bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Title: **Large field of view-spatially resolved transcriptomics at nanoscale resolution**

Short title: **DNA nanoball stereo-sequencing**

Ao Chen[1,2,*], Sha Liao[1,*], Mengnan Cheng[1,3,*], Kailong Ma[1,*], Liang Wu[1,3,4,*], Yiwei Lai[5,6,*], Jin Yang[7], Wenjiao Li[1], Jiangshan Xu[1,3], Shijie Hao[1,3], Xi Chen[1], Xing Liu[1], Xin Huang[3,7], Feng Lin[1], Xiaoxuan Tang[1], Zhao Li[1], Yang Hong[1], Defeng Fu[1], Yujia Jiang[1,8], Jian Peng[1], Shuai Liu[1], Mengzhe Shen[1], Chuanyu Liu[1,9], Quanshui Li[1], Zhifeng Wang[1,4], Zhaohui Wang[1], Xin Huang[1], Yue Yuan[1,3], Giacomo Volpe[5,6], Carl Ward[5,6], Pura Muñoz-Cánoves[10,11], Jean Paul Thiery[12], Fuxiang Zhao[1], Mei Li[1], Haoyan Kuang[1], Ou Wang[1], Haorong Lu[13], Bo Wang[13], Ming Ni[7], Wenwei Zhang[1,14], Feng Mu[7], Ye Yin[15], Huanming Yang[1,16], Michael Lisby[2], Richard J. Cornall[17], Mathias Uhlen[18,19], Miguel A. Esteban[1,5,6,20,†], Yuxiang Li[1,†], Longqi Liu[1,9,†], Jian Wang[1,16,†], Xun Xu[1,21,†]

[1]BGI-ShenZhen, Shenzhen 518103, China.

[2]Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark.

[3]BGI Education Center, University of Chinese Academy of Sciences (CAS), Shenzhen 518083, China.

[4]Shenzhen Key Laboratory of Single-Cell Omics, BGI-Shenzhen, Shenzhen 518120, China.

[5]Laboratory of Integrative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China.

[6]CAS Key Laboratory of Regenerative Biology and Guangdong Provincial Key Laboratory of Stem Cells and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Guangzhou 510530, China.

[7]MGI, BGI-Shenzhen, Shenzhen 518083, China.

[8]BGI College & Henan Institute of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou 450000, China.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[9]Shenzhen Bay Laboratory, Shenzhen 518000, China.

[10]Department of Experimental and Health Sciences, Pompeu Fabra University (UPF), ICREA and CIBERNED, Barcelona E-08003, Spain.

[11]Spanish National Center on Cardiovascular Research (CNIC), Madrid E-28029, Spain.

[12]Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou 510005, China.

[13]China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China.

[14]Shenzhen Key Laboratory of Neurogenomics, BGI-ShenZhen, Shenzhen 518103, China.

[15]BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China.

[16]James D. Watson Institute of Genome Sciences, Hangzhou 310058, China.

[17]Medical Research Council Human Immunology Unit, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7BN, United Kingdom.

[18]Department of Protein Science, Science for Life Laboratory, KTH-Royal Institute of Technology, 17121 Stockholm, Sweden.

[19]Department of Neuroscience, Karolinska Institute, 171 77 Stockholm, Sweden.

[20]Institute of Stem Cells and Regeneration, Chinese Academy of Sciences, Beijing 100101, China.

[21]Guangdong Provincial Key Laboratory of Genome Read and Write, Shenzhen 518120, China.

*These authors contributed equally to this work.

†Corresponding authors. Email: miguelesteban@genomics.cn (M.A.E.), liyuxiang@genomics.cn (Y.L.), liulongqi@genomics.cn (L.L.), wangjian@genomics.cn (J.W.), xuxun@genomics.cn (X.X.).

ABSTRACT

High-throughput profiling of *in situ* gene expression represents a major advance towards the systematic understanding of tissue complexity. Applied with enough

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

capture area and high sample throughput it will help to define the spatio-temporal dynamics of gene expression in tissues and organisms. Yet, current technologies have considerable bottlenecks that limit widespread application. Here, we have combined DNA nanoball (DNB) patterned array chips and *in situ* RNA capture to develop Stereo-seq (Spatio-Temporal Enhanced REsolution Omics-sequencing). This approach allows high sample throughput transcriptomic profiling of histological sections at unprecedented (nanoscale) resolution with areas expandable to centimeter scale, high sensitivity and homogenous capture rate. As proof of principle, we applied Stereo-seq to the adult mouse brain and sagittal sections of E11.5 and E16.5 mouse embryos. Thanks to its unique features and amenability to additional modifications, Stereo-seq can pave the way for the systematic spatially resolved-omics characterization of tissues and organisms.

MAIN TEXT

Single-cell RNA-sequencing (scRNA-seq) technologies have unveiled an unexpectedly large degree of heterogeneity in cell types and cell states among tissues (*1-5*). Because multicellular organization and cell-cell interactions determine individual cell function, spatially resolved transcriptomic information is necessary to fully understand tissue complexity in normal physiology or under perturbation. This is best illustrated in embryonic development, where the spatio-temporal coordination of cells and signaling cues determine the dynamics of self-organized patterning, but in fact applies to any tissue and biological process. For example, spatially resolved transcriptomic information of the complex interactions between cancer cells, immune cells and stromal cells in the tumor niche (*6, 7*) could be useful to improve anticancer therapies. Similarly, revealing the dynamic changes of gene expression responsible for the spreading pattern of certain neurodegenerative diseases (*8, 9*) could help prevent this process. A number of remarkable methodologies have been recently developed that enable the high-throughput transcriptomic profiling of *in situ* gene expression. These techniques can be grouped into three main categories depending on the RNA

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

capture approach: a) physical segmentation using laser microdissection, b) pre-designed targeted probes (for either single-molecule hybridization or amplification and *in situ* decoding) or c) DNA-barcoded probes. Examples of the former approach include LCM-seq (laser capture microscopy coupled with full-length mRNA-sequencing) (*10*) and GEO-seq (geographical position sequencing) (*11*), which are highly operable but have low capture throughput, limited resolution (center-to-center distance of each array component) and the accuracy of spatial information is subjected to experimental variability. Examples of techniques based on pre-designed probes are MERFISH (multiplexed error-robust FISH) (*12*) and seqFISH (sequential fluorescence in situ hybridization) (*13*) for hybridization-based methods, and FISSEQ (fluorescent in situ RNA sequencing) (*14*) and STARmap (spatially-resolved transcript amplicon readout mapping) (*15*) for *in situ* decoding. General caveats of these techniques are that they rely on a finite number of sequences, can only be practically applied to a small field of view and require complex instrumentation. As for the latter group, Slide-seq (*16*), spatial transcriptomics (ST, Visium) (*17*) and their optimized versions Slide-seqV2 and HDST (high-definition spatial transcriptome) (*18, 19*), respectively, use barcoded arrays and have attracted significant attention. Similarly, DBiT-seq (deterministic barcoding in tissue for spatial omics sequencing) uses microfluidic channels for delivering barcoded probes directly into tissues and can simultaneously capture transcripts and selected protein targets (*20*). Although unbiased in the nature of captured transcripts, barcoded probe-based techniques suffer from an inadequate balance between resolution, average numbers of gene/transcript per bin and capture rate across multiple bins. Besides, as for hybridization-based methods, the measured areas are small (e.g., Visium 42.5 mm^2, Slide-seq/Slide-seqV2 7 mm^2, HDST 13.68 mm^2 and DBiT-seq 25 mm^2) (*17-20*) and face a major bottleneck of scalability. Increasing the resolution is important because it can help define transcriptomic gradients within tissues more accurately and allow detection of the subcellular localization of transcripts. This could also reduce problems associated with RNA diffusion, capture rate or disturbed patterns caused by unmeasured inter-bin space. Likewise, enlarging the capture area could allow the

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

measurement of whole mammalian embryos, human organs, or model organisms or the simultaneous assessment of multiple smaller samples (e.g., biopsies) on the same chip, avoiding the batch effect problem caused by separate runs and facilitating automation.

Our aim was to develop a spatially resolved transcriptomic technology that combines billions of tightly packed barcodes in known positions with large-captured areas, high capture rate and high reproducibility. We deposited DNA nanoball (DNB) containing random barcoded sequences onto a modified silicon surface (chip) photolithographically etched with a grid-patterned array of spots where the DNB are efficiently docked in (*21*), each spot being approximately 220 nm in diameter and with a center-to-center distance of 500 or 715 nm (**Fig. 1**, step 1). The use of random barcode-labeled DNB generated by rolling circle amplification allows a much larger spatial barcode pool size (4^{25}) compared with bead-based approaches and at the same time maintains sequence fidelity. The array was then microphotographed, incubated with primers and sequenced to obtain the data matrix containing the coordinate identity (CID) of every etched DNB (**Fig. 1**, step 2). Then, molecular identifiers (MID) and polyT sequence-containing oligonucleotides were ligated on each spot through hybridization with the CID (**Fig. 1**, step 3). This DNB-based strategy enables the generation of large chips containing barcoded probes at a much higher density than any other previous method. The next step consisted in the capture of tissue polyA-tailed RNA, which was achieved *in situ* by loading fresh-nitrogen frozen tissue sections onto the chip surface, followed by fixation, permeabilization and finally reverse transcription and amplification (**Fig. 1**, step 4). Amplified cDNAs were collected, used as template for library preparation and sequenced together with the CID (**Fig. 1**, step 5). Computational analysis of the sequencing data allows spatially resolved transcriptomics with a resolution of 500 or 715 nm and minimal inter-bin space (**Fig. 1**, step 6), resulting in a significantly higher number of spots per 100 μm^2 (average size of a mammalian single cell) than any other available method (**Fig. 2a**). Currently, we have used Stereo-seq chips of 50, 100 and 200 mm^2 in tissues like the mouse olfactory bulb (~11.7 mm^2), half mouse brain (~26.7 mm^2) or a whole E11.5

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(~10.5 mm^2) and E16.5 (~76.9 mm^2) mouse embryos (**fig. S1a**). Notably, we have also developed DNB patterned array chips of up to 42.25 cm^2 for potential application to large tissues such as whole human brain tissue sections (**fig. S1a, b**). We termed our approach Stereo-seq (Spatio-Temporal Enhanced REsolution Omics-sequencing),

We first profiled the mouse olfactory bulb with Stereo-seq because it is a widely used model tissue for spatially resolved transcriptomic approaches (*18, 19*). We used patterned Stereo-seq chips (715 nm resolution) in combination with DAPI staining, which confirmed that the spatial transcriptomic information reflects the cellular structure of the tissue (**Fig. 2b**). Stereo-seq captured numbers ranging on average from 9,342 genes and 34,435 transcripts per 143 μm bin (bin 200, 200 × 200 DNB) to 275 genes and 346 transcripts per 14 μm bin (bin 20, 20 × 20 DNB, equivalent to ~1 cell) (**Fig. 2c**). Our reanalysis of previously reported datasets showed that Visium captures an average of 5,735 genes and 25,598 transcripts at 100 μm resolution (mouse brain), Slide-seq 37 genes and 45 transcripts at 10 μm resolution (hippocampus) (*16*), Slide-seqV2 426 genes and 494 transcripts at 10 μm resolution (hippocampus) (*18*), and DBiT-seq 3,301 genes and 7,603 transcripts at 100 μm resolution (E10 mouse embryo) and 1,801 genes and 3,572 transcripts at 20 μm resolution (E10 mouse embryo eye/brain) (**Fig. 2c**). Gene and transcript capture of Stereo-seq per 2 μm-diameter bin (bin 3, 3 × 3 DNB, subcellular resolution) was substantially higher than HDST (mouse olfactory bulb; 2 genes and 2 transcripts at 2 μm resolution) (*19*) (**Fig. 2d**), which is so far the only sequencing-based spatially resolved transcriptomic technology with subcellular resolution. We then performed unsupervised clustering to computationally reconstruct the spatial identity of the olfactory bulb profiled with Stereo-seq. We observed clustering consistent with known anatomical structures including the outer plexiform layer (OPL), granular cell zone deep (GCL-D), granular cell layer internal (GCL-I), internal plexiform layer (IPL), subependymal zone (SEZ), olfactory nerve layer (ONL), glomerular layer (GL) and granular cell layer externa (GCL-E) (**Fig. 2e**). This also identified two distinct sublayers within the GL and five in the ONL, suggesting potential sub-regional differences. The consistency between biological replicates using two adjacent sections

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

of the mouse olfactory bulb profiled with Stereo-seq was high (**fig. S1c-f)**. In addition, there was also high correlation between similar anatomical areas in each of the olfactory bulb halves from a single section (**fig. S2a**). Analysis of reported mouse olfactory bulb datasets obtained with HDST or Slide-seqV2 did not reflect tissue organization as well as Stereo-seq based on transcript capture intensity when compared at the same resolution (10 μm), nor was there such a high degree of comparability in the captured transcripts between the two halves (**fig. S2b, c**). Similarly, the numbers of transcripts captured across the tissue were more homogeneous in Stereo-seq than HDST and Slide-seqV2 (**fig. S2a-c lower panels**), which could be related to the ultrahigh density-patterned distribution of the barcoded spots on the Stereo-seq chip. In addition, we noticed a higher number of mitochondrial gene reads captured using Slide-seqV2 than with Stereo-seq (**fig. S2d**). An excessive mitochondrial gene capture reduces the number of valid transcripts for anatomical area or cell identity characterization. Notably, the distribution of specific markers (*Pcp4*, *Slc17a7*, *Sox11* and *Th*) showed clearer patterns using Stereo-seq than HDST or Slide-seqV2 (**Fig. 2f and fig. S3a-d**), producing images with remarkable similarity to *in situ* hybridization (ISH) data of mouse brain from the Allen Brain Atlas (ABA) (*22*). Therefore, Stereo-seq has an unprecedented resolution, high sensitivity and reproducibility for dissecting spatial transcriptomic patterns in heterogeneous tissues.

Next, we sought to apply Stereo-seq to a more complex tissue such as the mouse brain to further demonstrate its applicability. Characterizing the diverse cell types of the brain in a spatially resolved manner across different brain structures is fundamental to understand normal brain function as well as the mechanisms of psychiatric and neurodegenerative disorders. Massive ISH and immunostaining projects such as ABA (*22*) and the HPA (Human Protein Atlas) Brain Atlas (*23*), respectively, were developed to answer these questions but only test one gene or protein at a time and lack perturbation context. We applied Stereo-seq (500 nm resolution) to a half mouse brain section (**Fig. 3a**). Unsupervised clustering identified 15 major spatial clusters consistent with eight anatomical structures including

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

meningeal layer 1 and 2, fiber tracts 1 and 2, neocortex 1-4, retrohippocampal region, hippocampus 1 and 2, thalamus, midbrain 1 and 2, and substantia nigra (**Fig. 3b**). Analysis of differentially expressed genes (DEG) (**fig. S4a**) and Gene Ontology (GO) analysis confirmed the unique identity of these structures (**fig. S4b**). For example, GO terms related to the vasculature were enriched in the meningeal layer and dopamine-related processes in the substantia nigra. Similarly, by integrating our Stereo-seq results with a published scRNA-seq dataset of mouse neocortex (*24*), we successfully transferred the single-cell type labeling and identified the expected layer-specific gene expression pattern (**Fig. 3c and fig. S4c**). The spatial distribution of the known markers *Cux2* (L2, L3), *Rorb* (L4, L5), *Etv1* (L5) and *Ctgf* (L6) using Stereo-seq confirmed this result (**Fig. 3d**). In contrast, the pan-neuronal marker *Mef2c* was expressed across all layers (**fig. S4d**). We studied additional cell type markers for other anatomical structures including: 1) different layers of the hippocampus (*Prox1*, *Spink8 and Hpca*), 2) retrohippocampus (*Dcn*), 3) thalamus (*Pcp4*), 4) midbrain (*Gbx2*), 5) substantia nigra (*Th*), 6) fiber tracts (*Mbp*) and 7) meninges (*Ptgds* and *Igf2*) (**Fig. 3e, f and fig. S4d**). We then zoomed into the hippocampus at single DNB resolution (500 nm), where we could differentiate the distribution of two adjacent cell types: dentate gyrus granule cells (DG-GC, *Prox1*$^{+}$) and meningeal layer vascular leptomeningeal cells (ML-VLMC, *Ptgds*$^{+}$) (**Fig. 3e**). The same resolution confirmed the differential distribution of pyramidal neurons of the cornu ammonis 1 (CA1-Ex, *Spink8*$^{+}$) in the hippocampus and oligodendrocytes (*Mbp*$^{+}$) in fiber tracts of the adjacent corpus callosum (CC-oligo) (**Fig. 3f**). These results demonstrate the ability of Stereo-seq to spatially characterize the transcriptomic pattern of complex tissues in an unbiased manner, and the ability to assign individual cell types to this information.

We also used Stereo-seq (715 nm resolution) to profile E11.5 and E16.5 mouse embryos (**fig. S5a, b**). The latter exceeds by far the maximum size captured by other spatially resolved transcriptomic techniques. Comparison of our E11.5 embryo data and an E10 dataset generated with DBiT-seq (*20*) showed remarkable differences in overall transcript capture at the same resolution (50 μm, bin 70, 70 × 70 DNB) (**fig. S5c**). This discrepancy is due to the expected variability in microfluidic flow using

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

DBiT-seq that causes a high degree of variation in transcript detection within the same tissue slice. This caveat and the significant inter-bin space (50 μm) (**fig. S5c**) complicate the study of transcriptomic gradients in the developing embryo and other tissues using DBiT-seq. Importantly, unsupervised clustering of our Stereo-seq data for E11.5 and E16.5 embryos identified structures consistent with major developing organs (e.g., central nervous system, heart, liver, skin, bone and muscle) (**Fig. 4a and fig. S5d, e**). These structures expressed marker genes consistent with such identities (**fig. S6a, b**). We then focused on the developing central nervous system and axial skeleton for a more detailed analysis. In the E11.5 embryo, we detected two major areas of neurogenesis (anterior and posterior) based on the expression of the neural progenitor transcription factor *Pax6* (*25*) (**fig. S7a**). GO analysis of the up-regulated DEG showed enrichment of terms related to neural development in the anterior region and terms related to chordate development in the posterior region (**fig. S7b, c**). Similarly, the spatial distribution of *Hox* genes correlated with known specification patterns along the entire dorsal area (*26*) (**Fig. 4b and fig. S7b, d**). Moreover, *Msx1* and *Msx2* were concentrated in the developing craniofacial area (*27*), where we also noticed high levels of Wnt pathway activity based on the accumulation of *Wnt5a* (encoding a Wnt ligand) and the β-catenin interacting transcription factor *Lef1* (*28*) (**fig. S6a**). In contrast, *Pax1* and *Pax9*, which are involved in vertebral column formation (*29*), were more prevalent in the dorsal area. The same was observed for the neurofilament genes *Nefl* and *Nefm* (*30*) (**fig. S6a**). These findings prove the utility of Stereo-seq for detecting transcription factor gradients and signaling cues in developing tissues. Likewise, in the E16.5 embryo brain we could discern structures corresponding to the developing cortex, striatum and lateral ventricle (**Fig. 4c, d**). We then integrated the Stereo-seq data of the anterior neurogenesis area in the E11.5 embryo with the developing brain of the E16.5 embryo using uniform manifold approximation and projection representation (UMAP). The biggest co-distribution was in cluster 2 (**Fig. 4e, f**), which was enriched in genes consistent with apical precursors (AP) of neuroepithelial origin and the more differentiated radial glial precursors (RGP; the source of neural stem cells) (*31*) in the E11.5 and E16.5 embryo,

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

respectively (**Fig. 3f, g and fig. S7e, f**). Analysis of the pseudotime trajectory for cluster 2 using Monocle2 showed the differentiation path from AP to RGP (**Fig. 4h**). Genes associated with the progressive acquisition of a RGP phenotype were increased along this trajectory (**Fig. 4i**). GO analysis of the DEG between both embryonic time points in cluster 2 also confirmed an ulterior state of neural development in the E16.5 embryo (**fig. S7g**). Moreover, cell cycle stage prediction demonstrated that while AP are actively proliferating RGP are entering a quiescent stage (**Fig. 4j and fig. S7g**). Hence, Stereo-seq can be used to monitor the spatio-temporal dynamics in gene expression of developing embryos and our dataset provides a unique panoramic resource for further investigation.

In summary, Stereo-seq is a spatially resolved transcriptomic technology with nanoscale resolution, centimeter level field of view, high sensitivity, minimal inter-bin space and high bin-to-bin reproducibility. These parameters are all fundamental for reliably dissecting the transcriptomic heterogeneity of complex tissues and organisms. Being based on a standard DNB sequencing approach (*21*), it is relatively simple to increase the sample throughput and systematically analyze hundreds of serial slides from the same tissue in a short time. This capacity is necessary for properly understanding the dynamics of gene expression in tissues and for achieving the ultimate goal of 3D-transcriptomic reconstructions. In this regard, the precise spatial coordinates retrieved by Stereo-seq allow optimal integration of different slides, which might not be easily achieved with accuracy by techniques using randomly distributed beads (*18, 19*). Additional optimizations of Stereo-seq will increase gene and transcript capture, facilitating both the assignment of individual cell identities without the need for scRNA-seq data and the widespread identification of subcellular transcriptomic features. The former could also be achieved through immunostaining of the Stereo-seq slide or through combination with oligonucleotide-labeled antibodies (*32, 33*) for simultaneous protein detection of surface markers, which can be easily incorporated into the DNB. Likewise, we anticipate that Stereo-seq will have other applications beyond RNA-sequencing, in particular spatially resolved epigenomics (e.g., chromatin accessibility profiling and measurement of DNA

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

methylation) and genome sequencing. Stereo-seq and its future technical variations will thus be highly transformative for multiple research fields through the generation of comprehensive healthy and diseased body atlases as well as evolution and organ developmental atlases. Furthermore, Stereo-seq has the potential to move into routine clinical practice as an extraordinary diagnostic tool complementary to medical imaging and histopathology data.

ACKNOWLEDGMENTS

We thank the MOTIC China Group CO., Ltd. for providing technical support. **Funding:** This work was supported by the Shenzhen Key Laboratory of Single-Cell Omics (ZDSYS20190902093613831) and the Guangdong Provincial Key Laboratory of Genome Read and Write (2017B030301011); Longqi Liu was supported by the National Natural Science Foundation of China (No.31900466); Miguel A. Esteban's lab at the Guangzhou Institutes of Biomedicine and Health was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16030502). **Author contributions:** A.C., L.L., Y.Li, X.X. and J.W. conceived the idea; A.C., L.L., Y. Li, X.X., J.W. and M.A.E. supervised the work; A.C., S. Liao., M.C. and K.M. designed the experiment; S. Liao, J.Y. and W.L. generated the Stereo-seq chip with the help from X.C., X.H.[3,7] and D.F.; M.C., J.X. and F.L. performed the majority of the experiments; K.M., L.W., X.L., X.T. and M. Li performed data preprocessing and quality evaluation; L.W., Y.Lai and S.H. analyzed the data; Z.L., Y.J., J.P., S.Liu, C.L., Zhi.Wang, Y.Yuan., Y.H., Q.L., Zhao.Wang, X.H.[1], F.Z., H.K., O.W., B.W. and H.L. provided technical support; G.V., C.W., , M.N., W.Z., F.M., Y.Yin, H.Y., P.M.-C., J.P.T, M.Lisby, R.J.C and M.U. gave the relevant advice; M.A.E., L.L., Y.Lai, W.L. and A.C. wrote the manuscript. **Competing interests:** The chip, procedure and applications of Stereo-seq are covered in pending patents. Employees of BGI have stock holdings in BGI.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

FIGURE LEGENDS

Figure 1. Schematic representation of the Stereo-seq procedure. Step 1, design of the DNB patterned array chip. Step 2, *in situ* sequencing to determine the spatial coordinates of uniquely barcoded oligonucleotides on the chip. Step 3, preparation of Stereo-seq capture probes by ligating the MID-polyT containing oligonucleotides to each spot. Step 4, subsequent *in situ* RNA capture from tissue placed on the Stereo-seq chip. Step 5, cDNA amplification, library construction and sequencing. Step 6, data analysis to generate the spatially resolved transcriptome of the profiled tissue.

Figure 2. Advanced *in situ* RNA capture from a complex tissue by Stereo-seq.

a. Stereo-seq achieves a higher number of spots per 100 μm^2 (left panel) and larger capture area (right panel) than other reported methods (*16, 18-20*).

b. Top: DAPI staining of a mouse olfactory bulb tissue section analyzed by Stereo-seq. Bottom: spatial heatmap of the same section indicating the number of captured transcripts at bin 50 (~36 μm resolution, 50 × 50 DNB). Scale bars 500 μm.

c. Boxplots showing the number of genes (left panel) and transcripts (right panel) captured by Stereo-seq from the mouse olfactory bulb at the indicated resolution in comparison with reported Slide-seqV2 and HDST datasets. Mouse brain data for Visium were taken from the 10x Genomics website.

d. Boxplots showing the number of genes (left panel) and transcripts (right panel) captured by Stereo-seq at bin 3 (~2 μm resolution, 3 × 3 DNB) resolution in the mouse olfactory bulb compared with HDST.

e. Unsupervised clustering of a mouse olfactory bulb section analyzed by Stereo-seq at bin 50. Scale bar 500 μm.

f. Left: spatial expression heatmap of the granule cell marker (*Pcp4*) and the mitral and tufted cell marker (*Slc17a7*) in a mouse olfactory bulb section analyzed by Stereo-seq at bin 50. Right: corresponding ISH results for these genes taken from ABA (*22*). Scale bar 500 μm.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 3. Stereo-seq dissects the spatial transcriptomic organization of mouse brain at unprecedented resolution.

a. Spatial heatmap indicating the number of transcripts captured by Stereo-seq in a mouse half brain section at bin 50. Scale bar 500 μm.

b. Unsupervised clustering of a mouse brain analyzed by Stereo-seq data identifies different anatomical structures at bin 50. The black square indicates a region of the neocortex selected for downstream analysis. Scale bar 500 μm.

c. Spatial visualization at bin 50 of cell type annotation (transferred from a published scRNA-seq dataset (*24*)) for the same area marked in **b**. Scale bar 500 μm. Astro, Astrocyte; CR, Cajal-Retzius cell; Endo, endothelial cell; L2/3, layer 2/3 excitatory neuron; L4, layer 4 excitatory neuron; L5, layer 5 excitatory neuron; L6, layer 6 excitatory neuron; Lamp5, *Lamp5*$^{+}$ inhibitory neuron; Microglia, microglia cell; Meis2, *Meis2*$^{+}$ inhibitory neuron; NP, near-projecting neuron; Oligo, oligodendrocyte; Peri, pericyte; Pvalb, *Pvalb*$^{+}$ inhibitory neuron; Serpinf1, *Serpinf1*$^{+}$ inhibitory neuron; SMC, smooth muscle cell; Sncg, *Sncg*$^{+}$ inhibitory neuron; Sst, *Sst*$^{+}$ inhibitory neuron; Vip, *Vip*$^{+}$ inhibitory neuron; VLMC, vascular leptomeningeal cell.

d. Spatial expression heatmap at bin 50 of layer-specific genes in the mouse neocortex L2/3, L4, L5 and L6 layers shown in **c**. Scale bar 500 μm.

e. Spatial expression heatmap of the indicated genes related to DG-GC in the dentate gyrus (*Prox1*, *Zbtb20* and *Rfx3*) of the hippocampus and ML-VLMC in the neighboring meningeal layer (*Ptgds*, *Dcn* and *Igf2*) analyzed by Stereo-seq at bin 50. Scale bar 500 μm. The scatter plot in the center is a magnification of the area indicated by the black square and highlights single-DNB (500 nm) that captured the indicated genes. Gray dots are DNB that did not capture the depicted marker genes (also in **f**), DNB that did not capture any genes are not shown. Scale bar 100 μm.

f. Spatial expression heatmap of the indicated genes related to oligodendrocytes (oligo) from the corpus callosum (CC-oligo; *Mbp* and *Mobp*) and CA1-Ex from the adjacent cornu ammonis 1 (*Pou3f1*, *Spink8* and *Hpca*) of the hippocampus analyzed by Stereo-seq at bin 50. Scale bars 500 and 100 μm.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 4. Stereo-seq reconstructs the spatio-temporal gene expression dynamics of the developing mouse embryo.

a. Unsupervised clustering identifies different anatomical structures in an E11.5 embryo section analyzed by Stereo-seq at bin 50. Scale bar 1 mm. The detailed annotation is shown in **fig. S5d**.

b. Spatial heatmap of the module score representing the aggregated expression level of anterior, central and posterior *Hox* family genes in an E11.5 embryo section analyzed by Stereo-seq at bin 50. Scale bar 1 mm.

c. Top: unsupervised clustering of Stereo-seq data identifies different anatomical structures in the E16.5 embryo section at bin 100 (~71 μm resolution, 100 × 100 DNB). The detailed annotation is shown in **fig. S5e**. The black square indicates the embryonic brain. Scale bar 1 mm. Bottom: re-clustering of the embryonic brain in the E16.5 embryo section marked above identifies different regions at bin 50. Scale bar 500 μm.

d. Spatial heatmap showing the expression of canonical marker genes for neocortex (*Mef2c*), striatum (*Ebf1*) and posterior horn of the lateral ventricle (*Dbi*) in the E16.5 embryonic brain analyzed by Stereo-seq at bin 50. Scale bar 500 μm.

e. UMAP projection of the integrated E11.5 mouse embryo anterior neural tube with the E16.5 embryonic brain Stereo-seq data identifies neuroepithelial populations.

f and g. UMAP projection (**f**) and spatial expression heatmap (**g**) showing the expression of neuroepithelial cell marker genes (*Fabp7* and *Vim*) in the E11.5 embryo anterior neural tube and the E16.5 embryonic brain. Scale bar 1 mm (**g** left panels) and 500 nm (**g** right panels).

h. Pseudotime trajectory analysis of the neuroepithelial cell state transition from E11.5 (AP) to E16.5 (RGP) captured by Stereo-seq.

i. Expression dynamics of representative genes (*Ptprz1*, *Fabp7* and *Mt3*) related to the AP to RGP transition captured by Stereo-seq.

j. Cell cycle state determined by cell cycle scoring of neuroepithelial cells in the E11.5 to E16.5 embryos along the pseudotime trajectory captured by Stereo-seq.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Size of stereo-seq chips and assessment of reproducibility, related to Figure 2.

a. Stereo-seq chips of different sizes ranging from 50 mm^2 to 42.25 cm^2.

b. Representative images showing a mouse olfactory bulb tissue section, a whole mouse brain tissue section and a whole E16.5 mouse embryo section placed on Stereo-seq chips. The area of each chip is highlighted above in red. MOB, mouse olfactory bulb.

c. Spatial heatmap indicating the number of transcripts captured by Stereo-seq as in **Fig. 2b** from an independent (replicate 2) mouse olfactory bulb section at bin 50. Scale bar 500 μm.

d. Unsupervised clustering as in **Fig. 2e** of an independent (replicate 2) mouse olfactory bulb section analyzed by Stereo-seq at bin 50. Scale bar 500 μm.

e. Pearson correlation coefficient ($r^2 = 0.9666$) of pseudo-bulk profiles from the two independent Stereo-seq experiments of mouse olfactory bulb at bin 50.

f. Boxplots showing the number of genes (left) and transcripts (right) captured by Stereo-seq from the two independent experiments of mouse olfactory bulb at single-DNB resolution.

Figure S2. Robustness of Stereo-seq, related to Figure 2.

a. Top: spatial heatmap of the mouse olfactory bulb indicating the number of transcripts captured by (**a**) Stereo-seq in the non-overlapping bin 14 (~10 μm resolution, 14 × 14 DNB), (**b**) a reported HDST dataset in the non-overlapping bin 5 (10 μm) (*19*) and (**c**) a reported Slide-seqV2 dataset (10 μm) (*18*). Middle: hierarchical clustering shows Pearson's correlation between similar regions of the mouse olfactory bulb halves from the same section. Bottom: line chart representing transcript intensity in the three regions highlighted by color-matched lines across the spatial heatmaps in the top panel. Scale bars 500 μm.

d. Boxplot (left panel) and spatial heatmap (middle and right panels) showing the comparison of mitochondrial gene ratio captured by Stereo-seq at bin 14 and

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Slide-seqV2. Scale bars 500 μm.

Figure S3. Spatial expression heatmaps in mouse olfactory bulb comparing different techniques, related to Figure 2.

a. Spatial expression heatmap of the indicated genes (*Slc17a7* and *Pcp4*) as in **Fig. 2f** from an independent (replicate 2) mouse olfactory bulb section. Scale bar 500 μm.

b. Left and middle: spatial expression heatmap of the immature neuron marker (*Sox11*) and periglomerular cell marker (*Th*) in two independent Stereo-seq experiments of mouse olfactory bulb at bin 50. Right: corresponding ISH results for the same genes taken from ABA (*22*).

c. Spatial expression heatmap of the indicated genes (*Pcp4*, *Slc17a7*, *Sox11* and *Th*) in a previously reported mouse olfactory bulb dataset profiled with HDST (*19*) and (**d**) in Slide-seqV2 (*18*). Scale bar set at 500 μm.

Figure S4. Stereo-seq detects different anatomical structures in the mouse brain, related to Figure 3.

a. Heatmap showing the mean expression level of DEG between the indicated anatomical structures of the mouse brain identified by Stereo-seq in **Fig. 3b**. ML, meningeal layer. RHP, retrohippocampus. FT, fiber tracts. HP, hippocampus. MB, midbrain. SN, substantia nigra.

b. GO analysis of DEG detected between the indicated anatomical structures identified by Stereo-seq in **Fig. 3b**. Selected GO terms are highlighted (Benjamini-Hochberg corrected *P* value).

c. Heatmap showing the mean expression level of the indicated DEG between different layers of the mouse neocortex identified by Stereo-seq in **Fig. 3c**. 3110035E*: 3110035E14Rik.

d. Spatial heatmap showing the expression of selected genes (*Igf2*, *Mef2c*, *Dcn*, *Mobp*, *Hpca*, *Pcp4*, *Gbx2* and *Th*) representing different anatomical structures of a mouse half brain section analyzed by Stereo-seq at bin 50. Scale bar set at 500 μm.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure S5. Stereo-seq detects different developing structures in mouse embryos at different stages, related to Figure 4.

a. Hematoxylin and eosin staining of an E11.5 embryo section. Scale bar 1 mm.

b. Hematoxylin and eosin staining of an E16.5 embryo section. Scale bar 1 mm.

c. Spatial heatmap indicating the number of transcripts captured by Stereo-seq of an E11.5 embryo at bin 70 (~50 μm resolution, 70 × 70 DNB) or a previously reported DBiT-seq dataset (50 μm) (*20*). The line chart in the bottom panel represents the transcript intensity in the three regions highlighted by color-matched lines in the top panel.

d. Unsupervised clustering identifies different developing structures in an E11.5 mouse embryo section analyzed by Stereo-seq at bin 50. Scale bar 1 mm.

e. Unsupervised clustering identifies different developing structures in an E16.5 mouse embryo section analyzed by Stereo-seq at bin 100. Scale bar 1 mm.

Figure S6. Stereo-seq resolves the developmental structure-specific gene expression patterns in mouse embryos, related to Figure 4.

a. Spatial expression heatmap of the indicated genes representing different developing structures in an E11.5 embryo section analyzed by Stereo-seq at bin 50. Scale bar 1 mm.

b. Spatial expression heatmap of the indicated genes representing different developing structures in an E16.5 embryo section analyzed by Stereo-seq at bin 100. Scale bar 1 mm.

Figure S7. Stereo-seq captures dynamic gene expression patterns in developing mouse embryos, related to Figure 4.

a. Spatial expression heatmap of *Pax6* indicating the anterior and posterior neural tube of an E11.5 embryo section analyzed by Stereo-seq at bin 50. Scale bar 1 mm.

b. Volcano plot showing DEG between the anterior and posterior neural tube regions shown in **a**. Differentially expressed *Hox* family genes are highlighted.

c. GO analysis of DEG between the anterior and posterior neural tube regions in **a**.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Selected GO terms are highlighted (Benjamini-Hochberg corrected *P* value).

d. Spatial expression heatmap showing the indicated *Hox* family genes representing anterior (*Hoxa3*), central (*Hoxa7*) and posterior (*Hoxc10*) regions in an E11.5 embryo section analyzed by Stereo-seq at bin 50.

e. UMAP projection of selected gene expression in the integrated E11.5 anterior neural tube and E16.5 embryonic brain data as shown in **Fig 4e.**

f. Spatial expression heatmap indicating the neuroepithelial cells (*Sox2* and *Pax6*) in the E11.5 embryo (left, scale bar 1 mm) and E16.5 embryonic brain (right, scale bar 500 μm) analyzed by Stereo-seq at bin 50.

g. GO analysis of DEG between the E11.5 and E16.5 neuroepithelial cells detected by Stereo-seq. Selected GO terms are highlighted (Benjamini-Hochberg corrected *P* value).

REFERENCES AND NOTES

1. A. Tanay, A. Regev, Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331-338 (2017).
2. J. E. Park *et al.*, A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, eaay3224 (2020).
3. J. A. Harris *et al.*, Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195-202 (2019).
4. M. Hagemann-Jensen *et al.*, Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nature biotechnology* **38**, 708-714 (2020).
5. L. Zhu *et al.*, Single-Cell Sequencing of Peripheral Mononuclear Cells Reveals Distinct Immune Response Landscapes of COVID-19 and Influenza Patients. *Immunity* **53**, 685-696 e683 (2020).
6. A. L. Ji *et al.*, Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* **182**, 497-514 e422 (2020).
7. Y. Sun *et al.*, Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. *Cell*, (2020). *Epub ahead of print*.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

8. W. T. Chen *et al.*, Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. *Cell* **182**, 976-991 e919 (2020).

9. E. Lein, L. E. Borm, S. Linnarsson, The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* **358**, 64-69 (2017).

10. S. Nichterwitz *et al.*, Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nature communications* **7**, 12139 (2016).

11. G. Peng *et al.*, Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature* **572**, 528-532 (2019).

12. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).

13. C. L. Eng *et al.*, Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235-239 (2019).

14. J. H. Lee *et al.*, Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360-1363 (2014).

15. X. Wang *et al.*, Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).

16. S. G. Rodriques *et al.*, Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467 (2019).

17. P. L. Stahl *et al.*, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82 (2016).

18. R. R. Stickels *et al.*, Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature biotechnology*, (2020). *Epub ahead of print*.

19. S. Vickovic *et al.*, High-definition spatial transcriptomics for in situ tissue profiling. *Nature methods* **16**, 987-990 (2019).

20. Y. Liu *et al.*, High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **183**, 1665-1681 e1618 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

21. R. Drmanac *et al.*, Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78-81 (2010).

22. E. S. Lein *et al.*, Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168-176 (2007).

23. E. Sjostedt *et al.*, An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367**, eaay5947 (2020).

24. B. Tasic *et al.*, Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature neuroscience* **19**, 335-346 (2016).

25. X. Zhang *et al.*, Pax6 is a human neuroectoderm cell fate determinant. *Cell stem cell* **7**, 90-100 (2010).

26. P. Philippidou, J. S. Dasen, Hox genes: choreographers in neural development, architects of circuit organization. *Neuron* **80**, 12-34 (2013).

27. M. Medio *et al.*, Wnt/beta-catenin signaling and Msx1 promote outgrowth of the maxillary prominences. *Frontiers in physiology* **3**, 375 (2012).

28. F. Mu *et al.*, The Wnt/beta-Catenin/LEF1 Pathway Promotes Cell Proliferation at Least in Part Through Direct Upregulation of miR-17-92 Cluster. *Frontiers in genetics* **10**, 525 (2019).

29. H. Peters *et al.*, Pax1 and Pax9 synergistically regulate vertebral column development. *Development* **126**, 5399-5408 (1999).

30. A. Laser-Azogui, M. Kornreich, E. Malka-Gibor, R. Beck, Neurofilament assembly and function during neuronal development. *Current opinion in cell biology* **32**, 92-101 (2015).

31. S. A. Yuzwa *et al.*, Developmental Emergence of Adult Neural Stem Cells as Revealed by Single-Cell Transcriptional Profiling. *Cell reports* **21**, 3970-3986 (2017).

32. M. Stoeckius *et al.*, Simultaneous epitope and transcriptome measurement in single cells. *Nature methods* **14**, 865-868 (2017).

33. V. M. Peterson *et al.*, Multiplexed quantification of proteins and transcripts in single cells. *Nature biotechnology* **35**, 936-939 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

MATERIALS AND METHODS

Tissue collection

All relevant procedures involving animal experiments presented in this paper are in compliance with ethical regulations regarding animal research and were conducted under the approval of the Animal Care and Use committee of the Guangzhou Institutes of Biomedicine and Health (license number IACUC2021002). Mouse olfactory bulb and brain were dissected from 3-week-old C57BL/6J female mice. E11.5 and E16.5 embryos were collected from pregnant C57BL/6J female mice. After collection, tissues were snap-frozen in liquid nitrogen prechilled isopentane in Tissue-Tek OCT (Sakura, 4583) and transferred to a -80℃ freezer for storage before the experiment. Cryosections were cut at a thickness of approximately 10 μm in a Leika CM1950 cyrostat; mouse olfactory bulb and mouse brain were cut coronally, mouse embryos were cut sagittally.

Stereo-seq chip preparation

Generation of Stereo-seq capture chips. To generate the DNB array for *in situ* RNA capture, we first synthesized random 25-nt CID-containing oligos. These oligos were circularized with T4 ligase and splint oligos. DNB were then generated by rolling circle amplification and were loaded onto the patterned chips (65 mm × 65 mm). Next, to determine the distinct DNB-CID sequences at each spatial location, single-end sequencing was performed using sequencing primers in a MGI DNBSEQ-Tx sequencer with sequencing strategy SE25. After sequencing, polyT and 10 bp MID-containing oligos were hybridized and ligated to the DNB on the chip. This procedure produces capture probes containing a 25 bp CID barcode, a 10 bp MID and a 22 bp polyT ready for *in situ* capture.

Calling of CID. CID sequences together with their corresponding coordinates for each DNB were determined using a base calling method according to manufacturer's instruction of DNBSEQ™ sequencer. After sequencing, the capture chip was split into smaller size chips (5 mm × 10 mm, 10 mm × 10 mm, 10 mm × 20 mm). At this stage, we filtered out all duplicated CID that corresponded to non-adjacent spots.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Stereo-seq library preparation and sequencing

Tissue processing. Tissue sections were adhered to the Stereo-seq chip surface and incubated at 37℃ for 3 minutes. Then, tissues were fixed in methanol and incubated for 30 minutes at -20℃, after which they were ready to be used for Stereo-seq. Where indicated, the same sections were stained with DAPI, while tissue sections adjacent to the ones used for Stereo-seq were subjected to tissue histology examination using hematoxylin and eosin staining. DAPI and hematoxylin and eosin staining were performed according to standard procedure. Imaging for both procedures was performed with a Ti-7 Nikon Eclipse microscope.

In situ reverse transcription. Tissue sections placed on the chip were permeabilized using 0.1% pepsin (Sigma, P7000) in 0.01 M HCl buffer (pH = 2), incubated at 37℃ for 12 minutes and then washed with 0.1x SSC buffer (Thermo, AM9770) supplemented with 0.05 U/μl RNase inhibitor (NEB, M0314L). RNA released from the permeabilized tissue and captured by the DNB was reverse transcribed overnight at 42℃ using SuperScript II (Invitrogen, 18064-014, 10 U/μL reverse transcriptase, 1 mM dNTPs, 1 M betaine solution PCR reagent, 7.5 mM $MgCl_2$, 5 mM DTT, 2 U/μL RNase inhibitor, 2.5 μM Stereo-TSO and 1x First-Strand buffer). After reverse transcription, tissue sections were washed twice with 0.1x SSC buffer and digested with Tissue Removal buffer (10 mM Tris-HCl, 25 mM EDTA, 100 mM NaCl, 0.5% SDS) at 37℃ for 30 minutes. cDNA-containing chips were then subjected to Exonuclease I (NEB, M0293L) treatment for 1 hour at 37℃ and were finally washed once with 0.1x SSC buffer.

Amplification. The resulting cDNAs were amplified with KAPA HiFi Hotstart Ready Mix (Roche, KK2602) with 0.8 μM cDNA-PCR primer. PCR reactions were conducted as: first incubation at 95℃ for 5 minutes, 15 cycles at 98℃ for 20 seconds, 58℃ for 20 seconds, 72℃ for 3 minutes and a final incubation at 72℃ for 5 minutes.

Library construction and sequencing. The concentrations of the resulting PCR products were quantified by Qubit™ dsDNA Assay Kit (Thermo, Q32854). A total of 20 ng of DNA were then fragmented with *in-house* Tn5 transposase at 55℃ for 10 minutes, after which the reactions were stopped by the addition of 0.02% SDS buffer

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and gently mixing at 37℃ for 5 minutes after fragmentation. Fragmentation products were amplified as described below: 25 μL of fragmentation product, 1x KAPA HiFi Hotstart Ready Mix and 0.3 μM Stereo-Library-F primer, 0.3 μM Stereo-Library-R primer in a total volume of 100 μl with the addition of nuclease-free H_2O. The reaction was then run as: 1 cycle of 95℃ 5 minutes, 13 cycle of (98℃ 20 seconds, 58℃ 20 seconds and 72℃ 30 seconds) and 1 cycle 72℃ 5 minutes. PCR products were purified using the Ampure XP Beads (Vazyme, N411-03) (0.6x and 0.15x), used for DNB generation and finally sequenced (paired-end 50 bp or paired-end 100 bp) on a MGI DNBSEQ-Tx sequencer.

Stereo-seq data analysis

Raw data processing. Fastq files were generated using a MGI DNBSEQ-Tx sequencer. CID and MID are contained in the forward reads (CID: 1-25bp, MID: 26-35bp) while the reverse reads consist of the cDNA sequences. CID sequences on the forward reads were first mapped to the designed coordinates of the *in situ* captured chip, allowing 1 base mismatch to correct for sequencing and PCR errors. Reads with MID containing either N bases or more than 2 bases with quality score lower than 10 were filtered out. CID and MID associated with each read were appended to each read header. Retained reads were then aligned to the reference genome (mm10) using STAR (*34*) and mapped reads with MAPQ ≥10 were counted and annotated to their corresponding genes using an *in-house* script (available at https://github.com/BGIResearch/handleBam). MID with the same CID and the same gene locus were collapsed, allowing 1 mismatch to correct for sequencing and PCR errors. Finally, this information was used to generate a CID-containing expression profile matrix.

Unsupervised clustering of Stereo-seq data. Expression profile matrix was divided into non-overlapping bins covering an area of X × X DNB, with X ∈ (*1, 3, 14, 50, 70, 100*) and the transcripts of the same gene aggregated within each bin. After this step, data normalization and unsupervised clustering were performed using Seurat (*35*). In brief, we first used the SCTransform function for normalization and identification of

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

highly variable genes. Then, we applied the runPCA function for the dimension reduction. With these settings we ran the runUMAP function to obtain two-dimensional data projections with the following parameter: dims = 15, followed by the FindNeighbors and the FindClusters functions to identify all clusters within the dataset. To assess the identity of each cluster, we determined the marker genes for each of them using the FindAllMarkers function with the following parameters: only.pos = TRUE, min.pct = 0.1, logfc.threshold = 0.1. Finally, GO analysis was performed using clusterProfiler (*36*).

Deconvolution of cell types within each bin of Stereo-seq. We applied SPOTlight software (*37*) to deconvolute and map all cell types in the murine neocortex. Mouse visual cortex scRNA-seq data were retrieved from GSE71585 (*24*) and analyzed using Seurat with SCTransform normalization, while the FindAllMarkers function was used to identify the positive markers for each cluster with the following parameters: logfc.threshold = 1 and min.pct = 0.9. We chose the cell type with the highest probability out of all cell types provided by spotlight_deconvolution function with commendatory parameters as the final cell type for each bin.

Comparison of Stereo-seq with other published methods. We chose three y coordinates of each mouse olfactory bulb sample from Stereo-seq, HDST (*19*) and Slide-seqV2 (*18*) to obtain the transcript captured along the horizontal lines (as displayed in **fig. S2a-c** and **fig**. **S5c**). transcript captured was displayed in a line chart in which the x axis corresponds to x coordinates of binned CID while the y axis corresponds to the transcript captured of the bin. To ensure that proper comparisons were made, the data of Stereo-seq were binned into bin 14 (~10 μm) and the data from HDST were binned into bin 5 (10 μm). Six regions of Stereo-seq, Slide-seqV2 and HDST were selected from each dataset and subjected to correlation analysis by determining the coefficient of Pearson's correlation (r^2) among those regions. Mouse olfactory bulb data were used to compare the mitochondrial gene ratio between Stereo-seq and Slide-seqV2 by applying the PercentageFeatureSet function from Seurat. For Stereo-seq and DBiT-seq (*20*) comparison as above, we applied the same strategy but at a lower resolution (50 μm, equivalent to Stereo-seq bin 70).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Integrated analysis of E11.5 anterior neural tube and E16.5 embryonic brain Stereo-seq data. Bin 50 data corresponding to the anterior neural tube (E11.5) and the embryonic brain (E16.5) were extracted, SCTransform normalized, integrated and subjected to unsupervised clustering using the default Seurat parameters. After this, we identified the DEG between anterior neural tube and embryonic brain in cluster 2 by running the FindMarker function with the following parameters: max.cells.per.ident = 89, min.pct = 0.1, pseudocount.use = 0.1, logfc.threshold = 0.1 and assay = "RNA".

Trajectory analysis of neuroepithelial cells. To define the developmental trajectory of neuroepithelial cells from E11.5 to E16.5 mouse embryos, we performed pseudotemporal ordering of the cells by Monocle2 (*38*). In brief, after identifying the variable features in the normalized expression profile matrixes of the integrated data, DDRTree method was used to construct the pseudotime developmental trajectories. Cell cycle analysis along the pseudotime trajectory was performed using the CellCycleScoring function of Seurat.

REFERENCES AND NOTES

34. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

35. T. Stuart *et al.*, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).

36. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).

37. M. Elosua, P. Nieto, E. Mereu, I. Gut, H. Heyn, SPOTlight: Seeded NMF regression to Deconvolute Spatial Transcriptomics Spots with Single-Cell Transcriptomes. *bioRxiv*, 2020.2006.2003.131334 (2020).

38. C. Trapnell *et al.*, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology* **32**,

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

381-386 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

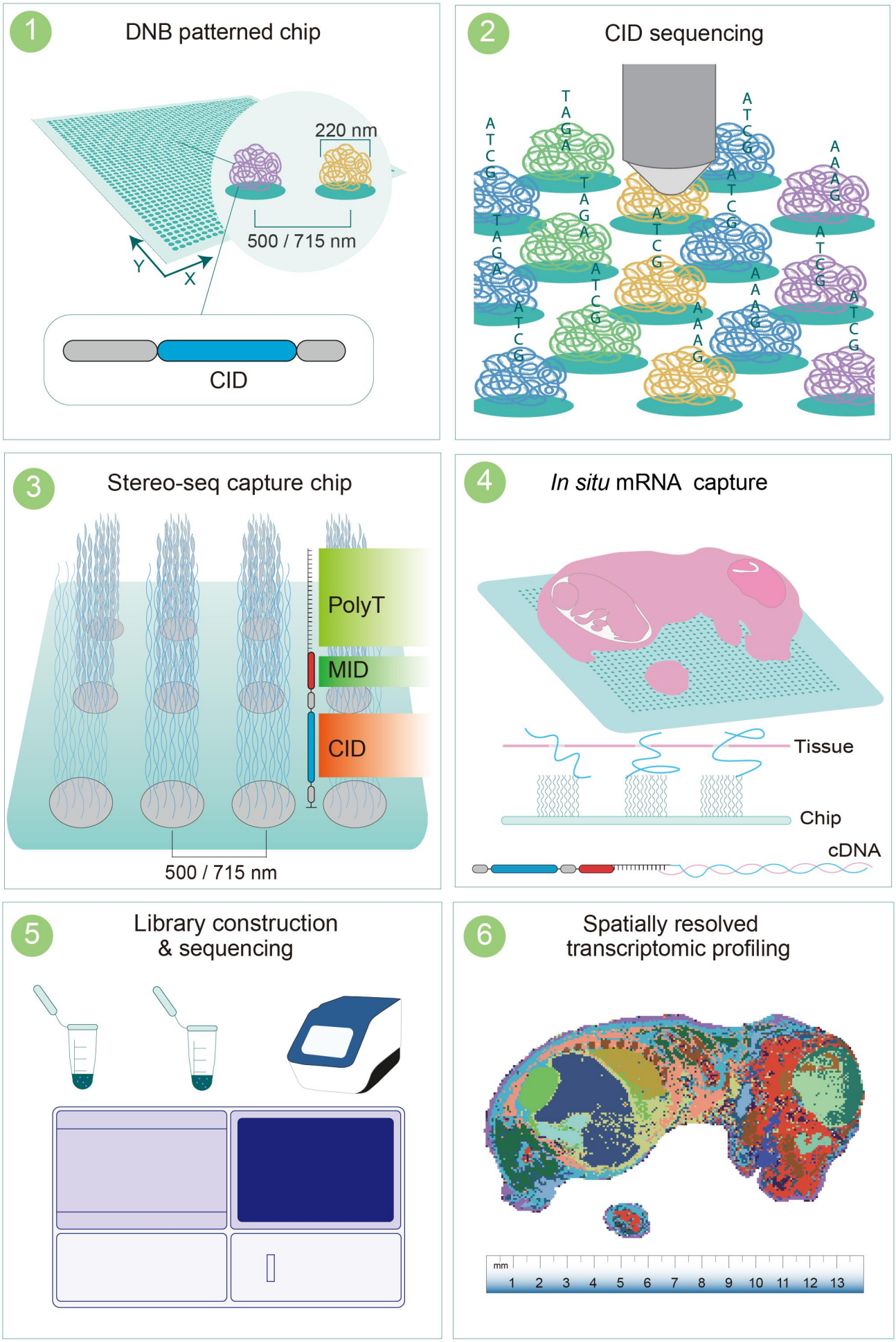


Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

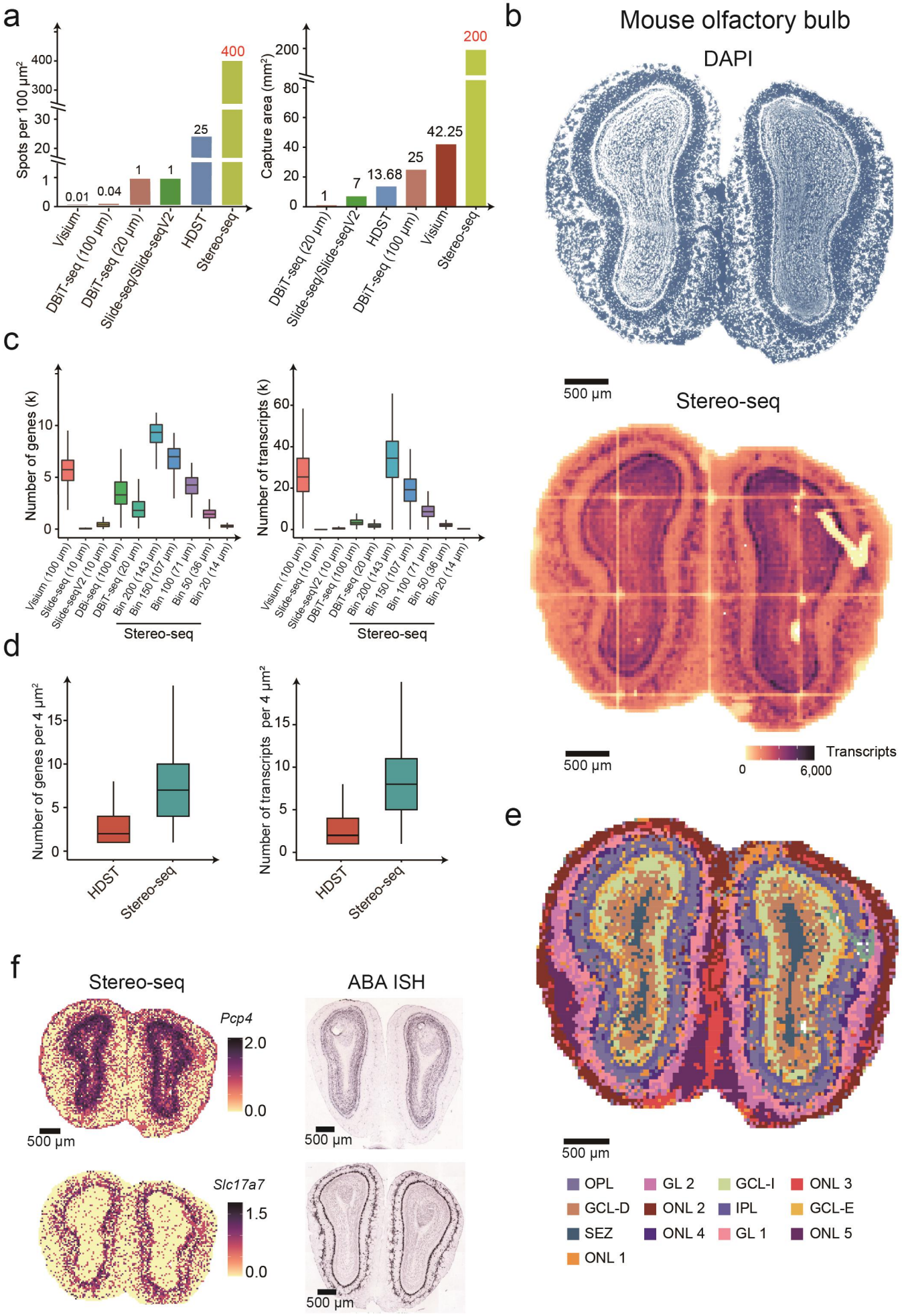


Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

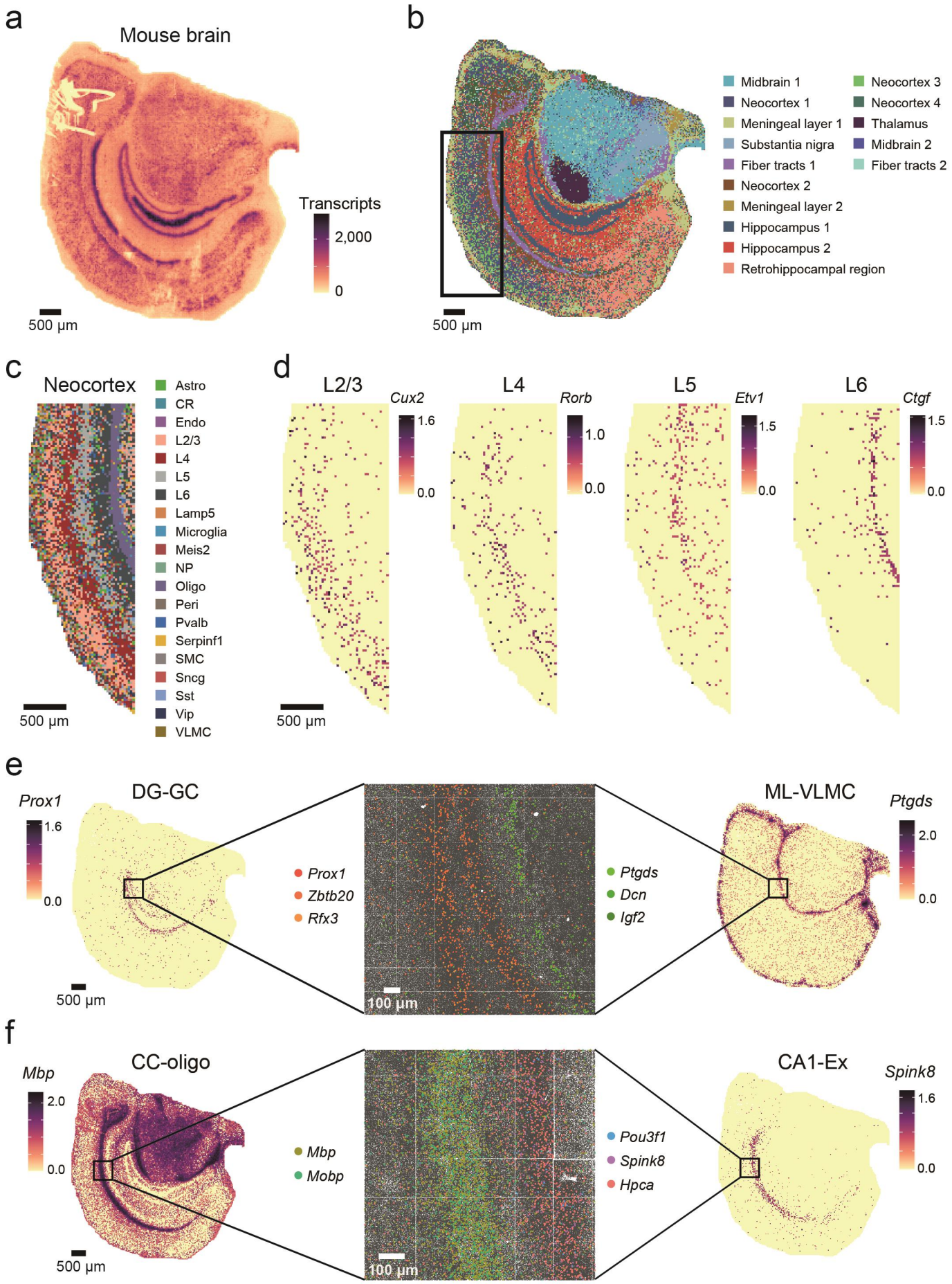


Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

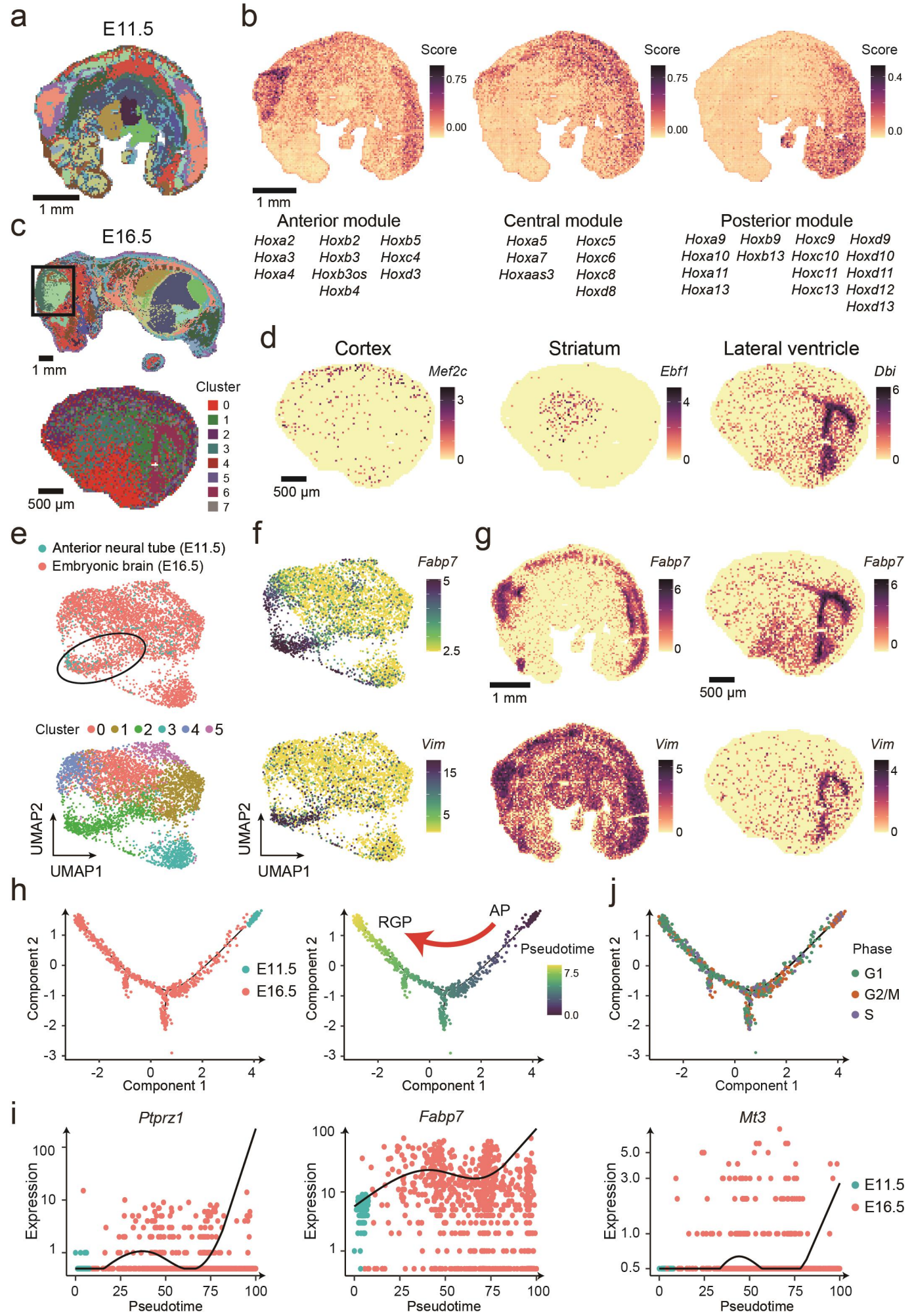
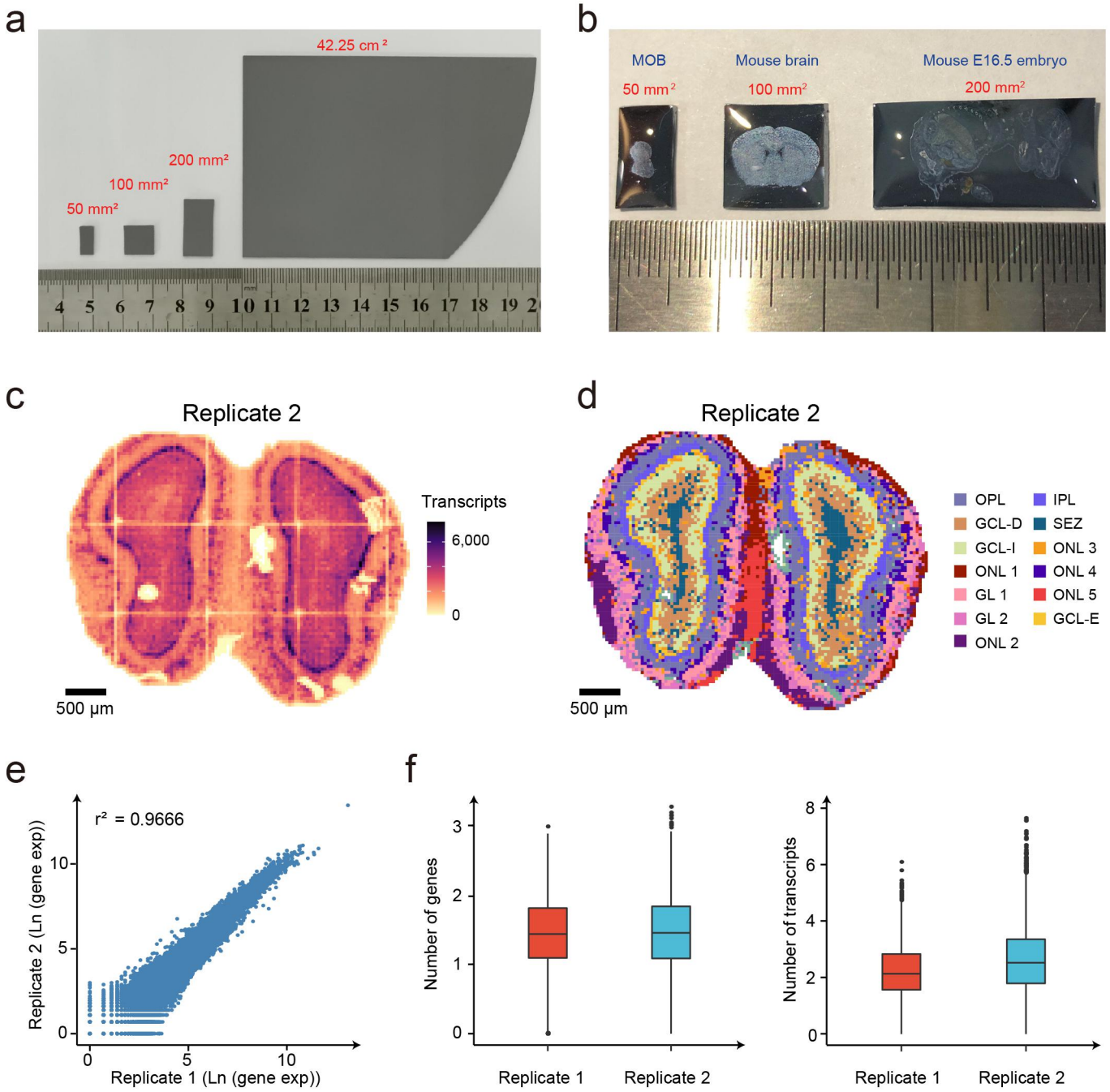


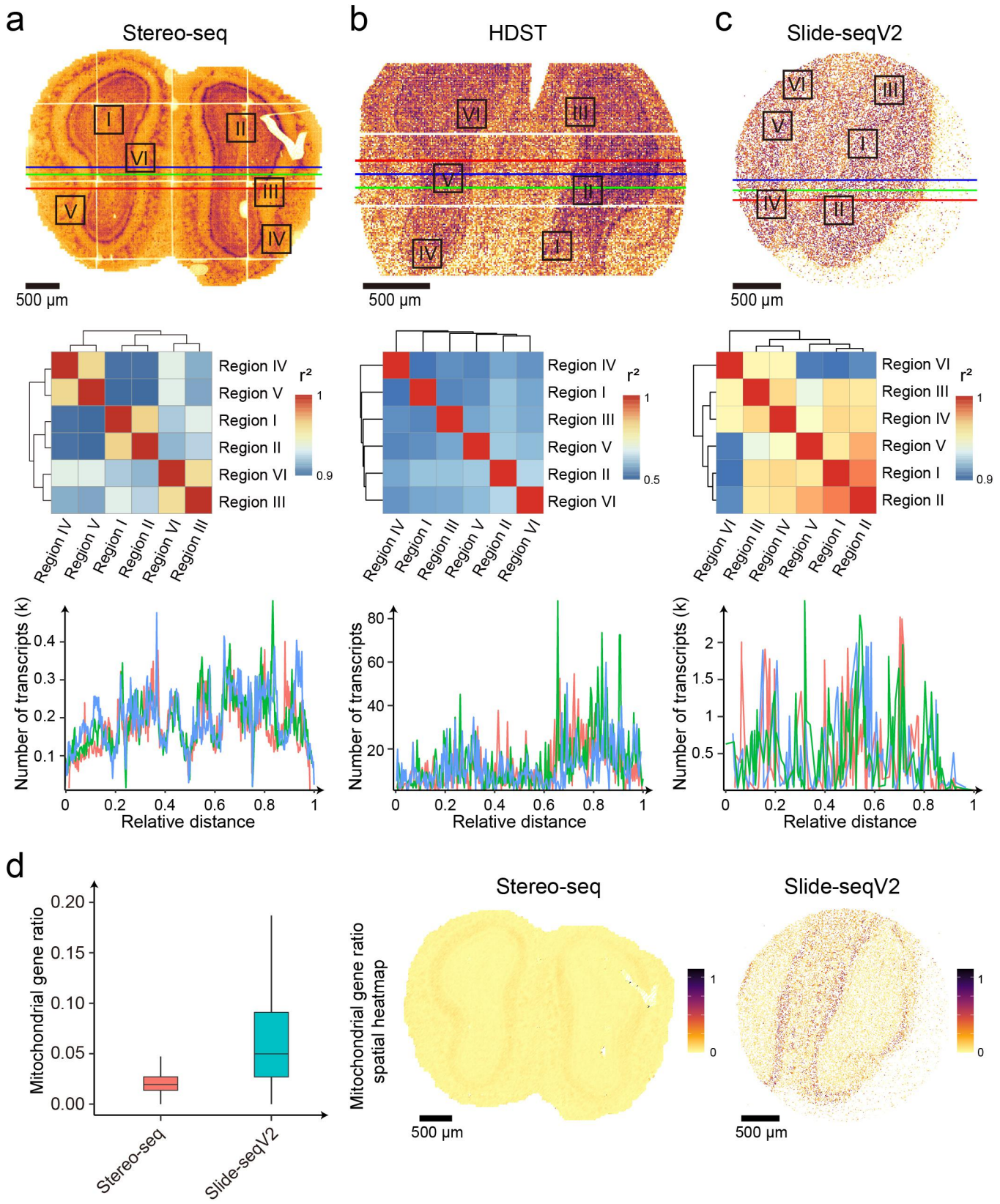
Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



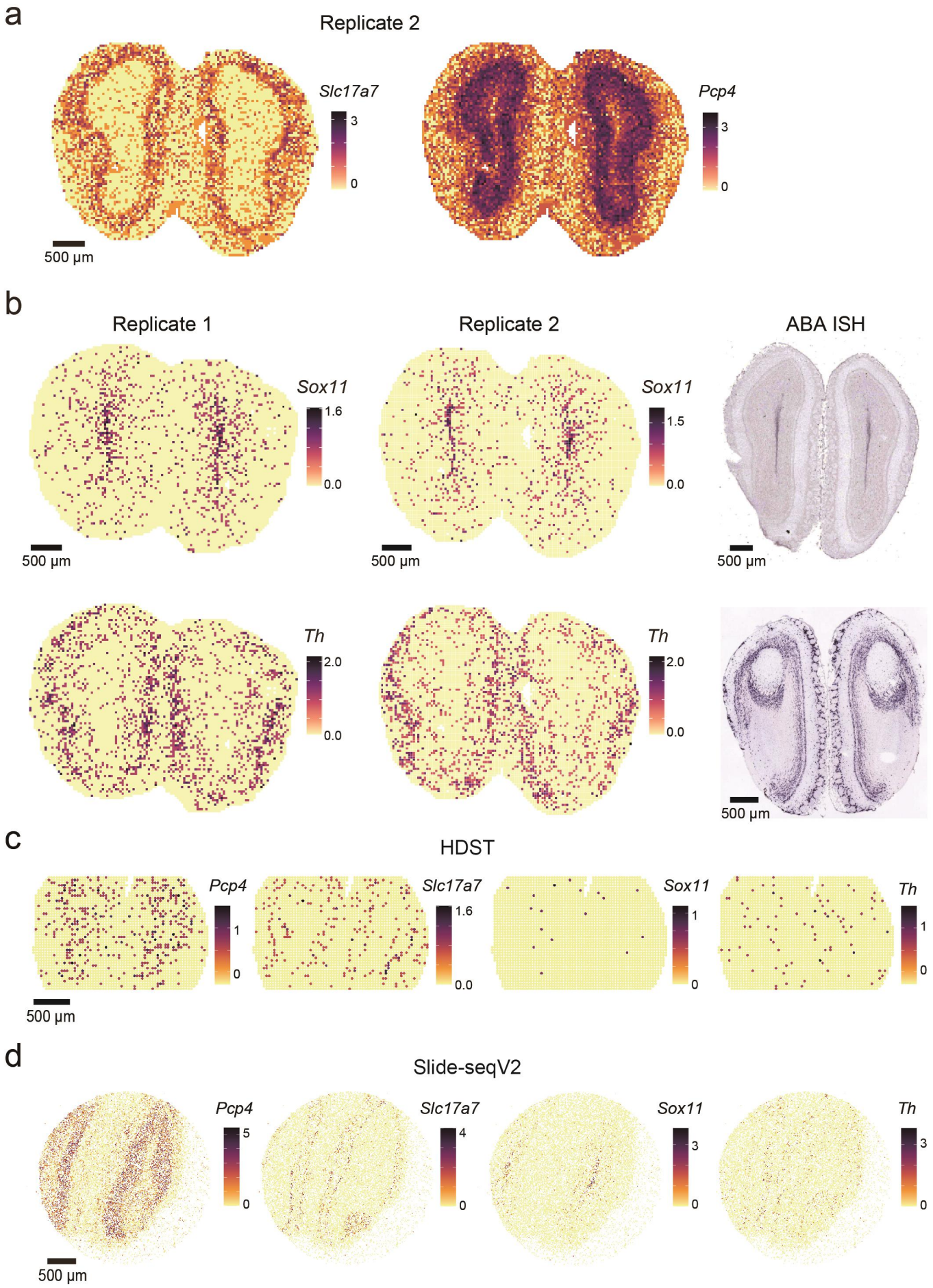
Supplementary Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



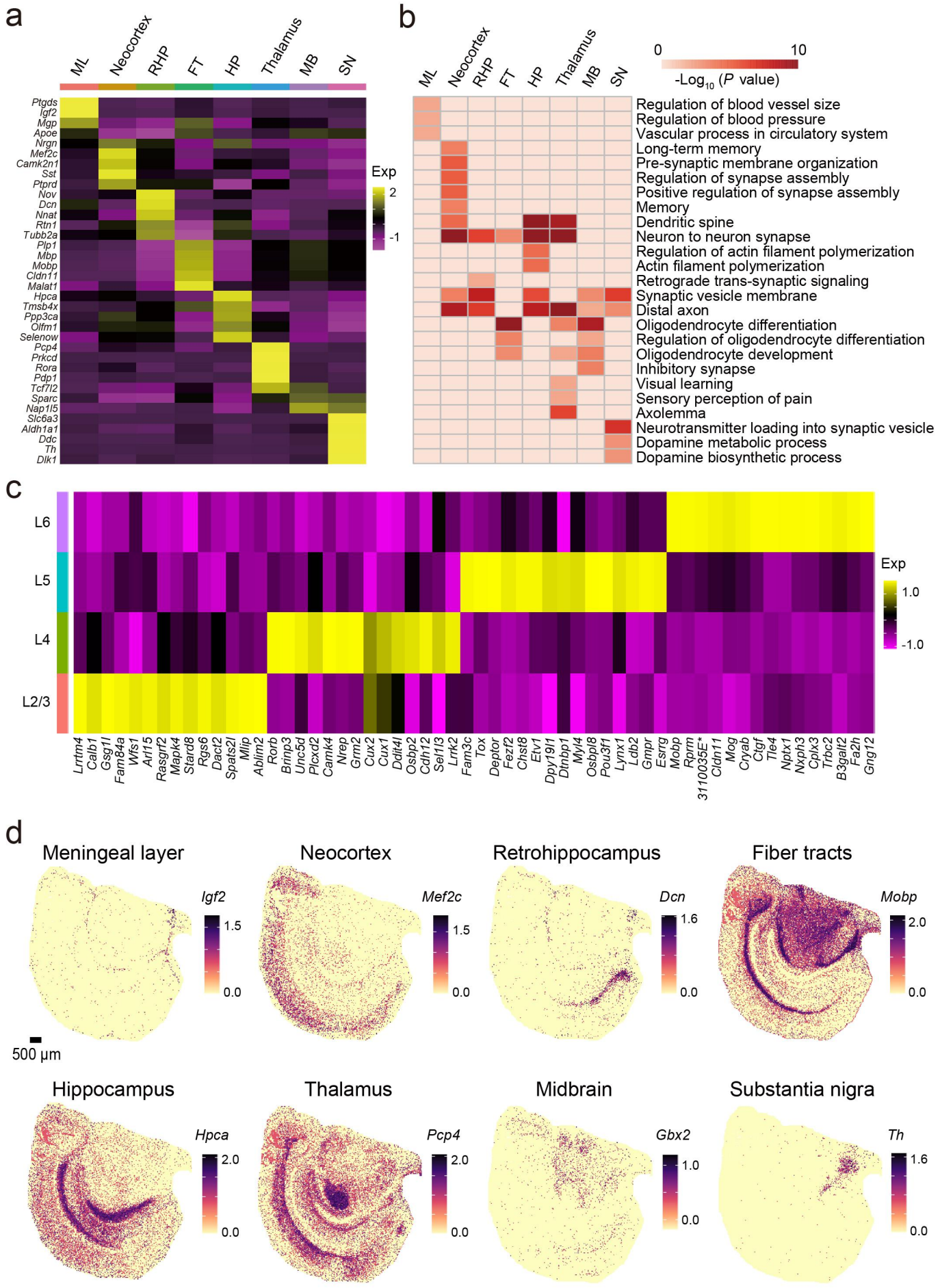
Supplementary Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



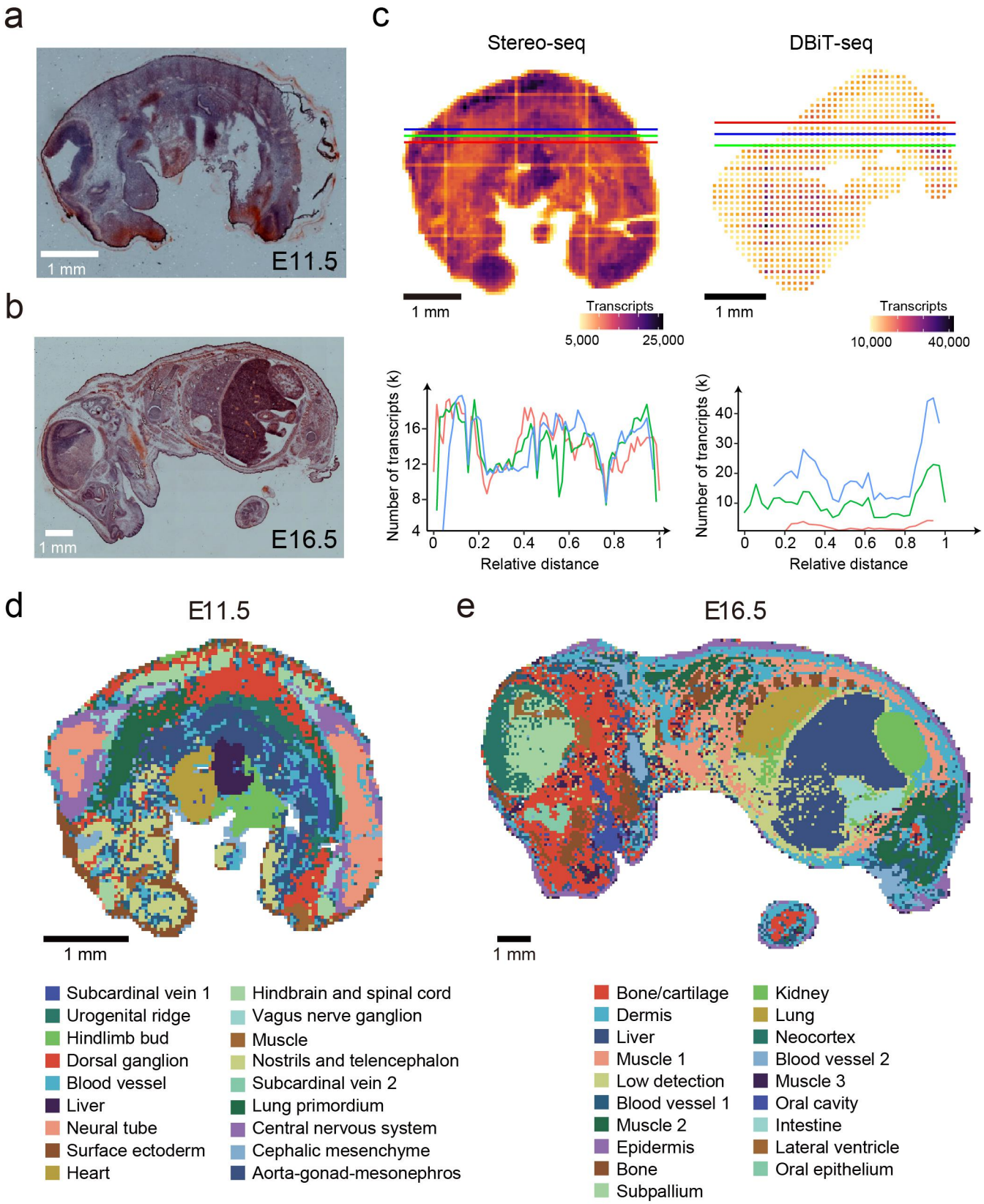
Supplementary Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



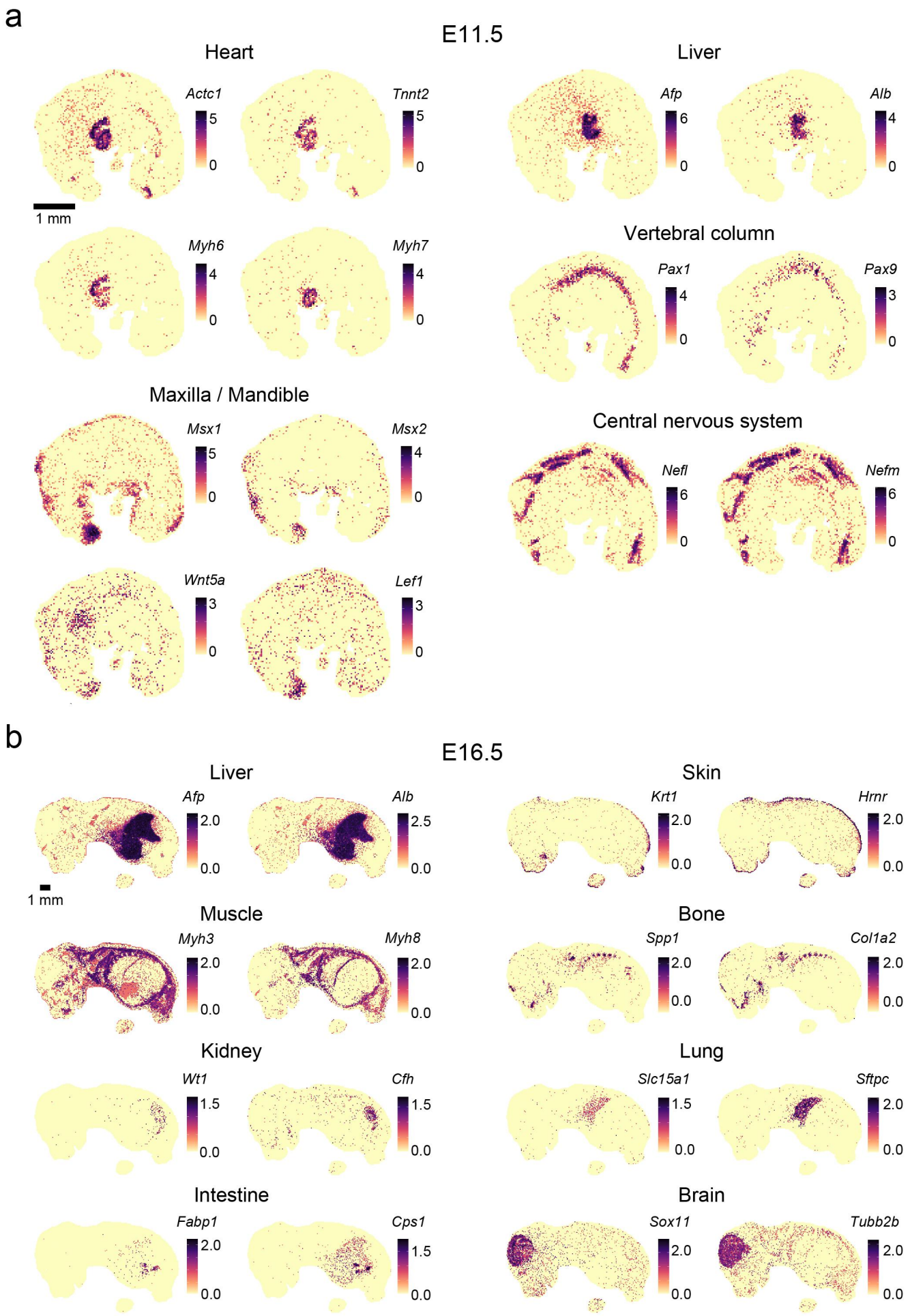
Supplementary Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



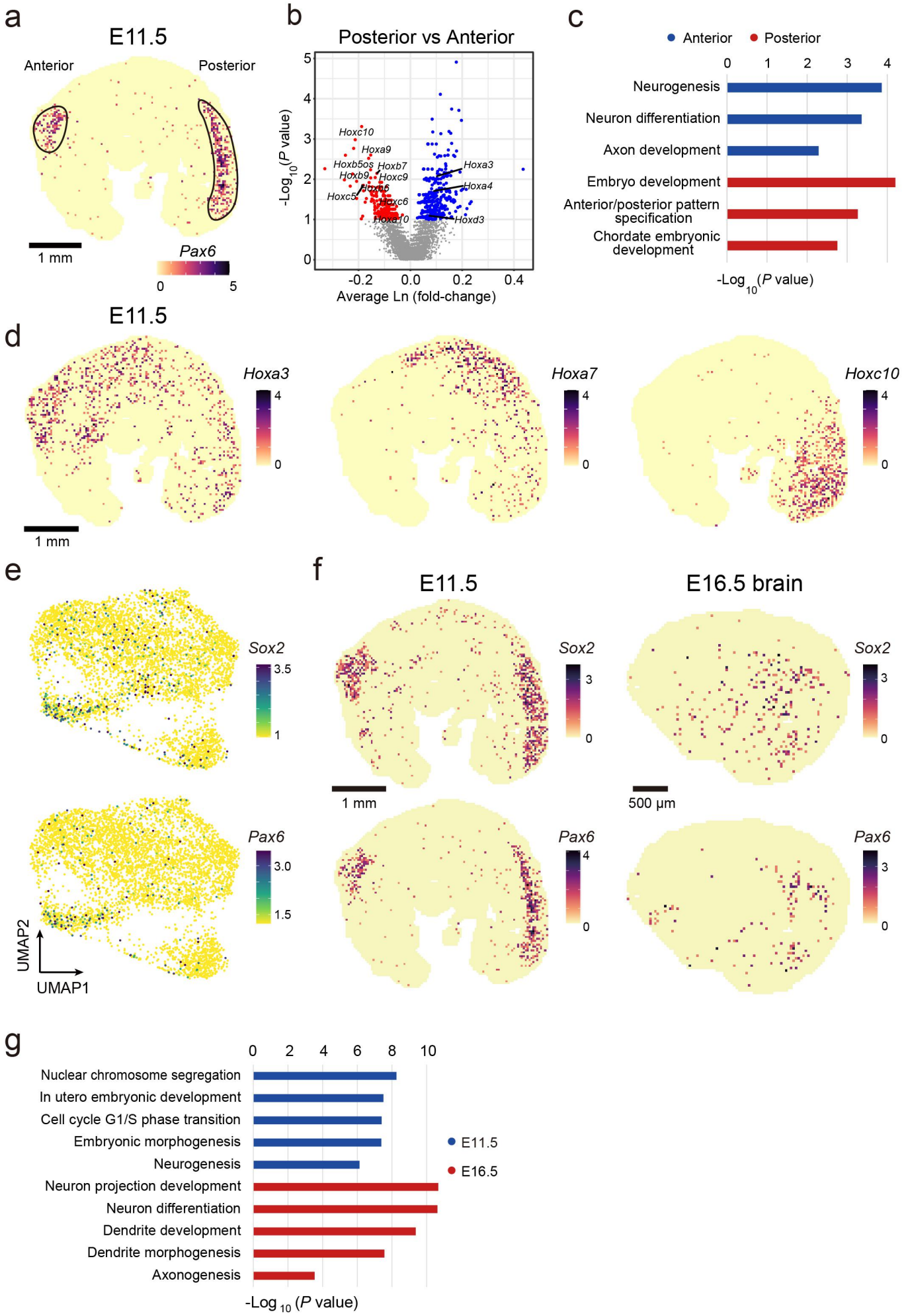
Supplementary Figure 5

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



Supplementary Figure 6

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.17.427004; this version posted January 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



Supplementary Figure 7