Supplemental Information

Dissecting cell-type composition and activity-dependent transcriptional state in mammalian brains by massively parallel single-nucleus RNA-Seq

Peng Hu[1,2,3], Emily Fabyanic[1,2,3], Deborah Kwon[1,2], Sheng Tang[1,2], Zhaolan Zhou[1,2], and Hao Wu[1,2,4,*]

[1]Department of Genetics
[2]Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA 19104, USA
[3]These authors contributed equally to this work
[4]Lead Contact
[*]Correspondence: haowu2@pennmedicine.upenn.edu (H.W.)

Supplemental Figures

Figure S1: Quality control and validation of sNucDrop-Seq. Related to Figure 1.

Figure S2: Experimental reproducibility, subtype-specific marker gene expression and alternative mRNA processing events. Related to Figure 1 and 2.

Figure S3: Sub-type expression signatures of excitatory neurons identified by sNucDrop-Seq agree with previous studies. Related to Figure 3.

Figure S4: Expression of IEG genes and PTZ-induced differential gene expression in cortical cell types. Related to Figure 3 and 4.

Supplemental Tables

Table S1: Sample information and sequencing metrics. Related to Figure 1.

Table S2: Co-expression patterns and statics of selected activity-regulated genes (ARGs) in excitatory neuronal sub-clusters. Related to Figure 3.

Figure S1

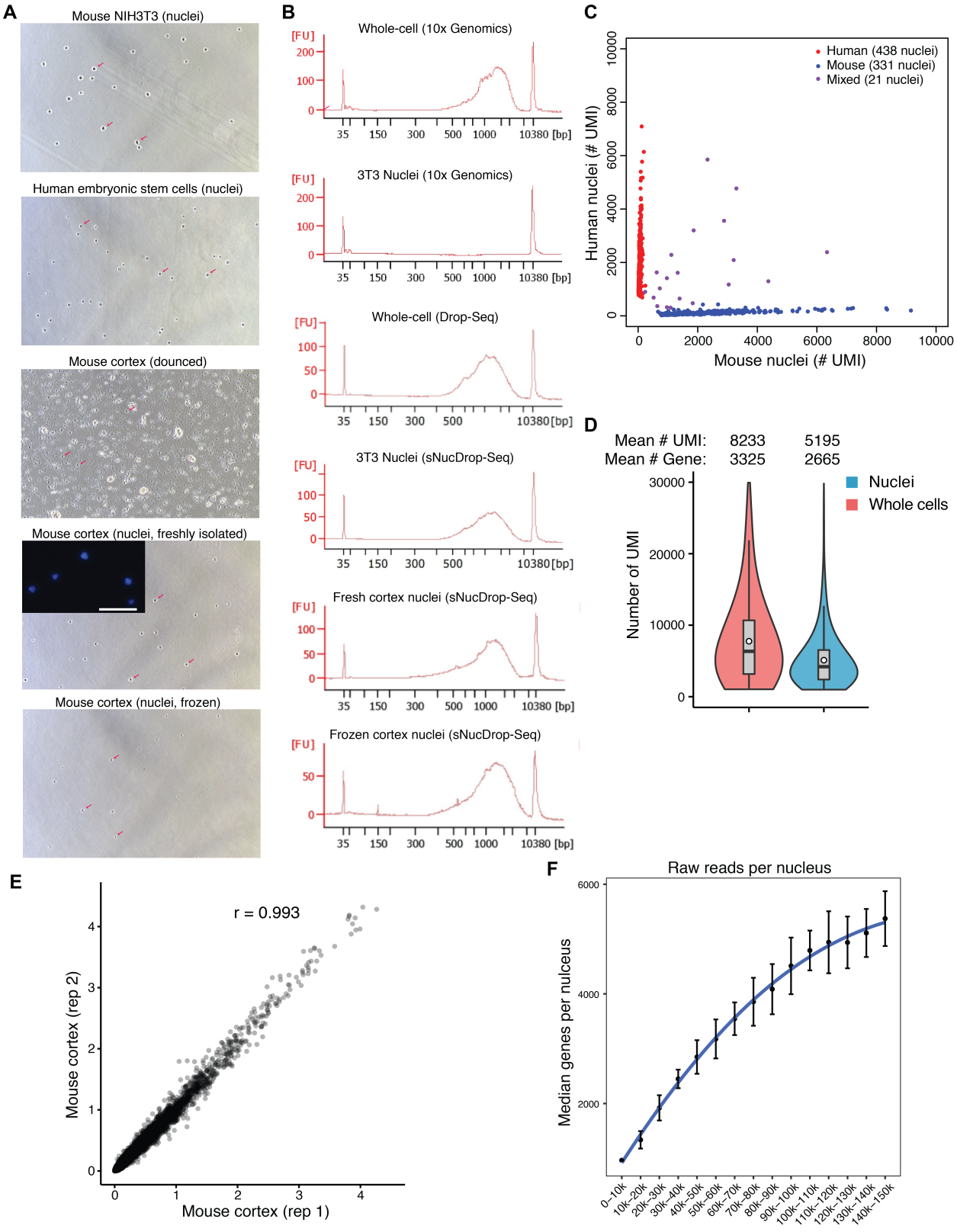

A
Mouse NIH3T3 (nuclei)
Human embryonic stem cells (nuclei)
Mouse cortex (dounced)
Mouse cortex (nuclei, freshly isolated)
Mouse cortex (nuclei, frozen)
B
Whole-cell (10x Genomics)
3T3 Nuclei (10x Genomics)
Whole-cell (Drop-Seq)
3T3 Nuclei (sNucDrop-Seq)
Fresh cortex nuclei (sNucDrop-Seq)
Frozen cortex nuclei (sNucDrop-Seq)
[FU]
35 150 300 500 1000 10380 [bp]
C
Human (438 nuclei)
Mouse (331 nuclei)
Mixed (21 nuclei)
Human nuclei (# UMI)
Mouse nuclei (# UMI)
D
Mean # UMI: 8233 5195
Mean # Gene: 3325 2665
Nuclei
Whole cells
Number of UMI
E
r = 0.993
Mouse cortex (rep 2)
Mouse cortex (rep 1)
F
Raw reads per nucleus
Median genes per nulceus
0–10k
10k–20k
20k–30k
30k–40k
40k–50k
50k–60k
60k–70k
70k–80k
80k–90k
90k–100k
100k–110k
110k–120k
120k–130k
130k–140k
140k–150k

Figure S1: Quality control and validation of sNucDrop-Seq. Related to Figure 1.

(A) After sucrose gradient centrifugation and filtering through cell mesh, mouse (NIH3T3) or human (H7 embryonic stem cell (ESC) line) nuclei were visualized by phase-contrast microscopy (10x). After dounce homogenization, mouse cortical nuclei were visualized by phase-contrast microscopy (10x) before (top) or after (bottom) sucrose gradient centrifugation. Red arrows indicate nuclei before or after sucrose gradient centrifugation. The inlet indicates fluorescent image of purified nuclei stained with DNA intercalating dye Hoechst 33342 (10 ng/μL). Scale bar, 50 μm.

(B) Bioanalyzer electropherogram of amplified cDNA shown for samples prepared from whole-cell or nuclei by different platforms (10x Genomics platform or Drop-Seq/sNucDrop-Seq). FU, fluorescence units.

(C) Multi-species nuclei-mixing experiment measures sNucDrop-Seq specificity, by sequencing a mix of human (ESCs) and mouse (NIH3T3) nuclei. Scatter plot shows the number of transcripts (UMIs) associated with annotated human (y-axis) or mouse (x-axis) transcripts for each nucleus (dot). Nuclei with >80% human transcripts are labeled as human (red), and nuclei with >80% mouse transcripts are labeled as mouse (blue). Nuclei with a relatively high percentage of both human and mouse transcripts are labeled as mixed (purple). Of the 790 nuclei that passed quality filter (>800 UMIs), 21 (2.66%) had a mixed phenotype.

(D) Violin plots illustrating number of transcripts (UMIs) detected by sNucDrop-Seq of nuclei (blue) or whole cells (red) isolated from mouse 3T3 cells by Drop-Seq. Center line: median; circle: mean; limits: first and third quartile; whiskers, ±1.5 Inter quartile range (IQR). Indicated on top are the number of cells or nuclei (>= 800 genes detected), mean number of UMIs per cells/nuclei, and mean number of genes per cells/nuclei.

(E) Scatter plot showing the high correlation of average expression levels [log (normalized UMI counts + 1)] between two sNucDrop-Seq biological replicates of mouse cortical cells.

(F) Median number of genes detected per nucleus at scaling ranges of raw reads per nucleus. Data from two independent experiments were included, mean±s.e.m.

Figure S2

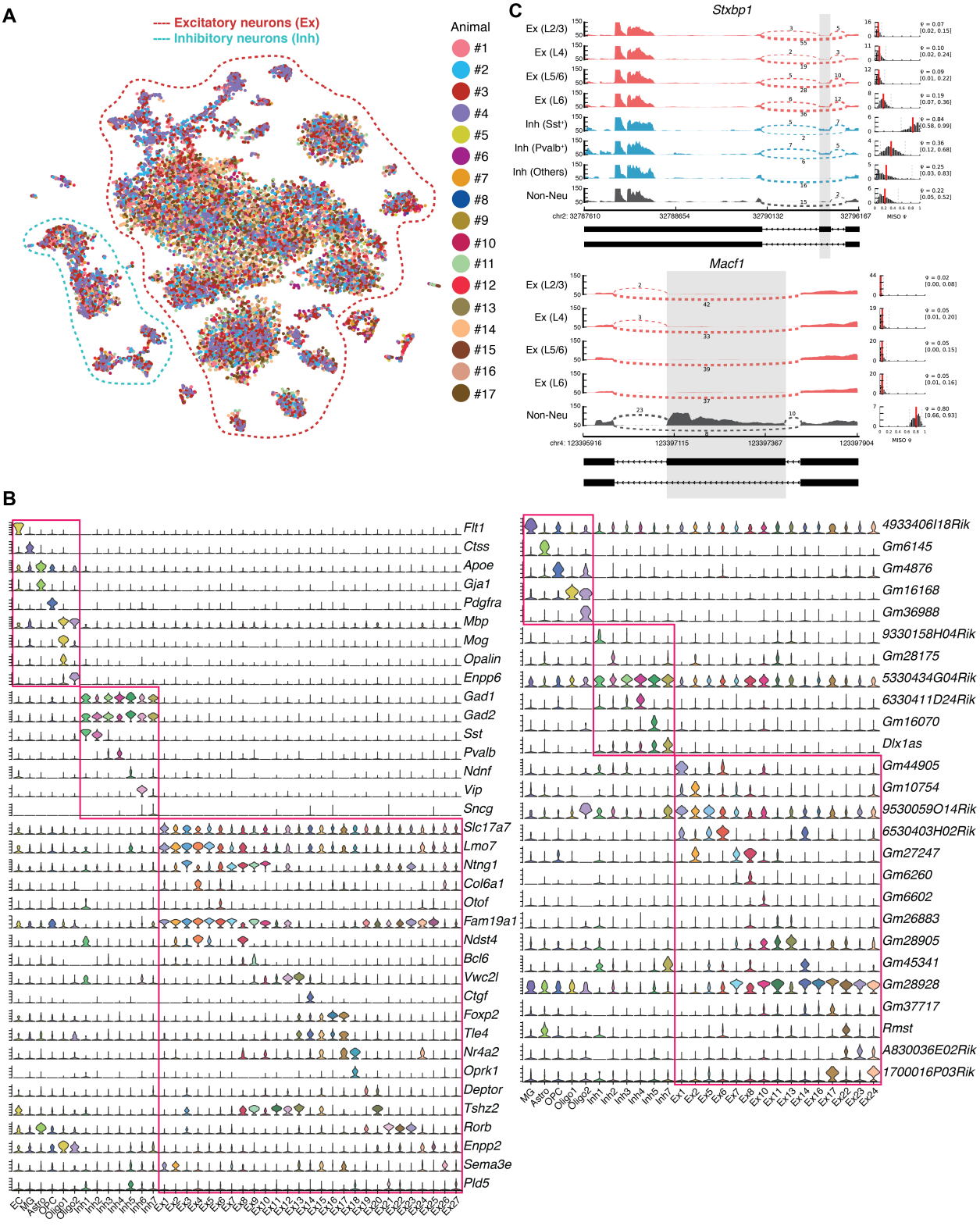

A
---- Excitatory neurons (Ex)
---- Inhibitory neurons (Inh)
Animal
#1
#2
#3
#4
#5
#6
#7
#8
#9
#10
#11
#12
#13
#14
#15
#16
#17
C
Stxbp1
Ex (L2/3)
Ex (L4)
Ex (L5/6)
Ex (L6)
Inh (Sst+)
Inh (Pvalb+)
Inh (Others)
Non-Neu
chr2: 32787610
32788654
32790132
32796167
MISO Ψ
Macf1
chr4: 123395916
123397115
123397367
123397904
B
Flt1
Ctss
Apoe
Gja1
Pdgfra
Mbp
Mog
Opalin
Enpp6
Gad1
Gad2
Sst
Pvalb
Ndnf
Vip
Sncg
Slc17a7
Lmo7
Ntng1
Col6a1
Otof
Fam19a1
Ndst4
Bcl6
Vwc2l
Ctgf
Foxp2
Tle4
Nr4a2
Oprk1
Deptor
Tshz2
Rorb
Enpp2
Sema3e
Pld5
4933406I18Rik
Gm6145
Gm4876
Gm16168
Gm36988
9330158H04Rik
Gm28175
5330434G04Rik
6330411D24Rik
Gm16070
Dlx1as
Gm44905
Gm10754
9530059O14Rik
6530403H02Rik
Gm27247
Gm6260
Gm6602
Gm26883
Gm28905
Gm45341
Gm28928
Gm37717
Rmst
A830036E02Rik
1700016P03Rik

Figure S2: Experimental reproducibility, subtype-specific marker gene expression and alternative mRNA processing events. Related to Figure 1 and 2.

(A) The tSNE plot (the same plot as in Figure 1C) with all nuclei colored according to animal identification (ID) code. Clusters corresponding to excitatory (red dashed line) and inhibitory (blue dashed line) neurons are grouped together.

(B) Protein-coding (left) and non-coding (right) marker genes identified for neuronal and non-neuronal subtypes. Violin plots illustrating select protein-coding (left) and non-coding marker gene expression for excitatory neuronal (Ex1-27), inhibitory neuronal (Inh 1-7) and non-neuronal (Astro, OPC, Oligo, MG, EC) cell clusters.

(C) Cell-type-specific alternative splicing events revealed by sNucDrop-Seq. Sashimi plot of exon skipping events associated with the *Stxbp1* (top) or *Macf1* (bottom) in each cell type. Left panel: upper histograms show the read coverage along the exons. Junction reads are visualized by a dashed curve to indicate the splice event, and the number of junction reads were indicated in the center of dashed curve. The bottom schematic diagram shows the exon (wider bars) and intron (black line) structure of two annotated isoforms (inclusion and exclusion of an exon), with the direction of transcription indicated by arrows along the intron. The variably present exon is highlighted with grey rectangle. Right panel: the posterior distribution of MISO expression estimates (Ψ or PSI: percentage spliced in) that represent the fraction of the inclusion isoform for a gene in each cell type. The red line indicates the Ψ estimate, and the lower and upper 95% confidence intervals are plotted as dotted grey lines. The actual value of Ψ and 95% CI are showed in the right of the histograms.

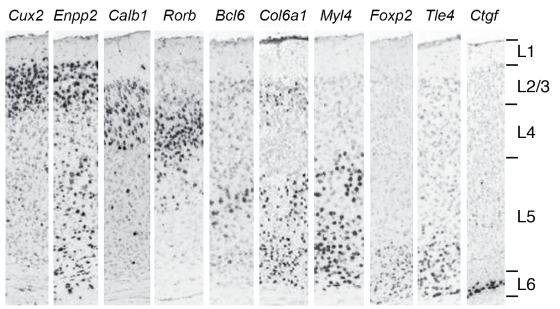
Figure S3

A

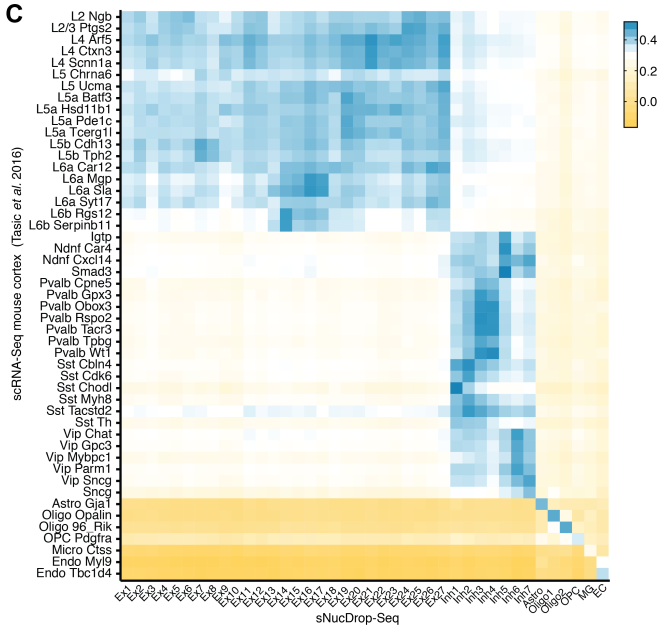
15,353 nuclei (ex neurons)

L2/3 (36.7%)	Ex25 (n=3,628; 23.6%): *Enpp2, Calb1* Ex6 (n=286; 1.86%): *Otof, Ngb* Ex24 (n=212; 1.38%): *Ptgs2, Calb1* Ex5 (n=461; 3.00%): *Otof, Ptgs2* Ex26 (n=788; 5.13%): *Inhba, Ngb* Ex27 (n=716; 4.66%): *Inhba, Enpp2*
L4 (17.7%)	Ex22 (n=976; 6.36%): *Rorb, Cux2* Ex21 (n=1,104; 7.19%): *Rorb, Tph2* Ex23 (n=632; 4.12%): *Rorb, Hsd11b1*
L5 (20.5%)	Ex20 (n=180; 1.17%): *Batf3, Deptor* Ex1 (n=164; 1.07%): *Arf5, Col6a1* Ex9 (n=88; 0.57%): *Bcl6, Hsd11b1* Ex19 (n=883; 5.75%): *Deptor, Batf3* Ex12 (n=142; 0.92%): *Pde1c, Batf3* Ex7 (n=1,033; 6.73%): *Tcerg1l, Bcl6* Ex10 (n=169; 1.10%): *Qrfpr, Ntng1* Ex8 (n=77; 0.50%): *Stac, Bcl6* Ex3 (n=128; 0.83): *Chrna6, Ntng1* Ex2 (n=276; 1.80%): *Bcl11b, Stac*
L6a (19.2%)	Ex4 (n=119; 0.78%): *Col6a1, Sla* Ex13 (n=335; 2.18%): *Myl4, Tle4* Ex16 (n=1,847; 12.0%): *Foxp2, Ly6d* Ex17 (n=91; 0.59%): *Foxp2, Tle4* Ex11 (n=423; 2.76%): *Mgp, Ddit4l* Ex15 (n=128; 0.83%): *Mgp, Tle4*
L6b (3.04%)	Ex14 (n=305; 1.99%): *Ctgf, Tnmd* Ex18 (n=162; 1.06%): *Rgs12, Cdh13*

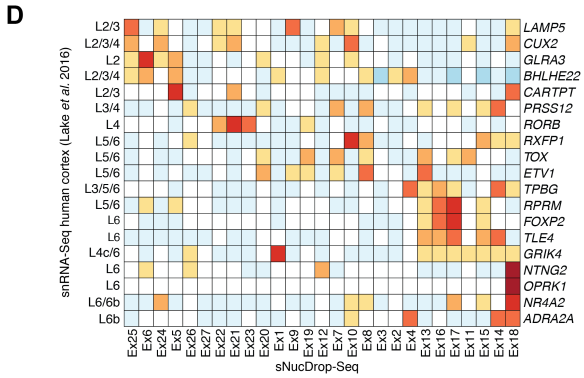
B



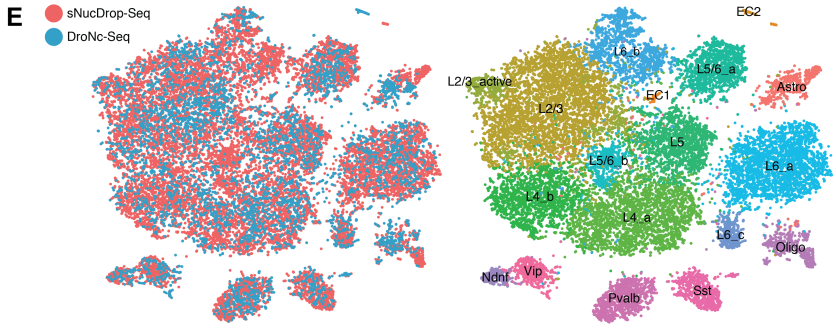
C



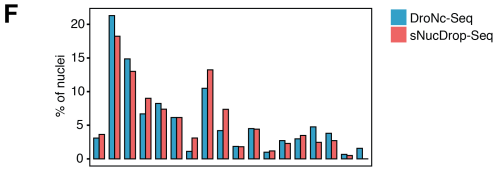
D



E



F



G

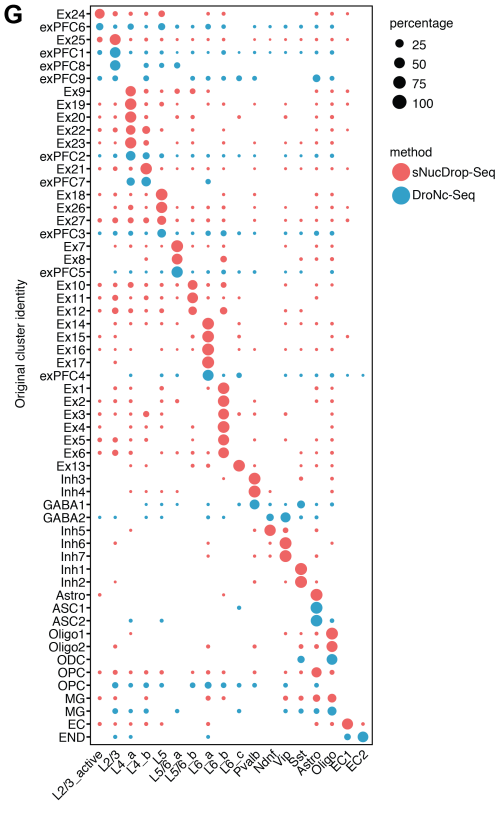


Figure S3: Sub-type expression signatures of excitatory neurons identified by sNucDrop-Seq agree with previous studies. Related to Figure 3.

(A) Summary of cortical excitatory neuronal subtypes identified by sNucDrop-Seq. Glutamatergic neuronal subtypes are grouped according to cortical layer distribution. Also shown are number of nuclei per subtype and representative marker genes for each subtype.

(B) RNA in situ hybridization (ISH) showing layer-specific expression of selected markers in the mouse adult cortex (postnatal day 56, Allen Brain Atlas).

(C) Cell-type-specific expression signatures (identified by sNucDrop-Seq) agree with previously published work. Pairwise correlations of the average expression (Methods) for the genes in each cell-type signature defined by sNucDrop-Seq and cell-types defined by single-cell RNA-Seq in the mouse visual cortex (Tasic et al., 2016).

(D) Heatmap showing select human marker gene expression for cortical excitatory neuronal sub-populations identified in Figure 1C. The human marker gene list is derived from single-nucleus RNA-Seq of adult human cortex (Lake et al., 2016).

(E) Integrative analysis of sNucDrop-Seq and DroNc-Seq data sets of adult mouse cortex. A tSNE plot of 17,876 sNucDrop-Seq (fresh mouse whole cortex) and 5,337 DroNc-Seq (frozen mouse prefrontal cortex) (Habib et al., 2017) nuclei profiles (>800 genes per nucleus). Nuclei (dots) are colored by method (left) and cluster membership (right) that is labeled *post hoc* according to cell-type, cortical layer distribution (L2/3, L4, L5/6, and L6), and transcriptional state of neuronal activity regulated genes (L2/3_active versus L2/3).

(F) Fraction of either sNucDrop-Seq (red) or DroNc-Seq (blue) nuclei from each cortical cell-types. Cell types are defined as in Figure S3E.

(G) Mapping of original cortical cell-type clusters defined by sNucDrop-Seq (red) and DroNc-Seq (blue) to clusters defined by combined data sets in Figure S3E. Dot plot shows the proportion of nuclei in each cluster defined by two different methods that were classified to each cluster defined by combined data sets. DroNc-Seq clusters: exPFC, excitatory neurons; GABA, inhibitory neurons; ASC, astrocytes; OPC, oligodendrocyte precursor cells; ODC, oligodendrocytes; MG, microglia; END, endothelial cells. sNucDrop-Seq clusters: Ex, excitatory neurons; Inh, inhibitory neurons; Astro, astrocytes; OPC, oligodendrocyte precursor cells; Oligo, oligodendrocytes; MG, microglia; EC, endothelial cells;

Figure S4

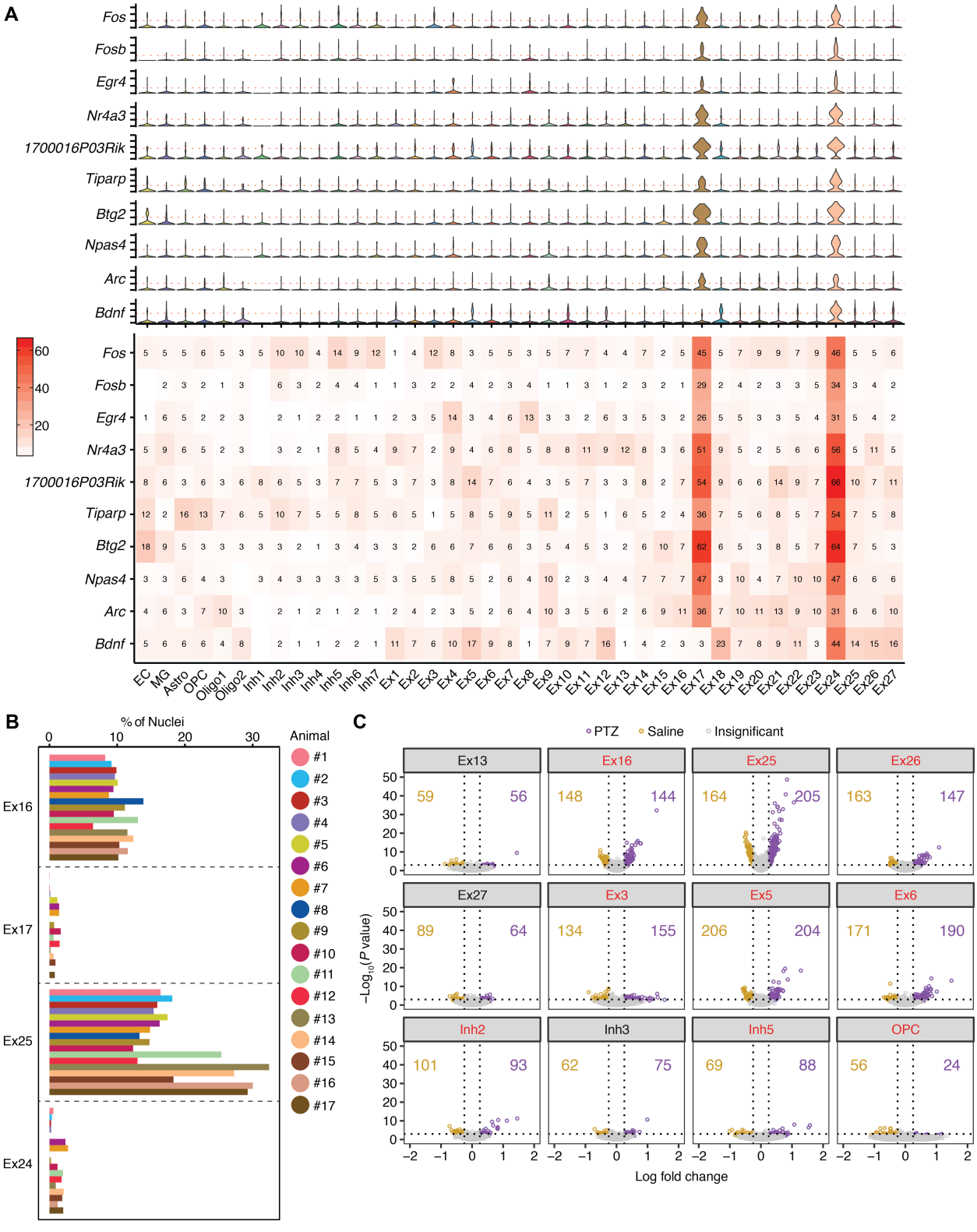

A
Fos
Fosb
Egr4
Nr4a3
1700016P03Rik
Tiparp
Btg2
Npas4
Arc
Bdnf
60
40
20
EC
MG
Astro
OPC
Oligo1
Oligo2
Inh1
Inh2
Inh3
Inh4
Inh5
Inh6
Inh7
Ex1
Ex2
Ex3
Ex4
Ex5
Ex6
Ex7
Ex8
Ex9
Ex10
Ex11
Ex12
Ex13
Ex14
Ex15
Ex16
Ex17
Ex18
Ex19
Ex20
Ex21
Ex22
Ex23
Ex24
Ex25
Ex26
Ex27
B
% of Nuclei
0
10
20
30
Animal
#1
#2
#3
#4
#5
#6
#7
#8
#9
#10
#11
#12
#13
#14
#15
#16
#17
Ex16
Ex17
Ex25
Ex24
C
PTZ
Saline
Insignificant
Ex13
Ex16
Ex25
Ex26
Ex27
Ex3
Ex5
Ex6
Inh2
Inh3
Inh5
OPC
59
56
148
144
164
205
163
147
89
64
134
155
206
204
171
190
101
93
62
75
69
88
56
24
−Log10(P value)
Log fold change

Figure S4: Expression of IEG genes and PTZ-induced differential gene expression in cortical cell types. Related to Figure 3 and 4.

(A) Violin plots (top) showing the distribution of the expression level of select IEGs in all cortical cell-types. Dashed line (mean + 2 SD) marks the background level of IEG expression that is used to call IEG-expressing nuclei. Heatmap (bottom) showing percentage of nuclei expressing select IEGs in all cortical cell-types. The number indicates the percentage of IEG-expressing nuclei over total number of nuclei in each cluster. SD, standard deviation.

(B) Percentage of nuclei for each animal (colored according to animal ID in Figure S2A) in clusters Ex16, Ex17, Ex 24 and Ex 25.

(C) Identification of differentially expressed genes between PTZ and saline treatment for representative clusters. Significant genes with p-values less than 0.01 and absolute natural log fold changes greater than 0.25 was colored (purple or yellow). The cluster identified by GSEA as enriched for activity-dependent genes (FDR<0.2) is highlighted in red.

Table S1: Sample information and sequencing metrics. Related to Figure 1.

Sample name[1]	Data type	Animal id[2]	Mouse strain	Cell-type/ Brain region	Sex	Age (approx.)	Drug treatment	Total reads	Uniquely mapped reads	UMI #	Cell/nucleus # (>800 genes)	Read # (x1000) per cell/nucleus[3]
cell-3T3	Drop-Seq	N/A	N/A	NIH3T3	N/A	N/A	N/A	95,093,080	62,024,451	9,550,135	1,160	25.1
nuclei-3T3	sNucDrop-Seq	N/A	N/A	NIH3T3	N/A	N/A	N/A	152,135,088	99,618,264	10,309,036	1,987	22.7
nuclei-ctx-1	sNucDrop-Seq	#8	C57BL/6J	whole cortex	M	6-10 wk	N/A	26,721,375	17,753,847	1,371,024	185	36.6
nuclei-ctx-2	sNucDrop-Seq	#9	C57BL/6J	whole cortex	M	6-10 wk	N/A	50,695,703	35,815,095	3,136,674	474	32.2
nuclei-ctx-3	sNucDrop-Seq	#5	C57BL/6J	whole cortex	M	6-10 wk	N/A	28,669,035	28,669,035	2,019,055	276	30.2
nuclei-ctx-4	sNucDrop-Seq	#10	C57BL/6J	whole cortex	M	6-10 wk	N/A	94,491,277	73,085,661	4,673,340	529	60.3
nuclei-ctx-5	sNucDrop-Seq	#6	C57BL/6J	whole cortex	M	6-10 wk	N/A	80,648,373	65,889,268	7,565,915	1,311	24.1
nuclei-ctx-6	sNucDrop-Seq	#12	C57BL/6J	whole cortex	M	6-10 wk	N/A	92,204,765	75,715,967	7,116,538	817	45.2
nuclei-ctx-7	sNucDrop-Seq	#7	C57BL/6J	whole cortex	M	6-10 wk	N/A	106,311,622	88,142,394	4,853,611	950	40.2
nuclei-ctx-8	sNucDrop-Seq	#15	C57BL/6J	whole cortex	M	6-10 wk	N/A	64,698,455	52,819,015	4,321,996	996	19.6
nuclei-ctx-9	sNucDrop-Seq	#11	C57BL/6J	whole cortex	M	6-10 wk	N/A	96,415,484	73,829,793	3,219,678	1,584	7.3
nuclei-ctx-10	sNucDrop-Seq	#16	C57BL/6J	whole cortex	M	6-10 wk	N/A	63,396,599	48,694,637	901,662	537	8.5
nuclei-ctx-11	sNucDrop-Seq	#17	C57BL/6J	whole cortex	M	6-10 wk	N/A	66,081,696	53,142,553	1,538,201	675	10.8
nuclei-ctx-12	sNucDrop-Seq	#14	C57BL/6J	whole cortex	M	6-10 wk	N/A	122,608,270	93,228,418	5,200,104	1,929	8.8
nuclei-ctx-13	sNucDrop-Seq	#13	C57BL/6J	whole cortex	M	6-10 wk	N/A	75,667,980	56,551,517	3,756,479	1,359	9.6
nuclei-ctx-saline1	sNucDrop-Seq	#1	C57BL/6J; FVB/NJ	whole cortex[4]	M	10 wk	Saline	92,873,910	75,540,881	6,754,151	2,490	10.4
nuclei-ctx-PTZ1	sNucDrop-Seq	#2	C57BL/6J; FVB/NJ	whole cortex[4]	M	10 wk	PTZ	69,812,807	57,531,848	5,146,303	2,326	6.9
nuclei-ctx-saline2	sNucDrop-Seq	#3	C57BL/6J; FVB/NJ	whole cortex[4]	M	10 wk	Saline	85,944,486	70,173,310	6,424,499	2,489	8.8
nuclei-ctx-PTZ2	sNucDrop-Seq	#4	C57BL/6J; FVB/NJ	whole cortex[4]	M	10 wk	PTZ	84,952,663	70,034,705	4,809,316	2,093	9.2

1: sample id in the GEO database (GSE106678).
2: animal id in Figure S2A and S4B.
3: only cells or nuclei with >800 genes detected were included for calculation.
4: a small portion of striatum was included during dissection for comparison.

Table S2: Coexpression patterns and statics of selected activity-regulated genes (ARGs) in excitatory neuronal sub-clusters. Related to Figure 3.

ARG coexpression*	Ex17 (n=91 nuclei)	Ex16 (n=1,847 nuclei)	Odds ratio (Ex17/ Ex16)	*P*-value (Ex17 vs Ex16)**
Fos-/Egr1-	12.1%	65.7%	0.07	4.86E-25
Fos-/Egr1+	33.0%	27.5%	1.30	2.80E-01
Fos+/Egr1-	12.1%	3.6%	3.71	6.83E-04
Fos+/Egr1+	42.9%	3.2%	22.62	1.01E-29
Fos-/Nr4a3-	17.6%	86.4%	0.03	1.21E-45
Fos-/Nr4a3+	27.5%	6.9%	5.12	6.47E-09
Fos+/Nr4a3-	26.4%	5.8%	5.76	1.68E-09
Fos+/Nr4a3+	28.6%	0.9%	42.67	1.56E-25
Fos-/Pcsk1-	20.9%	81.1%	0.06	4.82E-33
Fos-/Pcsk1+	24.2%	12.2%	2.30	1.93E-03
Fos+/Pcsk1-	25.3%	5.5%	5.78	3.13E-09
Fos+/Pcsk1+	29.7%	1.2%	33.23	1.13E-24
ARG coexpression	**Ex24 (n=212 nuclei)**	**Ex25 (n=3,628 nuclei)**	**Odds ratio (Ex24/ Ex25)**	***P*-value (Ex24 vs Ex25)**
Fos-/Egr1-	22.2%	71.9%	0.11	8.55E-48
Fos-/Egr1+	24.5%	22.3%	1.13	4.47E-01
Fos+/Egr1-	14.2%	3.1%	5.08	8.29E-11
Fos+/Egr1+	39.2%	2.6%	23.87	4.07E-62
Fos-/Nr4a3-	19.8%	88.8%	0.03	1.50E-108
Fos-/Nr4a3+	26.9%	5.5%	6.36	7.34E-22
Fos+/Nr4a3-	20.3%	5.0%	4.79	9.98E-14
Fos+/Nr4a3+	33.0%	0.7%	68.08	3.16E-71
Fos-/Pcsk1-	20.8%	77.6%	0.08	3.87E-64
Fos-/Pcsk1+	25.9%	16.7%	1.75	9.73E-04
Fos+/Pcsk1-	17.5%	4.3%	4.70	5.53E-12
Fos+/Pcsk1+	35.8%	1.5%	37.56	9.46E-67

*: The ARG coexpression patterns are defined based on the expression level of ARGs, as described in Figure 3D.
**: The *P*-value associated with the odds ratio was calculated by Fisher's exact test.