Supplementary Figures For manuscript:

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

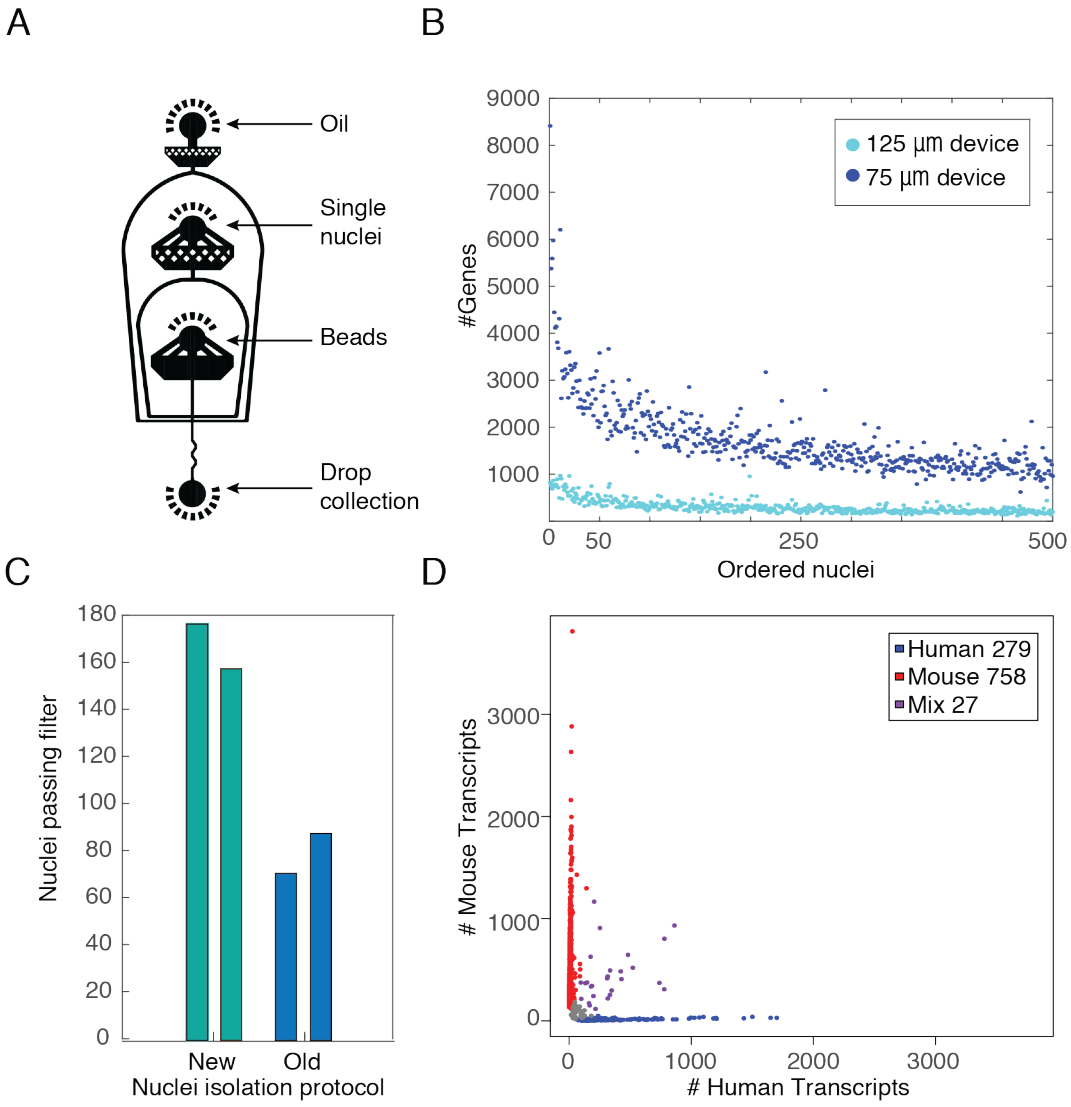
[5]Howard Hughes Medical Institute, Department of Biology, Koch Institute of Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge MA 02140

[&]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)

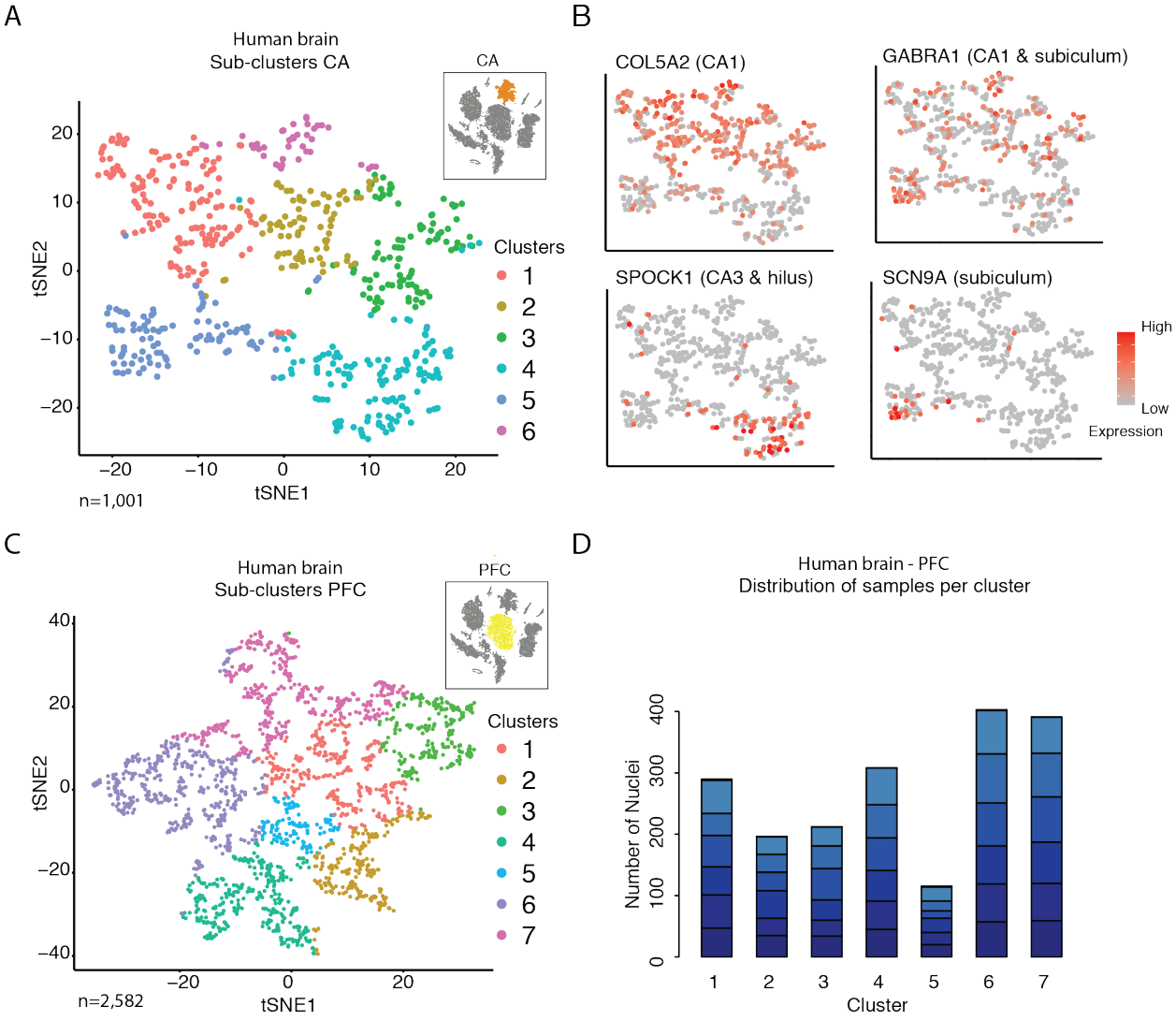
Figure S1



Supplementary Figure 1. DroNc-Seq. (**a**) CAD schematic of DroNc-Seq microfluidics device. (b-d) DroNc-seq performance in 3T3 nuclei (channels are nominally 70 μm wide). (**b**) Number of detected genes. Scatter plot shows the number of detected genes (Y axis; defined as a gene with at least two different UMIs detected, **Methods**) across nuclei, ranked in decreasing order (X axis), using DroNc-Seq 75µm microfluidics device (blue, **Methods**) and Drop-Seq[1] 125µm device (light blue). (**c**) Number of successfully profiled nuclei. Bar plot shows the number of

nuclei passing library quality filter out of 1,288 (±114) nuclei per library (**Methods**, Y axis), using either sNuc-Seq[2] ("old") or DroNc-Seq ("new") nuclei isolation protocol (**Methods)**. (**d**) Single nucleus specificity in DroNc-seq, estimated from mixtures of human 293 and mouse 3T3 nuclei. Scatter plot shows the number of UMIs associated with human (Y axis) or mouse (X axis) transcripts for each nucleus barcode (dot). Barcodes associated with a high number of both human and mouse transcripts (purple) reflect nuclei doublets. We find 2.5% (27/1,064) nuclei to be human-mouse, and thus estimate the expected doublet rate at our current loading and flow parameters to be 5%.

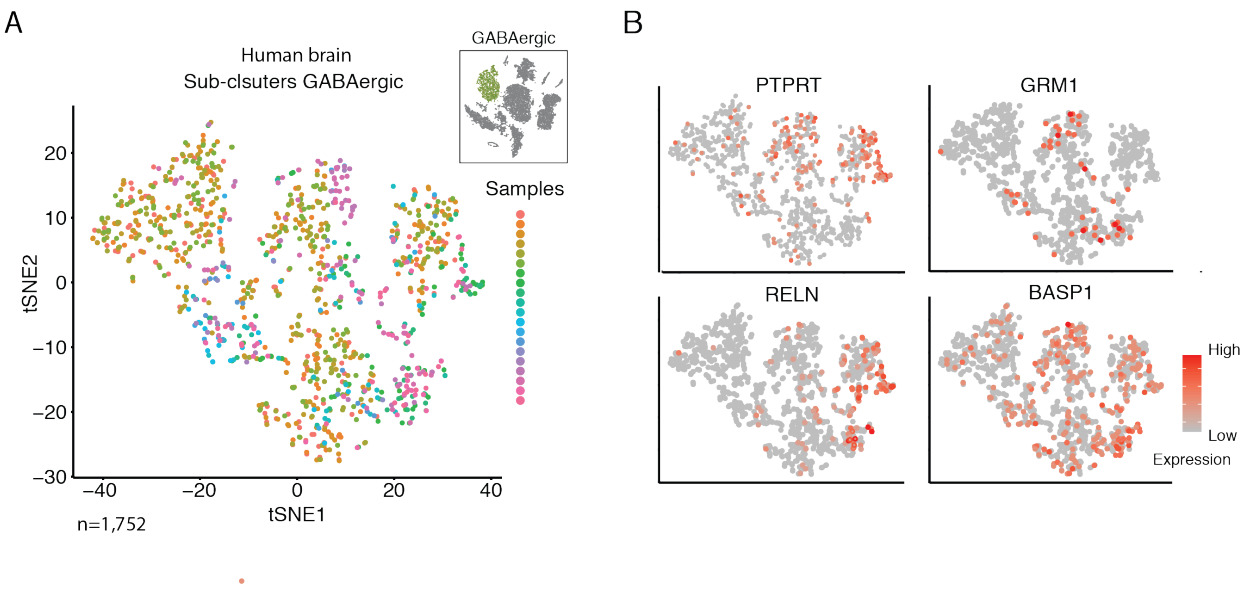
Figure S2



Supplementary Figure 2. DroNc-Seq identifies sub-clusters for major cell types in the adult human brain. (**a,b**) Sub-clusters of the human CA pyramidal neurons cluster. tSNE embedding of DroNc-Seq profiles from the human CA pyramidal neurons cluster (cluster 2 in **Fig. 2a**; inset), color coded by sub-clusters (a) or by the expression of genes differentially expressed between the CA sub-clusters (b), with the known anatomical position of the gene according to the human Allen Brain Atlas[3] noted in parentheses. Genes are: GABRA1 (highest in

the CA1 and subiculum), SCN9A (subiculum), SPOCK1 (CA3 and hilus), and CL5A2 (CA1 based on the mouse Allen Brain Atlas[4]), showing that DroNc-Seq can distinguish between pyramidal neurons that are spatially separated along the anatomical sub-regions of the CA. (**c,d**) Sub-clusters of the human prefrontal cortex (PFC) pyramidal neurons cluster. (**c**) tSNE embedding of DroNc-Seq nuclei profiles from the human prefrontal cortex (PFC) pyramidal neurons cluster (cluster 1 in **Fig. 2a**, inset), color coded by sub-clusters. (**d**) Number of nuclei (Y axis) from each sample (color code) associated with each cluster (X axis), showing that each cluster is supported by multiple samples.

Figure S3



Supplementary Figure 3. Sub-clusters of GABAergic neurons in the human brain. tSNE embedding (as in **Fig. 2g**) of DroNc-Seq nuclei profiles from the human GABAergic neuron cluster (cluster 3 in **Fig. 2a**, inset). (**a**) Color coded by the sample of origin, showing that each cluster is supported by multiple samples. (**b**) Colored by the expression level of known GABAergic marker genes or genes differentially expressed between the sub-clusters, showing unique combinatorial expression of patterns across clusters.

References

1 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).

2 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).

3 Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399, doi:10.1038/nature11405 (2012).

4 Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168-176, doi:10.1038/nature05453 (2007).