ALLEN INSTITUTE for BRAIN SCIENCE

Allen Cell Types Database

TECHNICAL WHITE PAPER: TRANSCRIPTOMICS

OVERVIEW

The complexity of a cell's function can, at some level, be described by its transcriptional signature. Recent technical advances in RNA sequencing (RNA-Seq) allow the creation of a transcriptomic inventory at the cellular level. This Technical White Paper describes the methods used for generating transcriptional data from nuclei or whole cells derived from both mouse and human tissue, using standardized laboratory operating procedures.

For this particular study, the anatomical areas selected are the mouse visual system, non-visual sensorimotor cortical areas, and cortical tissue from the human brain made available through tissue donation. Specific structures include the dorsal part of the mouse lateral geniculate complex (LGd), human and macaque lateral geniculate nucleus (LGN), mouse primary visual cortex (VISp), mouse secondary motor area (MOs), also functionally defined as the anterior lateral motor area (ALM), and the middle temporal gyrus of human neocortex (MTG). Methods used in the data collection workflow include tissue preparation, RNA amplification and library preparation for RNA-Seq, RNA-Seq data processing, and clustering. The predominant source of tissue from mouse brain was acquired from animals between the ages of P51-P59, with some additional collection of data from animals up to age P91. Images of each tissue slice were acquired to aid in brain region identification and registration to a standard spatial reference, the Allen Mouse Common Coordinate Framework (CCF). The anatomical regions of interest listed above were microdissected, treated with protease and single cells were collected by fluorescence-activated cell sorting (FACS). For tissue from human MTG and LGN, slices were prepared from adult postmortem human brain tissue using a vibrating microtome. Macaque LGN was similarly prepared, but used fresh tissue. Sections were stained with a fluorescent Nissl dye and individual cortical or LGN layers were microdissected and nuclei were isolated via dounce homogenization. Nuclei were then stained with an antibody against NeuN to label neuronal nuclei. FACS was used to capture single NeuN-positive and NeuN-negative nuclei. Cellular and nuclear mRNA from both mouse and human was reverse transcribed to cDNA, then amplified using the SMART-Seq v4 kit (Takara), followed by library construction using the Nextera XT kit (Illumina). Single cell/nucleus libraries were sequenced on a HiSeq 2500 instrument (Illumina) to generate 50 base-pair paired-end reads. The reads were aligned and after alignment, quality control (QC) analyses were performed.

EXPERIMENTAL DESIGN

Part of this research goal is to understand cortical diversity at the cellular level, by creating a cell census of the visual cortex and its input, LGd. The core and shell regions of mouse dorsal LGd receive different inputs from the retina: on-off direction-selective retinal ganglion cells project to the shell but not core region (Piscopo, *et al*. 2013). In addition, approximately 20% of neurons in rodent LGd are GABAergic (Gabbott, *et al*. 1986) and 20% are located within the shell region (Piscopo, *et al*. 2013). The sampling strategy for LGd leveraged four mouse Cre lines to capture pan-neuronal (*Snap25-IRES2-Cre;Ai14),* glutamatergic (*Slc17a6-IRES2-Cre;Ai14*) and GABAergic (*Gad2-IRES-Cre;Ai14*, *Slc32a1-IRES2-Cre;Ai14*) neurons from core and shell regions of LGd based on their relative proportions *in vivo*. Monte Carlo simulations were used to estimate the number of cells needed to capture (with 95% confidence) at least 16 cells of a cell type as rare as 2% of all LGd neurons. With 16 cells,

discrimination between cell types as similar as *Sst*+ interneuron subtypes was expected (ie, Sst Cbln4 versus a mixture of Sst Myh8, Sst Cdk6, and Sst Th), as identified in mouse primary visual cortex (Tasic, *et al*. 2016).

An initial study of transcriptionally-defined cell diversity in mouse primary visual cortex identified 42 neuronal types and 7 non-neuronal types (Tasic, *et al*. 2016). The current cortical survey builds upon this initial study to provide a more comprehensive overview of cell types in VISp and also a comparison with an alternate cortical region, secondary motor area (MOs, which includes ALM) (Tasic, *et al*. 2017). The current dataset contains 133 putative types (117 neuronal and 16 non-neuronal), sampled to sufficient depth to capture substantial transcriptomic diversity. The sampling strategy for neocortex leveraged four mouse Cre lines to capture pan-neuronal (*Snap25-IRES2-Cre;Ai14),* glutamatergic (*Slc17a7-IRES2-Cre;Ai14*) and GABAergic (*Gad2-IRES-Cre;Ai14*, *Slc32a1-IRES2-Cre;Ai14*) neurons from 4-5 layers of cortex (L1, L2/3, L4 (in V1 only), L5, and L6) based on their relative proportions *in vivo*. Additional cells were isolated from transgenic Cre recombinase lines selected to enrich for broad classes of cell types that were underrepresented in the initial 4 Cre lines. Similar to the preliminary study, the number of cells captured reflected the coverage goal (with 95% confidence) of at least 16 cells of a given type, at a frequency of 0.4% of glutamatergic neurons and 0.1% of GABAergic neurons.

A low-bias approach to profile cell type diversity in human cortex was sought, constrained by the challenge of working with precious and limited tissue sources. Individual layers of cortex were dissected from tissue covering the MTG derived from human brain, and nuclei were dissociated and sorted using the neuronal marker NeuN. In total, nuclei were sampled from four postmortem donor brains and four neurosurgical donor brains. Profiled nuclei included approximately 90% neurons and 10% glia, and layer sampling was based on the relative number of neurons in each layer. Additional nuclei were sampled from deep layers of cortex to reflect the transcriptomic diversity of neurons in those layers. More than 15,000 neurons were sampled, resulting in an estimated detection of neuronal cell types as rare as 0.2% of all neurons. This estimate was based on Monte Carlo simulations as described for sampling of mouse LGd.

A similar low-bias sampling strategy was employed for human and macaque LGN. Briefly, individual layers of LGN were dissected from tissue sections, processed for nuclei isolation, and isolated nuclei were sorted using NeuN staining. For human LGN, nuclei were sampled from four postmortem tissue donors and nuclei were ~90% neuronal and ~10% non-neuronal. For macaque, nuclei were sampled from 3 specimens and only neuronal nuclei were captured. In total, >1,500 nuclei were sampled from human LGN and >1,000 were sampled from macaque LGN.

TISSUE PREPARATION

Single Cell Samples from Mouse VISp, MOs, and LGd

Tissue samples were obtained from adult (postnatal day P53-P59) mice (LGd was isolated from male mice, and tissue from both males and females was used for cortical samples) carrying one or two recombinase transgenes (Cre, FlpO) and a recombinase-dependent reporter transgene (see **Table 1** and **Table 2**). Detailed descriptions of recombinase and reporter lines can be found at http://connectivity.brain-map.org/transgenic. In addition, retrogradely labeled cells were isolated from reporter mice infected with AAV2-retro-Ef1a-Cre, SADB-rabies-dGdL-Cre, or CAV-Cre. For MOs (ALM) retrograde experiments, AAV2-retro-CAG-GFP or AAV2-retro-CAG-tdTomato were injected into C57BL/6J mice.

Mice were anesthetized with 5% isoflurane and intracardially perfused with either 25 or 50 ml of ice cold, oxygenated artificial cerebral spinal fluid (ACSF.I) at a flow rate of 9 ml per minute until the liver appeared clear, or the full volume of perfusate had been flushed through the vasculature. The brain was then rapidly dissected and mounted for coronal slice preparation (rostral end at base for LGd and VISp collections, caudal end at base for MOs collections) on the chuck of a Compresstome VF-300 vibrating microtome (Precisionary Instruments) (see **Figure 1** for workflow). Using a custom photodocumentation system (Mako G125B PoE camera with custom integrated software), a blockface image of the coronal or semi-coronal brain surface was acquired before each section was sliced at 250 µm intervals. The slice was then hemisected along the midline, and the left hemisphere for LGd or both hemispheres for cortical samples were transferred to ACSF.I.

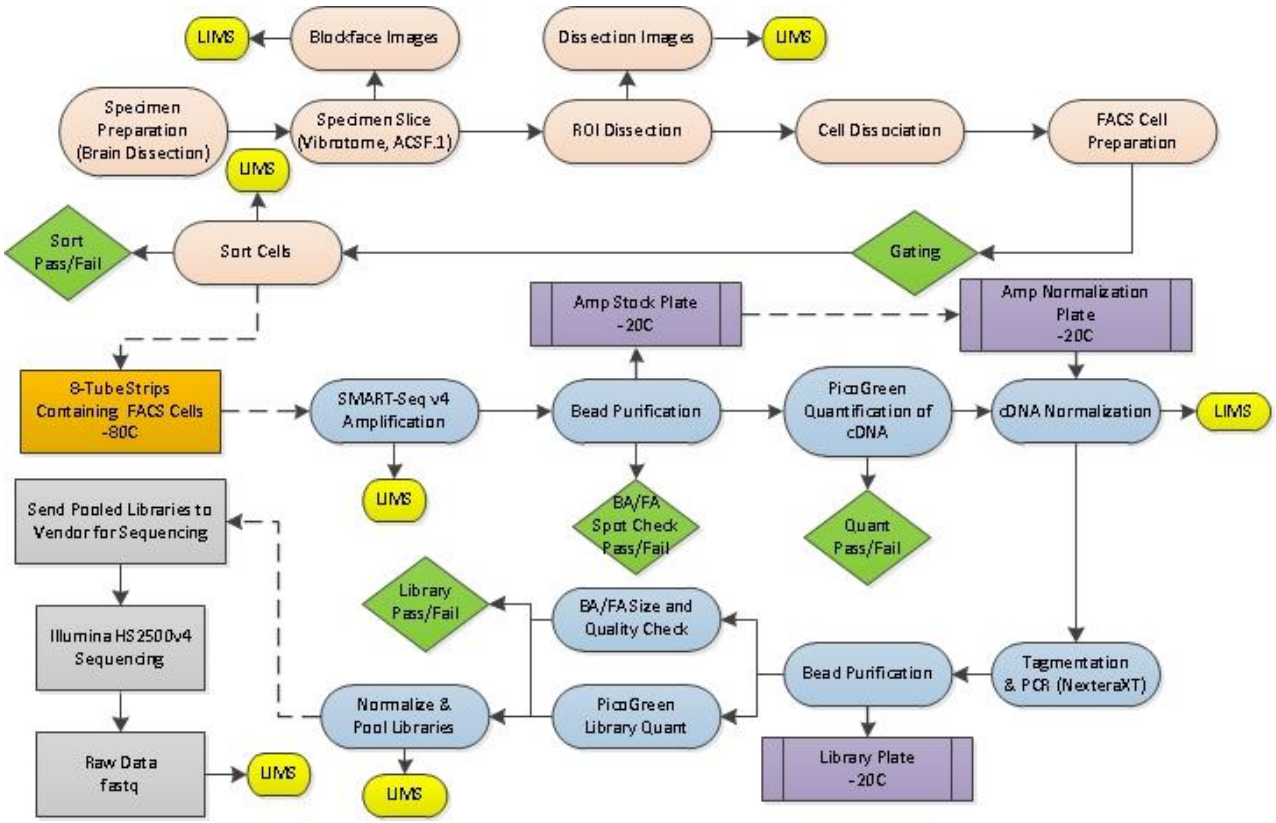

Figure 1. Workflow for tissue preparation and RNA-Seq data generation.
The main steps of the entire workflow include brain dissection, Region of Interest (ROI) dissection, cell sorting, SMART-Seq v4 Ultra amplification, bead purification, PicoGreen quantification, cDNA normalization, tagmentation, PCR, bead purification, and library normalization and pooling. Quality control checkpoints are indicated in green. Interface points with the Laboratory Information Management System (LIMS) are shown in yellow. Abbreviations: BA/FA, Bioanalyzer/Fragment Analyzer; Quant, quantification.

Each slice-hemisphere was transferred into a Sylgard-coated dissection dish containing 3 ml of chilled, oxygenated ACSF.I. Brightfield and fluorescent images between 4X and 20X were obtained of the intact tissue with a Nikon Digital Sight DS-Fi1 or a Sentech STC-SC500POE camera mounted to a Nikon SMZ1500 dissecting microscope. To guide anatomical targeting for dissection, boundaries were identified by trained anatomists, comparing the blockface image and the slice image to a matched plane of the Allen Reference Atlas. For LGd, samples for RNA-Seq were targeted for either core or shell enrichment (see **Table 1**). In general, three to five slices were sufficient to capture the targeted region of interest, allowing for expression analysis along the anterior/posterior axis. The region of interest was then dissected and both brightfield and fluorescent images of the dissections were acquired for secondary verification. The dissected regions were transferred in ACSF.I to a microcentrifuge tube, and stored on ice. This process was repeated for all slices containing the target region of interest, with each region of interest deposited into a new microcentrifuge tube.

After all regions of interest were dissected, the ACSF.I was removed and 1 ml of 2 mg/ml pronase in ACSF.I was added. Tissue was digested at room temperature (approximately 22°C) for a duration that consisted of adding 15 minutes to the age of the mouse (in days; *i.e.*, P53 specimen had a digestion time of 68 minutes). After digestion, the pronase solution was removed and replaced by 1 ml of ACSF.I supplemented to a concentration of 1% FBS (Fetal Bovine Serum). The tissue was washed two more times with the same solution with the third wash being 500 µl for final sample volume. The sample was then triturated using fire-polished glass pipettes of decreasing bore sizes (600, 350 and 150 µm for LGd, and 300 and 150 µm for cortical samples). For triturating larger regions of interest (ROI), pipettes with three decreasing bore sizes were used (600, 300 and 150 µm). The cell suspension was incubated on ice in preparation for fluorescence-activated cell sorting (FACS). *Note: Samples collected after 12/16/2016 had 0.0132M trehalose added to all solutions used after the point of slicing to improve cell viability and yield.*

Table 1. Summary of mouse lines used for cortical sampling.

	Genotype	Cells ALM	Cells VISp
Retro-gradely labeled Cells		331	0
	Ai14(RCL-tdT)/Ai14(RCL-tdT)	1013	311
	Ai14(RCL-tdT)/wt	37	1148
	Ai14(RCL-tdT)/wt;Ai148(TIT2L-GC6f-ICL-tTA2)/wt	0	90
Cre or Flp Line Cells	Calb1-IRES2-Cre/wt;Ai14(RCL-tdT)/wt	0	157
	Chat-IRES-Cre-neo/wt;Ai14(RCL-tdT)/wt	17	39
	Chrna2-Cre_OE25/wt;Ai14(RCL-tdT)/wt	112	99
	Chrna2-Cre_OE25/wt;Pvalb-T2A-Dre/wt;Ai66(RCRL-tdT)/wt	47	60
	Chrnb3-Cre_SM93/wt;Ai14(RCL-tdT)/wt	6	41
	Crh-IRES-Cre_ZJH/wt;Sst-IRES-FlpO/wt;Ai65(RCFL-tdT)/wt	5	75
	Ctgf-T2A-dgCre/wt;Ai14(RCL-tdT)/wt	228	166
	Ctgf-T2A-dgCre/wt;Snap25-LSL-F2A-GFP/wt	0	195
	Cux2-CreERT2/wt;Ai14(RCL-tdT)/wt	62	0
	Esr2-IRES2-Cre/wt;Ai14(RCL-tdT)/wt	0	38
	Etv1-CreERT2/wt;Pvalb-T2A-FlpO/wt;Ai65(RCFL-tdT)/wt	10	82
	Gad2-IRES-Cre/wt;Ai14(RCL-tdT)/wt	719	1423
	Glt25d2-Cre_NF107/wt;Ai14(RCL-tdT)/wt	0	4
	Gng7-Cre_KH71/wt;Ai14(RCL-tdT)/wt	177	103
	Htr3a-Cre_NO152/wt;Ai14(RCL-tdT)/wt	149	154
	Htr3a-Cre_NO152/wt;Pvalb-T2A-Dre/wt;Ai66(RCRL-tdT)/wt	96	30
	Ndnf-IRES2-dgCre/wt;Ai14(RCL-tdT)/wt	123	113
	Ndnf-IRES2-dgCre/wt;Slc32a1-IRES2-FlpO/wt;Ai65(RCFL-tdT)/wt	0	55
	Nkx2-1-CreERT2/wt;Ai14(RCL-tdT)/wt	64	55
	Nos1-CreERT2/wt;Ai14(RCL-tdT)/wt	54	64
	Nos1-CreERT2/wt;Sst-IRES-FlpO/wt;Ai65(RCFL-tdT)/wt	25	7
	Nr5a1-Cre/wt;Ai110(RCL-FnGF-nT)/wt	0	105
	Nr5a1-Cre/wt;Ai14(RCL-tdT)/wt	0	134
	Ntsr1-Cre_GN220/wt;Ai14(RCL-tdT)/wt	130	99
	Oxtr-T2A-Cre/wt;Ai110(RCL-FnGF-nT)/wt	0	154
	Oxtr-T2A-Cre/wt;Ai14(RCL-tdT)/wt	0	132
	Oxtr-T2A-Cre/wt;Pvalb-T2A-FlpO/wt;Ai65(RCFL-tdT)/wt	13	22
	Pdyn-T2A-CreERT2/wt;Ai14(RCL-tdT)/wt	10	105
	Penk-IRES2-Cre-neo/wt;Ai14(RCL-tdT)/wt	0	128
	Pvalb-IRES-Cre/wt;Ai14(RCL-tdT)/wt	465	498
	Rasgrf2-T2A-dgFlpO/wt;Ai65F/wt	25	84
	Rbp4-Cre_KL100/wt;Ai14(RCL-tdT)/wt	625	981
	Rorb-IRES2-Cre-neo/wt;Ai14(RCL-tdT)/wt	97	117
	Rorb-P2A-FlpO/wt;Ai65F/wt	147	184
	Scnn1a-Tg2-Cre/wt;Ai14(RCL-tdT)/wt	0	43
	Scnn1a-Tg3-Cre/wt;Ai14(RCL-tdT)/wt	30	136
	Sim1-Cre_KJ18/wt;Ai14(RCL-tdT)/wt	103	97
	Slc17a6-IRES-Cre/wt;Ai14(RCL-tdT)/wt	9	32
	Slc17a7-IRES2-Cre/wt;Ai14(RCL-tdT)/wt	679	1244
	Slc17a8-iCre/wt;Ai14(RCL-tdT)/wt	94	56
	Slc17a8-IRES2-Cre/wt;Ai14(RCL-tdT)/wt	0	154
	Slc32a1-IRES-Cre/wt;Ai14(RCL-tdT)/wt	909	1200
	Snap25-IRES2-Cre/wt;Ai14(RCL-tdT)/wt	1036	2849
	Sst-IRES-Cre/wt;Ai14(RCL-tdT)/wt	198	368
	Sst-IRES-Cre/wt;Pvalb-T2A-Dre/wt;Ai66(RCRL-tdT)/wt	47	32
	Tac1-IRES2-Cre/wt;Ai14(RCL-tdT)/wt	0	141
	Tac1-IRES2-Cre/wt;Sst-IRES-FlpO/wt;Ai65(RCFL-tdT)/wt	19	33
	Th-Cre_FI172/wt;Ai14(RCL-tdT)/wt	15	66
	Tlx3-Cre_PL56/wt;Ai14(RCL-tdT)/wt	102	136
	Trib2-F2A-CreERT2/wt;Snap25-LSL-F2A-GFP/wt	111	167
	Vip-IRES-Cre/wt;Ai14(RCL-tdT)/wt	228	348
	Vipr2-IRES2-Cre/wt;Ai14(RCL-tdT)/wt	149	137
	Vipr2-IRES2-Cre/wt;PhiC31-neo/Ai14(RCL-tdT)	0	125
	Vipr2-IRES2-Cre/wt;Pvalb-T2A-FlpO/wt;Ai65(RCFL-tdT)/wt	0	4
	Vipr2-IRES2-Cre/wt;Slc32a1-T2A-FlpO/wt;Ai65(RCFL-tdT)/wt	0	42

FACS preparation involved adding 2 µl of 4'-6-diamidino-2-phenylindole (DAPI) (2 mg/ml) to the 1.0 ml or 1 µl to the 500 µl ACSF.I-1% FBS cell suspension. The suspension was then filtered through a fine-mesh cell strainer (35 µm for samples collected before 2/22/2017, 70 µm for samples collected after 2/22/2017) and sorted by excluding DAPI positive events and debris, and gating to include red fluorescent events (tdTomato-positive cells) or green fluorescent events (GFP-positive cells). Single cells were collected into strip tubes containing 11.5 µl of collection buffer (SMART-Seq v4 lysis buffer 0.83x, (Takara #634894), RNase Inhibitor (0.17U/µl) and ERCCs (External RNA Controls Consortium) (Baker, *et al*.; Risso, *et al*.) (MIX1 at $1x10^{-8}$)). After sorting, the cells were subjected to centrifugation and then stored at -80°C.

Single Nucleus Samples from Human MTG and LGN

To prepare and archive tissues from suitable cases, whole postmortem brain specimens were bisected through the midline and individual hemispheres were embedded in alginate for slabbing. Coronal brain slabs were cut at 0.5-1cm intervals through each hemisphere and the slabs were then frozen in a bath of dry ice and isopentane, vacuum sealed in freezer bags to prevent frost damage, and stored at -80°C until use. MTG was identified on slabs of interest and selected slabs were transferred to -20°C overnight to equilibrate them for blocking. Slabs were then placed on a -20°C cold table to maintain temperature during sub-blocking. For MTG, the portion of the slab corresponding to this region was subdivided into small blocks for further sectioning. Each block was individually vacuum sealed and returned to the -80°C freezer until the time of processing. For LGN, the entire structure was removed from the slab and kept as a single intact block, which was vacuum sealed and returned to the -80°C freezer until the time of use. To section tissue on the vibrating microtome, tissue blocks were removed from the -80°C freezer and rapidly thawed in ice-cold buffer containing PBS supplemented with 10mM DL-Dithiothreitol (DTT). Thawed blocks were mounted on a vibrating microtome and sectioned at 500µm in the coronal plane in the same cold buffer solution.

Neurosurgical donor tissue (MTG only) was received from patients undergoing surgery for epilepsy or brain tumors. The tissue blocks received were distal, normal cortical tissue removed in order to access underlying pathological brain tissues. Tissue was transported in chilled ACSF and sectioned at 350µm on a VF-300 compresstome. Sections used for transcriptomic analysis were transferred to microcentrifuge tubes and flash-frozen in a slurry of dry ice and ethanol. Frozen sections were stored at -80°C until use.

To facilitate microdissection of individual cortical or LGN layers, sections were transferred to a fluorescent Nissl staining solution (Neurotrace 500/525, ThermoFisher Scientific, 1:100 dilution) prepared in PBS with 10mM DTT and 0.5% RNasin Plus RNase inhibitor (Promega) and were kept on ice. After staining for 5 minutes, sections were visualized under a fluorescence dissecting microscope (Leica) and individual cortical or LGN substructures (layers, regions) were microdissected using a needle blade micro-knife (Fine Science Tools). Microdissected tissue pieces were transferred to microcentrifuge tubes containing PBS supplemented with 10mM DTT and 0.5% RNasin Plus on ice. A fluorescent image of each section was acquired before and after dissection. For LGN, a brightfield image of each tissue section was also acquired before and after dissection.

After dissection was completed, microdissected tissue pieces were transferred into nuclei isolation medium containing 10mM Tris pH 8.0, 250mM sucrose, 25mM KCl, 5mM $MgCl_2$, 0.1% Triton-X 100, 0.5% RNasin Plus, 1X protease inhibitor (Promega), and 0.1mM DTT and placed into a 1ml dounce homogenizer (Wheaton). Tissue was homogenized to liberate nuclei using 10 strokes of the loose dounce pestle followed by 10 strokes of the tight pestle. Homogenate was strained through a 30µm cell strainer and centrifuged at 900x*g* for 10 minutes to pellet nuclei. Nuclei were then resuspended in staining buffer containing PBS supplemented with 0.8% nuclease-free BSA and 0.5% RNasin Plus. Mouse monoclonal anti-NeuN-PE-conjugated primary antibody (EMD Millipore, FCMAB317PE) was applied to nuclei preparations at a dilution of 1:500 and samples were incubated for 30 min at 4°C. Control samples were incubated with mouse IgG1,k Isotype control (BD Biosciences, 550671). After primary antibody incubation, samples were centrifuged for 5 min at 400x*g* to pellet nuclei and pellets were resuspended in PBS with 0.8% BSA and 0.5% RNasin Plus. Samples were supplemented with DAPI at a concentration of 0.1µg/ml prior to FACS. Single nuclei were captured by gating on DAPI-positive events, excluding debris and doublets, and then gating on PE (NeuN) signal, which allowed for the isolation of either NeuN-positive (neuronal) or NeuN-negative events (non-neuronal). Strip tubes

containing FACS isolated single nuclei were briefly centrifuged and frozen at -80°C until use. The total number of NeuN-positive and and NeuN-negative nuclei included in this resource from each layer from each donor are included in **Table 2** and **Table 3**, respectively. The total number of nuclei from each donor collected from each LGN region are included in **Table 4**.

Table 2. Summary of NeuN-positive human MTG sampling.

Layer	H16.03.004	H16.06.002	H16.06.008	H16.06.009	H16.24.010	H200.1023	H200.1025	H200.1030
L1	0	0	0	0	24	211	199	562
L2	0	0	0	0	20	1004	140	821
L3	0	0	0	0	25	1320	263	1687
L4	0	0	0	0	243	1142	82	1264
L5	208	86	197	220	31	1039	194	1512
L6	0	0	0	0	16	1143	344	1008

Table 3. Summary of NeuN-negative human MTG sampling.

Layer	H16.03.004	H16.06.002	H16.06.008	H16.06.009	H16.24.010	H200.1023	H200.1025	H200.1030
L1	0	0	0	0	1	20	14	45
L2	0	0	0	0	4	68	16	65
L3	0	0	0	0	2	50	21	115
L4	0	0	0	0	3	59	4	79
L5	0	11	0	0	2	39	19	100
L6	0	0	0	0	0	75	38	73

Table 4. Summary of Human LGN sampling.

Donor	Koniocellular - Anterior	Koniocellular - Posterior	Magnocellular - Anterior	Magnocellular - Posterior	Parvocellular - Anterior	Parvocellular - Posterior
H17.24.001	32	0	79	63	65	75
H200.1023	51	33	122	93	142	86
H200.1025	31	38	74	53	88	94
H200.1030	20	20	87	85	72	73

Single Nucleus Samples from Macaque LGN

Macaque tissue (*Macaca nemestrina*) specimens were received from the Tissue Distribution Program at the University of Washington National Primate Research Center. Three female specimens ranging in age from 2.3-17.9 years were used in the study. At the time of necropsy, macaque LGN tissue was removed and placed in carbogenated ACSF. Tissue was transported in carbogenated ACSF from the University of Washington to the Allen Institute in chilled ACSF and sectioned at 250-350µm on a VF-300 compresstome. Sections used for transcriptomic analysis were transferred to individual microcentrifuge tubes and flash-frozen in a slurry of dry ice and ethanol. Frozen sections were stored at -80°C until use. To isolate nuclei, tissue sections were thawed and processed exactly as described above for human MTG and LGN. The total number of nuclei from each donor collected from each LGN region are included in **Table 5**.

Table 5. Summary Macaque LGN sampling.

Donor	Magnocellular	Parvocellular
Q17.27.002	347	255
Q18.27.001	219	271

RNA AMPLIFICATION, LIBRARY PREPARATION, AND RNA SEQUENCING

Single Cell Samples from Mouse Cortex

SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (Takara #634894) was used per manufacturer's instructions for cDNA synthesis of single cell RNA and subsequent amplification. Single cells were stored in 8-strips at -80°C in 11.5 µl of collection buffer (SMART-Seq v4 lysis buffer at 0.83x, RNase Inhibitor at 0.17 U/µl, and ERCC MIX1 at final $1x10^{-8}$ dilution as described above). Twelve to 36 8-strips were processed at a time (the equivalent of 1-3 96-well plates). At least 1 control strip was used per amplification set, containing 2 wells without cells (termed ERCC), 2 wells without cells or ERCC (termed NTC), and either 4 wells of 10 pg of Mouse Whole Brain Total RNA (Zyagen, MR-201) or 2 wells of 10 pg of Mouse Whole Brain Total RNA (Zyagen, MR-201) and 2 wells of 10 pg Control RNA provided in the Takara kit. Mouse whole cells were subjected to 18 PCR cycles after the reverse transcription step. AMPure XP Bead (Agencourt AMPure beads XP PCR, Beckman Coulter A63881) purification was done using the Agilent Bravo NGS Option A instrument. A bead ratio of 1x was used (50 µl of AMPure XP beads to 50 µl cDNA PCR product with 1 µl of 10x lysis buffer added, as per Takara instructions), and purified cDNA was eluted in 17 µl elution buffer provided by Takara. All samples were quantitated using PicoGreen® on Molecular Dynamics M2 SpectraMax instrument. The samples were then either run on the Agilent Bioanalyzer 2100 using High Sensitivity DNA chips or the Advanced Analytics Fragment Analyzer (96) using the High Sensitivity NGS Fragment Analysis Kit (1bp-6000bp) to qualify cDNA size distribution. Purified cDNA was stored in 96-well plates at -20°C until library preparation.

All samples proceeded through NexteraXT DNA Library Preparation (Illumina FC-131-1096) using NexteraXT Index Kit V2 Sets A-D (FC-131-2001, 2002, 2003, or 2004) or or custom 8bp Unique Design index primers designed and manufactured by IDT (Integrated DNA Technologies). NexteraXT DNA Library prep was done at either 0.5x volume manually or 0.4x volume on the Mantis instrument (Formulatrix). Three different cDNA input amounts were used in generating the libraries: 75pg, 100pg, and 125pg. An aliquot of all amplified cDNA samples was first normalized to 30 pg/µl, 45 pg/ul, or 50 pg/ul with Nuclease-Free Water (Ambion), then this normalized sample aliquot was used as input material into the NexteraXT DNA Library Prep (for a total of 75pg, 100pg, or 125pg input). AMPure XP bead purification was done using the Agilent Bravo NGS Option A instrument. A bead ratio of 0.9x was used (22.5 ul of AMPure XP beads to 25 ul library product, as per Illumina protocol), and all samples were eluted in 22 µl of Resuspension Buffer (Illumina). All samples were quantitated using PicoGreen using Molecular Bynamics M2 SpectraMax instrument. All samples were run on either the Agilent Bioanalyzer 2100 using High Sensitivity DNA chips or the Advanced Analytics Fragment Analyzer (96) using the High Sensitivity NGS Fragment Analysis Kit (1bp-6000bp) for sizing. Molarity was calculated for each sample using average size as reported by Bioanalyzer or Fragment Analyzer and pg/µl concentration as determined by PicoGreen. Samples (5 µl aliquot) were normalized to 2-10 nM with Nuclease-free Water (Ambion), then 2 µl from each sample within one 96-index set was pooled to a total of 192 µl at 2-10 nM concentration. Libraries were further multiplexed at either 96 samples/lane, 192 samples/lane, or 384 samples/lane by pooling multiple 96-sample libraries using compatible Index Sets. A portion of the final library pool was sent to an outside vendor for sequencing on an Illumina HiSeq 2500 instrument. All of the library pools were run using Illumina High Output V4 chemistry. The Broad Institute Inc, Clinical Research Sequencing Platform LLC, performed the majority of the RNA-Sequencing services. Covance Genomics Laboratory, Seattle subsidiary of LabCorp Group of Holdings also performed a portion of the RNA-Sequencing services. Statistics related to RNA-amplification yield, average mapped reads, and other QC statistics can be found in **Tables 7-12** in the RNA-SEQ DATA PROCESSING section below.

Single Cell Samples from Mouse LGd

SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (Takara #634894) was used per manufacturer's instructions for cDNA synthesis of single cell RNA and subsequent amplification. Single cells were stored in 8-strips at -80°C in 11.5µl of collection buffer (SMART-Seq v4 lysis buffer at 0.83x, RNase Inhibitor at 0.17U/µl, and ERCC MIX1 at final $1x10^{-8}$ dilution as described above). Twelve to 24 8-strips were processed at a time (the equivalent of 1-2 96-well plates). At least 1 control strip was used per amplification set, containing 2 wells without cells (termed ERCC), 2 wells without cells or ERCC (termed NTC), and 4 wells of 10pg of Mouse Whole Brain Total RNA (Zyagen, MR-201). AMPure XP Bead (Agencourt AMPure beads XP PCR, Beckman Coulter A63881) purification was done manually for the first amplification set but then was done using the Agilent Bravo NGS Option A instrument. A bead ratio of 1x was used (50µl of AMPure XP beads to 50µl cDNA PCR product with 1µl of 10x lysis buffer added, as per Takara instructions), and purified cDNA was eluted in 17µl elution

buffer provided by Takara. All samples were quantitated using PicoGreen® on Molecular Dynamics M2 SpectraMax instrument. A portion of the samples, and all controls, were run on the Agilent Bioanalyzer 2100 using High Sensitivity DNA chips to qualify cDNA size distribution. Purified cDNA was stored in 96-well plates at -20°C until library preparation.

All samples were proceeded through NexteraXT DNA Library Preparation (Illumina FC-131-1096) using NexteraXT Index Kit V1 or V2 Set A (FC-131-1002 or FC-131-2001). NexteraXT DNA Library prep was done at either 1x, 0.5x, or 0.25x volume (applied to input and all reagents), but otherwise followed manufacturer's instructions. An aliquot of all samples was first normalized to 30pg/µl with Nuclease-Free Water (Ambion), then this normalized sample aliquot was used as input material into the NexteraXT DNA Library Prep. See **Table 7** for a summary of library prep conditions applied to the samples. AMPure XP bead purification was done using 0.9x bead ratio to sample volume, and all samples were eluted in 22µl of Resuspension Buffer (Illumina). As with the Amplification sets, manual bead purification was done for the first Library set, but thereafter bead purification was performed by Agilent Bravo NGS Option A instrument. All samples were run on Agilent Bioanalyzer 2100 using High Sensitivity DNA chips (for sizing), and all samples were quantitated using PicoGreen using Molecular Bynamics M2 SpectraMax instrument. Molarity was calculated for each sample using average size as reported by Bioanalyzer and pg/µl concentration as determined by PicoGreen. Samples (5µl aliquot) were normalized to 2-5nM with Nuclease-free Water (Ambion), then 2µl from each sample within one 96-index set was pooled to a total of 192µl at 2-5nM concentration. A portion of this library pool was sent to an outside vendor for sequencing on an Illumina HiSeq 2500 instrument. Most of the library pools were run using Illumina High Output V4 chemistry, although a few sets were sequenced using the Rapid Run V1 chemistry. Covance Genomics Laboratory, Seattle subsidiary of LabCorp Group of Holdings, performed the majority of RNA sequencing services, with some also provided by EA Genomic Services.

Single Nucleus Samples from Human MTG, LGN and Macaque LGN

SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (Takara #634894) was used per the manufacturer's instructions for cDNA synthesis of single cell RNA and subsequent amplification. FACS-collected single nuclei were stored in 8-strips at -80°C in 11.5 µl of collection buffer (SMART-Seq v4 lysis buffer at 0.83x, RNase Inhibitor at 0.17 U/µl, and ERCC MIX1 at final $1x10^{-8}$ dilution as described above). Twelve to 36 8-strips were processed at a time (the equivalent of 1-3 96-well plates). At least 1 control strip was used per amplification set, containing 2 wells without cells (termed ERCC), 2 wells without cells or ERCC (termed NTC), and either 4 wells of 10 pg of Human Universal Reference Total RNA (Takara 636538) or 2 wells of 10 pg of Human Universal Reference and 2 wells of 10 pg Control RNA provided in the Takara kit. Human and macaque nuclei samples were subjected to 20 cycles (Macaque LGN nuclei) or 21 cycles (Human LGN nuclei) of PCR amplification after the reverse transcription step. AMPure XP Bead (Agencourt AMPure beads XP PCR, Beckman Coulter A63881) purification was done using the Agilent Bravo NGS Option A instrument. A bead ratio of 1x was used (50 µl of AMPure XP beads to 50 µl cDNA PCR product with 1 µl of 10x lysis buffer added, as per Takara instructions), and purified cDNA was eluted in 17 µl elution buffer provided by Takara. All samples were quantitated using PicoGreen® on a Molecular Dynamics M2 SpectraMax instrument. All samples and controls were either run on the Agilent Bioanalyzer 2100 using High Sensitivity DNA chips or the Advanced Analytics Fragment Analyzer (96) using the High Sensitivity NGS Fragment Analysis Kit (1bp-6000bp) to qualify cDNA size distribution. Purified cDNA was stored in 96-well plates at -20°C until library preparation.

All samples proceeded through NexteraXT DNA Library Preparation (Illumina FC-131-1096) using NexteraXT Index Kit V2 Set A, B, C, or D (FC-131-2001, 2002, 2003, or 2004) or custom 8bp Unique Design index primers designed and manufactured by IDT (Integrated DNA Technologies). NexteraXT DNA Library prep was done at either 0.5x volume manually or 0.4x volume on the Mantis instrument (Formulatrix). Reduction in volume was applied to input and all reagents, but otherwise the manufacturer's instructions were followed. Three different cDNA input amounts were used in generating the MTG libraries: 75pg (2,139), 100pg (9,351), and 125pg (12,179). An aliquot of all amplified cDNA samples was first normalized to 30 pg/µl, 45 pg/ul, or 50 pg/ul with Nuclease-Free Water (Ambion), then this normalized sample aliquot was used as input material into the NexteraXT DNA Library Prep (for a total of 75pg, 100pg, or 125pg input). For human LGN, cDNA input was 125pg or 100pg and for macaque LGN it was 100pg. AMPure XP bead purification was done using the Agilent Bravo NGS Option A instrument. A bead ratio of 0.9x was used (22.5 ul of AMPure XP beads to 25 ul library

product, as per Illumina protocol), and all samples were eluted in 22 µl of Resuspension Buffer (Illumina) or Elution Buffer (Qiagen 19086). All samples were run on either the Agilent Bioanalyzer 2100 using High Sensitivity DNA chips or the Advanced Analytics Fragment Analyzer (96) using the High Sensitivity NGS Fragment Analysis Kit (1bp-6000bp) for sizing. All samples were quantitated using PicoGreen on the Molecular Dynamics M2 SpectraMax instrument. Molarity was calculated for each sample using average size as reported by Bioanalyzer or Fragment Analyzer and pg/µl concentration as determined by PicoGreen. Samples (5 µl aliquot) were normalized to 2-10 nM with Nuclease-free Water (Ambion), then 2 µl from each sample within one 96-index set was pooled to a total of 192 µl at 2-10 nM concentration. Libraries were further multiplexed at either 96 samples/lane or 192 samples/lane. A portion of the final library pool was sent to an outside vendor for sequencing on an Illumina HiSeq 2500 instrument. All of the library pools were run using Illumina High Output V4 chemistry. RNA sequencing services were provided by Covance Genomics Laboratory, Seattle subsidiary of LabCorp Group of Holdings, and The Broad Institute Genome Sequencing Platform.

ANALYSIS: DATA PROCESSING FOR RNA SEQUENCES

RNA-Seq alignment and data processing was done the same way for all single cell and single nucleus RNA-Seq data, except that different versions of the genome and transriptome were used for each species. For mouse, raw read (fastq) files were aligned to the mm10 mouse genome sequence (Genome Reference Consortium, 2011) with the RefSeq transcriptome version GRCm38.p3 (current as of 01/15/2016) and updated by removing duplicate Entrez gene entries from the gtf reference file. For human, raw read files were aligned to the GRCh38 human genome sequence (Genome Reference Consortium, 2011) with the RefSeq transcriptome version GRCh38.p2 (current as of 4/13/2015) and likewise updated by removing duplicate Entrez gene entries from the gtf reference file. For macaque, raw read files were aligned to a new assembly and annotation of the rhesus macaque genome from the University of Nebraska Non-human Primate Genome center (https://www.unmc.edu/rhesusgenechip/index.htm; MacaM_v7.8.2, released 07/17/2016). The details of this work are described in Zimin, *et al.* 2014.

In all samples, Illumina sequencing adapters were clipped from the reads using the fastqMCF program (Aronesty, *et al*., 2011) for alignment. After clipping, the paired-end reads were mapped using Spliced Transcripts Alignment to a Reference (STAR v2.5.3) (Dobin, *et al*., 2013) using default settings. STAR uses and builds it own suffix array index which considerably accelerates the alignment step while improving on sensitivity and specificity, due to its identification of alternative splice junctions. Reads that did not map to the genome were then aligned to synthetic constructs (i.e. ERCC) sequences and the *E.coli* genome (version ASM584v2). The final results files included quantification of the uniquely mapped reads (raw exon and intron counts for the transcriptome-mapped reads). Also, part of the final results files are the percentages of reads mapped to the transcriptome, to ERCC spike-in controls, and to *E.coli*. For all data other than macaque LGN, reads were aligned to tRNA, ncRNA, and rRNA in addition to mRNA.

Pre-classification was performed to facilitate quality analysis. Cells and nuclei were classified as Glutamatergic, GABAergic, or Non-Neuronal based on qualitative post-hoc annotation of clusters (see below for description of clustering strategies). Clusters were assigned to a broad class based on expression of canonical markers and all cells or nuclei in a cluster were assigned the same broad type. Cells not assigned to a cluster were designated “unassigned” and include cells that failed QC for any reason, as well as those that passed QC but were excluded for other reasons (e.g., were initially assigned to a cluster only found in one human donor). QC statistics for each data set grouped by broad class are shown in **Tables 7-12**. While not shown in the tables, the average fraction of reads aligned to *E. coli* and ERCC standards is <0.02% for all mouse and human data analyzed. QC statistics for controls run alongside single cell and single nucleus RNA-Seq data are shown in **Table 13**.

Table 7. QC Statistics for data from broad classes in mouse cortex (VISp).

		Mouse VISp			
		GABAergic	Glutamatergic	unassigned	Non-Neuronal
	Sample Count	6,125	7,366	1,164	758
Mean Values	Total Reads	2,383,336	2,432,407	1,948,033	2,240,754
	% Total Reads Aligned	93.51	94.13	83.09	89.96
	% Reads Aligned to Exons	69.09	66.73	52.04	57.89
	% Reads Aligned to Introns	10.51	13.38	14.27	15.34
	Amp Yield (ng)	5.57	6.78	5.79	2.08
	Genes Detected (CPM>0, exon + intron reads)	9,671	10,625	8,770	5,464
	Genes Detected (FPKM>0, exon reads)	8,696	9,594	7,570	4,084

Table 8. QC Statistics for data from broad classes in mouse cortex (ALM).

		Mouse ALM			
		GABAergic	Glutamatergic	unassigned	Non-Neuronal
	Sample Count	4,409	4,539	495	625
Mean Values	Total Reads	2,645,091	2,753,395	1,856,274	2,406,494
	% Total Reads Aligned	88.96	89.19	75.14	89.86
	% Reads Aligned to Exons	64.30	65.84	47.12	59.86
	% Reads Aligned to Introns	10.18	10.13	8.20	14.14
	Amp Yield (ng)	33.62	25.19	12.04	2.14
	Genes Detected (CPM>0, exon + intron reads)	10,039	11,846	6,361	5,628
	Genes Detected (FPKM>0, exon reads)	9,107	10,897	5,230	4,268

Table 9. QC Statistics for data from broad classes in human cortex (MTG).

		Human MTG			
		GABAergic	Glutamatergic	unassigned	Non-neuronal
	Sample Count	4,099	10,323	599	907
Mean Values	Total Reads	2,614,747	2,640,341	2,730,801	2,501,674
	% Total Reads Aligned	90.80	92.07	83.51	87.20
	% Reads Aligned to Exons	31.14	28.37	24.16	27.13
	% Reads Aligned to Introns	41.52	44.92	40.84	39.12
	Amp Yield (ng)	8.67	15.63	15.28	4.45
	Genes Detected (CPM>0, exon + intron reads)	8,437	9,937	9,772	6,186
	Genes Detected (FPKM>0, exon reads)	4,594	6,123	5,737	2,428

Table 10. QC Statistics for data from broad classes in mouse LGd

		Mouse LGd			
		GABAergic	Glutamatergic	unassigned	Non-Neuronal
	Sample Count	688	1,090	122	96
Mean Values	Total Reads	2,462,402	2,486,381	2,254,053	2,422,032
	% Total Reads Aligned	92.34	93.05	86.99	89.45
	% Reads Aligned to Exons	63.16	58.74	53.24	60.58
	% Reads Aligned to Introns	15.17	20.53	16.13	13.67
	Amp Yield (ng)	6.68	10.55	6.99	3.00
	Genes Detected (CPM>0, exon + intron reads)	10,880	11,945	9,890	7,369
	Genes Detected (FPKM>0, exon reads)	9,601	10,776	8,613	6,115

Table 11. QC Statistics for data from broad classes in human LGN

		Human LGN			
		GABAergic	Glutamatergic	unassigned	Non-Neuronal
	Sample Count	231	1,027	23	295
Mean Values	Total Reads	2,534,596	2,462,098	2,018,440	2,341,087
	% Total Reads Aligned	88.65	90.49	79.79	84.39
	% Reads Aligned to Exons	34.31	30.62	19.98	30.58
	% Reads Aligned to Introns	36.31	40.96	30.62	33.37
	Amp Yield (ng)	11.35	20.95	10.70	4.48
	Genes Detected (CPM>0, exon + intron reads)	9,804	11,676	9,546	6,513
	Genes Detected (FPKM>0, exon reads)	5,942	7,916	3,869	2,754

Table 12. QC Statistics for data from broad classes in macaque LGN.

		Macaca nemestrina LGN		
		GABAergic	Glutamatergic	unassigned
	Sample Count	110	969	13
Mean Values	Total Reads	1,160,700	1,158,442	1,212,880
	% Total Reads Aligned	86.42	85.82	79.81
	% Reads Aligned to Exons	12.11	11.70	22.01
	% Reads Aligned to Introns	47.43	45.03	30.43
	Amp Yield (ng)	11.12	15.05	2.42
	Genes Detected (CPM>0, exon + intron reads)	9,018	9,647	6,057
	Genes Detected (FPKM>0, exon reads)	4,705	5,385	2,379

Note: All Macaque LGN samples processed at index192 with target read depth of 1.2 million reads per sample.

Table 13. QC Statistics for control data (mouse and human).

		Mouse Controls				Human Controls		
		Control Total RNA	ERCC	Mouse Whole RNA	NTC	ERCC	NTC	UHR
	Sample Count	573	750	927	744	602	602	638
Mean Values	Total Reads	2,232,041	559,720	2,343,915	435,688	1,298,254	817,241	2,398,408
	% Total Reads Aligned	89.74	39.98	85.77	17.60	50.77	22.38	89.89
	% Reads Aligned to Exons	60.45	1.14	34.67	0.96	0.37	0.70	63.90
	% Reads Aligned to Introns	8.22	0.42	17.26	0.53	2.82	4.08	6.81
	Amp Yield (ng)	15.31	2.42	8.39	2.44	2.05	1.91	6.79
	Genes Detected (CPM>0, exon + intron reads)	8,788	895	9,366	841	1,754	1,749	7,338
	Genes Detected (FPKM>0, exon reads)	7,128	469	6,486	441	350	379	5,364

Abbreviation: NTC, no template control.

ANALYSIS: CELL TYPE CLASSIFICATION

Clustering Mouse Cortical Single Cell RNA-Seq Data

Cells that passed QC criteria were evaluated for similarity using the iterative clustering R package **hicat**, available via Github (https://github.com/AllenInstitute/). In brief, cells were grouped into very broad categories using known markers, then clustered using high variance gene selection, dimensionality reduction, dimension selection, Jaccard-Louvain or hierarchical (Ward) clustering, and cluster merging based on overall significance of differentially expressed genes. This process was repeated within each resulting cluster until no more child clusters met differential gene expression or cluster size termination criteria. The entire clustering procedure was repeated 100 times using 80% of all cells sampled at random, and the frequency with which cells co-cluster was used to generate a final set of clusters, again subject to differential gene expression and cluster size termination criteria. For dimension reduction, hicat provides two modes: PCA for principal component analysis, and WGCNA, which identified co-expressed gene modules, and uses the module eigengene as reduced dimensions. Each of these two dimension reduction approaches has its own strengths: WGCNA mode is good for detecting rare clusters, and provides cleaner cluster boundaries, while PCA mode is more scalable to large datasets, captures combinatorial marker expression pattern, and more sensitive to low depth datasets. To combine the strengths of both methods, we run the whole clustering pipeline using both modes on subsampled cells 100 times, and combine the two resulting cell-cell co-clustering matrices (the new matrix contain the element-by-elment min values of the two matrices) to derive the final consensus clusters. The key strength of this approach is to provide fine resolution cell type categorization that withstands rigious statistical test to ensure reproducibility and biological relevance of the results.

Cell groups were arranged by transcriptomic similarity based on hierarchical clustering. First, the average expression level of the top differentially expressed genes between clusters was calculated for each cluster. For cells originating from human MTG, differentially expressed genes were defined as the 1200 genes with the highest beta score (see Bakken, *et al*. 2018) across all clusters. For cells from mouse VISp/ALM, the most significant differentially expressed genes in both directions for every pair of clusters was computed using the R package **limma**, and up to 50 from each comparison were pooled for a total of 4020 genes. Next, median

expression of marker genes per cluster was calculated as the cluster centroid, then a correlation-based distance matrix (Dxy = (1-correlation(x,y))/2) was determined, followed by a complete-linkage hierarchical clustering using the **hclust** R function with default parameters to infer a cell type taxonomy tree. The resulting dendrogram branches were reordered to visualize inhibitory, excitatory, and non-neuronal groups, while retaining the tree structure. For mouse VISp/ALM, the confidence for each branch of the tree wass estimated by a bootstrap resampling approach using R package **pvclust** (v2.0). Cluster nomenclature is a combination of cell type subclass definition and exemplar marker(s). Cell type subclasses were defined based on the major splits of the cell type taxnomony tree. Excitatory cell types also contain region information in the cluster names. The marker genes included in cluster names were selected to be unique either individually or as a combination within the cell type cohort. Differentially expressed genes are considered at different levels of taxonomy: *globally specific*, *within-subclass specific*, and *specific* compared to the *nearest sibling cluster*. Marker genes were also evaluated based on the 'completeness' of expression within the cluster that would be named by that gene. From this list of markers, specificity was assigned by examining gene expression at the single-cell level in given clusters, with preference to globally unique genes (e.g., *Chodl*, included in the *Sst-Chodl* cluster name), and putative markers that are expressed in all or a large proportion of cells within the cluster.

Clustering of Human MTG Single Nucleus RNA-Seq Data

Nuclei and cells were grouped into transcriptomic cell types using an iterative clustering procedure based on community detection in a nearest neighbor graph as described in Bakken, *et al.* 2018. Briefly, intronic and exonic read counts were summed, and $\log_2$(CPM+1)-transformed expression was centered and scaled across samples. Differentially expressed genes were selected while accounting for gene dropouts, and principal components analysis (PCA) was used to reduce dimensionality. Nearest-neighbor distances between samples were calculated using up to 20 principal components, Jaccard similarity coefficients were computed, and Louvain community detection was used to cluster this graph with 15 nearest neighbors. Clusters were merged if they lacked marker genes, and clustering was then applied iteratively to each sub-cluster. Cluster dendrograms reflect this analysis process and nomenclature for mouse cortical cell RNA-Seq data display.

Cluster names were defined using an automated strategy which combined molecular information (marker genes) and anatomical information (layer of dissection). Clusters were assigned a broad class of interneuron, excitatory neuron, microglia, astrocyte, oligodendrocyte precursor, oligodendrocyte, or endothelial cell based on maximal median cluster CPM of GAD1, SLC17A7, TYROBP, AQP4, PDGFRA, OPALIN, or NOSTRIN, respectively. Enriched layers were defined as the range of layers which contained at least 10% of the total cells from that cluster. Clusters were then assigned a broad marker, defined by maximal median CPM of PAX6, LAMP5, VIP, SST, PVALB, LINC00507, RORB, THEMIS, FEZF2, TYROBP, FGFR3, PDGFRA, OPALIN, or NOSTRIN. Finally, clusters in all broad classes with more than one cluster (e.g., interneuron, excitatory neuron, and astrocyte) were assigned the genes showing greatest specificity of expression in that cluster.

Clustering of LGd/LGN RNA-Seq Data – Mouse, Macaque, and Human

Comparable clustering strategy was used for human MTG, mouse LGd, and human, and macaque LGN. For macaque and human, cluster names were defined by the region sampled and the broad cell class (Inh – GABAergic neurons; Exc – glutamatergic neurons; Astro – astrocytes; OPC – oligodendrocyte precursor cells; Oligo – oligodendrocytes; Micro – microglia). Putative koniocellular glutamatergic neurons were labeled by the expressed marker gene *PRKCG* (Protein Kinase C Gamma). Finally, all clusters were labeled with the gene identified as having the most highly selective expression in that cluster. Cluster dendrograms reflect this analysis process and nomenclature for human cortical cell RNA-Seq data display.

Clusters were first labeled based on these brain regions (LGd – dorsal lateral geniculate nucleus; vLGd – ventral LGd; RT – reticular nucleus; LP – lateral posterior nucleus; and HIP – hippocampus). Next, clusters were assigned to the same broad classes defined for macaque and human clusters. Finally, all clusters were labeled with one or two genes with the most highly selective expression in that cluster.

REFERENCES

Aronesty E (2011) ea-utils: Command-line tools for processing biological sequencing data. Expression Analysis http://code.google.com/p/ea-utils.

Baker SC, Bauer SR, Beyer RP, Brenton JD, Bromley B, Burrill J, Causton H, Conley MP, Elespuru R, Fero M, Foy C, Fuscoe J, Gao X, Gerhold DL, Gilles P, Goodsaid F, Guo X, Hackett J, Hockett RD, Ikonomi P, Irizarry RA, Kawasaki ES, Kaysser-Kranich T, Kerr K, Kiser G, Koch WH, Lee KY, Liu C, Liu ZL, Lucas A, Manohar CF, Miyada G, Modrusan Z, Parkes H, Puri RK, Reid L, Ryder TB, Salit M, Samaha RR, Scherf U, Sendera TJ, Setterquist RA, Shi L, Shippy R, Soriano JV, Wagar EA, Warrington JA, Williams M, Wilmer F, Wilson M, Wolber PK, Wu X, Zadro R (2005) The External RNA Controls Consortium: a progress report. *Nature Methods* 2**:**731-734.

Bakken TE, Hodge RD, Miller JA, Yao Z, Nguyen TN, Aevermann B, Barkan E, Bertagnolli D, Casper T, Dee N, Garren E, Goldy J, Graybuck LT, Kroll M, Lasken RS, Lathia K, Parry S, Rimorin C, Scheuermann RH, Schork NJ, Shehata SI, Tieu M, Phillips JW, Bernard A, Smith KA, Zeng H, Lein ES, Tasic B (2018) Equivalent high-resolution identification of neuronal cell types with single-nucleus and single-cell RNA-sequencing. *bioRxiv, in review*.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15-21.

Gabbott PLA, Somogyi J, Stewart MG, Hámori J (1986) A quantitative investigation of the neuronal composition of the rat dorsal lateral geniculate nucleus using GABA-immunocytochemistry. *Neuroscience* 19:101–111.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10:R25.

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202-1214.

Piscopo DM, El-Danaf RN, Huberman AD, Niell CM (2013) Diverse visual features encoded in mouse lateral geniculate nucleus. *Journal of Neuroscience* 33:4642–4656.

Risso D, Ngai J, Speed TP, Dudoit S (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology* 32**:**896-902.

Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, Bertagnolli D, Goldy J, Shapovalova N, Parry S, Lee C, Smith K, Bernard A, Madisen L, Sunkin SM, Hawrylycz M, Koch C, Zeng H (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience* 19:335-346.

Tasic B, Yao Z, Smith KA, Graybuck L, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, Penn O, Bakken T, Menon V, Miller JA, Fong O, Hirokawa KE, Lathia K, Rimorin C, Tieu M, Larsen R, Casper T, Barkan E, Kroll M, Parry S, Shapovalova NV, Hirchstein D, Pendergraft J, Kim TK, Szafer A, Dee N, Groblewski P, Wickersham I, Cetin A, Harris JA, Levi BP, Sunkin SM, Madisen L, Daigle TL, Looger L, Bernard A, Phillips J, Lein E, Hawrylycz M, Svoboda K, Jones AR, Koch C, Zeng H (2017) Shared and distinct transcriptomic cell types across neocortical areas. *bioRxiv, in review.*

Zimin AV, Cornish AS, Maudhoo MD, Gibbs RM, Zhang X, Pandey S, Meehan DT, Wipfler K, Bosinger SE, Johnson ZP, Tharp GK, Marcais G, Roberts M, Ferguson B, Fox H, Treangen T, Salzberg SL, Yorke JA, Norgren RB (2014) A new rhesus macaque assembly and annotation for next-generation sequencing analyses. *Biology Direct*, 9:20.