

18. T. Alerstam, D. Christie, A. Ulfstrand, *Bird Migration* (Cambridge Univ. Press, 1993).
19. F. Liechti, W. Witvliet, R. Weber, E. Bächler, *Nat. Commun.* **4**, 2554 (2013).
20. N. C. Rattenborg, *Naturwissenschaften* **93**, 413–425 (2006).
21. J. A. Lesku et al., *Science* **337**, 1654–1658 (2012).
22. H. Weimerskirch, M. Louzao, S. de Grissac, K. Delord, *Science* **335**, 211–214 (2012).
23. U. C. Mohanty, M. Mohapatra, O. P. Singh, B. K. Bandyopadhyay, L. S. Rathore, *Monitoring and Prediction of Tropical Cyclones in the Indian Ocean and Climate Change* (Springer, Dordrecht, Netherlands, 2014).
24. H. J. J. Jonker, T. Heus, P. P. Sullivan, *Geophys. Res. Lett.* **35**, L07810 (2008).

ACKNOWLEDGMENTS

The tracking data presented in the paper are available from the Dryad Digital Repository. We thank the Forces Armées de la Zone Sud de l'Océan Indien for transport and logistical support on Europa Island and the TAAF Administration for allowing us to work on Europa Island. We thank the fieldworkers involved in the study on Europa, in particular J. B. Pons and R. Weimerskirch; R. Spivey for help with preparing the electrocardiogram and acceleration tags and for the data processing of the heart rate recording; and A. Corbeau for help with data analyses. The study is a contribution to the Program EARLYLIFE funded by a European Research Council Advanced Grant under the European Community's Seven Framework Program FP7/2007–2013 (grant agreement ERC-2012-ADG_20120314 to

H.W.). We thank Y. Ropert-Coudert, Y. Cherel, and two anonymous reviewers for helpful comments on earlier versions of the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/353/6294/74/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Table S1
References (25–34)

11 February 2016; accepted 20 May 2016
10.1126/science.aaf4374

TRANSCRIPTION

Visualization and analysis of gene expression in tissue sections by spatial transcriptomics

Patrik L. Ståhl,^{1,2*} Fredrik Salmén,^{2†} Sanja Vickovic,^{2,†} Anna Lundmark,^{2,3,‡} José Fernández Navarro,^{1,2} Jens Magnusson,¹ Stefania Giacomello,² Michaela Asp,² Jakub O. Westholm,⁴ Mikael Huss,⁴ Annelie Mollbrink,² Sten Linnarsson,⁵ Simone Codeluppi,^{5,6} Åke Borg,⁷ Fredrik Pontén,⁸ Paul Igor Costea,² Pelin Sahlén,² Jan Mulder,⁹ Olaf Bergmann,¹ Joakim Lundberg,^{2,‡} Jonas Frisén¹

Analysis of the pattern of proteins or messenger RNAs (mRNAs) in histological tissue sections is a cornerstone in biomedical research and diagnostics. This typically involves the visualization of a few proteins or expressed genes at a time. We have devised a strategy, which we call “spatial transcriptomics,” that allows visualization and quantitative analysis of the transcriptome with spatial resolution in individual tissue sections. By positioning histological sections on arrayed reverse transcription primers with unique positional barcodes, we demonstrate high-quality RNA-sequencing data with maintained two-dimensional positional information from the mouse brain and human breast cancer. Spatial transcriptomics provides quantitative gene expression data and visualization of the distribution of mRNAs within tissue sections and enables novel types of bioinformatics analyses, valuable in research and diagnostics.

Tissue transcriptomes are typically studied by RNA-sequencing (RNA-seq) (*1*) of homogenized biopsies, which results in an averaged transcriptome and loss of spatial information. The positional context of gene expression is of key importance to understand-

ing tissue functionality and pathological changes. Several strategies have recently been developed with this aim (*2–5*), but they have limitations in the number of transcripts that can be analyzed, rely on rich preexisting data sets, and/or are costly and labor-intensive, and none of them are operational in the standard research and diagnostic setting of regular histological tissue sections.

We asked whether it would be possible to introduce positional molecular barcodes in the complementary DNA (cDNA) synthesis reaction within the context of an intact tissue section before RNA-seq. We first assessed whether it was feasible to generate cDNA from messenger RNA (mRNA) in tissue sections on a surface. We immobilized reverse-transcription oligo(dT) primers on glass slides and placed on the slides sections of adult mouse olfactory bulb, a brain region with clear histological landmarks and ample gene-expression reference data. The tissue was fixed, stained, and imaged (Fig. 1A) (*6*).

After permeabilization, we added reverse-transcription reagents on top of the tissue. We used fluorescently labeled nucleotides to visualize the

synthesized cDNA (Fig. 1A and fig. S1). The tissue was then enzymatically removed, which left cDNA coupled to the arrayed oligonucleotides on the slide (*6*). The fluorescent cDNA showed a pattern in detail corresponding to the tissue structure revealed by the general histology (Fig. 1, B and C), and the cDNA was strictly localized directly under individual cells (Fig. 1, D to G). By comparing the hematoxylin-and-eosin and fluorescent signals, we could measure the average distance of diffusion outside the border of a cell to $1.7 \pm 2 \mu\text{m}$ (mean \pm SD) (fig. S1, E to H).

The realization that it is possible to capture mRNA in tissue sections with minimal diffusion and maintained positional representation motivated us to array oligonucleotides with positional barcodes (Fig. 2A), and we denoted this strategy “spatial transcriptomics.” We deposited ~200 million oligonucleotides in each of 1007 features, with a diameter of $100 \mu\text{m}$ and a center-to-center distance of $200 \mu\text{m}$, over an area of $6.2 \text{ mm} \times 6.6 \text{ mm}$ (fig. S2).

After capturing and reverse-transcribing mRNA, we generated sequencing libraries based on amplification by *in vitro* transcription (fig. S3, A and B) (*7, 8*). Comparison with data from RNA extracted and fragmented in solution revealed that ~95% of the genes found with one of the methods was also found with the other (fig. S3C). The correlation between the surface and in-solution libraries was $r = 0.94$, with even representation of genes having high or low expression (fig. S3D). Replicates of surface-based experiments of adjacent tissue sections showed a correlation of $r = 0.97$ (fig. S3E). Thus, cDNA synthesis from tissue with arrayed oligonucleotides on a surface is efficient and does not introduce bias compared with in-solution protocols (fig. S3F and table S1).

We sorted the RNA-seq data to its corresponding array features by using the spatial barcodes and aligned the tissue image with the features of the array, which enabled visualization and analyses. Examples of gene-expression patterns revealed by spatial transcriptomics and validation by *in situ* hybridization are shown in Fig. 2B and fig. S4, A to C. Transcripts expressed at very low levels, such as olfactory receptor mRNAs (*9*), were also detected with spatial transcriptomics (fig. S4D).

The number of genes (*10*) (Fig. 2C) and unique transcripts (fig. S5A) per individual feature varied between cell layers with different cell density (Fig. 2D and table S2). For the vast majority of genes, the coefficient of variation decreased as the average expression increased (fig. S5B). The number of

*Department of Cell and Molecular Biology, Karolinska Institute, SE-171 77 Stockholm, Sweden. [†]Science for Life Laboratory, Division of Gene Technology, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden. [‡]Department of Dental Medicine, Division of Periodontology, Karolinska Institute, SE-141 04 Huddinge, Sweden. ⁴Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Box 1031, SE-171 21 Solna, Sweden. ⁵Division of Molecular Neuroscience, Department of Medical Biochemistry and Biophysics, Karolinska Institute, SE-17177 Stockholm, Sweden. ⁶Department of Physiology and Pharmacology, Karolinska Institute, SE-17177 Stockholm, Sweden. ⁷Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, SE-223 81 Lund, Sweden. ⁸Department of Immunology, Genetics and Pathology, Uppsala University, SE-751 85 Uppsala, Sweden. ⁹Science for Life Laboratory, Department of Neuroscience, Karolinska Institute, SE-171 77 Stockholm, Sweden.
*These authors contributed equally to this work. †These authors contributed equally to this work. [‡]Corresponding author. Email: joakim.lundberg@scilifelab.se

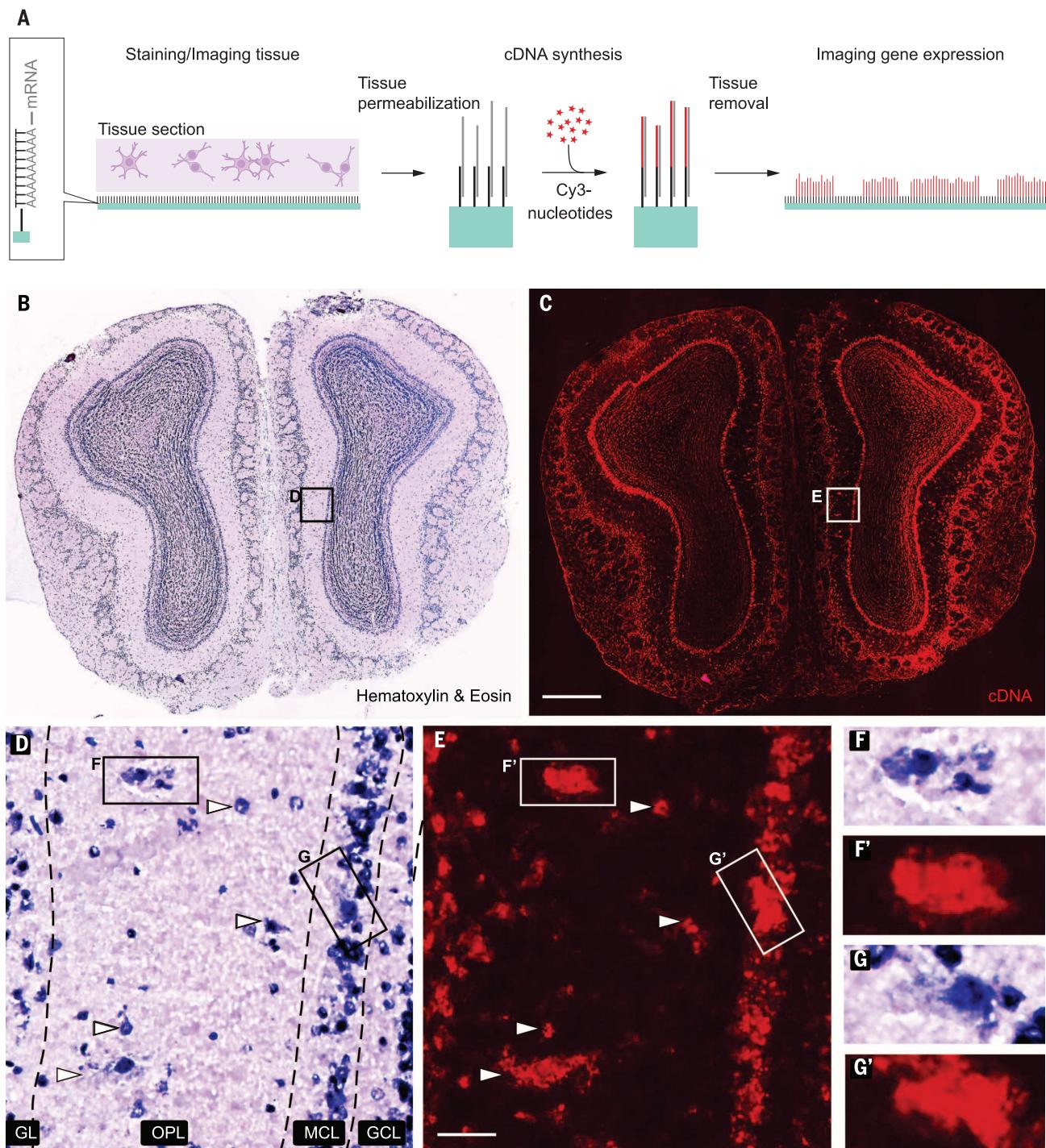


Fig. 1. Spatially localized cDNA synthesis. (A) The tissue is sectioned, placed onto oligo(dT) primers, stained, and imaged. cDNA synthesis with Cy3-labeled nucleotides reveals fluorescent cDNA after tissue removal. (B) Hematoxylin-and-eosin staining of olfactory bulbs and (C) fluorescent cDNA after tissue removal. Scale bar, 500 μ m. (D and E) Magnification of boxes in (B) and (C). Cell layers: GL; OPL, outer plexiform layer; MCL; and GCL. Arrowheads and boxes indicate individual cells and corresponding cDNA with overlapping positions. Scale bar, 40 μ m. (F to G') Cells in (D) and (E) magnified showing cytoplasm and corresponding cDNA.

genes and transcripts captured was at least twice as high as when using laser capture microdissection (11), and spatial transcriptomics detected almost twice as many genes as examination by *in situ* hybridization in the Allen Brain Atlas (fig. S5, C and D). Furthermore, we compared spatial

transcriptomics with the near-100% sensitivity of single-molecule fluorescent *in situ* hybridization in adjacent tissue sections. The sensitivity of spatial transcriptomics was $6.9 \pm 1.5\%$ of single-molecule fluorescent *in situ* hybridization (fig. S6). By comparison, single-cell RNA sequencing

has been reported to have about 5 to 40% sensitivity (12).

To further assess the potential lateral diffusion of transcripts, we investigated the distribution of the expression of 10 different genes with highly enriched expression in the mitral cell layer (MCL),

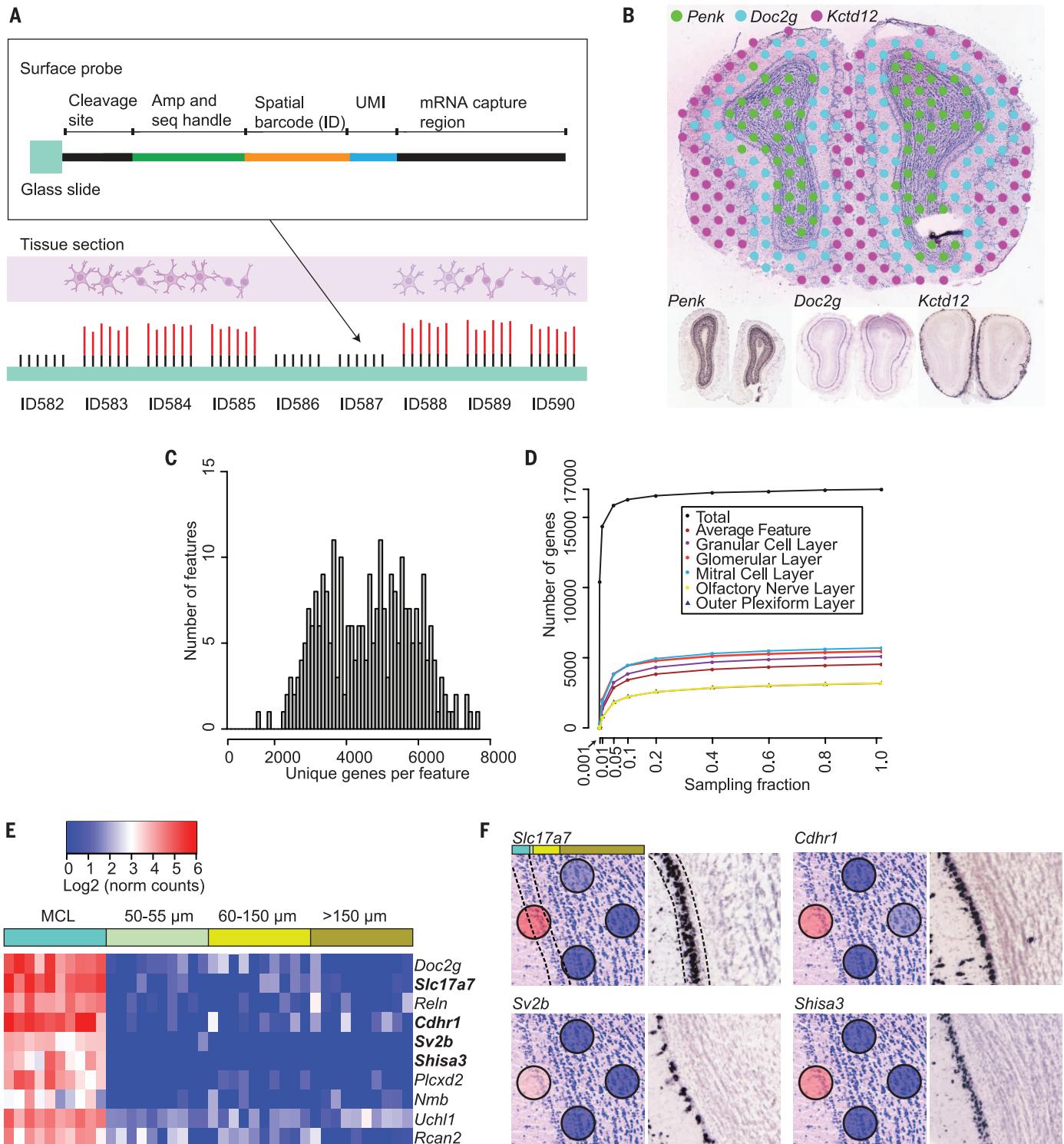


Fig. 2. Spatially resolved gene expression. (A) Each array feature contains unique DNA-barcoded probes containing a cleavage site, a T7 amplification and sequencing handle, a spatial barcode, a unique molecular identifier (UMI), and an oligo(dT) VN-capture region, where V is anything but T and where N is any nucleotide. cDNA (red) is generated from captured mRNA by reverse transcription. (B) Visualization of the expression of three genes by spatial transcriptomics (top) and in situ hybridization (bottom). *Penk* and *Kctd12* in situ images are from the Allen Institute. Cutoff normalized counts, *Penk*, 8; *Doc2g*,

13; and *Kctd12*, 19. (C) Distribution of unique genes per feature under the tissue. (D) Number of genes detected for different layers and entire tissue over sequencing depth. (E) Lateral diffusion of transcripts from genes enriched in MCL. The genes are expressed in MCL features but are not separable from the background in features adjacent to the MCL. (F) Spatial expression and in situ hybridization of four genes in (E). The leftmost feature overlaps the MCL, and the three rightmost features are situated in the GCL. The colored bar depicts the distances from feature center in (E).

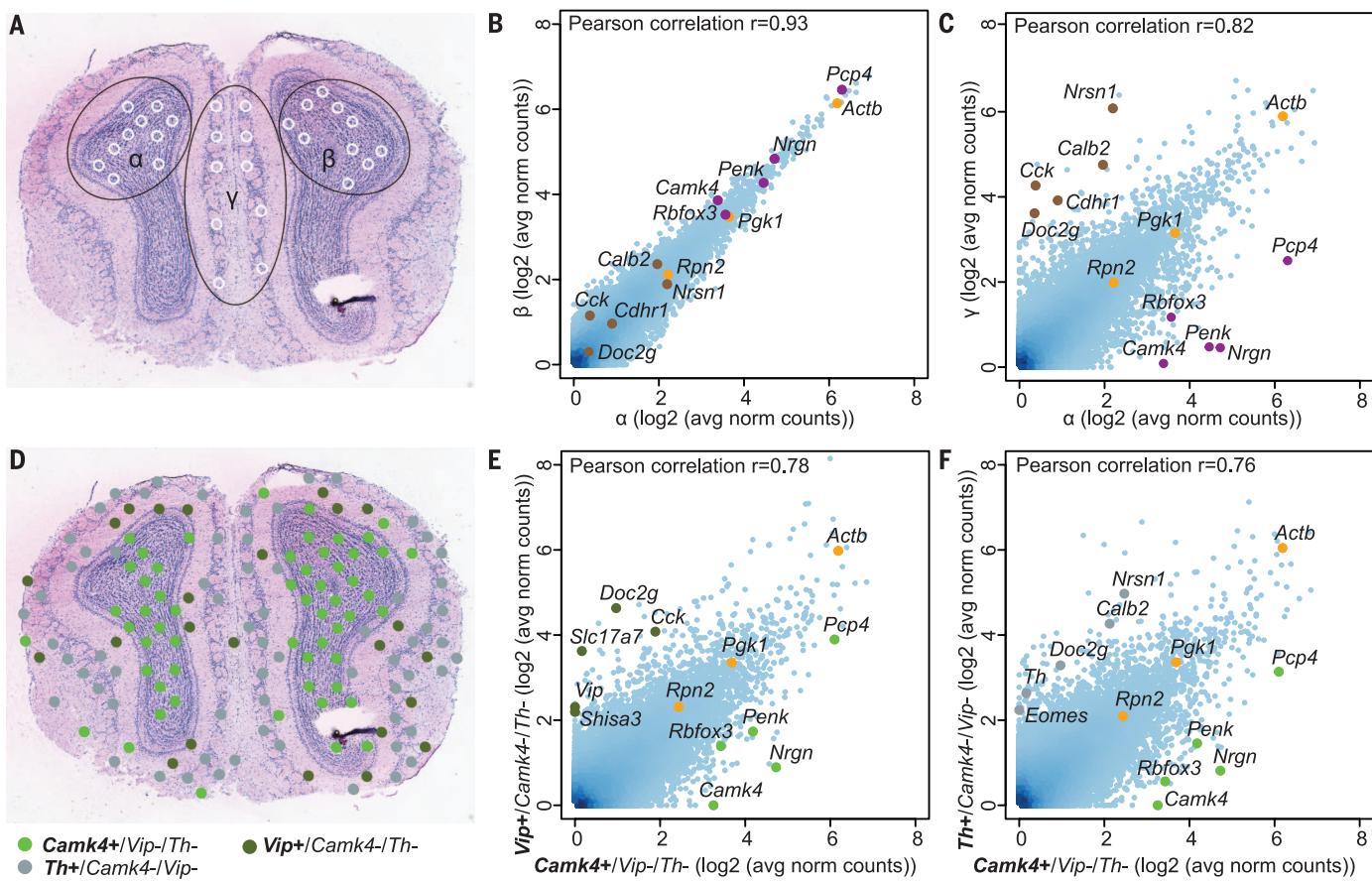


Fig. 3. Visualization and bioinformatics analyses of tissue domains defined by morphology or gene expression profile. (A) Ten selected features in areas α (GCL), β (GCL), or γ (GL) are indicated. (B) Scatterplot of gene expression in areas α and β shows similar expression of layer-specific genes. Examples of genes are indicated with purple and brown dots. Housekeeping genes are orange. (C) Scatterplot of gene expression in areas α and γ shows a difference in gene expression. Examples from the 170 differentially expressed

genes are labeled. (D) The spatial expression of three interneuron-marker-gene profiles. Ten features with the different expression profiles were randomly selected for differential expression analysis. (E) Comparing the 10 $Camk4^+/Vip^-/Th^-$ features with the 10 $Vip^+/Camk4^-/Th^-$ features. Examples, out of the 196 differentially expressed genes, are labeled. (F) Comparing the 10 $Camk4^+/Vip^-/Th^-$ features with the 10 $Th^+/Camk4^-/Vip^-$ features. Examples from the 328 differentially expressed genes are labeled.

and we asked whether they could be detected in the adjacent granular cell layer (GCL). All these genes were confirmed to be highly expressed in the MCL by spatial transcriptomics, but they were undetectable or detected at very low levels within the GCL, even with the border of the feature 0 to 5 μ m and the center of the feature 50 to 55 μ m from the MCL (Fig. 2, E and F, and fig. S7A). Furthermore, we compared the distribution of transcripts between areas obtained with laser capture microdissection (6) where there is no diffusion of transcripts and with spatial transcriptomics features, and we did not find evidence for a difference between these methods in terms of mRNA diffusion (fig. S7, B and C).

A common goal of gene expression analysis of tissues is to define the transcriptome of specific areas. Analysis between homologous regions revealed very similar expression profiles (Fig. 3, A and B, and fig. S8), with no differentially expressed genes. In contrast, comparison of different domains revealed different gene expression profiles (Fig. 3, A and C, and fig. S8). This included genes with previously known restricted expression, such as

Doc2g in the glomerular layer (GL) and *Penk* in the GCL (13), as well as novel layer-specific gene expression profiles (Fig. 3C).

It is valuable to explore the gene expression pattern of populations of cells or tissue domains that can be defined by a combination of markers. Spatial transcriptomics offers an alternative approach that circumvents multiplex labeling and cell isolation. Any combination of presence or absence of expression for a set of genes can be used to define a marker profile of interest for further analysis. Features were selected on the basis of the presence and/or absence of the three interneuron-marker genes *Camk4*, *Th*, and *Vip*. The distribution of features, where one of the genes is expressed alone, is shown in Fig. 3D. Comparing gene expression revealed specific transcriptomes defined by these interneuron-marker profiles (Fig. 3, E and F, and fig. S8).

To further explore gene expression profiles in spatially defined domains within the olfactory bulb, we used principal component analysis (fig. S9) or the t-distributed stochastic neighbor embedding (t-SNE) (14, 15) machine-learning algorithm for

dimensionality reduction, followed by hierarchical clustering (Fig. 4A). When placing back the clustered features on the tissue images, it was apparent that each cluster of features largely corresponded to well-defined morphological layers (Fig. 4B). The clusters were then compared with each other, which allowed the identification and visualization of cluster-specific marker genes (fig. S10, A and B). This proved to be an efficient, unbiased way to identify genes with expression enriched in the cell layers of interest. Furthermore, we investigated the gene expression pattern in 10 sections from a total of five animals, as well as the feature-to-feature correlation at the same location in two adjacent sections (fig. S10, C to E).

Analysis of the histology and a set of markers are routine in cancer diagnostics, although analysis of the expression of panels of genes has started to enter the clinic. We asked whether adding a spatial dimension to gene expression analysis may add information in cancer diagnostics and applied spatial transcriptomics to breast cancer biopsies. In Fig. 4, C and D (see also fig. S11, A and B), an area with invasive ductal

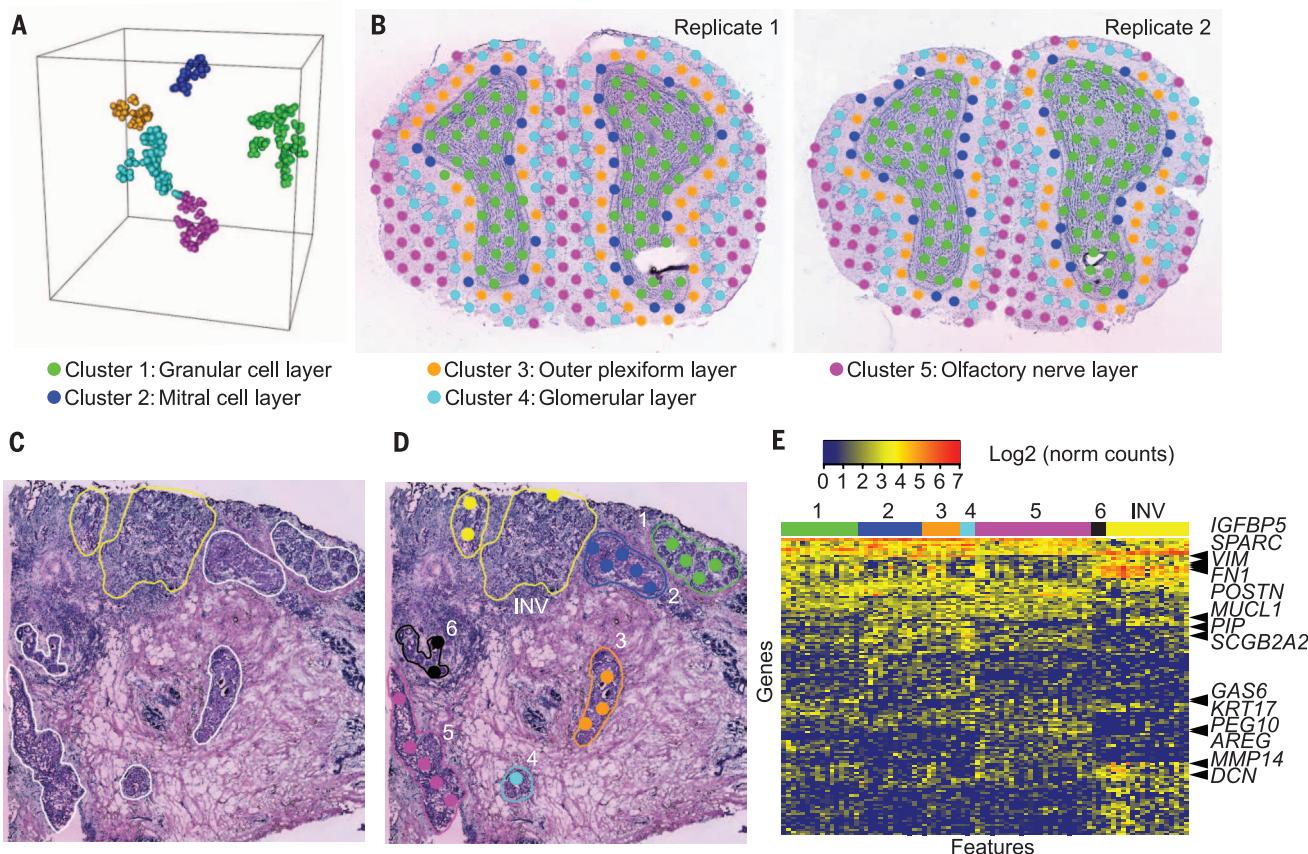


Fig. 4. Comparative analyses of tissue domains. (A) t-SNE analysis and hierarchical clustering of 551 features from two replicates creates five distinct clusters. (B) The features placed back onto the two tissue images. (C and D) Histological section of a breast cancer biopsy (C) containing invasive ductal cancer (INV) and six separate areas of ductal cancer in situ (1 to 6), with analyzed spatial transcriptomics features in (D). INV areas without, or with minimal, stromal infiltration were selected. (E) Gene expression heat map over the different areas in four adjacent sections (D) and (fig. S11).

cancer, as well as six separate areas of ductal cancer in situ, were identified on the basis of morphological criteria. Spatial transcriptomics analysis of the invasive component revealed high expression of extracellular matrix-associated genes (Fig. 4E). Analysis of the ductal cancer in situ areas revealed a surprisingly high degree of heterogeneity in gene expression between these regions, probably reflecting different subclones, with varying expression of several genes implicated in cancer progression (Fig. 4E and fig. S1C). For example, expression of *KRT17* and *GAS6*, implicated in epithelial-to-mesenchymal transition (16, 17), was high only in areas 1 and 5 (Fig. 4, C to E, and fig. SII). Thus, spatial transcriptomics revealed unexpected heterogeneity within a biopsy, which would not be possible to detect with regular transcriptome analysis and which may give more detailed prognostic information.

Spatial transcriptomics calls for only a few extra steps compared with RNA-seq analysis of homogenized tissue, with the benefit of providing spatial information enabling additional levels of analysis. In contrast to standard methods, different domains of the tissue are processed in the same reaction in spatial transcriptomics, which removes technical variation between samples. A unique feature of spatial transcriptomics is that any gene ex-

pression profile can be selected to specify a molecularly defined domain for further analysis. Finally, in contrast to when different regions of a tissue are dissected for analysis, the information for the whole section is maintained; hence, the analysis is not limited to the initially selected regions. An individual spatial transcriptomics experiment thus serves as a permanent resource to investigate gene expression patterns for future research questions.

REFERENCES AND NOTES

1. Z. Wang, M. Gerstein, M. Snyder, *Nat. Rev. Genet.* **10**, 57–63 (2009).
2. N. Crosetto, M. Bilenko, A. van Oudenaarden, *Nat. Rev. Genet.* **16**, 57–66 (2015).
3. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, *Nat. Biotechnol.* **33**, 495–502 (2015).
4. P. A. Combs, M. B. Eisen, *PLOS ONE* **8**, e17820 (2013).
5. K. Achim et al., *Nat. Biotechnol.* **33**, 503–509 (2015).
6. Materials and methods are available as supplementary materials on *Science Online*.
7. T. Hashimshony, F. Wagner, N. Sher, I. Yanai, *Cell Reports* **2**, 666–673 (2012).
8. R. N. Van Gelder et al., *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1663–1667 (1990).
9. R. Vassar et al., *Cell* **79**, 981–991 (1994).
10. K. D. Pruitt et al., *Nucleic Acids Res.* **42** (D1), D756–D763 (2014).
11. S. Zechel, P. Zajac, P. Lönnerberg, C. F. Ibáñez, S. Linnarsson, *Genome Biol.* **15**, 486 (2014).
12. D. Grün, A. van Oudenaarden, *Cell* **163**, 799–810 (2015).
13. E. S. Lein et al., *Nature* **445**, 168–176 (2007).
14. A. Mahfouz et al., *Methods* **73**, 79–89 (2015).
15. L. J. P. van der Maaten, G. E. Hinton, *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
16. M. Kittaneh, A. J. Montero, S. Glück, *Biomarkers Cancer* **5**, 61–70 (2013).
17. C. Gjerdrum et al., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1124–1129 (2010).

ACKNOWLEDGMENTS

We thank K. Meletis and M. Nilsson for discussions. This study was supported by Knut och Alice Wallenberg Foundation, the Swedish Foundation for Strategic Research, the Swedish Research Council, the Swedish Cancer Society, the Karolinska Institute, Tobias Stiftelsen, Torsten Söderbergs Stiftelse, Ragnar Söderbergs Stiftelse, StratRegen, Åke Wiberg Foundation, and the Jeansson Foundations. P.L.S. was supported by a postdoctoral fellowship from the Swedish Research Council. We thank the Swedish National Genomics Infrastructure hosted at SciLifeLab, as well as the Swedish National Infrastructure for Computing—Uppsala Multidisciplinary Center for Advanced Computational Science and Bioinformatics Long-Term Support for providing sequencing and computational assistance and infrastructure. The sequencing data are deposited at the National Center for Biotechnology Information, NIH, with BioProject ID PRJNA316587. Gene counts and scripts can be downloaded from www.spatialtranscriptomicsresearch.org. P.L.S., F.S., J.L., and J.F. are authors on patents applied for by Spatial Transcriptomics AB covering the technology.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/353/6294/78/suppl/DC1
Materials and Methods
Figs. S1 to 11
Tables S1 and S2
References (18–25)

12 January 2016; accepted 31 May 2016
10.1126/science.aaf2403