until nearly all variable combinations are considered (figs. S1 to S20). Moreover, alternative criteria for selecting candidate views may be desirable to address specific objectives. For example, a more robust but less specifically predictive model could be constructed by selecting variable combinations that are maximally distinct. With enough data, it should even be possible to identify optimal weightings of the different views or have such weightings be state-dependent (e.g., to correct for the state-dependent biases of individual views). Regardless of details, the implementation of MVE demonstrated here is intended to be as simple as possible.

The main innovation of MVE is to leverage the interconnectedness (the shared information) of complex systems. As seen in Fig. 3, improvements in forecast skill can be especially evident for short time series (~25 time points). This result is especially promising given that many current ecological data sets are wide in scope, with many different variables being tracked, but shallow in terms of time series length. Furthermore, the noise-mitigating aspects of MVE are potentially useful for many other applications such as reconstructing historical behavior, signal processing (31), or nonlinear system control (32). Although the high-dimensionality of complex systems is typically perceived as an obstacle, such complexity is actually an advantage, leading to better clarity and prediction.

REFERENCES AND NOTES
22. For nearest-neighbor methods, asymptotic convergence requires only that the selection of a satisfies Stone’s consistency theorem (23). The square root was chosen for simplicity and for its use in the machine learning literature.

24. See supplementary materials on Science Online.

ACKNOWLEDGMENTS
We thank S. Glaser, C. Hsieh, E. Deyle, and S. Munch for suggestions and feedback on early drafts of this work.

SINGLE-CELL GENOMICS

Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons

Naomi Habib,1,2,3* Yiqing Li,1,2,3,4* Matthias Heidenreich,1,2,3 Lukasz Swiech,1,2,3 Inbal Avraham-Davidi,1 John J. Trombetta,1 Cynthia Hesson,1 Feng Zhang,1,2,3,5,6† Avig Regev1,7†

Single-cell RNA sequencing (RNA-seq) provides rich information about cell types and states. However, it is difficult to capture rare dynamic processes, such as adult neurogenesis, because isolation of rare neurons from adult tissue is challenging and markers for each phase are limited. Here, we develop Div-Seq, which combines scalable single-nucleus RNA-seq (sNuc-Seq) with pulse labeling of proliferating cells by 5-ethyl-2'-deoxyuridine (EdU) to profile individual dividing cells. sNuc-Seq and Div-Seq can sensitively identify closely related hippocampal cell types and track transcriptional dynamics of newborn neurons within the adult hippocampal neurogenic niche, respectively. We also apply Div-Seq to identify and profile rare newborn neurons in the adult spinal cord, a noncanonical neurogenic region. sNuc-Seq and Div-Seq open the way for unbiased analysis of diverse complex tissues.

,*These authors contributed equally to this work.
†Corresponding author. Email: zhang@broadinstitute.org (F.Z.); aregev@broadinstitute.org (A.R.)

This work was supported by U.S. Department of Defense Strategic Environmental Research and Development Program 15 RC-2509, Leiferson Foundation award D2028335, NSF grant DEB-1020372, the McQuown Chair in the Natural Sciences, and the Sugihara Family Trust. Mesocosm data are available in the appendix of (28); model simulation data are available in Data S1.

SUPPLEMENTARY MATERIALS
www.sciencemag.org/content/353/6302/925/suppl/DCC
Materials and Methods
Figs. S1 to S20
References
Data S1
6 May 2016; accepted 1 August 2016
10.1126/science.aag0863

S
ingle-cell RNA sequencing (scRNA-Seq) has extended our understanding of heterogeneous tissues, including the central nervous system (CNS) (1–3). However, dynamic processes, such as adult neurogenesis, remain challenging to study by scRNA-Seq. First, scRNA-Seq requires enzymatic tissue dissociation (Fig. 1A), which may compromise the integrity of neurons and their RNA content, skew data toward easily dissociated cell types, and is restricted to fetal or young animals (4). Second, it is difficult to capture rare cell types, such as adult newborn neurons (4), because of limitations in cell tagging and isolation at each phase of the dynamic process.

We therefore developed Div-Seq, a method for RNA-seq of individual, recently divided cells. Div-Seq relies on sNuc-Seq, a single-nucleus isolation and RNA-seq method compatible with frozen or fixed tissue (Fig. 1A), which enables enrichment of rare labeled cell populations by fluorescence-activated cell sorting (FACS) (fig. S1). Div-Seq combines sNuc-Seq with pulse labeling of dividing cells by 5-ethyl-2'-deoxyuridine (EdU) (5, 6).

We validated that sNuc-Seq on population of nuclei faithfully represents tissue-level RNA
(7) (fig. S2, A and B), in agreement with earlier studies on the feasibility of single-nuclei sequencing (7, 8). Next, we analyzed 1367 single nuclei from hippocampal anatomical subregions [dorsal ganglion (DG), CA1, CA2, and CA3] from adult mice, including enrichment of genetically tagged low-abundance γ-aminobutyric acid–releasing (GABAergic) neurons (9) (fig. S1). sNuc-Seq robustly generated high-quality data across animal age groups (including 2-year-old mice; figs. S2, C to H, and S3), detecting 5100 expressed genes per nucleus on average, with complexity comparable

![Fig. 1. sNuc-Seq identifies cell types in adult mouse brain.](image1)

**A** sNuc-Seq workflow: Enzymatic digestion, Fixation and gradient centrifugation, Nuc-Seq RNA-Seq, FACS sorting, and profiled by modified Smart-seq2 protocol (21). (B) Major cell types identified from sNuc-Seq data reflected by clusters, shown as two-dimensional (2D) tSNE embedding of 1188 nuclei from adult mouse hippocampus. Axes: 2D coordinates from tSNE algorithm. (C) Cluster-specific genes across single nuclei. Color bar matches cluster color in (B). TPM: transcripts per million. (D) Identification of DG granule cell, CA1, CA2, and CA3 pyramidal cell clusters by marker genes, shown as (i) ISH image in hippocampus section (10) (arrowhead: high expression; scale bar, 400 μm); (ii) histogram quantifying expression in relevant cluster; and (iii) 2D embedding of nuclei [as in (B)] colored by relative expression.

![Fig. 2. sNuc-Seq and biSNE distinguish cell subtypes and spatial expression patterns.](image2)

**A** Pyramidal neurons biSNE subclusters: (A) Pyramidal CA1 and CA3 biSNE subclusters. Shown is a 2D embedding of the CA1 and CA3 pyramidal nuclei (colored by cluster). Insets: the CA1 cluster (orange) and CA3 cluster (green) within all other clusters from Fig. 1B. (B) Mapping of CA1 and CA3 pyramidal subclusters to subregions. Subcluster assignments are numbered and color coded as in (A). (C) Predictions by CA1 and CA3 subcluster spatial mapping match with Allen ISH data (10). Left illustrations: boxes: predicted differential expression regions; arrowhead: high expression; asterisk: low expression. (D) Mutually exclusive expression of Penk (facing up) and its receptor Oprd1 (facing down) across neuronal subclusters. Red line: median; box: 75 and 25% quantile. Single and double asterisks: GABergic clusters associated with Pvalb or Vip markers, respectively. (E) Co-FISH of Penk or Oprd1 with markers of GABergic subtypes [Pvalb and Vip as in (D)]. Arrowheads: coexpression. Scale bars, 20 μm.
to that of single-neuron RNA-Seq from young mice (J–3) (fig. S3, A to C).

Analysis of sNuc-Seq data revealed distinct nuclei clusters (Fig. 1, B to D; figs. S4, S5, and S6, A to C; and table S1) corresponding to known cell types and anatomical distinctions in the hippocampus. Analysis was consistent with microdissections, in situ hybridization [Allen Brain Atlas ISH (10), fig. S5], and bulk RNA-Seq (11) (fig. S8D). We captured finer distinctions between closely related cells using a new clustering algorithm, biSNE (biclustering on stochastic

**Fig. 3. Transcriptional dynamics of adult neurogenesis by Div-Seq.** (A) Div-Seq method: EdU is injected into adult mice and incorporates into dividing cells (5). Isolated EdU-labeled nuclei are fluorescently tagged by click-IT chemistry and captured by FACS for sNuc-Seq. (B) Adult neurogenesis in the DG (4). Tan box: timing of EdU labeling. NSC: neuronal stem cell. Bottom panel: EdU labeling and tissue dissection (gray) time course. (C) A continuous trajectory of newborn cells in the DG. biSNE 2D embedding of neuronal lineage nuclei (n = 269). Arrow: direction of trajectory determined by labeling time and marker expression. Top: Colored by labeling time (1 to 14 days). Bottom: Expression of markers, shown as (i) average expression along the trajectory (left colorbar) and (ii) 2D embedding colored by the expression level (right colorbar). Markers (clockwise from top left): Sox9 (NSC), Notch1 (proliferation/differentiation), Neurod1 (immature neurons), Eomes/Tbr2 (neuronal precursor). (D) Expression waves along the trajectory. Left: average expression of cluster genes along the trajectory. Middle: heatmap of average expression of each gene along the trajectory and neurogenic stages [labeled as in (B)]. Right: representative enriched biological pathways.

**Fig. 4. Dynamics of adult newborn GABAergic neurons in SC.** (A) Div-Seq in SC captures oligodendrocytes precursor cells (OPCs) and immature neurons. Distribution of cell types in non-EdU-labeled and 6 to 7 days EdU-labeled nuclei. (B) Div-Seq captured nuclei expressing marker genes of immature (Sox11) and GABAergic (Gad1) neurons. Box plots for immature neurons, mature neurons, and OPCs. Red: median; box: 75 and 25% quartiles. (C) Newborn cells in SC form a continuous trajectory. Two-dimensional embedding of 1 to 7 days EdU-labeled and nonlabeled nuclei (n = 184 neuronal lineage nuclei), colored by labeling time. Trajectory directionality is based on EdU labeling time and marker genes. (D) Dynamically expressed genes shared in SC and DG neurogenesis (347 genes from fig. S22B and Fig. 3D). (E) Gradual transition from a glia-like to neuronal state. Neuronal trajectories in the SC [as in (C)] and DG. BiSNE map of average expression of each gene along the trajectory and neurogenic stages. (F) Region-specific gene expression in immature neurons (6 to 7 days after EdU labeling). A total of 236 genes differentially expressed between SC and DG (t-test false discovery rate <0.05, log-ratio >1) in olfactory bulb (OB), SC, and DG. Box: average expression of example genes up-regulated in OB and SC compared to DG.
neighbor embedding) (fig. S7), which partitioned the GABAergic neurons into subclusters (fig. S8 and table S2) and associated each subcluster with combinations of canonical markers (fig. S8C). We validated selected markers using fluorescent RNA in situ hybridization (FISH) (fig. S9).

BiSNE also distinguished between spatial hippocampal subregions with divergent transcriptional profiles. BiSNE partitioned glutamatergic cells into subclusters (fig. 2A and fig. S10), which were further mapped to hippocampal subregions, with ISH of spatial landmark genes (20) (fig. 2B and figs. S11 to S13). We validated our mapping by confirming expression patterns using the Allen ISH data set (10) (fig. 2C and fig. S14). Although some subregions were assigned to a single subcluster (e.g., dorsal lateral CA1, fig. 2B), most subregions were assigned partially overlapping subclusters, suggesting a gradual transition of transcriptional profiles between neighboring subregions. This extends current anatomical resolutions of the hippocampus (fig. S15) and supports the notion that cellular diversity does not always partition into discrete subtypes (12).

We identified genes that are indicative of specific cell type or position (tables S1 to S8). For example, *Penk*, encoding an opioid neuropeptide (enkephalin), and its receptor *Oprd1* (13) were expressed in mutually exclusive subclusters of cells (fig. 2D), which we validated by FISH and the Allen ISH data set (10) (fig. 2E and fig. S16). These cell types and spatial positions may be involved in enkephalin signaling within the hippocampus.

Next, to study transcriptional dynamics during adult neurogenesis, we developed Div-Seq by combining sNuc-Seq with EdU labeling of dividing cells (fig. 3A). Unlike genetic labeling (2, 14), EdU tags proliferating cells at any time window, marking stem cells and their progeny with high temporal resolution. We applied Div-Seq in the DG, a canonical neurogenic niche (4), over multiple time points (1 to 14 days after cell division; fig. 3B, fig. S17, and table S4). Div-Seq enriched for diverse newborn cell types and neurogenic stages (fig. S17, F and G), from proliferating stem cells to immature neurons (4).

BiSNE analysis of neuronal lineage nuclei placed the DG newborn neurons on a continuous trajectory. The order of nuclei along the trajectory matched the EdU labeling time (fig. 3C), was independent of animal age (fig. S17H), and recapitulated known dynamics of neurogenesis markers (2, 3, 15) (fig. S18A), indicating that the trajectory indeed captured the neuronal maturation process.

To characterize the transcriptional program of adult neurogenesis, we identified and clustered genes with dynamic expression patterns along the trajectory (fig. 3D and table S5). We found major coordinated transcriptional waves, involving hundreds of genes, and aligned with known transitions between neurogenic stages, with expression shifts from proliferation to neuronal differentiation (consistent with (2)), and then to neuronal integration and maturation (fig. 3D). We identified genes with restricted expression in specific stages of neurogenesis (figs. S18 and S19), including transcription factors and chromatin regulators (fig. S18B). We confirmed the early neurogenic stage-specific expression of the axon guidance molecule *Draxin* and the ribonucleotide reductase *Rmr2* by FISH (fig. S19).

Accumulating evidence suggests that adult neurogliogenesis occurs in multiple noncanonical regions (16); but traditional methods are limited for the characterization of rare newborn cells and can lead to less definitive findings, as in the spinal cord (SC) (17, 18). We applied Div-Seq over multiple time points (1 to 7 days) in the SC (fig. S20). SC nuclei 6 to 7 days after labeling (fig. 4A and B) comprised a diverse population of newborn cells including oligodendrocyte precursor cells (OPCs, 44%) and immature neurons (19%), in contrast to 4% OPCs and no immature neurons in the non–EdU-labeled population. The SC newborn neurons expressed the GABAergic markers *Gad1* and *Gad2*, suggesting GABAergic neurogenesis (consistent with (18)) (fig. 4B). Notably, we found a set of immature neuronal markers (10%) at 23 to 24 days after EdU labeling (fig. S21), suggesting survival of newborn neurons in the SC.

The full set of neuronal lineage nuclei (fig. S20A) map to a continuous trajectory (fig. 4C) that matched labeling time and expression dynamics of known markers (fig. S20C). Comparison of dynamically expressed genes along the SC and DG trajectories (fig. S20B) identified 347 (28%) common neurogenesis genes (fig. 4D and fig. S20C) but also revealed notable distinctions in the expression dynamics and branching along the DG and SC trajectories (fig. 4, D and E) (6), which can result from differences in time scales, cell populations, or parallel gliogenesis and neurogenesis processes.

The immature neurons from SC and DG are composed of different neuronal types (GABAergic in SC, granule cells in DG). To identify candidate genes driving neuronal lineage specification, we first identified differentially expressed genes between SC and DG (t test) and then compared their expression patterns to those of newborn neurons in the olfactory bulb (OB), where GABAergic neurons image. N.H. is a Howard Hughes Medical Institute Scholar. J.E. is a Howard Hughes Medical Institute Scholar. S.V. is a Howard Hughes Medical Institute Scholar. J.E. and S.V. designed and carried out experiments. J.E. and S.V. analyzed data. J.E. and S.V. wrote the paper. We thank N. Friedman, A. Shalek, D. Gennert, B. P. Roques, M. C. Fournié-Zaluski, M. Wurm, Nat. Rev. Drug Discov. 11, 291–302 (2012).

ACKNOWLEDGMENTS

Raw data are deposited to the Gene Expression Omnibus with accession no. GSE84371. annotated data are available at https://portals.broadinstitute.org/single_cell, and software tools are available at https://github.com/yning/nucseq analysis. We thank N. Friedman, A. Shalek, D. Gennert, T. Blosser, S. Kadosch, O. Rosen, Z. Wang, P. Rogers, and L. Gaffney for support and J. Campbell for dissociated neurons image. N.H. is a Howard Hughes Medical Institute fellow in the Helen Hay Whitney Foundation. M.H. is supported by the Human Frontier Science Program. This work was supported by the National Institutes of Health grants 1R01MH104050-01 and 1R01MH104049-01, the National Science Foundation (NSF, the New York Stem Cell, Simons, Paul G. Allen Family, and Vaille Foundations; and James and Patricia Poltras, Robert Metcalfe, and David Cheng. A.R. is a Howard Hughes Medical Institute Investigator on the scientific advisory board of Syros Pharmaceuticals and Thermo Fisher and a consultant for Driver group. N.H., Y.-L. A., and F.Z. are inventors on provisional patent application 62/311,292 applied for by the Broad Institute and MIT, that covers the methods described in this paper. All DNA constructs are available from Addgene subject to a material transfer agreement with Addgene.


REFERENCES AND NOTES

6. Materials and methods are available as supplementary materials on Science Online.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/353/6302/925/suppl/DC1

Materials and Methods

Supplementary Text

Figs. S1 to S24
 Tables S1 to S7

References (22–63)

22 October 2015; accepted 19 July 2016
Published online 28 July 2016
10.1126/science.aad7038
Editor's Summary

Visualizing gene expression in nuclei

Gene expression can vary greatly within a single cell. Using techniques that they developed for sequencing single nuclei and labeling proliferating cells in vivo, Habib et al. performed RNA sequencing of 1402 single nuclei from the adult mouse hippocampus. Combining this approach with a clustering algorithm for single-cell and -nucleus RNA sequencing data delineated specific cell types during cell differentiation and development. By providing polyadenylated RNA from nuclei alone, as opposed to cytoplasmic RNA, these methods open the application of single-cell transcriptomics to tissues in which individual cells are difficult to isolate.

Science, this issue p. 925