nature biotechnology

LETTERS

https://doi.org/10.1038/s41587-020-0739-1

Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2

Robert R. Stickels[1,2,3,6], Evan Murray[1,6], Pawan Kumar[1], Jilong Li[1], Jamie L. Marshall[1], Daniela J. Di Bella[4], Paola Arlotta[5], Evan Z. Macosko[1,5,7] ✉ and Fei Chen[1,4,7] ✉

Measurement of the location of molecules in tissues is essential for understanding tissue formation and function. Previously, we developed Slide-seq, a technology that enables transcriptome-wide detection of RNAs with a spatial resolution of 10 µm. Here we report Slide-seqV2, which combines improvements in library generation, bead synthesis and array indexing to reach an RNA capture efficiency ~50% that of single-cell RNA-seq data (~10-fold greater than Slide-seq), approaching the detection efficiency of droplet-based single-cell RNA-seq techniques. First, we leverage the detection efficiency of Slide-seqV2 to identify dendritically localized mRNAs in neurons of the mouse hippocampus. Second, we integrate the spatial information of Slide-seqV2 data with single-cell trajectory analysis tools to characterize the spatiotemporal development of the mouse neocortex, identifying underlying genetic programs that were poorly sampled with Slide-seq. The combination of near-cellular resolution and high transcript detection efficiency makes Slide-seqV2 useful across many experimental contexts.

The ab initio identification of spatially defined gene expression patterns can provide insights into the development and maintenance of complex tissue architectures and the molecular characterization of pathological states. We recently developed Slide-seq[1], a spatial genomics technology that quantifies expression across the genome with 10-µm spatial resolution. While recent developments in imaging-based transcriptomics have enabled the identification of tens to hundreds of preselected genes in fixed specimens[2–5], array-based approaches[1,6,7] such as Slide-seq critically decouple the imaging from molecular sampling while simultaneously allowing for transcriptome-wide identification of molecular patterns in diverse tissue sections[1,6,7]. In Slide-seq, densely barcoded bead arrays, termed 'pucks', are fabricated by split-pool phosphoramidite synthesis and indexed up front using a sequencing-by-ligation strategy. Once the arrays are indexed, Slide-seq assays are performed with equipment found in a standard molecular biology laboratory, enabling the facile reconstruction of three-dimensional (3D) tissue volumes that are tens or even hundreds of cubic millimeters in size.

However, Slide-seq's low transcript detection sensitivity limited the range of biological problems to which the technology could be applied. Through improvements to the barcoded bead synthesis, the array sequencing pipeline and the enzymatic processing of cDNA, we report here a version of Slide-seq with an order of magnitude higher sensitivity. With our new protocol, termed Slide-seqV2, we demonstrate a range of new analytical possibilities by leveraging its improved capture efficiency, including the identification of process-localized genes in neurons and the analysis of developmental trajectories in situ.

We increased the yield of Slide-seq capture by improving the array generation pipeline and the library preparation strategy (Fig. 1a). First, we developed a new strategy to spatially index barcoded bead arrays using a monobase-encoding scheme with sequencing by ligation using sequential interrogation by offset primers[8,9] (Supplementary Fig. 1a–c and Methods). We were motivated to develop the monobase-encoding scheme for two reasons. First, SOLiD dibase encoding utilizes proprietary cleavage chemistry that is not commercially available. Second, computational matching between dibase-sequenced barcodes and Illumina sequencing requires conversion between colorspace and basespace and is not error robust. Our open-source monobase-sequencing strategy, which uses only readily available reagents, performed equivalently to SOLiD in array indexing (Supplementary Fig. 1d,e). In addition, we optimized the conditions for split-pool synthesis of the 10-µm polystyrene barcoded beads (Methods), which improved the clonality of our barcodes (Supplementary Fig. 2). Together, these strategies enabled more efficient recovery of gene expression on Slide-seqV2 arrays per Illumina read.

Next, we optimized the enzymatic library preparation steps of Slide-seqV2. We hypothesized that, due to a tissue's inhibitory presence during reverse transcription, the template-switching reaction that adds a 3′ priming site for whole-transcriptome amplification was inefficient. We therefore added another second-strand synthesis step[10] after reverse transcription to increase the number of cDNAs that could be amplified by PCR. We performed Slide-seqV2 on embryonic day (E) 12.5 mouse embryos and obtained ~9.3-fold more transcripts (unique molecular identifiers, UMIs) per bead than with the original Slide-seq protocol (Fig. 1b; median UMIs, Slide-seqV2 = 550 and Slide-seq = 59). Similarly, in the adult mouse hippocampus, we observed an 8.9-fold increase in the number of UMIs per bead, with the majority of the improvement (4.6-fold; Supplementary Table 1) attributable to the additional second-strand synthesis step and the remaining improvement largely due to improvements in bead barcode synthesis. In the mouse hippocampus, the capture efficiency of Slide-seqV2 was higher than that of a recently released commercial spatial transcriptomics (ST) technology (mean UMIs, Slide-seqV2 = 45,772 and 10x Genomics Visium = 27,952 for equal feature size (110 µm diameter binned area; Supplementary Fig. 3a–d) while maintaining 30-fold-improved spatial resolution per feature (30.25-fold by area

[1]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [2]Graduate School of Arts and Sciences, Harvard University, Cambridge, MA, USA. [3]Division of Medical Science, Harvard Medical School, Boston, MA, USA. [4]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [5]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [6]These authors contributed equally: Robert R. Stickels, Evan Murray. [7]These authors jointly supervised this work: Evan Z. Macosko, Fei Chen. ✉e-mail: emacosko@broadinstitute.org; chenf@broadinstitute.org

per feature; 10-μm feature size for Slide-seqV2 compared to 55-μm feature size for 10x Genomics Visium data). We also compared the sensitivity of Slide-seqV2 to that of HDST, another ST technology with high spatial resolution[7]. We found that, in the mouse olfactory bulb, Slide-seqV2 recovered many more transcripts per 10-μm feature than HDST (44.9-fold mean difference; 494 UMIs for Slide-seqV2 versus 11.5 UMIs for HDST; Supplementary Fig. 3e,f).

Next, we sought to quantify the absolute sensitivity of Slide-seqV2 relative to other molecular technologies that measure RNA content in cells and tissues. We compared counts of CA1 marker genes (*Atp2b1*, *Ocaid2* and *Slc17a7*) in an equal number of cells that were measured by the following: (1) Slide-seqV2; (2) Drop-seq, a high-throughput single-cell RNA (scRNA)-seq method[11,12]; and (3) single-molecule fluorescence in situ hybridization (smFISH; refs. [5,13,14]; Methods). We found that Slide-seqV2 detected similar patterns to smFISH (Fig. 1c and Supplementary Figs. 4 and 5a) and similar numbers of UMIs when compared to Drop-seq for the three genes measured (equivalent area in Slide-seqV2 to number of cells taken from Drop-seq; mean ± s.d., scRNA-seq = 33.5 ± 1.4, 2.1 ± 1.5 and 1.2 ± 1.5; Slide-seqV2 = 15.7 ± 1.5, 2.3 ± 2.4 and 1.9 ± 2.6; Fig. 1d; $n = 6$; Supplementary Table 2). To more thoroughly characterize the sensitivity of Slide-seqV2, we compared the total UMI counts per gene for all genes detected in CA1 excitatory neurons in Drop-seq[11] to the UMIs detected for an equivalent number of CA1 cells by area in Slide-seqV2. We found that, across the genome, Slide-seqV2 detected approximately 44% ± 26% of the counts of Drop-seq (median ± median absolute deviation (MAD); Supplementary Fig. 5b), demonstrating that the Slide-seqV2 capture efficiency approaches that of modern single-cell technologies. Lastly, we found Slide-seqV2 to be highly reproducible between replicates ($\rho = 0.98$; Supplementary Fig. 5c).

We next applied Slide-seqV2 to gain insight into biological problems where higher capture sensitivity is important. Neurons actively transport specific mRNAs to dendrites and postsynaptic densities, where they play critical roles in synaptic development and plasticity[15–17]. Previous studies have explored dendritic enrichment through physical microdissection or cell culture, but none has systematically identified the distribution of dendritically localized transcripts in situ. Dendritic mRNAs constitute only a tiny fraction of neuronal transcripts[18], necessitating higher-sensitivity methods for their detection. To identify dendritically localized mRNAs from our mouse hippocampal Slide-seqV2 dataset, we took advantage of the stereotyped architecture of the CA1 neuropil to reduce the spatial localization of transcripts to a one-dimensional (1D) profile perpendicular to the CA1 soma layer (from the stratum oriens to the stratum pyramidale and across the stratum radiatum; Fig. 2a,b). For each gene detected in Slide-seqV2 ($n = 4$ sections), we calculated the spatial expression as a function of distance from the soma (representative spatial expression profiles shown in Fig. 2b, bottom).

To select for dendritically localized mRNA, we performed differential expression analysis, comparing the proximal neuropil (stratum radiatum) to the soma (stratum pyramidale). The CA1 neuropil contains glial cell types (that is, microglia and astrocytes) that also contribute RNA and interfere with analysis; we therefore included only genes expressed in CA1 pyramidal cells (>0.5 transcripts per million in CA1 pyramidal neurons) and excluded those that are markers of non-neuronal cell types (Methods and Supplementary Table 3) based on existing scRNA-seq data of the hippocampus[11]. After filtering, differential expression analysis between the proximal neuropil and the soma revealed 213 significant genes with greater than twofold dendritic enrichment ($q < 0.05$, unpaired t-test, $n = 4$ sections; Fig. 2c and Supplementary Table 3). These genes overlapped significantly ($P < 10^{-16}$, hypergeometric test; Supplementary Fig. 6a) with lists of dendritically enriched RNAs from two previous studies[19,20], suggesting that Slide-seqV2 can discover dendritically enriched genes.

Next, we asked whether functionally related genes showed similarities in their dendritic enrichment. First, we grouped dendritically enriched genes according to their 1D spatial expression profile (Fig. 2d). Using unsupervised clustering, we identified four spatial expression clusters of dendritically localized genes in CA1 neuropil, with clusters having different degrees of dendritic enrichment (Fig. 2e and Supplementary Table 3). To identify whether this observed spatial diversity in localization was related to protein function, we used Gene Ontology (GO) to determine the cellular components of each spatial cluster (Supplementary Fig. 6b and Methods). We found that each cluster was enriched for ontologically distinct groups of genes. Specifically, the first two clusters were enriched for components of the cellular respiration machinery and ubiquitin ligases, while clusters 3 and 4 were enriched for ribosomal subunits. Slide-seqV2's genome-wide capture allowed us to visualize the heterogeneity in dendritic trafficking across two synaptic and two cytoskeletal genes chosen from each cluster (Fig. 2f; spatial reconstructions of all 213 genes are shown in Supplementary Dataset 1). Taken together, these data demonstrate Slide-seqV2's ability to characterize process-localized mRNAs, which appear to display significant heterogeneity among the various trafficked synaptic mRNA components.

The specificity of dendritically enriched genes for specific cell types (for example, CA1 versus other pyramidal cells) has not been widely examined, in part because traditional approaches have measured dendritic trafficking in vitro or only from a single in vivo cell type. To explore this question, we integrated Slide-seqV2 data, which span multiple hippocampal fields, with scRNA-seq data from the same tissue. From an existing hippocampal single-cell dataset[11], we computed differential expression between CA1 and other hippocampal pyramidal cells for all genes (Methods). The dendritically localized set identified by Slide-seqV2 showed a significant depletion of differentially expressed genes relative to somatically enriched genes ($P < 0.05$, Wilcoxon rank-sum test; Fig. 2g). These results suggest that, in the hippocampus, dendritically localized transcripts are more likely to be broadly expressed genes rather than markers of specialized neuronal cell types.

Dendritically enriched genes are expressed and trafficked from the soma compartment; we therefore asked whether cell-type-specific expression changes are reflected in the dendritic compartment. When dendritically enriched genes were examined in Slide-seqV2 in both CA3 and CA1 neurons, the variance in fold change in the soma for dendritic genes was much larger than the variance in fold change observed for dendrites (Fig. 2h; two-sample F-test, $P = 3 \times 10^{-9}$). Furthermore, soma fold change explained only 13% of the variance in dendritic expression between CA3 and CA1 dendrites. These results indicate that expression in dendrites is relatively buffered from the soma, suggesting the existence of distinct regulatory mechanisms in these two neuronal compartments.

During development, dynamic changes in gene expression across time and space help give rise to complex tissue architectures and terminally differentiated cell types. An array of computational strategies have been developed to identify and explore developmental trajectories from scRNA-seq data[21–24] based on similarities in gene expression between individual profiles. More recently, an additional approach called RNA velocity was developed that dynamically models expression trajectories by the relative quantities of spliced and unspliced transcripts for each gene[25]. Inspired by this work, we reasoned that the combination of Slide-seqV2's enhanced capture efficiency and its near-single-cell resolution might allow us to exploit these powerful algorithms directly on our spatial data to learn how developmental processes proceed across a tissue section.

In the embryonic mouse neocortex, neuronal development progresses along a radial axis that begins in the ventricular zone (VZ) and moves through the subventricular zone (SVZ), intermediate zone (IZ) and finally the cortical plate (CP), where neurons

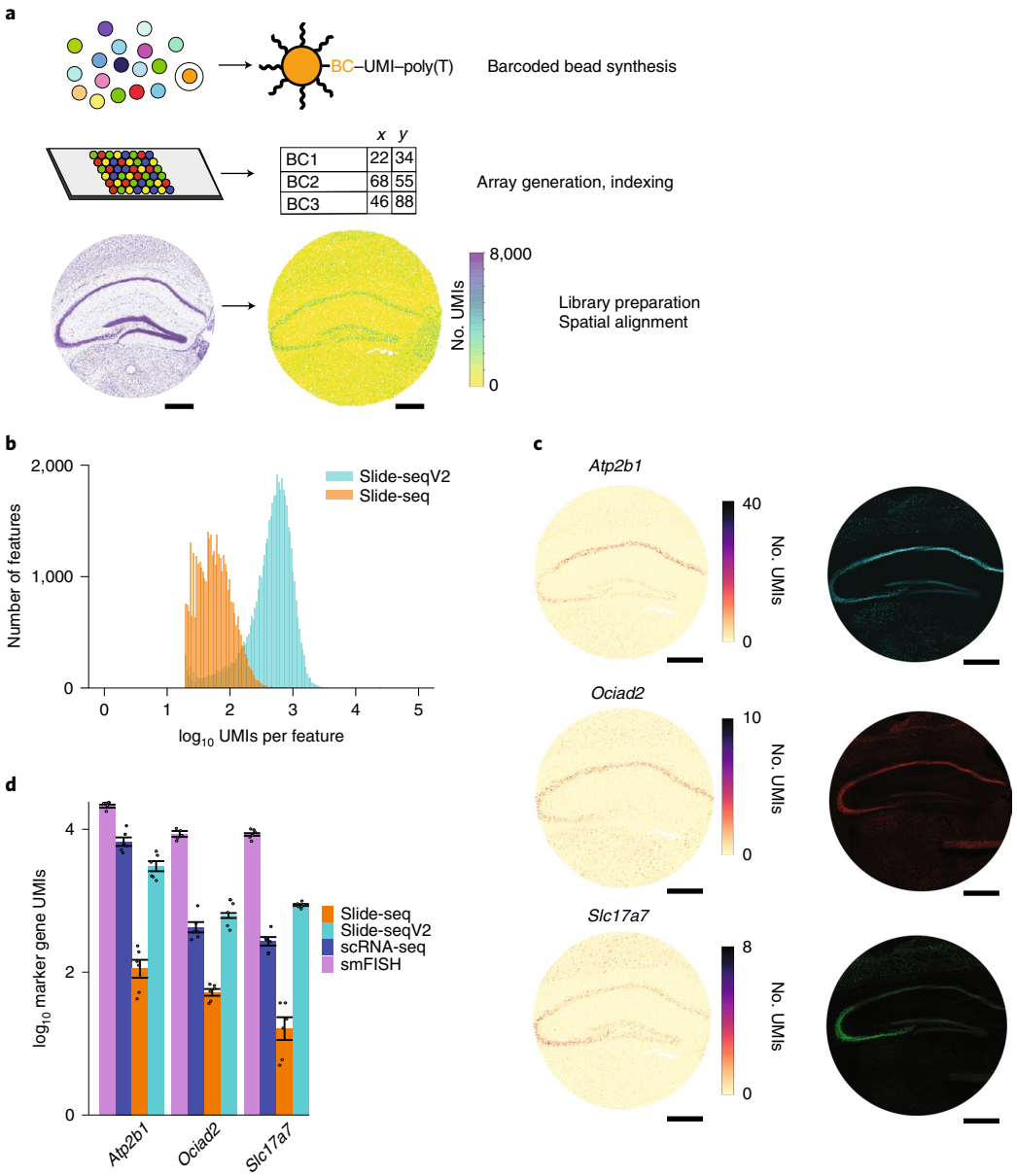

Fig. 1 | Highly improved mRNA detection sensitivity in Slide-seqV2. a, Overview of the Slide-seq method. An example array is shown of mouse hippocampus generated with Slide-seqV2 in which each bead is colored by the number of UMIs. BC, barcode. **b**, Histogram of the number of UMIs per bead for Slide-seq (red) versus Slide-seqV2 (blue) on serial mouse embryo sections. **c**, Images of hippocampus marker genes in Slide-seqV2 (left) versus hybridization chain reaction (HCR) FISH images (right). Data represent $n=1$ HCR experiment on a serial section of Slide-seq. **d**, Comparison of marker gene counts in the mouse hippocampus CA1 across four modalities ($n=6$ measurements per modality; mean ± s.d. data are reported in Supplementary Table 2). For smFISH, Slide-seqV2 and Slide-seq, all transcript counts within a fixed area of the CA1 were summed together. For scRNA-seq, we summed the counts for the number of CA1 pyramidal cells counted within this area. Scale bars, 500 µm.

integrate into cortical layers in a birthdate-dependent manner[26]. We wondered whether Slide-seqV2 data could be used to successfully recover this highly spatially organized developmental trajectory[27]. We first applied unsupervised clustering[28] to Slide-seqV2 data from an E15 developing mouse brain to characterize gene expression gradients in the neocortex. We annotated clusters corresponding to cell types in different developing brain regions, including the cortex and striatum (Fig. 3a and Supplementary Fig. 7a,b). Segregating just the radially developing cortex (Fig. 3a, black box), we reclustered the beads to reveal populations representing the VZ, SVZ, IZ, CP, early cortical layers (L5/6) and Cajal Retzius (CR) cells (Fig. 3a).

To determine whether Slide-seqV2 data can identify developmental trajectories, we first applied scVelo[29], a recently developed trajectory inference method that leverages splicing information[25], to order our beads along predicted latent time (LT). Projection of each bead's LT value onto spatial coordinates successfully recapitulated the established radial developmental axis of the neocortex (Fig. 3b). A very similar trajectory was recovered using the pseudotime ordering generated by Monocle3 (refs. [22,30]; Supplementary Fig. 7c).

During the course of a developmental process, each stage of maturation can proceed at a different rate. We wondered whether Slide-seqV2's spatial information could be exploited to identify the relative rates of differentiation across the radial axis of neocortical development. To accomplish this, we took the spatial derivative of the scVelo-generated LT (Methods), recovering regions where LT changed most dramatically (Fig. 3b, magnitude of arrows representing

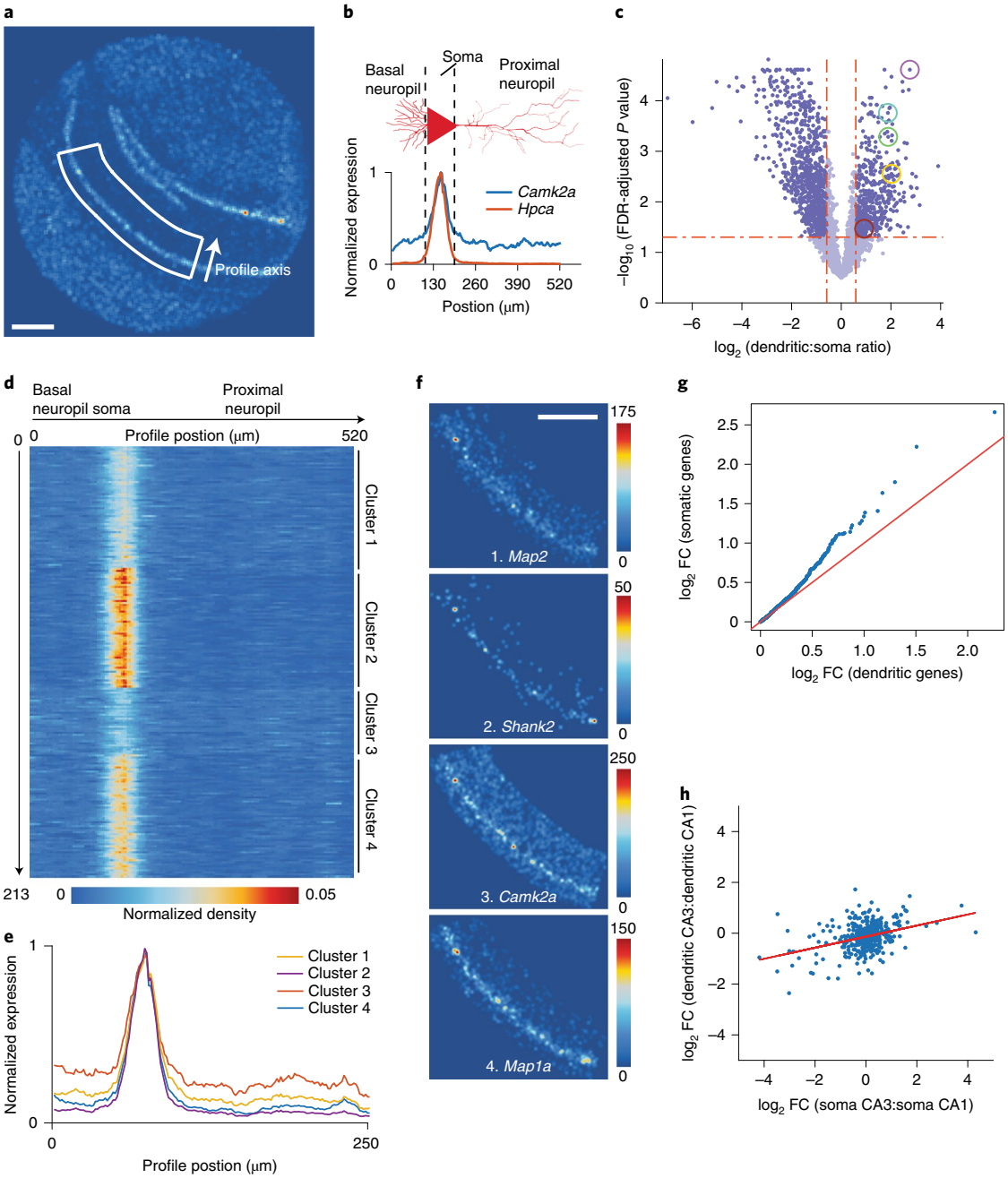

Fig. 2 | Slide-seqV2 reveals spatial patterning of dendritically enriched mRNAs. a, Spatial heat map of the number of UMIs for a hippocampal Slide-seqV2 dataset. **b**, Top, schematic of linear spatial profiling across CA1 soma and dendrites. Bottom, spatial profiles of a CA1 marker (*Hpca*, red) and a classically dendritically localized gene (*Camk2a*, blue). **c**, Differentially expressed genes in the soma versus proximal dendrites. Two-tailed, two-sample *t*-tests were performed ($n=5$ tissue sections), and genes with a false discovery rate (FDR)-corrected *P* value of <0.05 and a fold change >2 are highlighted. Several classically known dendritically expressed genes are circled. *Camk2a*, FDR-adjusted $P=2.8\times10^{-3}$, yellow; *Ef1a*, FDR-adjusted $P=5\times10^{-4}$, green; *Prkcz*, FDR-adjusted $P=0.034$, red; *Map2*, FDR-adjusted $P=1.7\times10^{-4}$, teal; and *Ddn*, FDR-adjusted $P=2\times10^{-5}$, purple. **d**, Expression heat map of 237 dendritically enriched RNAs across the neuronal profile axis. Genes are shown clustered by their spatial profile (*k*-means clustering, four clusters). Rows are normalized and summed to 1. **e**, Average spatial expression profile of each of the four gene clusters identified in **d** across the CA1. **f**, Slide-seqV2 reconstruction images of one synaptic protein-encoding gene from each of the four clusters in **d**. Scale bars, 500 μm for all Slide-seqV2 reconstructions. The color bar represents the total number of UMIs detected for a gene. **g**, Quantile–quantile plot of the $\log_2$ fold change ($\log_2$ FC) between CA1 and CA3/dentate pyramidal cell types (defined by scRNA-seq[12]) of dendritic (*x* axis) compared to somatic (*y* axis) gene sets defined by the analysis in **c**. **h**, Ratio of expression between CA3 and CA1 regions in the soma and dendrites for Slide-seqV2 data. Linear fit is shown in red (slope = 0.22, $R^2=0.13$).

the magnitude of the derivative). We found that the spatial rate of change was most pronounced at the earlier stages of the trajectory, decreasing as cells progressed from the VZ to the SVZ/IZ and largely terminating in the CP.

Because each bead's physical position is strongly predictive of its LT value, we reasoned that combining spatial and LT information could give us considerably greater statistical power to identify gene expression changes across this developmental process. The scVelo

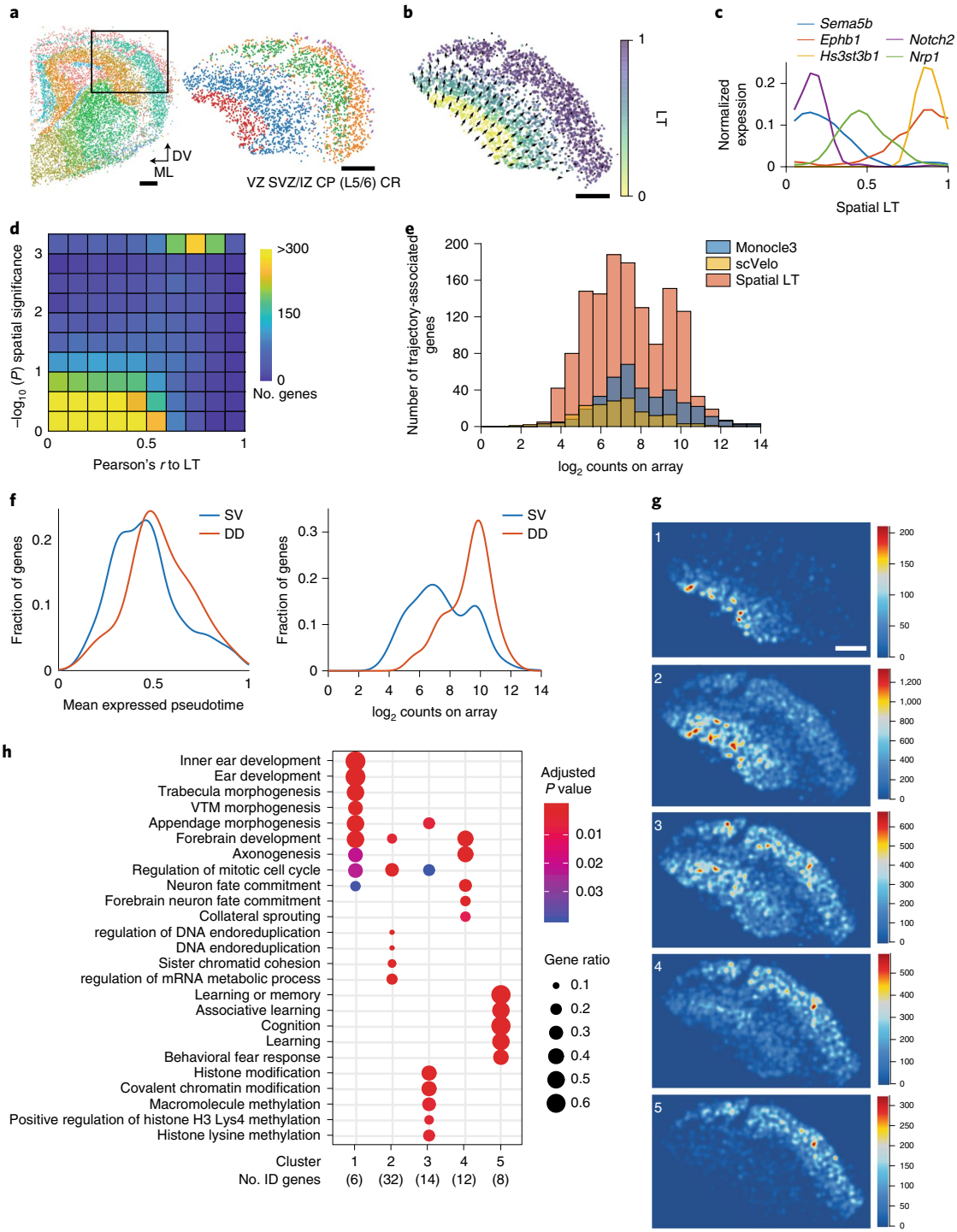

Fig. 3 | Slide-seqV2 of developing mouse cortex reconstructs spatial developmental trajectories. a, Left, unsupervised cluster analysis of Slide-seqV2 data obtained from a section of E15 mouse brain. The black box delineates the region used in the analysis. Scale bar, 200 μm. ML, medial–lateral axis; DV, dorsal–ventral axis. Right, beads present within the black box colored by their annotated cluster identities and subsetted by clusters of cortical identity. Red, VZ; blue and purple, SVZ and IZ, respectively; green and orange, CP and L5/6, respectively; pink, CR cells. These reflect the layers present in the mouse cortex at this time point. Scale bar, 200 μm. **b**, Beads within the anatomical region of developing cortex colored by their assigned LT metric from scVelo. Arrow size and direction correspond to the direction and magnitude of the spatial derivative of the LT in physical space. Scale bar, 200 μm. **c**, Expression profiles of sample genes jointly identified by Slide-seqV2, scVelo and Monocle3 across the Slide-seqV2-generated spatial LT axis. **d**, Two-dimensional (2D) density plot quantifying the relationship between a gene's correlation with scVelo LT (*x* axis) and spatial significance (permutation test, one-sided, *y* axis; Methods). Each square is colored by the number of genes found in that bin. **e**, Stacked histogram of the number of genes associated with the developmental trajectory by Monocle3 (blue), scVelo (yellow) and spatial LT (red) binned by expression level. **f**, Left, density plot of all spatial LT genes (SV) compared to DD LT genes (DD) across mean-expressed LT value; right, density plot of all spatial LT genes (SV) compared to DD LT genes (DD) for summed gene expression across array. **g**, Slide-seqV2 reconstruction images of metagenes associated with each spatial cluster of DD genes Scale bar, 200 μm. **h**, GO classifications using over-representation analysis (Methods) for biological process terms for each spatial cluster in **g** (hypergeometric test, FDR-corrected *P* value).

method was able to identify 179 genes with significant loading on LT, while the Monocle3 approach identified 377 genes. We previously demonstrated that we could leverage the spatial dimension of Slide-seq to systematically discover non-random spatial gene expression patterns[1]. Leveraging this, we identified 1,349 spatially varying genes in the developing neocortex ($P < 0.005$; Methods and Supplementary Table 4; spatial expression plots of all genes are in Supplementary Dataset 2). Among these were genes that are known to be involved in cortical development and are shared among Slide-seq and the trajectory inference methods, including *Sema5b* and *Nrp1*, which encode proteins involved in axonal guidance[31,32] (Fig. 3c). We noted that these genes correlated strongly with the spatial LT axis. Thus, to systematically find genes that varied along this axis, we correlated the expression of these 1,349 non-random genes with a spatial LT axis that was created by fitting a surface to the LT values in physical space (Methods). Of the 1,349 spatially variable genes, 1,043 correlated significantly with LT (pFDR < 0.005), while very few of the non-spatially variable genes showed a significant LT relationship (Fig. 3d). In addition, the 1,043 genes were highly overlapping with the genes identified by the trajectory inference methods across the range of expression levels (Fig. 3e). Among these genes were 76.5% of the scVelo-identified genes (137/179; Supplementary Fig. 8 and Supplementary Table 5) and 75.6% of the Monocle3-defined genes (285/377; Supplementary Fig. 8 and Supplementary Table 5). These results gave us confidence that the 1,043 genes found using our spatial LT approach were truly associated with neocortical development.

We applied this spatial LT approach to the embryonic eye at E12.5, a critical period of cell differentiation and migration[33]. We recovered the radial axis of ocular lens development, with spatial trajectories from the lens placode inward to the developing lens (Supplementary Fig. 7d,e). Similarly to the cortex, we found that spatial LT provides a highly sensitive method to detect developmental genes along this trajectory. With this approach, we identified >1,000 genes as uniquely spatially variable (Supplementary Figs. 7f and 8). Many genes recovered are important in ocular development, including genes essential to the development of the lens and associated structures such as the lens placode (*Pax6*)[34] and the primordial optic cup (*Vax2*)[35]. This list also included many from the crystallin (*Crybb3*)[36] and aldehyde dehydrogenase (*Aldh1a1*)[37,38] gene families whose products form the fiber network of the lens[36] and pattern signaling networks across the optic area[37,38]. Additionally, we found genes (*Aldh1a3* and *Col9a1*) identified as genetic drivers of ocular development disorders[39,40] that were spatially enriched and whose protein products differentially pattern the lens placode, giving rise to distinct structures within the eye[33] (Supplementary Fig. 7g). These results suggest that spatial LT, in combination with Slide-seqV2, can add spatial context to understanding the molecular signatures of genetic drivers of disease.

Inspired by the spatial enrichment of molecular signatures of genetic diseases in the developing eye, we next focused on neurodevelopmental disorders (DD), a class of diseases frequently caused by pathogenic mutations in protein-coding genes[41] that often disrupt the normal process of neocortical development. We asked how a set of 299 DD-associated genes, recently discovered by exome sequencing of DD parent–offspring trios[42], distributed on our spatial LT trajectory. A total of 74 of the 299 DD-associated genes were found in the spatial LT gene set (1.87-fold enrichment, $P = 3.2 \times 10^{-8}$). These genes were expressed later in average LT compared with all spatial LT genes (Fig. 3f, left). Interestingly, the average expression of the 74 DD-associated spatial LT genes was much higher than that of all spatial LT genes (Fig. 3f, right). These 74 genes could be clustered into five groups based on their spatial expression patterns (Fig. 3g and Methods). The individual clusters were enriched for distinct GO functional terms, suggesting that these genes participate in distinct developmental processes and pathways (Fig. 3h), ranging from chromatin modification to the establishment of neuronal states. Once additional phenotypic data become available about the relative clinical differences among these DD-associated genetic disorders, they will help reveal how such phenotypes differentially load onto the spatial LT axis.

Here we described Slide-seqV2, a high-resolution spatial genomics technology with nearly an order of magnitude higher sensitivity than the original Slide-seq protocol. In particular, we demonstrated how the higher capture efficiency of Slide-seqV2 substantially expands the scope of possible analyses, including the discovery of genes with distinct patterns of subcellular localization and the tracing of developmental programs involved in fate specification through space.

To facilitate adoption of the technology, we have generated a streamlined pipeline, which is available on our GitHub repository (https://github.com/MacoskoLab/slideseq-tools), for image processing and merging of short-read sequencing and imaging data (Supplementary Fig. 9 and Methods). With minimal user intervention required, this pipeline provides barcode locations and gene expression matrices, as well as statistics related to the alignment of imaging and short-read data. The combination of efficient molecular biology workflows, open-source sequencing chemistry for array indexing and easy-to-use software for merging imaging and sequencing data should support the wide application of Slide-seqV2. We anticipate that the technical and computational improvements here will substantially accelerate the adoption of Slide-seqV2 across the academic community.

Online content

Any methods, additional references, Nature Research 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/s41587-020-0739-1.

Received: 12 March 2020; Accepted: 2 October 2020;
Published online: 07 December 2020

References

1. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
2. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
3. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
4. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
5. Codeluppi, S. et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935 (2018).
6. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
7. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
8. Drmanac, R. et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
9. Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
10. Hughes, T. K. et al. Second-strand synthesis-based massively parallel scRNA-seq reveals cellular states and molecular features of human inflammatory skin pathologies. *Immunity* **53**, 878–894 (2020).
11. Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030 (2018).
12. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
13. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).

14. Choi, H. M. T. et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).
15. Govindarajan, A., Israely, I., Huang, S.-Y. & Tonegawa, S. The dendritic branch is the preferred integrative unit for protein synthesis-dependent LTP. *Neuron* **69**, 132–146 (2011).
16. Richter, J. D. & Klann, E. Making synaptic plasticity and memory last: mechanisms of translational regulation. *Genes Dev.* **23**, 1–11 (2009).
17. Huber, K. M., Kayser, M. S. & Bear, M. F. Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression. *Science* **288**, 1254–1257 (2000).
18. Kosik, K. S. Life at low copy number: how dendrites manage with so few mRNAs. *Neuron* **92**, 1168–1180 (2016).
19. Ainsley, J. A., Drane, L., Jacobs, J., Kittelberger, K. A. & Reijmers, L. G. Functionally diverse dendritic mRNAs rapidly associate with ribosomes following a novel experience. *Nat. Commun.* **5**, 4510 (2014).
20. Tushev, G. et al. Alternative 3′ UTRs modify the localization, regulatory potential, stability, and plasticity of mRNAs in neuronal compartments. *Neuron* **98**, 495–511 (2018).
21. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).
22. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
23. Setty, M. et al. Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* **34**, 637–645 (2016).
24. Welch, J. D., Hartemink, A. J. & Prins, J. F. SLICER: inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. *Genome Biol.* **17**, 106 (2016).
25. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
26. Lodato, S. & Arlotta, P. Generating neuronal diversity in the mammalian cerebral cortex. *Annu. Rev. Cell Dev. Biol.* **31**, 699–720 (2015).
27. Telley, L. et al. Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. *Science* **364**, eaav2522 (2019).
28. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
29. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020-0591-3 (2020).
30. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
31. Ruediger, T. et al. Integration of opposing semaphorin guidance cues in cortical axons. *Cereb. Cortex* **23**, 604–614 (2013).
32. Polleux, F., Giger, R. J., Ginty, D. D., Kolodkin, A. L. & Ghosh, A. Patterning of cortical efferent projections by semaphorin–neuropilin interactions. *Science* **282**, 1904–1906 (1998).
33. Heavner, W. & Pevny, L. Eye development and retinogenesis. *Cold Spring Harb. Perspect. Biol.* **4**, a008391 (2012).
34. Ashery-Padan, R., Marquardt, T., Zhou, X. & Gruss, P. Pax6 activity in the lens primordium is required for lens formation and for correct placement of a single retina in the eye. *Genes Dev.* **14**, 2701–2711 (2000).
35. Barbieri, A. M. et al. A homeobox gene, *vax2*, controls the patterning of the eye dorsoventral axis. *Proc. Natl Acad. Sci. USA* **96**, 10729–10734 (1999).
36. Andley, U. P. Crystallins in the eye: function and pathology. *Prog. Retin. Eye Res.* **26**, 78–98 (2007).
37. Niederreither, K., Subbarayan, V., Dollé, P. & Chambon, P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat. Genet.* **21**, 444–448 (1999).
38. Fan, X. et al. Targeted disruption of *Aldh1a1* (*Raldh1*) provides evidence for a complex mechanism of retinoic acid synthesis in the developing retina. *Mol. Cell. Biol.* **23**, 4637–4648 (2003).
39. Snead, M. P. et al. Stickler syndrome, ocular-only variants and a key diagnostic role for the ophthalmologist. *Eye* **25**, 1389–1400 (2011).
40. Fares-Taie, L. et al. *ALDH1A3* mutations cause recessive anophthalmia and microphthalmia. *Am. J. Hum. Genet.* **92**, 265–270 (2013).
41. Martin, H. C. et al. Quantifying the contribution of recessive coding variation to developmental disorders. *Science* **362**, 1161–1164 (2018).
42. Kaplanis, J. et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757–762 (2020).

Methods

Barcoded beads. Bead barcodes were synthesized either by the ChemGenes Corporation or in house on an Akta Oligopilot 10 on one of two polystyrene supports, Agilent PLRP-S-1000A 10-μm particles or 10-μm custom polystyrene from AMBiotech. Oligonucleotide synthesis was performed as described below. Beads were used with one of the two following sequences: ChemGenes Corporation beads (5′-TTTTTTTTCTACACGACGCTCTTCCGATCTJJJJJ JJJTCTTCAGCGTTCCCGAGAJJJJJJNNNNNNNNT$_{30}$-3′) and custom synthesis beads (5′-TTT_PC_GCCGGTAATACGACTCACTATAGGGCTACACGACG CTCTTCCGATCTJJJJJJJTCTTCAGCGTTCCCGAGAJJJJJJTCNNNNN NNNT$_{25}$-3′ (vs1) and 5′-TTT_PC_GCCGGTAATACGACTCACTATAGGGCT ACACGACGCTCTTCCGATCTJJJJJJJTCTTCAGCGTTCCCGAGAJJJJJN NNNNNNVVT$_{30}$-3′ (vs2)); PC, a photocleavable linker; J, bases generated by split-pool barcoding, such that every oligonucleotide on a given bead has the same J bases; N, bases generated by standard base mixing of a 1:1:1:1 ratio of A, C, T and G, such that every oligonucleotide on a given bead has different N bases; T_X, a sequence of X thymidines; V, an A, C or G but not T.

Bead synthesis. PLRP-S resin (~10-μm mean particle diameter; Agilent) was functionalized with a non-cleavable linker by ChemGenes. The functionalized beads were then used as a solid support for reverse-direction phosphoramidite synthesis (5′ to 3′) on an Akta OligoPilot 10 using a standard solid-phase DNA synthesis protocol. 5′-CE (b-cyanoethyl) phosphoramidites were purchased from Glen Research and were dissolved in anhydrous acetonitrile to obtain a concentration of 0.1 M. Successive phosphoramidites were coupled for 5 min using 5-benzylmercaptotetrazole (0.30 M in acetonitrile) as an activator. Oxidation of the phosphite backbone to a phosphate backbone was achieved using iodine. Failure sequences were capped using acetic anhydride. Dichloroacetic acid was used as a detritylation reagent. For split-pool synthesis cycles, beads were suspended in acetonitrile and divided into four equal portions. These bead aliquots were then placed in four separate synthesis columns and reacted with dG, dC, dT or dA phosphoramidites. After each cycle, beads were pooled, suspended in acetonitrile and aliquotted into four equal portions. The split-pool procedure was repeated 15 times in total (two blocks of eight and seven cycles) to obtain 4^{15} (~10^9) unique barcode sequences. After synthesis completion, the protecting groups from the nucleobases and phosphate backbone were removed by treating beads with 30% ammonium hydroxide containing 10% diethylamine for 40 h at room temperature. The beads were centrifuged, and the supernatant was discarded. Beads were then washed three times with 1% acetone in acetonitrile, three times with water and three times with a buffer consisting of 10 mM Tris and 1 mM EDTA pH 8.

Puck preparation. Puck preparation was performed as described previously[1], with the following modification. Beads were pelleted and resuspended in water with 10% DMSO at a concentration between 20,000 and 50,000 beads per μl. Then, 10 μl of the resulting solution was pipetted into each position on the gasket. The coverslip gasket filled with beads was centrifuged at 850*g* for at least 30 min at 40 °C until the surface was dry.

Puck sequencing. Puck sequencing was performed in a Bioptechs FCS2 flow cell using an RP-1 peristaltic pump (Rainin) and a modular valve positioner (Hamilton MVP). Flow rates between 1 ml min^{-1} and 3 ml min^{-1} were used during sequencing. Imaging was performed using a Nikon Eclipse Ti microscope with a Yokogawa CSU-W1 confocal scanner unit and an Andor Zyla 4.2 Plus camera. Images were acquired using a Nikon Plan Apo ×10, 0.45-NA objective. After each ligation, images were acquired in the following channels: 488-nm excitation with a 525/36-nm emission filter (MVI, 77074803), 561-nm excitation with a 582/15-nm emission filter (MVI, FF01-582/15-25), 561-nm excitation with a 624/40-nm emission filter (MVI, FF01-624/40-25) and 647-nm excitation with a 705/720-nm emission filter (MVI, 77074329). The final stitched images varied in size depending on the size of the Slide-seq array. For the arrays presented in this work the final stitched images were 6,030 pixels by 6,030 pixels.

Pucks were sequenced using a sequencing-by-ligation approach with a SOLiD dibase-encoding strategy previously described[1,43] and with a monobase-encoding strategy developed for this work. Fluorescent oligonucleotides were synthesized on an Akta OligoPilot 10 or were obtained from IDT (Supplementary Table 6). A total of eight fluorescent oligonucleotides were used and are referred to as 5(base) or 3(base) to indicate the corresponding mode of ligation and the identity of the interrogated base. Each sequencing oligonucleotide interrogates the +2 base from the ligation junction with each base identity corresponding to a fluorescent channel (A, FAM; C, Cy3; T, Cy5; G, Texas Red or AqP593).

The monobase-sequencing strategy interrogated 14 split-pool bases using three modes of sequencing by ligation. This strategy is motivated by the need to eliminate the use of proprietary cleavage reagents from SOLiD and to allow for sequencing using commercially available oligonucleotides. The overall sequencing strategy (Supplementary Fig. 1a) consisted of hybridization of a sequencing primer to interrogate the +2 base from the ligation junction followed by a ligation to interrogate a split-pool base followed by dehybridization of the sequencing primer and subsequently ligated sequencing oligonucleotide using formamide before moving onto the next ligation. The three ligation modes are 5′ ligation (ligation at the 5′ end of a hybridized sequencing primer), 3′ ligation (ligation at the 3′ end of a hybridized sequencing primer) and SEDAL (sequencing primer hybridization and ligation with a degenerate primer in a single solution).

On each bead sequence, there were two primer binding sites, a TruSeq primer site (T) and a Universal Primer site (UP; Supplementary Fig. 1b). All primer sequences are listed in Supplementary Table 6. Sequencing started with 5′ ligation on the TruSeq primer. First, the T−1 primer (primer T shortened by one base on the 5′ end) was hybridized, and a ligation was performed to interrogate the first J. After stripping with formamide, the T primer was hybridized (primer T) and the second J was interrogated. After stripping with formamide, the T+1 primer was hybridized (the + primers represent the T primer with an added N (representing a random nucleotide generated by basemixing of all four bases at an equal concentration) on the 5′ end). For the rest of the 5′ ligations, the same steps were repeated with T+2, UP−1, UP, UP+1 and UP+2, where UP primers followed the same conventions as TruSeq primers.

Next, sequencing proceeded with 3′ ligations using the 3UP primer series and 3(base) sequencing oligonucleotides. These ligations were performed for 3UP+1, 3UP and 3UP–1 (where 3UP+1 is UP plus an N base on the 3′ end, 3UP is the UP primer and 3UP–1 is the UP primer shortened by one base on the 3′ end).

Lastly, sequencing was performed with SEDAL, which utilizes degenerate primers in solution with the 5′ sequencing oligonucleotides. As the number of N bases added to the end of the hybridized primer increases, the sequencing efficiency decreases. Empirically, we found that we could not use prehybridized + primers beyond two bases (T+2 and UP+2). To overcome this, we included shortened primers with additional N bases (+3 and +4) in solution with the fluorescent sequencing oligonucleotides. We performed three rounds of SEDAL with three separate primers (T+3, UP+3 and UP+4).

This sequencing approach is outlined schematically in Supplementary Fig. 1b. For 5′ and 3′ ligation modes, 5 μM of primer was injected into the flow cell in 4× SASC for 40 min. Subsequently, the flow cell was washed in 5 ml of wash buffer (50 mM Tris-acetate and 0.05% Triton X-100). Ligation mix (5′ ligation mix: 1× T4 DNA ligase buffer (NEB), 6 U μl^{-1} T4 DNA ligase (NEB), 20 μM each of 5T, 5A, 5G and 5C oligonucleotides; 3′ ligation mix: 1× T4 DNA ligase buffer (NEB), 6 U μl^{-1} T4 DNA ligase (NEB), 20 μM each of 3T, 3A, 3G and 3C oligonucleotides; SEDAL ligation mix: 1× T4 DNA ligase buffer (NEB), 6 U μl^{-1} T4 DNA ligase (NEB), 5 μM primer, 5 μM each of 5T, 5A, 5G and 5C oligonucleotides) was then added to the chamber and allowed to sit for 40 min, at which point the flow was reversed to return the ligation mix to its original reservoir. Ligation mix was reused for a complete sequencing run before being replenished. After a subsequent wash, pucks were imaged as described above and then stripped using 10 ml of 80% formamide for 20 min. For SEDAL ligations, 5 μM of primer was added to the ligation mix, and this mixture was introduced to the chamber and allowed to sit for 2 h.

Bead barcodes consisted of 15 'J' bases, of which 14 were used. To sequence these barcodes, we performed three rounds of SEDAL, eight rounds of 5′ ligation and three rounds of 3′ ligation. The bases corresponding to each ligation mode are outlined in Supplementary Fig. 1a,b. The 14 primers necessary for this process were obtained from IDT (Supplementary Table 6).

Microscopy. Imaging was performed using a Nikon Eclipse Ti microscope with a Yokogawa CSU-W1 confocal scanner unit and an Andor Zyla 4.2 Plus camera. Images were acquired using a Nikon Plan Apo ×10, 0.45-NA objective. After each ligation, as mentioned in the Puck sequencing section, we acquired four images, one using a 488-nm laser and a 525/36-nm emission filter (MVI, 77074803), one using a 561-nm laser and a 582/15-nm emission filter (MVI, FF01-582/15-25), one using a 561-nm laser and a 624/40-nm emission filter (MVI, FF01-624/40-25) and one using a 647-nm laser and a 705/72-nm emission filter (MVI, 77074329). The final stitched images were 6,030 pixels by 6,030 pixels. All images acquired for hybridization chain reaction (HCR) experiments were imaged using a Nikon Plan Apo ×40 1.15-NA water immersion objective.

Image processing and basecalling. Image processing was performed as previously described, and we have made an easy-to-use image processing and basecalling MATLAB package that has been deposited at https://github.com/MacoskoLab/PuckCaller/. Input images are four-channel sequencing images for each puck for each time point of sequencing. For each bead, the sequence string for the bead barcode is output. For monobase imaging, the images were directly convertible to basespace rather than colorspace; thus, we omitted the step of conversion of Illumina reads to colorspace before comparison to the in situ indexing data as previously described. Metadata on all pucks used are shown in Supplementary Table 7.

Slide-seq tools. We developed the Slide-seq tools pipeline for processing Slide-seq data. The scripts, documentation and example data are available at https://github.com/MacoskoLab/slideseq-tools. The Slide-seq tools included several analysis steps, and the workflow is illustrated in Supplementary Fig. 9.

1. Extract Illumina barcodes. This step runs run_barcodes2sam.py and calls the ExtractIlluminaBarcodes function in Picard tools (https://github.com/broadinstitute/picard) to extract the barcode for each read in an Illumina lane from Illumina binary basecall (BCL) files.

2. Convert Illumina basecalls to bam. This step runs run_processbarcodes.py and calls the IlluminaBasecallsToSam function in Picard tools to collect, demultiplex and sort reads across all the tiles of a lane by barcode to produce an unmapped bam file.
3. Pre-alignment. This step runs run_alignment.py, calls the functions of TagBamWithReadSequenceExtended, FilterBam, TrimStartingSequence and PolyATrimmer in the Drop-seq tools (https://github.com/broadinstitute/Drop-seq) to tag unmapped bam files with a bead barcode (XC) and a molecular barcode (that is, UMI marked by tag XM), filters low-quality reads, trims reads with a starting sequence and poly(A) tail and calls the SamToFastq function in Picard tools to convert bam files to fastq files.
4. Align reads to the reference genome. This step runs run_alignment.py and calls the STAR aligner[44] to align reads in fastq to a reference genome.
5. Post-alignment. This step runs run_alignment.py to call SortSam and MergeBamAlignment in Picard tools to sort aligned bam files and to merge unmapped bam and aligned bam files and also calls the functions TagReadWithInterval and TagReadWithGeneFunction in Drop-seq tools to tag reads with interval and gene identity.
6. Generate alignment reports and plots. This step runs generate_plots.py and calls the CollectRnaSeqMetrics function in Picard tools and the functions BamTagHistogram, BaseDistributionAtReadPosition and GatherReadQualityMetrics in Drop-seq tools to generate a few reports based on the aligned bam file, such as read quality and mapping rate, base distribution across the reads and data on the composition and quality of the bead and molecular barcodes.
7. Select top cells by the number of transcripts. This step runs run_analysis_spec.py and calls the SelectCellsByNumTranscripts function in the Drop-seq tools to select top cells by the number of transcripts that the user specifies when submitting a request.
8. Match Illumina barcodes to bead barcodes. This step runs cmatcher.cpp to calculate Hamming distances between each Illumina barcode and all of the bead barcodes from in situ sequencing. The list of uniquely matched Illumina barcodes with a Hamming distance ≤1 along with the matched bead barcodes is output.
9. Generate reports and plots on matched barcodes. This step runs generate_plots_cmatcher.py and calls the CollectRnaSeqMetrics function in Picard tools and the functions DigitalExpression, BamTagHistogram, BaseDistributionAtReadPosition, GatherReadQualityMetrics and SingleCellRnaSeqMetricsCollector in Drop-seq tools to generate the digital gene expression matrix. Quality metrics, including a histogram of Hamming distance between Illumina and bead barcode matches, a color-scaled number of UMIs per bead and other reports, are also generated.

Supplementary Table 8 shows the running time of the Slide-seq tools on four libraries: 190926_01, 190926_02, 190926_03 and 190926_06. The Illumina platform was NovaSeq, and there were two lanes in the experiment. Reads were aligned to the GRCm38.81 genome sequence. The read base quality for alignment and the minimum number of transcripts per cell for selecting top cells were each set to ten. Reads aligned to both exons and introns were included in the gene expression analysis. To speed up the process, Slide-seq tools split each lane of NovaSeq data into ten slices, ran the alignment steps on the slices in parallel and combined the alignment outputs.

Slide-seqV2 library preparation. *RNA hybridization.* Pucks in 1.5-ml tubes were immersed in 200 µl of hybridization buffer (6× SSC with 2 U μl^{-1} Lucigen NxGen RNase inhibitor) for 30 min at room temperature to allow for binding of the RNA to the oligonucleotides on the beads.

First-strand synthesis. First-strand synthesis was performed by incubating the pucks in RT solution (115 µl water, 40 µl Maxima 5× RT buffer (Thermo Fisher, EP0751), 20 µl of 10 mM dNTPs (NEB, N0477L), 5 µl RNase inhibitor (Lucigen, 30281), 10 µl of 50 µM template switch oligonucleotide (Qiagen, 339414YCO0076714) and 10 µl Maxima H Minus reverse transcriptase (Thermo Fisher, EP0751)) for 1.5 h at 52 °C.

Tissue digestion. Two hundred microliters of 2× tissue digestion buffer (200 mM Tris-Cl pH 8, 400 mM NaCl, 4% SDS, 10 mM EDTA and 32 U ml^{-1} proteinase K (NEB, P8107S)) was then added directly to the RT solution, and the mixture was incubated at 37 °C for 30 min.

Second-strand synthesis. The solution was then pipetted up and down vigorously to remove beads from the surface, and the glass substrate was removed from the tube using forceps and discarded. Two hundred microliters of wash buffer (10 mM Tris pH 8.0, 1 mM EDTA and 0.01% Tween-20) was then added to the 400 µl of tissue clearing and RT solution mix, and the tube was centrifuged for 3 min at 3,000*g*. The supernatant was then removed from the bead pellet, and the beads were resuspended in 200 µl of wash buffer and centrifuged again. This was repeated a total of three times. The supernatant was then removed from the pellet. The beads were resuspended in 200 µl of ExoI mix (170 µl water, 20 µl ExoI buffer and 10 µl ExoI (NEB, M0568)) and incubated at 37 °C for 50 min.

After ExoI treatment, the beads were centrifuged for 3 min at 3,000*g*. The supernatant was removed from the bead pellet, and the beads were resuspended in 200 µl of wash buffer and centrifuged again. This was repeated a total of three times. The supernatant was then removed from the bead pellet, and the bead pellet was resuspended in 200 µl of 0.1 N NaOH and incubated for 5 min at room temperature. To quench the reaction, 200 µl of wash buffer was added and beads were centrifuged for 3 min at 3,000*g*. The supernatant was then removed from the bead pellet, and the beads were resuspended in 200 µl of wash buffer and centrifuged again. This was repeated a total of three times. Second-strand synthesis was then performed on the beads by incubating the pellet in 200 µl of second-strand synthesis mix (133 µl water, 40 µl Maxima 5× RT buffer, 20 µl of 10 mM dNTPs, 2 µl of 1 mM dN-SMRT oligonucleotide and 5 µl Klenow enzyme (NEB, M0210)) at 37 °C for 1 h.

After second-strand synthesis, 200 µl of wash buffer was added, and the beads were centrifuged for 3 min at 3,000*g*. The supernatant was then removed from the bead pellet, and the beads were resuspended in 200 µl of wash buffer and centrifuged again. This was repeated a total of three times.

Library amplification. Water (200 µl) was added to the bead pellet, and the beads were moved into a 200-µl PCR strip tube, pelleted in a minifuge and resuspended in 200 µl of water. The beads were then pelleted and resuspended in library PCR mix (22 µl water, 25 µl of Terra Direct PCR mix buffer (Takara Biosciences, 639270), 1 µl of Terra polymerase (Takara Biosciences, 639270), 1 µl of 100 µM TruSeq PCR handle primer (IDT) and 1 µl of 100 µM SMART PCR primer (IDT)), and PCR was performed according to the following program: 95 °C for 3 min; four cycles of 98 °C for 20 s, 65 °C for 45 s and 72 °C for 3 min; nine cycles of 98 °C for 20 s, 67 °C for 20 s and 72 °C for 3 min; 72 °C for 5 min; hold at 4 °C.

PCR cleanup and Nextera tagmentation. The PCR product was then purified by adding 30 µl of AMPure XP beads (Beckman Coulter, A63880) to 50 µl of PCR product. The samples were cleaned according to the manufacturer's instructions and resuspended in 50 µl of water. The cleanup was repeated, resuspending in a final volume of 10 µl. One microliter of the library was quantified on a Bioanalyzer High Sensitivity DNA chip (Agilent, 5067-4626). Then, 600 pg of PCR product was prepared as Illumina sequencing libraries through tagmentation with a Nextera XT kit (Illumina, FC-131-1096). Tagmentation was performed according to the manufacturer's instructions, and the library was amplified with TruSeq5 and N700 series barcoded index primers. The PCR program was as follows: 72 °C for 3 min; 95 °C for 30 s; 12 cycles of 95 °C for 10 s, 55 °C for 30 s, 72 °C for 30 s and 72 °C for 5 min; hold at 10 °C.

Samples were cleaned with AMPure XP beads (Beckman Coulter, A63880) in accordance with the manufacturer's instructions at a 0.6× bead per sample ratio (30 µl of beads to 50 µl of sample) and resuspended in 10 µl of water. Library quantification was performed using a Bioanalyzer. Finally, the library concentration was normalized to 4 nM for sequencing. Samples were sequenced on an Illumina NovaSeq S2 flow cell 100 cycle kit with 24 samples per run (six samples per lane) with the read structure 44 bases for read 1, 8 bases for the i7 index read and 50 bases for read 2. Each puck received ~200 million reads, corresponding to ~3,000 reads per bead.

Animal handling. All procedures involving animals at the Broad Institute were conducted in accordance with the US National Institutes of Health Guide for the Care and Use of Laboratory Animals under protocol number 0120-09-16. All procedures involving animals at Harvard University were handled according to protocols approved by the Institutional Animal Care and Use Committee of Harvard University (protocol number 11-03) and followed the guidelines set forth in the US National Institute of Health Guide for the Care and Use of Laboratory Animals. Wild-type C57BL/6 mice (Charles River Laboratories) were housed in a 12-h light/12-h dark cycle with ad libitum access to food and water. We set harem breeding cages and defined morning of plug detection as E0.5.

Transcardial perfusion. C57BL/6 mice were anesthetized by administration of isoflurane in a gas chamber flowing 3% isoflurane for 1 min. Anesthesia was confirmed by checking for a negative tail pinch response. Animals were moved to a dissection tray, and anesthesia was prolonged via a nose cone flowing 3% isoflurane for the duration of the procedure. Transcardial perfusions were performed with ice-cold pH 7.4 HEPES buffer containing 110 mM NaCl, 10 mM HEPES, 25 mM glucose, 75 mM sucrose, 7.5 mM $MgCl_2$ and 2.5 mM KCl to remove blood from the brain and other organs sampled. The appropriate organs were removed, frozen for 3 min in liquid nitrogen vapor and moved to −80 °C for long-term storage.

Tissue handling. Fresh frozen tissue was warmed to −20 °C in a cryostat (Leica, CM3050S) for 20 min before handling. Tissue was then mounted on a cutting block with optimal cutting temperature compound and sliced at a 5° cutting angle at 10-µm thickness. Pucks were placed on the cutting stage and tissue was maneuvered onto the pucks. The tissue was then melted onto the puck by moving the puck off the stage and placing a finger on the bottom side of the glass. The puck was then removed from the cryostat and placed into a 1.5-ml Eppendorf tube.

The sample library was prepared as described below. The remaining tissue was returned to −80 °C and stored for processing at a later date.

Diffusion analysis. Diffusion was determined as previously described by measuring features across the CA1 of the mouse hippocampus[1] for Slide-seqV2 data (Puck 200115_08).

Comparison of counts for Slide-seq, Slide-seqV2, FISH and scRNA-seq. For Slide-seq and Slide-seqV2, we subsetted a region of the CA1 and took the total number of counts for each of the marker genes. For the scRNA-seq data, we used the mouse brain scRNA-seq data from ref. [11] and sampled a number of CA1 profiles equal to the number of cells found within the Slide-seq subsetted region. For the smFISH data, we generated the data by using HCRv3.0 with probe sets against each of the genes chosen in 488 nm, 594 nm and 647 nm (*Slc17a7*, *Ociad2* and *Atp2b1*, respectively). Following recommendations from the manufacturer, we used the suggested number of probes per gene for HCRv3, namely 20 probe pairs per gene (40 probes total). Probe sequences were designed by Molecular Instruments. We stained the tissue with DAPI for segmentation purposes and for counting nuclei. We performed counting of smFISH data using a custom pipeline implemented using the Starfish package (https://github.com/spacetx/starfish) in Python.

Comparison of Slide-seqV2 to 10x Visium technology and HDST. To compare Slide-seqV2 to 10× Genomics Visium data (https://www.10xgenomics.com/solutions/spatial-gene-expression/), we downloaded available coronal mouse hippocampus data and plotted the number of UMIs per spatial feature. We next binned Slide-seqV2 data from the same region to equivalent feature size (110 µm), accounting for the Visium data 55-µm feature size and 55-µm feature spacing, merging the counts of Slide-seqV2 beads (10 µm original) within each of the larger features generated (110 µm).

To compare Slide-seqV2 to HDST, we first obtained HDST data from supplementary data provided in the original HDST publication[7]. We used the data at 10-µm feature size (5× binned) and compared it to Slide-seqV2 data collected from an equivalent region (mouse olfactory bulb). We then plotted the number of counts per feature and took the mean to obtain the difference in average counts between Slide-seqV2 and HDST.

Spatial comparison of Slide-seqV2 to smFISH. To compare Slide-seqV2 to smFISH, we used a dataset recently obtained from mouse cortex using osmFISH, an optimized method of cyclic smFISH[5]. A spatial profile was taken along the length of the cortex for osmFISH and Slide-seqV2 data perpendicular to the expression of the layer marker *Lamp5*, traversing from layer 6 to layer 1 of the cortex. Each gene for both datasets was normalized along this profile. Genes in osmFISH with a coefficient of variation along the profile of >50% were selected for analysis (this enriches for cortical layer markers; other genes do not have stereotyped patterns along this spatial dimension and are not compared). Both datasets were downsampled to 50 µm and aligned spatially along the profile by aligning the positions of layer 1 and layer 4. To analyze spatial correlation, the Pearson correlation of each gene's spatial profile against all genes was calculated between Slide-seqV2 and osmFISH as well as between osmFISH and osmFISH.

Hippocampal Slide-seqV2. Slide-seqV2 was performed on the mouse hippocampus ($n = 4$ sections, two mice). A spline was fit along the pyramidal cell layer of the CA1. Beads were averaged to a profile perpendicular to this spline ~100 µm into the basal neuropil and ~400 µm to the proximal neuropil to form a spatial profile of gene expression along the CA1 neuropil axis.

Dendritic enrichment analysis. To test for dendritic enrichment, for each gene the gene expression in the soma layer (defined as ±32.5 µm from the peak of the profile counts for all genes) was compared against the gene expression in the proximal dendrites (greater than 32.5 µm away from the peak of the CA1 layer). We leveraged existing scRNA-seq data[11] to exclude marker genes from cell types outside of CA1 using differential expression. Specifically, all genes with twofold-higher expression in cell type clusters other than CA1 were excluded from the analysis. For each gene, the gene expression in the soma layer was normalized to the total number of UMI counts in the soma layer, and the gene expression in the proximal dendrite layer was normalized to the total number of UMI counts in the proximal dendrite layer. A two-sample *t*-test was performed to identify differentially expressed genes, and pFDR was calculated as described previously[45].

Spatial clustering of dendritically enriched genes. For the 213 genes identified to be dendritically enriched, we clustered genes by their spatial profile along the CA1–neuropil axis via *k*-means clustering. The gap-statistic was used to determine the optimal number of clusters ($k = 4$).

Gene ontology analysis. For each cluster identified by spatial profiling, GO analysis was performed using the clusterProfiler[46] package in R. Cellular components were from the org.Mm.eg.db[47] (http://bioconductor.org/packages/release/data/annotation/html/org.Mm.eg.db.html). Genome-wide annotation for mouse in Bioconductor was used for the ontology database. For Fig. 3, GO analysis was performed using biological processes from org.Mm.eg.db.

Cell-type specificity of dendritic genes. To explore the relationship between the cell-type specificity of gene expression and dendritic localization (Fig. 2g), we first computed differential expression between CA1 and other hippocampal principal cells from an existing single-cell dataset[11] using the FindMarkers() function in Seurat. Next, we compared the $\log_2$ fold change of the genes in the dendritic and somatic gene sets (see "Dendritic enrichment analysis") with a Wilcoxon rank-sum test and visualized the comparison using a quantile–quantile plot. To calculate dendritic and soma expression in CA3, the same procedure as in CA1 was carried out (see "Dendritic enrichment analysis"). Soma fold change was calculated as the ratio of counts in CA3 soma to CA1 soma after normalizing by the total number of UMIs in each compartment. Dendrite fold change was calculated as the ratio of counts in CA3 soma to CA1 soma after normalizing by the total number of UMIs in each compartment.

Embryo samples. Whole-mount frozen embryos were obtained from a commercial source (Zyagen). The pregnant mice (C57BL/6NCrl) were bred and maintained by Charles River Laboratories. The timed-pregnant mice (day 10) were shipped to Zyagen the same day. The mice were killed on the appropriate day for embryo collection.

Trajectory analysis. Trajectory analysis was performed using scVelo[29]. We first loaded intronic and exonic gene expression matrices, UMAP coordinates created in Seurat from the original clustering of the Slide-seqV2 data, cluster IDs, and spatial coordinates of each bead from Slide-seqV2 into a scanpy object using a custom Python environment. We next applied the LT method developed in scVelo to our Slide-seqV2 expression data and plotted each bead using the Slide-seqV2 coordinates with the shading defined by the LT ordering. Plots of individual expression of genes over LT were generated using plotting functions in scVelo plotting the expression of each individual gene over the LT axis with coloring of each bead by cluster identity to the original clustering of the data. Plots showing expression for each of the genes on the puck were performed using a custom Python script. The gene lists for LT were called as velocity loading genes from the scVelo pipeline using standard parameters and a likelihood cutoff of >0.1.

Monocle3[22,30] was run on the data by importing the UMAP and PCA coordinates from Seurat into a Single Cell Experiment object. The analysis was performed in accordance with the monocle tutorial found at http://cole-trapnell-lab.github.io/monocle-release/monocle3/. The *q*-value cutoff for gene selection was $q < 0.005$.

Fitting a spatial surface to latent time. LT data scores generated from scVelo and spatial coordinates were taken as a 3D set of points (x, y and LT score), and a surface was fit over the set of points for a region of the cortex. A grid was created (80 µm × 80 µm for cortex), and the spatial derivative was taken over the grid using MATLAB's differentiate function. The *fx*, *fy* values of the surface were extracted from MATLAB and imported into Python. The plot for Fig. 3b was generated using a custom Python script where the magnitude of the arrows represents the magnitude of the derivative at each of the points in the grid. The position of the underlying beads is from Slide-seqV2, and the color scale is from scVelo's LT output.

Spatially non-random gene analysis. The test for spatial non-randomness was performed as previously described[1] with the following modifications. Genes were identified as spatially non-random using a custom MATLAB application. The set of pairwise Euclidean distances between all beads was calculated. Candidate genes for the statistical significance analysis were required to have at least one transcript on at least ten beads. To determine whether a transcript had a significantly non-random spatial distribution within a particular set of beads, we compared the distribution of pairwise distances between the beads expressing at least one count of that transcript to the distribution of pairwise distances between an identical number of beads sampled randomly from all mapped beads on the puck with probability proportional to the total number of transcripts on the bead. Specifically, we generated 1,000 such random samples, and for each sample we calculated the distribution of pairwise distances. We then calculated the average distribution of pairwise distances, averages pairwise across all 1,000 samples. Finally, we calculated the L1 norm between the distribution of pairwise distances for each of the 1,000 random samples and the average distribution, and the L1 norm between the distribution of pairwise distances for the true sample of beads and the average distribution. We defined P to be the fraction of random samples having distributions closer to the average distribution (under the L1 norm) than the true sample and considered any genes with values $P \leq 0.005$.

Spatial correlation to latent time. Spatially identified genes were binned along 20 spatial contours of the same LT as fitted by the surface described above. Expression was normalized for each bin by the total number of counts observed. For each gene, the Pearson correlation coefficient and the P value of the correlation between

the binned expression in the spatial LT axis was correlated with a linear function of slope 1. pFDR was calculated as described previously[45].

Spatial clustering of neurodevelopmental disorder genes. For the 74 genes identified to be involved in developmental disorders that load onto pseudotime, the spatial correlation between each gene was determined by convolving the spatial expression of each gene with an integralBoxFilter of size 70 µm, and then the 2D cross-correlation for each gene against each other gene was calculated with the MATLAB function corr2. The spatial cross-correlation matrix was clustered using *k*-means, and the gap-statistic was used to determine the optimal number of clusters ($k = 6$). The five clusters with more than two genes were analyzed with GO analysis and are visualized in Fig. 3g.

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

Data availability

All data are available at https://singlecell.broadinstitute.org/single_cell/study/SCP815/sensitive-spatial-genome-wide-expression-profiling-at-cellular-resolution#study-summary.

Code availability

Code related to this manuscript can be found at https://github.com/MacoskoLab/slideseq-tools and https://github.com/rstickels/Slide_seqv2. The following package version numbers were used for data processing and associated analyses: https://github.com/broadinstitute/Drop-seq (Drop-seq-tools-2.3.0), https://broadinstitute.github.io/picard/ (picard-2.18.14), https://github.com/alexdobin/STAR (STAR-2.5.2a), https://github.com/theislab/scvelo (0.1.25), https://github.com/cole-trapnell-lab/monocle3 (beta) and https://github.com/satijalab/seurat (2.3.4). MATLAB 2017a, R3.5.3 and Python 3.7 were used for processing data.

References

43. McKernan, K. J. et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* **19**, 1527–1541 (2009).
44. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
45. Storey, J. D. A direct approach to false discovery rates. *J. R. Stat. Soc. B Stat. Methodol.* **64**, 479–498 (2002).
46. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
47. Carlson M.org.Mm.eg.db genome wide annotation for mouse. R package version 3.8.2 http://bioconductor.org/packages/release/data/annotation/html/org.Mm.eg.db.html (2019).

Acknowledgements

We thank J. Dimidschstein and G. Fishell for their helpful advice on the developmental trajectory analysis. This work was supported by an NIH New Innovator Award (DP2 AG058488-01 to E.Z.M.), an NIH Early Independence Award (DP5, 1DP5OD024583 to F.C.), the NHGRI (R01, R01HG010647 to E.Z.M. and F.C.), the Burroughs Wellcome Fund CASI award (to F.C.) and the Schmidt Fellows Program at the Broad Institute and the Stanley Center for Psychiatric Research.

Author contributions

F.C. and E.Z.M. supervised the work. R.R.S. and E.M. performed experiments. D.J.D. and P.A. contributed to experiments on the developing neocortex. R.R.S., F.C. and E.Z.M. analyzed the data. J.L. developed the Slide-seq tools software package. P.K. developed the bead synthesis protocol. J.L.M. performed optimization experiments. F.C., E.Z.M., R.R.S. and E.M. wrote the manuscript with input from all authors.

Competing interests

R.R.S., F.C. and E.Z.M. are listed as inventors on a pending patent application related to the development of Slide-seq.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0739-1.

Correspondence and requests for materials should be addressed to E.Z.M. or F.C.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Fei Chen, Evan Macosko

Last updated by author(s): Sep 23, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used custom code for array generation and alignment, all code deposited on github. Additional tools used Drop-seq tools: https://github.com/broadinstitute/Drop-seq Picard: https://broadinstitute.github.io/picard/ STAR: https://github.com/alexdobin/STAR
Data analysis	We used matlab, python and R to perform analyses. Code related to this manuscript can be found at: https://github.com/MacoskoLab/slideseq-tools and https://github.com/rstickels/Slide_seqv2 R version 3.5 Matlab 2017a Python 3.7

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data is available at: https://singlecell.broadinstitute.org/single_cell/study/SCP815/sensitive-spatial-genome-wide-expression-profiling-at-cellular-resolution#study-summary.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We selected sample size up front for tech development due to sequencing cost. No hypothesis was tested, so we did not perform sample size calculation for power on hypothesis. Analysis performed used all available samples.
Data exclusions	No data excluded
Replication	All attempts at replication successful. Slide-seqV2 on mouse hippo-campus replicated >3 times in this manuscript.
Randomization	No randomizatioon
Blinding	No blinding

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	C57BL/6J, male and female were used, age >=60 days. Wild-type C57Bl/6 mice (from Charles River Laboratories) were housed in a 12:12 light-dark cycle with ad libitum access to food and water. For embyros, we set harem breeding cages and defined morning of plug detection as E0.5.
Wild animals	This study did not involve wild animals
Field-collected samples	No samples were field collected
Ethics oversight	All procedures involving animals at Harvard University were handled according to protocols approved by the Institutional Animal Care and Use Committee (IACUC) of Harvard University (protocol number 11-03) and followed the guidelines set forth in the National Institute of Health Guide for the Care and Use of Laboratory Animals.

Note that full information on the approval of the study protocol must also be provided in the manuscript.