Supplementary Materials for

Embryo-scale, single-cell spatial transcriptomics

Sanjay R. Srivatsan†, Mary C. Regier†, Eliza Barkan, Jennifer M. Franks, Jonathan S. Packer, Parker Grosjean, Madeleine Duran, Sarah Saxton, Jon J Ladd, Malte Spielmann, Carlos Lois, Paul D. Lampe, Jay Shendure*, Kelly R. Stevens*, Cole Trapnell*

†These authors contributed equally to this work.
*Corresponding author. Email: shendure@uw.edu (J.S.); ksteve@uw.edu (K.R.S.); coletrap@uw.edu (C.T.)

Published 2 July 2021, Science 373, 111 (2021)
DOI: 10.1126/science.abb9536

This PDF file includes:

Materials and Methods
Figs. S1 to S33
Tables S1 and S2
References

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/content/373/6550/111/suppl/DC1)

Data Files S1 to S3
MDAR Reproducibility Checklist
Materials and Methods

Overview
Sci-Space is a single cell technology wherein positional information is recorded and read out in conjunction with single cell transcriptomes. This process is mediated by the spatial transfer of hashing oligos (which resemble poly-adenylated (polyA) transcripts) onto cells. A combination of 3 hashing oligos is meant to uniquely specify the slide from which a cell originates as well as its position within a regular grid on that slide. After performing sci-RNA-seq library preparation and sequencing, the spatial barcodes found within each cell’s library are used to assign it to a slide as well as to position it within the grid. The molecular biology details needed to perform the sci-Space procedure and library preparation are outlined below as well as in a series of published protocols.io sites:

1. [Creating sci-Space Grids for Spatial Barcoding](36)
2. [Spatial Transfer of Oligonucleotides and Imaging](37)
3. [Single cell RNA sequencing library preparation (2-level sci-RNA-seq)](38)

Creating sci-Space grids for spatial barcoding

A thin membrane of dried agarose was fabricated on the surface of microscope slides (Superfrost Plus, Thermofisher). This agarose matrix absorbed and retained an array of spotted oligo hashes. To prepare nuclease-free agarose, 3% w/v low melting temperature agarose powder (SeaPlaque, Lonza, Bend, OR) was added to deionized water containing 0.1% v/v diethyl pyrocarbonate, incubated 2 hr at room temperature, and autoclaved for 15 min. The uniform thickness of the layer of agarose across the slide surface was patterned using spacers of two stacked 22 x 22 mm, number one thickness (0.15 ± 0.02 mm each) coverslips overhanging either end of the slide. Molding of the agarose was performed by pipetting a 300 uL volume of heated agarose solution into the center of the slide and slowly placing a second slide onto the agarose solution avoiding the formation of bubbles. The molding slide was allowed to rest on the cover glass spacers. After the agarose had gelled between the two slides (~30-60 min on ice) a razor blade was used to release the exposed edges of the agarose layer from the top, molding slide. The two slides were then carefully slid apart and the cover glass spacers were removed. The resulting thin layer of agarose gel was dried onto the bottom slide overnight in a biosafety cabinet. All agarose slides were UV-treated for 20-30 min prior to spotting to further protect against nuclease activity.

The space-grid array of hashing oligos and SYBR green reference points was spotted onto agarose-coated slides using a QArray2 microarray scanner (Genetix, New Milton, Hampshire, GB). A series of 384-well high sample recovery plates (Molecular Devices, San Jose, CA, X7020) was prepared containing a final concentration of 15uM spot oligo and 2.5uM sector oligo per well (Integrated DNA Technologies, Coralville, IA), and 0.5% v/v glycerol, with or without SYBR green dye ((5x) Thermofisher, S7585) to achieve the predetermined oligo and SYBR green reference point layout when a 21 x 21 spot/pin array was printed with 16 spotting pins (4 x 4 grid). These printing parameters gave space-grids containing 7056 (84 x 84) spots of unique oligo combinations. The spotting height was adjusted to ensure consistent contact of the spotting pins with the transfer slides’ agarose coating.
Testing blotting concentrations of hash oligos and SYBR green
Space-grids for testing hash oligo blotting concentrations were prepared as noted above using the QArray2 microarray scanner (Genetix, New Milton, Hampshire, GB). Each space-grid was given a single distinct DNA barcode sequence at a chosen final concentration (10μM, 20μM, 25μM or 50μM) with a single sector marked with 5x SYBR green. These space-grids were then blotted onto a series of mouse embryo sections ranging from E13 to E16 (Zyagen, San Diego, CA), by sandwiching the two slides (tissue and space-grid). First, permeabilization and hashing solution was prepared for each slide by mixing a unique slide-specific hash oligo (5μL at 10μM) in the 495uL permeabilization solution [10mM Tris/HCl pH 7.4, 10mM NaCl, 3mM MgCl2 with 1% v/v superase inhibitor (Invitrogen) and 0.1%v/v IGEPAL CA-630 (Sigma Aldrich)]. Following permeabilization, each slide was barcoded via transfer with a test space-grid. The transfer was then imaged and the cells were harvested by cell scraping into a solution of 5% paraformaldehyde (cat. no. 100504-940, VWR) in 1x PBS. This is described in detail in a subsequent section (“Spatial Transfer of Oligonucleotides and Imaging”). After 15 minutes of fixation on ice, cells were centrifuged (800g for 10 minutes), pooled and subjected to sci-RNA-seq2 library preparation (39). This is described in detail in a subsequent section (“Single cell RNA sequencing library preparation”).

To test co-cultured human and mouse cells, HEK293T cells and NIH3T3 cells were placed in a droplet on a coverslip coated with 1% gelatin. 4 coverslips were prepared and cells were allowed to attach overnight to the coverslip surface. Next, each coverslip was permeabilized as detailed above and labeled with a gel containing a single hash oligo. These nuclei were then scraped into paraformaldehyde, fixed for 15 minutes on ice and subject to sci-RNA-seq library preparation as detailed below.

Testing spotted space-grids
Hash oligos from three space-grids were dissolved in 500μL of permeabilization solution [10mM Tris/HCl pH 7.4, 10mM NaCl, 3mM MgCl2 with 1% v/v superase inhibitor (Invitrogen) and 0.1%v/v IGEPAL CA-630 (Sigma Aldrich)]. Concurrently, three aliquots of 2 million HEK293T cells were harvested and washed once with 1x PBS. The resuspended hash oligo solutions were then used to lyse and label the HEK293T nuclei. After a 3 minute incubation on ice, the nuclei suspension was chemically fixed with 5mL of 4% paraformaldehyde and incubated for 15 minutes on ice. Nuclei were then pelleted at 500g for 5 minutes, washed with 500μL Nuclei Buffer (NSB) [10mM Tris/HCl pH 7.4, 10mM NaCl, 3mM MgCl2 with 1% v/v superase inhibitor (Invitrogen) 1% v/v BSA (New England Biolabs)] and permeabilized by resuspension in 500μL NB + 0.2% Triton-X. These nuclei were centrifuged and washed with 500μL of NSB. These nuclei were then pelleted and 5000 nuclei from each sample was loaded into indexed reverse-transcription reactions. Reverse transcription was performed as described previously (39) were pooled and 25 nuclei were sorted into a 96 well plate containing 16μL of elution buffer per well. Libraries were prepared by performing an indexed PCR using 20uL of NEBNext High-Fidelity 2X PCR Master Mix (NEB), 2μL of 10μM indexed P5 primer and 2μL of 10μM indexed P7 primer. PCR was run for 18 cycles with the following settings: 72°C for 5 min, 98°C for 30 sec, 18 cycles of (98°C for 10 sec, 66°C for 30 sec, 72°C for 30 sec) and a final 72°C for 5 min. These libraries were then pooled and sequenced on a Nextseq 500 (Illumina, San Diego, CA) using a high output 75 cycle kit (Read 1: 18 cycles, Read 2: 52 cycles, Index 1: 10 cycles and Index 2: 10 cycles).
Designing and fabricating transfer clips

Clips were fabricated to securely hold the space-grid slides and tissue slides together during transfer of the oligo hashes from the spotted agarose to the nuclei within the embryo tissue sections. Clips were designed in SolidWorks v24 (Dassault Systèmes SolidWorks Corp., Waltham, MA). The clips spanned the stacked transfer and tissue slides’ width and included fastening features on each end with slight overhangs that fit over the top of the stacked slides. The clips were 3D printed on a Makergear M2 (Makergear, Beachwood, OH) printer using consumer grade poly(lactic acid) plastic filament (Makergear). Two clips were used per transfer, placed one on either side of the embryo section.

Spatial transfer of oligonucleotides and imaging

Protocols.io - Spatial Transfer of Oligonucleotides and Imaging (37)

Serial sections of an E14.0 mouse embryo were purchased (Zyagen, San Diego, CA) and stored at -80°C prior to use. Oligo hashes were transferred in their arrayed pattern from the space-grid slides to fresh-frozen embryo sections by diffusion through cell permeabilization buffer. First, the embryo slide was placed so that it rested (tissue facing up) with the tissue section between two transfer clips. Subsequently, 500μL of nuclei permeabilization buffer [10mM Tris/HCl pH 7.4, 10mM NaCl, 3mM MgCl2 with 1% v/v superase inhibitor (Invitrogen) and 0.1%v/v IGEPAL CA-630 (Sigma Aldrich)] with 5μL of slide-specific hashing oligo at 10μM and 5μL of 500μM stock DAPI, was pipetted gently onto the tissue section and across the long edge of the embryo slide nearest the user using an wide bore p1000 tip. A space-grid transfer slide was then positioned (agarose surface facing the tissue section) so that the arrayed oligos were aligned between the two transfer clips and spanned the tissue section’s extent. Placement of the space-grid slide was achieved by tilting the slide so that its long edge nearest the user contacted the edge of the tissue section slide and fit under the overhanging fastening teeth of the transfer clips. The space-grid slide was then rocked toward the embryo section slide until the two slides were face-to-face with the tissue section contacting the space-grid oligo array-laden agarose membrane. Excess buffer was allowed to wick into a laboratory wipe. When stacked, the slides snapped into the transfer clips and were thereby securely held together during transfer. The slide stack was moved to the microscope stage and the entire embryo section was imaged in GFP and DAPI channels. The transfer slide was then removed from the transfer clips and separated from the tissue.

Cells of the embryo section were then scraped using a cell scraper (Fisherbrand, GDPC240) from the slide into a 4% paraformaldehyde fixing solution. At this point, Slides 7-14 were subjected to sonication using the bioruptor sonicator (Diagenode, product #B01020001). This extra round of sonication aided in disassociating more nuclei. After fixation for 15 minutes on ice, cells were spun down in 1.5mL tubes in a chilled benchtop centrifuge at 800g for 10 minutes. The supernatant in each tube was removed and cells were pooled in 1mL of NSB [Nuclei Buffer (10mM Tris/HCl pH 7.4, 10mM NaCl, 3mM MgCl2) with 1% v/v superase inhibitor (Invitrogen) and 1% v/v BSA (New England Biolabs)], flash frozen and stored at -80C.

Single cell RNA sequencing library preparation

Protocols.io - Single cell RNA sequencing library preparation (2-level sci-RNA-seq) (38)

Frozen nuclei were thawed over ice and spun down at 800g for 8 minutes in a 15mL conical tube. Cells were then permeabilized in 400μL of permeabilization buffer (NSB + 0.25% Triton-X) for 3 minutes and
then spun down at 800g for 8 minutes. Following resuspension in 500µL NSB, two-level sci-RNA-seq libraries were prepared as previously described (16, 39). Briefly, nuclei were first sonicated for 12 seconds using the bioruptor sonicator on the low setting. This caused disruption of many nucleus aggregates that had formed. Cell counts were then obtained by staining nuclei with 0.4 % trypan blue (Sigma-Aldrich) and counted using a hemocytometer.

5000 nuclei in 2µL of NSB and 0.25µL of 10mM dNTP mix (Thermo Fisher Scientific, R0193) were then distributed onto a skirted twin.tec 96 well LoBind plate (Fisher Scientific, 0030129512). Next, 1µL of uniquely indexed oligo-dT at 25µM (39) was added to every well and mixed. This 96-well plate was then incubated at 55°C for 5 minutes and then immediately placed on ice. Next, 1.75µL of reverse transcription mix (1µL of Superscript IV first-strand buffer, 0.25µL of 100 mM DTT, 0.25µL of Superscript IV and 0.25µL of RNAseOUT recombinant ribonuclease inhibitor) was then added to every well and mixed. Plates containing the reverse transcription reactions were then incubated using a ramping reverse transcription protocol on a thermocycler:

1. 10°C for 2 minutes
2. 20°C for 2 minutes
3. 30°C for 2 minutes
4. 40°C for 2 minutes
5. 50°C for 2 minutes
6. 55°C for 15 minutes
7. 4°C forever

Wells were pooled and nuclei were transferred to a flow cytometry tube through a 0.35 µm filter cap and DAPI added to a final concentration of 3 µM. At this point pooled nuclei were seeded using one of 2 methods; (1) sorted on a BD FACS Aria II cell sorter (Slides 1-6) or (2) diluted (Slides 7-14).

Sorted cells from Slides 1-6 were prepared using 8 RT plates. After calculating the collision rate using the birthday problem calculator (40), we sorted 200 nuclei per well into 96-well LoBind plates with each well containing 5µL of EB buffer (Qiagen) and 0.75µL of second strand mix (0.5 µL of mRNA second strand synthesis buffer and 0.25µL of mRNA second strand synthesis enzyme, New England Biolabs).

Diluted nuclei originating from slides 7-14 were prepared using 10 RT plates were first diluted to 50 nuclei per µL in NSB. Diluted nuclei were then premixed with second strand-synthesis reagents (0.5 µL of mRNA second strand synthesis buffer per 5µL of nuclei suspension and 0.25µL of mRNA second strand synthesis enzyme per 5µL of nuclei suspension). 5.75µL of this nucleus/second-strand synthesis solution was dispensed into 96-well LoBind plates to seed 250 nuclei per well.

Second strand synthesis performed at 16°C for 150 minutes. Tagmentation was then performed by addition of 5.75µL of tagmentation mix per well (0.01µL of a custom n7-loaded Tn5 enzyme in 5.74µL 2x Nextera TD buffer, Illumina) and plates incubated for 5 minutes at 55°C. This reaction was terminated by addition of 12µL of DNA binding buffer (Zymo, D4004-1-L) and incubated for 5 minutes at room temperature. 36µL of Ampure XP beads (Beckman Coulter, A63880) were added to every well, DNA purified using the standard Ampure XP protocol eluting with 17µL of EB buffer. 16µL of this eluate was then transferred to a new 96 well LoBind plate for index PCR.
For PCR, 2µL of 10µM indexed P5, 2µL of 10µM indexed P7 (39) and 20µL of NEBNext High-Fidelity master mix (New England Biolabs, M0541L) were added to 16 uL of eluted template DNA. PCR indices were arrayed such that each well contained a unique combination of P5 and P7 barcodes. PCR was then performed using the following program:

1. 75C° for 3 minutes
2. 98C° for 30 seconds
3. 98C° for 10 seconds,
4. 66C° for 30 seconds
5. 72C° for 1 minute
6. Return to Step 3 for 17 times
7. 72C° for 5 minutes.
8. 4C° forever

After PCR, all wells were pooled, concentrated using a DNA clean and concentrator kit (Zymo Research, D4033) and purified via a 0.8X Ampure XP cleanup. Final library concentrations were determined by Qubit (Invitrogen), libraries visualized using a TapeStation D1000 DNA Screen tape (Agilent, 5067-5582) and libraries sequenced on a Nextseq 500 using a high output 75 cycle kit (Illumina, 20024906). Libraries were denatured and diluted to 2pM. Sequencing was performed using onboard primers using the following sequencing recipe:

Read 1: 18 cycles (8 base-pair UMI and 10 base-pair RT barcode
Index 1: 10 cycles (10 base-pair PCR index1)
Index 2: 10 cycles (10 base-pair PCR index2)
Read 2: 52 cycles (52 bases of transcript or hash-oligo)

Pre-processing of sequencing data
Sequencing data was processed as described previously (16). Briefly, sequencing runs were first demultiplexed using bcl2fastq v.2.19. Only barcodes that matched reverse transcription indices within an edit distance of 2 bp were retained. Following assignment of indices, polyA tails were trimmed using trim-galore (https://github.com/FelixKrueger/TrimGalore), and reads were mapped to a mouse transcriptome (mm-10), human-mouse transcriptome (hg-38 and mm-10) using the STAR aligner. Following alignment, reads were filtered for alignment quality, and duplicates were removed. Reads were considered duplicates if they (1) mapped to the same gene, (2) mapped to the same cell barcode and (3) contained the same unique molecular identifier (UMI). Reads that met the first two criteria, and differed by an edit distance of 1 from a previously observed UMI were also marked as duplicates and discarded. Non-duplicate reads were assigned to genes using bedtools to intersect with an annotated gene model. All 3’ UTRs in the gene model were extended by 100 bp to account for the possibility that some gene 3’ UTR annotations may be too short, causing genic reads to improperly be annotated as intergenic. Cell barcodes were considered to correspond to a bona fide cell if the number of unique reads associated with the barcode was greater than an interactively defined threshold on a knee plot. Reads from cells that passed this UMI count threshold were first aggregated into a sparse matrix format and then loaded and saved as a CDS object for analysis with Monocle3 or Seurat.
**Slide registration**

To map cell locations to an image of the embryo, fluorescent SYBR green spots with known positions are used to orient cells. Images of the hash array with fluorescent SYBR green spots on top of the DAPI-stained embryo section were taken with a 2.5x magnification (Zeiss Observer Z1 Microscope).

Prior to the transfer of the image of Slide 12, the Slide 12’s image file was accidentally erased and could not be recovered. The sequenced cells mapping to Slide 12 are included in the dataset with relative spatial positions based on the recovered hash oligos. However, for this slide we were unable to perform any analyses that relied on segmentation or slide registration.

The captured images were then used to orient the hash array to the embryo section. More specifically, co-registration of the imaged embryo sections and the oligo hash tagged transcriptomes was achieved through alignment of the SYBR green waypoints imaged during transfer to their position within an ideal space-grid layout. Coordinates for SYBR green spots imaged during oligo hash transfer to the embryo section and the corresponding coordinates in an image of an ideal space-grid were obtained in Fiji image processing software (24) using the Big Warp function of the BigDataViewer plugin (25). An affine matrix was computed using the coordinates as source (embryo image) and target (space-grid image) control points in the AffineTransformation function in the “vec2dtransf” and “imager” packages in R. The matrix was applied to the embryo section image. Sequenced nuclei were then mapped to the aligned space of the transformed image and space-grid using their space-grid hashes.

The following formula was used to calculate the number of microns per pixel and thereby estimate the size of each spot on the hash array.

\[
\text{microns per 1 pixel} = \left( \frac{\text{native camera pixel size}}{\text{objective / camera adaptor}} \right)
\]

One pixel was equal to 1.816 microns based on the camera pixel size of 4.54 for a Zeiss Axiocam 503 Mono Camera, an objective size of 2.5x and a camera adaptor size of 1.

**Assigning spatial labels from hash reads**

Reads from hash oligos were demultiplexed as described previously (16). Briefly, demultiplexed reads that matched combinatorial indexing barcodes were examined to identify hash reads. Reads were considered hash reads when they met two criteria: 1) the first 10 bp of read 2 matched a hash barcode in the experiment within an edit distance of two; and 2) contained a polyA track between base pairs 12 to 16 of read 2. These reads were then deduplicated by cell barcode and collapsed by UMIs to create a vector of hash oligo UMI counts for each nucleus in the experiment.

To assign each nucleus to the slide from which it came, we tested whether its sci-RNA-seq library was enriched for a particular hash barcode. We compared a nucleus’s hash UMIs against a ‘background distribution’, which under ideal circumstances, would be random and uniformly distributed. To estimate the background distribution, we simply aggregated the hash UMIs from cell indices for which fewer than 10 mRNA UMIs were collected, reasoning that these reflect library contributions from ambient reverse
transcriptase products, debris fragments, etc. We then compared the hash UMIs for nucleus to this background by a chi-squared test. After correcting the resulting p-values were corrected for multiple testing by the Benjamini-Hochberg procedure, we rejected the null hypothesis that originates from the background distribution at the specified FDR (5% FDR was used in this study). Those nuclei with hash counts deemed greater than background were also evaluated for enrichment for a single hash sequence. Enrichment ratios were calculated as the UMI count ratio of the most abundant vs. the second most abundant hash oligo. Specifically, if the UMI count for the most abundant hash in a nucleus is $\alpha$-fold higher than the second most abundant, is marked as a singleton. $\alpha$ was set to 5, which corresponded to a nadir between two modal outcomes -- separated unlabeled cells and singularly labeled cells. Cells that fell below 5-fold enrichment of a unique hash oligo were flagged as a multiplet or debris and discarded.

A cell’s spatial position within the grid consists of a specific combination of two oligonucleotides, a spot oligo and a sector oligo. To find a cell's position within the grid, we first took a single cell’s vector of spot counts and mapped these counts to their position in the sci-Space grid. Then we performed a 3x3 gaussian convolution on every position, allowing us to account for and integrate spatially local signal from neighboring positions and simultaneously reduce spurious background. Next, we performed an element-wise multiplication of convolved spot values by the matrix of sector counts measured for that cell. The product of spot and sector oligos were then ranked and a cell was mapped to the top ranking combination which matched two criteria: (1) the combination represented a valid pairing and (2) the combination mapped within the boundary of the imaged embryo. This boundary was determined by manually segmenting the outline of the DAPI stained image of the embryo.

Upon manual inspection of these draft spatial positions, we noticed a minority of nuclei which were clearly mismapped (e.g. cardiomyocytes outside of the heart). This seemed to occur through a nucleus’ absorption of a neighboring sector oligo or alternatively via barcode collision between cellular debris carrying a spatial index and the sequenced nucleus. To correct these misassignments, each slide was manually segmented into regions of interest (ROI) where a cell type was focally concentrated. Next we asked whether a nucleus of a given cell type had a plausible alternate mapping within these ROIs. A nucleus was moved from its draft position if the nucleus’ highest convolved spot value within the region of interest was within 5-fold of its highest spot value. The value of 5-fold was chosen by examining the distribution of all alternate mappings. This procedure was then repeated for each cluster using an automated algorithm to detect high density regions of clusters within a slide. Through this process, the spatial calls of 9.1% nuclei were remapped. Finally, we removed cardiomyocytes and hepatocytes that mapped outside of the heart and the liver, respectively, resulting in the removal of 369 nuclei.

**Estimating nuclei counts from embryo images**

Using python, the 2.5x magnification embryo DAPI stained images were preprocessed using a white top-hat transform followed by a histogram equalization to reduce uneven lighting and increase contrast respectively. The images were then thresholded using Otsu's method. In order to overcome the challenge of counting individual cells in nuclei clusters, the resulting binary masks were separated into ‘dense’ and ‘sparse’ nuclei masks using a connected components algorithm. The dense nuclei masks were used to isolate nuclei clusters in the original embryo images, which were then thresholded to a secondary value defined as Otsu’s value plus a constant intensity shift. The sparse and the dense nuclei masks were then distance transformed, using Euclidean geometry, to generate distance maps. A
peak finding algorithm was used to isolate the centroids of peaks in the distance maps and resulting unique centroids were counted as nuclei.

**Cell type classification**

Nearest neighbor classification was performed by aligning the cells from this study to cells from E13.5 time point from the MOCA single cell dataset (1). This MOCA dataset was chosen because it was prepared using nuclear sci-RNA-seq. It is our experience that alignment between datasets produced using the same technology are less sensitive to hyperparameter selection during alignment. The E13.5 time point was chosen because this time point most closely matched the E14.0 timepoint sequenced in this study.

Count matrices from the two datasets were subsetted for genes found in both datasets and then combined. The E13.5 time point was then downsampled and the two datasets were aligned using Seurat v3 (20, 28) dataset integration using reciprocal PCA. For each cell in the E14.0 time point the 10 nearest neighbors in UMAP space from the MOCA dataset were recorded. Each cell was then assigned the majority nearest neighbor label. Finally, to remove poor confidence cell type labels, the E14.0 data was clustered and cell type labels that did not account for more than 5% of a cluster were assigned “Unknown”. Garnett classification (22) was performed using a marker-free classifier trained on the E13.5 time point from the MOCA dataset. This classifier was then applied to the cells sequenced in this study.

This same process was repeated to match neurons from the developing mouse brain atlas dataset (DMBA) to neurons in the sci-Space dataset(21). Briefly, neurons and radial glia from the E13.5, E14.0 and E14.5 timepoints were used to perform nearest neighbor alignment using reciprocal PCA (20, 28). The majority label from a cell’s 5 nearest neighbors was used to transfer a number of different labels including cell type, anatomical dissection, age, and UMAP cluster in the original dataset. Finally, the different inferred labels were collated along with the top differential genes marking each cluster. The final annotation set consists of a combination of cell types transferred from the MOCA and DMBA 3938 datasets along with manual annotation matching a cluster’s differential gene expression.

**Immunostaining and adjacent image alignment**

Before immunostaining serial sections adjacent to sequenced sections were fixed in 4% paraformaldehyde (Electron Microscopy Sciences) in phosphate buffered saline (PBS, ThermoFisher) for three minutes at room temperature. Sections were then washed for five minutes three times in PBS-T (0.1% Tween-20, VWR), permeabilized for 10 minutes at room temperature in 0.1% Triton X-100 (VWR) in PBS, and washed for five minutes three times in PBS. Autofluorescence was quenched using TrueBlack Lipofuscin (Biotium) according to the manufacturer’s protocol. Briefly, sections were treated with TrueBlack Lipofuscin diluted 20X in 70% ethanol with a 30 second to three minute incubation, then washed for five minutes three times with PBS. Sections were next blocked with 2.5% normal donkey serum (NDS, Jackson ImmunoResearch Laboratories) in PBS for one hour at room temperature. Primary antibodies were applied in PBS containing 2.5% NDS as indicated in **Table S1** with an overnight incubation at 4°C. The following day sections were washed for five minutes three times in 2.5% NDS in PBS, then incubated for one hour at room temperature with Hoechst 33342, Trihydrochloride, Trihydrate (Invitrogen) counterstain and secondary antibodies diluted as indicated in
Table S1 in 2.5% NDS in PBS. Sections were next washed again in 2.5% NDS in PBS and then coverslipped with Fluoromount-G Mounting Medium (SouthernBiotech) prior to imaging.

Stained sections were imaged using a Ti-E inverted microscope (Nikon) and multi-field images were stitched in the NIS-Elements (Nikon) software. Each channel of the triple stained sections together with the counterstain was aligned to the DAPI image of the sequenced section using the StackReg plugin in Fiji (rigid body followed by affine). Each aligned channel of the stained section images were then separated from the counterstain and overlaid onto the DAPI channel image of the adjacent sequenced section.

Anatomical annotation and segmentation
Annotation was performed using The Atlas of Mouse Development (41) in conjunction with magnetic resonance images of the E14.5 embryo (42). Annotations were then confirmed using immunostained adjacent sections (when available). Anatomical segmentation was performed manually using the DAPI-stained embryo section and the Big Warp function of the BigDataViewer plugin in Fiji. The region of interest was demarcated by choosing a bounding set of points in clockwise or counter-clockwise order. These points were then used to construct a polygon using the spatial features package in R. These polygons were then scaled using the same affine transformation used for slide registration to put them on the same coordinate axis. For polygons with holes, the contour of the entire image was first segmented followed by segmentation of each cavity. This same process was repeated to segment and annotate the brain regions. However, the Allen Institute’s Anatomical Reference Brain Atlas (www.atlas.brain-map.org) was used as a guide to annotate the Pallium, Sub Pallium, Midbrain, Hindbrain, Thalamus and Hypothalamus of each brain.

Tissue domains
Tissue domains with similar cell type composition were identified using Giotto (26), an unsupervised tool for single-cell spatial expression analysis. The top 500 spatially autocorrelated genes (Moran’s I > 0.05, FDR < 0.001, expressed in at least 1% of cells) were identified from all slides. For each slide, spatial domains were identified using the Hidden Markov random field model with parameters k=50, beta=10. To identify consensus spatial domains that existed on multiple slides, UMAP dimensionality reduction was performed based on the absolute numbers of annotated cell-types in each domain from each slide. Community detection with louvain clustering (k=5) identified 22 clusters representing the 22 tissue domains.

Analysis of aggregated spatial positions
Nuclei mapping to the same spatial grid position within each slide were aggregated by summing gene expression counts from each nucleus. These spatial positions were then treated as the columns of a gene by position count matrix. This matrix was converted into a Monocle3 CDS object, and used as input for PCA, UMAP and louvain clustering.

To assess the ability of integration to recover a cell’s spatial position, we used the Seurat package (version 3.2.2) and applied this analysis to Slide 14. First, nuclei mapping to a spatial position were aggregated and used to create a Seurat object bearing the spatial coordinates of each position. These spatial positions were then processed as described by the Seurat spatial vignette. Briefly, this involved running SCTTransform, PCA, UMAP and clustering using the default parameters. The nuclei used to
aggregate each of these positions (nuclei from Slide 14) were also processed in the same way (SCTransform, PCA, UMAP and clustering). To integrate the two datasets we used the FindTransferAnchors() function in Seurat to find anchors between the two datasets using the SCTransform as the normalization method. Finally, these anchors were used as an input to the TransferData() function to return a set of predicted spatial positions for each cell. To determine the error associated between data integration and the pre-aggregation ground truth, a cell was assigned to the position with the highest transfer probability returned by TransferData() and the euclidean distance was calculated between the most probable transferred position and the ground truth position.

**Kriging gene expression**
Spatial grid positions were first collapsed such that non-overlapping sets of 4 adjacent positions were collapsed into a single spatial position. Each cell type within these spatial bins was then aggregated by summing the counts for each gene contributed by that cell type. These values were then kriged with the automap package in R using ordinary kriging via the autoKrig() function. Interpolated values were then rescaled to reflect the percentage of gene expression contributed by a cell type at each given position. Finally, a polygon object specific to each slide was used to clip the interpolated gene expression values.

**Spatial autocorrelation analysis**
Gene spatial autocorrelation was computed by first subsetting cell types for which there were more than 100 cells present on a slide. After setting a random seed, we estimated size factors, performed PCA and UMAP dimensionality reduction with a fixed set of parameters. UMAP was run using uwot’s implementation in R with the flag fast_sgd set to false. This ensured that UMAP dimensionality reductions were consistent between runs. Following UMAP, for each subset of cells, a gene’s spatial autocorrelation was computed using either its cell’s spatial coordinates or UMAP coordinates as the input into Monocle3’s graph_test() function. The resulting test statistic was corrected for multiple testing and genes with an FDR < 0.01 and a Moran’s I test statistic greater than 0.05 were reported as having statistically significant spatial autocorrelation.

**RNA Fluorescence in situ Hybridization (FISH) and analysis**
To validate spatial expression patterns across cell types identified in our dataset, an 8-probe RNAscope® HiPlex kit (Advanced Cell Diagnostics, Inc.) including probes against 6 transcripts to mark various cell populations: Gad2 -- GABAergic neurons; Slc17a7 -- VGlut1+ glutaminergic neurons; Slc17a6 -- VGlut2+ glutaminergic neurons; Pax6 -- radial glia; Lum -- fibroblasts; and finally, Cldn5 -- endothelial cells. The kit additionally assayed for Cyp26b1 and Hoxa10 to demonstrate their spatial localization within multiple cell types and neuronal cell subtypes as indicated by our sciSpace data. Briefly, serial sections near Slide 1 of the sci-Space dataset were assayed according to the manufacturer’s protocol. The fresh-frozen tissue sections were fixed using 4% paraformaldehyde (Electron Microscopy Sciences) in 1X PBS, dehydrated, and treated with the Protease IV kit component. Following the manufacturer’s specified hybridization steps, counterstaining, and coverslipping, the first four probes: 1) Cyp26b1, 2) Gad2, 3) Slc17a6, 4) Hoxa10 were imaged. Scans were obtained using an Aperio VERSA slide scanner (Leica Biosoftware) with 40x magnification and DAPI, FITC, Cy3, Cy5, and Cy7 filter sets. Coverslips were removed, the first four fluorophores were cleaved, the fluorophores for probes 5-8 were hybridized: 5) Pax6, 6) Slc17a7, 7) Lum, 8) Cldn5. Slides were re-coverslipped and imaged as before.
Slide scans were analyzed using QuPath 0.2.3 quantitative pathology and bioimage analysis freeware. Briefly, the two scans for each slide were imported and affine transform matrices for alignment were obtained using: Analyze/Interactive image alignment/autoalign/Estimate transform; with settings: Registration type -> Affine transform, Alignment type -> Image intensity, Pixel size -> 20. Positive signal foci were identified using Analyze/Cell detection/Cell detection with the setting parameters designated in Table S2. Parameters were manually adjusted to allow for detection of multiple foci in clusters with a bias toward avoiding false positive detection events. Coordinates of detected foci in each channel were output as .tsv files using the Counting function (Counting/Convert detections to points, followed by Counting/Save points).

Because the assay lacked a counterstain that could be used to delineate cell boundaries, we relied on transcript proximity to designate positions positive for expression and coexpression. Coordinates were analyzed in R to identify transcript locations within putative expressing cells and to estimate the positions of coexpression between transcripts. Briefly, foci located within an approximate sub-cellular length (10 um) of two other foci for the same probe were considered transcripts in an expressing cell, i.e. were designated as positive points of expression. To detect coexpression a similar analysis was performed where a positive marker gene (Gad2, Slc17a6, Pax6, Slc17a7, Lum, or Cldn5) position was considered to be coexpressed with Cyp26b1 when it lied within 10 um of two positive positions for Cyp26b1. For Hoxa10, which was more lowly expressed than the other assayed genes, a positive designation for expression was given when at least one other Hoxa10 position fell within 10 um of a focus, and a coexpression designation required one positive Hoxa10 position within 10um of a given positive marker gene position. Expression and coexpression mapping was compared between the RNAscope and sci-Space datasets.

**Variance decomposition model**

The variance of gene expression within and between cell populations was computed using the angular distance metric. Specifically, let $y = \text{the vector of log-scaled gene expression levels for each gene in a cell}$, normalized to be of unit magnitude, let $E[Y] = \text{the arithmetic mean of log-scaled gene expression levels normalized to be of unit magnitude}$, and let $Y = \text{the empirical distribution of y across all cells in a given analysis}$. Then:

$$\text{Var}(Y) := E[\text{angular\_distance}(Y - E[Y])^2]$$

where $\text{angular\_distance}(a, b) =$

$$(2/\pi) \cos^{-1}((A \cdot B) / (\|A\| \|B\|))$$

This variance statistic behaves similarly to the variance of a univariate distribution. Most importantly, it obeys the Law of Total Variance (fig. S28C). If cells sampled from $Y$ are partitioned into groups $X$, then:

$$\text{Var}(Y) = E[\text{Var}(Y | X)] + \text{Var}(E[Y | X])$$

$= \text{weighted average of within-group variance + variance of averages among groups}$
The “variance explained” by a grouping \( X \), e.g. grouping cells by cell type, can therefore be computed as:

\[
\text{Variance explained} = 1 - \frac{\text{within-group var.}}{\text{total var.}}
\]

\[
= 1 - \frac{E[\text{Var}(Y \mid X)]}{\text{Var}(Y)}
\]

This naive formula for variance explained is strongly affected by the number of groups in the grouping \( X \), similar to how the variance explained in a regression model is affected by the number of degrees of freedom in the model. If one divides a population of cells into random groups of two, variance explained will be high, even though the grouping has no biological meaning. We can correct for this by using an adjusted formula:

\[
\text{Variance explained} = 1 - \frac{E[\text{Var}(Y \mid X)]}{E[\text{Var}(Y \mid \text{permuted } X)]}
\]

where in the denominator, the group id associated with each cell is randomly permuted.

This adjusted formula corrects for the number of groups in \( X \), but is confounded by another factor. The observed variance between cell gene expression vectors is a result of both biological heterogeneity and technical factors such as the sparsity of the single cell RNA-seq data. If the same biological sample is profiled in two different experiments, and the median number of UMIs per cell is \( N \) in experiment 1 and \( 4N \) in experiment 2, then the variance explained by the grouping \( X = \text{cell type} \) will be lower in experiment 1 vs. 2 due to increased sparsity. To correct for sparsity and estimate the proportion of biological variance explained by grouping \( X \), we use a final adjusted formula:

\[
\text{Variance explained} = \frac{E[\text{Var}(Y \mid X)] - E[\text{Var}(\text{resampled } Y \mid X)]}{E[\text{Var}(Y \mid \text{permuted } X)]-E[\text{Var}(\text{resampled } Y \mid \text{permuted } X)]}
\]

In this formula, the “resampled \( Y \)” terms are computed by replacing each cell’s gene expression vector \( y \) with a sample from the distribution \( \text{Multinomial} \). The distribution of the multinomial is parameterized by the two parameters, \( n \) the number of draws and \( p \) a vector of probabilities \( (p_1, p_2, \ldots, p_i) \) or the probability of drawing each gene. For each simulated cell from a given group, \( n \) is set to match the number of UMIs measured from a given cell and \( p \) is calculated by dividing the number of UMIs for a
given gene in that cell grouping, by the total number of UMI$s from that group. Resampling in this manner is a way of estimating what variance one would observe if there were no biological heterogeneity in a group of cells and the only source of observed heterogeneity was sparsity.

**fig. S29** applies this formula to estimate the proportion of biological gene expression variance explained by cell type; the proportion of biological gene expression variance explained by cluster; the proportion of biological gene expression variance explained by spatial position, using the grouping \( X = \) a spatial spatial bin (non-overlapping 2*2 spot squares); the proportion of biological gene expression variance explained by spatial position and cell type where \( X = \) the combination of spatial bin and cell type. **Fig. 4B** applies this formula to estimate the proportion of the residual biological gene expression variance, after accounting for cell type, that is explained by spatial position. In these models, the grouping \( X = \) the tuple (cell type, spot id), and permuted \( X = \) the tuple (cell type, permuted spot id). By permuting the spot id but not the cell type annotation, we ensure that our estimate is of the variance explained by the cell type + space model relative to the cell-type-only model, rather than relative to a null model. To estimate the variance that can be explained by a null (shuffled) model, we perform this permutation procedure 50 times and take the average of these trials.

**Pairwise angular distance**
Only cell types with 100 cells or more, originating from a single slide were considered for this analysis. Each cell was size factor normalized, and scaled to the unit hypersphere. The pairwise angular distance was then calculated as detailed above. To test whether there was a relationship between angular distance and physical distance, a linear model was fit with angular distance as the response and physical distance as the sole predictor (angular.distance ~ distance). Reported p-values indicate the significance of coefficient for the physical distance predictor variable using the Wald linear regression test.

**Spatial gene modules**
Gene module analysis in **fig. S30** comparing cell type derived gene modules and spatial gene modules was performed on Slide 14. Briefly, we performed PCA, UMAP dimensionality reduction and clustering on nuclei mapping to Slide 14. Genes autocorrelated in the UMAP embedding were then calculated using Monocle3’s `graph_test()` function. Similarly, genes autocorrelated in spatial-position were calculated using the sci-Space derived spatial coordinates as the input into Monocle3’s `graph_test()` function. We performed module analysis on the union of genes with significant autocorrelation in the UMAP embedding and spatial position (FDR < 0.05). UMAP gene modules were recovered using Monocle3’s `find_gene_modules()` function on the cell*gene matrix. Spatial gene modules were recovered using Monocle3’s `find_gene_modules()` on the position*gene matrix. Gene expression from each set of discovered modules was then aggregated using Monocle 3’s `aggregate_gene_expression()` function and visualized as row- and column-clustered heatmaps.

**Neuronal trajectory analysis**
Neuronal trajectories mapping to the brain and classified as either Radial Glia, Neurons or Glial Cells were subset from the neural lineage partition. PCA, followed by UMAP dimensionality reduction was performed on this subset. Monocle3 was then used to learn a principle graph on the UMAP embedding. For pseudotime inference, the root of the trajectory was selected by manually identifying principal graph
nodes occupied by radial glia. Each trajectory was then scaled for display purposes. These nuclei were then mapped to brain regions based on segmentation performed using the Allen Institute’s Anatomical Reference Brain Atlas (www.atlas.brain-map.org). To calculate pseudotime dependent genes, nuclei mapping to the Pallium, Sub Pallium and Midbrain were first subsetted. For each brain region and the trajectory contained within it, pseudotime dependent genes were then recovered by fitting a natural spline with three degrees of freedom using pseudotime as the predictor. Genes with at least one significant knot (FDR < 0.01) were deemed pseudotime dependent. This process was repeated for all three trajectories. The results of these tests are provided in File S3. The intersection of significant genes across all three trajectories is shown in Fig. 5F. For gene ontology analysis, the dendrogram producing the row-clustering in Fig. 5F was cut to produce 4 groups. Each group was then provided as input into http://geneontology.org. Displayed gene ontology terms were chosen from a list of significant terms (FDR < 0.01).
Supplementary Figures

Supplementary Figure 1. Comparison of methods for spatial transcriptomics. Schematics and key attributes of the two major classes of contemporary spatial transcriptomics methods, as well as sci-Space. (A) Spatial transcriptome capture (STC) methods, e.g. the original “spatial transcriptomics” method (11) and Slide-seq (12). (B) Methods relying on multiplexed mRNA imaging, e.g. MERFISH (13), seqFISH (14), and FISSEQ (15). (C) sci-Space.
Supplementary Figure 2. Space-grid design. (A) Schematic of spotted oligos with SYBR green fluorescent dye marked positions labeled in green. All positions contain a combination of location-informative “hashing” sequences (single stranded DNA) with polyA tails. (B) Overlay of brightfield and fluorescence image of the same position. (C) Average radius and spot-to-spot distance computed from imaged slides. (D) Diagram of hierarchical barcoding approach where each position is marked by a unique combination of one of 16 sector barcodes (colors) and one of 1536 spot sequences. (E) An example displaying a single spot oligo barcode (white square) which is in 5 different sectors. Scale bar (B) = 0.1 mm.
Supplementary Figure 3. SYBR green waypoints transfer to DAPI stained embryo. (A) Permeabilized mouse embryo section receives (B) SYBR green waypoints spotted at a single section. (C) The resulting transfer and imaging shows the location of each waypoint on the DAPI stained section. Dashed white lines denote the approximate location of each sector. Scale bar in panel (C) = 0.5 mm.
Supplementary Figure 4. Labeling of cryosectioned tissues with hash oligos from an agarose coated slide is compatible with sci-RNA-seq. Slides from sections of the developing mouse embryo were first labeled with a slide specific oligo and then labeled with another hash oligo from a space-grid containing a single hash oligo at varying concentrations. Replicates are independent experiments performed on different days using tissue sections from a single batch. **(A)** RNA UMIs recovered per cell across stages and replicates. **(B)** Number of cells sampled from each slide across stages and replicates. **(C-E)** UMAP embedding colored by (C) embryonic stage, (D) replicate or (E) expression of skeletal muscle marker Titin (Ttn). **(F)** Correlation of RNA UMIs recovered per gene between replicates at different stages. **(G)** Hash UMIs recovered per cell of oligo spotted at 10μM, 20μM, 25μM and 50μM concentrations.
Supplementary Figure 5. Spotted space-grids are reproducible. (A) HEK293T nuclei were exposed to hash-oligos dissolved from one of 3 space-grids. (B) Correlation between spot oligo counts originating from different slides. (C) Distribution of sector oligos observed per cell, broken out by replicate. (D) Cartoon depicting control experiment with human and mouse cell lines grown on a glass slide and barcoded with a single hash-oligo. (E) Scatter plot depicting the number of human (X axis) or mouse (Y axis) unique molecular identifiers (UMIs) detected per cell. Nuclei were filtered for those mapping to a single slide. (F) The percentage of contaminating UMIs for each cell type.
Supplementary Figure 6. Configuration for transfer of space-grid oligos to the nuclei of fresh frozen cryosections. (A) Spatially indexed slides, “space-grids,” were fabricated by spotting unique combinations of hashing oligos onto agarose membrane-coated slides. (B) Permeabilized fresh-frozen tissue sections (C) received the spatially-defined pattern of oligos by diffusion from the space-grids when the oligo-laden agarose and tissue section were sandwiched together between their carrier slides.
Supplementary Figure 7: sci-Space workflow for sequencing library preparation and demultiplexing transcripts allows transcripts and spatial positions to be assigned to individual nuclei. (A) Hashing oligos or barcodes are transferred to nuclei as determined by nuclei positions relative to the barcode array. (B) Nuclei from each slide are dissociated and labelled with an additional slide-specific barcode. (C) Transcripts and barcodes are tagged with nuclei-specific indices according to the sci-RNA-seq protocol (39). Note that the green segment shown in the schematic includes combinatorial indexing barcodes introduced during sci-RNA-seq. (D) Barcodes and transcripts from all nuclei are pooled and sequenced. (E) Indices are used to demultiplex transcripts and barcodes, which allow for the assignment of each nucleus to its slide, sector, and spot of origin.
Supplementary Figure 8. Cellular hashing distinguishes low RNA UMI nuclei from aggregates and uniquely marks a nucleus’ position. (A) Cumulative distribution of RNA UMIs (unique molecular identifiers) identified after sequencing. Red line corresponds to a lenient cutoff used for initial calls. (B) Violin density plots displaying the distribution of cells which were labeled by single slide-specific hash oligo. Cells which failed to show enrichment of a single slide-specific hash oligo were filtered out. Red line corresponds to the same cutoff displayed in panel (A). (C-F) Heatmaps displaying the average hash oligo counts from the top spot and surrounding positions for 1000 randomly sampled cells from Slide 1 (C), Slide 4 (D), Slide 13 (E) and Slide 14 (F). Histograms reflect the marginal sums for rows or columns.
Supplementary Figure 9. sci-Space sequenced cells have complex transcriptomes and separate into the major cell types. (A) Boxplots displaying the number of unique molecular identifiers recovered (red) and genes detected (blue) for cells from each slide. (B-C) Louvain clustering result and the corresponding proportions of cells found in each cluster. (D-E) UMAP embedding colored by the slide of origin (D) or embryo of origin (E).
Supplementary Figure 10. Comparison of recovery of nuclei from sequencing and estimated nuclei present upon imaging. (A) Estimates of DAPI stained nuclei present per spot. Per spot estimates for Slide 3 are shown as an example. (B) Comparison of the total number of estimated nuclei present versus the number of nuclei recovered from each slide. Method for seeding cells during sci-RNA-seq is noted below the x-axis. An estimate of nuclei count from slides 5 and 12 could not be computed due to corruption or loss of the image file.
Supplementary Figure 11. Automated cell type annotation and concordance between methods. (A) UMAP embedding of a single cell alignment (20, 28) between the E13.5, E14.0 and E14.5 timepoints of the developing mouse brain dataset (21) and neural sci-Space lineages. (B) Concordance between Garnett annotations (22) trained on published annotated data (1) and labels from k-nearest neighbor transfer between the same dataset and this study. Heatmap is colored to depict the percent of cells in a Garnett annotation (rows) that have the same label as the nearest neighbor annotation (columns).
Supplementary Figure 12. DAPI stained images of sci-space sequenced slides. The slide number and the embryo from which they originate are displayed above each image. The image for Slide 12 was lost during data transfer. SYBR green point layouts varied between space-grid prints, and positioning of the sections relative to the oligo array also varied. Scale bars = 0.5 mm.
Supplementary Figure 13. Co-registration procedure of imaged section and space-grid. (A) DAPI stained E14.0 section (Slide 3) with SYBR green points imaged in the GFP-channel. Matched SYBR green waypoints between the image and (B) the intended SYBR pattern on an ideal space-grid are used to calculate an affine-transformation. (C) Co-registered imaging data with inferred positions overlayed with image with inset highlighted (D). Scale bars in panels (A) and (C) = 1 mm.
Supplementary Figure 14. Spatial distribution of nuclei per slide in addition to key spatial metrics. Histograms displaying the number of nuclei recovered per position for each slide are shown next to a heatmap of the cells recovered per spatial position. In text, above and beside each histogram, summary spatial statistics are noted for each slide. An asterisk is displayed next to slides that were processed using the optimized sci-Space protocol with an extra sonication step and nucleus seeding via dilution (as opposed to FACS).
Supplementary Figure 15. Segmented organs from each embryo section. Segmented regions are highlighted on each section and colored according to the denoted anatomical structure. 500um scale bar shown in the bottom right corner of each image.
Supplementary Figure 16. Alignment of immunostaining from adjacent cryosections. (A-D) Serial cryosections adjacent to the sequenced embryo tissue sections immunostained for sarcomeric alpha-actinin (green), cardiac alpha-myosin heavy chain (red), and E-cadherin (magenta) with a Hoechst (blue) counterstain. (E-H) Affine transform alignment allowed for overlay of the staining on the image of the sequenced and for identification of tissue structures, as seen in (I&J), the cardiac region insets from panels (D&H) respectively. Scale bars: (A-H) = 1 mm, (I, J) = 0.5 mm.
Supplementary Figure 17. Spatial position of cell types for Slide 1 from embryo 1. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 17. Spatial position of cell types for Slide 4 from embryo 1. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 17. Spatial position of cell types for Slide 5 from embryo 1. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 17. Spatial position of cell types for Slide 6 from embryo 1. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 17. Spatial position of cell types for Slide 7 from embryo 1. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 17. Spatial position of cell types for Slide 8 from embryo 1. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 17. Spatial position of cell types for Slide 9 from embryo 1. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 17. Spatial position of cell types for Slide 10 from embryo 1. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 17. Spatial position of cell types for Slide 11 from embryo 1. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 17 (continued). Spatial position of cell types for Slide 13 from embryo 1. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 17. Spatial position of cell types for Slide 14 from embryo 1. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 17. Spatial position of cell types for Slide 2 from embryo 2. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right). Cells of cranial region were visibly dislodged during the placement of the oligo grid slide. Consequently, cells outside the embryo boundary were excluded from subsequent spatial analyses.
Supplementary Figure 17. Spatial position of cell types for Slide 3 from embryo 2. Sequenced cells, broken out by cell type, are placed onto the imaged embryo at the spatial position to which they map on top. Top 16 abundant cell types are shown with a 500um scale bar (bottom right).
Supplementary Figure 18. Cell types enriched in annotated anatomical segments. Heatmap displaying the percentage of cells of a given type mapping to an anatomical annotation. Unannotated denotes regions of the embryo which were not assigned to an anatomical annotation. (OPCs - Oligodendrocyte Progenitor Cells; CTPs - Connective Tissue Progenitors).
Supplementary Figure 19. Tissue domains based on similar cell type compositions are found across the embryo. (A) UMAP dimensionality reduction displaying the tissue domains, identified using the hidden Markov random field model from the Giotto (26) package, from each slide based on similar cell type composition in spatial locations. Each point is a single domain from a single slide and is colored according to cluster assignment. (B) Cells are colored based on membership to a tissue domain in four representative slides. (C) Heatmap displaying the percentage of a cell-type that composes each tissue domain. (D) The top five cell types that compose each tissue domain are shown with absolute counts.
Supplementary Figure 20. Comparison of traditional ISH and digital in-situs. Expression for a given gene (rows) plotted for slides from embryo 1 (slides 11, 1, 14 and 8) (columns) compared to published in-situ hybridization data (27) on matched sections. Genes with high spatial autocorrelation in different tissues were chosen for display. 500um scale bar shown at the bottom right of sci-Space data.
Supplementary Figure 21. Comparison spatial single cell data and aggregated spatial data. 
(A,B) UMAP dimensionality reduction of cells aggregated by spatial positions from all slides. Aggregated spatial positions which contain cells originally mapping to a single UMAP cluster are colored red in panel (A), and spatial positions consisting of a single cell type label are colored by their corresponding label in panel (B). Positions containing cells with membership with multiple clusters or cell types are colored in grey. (C) UMAP dimensionality reduction of single cells sampled from each spatial position, colored by cell type label. (D,E) Seurat v3 aggregated spatial positions from Slide 14 only. (D) displayed in UMAP space and colored by cluster or (E) displayed in their spatial positions on the embryo. (F) Distance between the Seurat v3 highest probability mapping of a single cell onto the aggregated spatial positions and a cell’s recovered position by sci-Space. Distance calculated between positions on the spatial grid. (G,H) Comparison between sci-Space data (v1 refers to slides 1-6 and v1.1 refers to slides 7-14), 10x Visium (mouse cortex) data and Slide-seq-v2 (hippocampus) data displaying UMI/um^2 in panel (G) or genes detected per spatial feature in panel (H). (I) The current distribution of UMIs/cell of sci-Space data (4,831 reads/cell) versus deeper sequencing (43,536 reads/cell) performed on a small scale sci-Space experiment.
Supplementary Figure 22. Dissection labels transferred from a developing mouse brain atlas to cells in the sci-Space dataset. The sci-Space dataset was co-embedded with the developing mouse brain atlas dataset from La Manno et al (21). The anatomical dissection label was then transferred from the developing mouse brain atlas dataset to the sci-Space dataset. This was done using the majority label of each cell’s five nearest neighbors in the co-embedded space. Each facet shows a different slide with a contour of the brain shown to the right of each facet.
Supplementary Figure 23. Transfer of spatial labels from a developing mouse brain atlas to the sci-Space dataset -- Slide 14. The sci-Space dataset was co-embedded with the developing mouse brain atlas dataset from La Manno et al. (21). The highly resolved UMAP cluster label was then transferred from the developing mouse brain atlas dataset to the sci-Space dataset. This was done using the majority label of each cell’s five nearest neighbors in the co-embedded space. Each facet displays all sci-Space cells (in grey) with highlighted cells (in red) bearing the transferred UMAP cluster. Only clusters with greater than 5 cells are shown.
Supplementary Figure 23. Transfer of spatial labels from a developing mouse brain atlas to the sci-Space dataset -- Slide 13. The sci-Space dataset was co-embedded with the developing mouse brain atlas dataset from La Manno et al. (21). The highly resolved UMAP cluster label was then transferred from the developing mouse brain atlas dataset to the sci-Space dataset. This was done using the majority label of each cell’s five nearest neighbors in the co-embedded space. Each facet displays all sci-Space cells (in grey) with highlighted cells (in red) bearing the transferred UMAP cluster. Only clusters with greater than 5 cells are shown.
Supplementary Figure 24. Connective tissue progenitors are composed of multiple subtypes with distinct spatial distributions. (A) Sub-clustering of the differentiating mesenchyme cells with cells colored by cluster. (B) Position of 3 selected sub-clusters that show differing spatial restriction within the embryos, with subcluster 1 primarily in the limbs, subcluster 7 focused in the face, and subcluster 8 more dispersed.
**Supplementary Figure 25.** HoxA cluster gene expression signature is found across neuronal subtypes. Subtypes of the neurons from Slide 1 annotated using developing mouse brain and spinal cord atlases (21, 30) plotted (A) by UMAP and (B) spatially. (C) log_{10}-scale boxplot of the UMAP Moran’s I statistic for Hox genes displayed in (Fig. 3B) versus all other expression level-matched genes (p-value < 0.01, two sided t-test). (D) UMAP embedded expression patterns for HoxA genes. (E) UMAP embedding of neurons from the E13.5 stage of mouse development from a published single cell spinal cord dataset (30). Colors and labels mark neuron subtypes annotated by the authors. (F) Expression of the same HoxA cluster displayed in panels (A-B).
Supplementary Figure 26. Spatially restricted gene expression and comparison of sci-Space and RNA FISH detected gene coexpression. (A) Spatial patterns of Cyp26b1 and Hoxa10 gene expression for neurons from Slide 1 and (B) corresponding expression patterns across neuronal cell states/subtypes (UMAP embedded). (C) Cyp26b1 and marker gene coexpression patterns measured by sci-Space (top row) were validated by comparison with co-expression detected with RNA FISH (bottom row). Cell types were designated using the following marker genes: Cldn5 - endothelial cells, Gad2 - GABAergic (inhibitory) neurons, Lum - vascular and leptomeningeal cells (VLMCs), Pax6 - radial glia, Slc17a6 and Slc17a7 - Glutamatergic (excitatory) neurons. RNA FISH supported the spatial mapping of sci-Space transcriptomes expressing Cyp26b1 and Hoxa10 and the assayed marker genes. (D) Expression patterns (UMAP) for a subset of spatially but not cell state restricted gene across neuronal cell states (Slide 1). (E) Analysis of Slide 14 shows consistent spatial versus UMAP $\log_{10}$(Morans I) values as compared to Slide 1 (Fig. 3E).
Supplementary Figure 27. Similarity between pairs of transcriptomes as a function of spatial distance. (A,B) Pairwise angular distance between cells broken out by the distance in millimeters between cells mapping back to the spatial grid. Boxplots display all cell types with over 100 cells in (A) Slide 1 and (B) Slide 14. Stars denote significant distance coefficient in linear regression without the application of an estimate filtering step (*: p-value < 0.01, **: p-value < 0.001, ***: p-value < 0.0001, Wald linear regression test; Methods).
Supplementary Figure 28. Re-examining the effect of lineage on transcriptome in the developing *C. elegans* dataset using angular distance. **(A)** Pairwise angular distance between cells in the *C. elegans* AB lineage (2). Boxplots denote lineage relationship between the pair of cells. **(B)** Relationship between Jenson-Shannon distance and angular distance calculated between all pairs of cells displayed in panel (A). **(C)** Empirical confirmation of the Law of Total Variance. Red bars indicate the sum of unexplained and explained variance after grouping cells that shared a common parent. Blue bar denotes the average mean squared error from the global mean. **(D)** Variance explained for cells in the AB lineage after accounting for parental identity. Variance explained statistic was computed as described in Methods.
Supplementary Figure 29. Variance explained by the cell type label, spatial label, or both. Proportion of variance explained, apart from that attributable to sparse UMI sampling, as described in Methods. Each spatial bin corresponds to 4 adjacent spots that are collapsed. Spatial bins that only contained a single cell of a given cell type were excluded from the analysis.
Supplementary Figure 30. Modules of gene expression from UMAP projection and spatial positions. (A) Gene modules were recovered by using either the aggregated spot by gene count matrix (Spatial) or the cell-by-gene count matrix (UMAP). Panel A shows the distribution of genes per module and the adjusted rand index calculated on the two groupings. (B,C) Row- and column-clustered heatmap of aggregated (B) UMAP-derived module expression (rows) per UMAP cluster (columns) or (C) spatially-derived module expression (rows) per UMAP cluster (columns). Color bar on top corresponds to the majority cell type within the UMAP cluster. (D) Row- and column-clustered heatmap of aggregated, spatially-derived module expression (rows) per spatial position (columns). Color bar on top corresponds to the majority cell type within each spatial position.
Supplementary Figure 31. Subtypes of chondrocytes are spatially restricted. (A) Sub-clustered chondrocytes colored by cluster. Colored arrowheads highlight accumulations of cluster 6 (blue) and cluster 7 (purple). (B) Log2-fold change of genes expressed in cluster 3 (left) and cluster 6 (right) relative to a background distribution of sampled chondrocyte cells.
Supplementary Figure 32. Characteristics of neural pseudotemporal trajectory. (A-D) UMAP embedding of neural cells in the trajectory colored by (A) aggregate expression of G1/S markers, (B) aggregate expression of G2/M markers, (C) selected gene expression with documented roles in neuronal migration, or (D) age of transferred nearest neighbors in the developing brain atlas dataset (21). (E) Scatter plot displaying transferred age and pseudotime for each of the three trajectories with a least squares regression fit (blue).
Supplementary Figure 33. Spatial location of cortical neural trajectories. Neurons and radial glia identified in the neural trajectories are shown colored from navy blue (early in pseudotime) to yellow (late in pseudotime), while neurons and radial glia not found in the trajectory are colored grey.
Supplementary Table S1. Reagents and dilutions for immunostaining

<table>
<thead>
<tr>
<th>Host, Target or Counterstain</th>
<th>Vendor, Catalog #</th>
<th>Dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goat, E-Cadherin</td>
<td>R&amp;D Systems, AF748</td>
<td>1:100</td>
</tr>
<tr>
<td>Rabbit, Sarcomeric Alpha-Actinin</td>
<td>Abcam, ab68167</td>
<td>1:100</td>
</tr>
<tr>
<td>Mouse, MYH6</td>
<td>Abcam, ab50967</td>
<td>1:200</td>
</tr>
<tr>
<td>Donkey, Goat IgG</td>
<td>ThermoFisher, A11055</td>
<td>1:1000</td>
</tr>
<tr>
<td>Donkey, Rabbit IgG</td>
<td>ThermoFisher, A31573</td>
<td>1:1000</td>
</tr>
<tr>
<td>Donkey, Mouse IgG</td>
<td>ThermoFisher, A31570</td>
<td>1:1000</td>
</tr>
<tr>
<td>Hoechst 33342, Trihydrochloride, Trihydrate</td>
<td>Invitrogen, H3570</td>
<td>1:2000</td>
</tr>
<tr>
<td>Parameter</td>
<td>Cyp26b1</td>
<td>Gad2</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>Detection channel</td>
<td>Channel 2</td>
<td>Channel 3</td>
</tr>
<tr>
<td>Requested pixel size (um)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background radius (um)</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Median filter radius (um)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sigma (um)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum area (um^2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum area (um^2)</td>
<td>0.6</td>
<td>1</td>
</tr>
<tr>
<td>Threshold</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Split by shape</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>Cell expansion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Include cell nucleus</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>Smooth boundaries</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>Make measurements</td>
<td>False</td>
<td></td>
</tr>
</tbody>
</table>
References and Notes


36. S. Srivatsan, M. Regier, Creating sci-space grids for spatial barcoding v1 (protocols.io.Bm64k9gw), Protocols.io (2020); https://doi.org/10.17504/protocols.io.bm64k9gw.


