DroNc-Seq: Deciphering cell types in human archived brain tissues by massively-parallel single nucleus RNA-seq

Naomi Habib[1,2,*], Anindita Basu[1,3,*,&], Inbal Avraham-Davidi[1,*], Tyler Burks[1], Sourav R. Choudhury[1,2], François Aguet[1], Ellen Gelfand[1], Kristin Ardlie[1], David A Weitz[3,4], Orit Rozenblatt-Rosen[1],Feng Zhang[1,2,#], and Aviv Regev[1,5,#]

[1]Broad Institute of MIT and Harvard, Cambridge MA 02142

[2]McGovern Institute, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge MA 02140

[3]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138

[4]Department of Physics, Harvard University, Cambridge, MA 02138

[5]Howard Hughes Medical Institute, Department of Biology, Koch Institute of Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge MA 02140

[&]Current address: Department of Medicine, University of Chicago, Chicago IL 60637; Center for Nanoscale Materials, Argonne National Laboratory, Lemont IL 60439

[*]These authors contributed equally to this work

Correspondence to: zhang@broadinstitute.org (FZ) and aregev@broadinstitute.org (AR)

Single nucleus RNA-Seq (sNuc-Seq) profiles RNA from tissues that are preserved or cannot be dissociated, but does not provide the throughput required to analyse many cells from complex tissues. Here, we develop DroNc-Seq, massively parallel sNuc-Seq with droplet technology. We profile 29,543 nuclei from mouse and human archived brain samples to demonstrate sensitive, efficient and unbiased classification of cell types, paving the way for charting systematic cell atlases.

Single cell RNA-seq has become an instrumental approach to interrogate cell types, dynamic states and functional processes in complex tissues[1,2]. However, current protocols require the preparation of a single cell suspension from fresh tissue, a major roadblock in many cases, including clinical deployment, handling archived materials and application in tissues that cannot be readily dissociated. In particular, in the adult brain, harsh enzymatic dissociation harms the integrity of neurons and their RNA, biases data in favour of recovery of easily dissociated cell types, and can only be used on samples from young organisms, precluding, for example, those obtained from deceased patients with neurodegenerative disorders. To address this challenge, we[3] and others[4] developed single nucleus RNA-seq (*e.g.*, sNuc-Seq[3] and Div-Seq[3]) for analysis of RNA in single nuclei from fresh, frozen or lightly fixed tissues. sNuc-Seq can handle even minute samples of complex tissues that cannot be successfully dissociated, and provide access to archived or banked samples, such as fresh-frozen or lightly fixed samples. However, it relies on sorting nuclei by FACS into plates (96 or 384 wells), and thus cannot easily be scaled to profiling tens of thousands of nuclei (such as human brain tissue) or large numbers of samples (such as tumor biopsies from a patient). Conversely, massively parallel single

cell RNA-seq methods, such as Drop-Seq[5], InDrop[6] and related commercial tools[7,8] can be readily applied at this scale[9] in a cost-effective manner[10], but require a single cell suspension as input.

To address this challenge, we developed DroNc-seq (**Fig. 1a**), a massively parallel single nucleus RNA-seq method that combines the advantages of sNuc-Seq with the scale of droplet microfluidics to profile thousands of nuclei at very low cost and massive throughput. DroNc-Seq was modified from Drop-Seq[5] to accommodate for the smaller size and relatively lower amount of RNA in nuclei compared to cells. Specifically, we modified the microfluidics design (**Supplementary Fig. S1A, B**) to generate smaller co-encapsulation droplets (75 μm diameter) and flow parameters; we optimized the nuclei isolation protocol to reduce processing time and increase capture efficiency (**Supplementary Fig. S1C**); and we changed the downstream PCR conditions (**Methods**). We validated for single nucleus specificity using species-mixing experiments[5], in which we combine nuclei from human 293 cells and mouse 3T3 cells in one DroNC-seq run, to assess single nucleus purity, as previously performed for cells[5] (**Supplementary Fig. S1D**). Notably, the DroNc-Seq device and workflow are compatible with current Drop-Seq platforms.

DroNc-Seq robustly generated high quality expression profiles from nuclei isolated from a mouse cell line (3T3, 4,442 nuclei), adult mouse brain tissue (9,219), and adult human post-mortem frozen archived tissue (20,324 nuclei). It detected, on average 3,152 genes (6,614 transcripts) for 3T3 nuclei, 1,500 genes (2,614 transcripts) for nuclei from adult

mouse brain, and 1,000 genes (1,337 transcripts) for nuclei from human post mortem brain tissue (**Methods**, **Fig. 1b**).

To assess Dronc-Seq's throughput and sensitivity, we profiled the same 3T3 cell culture at both the single cell (with Drop-Seq) and single nucleus (with DroNc-Seq) levels, each sequenced to an average depth of 120,000 reads per nucleus or cell. Both methods yielded high quality libraries, detecting, on average, 4,770 and 3,152 genes for cells and nuclei, respectively (**Fig. 1c**). DroNc-Seq had somewhat reduced throughput, with 2,982 / 300,000 input nuclei passing filter (~1%), compared to 5,175 / 100,000 cells (5%) passing filter per run. The average expression profile of single nuclei was well-correlated with the average profile of single cells (Pearson r=0.87, **Fig. 1d**), albeit somewhat lower than the correlation between the average profiles of two replicates of Drop-Seq (r=0.99) or DroNc-Seq (r=0.99). Those genes with significantly higher expression in nuclei (*e.g.*, the lncRNAs *Malat1* and *Meg3*) or cells (mitochondrial genes Mt-nd1, Mt-nd2, Mt-nd4, Mt-cytb) (**Fig. 1d**) were consistent with their known distinct enrichment in nuclear *vs.* non-nuclear compartments (**Supplementary Table 1**). Interestingly, while in both methods over 85% of reads align to coding loci, in cells 80% of these reads map to exons, whereas in nuclei 56% map to exons and 32% to introns (**Fig. 1e**), reflecting the enrichment of nascent, pre-processed transcripts in the nuclear compartment[3,11-14].

Clustering[9] of 5,592 nuclei profiled from frozen adult mouse hippocampus (3 samples) and prefrontal cortex (3 samples) (each with >20,000 reads per nucleus**, Methods**) revealed groups of nuclei corresponding to known cell types (*e.g.*, GABAergic neurons)

and anatomical distinctions between the brain regions and within the hippocampus (*e.g.*, CA1, CA3, dentate gyrus; **Fig. 1f**). Neurons of the same class but from different brain regions (and different samples) group together, as was also the case for GABAergic neurons, glia and endothelial cells (**Fig. 1f-g**). Among the non-neural cells, different glia cell types, including astrocytes, oligodendrocytes and oligodendrocyte precursor cells (OPC), readily partitioned into separate clusters, despite their relatively low RNA levels and correspondingly lower numbers of detected genes (**Fig. 1f**). Finally, DroNc-Seq of mouse hippocampus compared well to sNuc-Seq of the same region[15], maintaining the ability to detect the same cell types and correlated cell-types specific signatures (**Fig. 1h, Supplementary Table 2**) with increased throughput, despite a lower number of genes detected per nucleus in the massively parallel setting.

To demonstrate the utility of DroncSeq on archived human tissue, we profile adult (40-65 years old) human post-mortem frozen brain tissue archived by the GTEx project[16]. We analysed 10,368 nuclei (each with >20,000 reads per nucleus) from five frozen post-mortem archived samples of adult human hippocampus and prefrontal cortex, revealing distinct nuclei clusters corresponding to the known cell types in these regions (**Fig. 2a**). We readily annotated each cell type cluster *post-hoc* by its unique expression of known canonical marker genes (**Fig. 2b**), including rare types, such as adult neuronal stem cells specifically found in the hippocampus (**Fig. 2a**, cluster 9). Although the human archived samples vary in the quality of the input material, DroNc-Seq yielded high-quality libraries of both neurons and glia cells from each sample (**Fig. 2c**, bottom), and each

cluster was supported by multiple samples (**Fig. 2c**, top), demonstrating the robustness and utility of DroNc-Seq for clinical applications.

Finally, we determined cell-type specific gene signatures for each human cell type cluster (**Fig. 2d**), as well as a pan-neuronal signature, a pan-glia signature, and signatures for neuronal stem cells and endothelial cells (**Supplementary Table 3**). Signatures are enriched for key relevant pathways (FDR<0.01, **Methods)**. For example: Neuronal stem cells signatures are enriched for the expression of genes regulated by NF-kB in response to TNF signalling[17]; Endothelial cells are enriched for the expression of immune pathways, such as MHC I genes and interferon signalling (**Fig. 2e**), consistent with the known role of interferon signalling in modulation of the blood brain barrier[18]. Moreover, we captured finer distinctions between closely related cells (**Fig. 2f** and **Supplementary Fig. 2**), such as, distinct sub-types of GABAergic neurons (**Fig. 2f**), each robustly identified across biological replicates (**Supplementary Fig. 3a**), and often from both brain regions (**Fig. 2g**). Two of the GABAergic neuron sub-clusters are specific to the hippocampus (**Supplementary Fig. 3a, Fig. 2f**, clusters 1 and 4); these too are supported by multiple samples (**Supplementary Fig. 3a**). We associated each GABAergic neuron sub-cluster with a distinct combination of canonical markers (**Fig. 2h**), as previously reported in the mouse brain[3,19,20].

In conclusion, DroNc-Seq is a massively-parallel single nucleus RNA-seq method, which is robust, cost-effective, and easy to use. Our results show that DroNc-Seq profiling from both mouse and human frozen archived brain tissues successfully identified cell types and sub-types, rare cells, expression signatures and activated pathways, opening the way to

systematic single nucleus analysis of complex tissues that are either inherently challenging to dissociate or already archived. This will help create vital atlases of human tissues and clinical samples.

Acknowledgements

We thank Karthik Shekhar, Christoph Muus and Eugene Drokhlyansky for helpful discussions, Timothy Tickle and Asma Bankapur for technical support, and Leslie Gaffney for help with graphics. Work was supported by the Klarman Cell Observatory, NIMH grant U01MH105960 and NCI grant 1R33CA202820-1 (to A.R.). Microfluidic devices were fabricated at the Center for Nanoscale Systems (CNS), Harvard University, member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), and supported by the National Science Foundation under NSF award no. 1541959. A.R. is an Investigator of the Howard Hughes Medical Institute. A.R. is a member of the Scientific Advisory Board for Thermo Fisher Scientific, Syros Pharmaceuticals and Driver Genomics. F.Z. is supported by the NIH through NIMH (5DP1-MH100706 and 1R01-MH110049); NSF; the New York Stem Cell Foundation; the Howard Hughes Medical Institute; the Simons, Paul G. Allen Family, and Vallee Foundations; the Skoltech-MIT Next Generation Program; James and Patricia Poitras; Robert Metcalfe; and David Cheng. F.Z. is a New York Stem Cell Foundation-Robertson Investigator. D.A.W. thanks NSF DMR-1420570, NSF DMR-1310266 and NIH P01HL120839 grants for their support. NH is a Howard Hughes Medical Institute fellow for the Helen Hey Whitney Foundation. N.H., A.B., I.A.D., D.A.W., F.Z. and A.R. are inventors on international patent application PCT/US16/59239 filed by Broad Institute, Harvard and MIT, relating to method of this manuscript. GTEx is supported by the Common Fund of the Office of the Director of the United States National Institutes of Health, through Contract HHSN268201000029C (to K.A LDACC, Broad Institute).

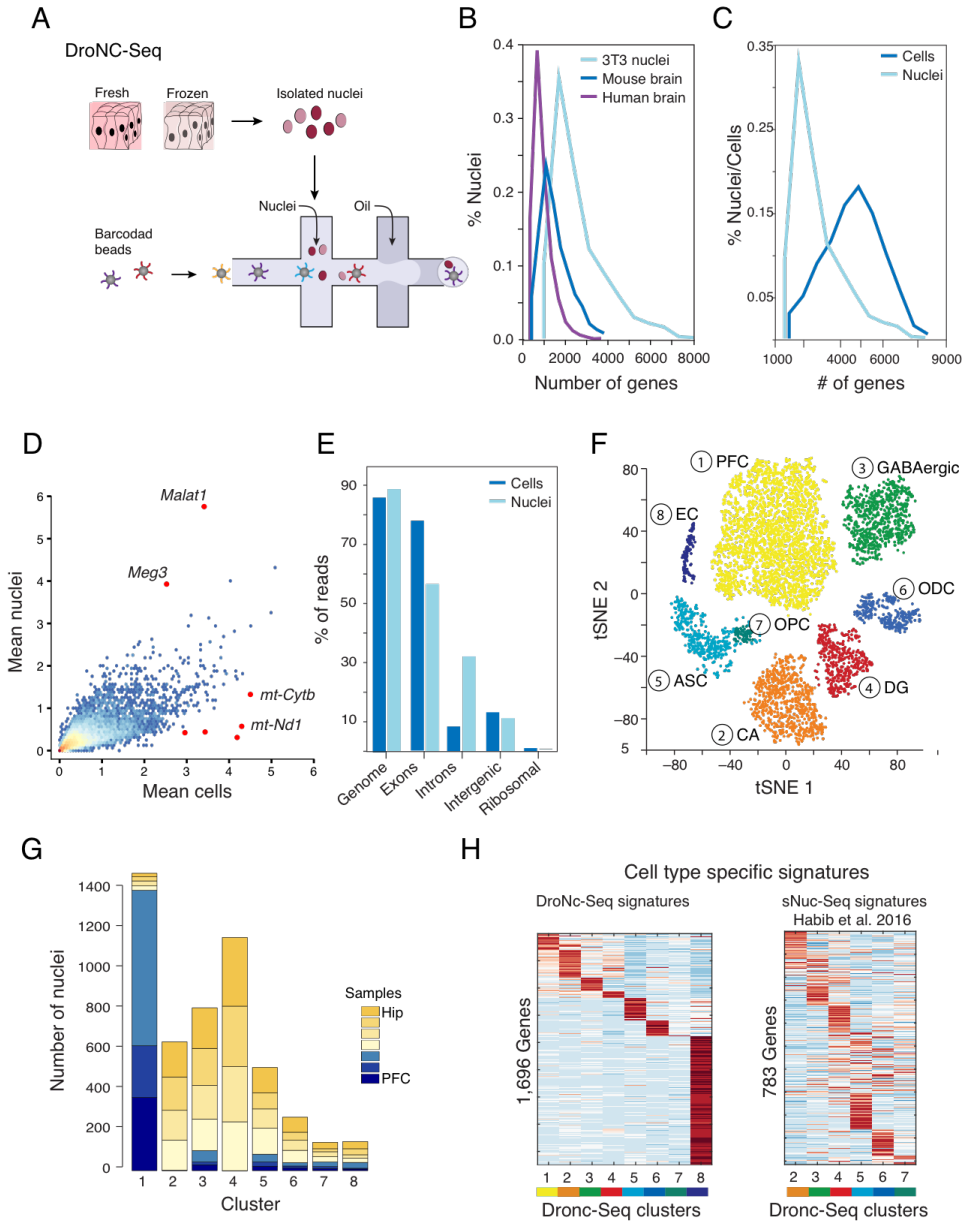

Figure 1. DroNc-Seq: Massively parallel single nucleus RNA-Seq. (a) Overview of DroNc-Seq. (**b-e**) Quality measures. (**b**) Distribution of number of genes detected (X axis) in DroNc-Seq of nuclei isolated from 3T3 mouse cells line, mouse frozen brain tissue, and human frozen archived brain tissue (**Methods**). **(c)** Distribution of number of genes detected per 3T3 cell (by Drop-Seq) or nucleus (by DroNc-Seq). (**d**) The percent of reads (Y axis) mapped to the: genome, exons, introns, intergenic regions and rRNA loci (X axis) of the mouse genome, for cells and nuclei. (**e**) Scatter plot comparing the average expression levels detected in single 3T3 nuclei (Y-axis, by DroNc-seq) and cells

(X-axis, by Drop-Seq). Red dots mark outlier genes highly expressed in one but not the other experiment. (**f-h**) DroNc-Seq analysis of adult frozen mouse hippocampus (hip) and prefrontal cortex (PFC) brain regions. (**f**) A 2 dimensional t-stochastic neighbourhood embedding (tSNE) plot of 5,592 DroNc-Seq nuclei profiles from adult frozen mouse hippocampus (hip) (3 samples) and prefrontal cortex (PFC) (2 samples, each with >20,000 reads per nucleus), colored by clustering and labelled *post hoc* by cell types and anatomical distinctions (1. PFC=pyramidal neurons from the PFC, 2. CA=pyramidal neurons from the hip CA, 3. GABAergic= GABAergic neurons, 4. DG=granule neurons from the hip dentate gyrus (DG), 5. ASC=astrocytes, 6. ODC=oligodendrocytes, 7. OPC=oligodendrocyte precursor cells, 8. EC= endothelial cells). (**g**) Number of nuclei (Y axis) from each sample (PFC = blue gradient, hip =yellow gradient) associated with each cluster (X axis), showing that each cluster is supported by multiple samples, and most by both brain regions. (**h**) Signatures of differentially expressed genes. Right: The average expression in each DroNc-Seq cluster (column) of signature genes (**Methods**, rows) that are differentially expressed in the DroNc-Seq data for each cell type cluster derived from the DroNc-Seq data (numbered as is **f**). Expression is centred per row (color bar). Right: The average expression in each relevant DroNc-Seq cluster (numbered as is **f**, column) of signature genes previously identified on sNuc-Seq profiles[3] for the corresponding cell types, showing that DroNc–seq captures similar diversity in nuclear RNA profiles between cell types.

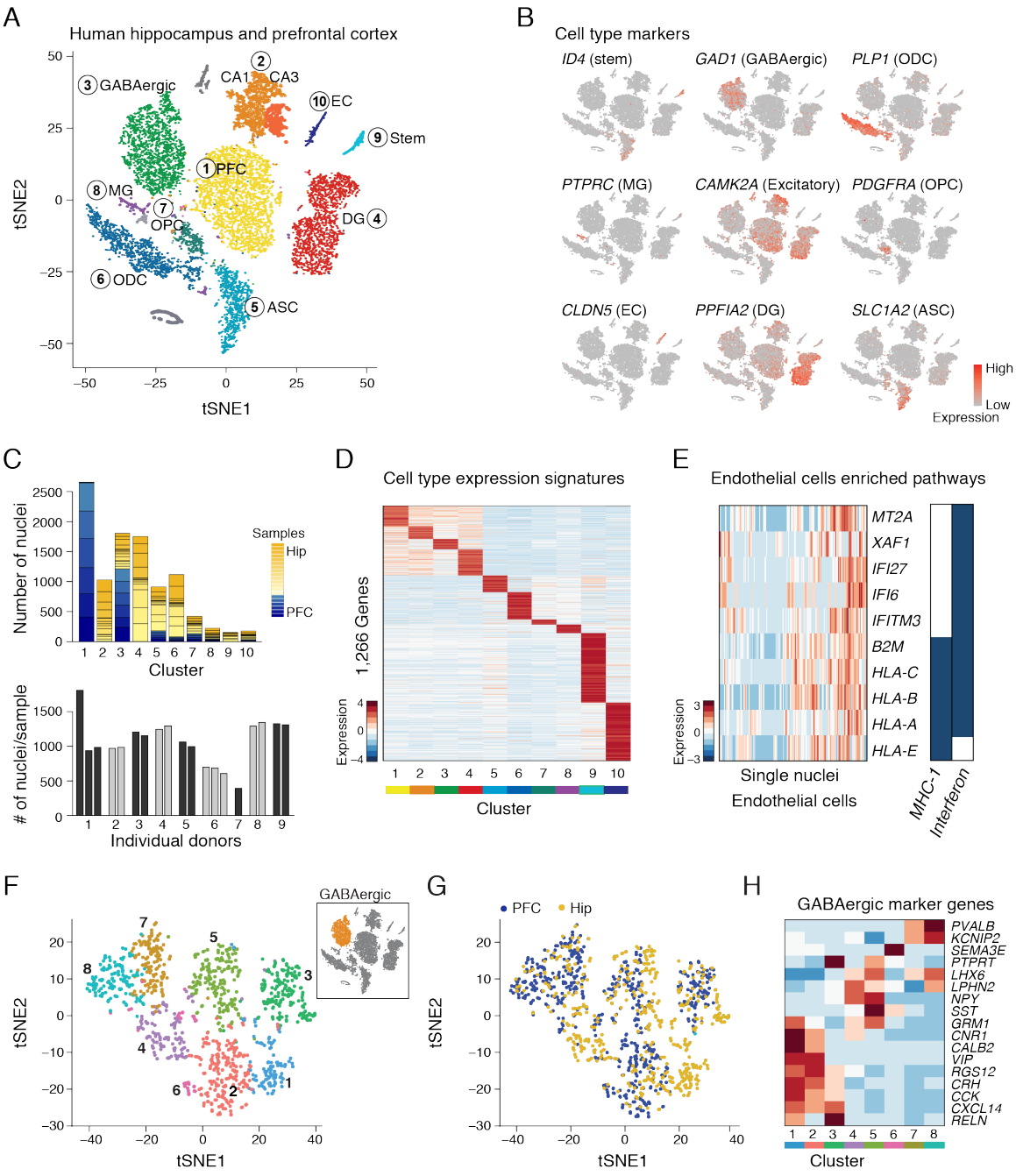

Figure 2. DroNc-Seq distinguished cell types and signatures in adult post-mortem human brain tissue. (**a**) Cell type clusters. tSNE embedding of 10,368 DroNc-Seq nuclei profiles from adult frozen human hippocampus and prefrontal cortex (PFC), each with >20,000 reads per nucleus. Clusters are color-coded and labelled *post-hoc* (1. PFC=pyramidal neurons from the PFC, 2. CA=pyramidal neurons from the hip CA, 3. GABAergic= GABAergic neurons, 4. DG=granule neurons from the hip dentate gyrus (DG), 5. ASC=astrocytes, 6. ODC=oligodendrocytes, 7. OPC=oligodendrocyte precursor cells, 8. MG=Microglia, 9. Stem=neuronal stem cells, 10. EC=endothelial cells) (**b**)

Marker genes. Shown is the same plot as in (a) but with cells colored by the expression level of known cell type marker genes. (ID4 - stem cells, GAD1 – GABAergic neurons, PLP1 – ODC, PTPRC – microglia, CAMK2 – excitatory neurons, PDGFRA – OPC, CLDN5 – EC, PPFIA2 – DG, SLC1A2 – ASC) (**c**) Successful DroNc-Seq across samples of different quality. Top: Number of nuclei (Y axis) from each sample (color code) associated with each cluster (X axis), showing that each cluster is supported by nuclei from multiple samples. Bottom: Number of nuclei passing quality filters (Y axis) recovered from each of 19 human tissue samples from 9 donors (X axis, sorted). (**d**) Cell type signatures. Heatmaps of the average expression of signature genes (rows; FDR <0.01) in nuclei in each of the clusters in (a). (**e**) Interferon signalling and MHC I genes in single endothelial cells. Shown is the expression of each gene across the nuclei in the endothelial cluster in (a). (**f-h**) Sub-types of GABAergic neurons. (**f,g**) tSNE embedding of DroNc-Seq nuclei profiles from the GABAergic neuronal cluster (in **a**), color coded by sub-clusters (f) or brain region (g). (**h**) Heatmap of the average expression of known marker genes of sub-types of GABAergic interneurons, in each of the nuclei sub-clusters in (f) (columns).

References

1 Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nature biotechnology* **34**, 1145-1160, doi:10.1038/nbt.3711 (2016).

2 Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331-338, doi:10.1038/nature21350 (2017).

3 Habib, N. *et al.* Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **353**, 925-928, doi:10.1126/science.aad7038 (2016).

4 Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586-1590, doi:10.1126/science.aaf1204 (2016).

5 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

6 Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201, doi:10.1016/j.cell.2015.04.044 (2015).

7 Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866 e1817, doi:10.1016/j.cell.2016.11.038 (2016).

8 Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882 e1821, doi:10.1016/j.cell.2016.11.048 (2016).

9 Shekhar, K. *et al.* Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* **166**, 1308-1323 e1330, doi:10.1016/j.cell.2016.07.054 (2016).

10 Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Molecular cell* **65**, 631-643 e634, doi:10.1016/j.molcel.2017.01.023 (2017).

11 Rabani, M. *et al.* High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159**, 1698-1710, doi:10.1016/j.cell.2014.11.015 (2014).

12 Rabani, M. *et al.* Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature biotechnology* **29**, 436-442, doi:10.1038/nbt.1861 (2011).

13 Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342, doi:10.1038/nature10098 (2011).

14 Cheadle, C. *et al.* Control of gene expression during T cell activation: alternate regulation of mRNA transcription and mRNA stability. *BMC genomics* **6**, 75, doi:10.1186/1471-2164-6-75 (2005).

15 Habib, N. *et al.* Div-Seq: A single nucleus RNA-Seq method reveals dynamics of rare adult newborn neurons in the CNS. *bioRxiv*, doi:10.1101/045989 (2016).

16 Consortium, G. T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660, doi:10.1126/science.1262110 (2015).

17 Widera, D., Mikenberg, I., Elvers, M., Kaltschmidt, C. & Kaltschmidt, B. Tumor necrosis factor alpha triggers proliferation of adult neural stem cells via IKK/NF-kappaB signaling. *BMC Neurosci* **7**, 64, doi:10.1186/1471-2202-7-64 (2006).

18 Daniels, B. P. & Klein, R. S. Knocking on Closed Doors: Host Interferons Dynamically Regulate Blood-Brain Barrier Function during Viral Infections of the Central Nervous System. *PLoS Pathog* **11**, e1005096, doi:10.1371/journal.ppat.1005096 (2015).

19 Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* **19**, 335-346, doi:10.1038/nn.4216 (2016).

20 Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142, doi:10.1126/science.aaa1934 (2015).