Supplementary Figure 1

snDrop-seq overview.

A. snDrop-seq method showing modifications needed to process nuclei, including bovine serum albumin (BSA) coating and droplet heating to ensure complete nuclear membrane lysis. B. Violin plot showing the ratio of area size of DAPI staining versus droplet size of untreated and heat-treated samples, confirming efficient nuclear lysis in snDrop-seq. Top left panel: 293T cells pre-stained with DAPI showing cellular, but not nuclear lysis inside droplets. Top right panel: DAPI staining in droplets after applying heat-treatment showing efficient lysis as evidenced by the overspreading of DNA stain inside the droplets. C. Collision plot for human (293T) and mouse (NIH3T3) nuclei analyzed by a single snDrop-seq run. Scatter plot shows the number of transcript reads mapping to human (hg19) and mouse (mm10) genomes for each cell barcode. Percent of collisions is indicated. D. Gene body coverage for visual cortex data sets (library occ5). E. Proportion of reads mapped to exons, 5' untranslated regions (UTR) 3' UTRs, introns, intergenic regions and that are unassigned across quality filtered data sets from the visual cortex (occ1-21, Table S1). F. Number of reads associated with (E).
Supplementary Figure 2

scTHS-seq overview.

A. For combinatorial barcoding, ~2000 nuclei are added per well to a 384-well plate containing uniquely barcoded transposons, followed by tagmentation, inactivation, pooling, and FACS redistribution into 96-well plates. In vitro transcription (IVT) and scTHS-seq RNA-seq is performed followed by PCR indexing, pooling and high throughput sequencing of libraries. Optimally ~25,000 total single-cell datasets are recovered per run. B. Customized scTHS-seq transposon design. Transposons contain a mosaic end for Tn5 binding, read primer site for sequencing, an r5 barcode unique to every transposon, an i5 connector sequence used in PCR amplification and a T7 promoter for IVT amplification. To generate the barcoded transposome complexes, each transposon is annealed to the RC mosaic end, then transposase is added. C. Depiction of sequencing ready fragments generated after library generation. Fragments contain i5/i7 adaptors and barcodes, the r5 transposon barcode, Read1 sequencing primer, and captured genomic DNA. The i5 and r5 barcodes are read in one sequencing read, with sequencing progressing through the linker region and ending with the r5 barcode. D. Confirmation of in vitro transcription and barcoded library generation in scTHS-seq. Left panel: gel image of whole reactions purified with SPRI binding buffer. Occipital lobe sample plate 5 and plate 6 quality controls (QC) consisted of two 100 nuclei samples, and two no template controls (NTCs). Right panel: gel images of occipital lobe samples from 12 wells from a 96-well plate, showing library size range that is selected for after pooling all samples. E. Identification of cell type specific DNA accessibility peaks over the promoter and regulatory region of Alzheimer's disease associated gene PICALM. Glial specific peaks are over-represented. The top tract shows all cells merged to generate peaks. Each cell subpopulation tract is represented by 100 randomly selected single cells having reads in the depicted region, where each row represents a single cell and each dot is a read. The color of highlighted peak regions corresponds to the cell type specificity of the peak, with each subpopulation tract title a specific color. Boxed regions highlight specific cell type specific peaks. The gray box highlights a glial cell specific peak. Nonspecific peaks are not highlighted.
snDrop-seq resolves neuronal and non-neuronal subpopulations.

**A.** Cluster dendrogram showing unsupervised grouping of all non-neuronal, excitatory and inhibitory neuronal cell types or subtypes across cortical and cerebellar regions. **B.** t-SNE plot for all data sets as shown in Fig. 1B. **C.** t-SNE plot as in (B) showing batch identity (see Table S1). **D.** t-SNE plot as in (B) showing individual patient identity (see Table S1). **E.** t-SNE plot as in (B) showing UMI read depth. Non-neuronal cell types show lower coverage than neuronal subtypes. **F-I.** Same as for (B-E) except showing only visual cortex (Occ or occipital lobe, BA17) data. Some patient-specific clustering can be seen for Ex3 neurons that may be associated with different visual cortical regions sampled between repositories. **J-M.** Same as for (B-E) except showing only frontal cortex (FCtx, BA6, BA8) data. **N-Q.** Same as for (B-E) except showing only cerebellar hemisphere (CBL) data.
<table>
<thead>
<tr>
<th>Method</th>
<th>Tissue/Stage</th>
<th>UMI Datasets</th>
<th>Clusters</th>
<th>Detected Molecules</th>
<th>Detected Genes</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>snDrop-seq</td>
<td>Human Cortex/CBL</td>
<td>35,289</td>
<td>34</td>
<td>Median 929</td>
<td>Median 719</td>
<td>Median 1.3</td>
</tr>
<tr>
<td>scDrop-seq</td>
<td>Mouse Retina (P14)</td>
<td>44,808</td>
<td>30</td>
<td>Median 712</td>
<td>Median 497</td>
<td>Median 1.4</td>
</tr>
<tr>
<td>10X Genomics</td>
<td>Mouse Brain (E18)</td>
<td>9093</td>
<td>22</td>
<td>Median 6354</td>
<td>Median 2597</td>
<td>Median 2.5</td>
</tr>
<tr>
<td>Fluidigm C1</td>
<td>Human Midbrain (w6-w11)</td>
<td>1695</td>
<td>25</td>
<td>Median 5274</td>
<td>Median 2293</td>
<td>Median 2.3</td>
</tr>
</tbody>
</table>

**Graphs:**
- **E:** Distribution of detected molecules.
- **F:** Distribution of detected genes.
- **G:** Coverage distribution.
- **H:** Genic length distribution.

Nature Biotechnology: doi:10.1038/nbt.4038
Supplementary Figure 4

A comparison of single-cell and single-nucleus RNA-seq methodologies.

A. snDrop-seq clusters from the human brain visualized by t-SNE (from Fig. 1B). B. Mouse retina data generated from single cells (sc) using the Drop-seq method and previously annotated\(^7\) were visualized using a maximum of 100 cells per cluster by t-SNE. C. Mouse embryonic brain (E18) data sets generated on the 10X Genomics platform were clustered and visualized by t-SNE. D. Embryonic human midbrain data sets generated on the Fluidigm C1 platform and previously annotated\(^6\) were visualized by t-SNE. E-H. Histograms of detected molecules (number of UMI) for associated quality filtered data sets. I-L. Histograms of detected genes for associated quality filtered data sets. M-P. Coverage (number of UMI per gene detected) for associated quality filtered data sets. Q. Scatter plot showing average expression values for protein-coding genes across quality filtered snDrop-seq samples (occ1-21, log transformed) and their associated genic length. R. Scatter plot showing average expression values for protein-coding genes across all quality filtered scDrop-seq samples (log transformed) and their associated genic length.
Supplementary Figure 5

scTHS-seq sequencing metrics.

**A-B.** Frequency of unique reads in each possible barcode combination in the human (hg38 noAlt) and mouse (mm10 noAlt) datasets. Clonal read statistics were rounded to the nearest whole percent before calculating. **C-D.** Frequency of percent clonal reads in each possible barcode combination in the human (hg38 noAlt) and mouse (mm10 noAlt) datasets. Clonal read statistics were rounded to the nearest whole percent before calculating. **E.** Collision density plot showing the ratio of unique reads mapped to human and unique reads mapped to mouse in each unique barcode combination (data with <500 reads were excluded). See Methods for determination of which data sets are considered collisions. **F.** Theoretical and experimentally derived collision rate percentages. Note that collision rates cannot be directly measured for mouse or human only nuclei. **G.** Violin plot of unique read counts for nuclei that pass filter and nuclei that pass filter with multi-mappers removed. The blue dots represent average number of unique reads and the red dots represent the median number of unique reads. **H.** Violin plot as in (G) showing only unique read count range of 0 to 80,000. **I.** Visualization of merged single cell data and individual single cell datasets from the visual cortex. All unique reads that passed filter were combined to generate the merged tract. Single cell datasets that had reads present in the chromosome 17 locus from 8,100,000 to 8,250,000 bp were randomly sampled and plotted in the single cell tracts section. Each black dot represents a unique read. UCSC genes are overlaid at the bottom.
Supplementary Figure 6

scTHS-seq resolves major neuronal and non-neuronal subpopulations.

A. t-SNE plot for all data sets. Cluster dendrogram showing unsupervised grouping of all non-neuronal, excitatory and inhibitory neuronal cell types or subtypes across cortical and cerebellar regions (top right). t-SNE plot showing batch identity (bottom left). t-SNE plot showing read depth (bottom right).

B. Same as for A except showing only visual cortex (Occ or occipital cortex, BA17) data.

C. Same as for A except showing only frontal cortex.

D. Same as for A except showing only cerebellum. For cluster proportions see Table S2.
snDrop-seq accurately resolves human brain cell identities.

A. Correlation heatmaps comparing averaged snDrop-seq data from all regions sampled with average expression values from RNA-seq data from: mouse pooled cortical cell types\(^{25}\) (left, nf = newly formed); human pooled temporal lobe cell types\(^{26}\) (right). B. Correlation heatmaps comparing average visual cortex data with average expression values from single-cell RNA-seq data from the mouse visual cortex\(^{37}\) (left) and human temporal lobe\(^{29}\) (right). C. Correlation heatmap of snDrop-seq-identified neuronal subtypes (visual cortex or frontal cortex) compared with subtypes previously identified using the C1 single nucleus Smart-seq\(^+\) pipeline\(^{4}\) (SNS, across cortical regions). D. Pair-wise correlation values for all cell types and subtypes resolved across cortical and cerebellar brain regions. E. Cell type sampling rates from single-cell RNA-seq on the temporal lobe\(^{39}\) compared with snDrop-seq sampling rates from the visual and frontal cortices.
Supplementary Figure 8

Subtype-specific gene expression.

A. Heatmap of top 10 differentially expressed marker genes enriched across excitatory neuron subtypes (Table S3). B. Heatmap of the top 10 differentially expressed marker genes enriched across inhibitory neuron subtypes (Table S3). C. Heatmap of the top 10 differentially expressed marker genes enriched across cerebellar clusters (Table S3). D. Top panel: RNA ISH counts showing number of positive cells for CBLN2 and PCP4 (chromogenic image shown) in image fields spanning the pial layer to the white matter. Lower panel: RNA ISH counts for SLC17A7 and EYA4 single positive cells. E. RNA ISH stains (Table S9) of the visual cortex showing select markers from Fig. 2C, and predicted spatial distribution of associated In neuron subtypes.
Supplementary Figure 9

Cerebellar subtype-specific expression.

A. Schematic of the cerebellar hemisphere cytoarchitecture as shown in Fig. 2E. B. Violin plots of expression values for type-specific marker genes specifically for cerebellar data. C. Protein staining (Table S10) for select cell-type specific markers shown in (B). Arrows indicate positively stained cells. D. Representative RNA ISH stainings for GAD1⁺SORCS3⁺ (left) and GAD1⁺SORCS3⁻ (right) Purk neurons that were quantified in Fig. 2H.
Supplementary Figure 10

Probabilistic inference of differential expression and differential accessibility signals.

A-C. Predicting differential accessibility of sites from differential expression. Astrocytes (Ast) and Oligodendrocytes (Oli) identified independently from snDrop-seq and scTHS-seq data analysis are used for training and testing. Genes highly upregulated in Ast snDrop-seq population predict sites highly accessibility in the Ast scTHS-seq population and similarly for Oli. A. A receiver operating characteristic (ROC) curve is used to illustrate the accuracy of predicting the differentially accessible sites within the proximity of the differentially-expressed genes. B. The ROC curve shows the performance of cell classification based on a multi-site predicted differential accessibility signature. In this case, we assess the ability of the average expected differential accessibility inferred from the differential expression signature to be able to distinguish Ast and Oli groups in the scTHS-seq data. C. Predicted differentially accessible sites in Ast and Oli are visualized with red indicating higher aggregate accessibility of oligodendrocyte-associated genes and blue for astrocyte. True cell population assignments are labeled. D-F. Predicting differential expression from differential accessibility profiles. D. ROC of predicting differential expression of an individual gene is shown. E. ROC curve and visualization of the combined differential expression scores inferred from the differential accessibility signature comparing Ast and Oli groups. F. Predicted differentially expressed genes in Ast and Oli are visualized with red indicating higher aggregate expression of oligodendrocyte-associated genes and blue for astrocyte. True cell population assignments are labeled.
Supplementary Figure 11

Transcription factors associated with cell-type-specific chromatin accessibility tend to be highly expressed in the associated cell type.

A. Oli-associated TFs are significantly upregulated in Oli based on the fold-change comparing TF expression in oligodendrocytes to neurons. B. Oli-associated TFs are significantly depleted in neurons based on the fold-change comparing neurons to oligodendrocytes.
Supplementary Figure 12

Destiny diffusion maps.

A. 3D visualization of Destiny components 1, 2, and 3 from visual cortex scTHS-seq data for OPCs and Oli only. Cell population assignments from the original clustering analysis are shown in blue for OPCs and red for Oli. B. 2D visualization of Destiny components 1, 2. C-D. Same plots as A-B colored by batch. E. Heatmap of all significantly (Z > 1.96) differentially upregulated genes corresponding to OPC, iOli and mOli.
Supplementary Figure 13

Pseudotemporal projection of all oligodendrocyte data across regions.

A. Single-cell ordering OPC and Oli data along a remyelination trajectory was generated using Monocle (see Methods). Top plot shows cluster identities (OPC, OPC_Cer and Oli), bottom plot shows regional identities (Cerebellum or CBL, Frontal Cortex and Visual Cortex). B. Pseudotemporal heatmap showing gene expression modules (I to VI), ordered based on expression dynamics progressing along the predicted remyelination trajectory, and generated from gene sets shown in Fig. S12E. Top 5 gene ontologies for each module are indicated.
Mapping of disease risk variants to specific brain cell types in four brain diseases and seven non-brain diseases.

Z-scores for the enrichment of GWAS SNPs in the open chromatin were overlaid onto the Ex, In, Oli, OPC, Ast, End, Mic cell populations (Table S8). Three autoimmune diseases (Crohn’s disease, Celiac disease, T1D) were included. Dark purple and purple represent a significant Z-score over 1.96, where light purple, gray and light green represent an insignificant Z-score, and green represents a significant negative association with a Z-score less than -1.96.

Supplementary Figure 14