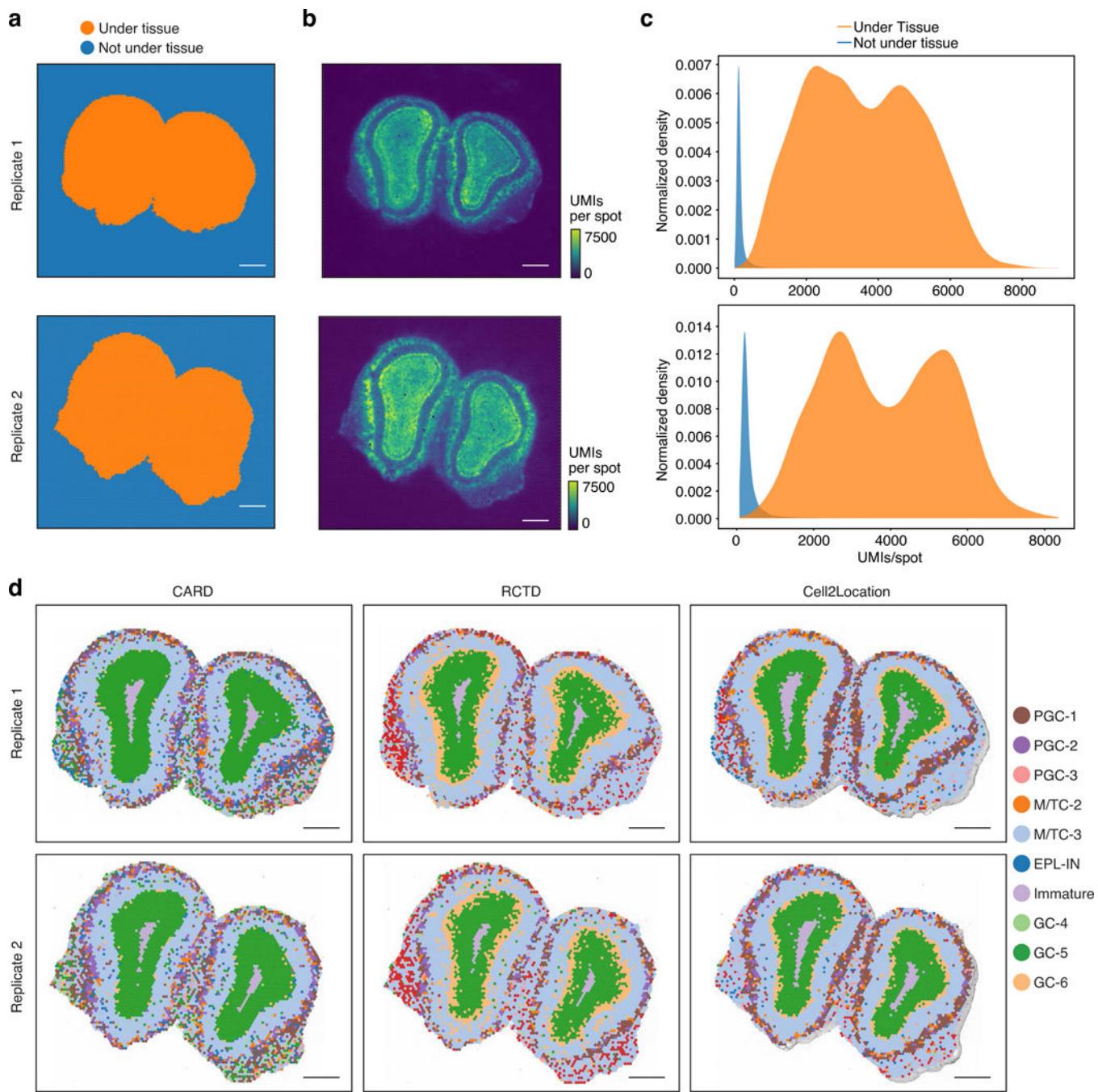
**a**, Custom sequence oligonucleotide microarrays containing two common sequences (Anchors 1 and 2) and a spatial barcode sequence unique to each spot on the array are hybridized with indicated oligonucleotides (Step 1). On-slide extension-ligation using a DNA polymerase and a DNA ligase yields a fully assembled mRNA capture probe containing a sequencing adaptor, a spatial barcode, UMIs, and an oligo(dT) sequence (Step 2). *In situ* reverse transcription in tissue sections placed on an Array-seq slide generates full-length cDNAs with a template switching oligo (TSO) sequence in 3' (Step 3). Full length cDNAs are eluted from the slides and processed for RNA sequencing library construction (Step 4). **b, c**, Ligation bias analysis after extension-ligation of mRNA capture probes. Bar plots showing the number of UMIs captured by spot across an Array-seq slide after grouping and normalizing by the last single (**b**) or two (**c**) bases of the spatial barcode sequence. Data are presented as mean values  $\pm$  SD (n = 3490 spots for A, 3756 for C, 2767 for G, and 3771 for T in **b**; and n = 647 for AA, 1000 for AC, 1000 for AG, 749 for AT, 1057 for CA, 789 for CC, 800 for CG, 1162 for CT, 997 for GA, 823 for GC, 1045 for GT, 789 for TA, 1144 for TC, 967 for TG, 770 for TT in **c**).



**Extended Data Fig. 2 | Reproducibility of Array-seq data from mouse main olfactory bulb sections.**

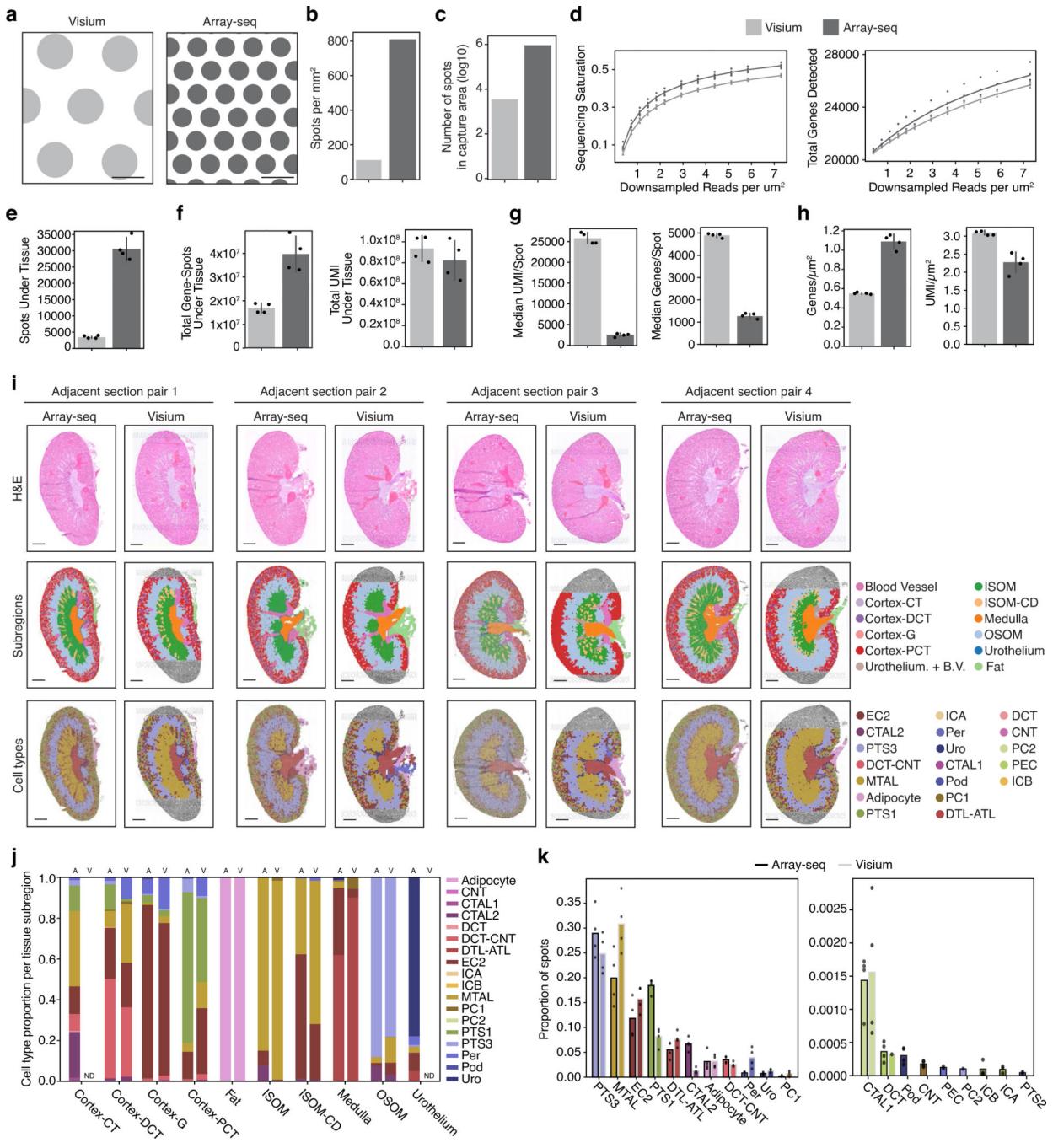
**a**, Correlation (Pearson's coefficient) between Array-seq replicates 1 and 2 from two sections obtained from two independent mouse main olfactory bulb (MOB) tissues. Shown are the normalized  $\log_{10}$  unique molecular identifier (UMI) counts which were averaged across all spots for each gene across both replicates. **b, d**, H&E images of independent MOB tissue sections which were placed onto separate gasket chambers of the same Array-seq slide. **c, e**, Unsupervised clustering highlights the histological layers of the MOB tissue

using the Leiden algorithm (**c**, showing clusters annotated by tissue subregions) or indicated algorithms for comparative analyses (**e**, showing raw clustering results across methods). ONL, Olfactory Nerve Layer; GL, Glomerular Layer; EPL, External Plexiform Layer; MCL, Mitral Cell Layer; IPL, Internal Plexiform Layer; GCL, Granule Cell Layer; RMS, Rostral Migratory Stream. **f**, Spatial plots of indicated MOB tissue subregions (left panels) and scaled  $\log_{10}$  expression of subregion-specific marker genes (right panels) overlaid on grayscale H&E images. Scale bars, 500  $\mu$ m.



**Extended Data Fig. 3 | mRNA diffusion and spatial cell type assignment analyses.**

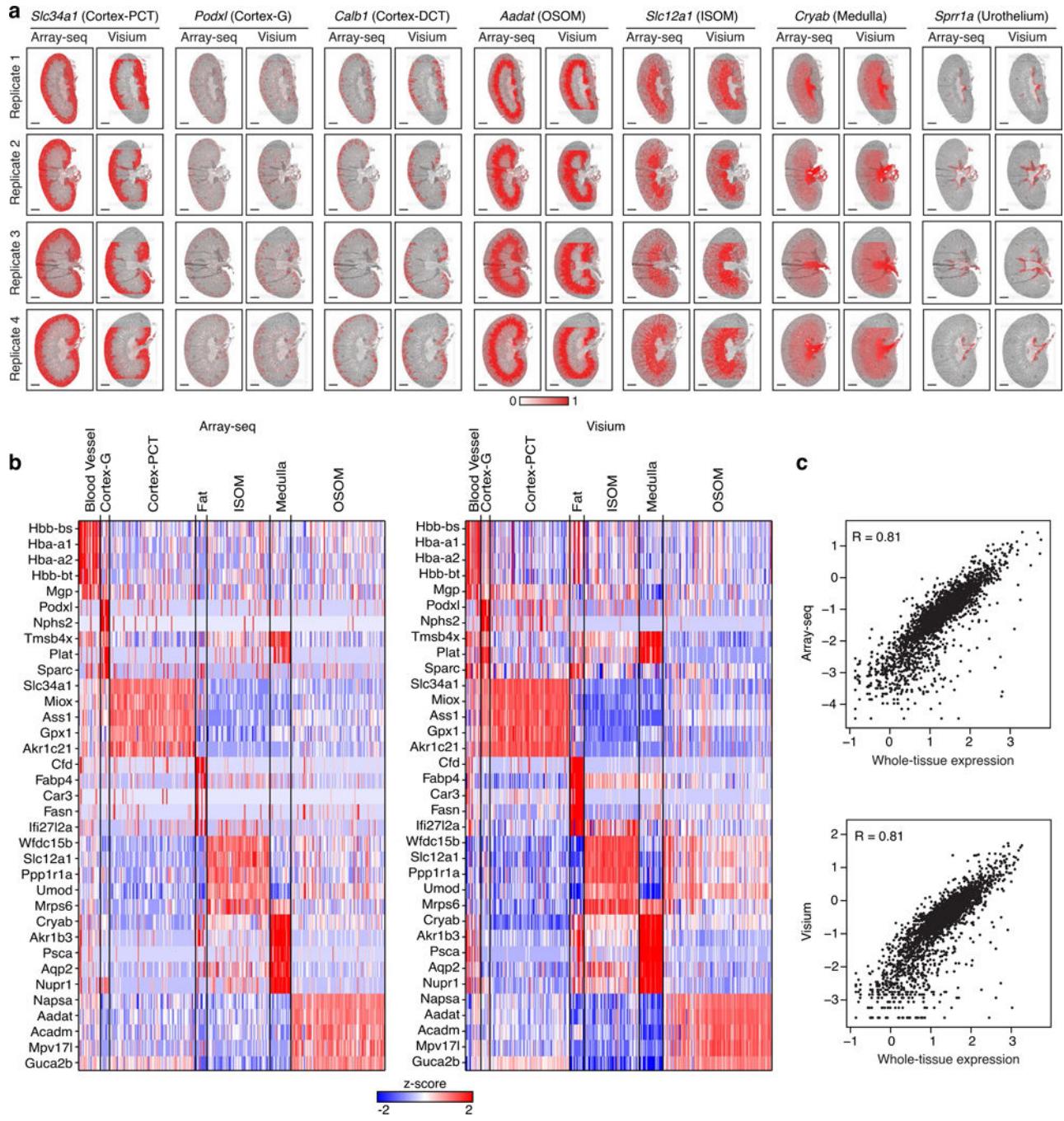
**a**, Images highlighting the spots of the Array-seq slide that were under (orange) or outside (blue) of the two MOB sections which were profiled. **b**, Spatial plots of total UMIs detected per spot across both MOB Array-seq data sets. **c**, Density plots (smoothed by kernel density estimation) of the distributions of UMIs per spot detected under (orange) and outside (blue) of the tissue sections. **d**, Cell type assignments across spots on the Array-seq slide for the indicated MOB cell types (colors) using indicated algorithms (top). Cell types with the highest percentage in inferred proportion per spot were assigned to each spot. EPL-IN, external plexiform layer interneuron; GC, granule cell; M/TC, mitral and tufted cell; PGC, periglomerular cell. Scale bars, 500  $\mu$ m.



**Extended Data Fig. 4 | Extended comparison of Array-seq and Visium mouse kidney spatial transcriptomics data.**

**a**, Representative images of virtually rendered Visium and Array-seq spot coverage (18.8% for Visium vs 60.1% for Array-seq). Scale bars, 50  $\mu$ m. **b, c**, Bar plots of the number of spots per mm<sup>2</sup> (**b**) and total active area (**c**) on indicated platforms. **d**, Downsampling analysis showing changes in sequencing saturation (left) and total genes detected across the entire section (right) using kidney section data from Array-seq (dark gray) and Visium (light gray) platforms ( $n = 4$  per platform). **e–h**, Bar plots of the numbers of spots under

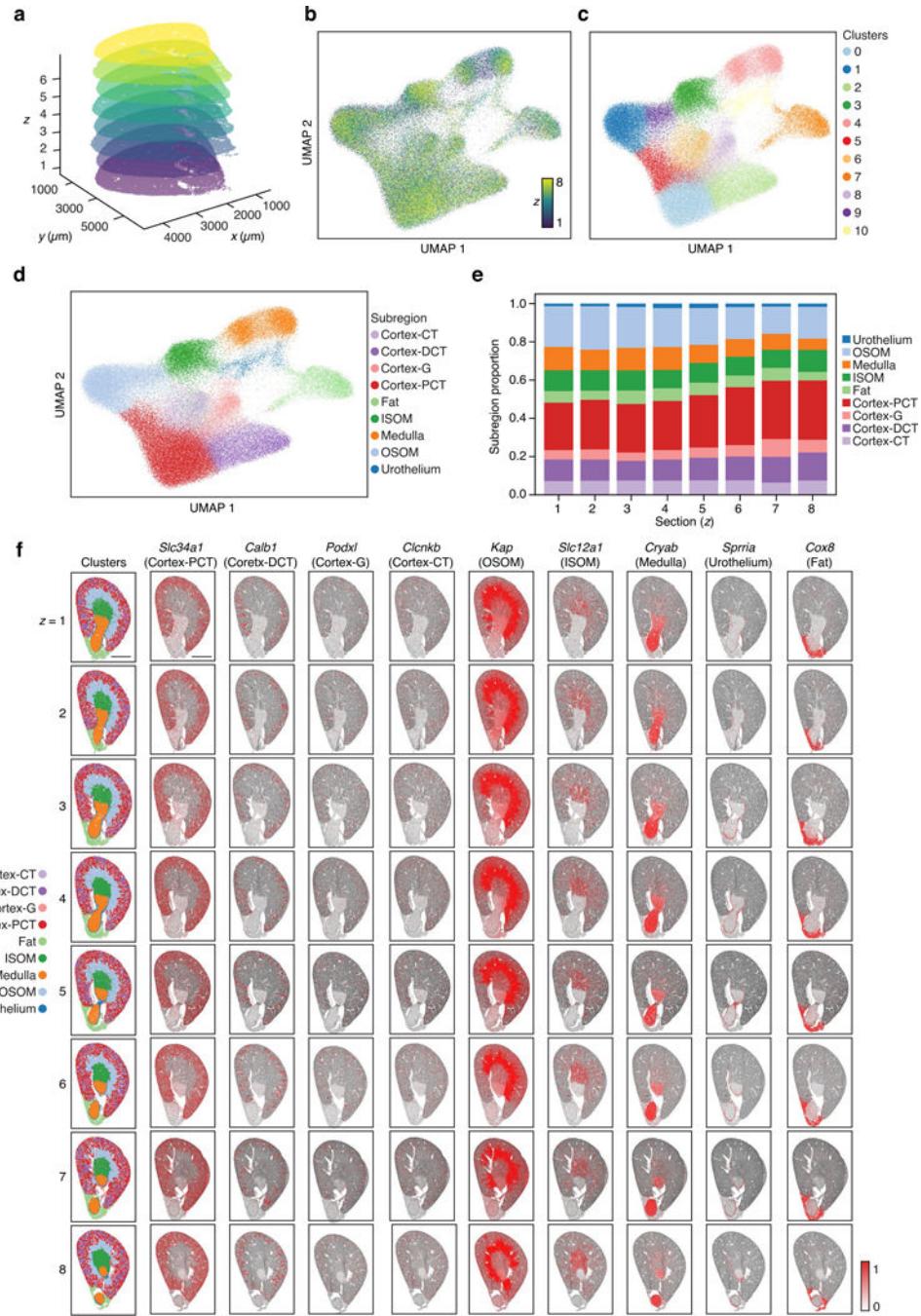
kidney tissue sections (**e**), total genes and UMIs detected (**f**), median genes and UMIs detected per spot (**g**), and genes and UMIs detected per  $\mu\text{m}^2$  of tissue on top of the capture area (**h**). Numbers in **e–h** were calculated using data that were downsampled to similar levels of sequencing depth. Data are presented as mean values  $\pm$  SD ( $n = 4$  per platform). **i**, H&E images (top) overlaid with annotations of tissue subregions (middle) and cell type assignments to spots (bottom) across replicates and platforms (columns). Subregions: CT, Connecting tubule; DCT, Distal Convolute Tubule; G, Glomerulus; PCT, Proximal Convolute Tubule; ISOM, Inner Stripe of Outer Medulla; CD, Collecting Duct; OSOM, Outer Stripe of Outer Medulla. Cell types: ATL, thin ascending limb of loop of Henle; CNT, connecting tubule; CTAL, thick ascending limb of loop of Henle in cortex; DCT, distal convolute tubule; DTL, descending limb of loop of Henle; EC, endothelial cells; ICA, type A intercalated cells of collecting duct; ICB, type B intercalated cells of collecting duct; MTAL, thick ascending limb of loop of Henle in medulla; PC1 and 2; principle cells; PEC, parietal epithelial cells; Per, pericytes; Pod, podocytes; PTS1 and 3, S1 and S3 segments of proximal tubule; Uro, urothelium. Scale bars, 1 mm. **j**, Bar plots of the proportions of cell type annotated spots for each to tissue subregion (bottom axis label) in Array-seq and Visium (top axis label; A =Array-seq, V = Visium) kidney tissue sections ( $n = 1$ , section pair 1). **k**, Bar plots of the proportion of spots under a tissue section which matched indicated cell types. Each spot was annotated with the most abundant cell type inferred computationally. Cell types commonly found are in the left panel and rare cell types are in the right panel (black border, Array-seq; gray border, Visium). Data are presented as mean values  $\pm$  SD ( $n = 4$  per platform).



#### Extended Data Fig. 5 |. Spatial marker gene expression analyses in mouse kidney sections.

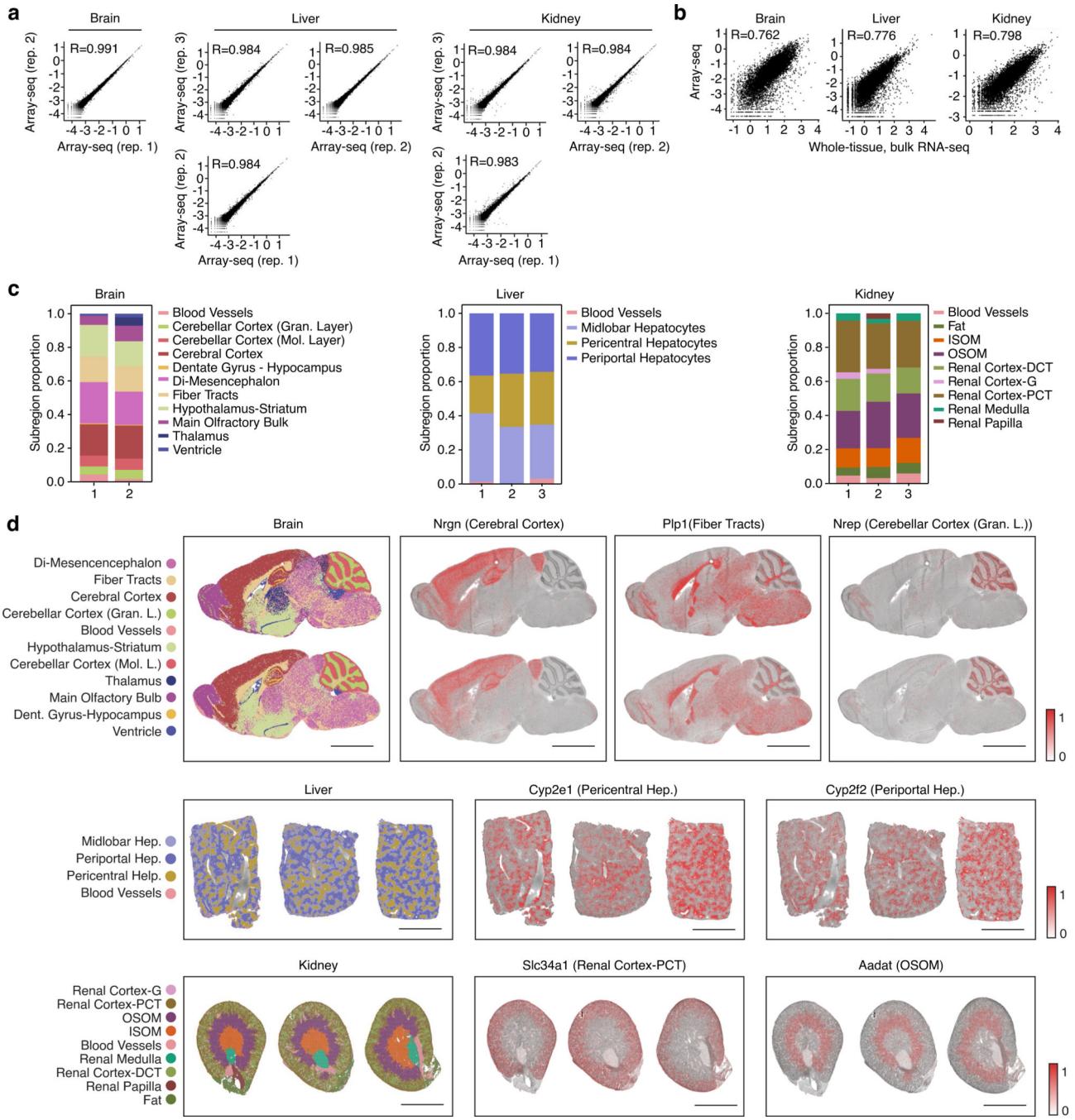
**a**, Scaled  $\log_{10}$  expression of indicated marker genes overlaid on grayscale H&E images of matching kidney sections for indicated tissue subregions and platforms (columns). PCT, Proximal Convolved Tubule; G, Glomerulus; DCT, Distal Convolved Tubule; ISOM, Inner Stripe of Outer Medulla; OSOM, Outer Stripe of Outer Medulla. Scale bars: 1 mm. **b**, Heatmaps of differentially expressed genes (rows) for spots (columns) corresponding to indicated kidney tissue subregions (top). Shown are the top five DE genes obtained with Array-seq data and plotted for both Array-seq (left) and Visium (right) datasets. Values are

z-scores of  $\log_{10}$  normalized UMI counts. **c**, Correlation (Pearson's coefficient) between Array-seq and Visium kidney gene expression and whole-kidney, bulk RNA-seq data. Shown are the normalized  $\log_{10}$  UMI counts which were averaged across all spots and replicates ( $n=4$ ) for indicated spatial platform (Y axis) or across independent replicates ( $n=4$ ) for bulk, whole-tissue RNA-seq (X axis).



**Extended Data Fig. 6 |. Three-dimensional Array-seq analysis of serial mouse kidney sections.**

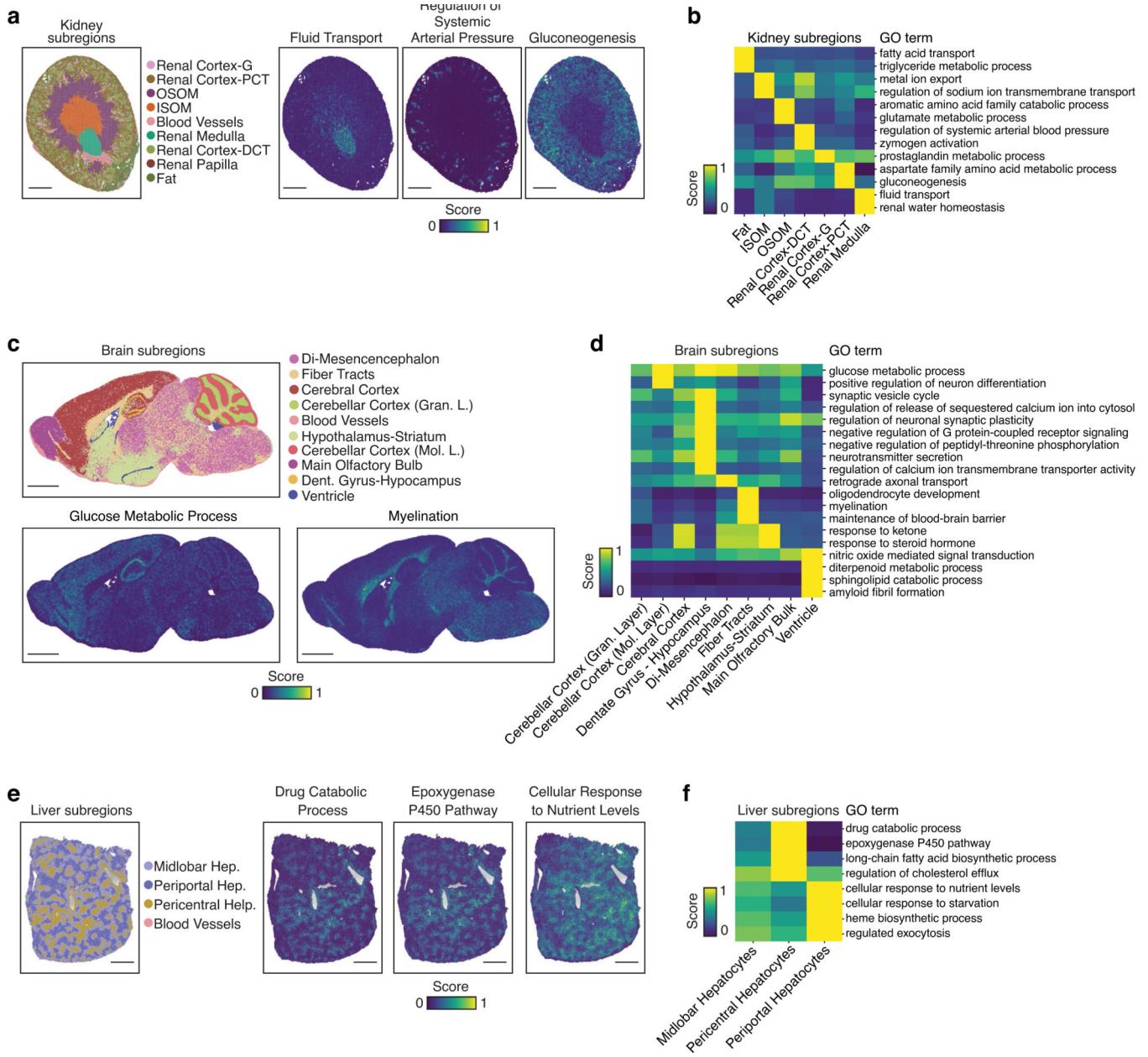
**a**, Images Array-seq data from eight mouse kidney sections aligned in a Z-stack and colored according to z positions within the stack. 80–120  $\mu\text{m}$  were skipped in between each section. **b–d**, Uniform Manifold Approximation and Projection (UMAP) plots of all spots from Array-seq profiles aggregating all eight serial kidney sections and colored by z position (**b**), Leiden clusters (**c**), or manually annotated clusters matching kidney tissue subregions (**d**). CT, Connecting tubule; DCT, Distal Convolute Tubule; G, Glomerulus; ISOM, Inner Stripe of Outer Medulla; ISOM, Inner Stripe of Outer Medulla; PCT, Proximal Convolute Tubule. **e**, Bar plots of the proportion of spots annotated as belonging to indicated kidney tissue subregions (y axis) for each tissue section (x axis). **f**, Spatial plots of indicated kidney tissue subregions (leftmost panels) and subregion marker genes (scaled  $\log_{10}$  expression) overlaid on grayscale H&E images. Consecutive kidney sections are shown from top (z = 1) to bottom (z = 8). Scale bars, 1 mm.



**Extended Data Fig. 7 | Reproducibility of Array-seq for multi-organ section profiling.**

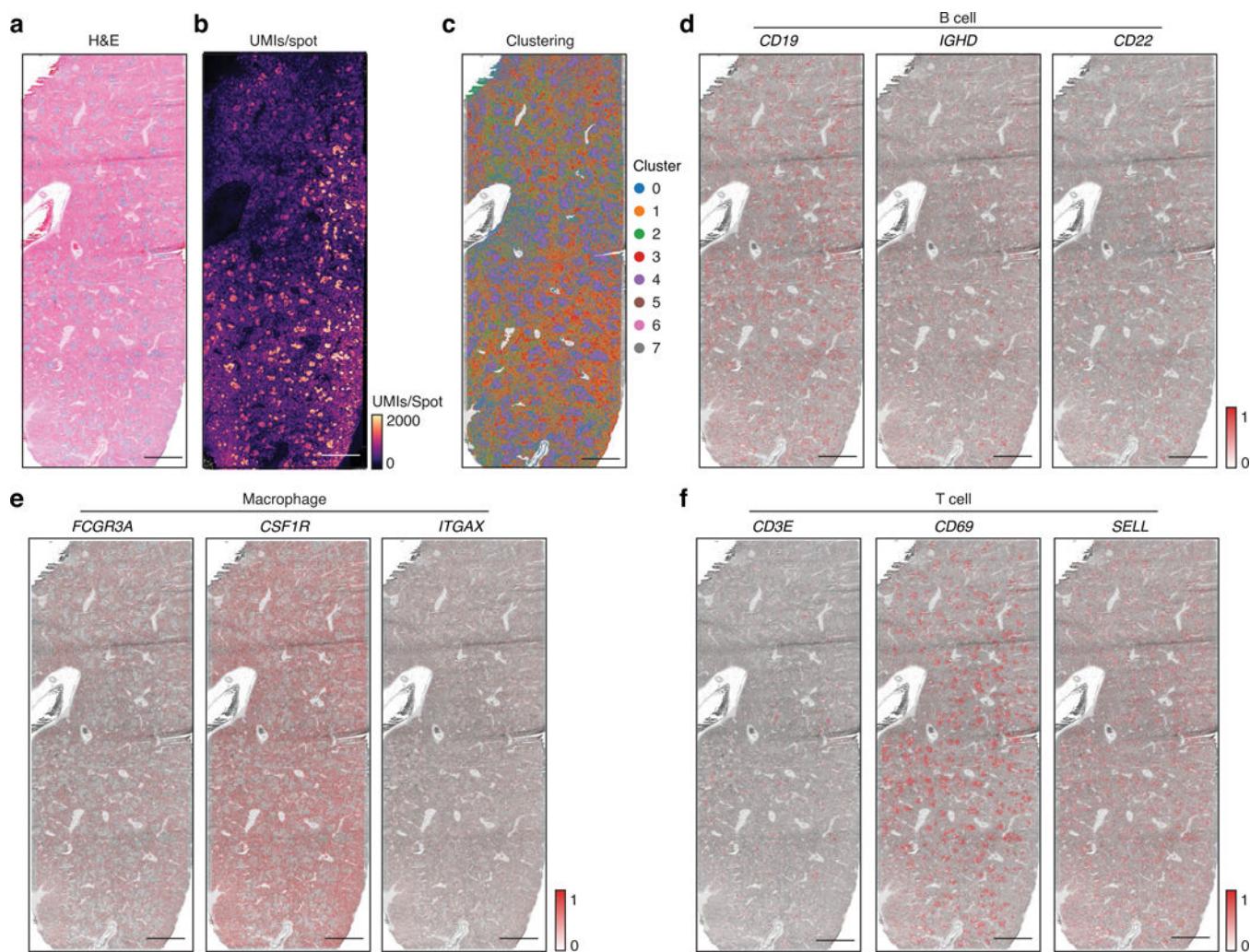
**a, b**, Correlation (Pearson's coefficient) between replicate Array-seq profiles (**a**), or between average Array-seq and bulk RNA-seq datasets (**b**) for indicated mouse tissue types. In **a**, shown are the normalized  $\log_{10}$  unique molecular identifier (UMI) counts which were averaged across all spots for each gene across each replicate. In **b**, Shown are the normalized  $\log_{10}$  unique molecular identifier (UMI) counts which were averaged across all spots for Array-seq data ( $n = 2$  for brain and 3 for liver and kidney sections) (y axis) or across independent organ samples ( $n = 4$ ) for bulk, whole-tissue RNA-seq data (x axis). **c**, Bar

plots of the proportion of the total Array-seq spots under tissue sections which matched indicated tissue subregions. Bars (x axis), replicate sections for each organ type. For brain: Gran. Layer, Granular Layer; Mol. Layer, Molecular Layer. For kidney: DCT, Distal Convolved Tubule; G, Glomerulus; PCT, Proximal Convolved Tubule; ISOM, Inner Stripe of Outer Medulla; OSOM, Outer Stripe of Outer Medulla. **d**, Spatial plots of indicated tissue subregions (leftmost panels) and subregion marker genes (scaled log<sub>10</sub> expression) overlaid on grayscale H&E images. Scale bars, 2 mm.



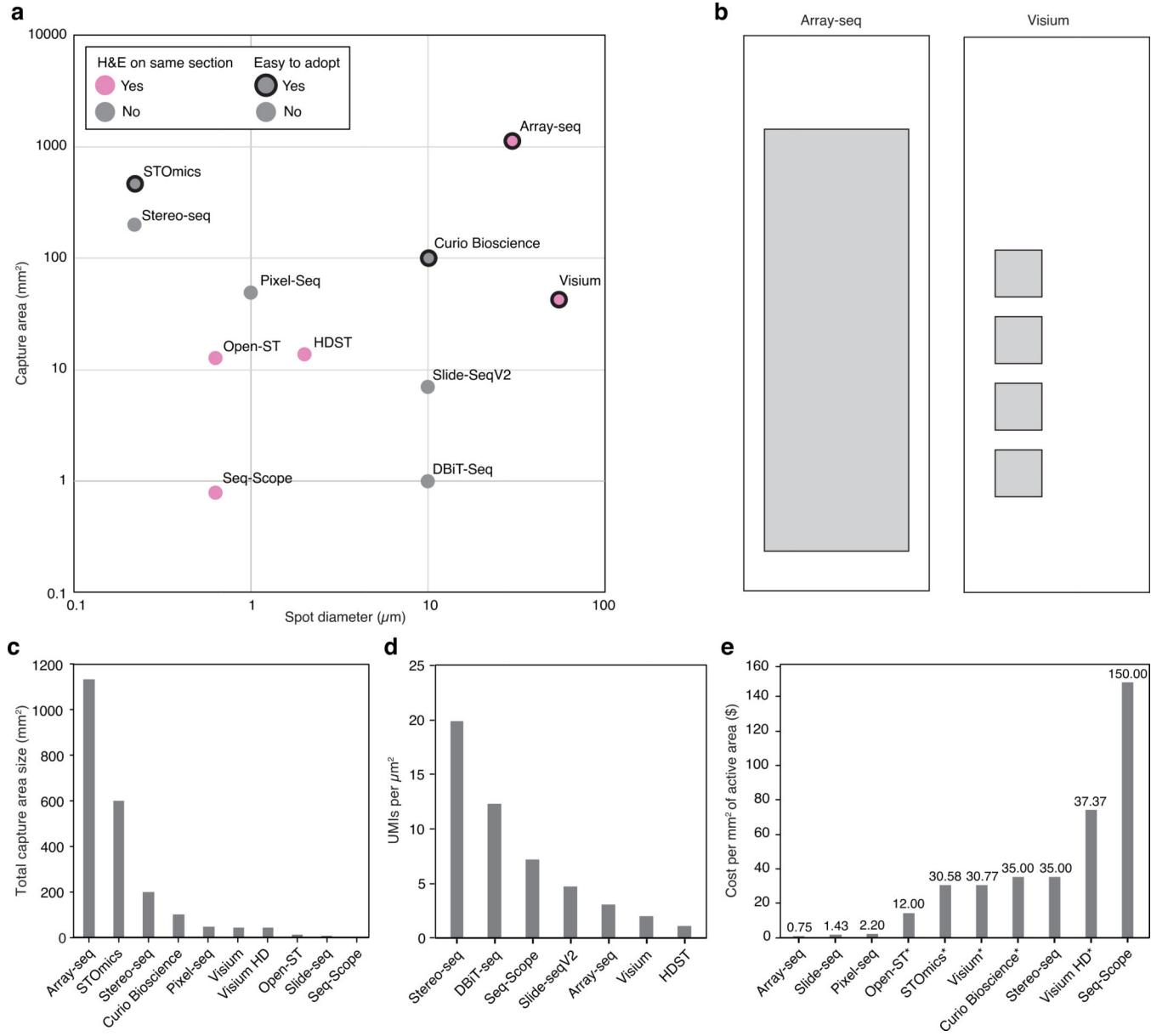
**Extended Data Fig. 8 |. Spatial enrichment of gene ontology gene sets in Array-seq profiles.**  
**a, c, e**, Spatial plots of indicated tissue subregions (left) and normalized enrichment score of indicated gene sets (right) in representative kidney (**a**), brain (**c**), and liver (**e**) sections.

For brain: Gran. L., Granular Layer; Mol. L., Molecular Layer; Dent., dentate. For kidney: DCT, Distal Convolved Tubule; G, Glomerulus; PCT, Proximal Convolved Tubule; ISOM, Inner Stripe of Outer Medulla; ISOM, Inner Stripe of Outer Medulla. Scale bars, 2 mm. **b**, **d–f**, Heatmap of enriched Gene Ontology (GO) terms (rows) in indicated tissue subregions (columns) in representative kidney (**b**), brain (**d**), and liver (**f**) sections. Values are row normalized enrichment scores.



#### Extended Data Fig. 9 |. Array-seq analysis of a whole-mount, human spleen section.

**a**, H&E image of a whole-mount, human spleen section mounted onto an Array-seq slide. **b–f**, Spatial plots of total unique molecular identifiers (UMIs) per spot (**b**), unsupervised, Leiden clustering results (**c**), and scaled gene expression for indicated marker genes for B cells (**d**), macrophages (**e**), and T cells (**f**). Scale bars, 5 mm.



**Extended Data Fig. 10 | Comparison between Array-seq and sequencing-based, spatial transcriptomics methods.**

**a**, Dot plot showing the total surface area available for spatial profiling (*y* axis) and the diameter of the barcoded spots, beads, or DNA species arranged on the slides or substrates used for mRNA capture (*x* axis) across indicated spatial transcriptomics methods compatible with fresh-frozen histological sections. Pink dots indicate that a method is compatible with H&E imaging on the same section that is used for spatial profiling. Easy-to-adopt indicates methods which can be readily deployed without the need for special expertise, instrumentation, or custom-made reagents. **b**, Diagram of Array-seq (left) and Visium (right) slides showing the mRNA capture area (grey) sizes and positions at scale. **c–e**, Bar plots of the total surface area available for mRNA capture (**c**), the sensitivity computed in total unique molecular identifiers (UMIs) detected per  $\mu\text{m}^2$  (**d**), and the cost per  $\text{mm}^2$  of active

surface area (**e**) for indicated method (x axis). In **d**, the sensitivity analysis was performed using publicly available, preprocessed datasets for each method on MOB tissue, except for Seq-Scope (mouse liver) and DBiT-seq (mouse embryo). In **e**, asterisks indicate that library preparation costs are included.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank members of the laboratory of N.C., V. Maran and M. Soumillon for valuable discussions. N.C. thanks S. Uematsu for the introduction to DNA microarrays. We thank UChicago Core facilities for support: the Single Cell Immunophenotyping Core, the Animal Resources Center, the Research Computing Center and the Integrated Light Microscopy Core; and SciStories for help with artwork. D.C. was supported by National Institutes of Health (NIH) grant T32-GM007281. N.C was supported by NIH grants DP2-AI145100 and U01-AI160418, the CZI grant DAF2020-217464 and a grant from the Chan Zuckerberg Initiative DAF (<https://doi.org/10.37921/767230ofotux>), an advised fund of Silicon Valley Community Foundation (funder <https://doi.org/10.13039/100014989>), the Agilent ACT-UR program (grant ID 4843), the UChicago Center for Interdisciplinary Study of Inflammatory Intestinal Disorders (NIDDK P30 DK042086), the UChicago Diabetes Research and Training Center (NIDDK P30 DK020595), the Duckworth Family Commercial Promise Cancer Research Award, the Robert Lavichant Faculty Innovation Award and funds from the Chicago Immunoengineering Innovation Center and the Pritzker School of Molecular Engineering at the University of Chicago.

## Data availability

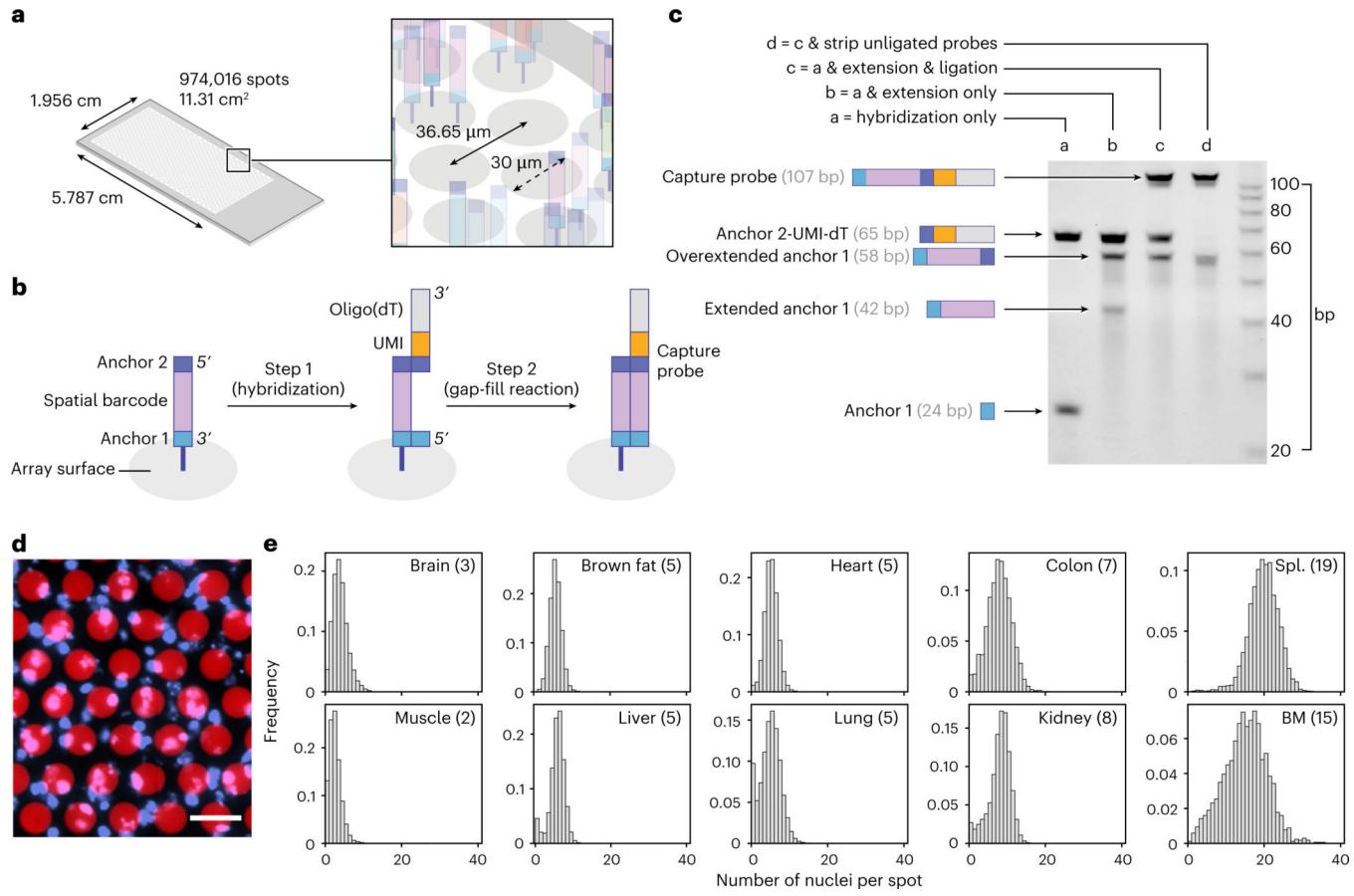
The sequencing data generated during this study have been deposited in the Gene Expression Omnibus under accession numbers GSE266246 (bulk, whole-tissue RNA-seq) and GSE266244 (ST). The uncropped scan of the DNA electrophoresis gel and the H&E images are available via Zenodo at <https://doi.org/10.5281/zenodo.13234161>. Source data are provided with this paper.

## References

1. Moses L. & Pachter L. Museum of spatial transcriptomics. *Nat. Methods* 19, 534–546 (2022). [PubMed: 35273392]
2. Lein E, Borm LE & Linnarsson S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* 358, 64–69 (2017). [PubMed: 28983044]
3. Rao A, Barkley D, Franca GS & Yanai I. Exploring tissue architecture using spatial transcriptomics. *Nature* 596, 211–220 (2021). [PubMed: 34381231]
4. Crosetto N, Bienko M. & van Oudenaarden A. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet* 16, 57–66 (2015). [PubMed: 25446315]
5. Stahl PL et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353, 78–82 (2016). [PubMed: 27365449]
6. Stickels RR et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol* 39, 313–319 (2021). [PubMed: 33288904]
7. Vickovic S. et al. High-definition spatial transcriptomics for *in situ* tissue profiling. *Nat. Methods* 16, 987–990 (2019). [PubMed: 31501547]
8. Liu Y. et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* 183, 1665–1681 (2020). [PubMed: 33188776]
9. Chen A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 185, 1777–1792 (2022). [PubMed: 35512705]

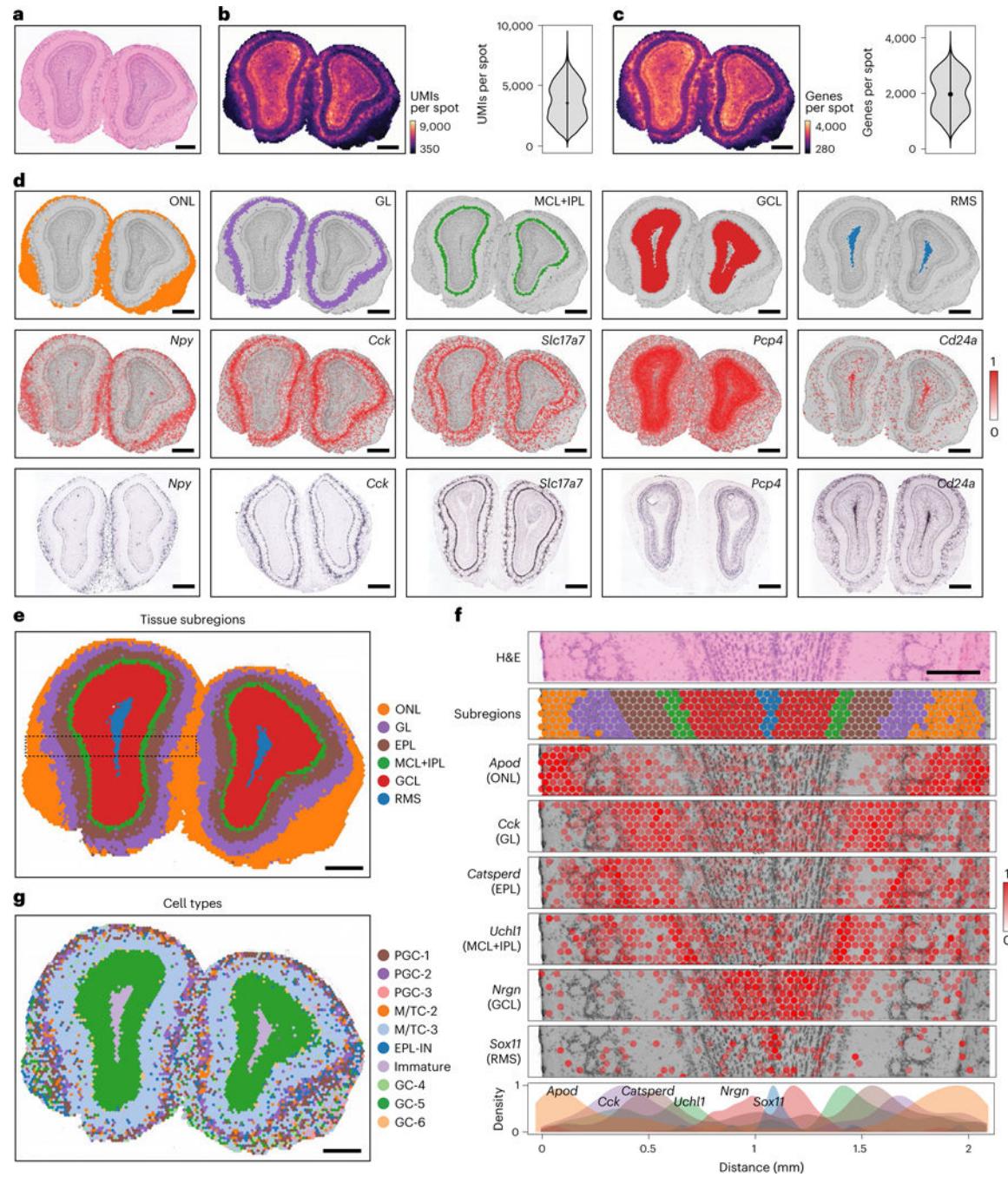
10. Cho CS et al. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* 184, 3559–3572 (2021). [PubMed: 34115981]
11. Fu X. et al. Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. *Cell* 185, 4621–4633 (2022). [PubMed: 36368323]
12. Salmen F. et al. Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. *Nat. Protoc.* 13, 2501–2534 (2018). [PubMed: 30353172]
13. Schena M, Shalon D, Davis RW & Brown PO Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470 (1995). [PubMed: 7569999]
14. Chee M. et al. Accessing genetic information with high-density DNA arrays. *Science* 274, 610–614 (1996). [PubMed: 8849452]
15. Ren B. et al. Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309 (2000). [PubMed: 11125145]
16. Islam S. et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11, 163–166 (2014). [PubMed: 24363023]
17. Hughes TR et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol* 19, 342–347 (2001). [PubMed: 11283592]
18. Chen X, Sun YC, Church GM, Lee JH & Zador AM Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res.* 46, e22 (2018). [PubMed: 29190363]
19. Traag VA, Waltman L. & van Eck NJ From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep* 9, 5233 (2019). [PubMed: 30914743]
20. Long Y. et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun* 14, 1155 (2023). [PubMed: 36859400]
21. Zhao E. et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol* 39, 1375–1384 (2021). [PubMed: 34083791]
22. Lein ES et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176 (2007). [PubMed: 17151600]
23. Ma Y. & Zhou X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat. Biotechnol* 40, 1349–1359 (2022). [PubMed: 35501392]
24. Tepe B. et al. Single-cell RNA-seq of mouse olfactory bulb reveals cellular heterogeneity and activity-dependent molecular census of adult-born neurons. *Cell Rep.* 25, 2689–2703 (2018). [PubMed: 30517858]
25. Cable DM et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol* 40, 517–526 (2022). [PubMed: 33603203]
26. Kleshchevnikov V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol* 40, 661–671 (2022). [PubMed: 35027729]
27. Kirita Y, Wu H, Uchimura K, Wilson PC & Humphreys BD Cell profiling of mouse acute kidney injury reveals conserved cellular responses to injury. *Proc. Natl Acad. Sci. USA* 117, 15874–15883 (2020). [PubMed: 32571916]
28. Hildebrandt F. et al. Spatial transcriptomics to define transcriptional patterns of zonation and structural components in the mouse liver. *Nat. Commun* 12, 7046 (2021). [PubMed: 34857782]
29. Alexandre YO & Mueller SN Splenic stromal niches in homeostasis and immunity. *Nat. Rev. Immunol* 23, 705–719 (2023). [PubMed: 36973361]
30. Cang Z. et al. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nat. Methods* 20, 218–228 (2023). [PubMed: 36690742]
31. Chen A. et al. Single-cell spatial transcriptome reveals cell-type organization in the macaque cortex. *Cell* 186, 3726–3743 (2023). [PubMed: 37442136]
32. Schott M. et al. Open-ST: high-resolution spatial transcriptomics in 3D. *Cell* 187, 3953–3972 (2024). [PubMed: 38917789]
33. Takahama M. et al. A pairwise cytokine code explains the organism-wide response to sepsis. *Nat. Immunol* 25, 226–239 (2024). [PubMed: 38191855]
34. Liu Y. et al. High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat. Biotechnol* 41, 1405–1409 (2023). [PubMed: 36823353]

35. Vickovic S. et al. SM-Omics is an automated platform for high-throughput spatial multi-omics. *Nat. Commun* 13, 795 (2022). [PubMed: 35145087]
36. Fan Y. et al. Expansion spatial transcriptomics. *Nat. Methods* 20, 1179–1182 (2023). [PubMed: 37349575]
37. Buschmann T. DNABarcodes: an R package for the systematic construction of DNA sample tags. *Bioinformatics* 33, 920–922 (2017). [PubMed: 28052927]
38. Soumillon M, Cacchiarelli D, Semrau S, van Oudenaarden A. & Mikkelsen T. Characterization of directed differentiation by high-throughput single-cell RNA-seq. Preprint at bioRxiv 10.1101/003236 (2014).
39. Pandey S. et al. A whole-tissue RNA-seq toolkit for organism-wide studies of gene expression with PME-seq. *Nat. Protoc* 15, 1459–1483 (2020). [PubMed: 32076350]
40. Kadoki M. et al. Organism-level analysis of vaccination reveals networks of protection across tissues. *Cell* 171, 398–413 (2017). [PubMed: 28942919]
41. Dobin A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013). [PubMed: 23104886]
42. Wolf FA, Angerer P. & Theis FJ SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15 (2018). [PubMed: 29409532]
43. van der Walt S. et al. scikit-image: image processing in Python. *PeerJ* 2, e453 (2014). [PubMed: 25024921]
44. Ni Z. et al. SpotClean adjusts for spot swapping in spatial transcriptomics data. *Nat. Commun* 13, 2971 (2022). [PubMed: 35624112]
45. Fang Z, Liu X. & Peltz G. GSEApY: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 10.1093/bioinformatics/btac757 (2023).



**Fig. 1 |. On-slide assembly of spatially barcoded mRNA capture probes using microarrays.**

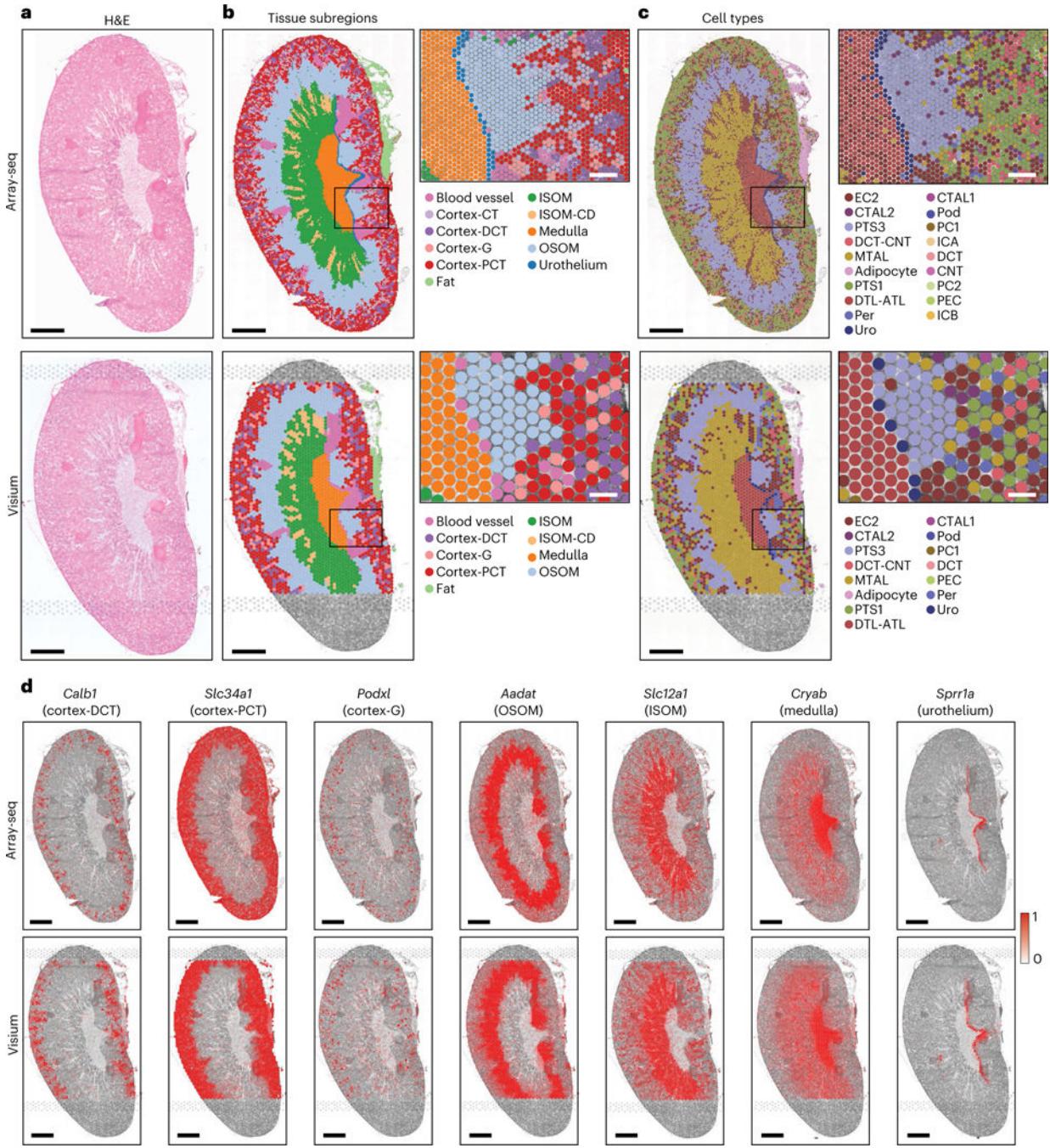
**a**, Schematic of a custom-sequence, large-format microarray comprising 974,016 spots of 30 μm in diameter that are arranged in 1,068 rows and 912 columns across 11.31 cm<sup>2</sup> in surface area. Each spot on the array carries a unique spatial barcode sequence. **b**, Overview of the on-slide assembly of mRNA capture probes by hybridization of indicated oligonucleotides (step 1) followed by an extension–ligation or ‘gap-fill’ reaction (step 2). **c**, Polyacrylamide gel electrophoresis (PAGE) analysis of the oligonucleotide products obtained after the indicated on-slide assembly procedures in lanes a through d. **d**, Representative fluorescence image of an area from a mouse brain section placed on an Array-seq slide hybridized with a Cy3-labeled, anchor 2 probe. Blue indicates the DAPI nuclear stain of the brain section. Scale bar, 50 μm. **e**, Histograms of the numbers of DAPI-positive nuclei per spot for indicated mouse organs. BM, bone marrow. Parentheses indicate median values.



**Fig. 2 |. Array-seq accurately captures region-specific expression patterns in tissues.**

**a**, H&E image of a section (10  $\mu\text{m}$ ) from the MOB system of the mouse brain profiled by Array-seq. **b,c**, Numbers of UMIs (**b**) and genes (**c**) detected per spot across the tissue section visualized as spatial plots (left) and violin plots (right). **d,e**, Unsupervised clustering highlights the histological layers of the MOB tissue (**d,e**), as confirmed by the spatial expression (**d**) and *in situ* hybridization (**d**, bottom; images from the Allen Brain Atlas database<sup>22</sup>) of indicated gene markers. ONL, olfactory nerve layer; GL, glomerular layer; EPL, external plexiform layer; MCL, mitral cell layer; IPL, internal plexiform layer; GCL, granule cell layer; RMS, radial glial cell layer.

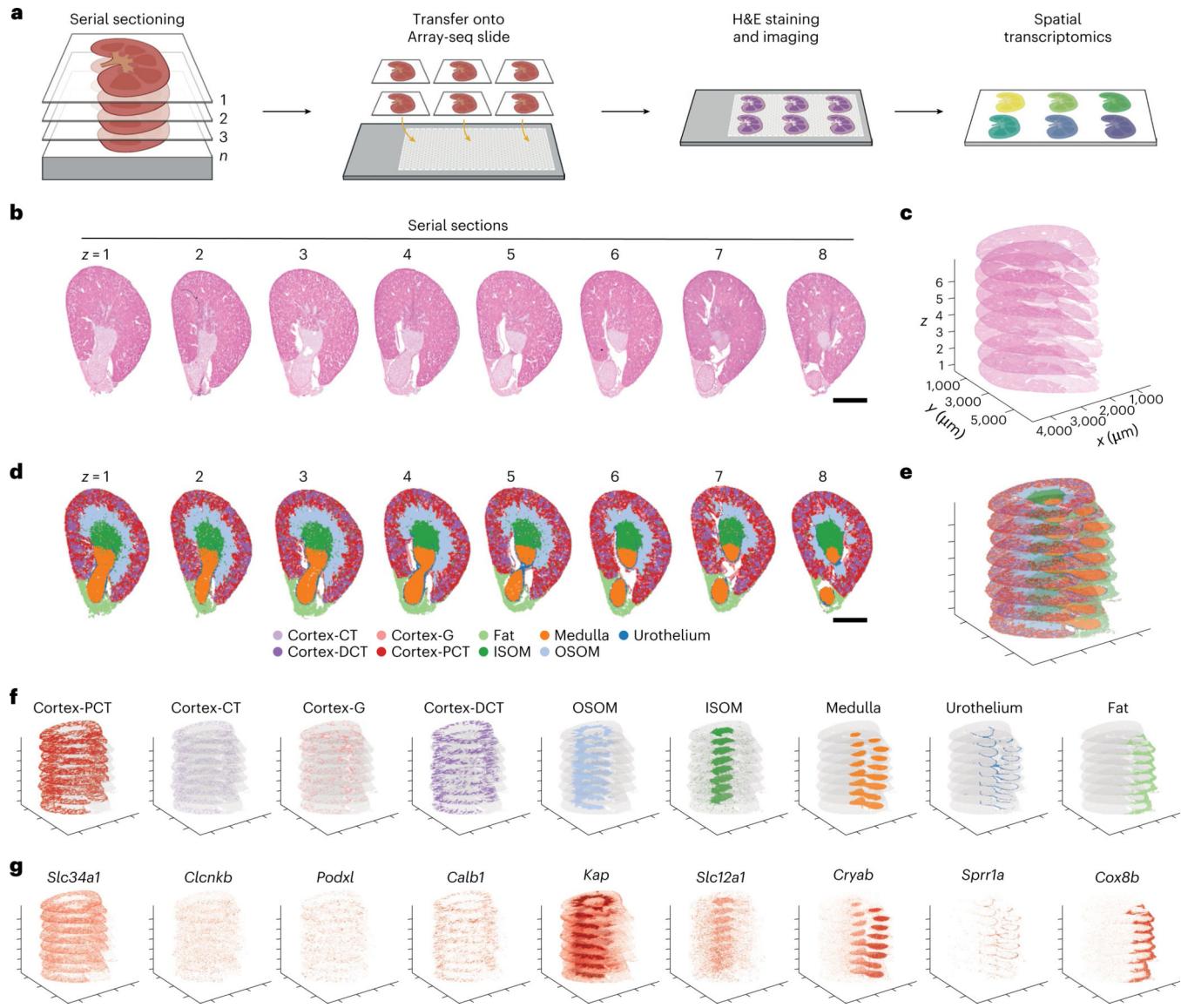
granule cell layer; RMS, rostral migratory stream. **f**, Magnified image of the MOB inset shown in **e** (dashed line). From top to bottom: H&E; subregion annotations; scaled  $\log_{10}$  expression of indicated gene markers overlaid on a grayscale H&E image; and line plots showing gene expression smoothed using kernel density estimation ( $y$  axis) of indicated genes across the selected tissue area ( $x$  axis). **g**, Cell-type assignments across spots on the Array-seq slide for the indicated MOB cell types. Cell types with the highest inferred proportion per spot were assigned to each spot. PGC, periglomerular cell; M/TC, mitral and tufted cell; EPL-IN, external plexiform layer interneuron; GC, granule cell. Scale bars, 500  $\mu\text{m}$  (**a–e,g**) and 200  $\mu\text{m}$  (**f**).



**Fig. 3 |. Side-by-side comparison of Array-seq and Visium data using adjacent mouse kidney sections.**

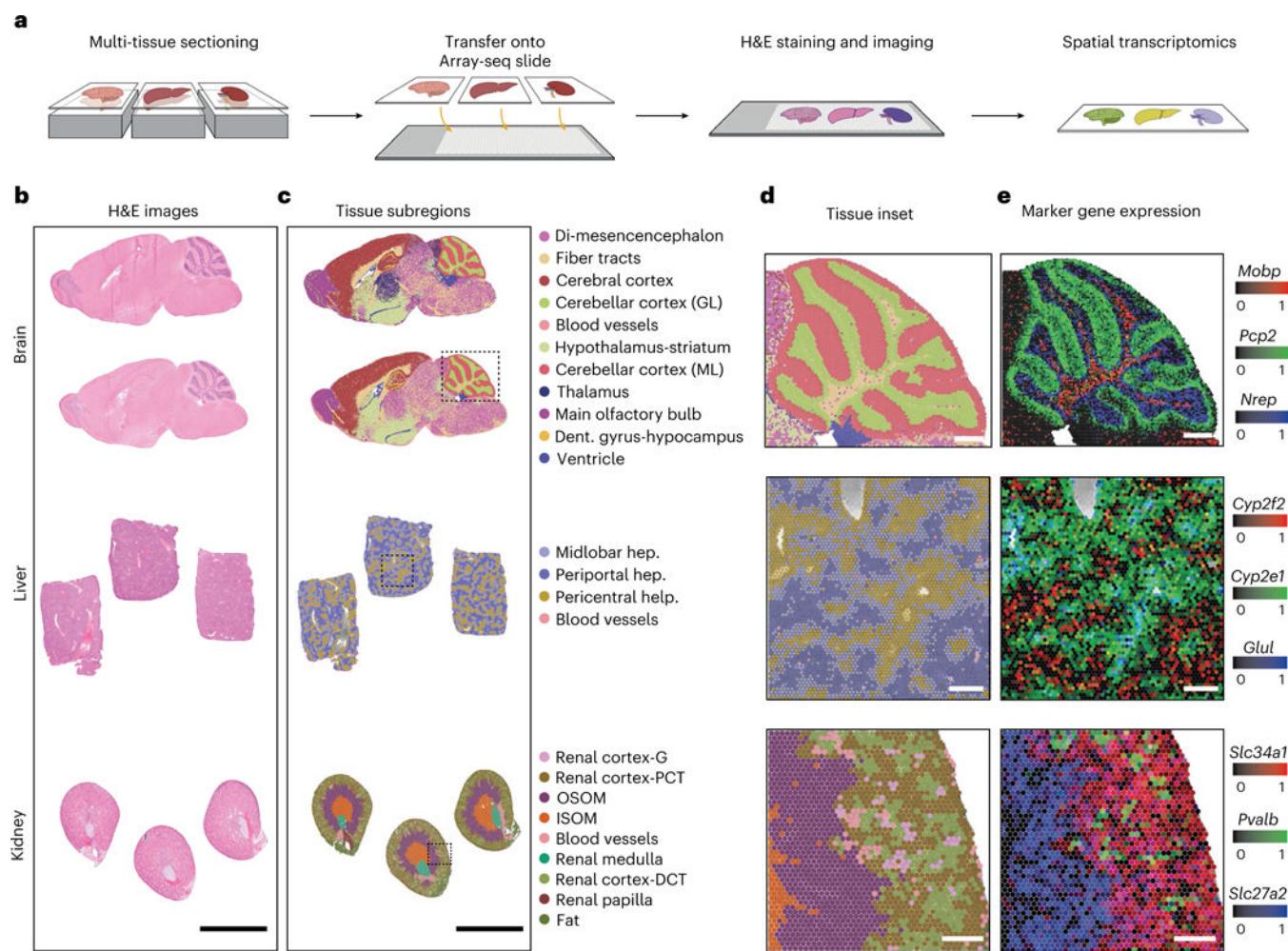
**a**, H&E images of two immediately adjacent sections from a mouse kidney placed onto Array-seq (top) and Visium (bottom) slides. **b,c**, Annotation of kidney tissue subregions (**b**) and cell types (**c**) for Array-seq (top) and Visium (bottom) datasets. Cell types with the highest inferred proportion per spot were assigned to each spot. Insets indicate the localization of the magnified images shown in **b** and **c**. Subregions: CT, connecting tubule; DCT, distal convoluted tubule; G, glomerulus; PCT, proximal convoluted tubule; ISOM/

OSOM, inner/outer stripe of outer medulla; CD, collecting duct. Cell types: ATL, thin ascending limb of loop of Henle; CNT, connecting tubule; CTAL, thick ascending limb of loop of Henle in cortex; DTL, descending limb of loop of Henle; EC, endothelial cell; ICA, type A intercalated cells of collecting duct; ICB, type B intercalated cells of collecting duct; MTAL, thick ascending limb of loop of Henle in medulla; PC1 and PC2, principle cells; PECs, parietal epithelial cells; Per, pericytes; Pod, podocytes; PTS1 and PTS3, S1 and S3 segments of proximal tubule; Uro, urothelium. **d**, Scaled  $\log_{10}$  expression of differentially expressed marker genes for indicated subregion (columns) in Array-seq (top) and Visium (bottom) data. Scale bars, 1 mm (**a–d**) and 200  $\mu\text{m}$  (insets shown in **b** and **c**).



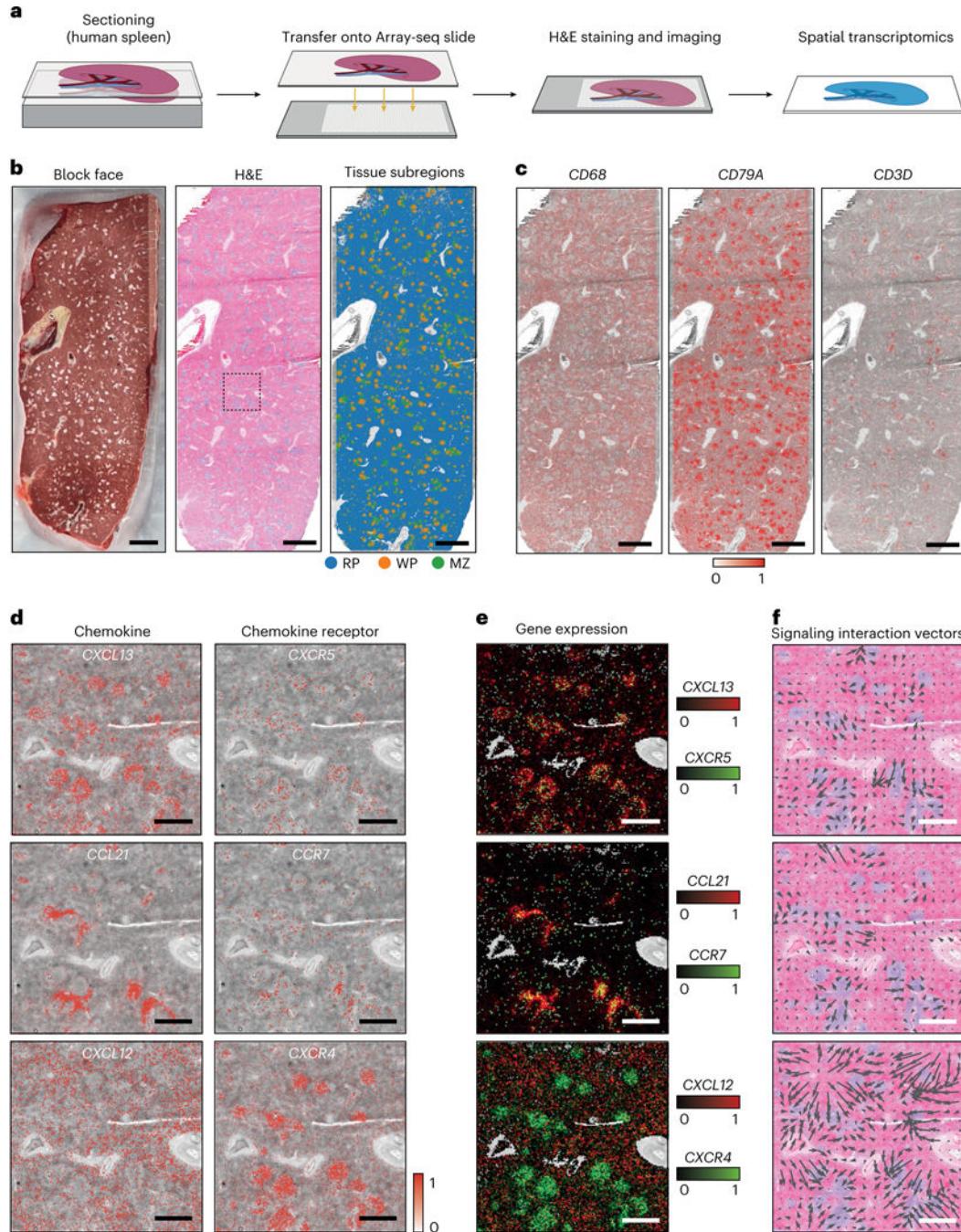
**Fig. 4 | Array-seq enables the 3D profiling of spatial transcriptomes.**

**a**, Schematic overview of the experimental workflow. **b,c**, H&E images of eight mouse kidney sections (10  $\mu\text{m}$  each) profiled by Array-seq slide shown in two dimensions (**b**) and aligned in a z-stack (**c**). 80–120  $\mu\text{m}$  were skipped between sections. **d,e**, Unsupervised clustering highlights the histological subregions of the kidney as shown in two (**d**) and three (**e**) dimensions. **f,g**, z-stack of H&E images (grayscale) from serial kidney sections overlaid with indicated tissue subregions (**f**) and marker genes (**g**, scaled  $\log_{10}$  expression). Scale bars, 1 mm (**b,d**).



**Fig. 5 |. Multi-tissue, ST profiling using Array-seq.**

**a**, Schematic overview of the experimental workflow. **b**, H&E images of tissue sections from mouse brain (top;  $n=2$ ), liver (middle;  $n=3$ ) and kidney (bottom;  $n=3$ ) organs that were placed onto a single Array-seq slide. **c**, Unsupervised clustering highlights the histological subregions of each organ type. Dent., dentate; GL, granular layer; Hep., hepatocyte; ML, molecular layer. **d,e**, Magnified images of indicated insets (dashed boxes in **c**) for brain (top), liver (middle) and kidney (bottom) showing tissue subregions (**d**) and scaled  $\log_{10}$  expression for indicated marker genes (**e**). Scale bars, 5 mm (**b,c**) and 250  $\mu\text{m}$  (**d,e**).



**Fig. 6 |. Array-seq profiling of a whole-mount section from a human spleen organ.**

**a**, Schematic overview of the experimental workflow. **b**, Images of the block face of an embedded, longitudinally sectioned human spleen (left), the H&E image from a section immediately adjacent to the block face view (middle) and manually annotated unsupervised clusters of the Array-seq data obtained on the same section as shown by H&E (right). MZ, marginal zone; RP, red pulp; WP, white pulp. **c**, Scaled log<sub>10</sub> expression of indicated marker genes for macrophages (*CD68*), B cells (*CD79A*) and T cells (*CD3D*). **d,e**, Scaled log<sub>10</sub> expression of genes encoding indicated chemokine ligand (left panels in **d**; red color in **e**)

and chemokine receptor (right panels in **d**; green color in **e**) pairs, overlaid on the grayscale H&E image of the inset shown in **b** (middle panel). **f**, Computationally inferred signaling vectors of the indicated chemokine ligand–receptor pairs. Scale bars, 5 mm (**b,c**) and 1 mm (**d–f**).