nature

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41586-018-0654-5

In the format provided by the authors and unedited.

Shared and distinct transcriptomic cell types across neocortical areas

Bosiljka Tasic[1]*, Zizhen Yao[1,4], Lucas T. Graybuck[1,4], Kimberly A. Smith[1,4], Thuc Nghi Nguyen[1], Darren Bertagnolli[1], Jeff Goldy[1], Emma Garren[1], Michael N. Economo[2], Sarada Viswanathan[2], Osnat Penn[1], Trygve Bakken[1], Vilas Menon[1,2], Jeremy Miller[1], Olivia Fong[1], Karla E. Hirokawa[1], Kanan Lathia[1], Christine Rimorin[1], Michael Tieu[1], Rachael Larsen[1], Tamara Casper[1], Eliza Barkan[1], Matthew Kroll[1], Sheana Parry[1], Nadiya V. Shapovalova[1], Daniel Hirschstein[1], Julie Pendergraft[1], Heather A. Sullivan[3], Tae Kyung Kim[1], Aaron Szafer[1], Nick Dee[1], Peter Groblewski[1], Ian Wickersham[3], Ali Cetin[1], Julie A. Harris[1], Boaz P. Levi[1], Susan M. Sunkin[1], Linda Madisen[1], Tanya L. Daigle[1], Loren Looger[2], Amy Bernard[1], John Phillips[1], Ed Lein[1], Michael Hawrylycz[1], Karel Svoboda[2], Allan R. Jones[1], Christof Koch[1] & Hongkui Zeng[1]

[1]Allen Institute for Brain Science, Seattle, WA, USA. [2]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. [3]Massachusetts Institute of Technology, Cambridge, MA, USA. [4]These authors contributed equally: Zizhen Yao, Lucas T. Graybuck, Kimberly A. Smith. *e-mail: bosiljkat@alleninstitute.org

Supplementary Information Guide

Shared and distinct transcriptomic cell types across neocortical areas

Bosiljka Tasic[#1], Zizhen Yao*[1], Lucas T. Graybuck*[1], Kimberly A. Smith*[1], Thuc Nghi Nguyen[1], Darren Bertagnolli[1], Jeff Goldy[1], Emma Garren[1], Michael N. Economo[2], Sarada Viswanathan[2], Osnat Penn[1], Trygve Bakken[1], Vilas Menon[1,2], Jeremy Miller[1], Olivia Fong[1], Karla E. Hirokawa[1], Kanan Lathia[1], Christine Rimorin[1], Michael Tieu[1], Rachael Larsen[1], Tamara Casper[1], Eliza Barkan[1], Matthew Kroll[1], Sheana Parry[1], Nadiya V. Shapovalova[1], Daniel Hirschstein[1], Julie Pendergraft[1], Heather A. Sullivan[3], Tae Kyung Kim[1], Aaron Szafer[1], Nick Dee[1], Peter Groblewski[1], Ian Wickersham[3], Ali Cetin[1], Julie A. Harris[1], Boaz P. Levi[1], Susan M Sunkin[1], Linda Madisen[1], Tanya L. Daigle[1], Loren Looger[2], Amy Bernard[1], John Phillips[1], Ed Lein[1], Michael Hawrylycz[1], Karel Svoboda[2], Allan R. Jones[1], Christof Koch[1], Hongkui Zeng[1].

[1]Allen Institute for Brain Science, Seattle, WA, USA

[2]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

[3]Massachusetts Institute of Technology, Cambridge, MA, USA

*These authors contributed equally to this work.

[#]Correspondence should be addressed to B.T. (bosiljkat@alleninstitute.org)

SUPPLEMENTARY TABLE LEGENDS

Supplementary Table 1. Cell counts and utilization per donor. ID, sex, age, genotype, and dissected region for each of the 352 donor animals in this study. Counts of cells assigned to GABAergic, glutamatergic, non-neuronal or low-quality clusters are also provided.

Supplementary Table 2. Summary of single cell numbers. Cell counts based on sample collection and analysis categories: QC metrics, source/labeling type, major class, and Retro-seq.

Supplementary Table 3. Area-specific genes identified by differential gene expression analysis of matched cell types. Differentially-expressed genes as identified by pairwise comparisons between the best-match types (for region-specific clusters, mostly glutamatergic), or between ALM and VISp portions of the same cluster (for shared clusters, mostly GABAergic). For shared glutamatergic types (CR–Lhx5, L6–CT–Nxph2–Sla, L6b–Col8a1–Rprm, L6b–P2ry12, L6b–Hsd17b2), we detect 10-34 DE genes between regions. It is possible that these clusters were not segregated into region-specific subclusters because our dataset contained too few cells from one region to provide sufficient statistical power for a split.

Supplementary Table 4. Comparison of cell types defined in this study to previous studies. Correspondence between the cell types in this study and select previously published studies. For glutamatergic types, we focused on VISp. For several single-cell transcriptomic studies, we computed the correspondence of types by mapping cells from these studies onto our types using a centroid classifier (**Methods**). For Zeisel et al., 2015[19], we excluded hippocampal excitatory, choroid and ependymal types. For other single cell or bulk RNA transcriptomic studies and non-transcriptomic studies, correspondence was assigned based on marker gene expression or transgenic driver use. In many cases, assignments to classes/subclasses were performed based on a single marker, and could be too rigid. The references (ordered by mapping strategy and then year) included in this table are: Hrvatin et al., 2018[40]; Paul et al., 2017[36]; Tasic et al., 2016[20]; Cadwell et al., 2016[38]; Zeisel et al., 2015[19]; Marques et al., 2016[78]; Matcovitch-Natan et al., 2016[88]; Hilscher et al., 2017[37]; Frazer 2016[22]; He et al., 2016[35]; Dehorter et al., 2015[89]; Sorensen et al., 2015[26]; Bortone et al., 2014[90]; Velez-Fort et al., 2014[91]; Zhang et al., 2014[79]; Pfeffer et al., 2013[92]; Taniguchi et al., 2013[87]; Zeng et al., 2012[93]; Rudy et al., 2011[16]; Cahoy et al., 2008[77]; Gerashchenko et al., 2008[94]; Gonchar et al., 2008[95]; Molyneaux et al., 2007[15]; von Engelhardt et al., 2007[96]; Xu et al., 2006[97]; and Oliva et al., 2000[98]. NA, no correspondence available/applicable; NP, correspondence analysis not performed.

Supplementary Table 5. Transgenic mouse lines and sources. Information on 47 driver lines and 6 reporter lines used in this study, with originating labs, references, generation method, RRID, and repository availability. All lines are available from the Jackson Laboratory or MMRRC, with the exception of *Rorb-P2A-Flpo* and *Ai110*, which are available directly from the Allen Institute. The references included in this table are: Daigle et al., 2018[99]; Rossi et al., 2011[100]; Gong et al., 2007[101]; Taniguchi et al., 2011[102]; Tasic et al., 2016[20]; Franco et al., 2012[103]; Gerfen et al., 2013[104]; Dhillon et al., 2006[105]; Madisen et al., 2015[106]; Hippenmeyer et al., 2005[107]; Harris et al., 2014[86]; Madisen et al., 2010[108]; Vong et al., 2011[109]; Grimes et al., 2011[110]; Tong et al., 2008[111]; and He et al., 2016[35].

Supplementary Table 6. Retrograde injection coordinates and injection site labeling details. Stereotaxic injection coordinates were used for retrograde injection into mouse brains, but injection attempts did not always result in labeling of the targeted structure. We provide injection coordinates, attempted targets (injection_target), and the results of careful inspection of injection targets from images taken during dissection. We identify targets with the highest level of viral labeling brightness (injection_verified_primary_target) and all structures that appeared to have viral labeling

(injection_verified_all_targets), as well as regions that have labeling along the injection tract (injection_tract_labeling). Only samples with 'OK' in the injection_qc_flags column were used to analyze retrograde connectivity.

Supplementary Table 7. Viruses and sources. Information on 6 viruses for retrograde tracing and the virus used for anterograde tracing, with their sources, packaging lines, genome references, serotypes, and availability from Addgene. The references included in this table are: Tervo et al., 2016[52]; Hnasko, et al. 2016[54]; Chatterjee, et al., 2018[53]; Kiyota, et al., 2009[112]; Mao et al., 2011[113]; Wertz et al., 2015[114]; and Oh et al., 2014[28].

Supplementary Table 8. Synthetic constructs used for alignments. Sequences of ERCC and other synthetic constructs used to quantify synthetic construct mapping (e.g., tdTomato CPM).

Supplementary Table 9. Summary of transcriptomic cell types, annotations, and markers. Metadata characteristics for each transcriptomic cell type identified by clustering, with associated class and subclass-level annotations, cell counts, ALM/VISp fraction, median gene detection, and marker genes.

Cell type annotations:

- cluster_id: Unique cluster number used to arrange clusters in figures.
- cluster_color: Color assigned to each cluster for use in figures.
- cluster_label: Cluster name based on subclass and marker gene expression.
- class_id: Unique number for each transcriptomic class.
- class_label: Broad class assigned to each cluster at the level of GABAergic/Glutamatergic cells.
- class_color: Color assignment for each class.
- subclass_id: Unique number for each transcriptomic subclass.
- subclass_label: Subclass assigned to each cluster at the level of Pvalb/L5 IT.
- sublcass_color: Color assignment for each subclass

Cluster metadata:

- n_cells: Number of cells assigned as core or intermediate to each cluster.
- fraction_ALM: Fraction of cells in each cluster that originated from dissections of ALM.
- fraction_VISp: Fraction of cells in each cluster that originated from dissections of VISp.
- median_genes: Median number of genes detected in each cluster.

Marker genes:

- markers_global_specific: Markers enriched in each cluster compared to all other clusters.
- markers_subclass_specific: Markers enriched in each cluster compared to all other clusters within the same subclass. For glutamatergic types, comparisons were restricted to clusters from the same subclass and cortical region.
- markers_vs_siblings: Markers enriched in each cluster compared to neighboring clusters that share a node in the dendrogram with the examined cluster (< 5 clusters).
- markers_node_specific: Markers enriched in a set of neighboring clusters that share a node in the dendrogram (< 5 clusters) compared to all other clusters within the same subclass in the dendrogram.

Supplementary Table 10. Metadata for all analyzed samples. 48 metadata characteristics for all 25,481 cells including 2 identifiers, 7 donor characteristics, 3 dissection annotations, 6 retrograde injection annotations, 3 FACS annotations, 4 molecular biology tracking annotations, 5 sequencing annotations, 11 alignment characteristics, 2 reporter expression levels, 3 cell type annotations, and 3 annotations related to cell type assignment confidence. Descriptions of each column are below:

Sample identifiers:

- sample_name: Unique sample identifier. Matches data released on brain-map.org.
- sample_id: Alternative unique identifier.

Donor characteristics:

- donor: Unique donor mouse identifier.
- sex: Sex of donor mouse. F, female; M, male.
- age_days: Age of the donor mouse in postnatal days.
- eye_condition: Annotation of eye conditions identified in each animal, followed by the side on which the eye was affected. L, left; R, right.
- genotype: Full genotype for each donor. See **Supplementary Table 5** for sources.
- driver_lines: Cre and/or Flp driver line(s) for each donor, when used. See **Supplementary Table 5** for sources.
- reporter_lines: Cre and/or Flp-dependent reporter(s) for each donor, when used. See **Supplementary Table 5** for sources.

Dissection annotations:

- brain_hemisphere: Hemisphere from which cells were collected. L, Left; R, Right.
- brain_region: Cortical region dissected. ALM, anterior lateral motor cortex; VISp, primary visual cortex.
- brain_subregion: Layer-enriching dissection from which cells were collected. Layers are abbreviated with L. Dissections that contained multiple layers are labeled with the upper-most layer followed by the lower-most layer. L1-L4, for example, contains L1, L2/3, and L4. L1-L6 is a dissection containing all layers.

Injection annotations:

- injection_label_direction: Type of injection experiment performed (Retrograde for all injections in this study), or No Injection if not used.
- injection_primary_target: Center of injection site, as verified by inspection of coronal target site slices. Abbreviations match the Allen Mouse Brain Atlas. These targets were used for **Figure 3** and **Extended Data Figure 10**.
- injection_secondary: All labeled regions, as verified by inspection of coronal target site slices. Abbreviations match the Allen Mouse Brain Atlas.
- injection_tract: Regions that show labeled cells along the need insertion injection tract. Abbreviations match the Allen Mouse Brain Atlas.
- injection_material: Virus injected. See **Supplementary Table 7** for further virus information.
- injection_exclusion_criterion: Quality control flag, used to exclude injections that are too far off-target, or have too much tract labeling. Cells used for analysis of projection targets are labelled “OK”.

FACS annotations:

- facs_date: The date on which cells were collected by FACS. All dissections were performed on the same date as the FACS date for each sample.
- facs_container: An ID for each 8-well strip collected by FACS.
- facs_sort_criteria: Gating criteria used to select cells for sorting.

Molecular biology annotations:

- rna_amplification_set: An ID for each plate, for matching to amplification control samples provided on GEO and on the Allen Institute website.
- rna_amplification_pcr_cycles: The number of PCR cycles used for RNA amplification – either 18x or 21x.
- library_prep_set: An ID for each library preparation set following amplification.
- library_prep_avg_size_bp: The average length of prepared fragments from each sample, as measured by fragment analyzer.

Sequencing annotations:

- seq_name: An ID for each sample per sequencing run
- seq_tube: The tube ID for mixed, multiplexed samples
- seq_batch: Sequencing batch identifier.
- total_reads: Number of total reads sequenced for each sample.
- complexity_cg: CG dinucleotide odds ratio.

Alignment characteristics:

- percent_exon_reads: Percent of reads that aligned to mouse mRNA exons.
- percent_intron_reads: Percent of reads that aligned to mouse intronic genomic regions.
- percent_intergenic: Percent of reads that aligned to mouse intergenic genomic regions.
- percent_rrna_reads: Percent of reads that aligned to mouse ribosomal RNAs (rRNA).
- percent_mt_exon_reads: Percent of reads that aligned to mouse mitochondrial (MT) mRNA exons.
- percent_reads_unique: Percent of unique reads based on removal of duplicate alignments (PCR duplicates).
- percent_synth_reads: Percent of reads that aligned to synthetic constructs, including ERCCs and transgene sequences.
- percent_ecoli_reads: Percent of reads that aligned to the *E. coli* genome.
- percent_aligned_reads_total: Percent of reads that aligned to any part of the mouse genome.
- genes_detected_cpm_criterion: Number of genes detected using a criteria of > 0 CPM in either exons or introns.
- genes_detected_fpkm_criterion: Number of genes detected using a criteria of > 0 FPKM for exons only. This value is used for computation of genes detected in this study.

Reporter expression:
- tdt_cpm: tdTomato expression in Counts per Million sequenced fragments (CPM).
- gfp_cpm: GFP expression in Counts per Million sequenced fragments (CPM).

Cell type annotations:
- class: Broad cell type class (e.g. GABAergic and Glutamatergic).
- sublcass: Cell type subclass annotations (e.g. Pvalb, L5 IT).
- cluster: Cluster name used in figures.

Cell Type confidence measures:
- confusion_score: The ratio of the probabilities that each cell was co-clustered with the cells from its second-best cluster and with the cells from its assigned cluster (**Methods**).
- cluster_correlation: Pearson correlation of each cell with the centroid of gene expression for its assigned cluster.
- core_intermediate_call: Core or Intermediate cell type calls based on centroid classifier validiation (**Methods**).

Supplementary Table 11. Complete statistics for Extended Data Figure 15b.

ADDITIONAL REFERENCES

88. Matcovitch-Natan, O. et al. Microglia development follows a stepwise program to regulate brain homeostasis. *Science* **353**, aad8670 (2016).

89. Dehorter, N. et al. Tuning of fast-spiking interneuron properties by an activity-dependent transcriptional switch. *Science* **349**, 1216–1220 (2015).

90. Bortone, D. S., Olsen, S. R. & Scanziani, M. Translaminar inhibitory cells recruited by layer 6 corticothalamic neurons suppress visual cortex. *Neuron* **82**, 474–485 (2014).

91. Vélez-Fort, M. et al. The stimulus selectivity and connectivity of layer six principal cells reveals cortical microcircuits underlying visual processing. *Neuron* **83**, 1431–1443 (2014).

92. Pfeffer, C. K., Xue, M., He, M., Huang, Z. J. & Scanziani, M. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat. Neurosci.* **16**, 1068–1076 (2013).

93. Zeng, H. & Madisen, L. Mouse transgenic approaches in optogenetics. *Prog. Brain Res.* **196**, 193–213 (2012).

94. Gerashchenko, D. et al. Identification of a population of sleep-active cerebral cortex neurons. *Proc. Natl. Acad. Sci.* **105**, 10227–10232 (2008).

95. Gonchar, Y., Wang, Q. & Burkhalter, A. Multiple distinct subtypes of GABAergic neurons in mouse visual cortex identified by triple immunostaining. *Front. Neuroanat.* **1**, 3 (2008).

96. von Engelhardt, J., Eliava, M., Meyer, A. H., Rozov, A. & Monyer, H. Functional characterization of intrinsic cholinergic interneurons in the cortex. *J. Neurosci.* **27**, 5633–5642 (2007).

97. Xu, X., Roby, K. D. & Callaway, E. M. Mouse cortical inhibitory neuron type that coexpresses somatostatin and calretinin. *J. Comp. Neurol.* **499**, 144–160 (2006).

98. Oliva, A. A., Jr, Jiang, M., Lam, T., Smith, K. L. & Swann, J. W. Novel hippocampal interneuronal subtypes identified using transgenic mice that express green fluorescent protein in GABAergic interneurons. *J. Neurosci.* **20**, 3354–3368 (2000).

99. Daigle, T. L. et al. A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. *Cell* **174**, 465–480.e22 (2018).

100. Rossi, J. et al. Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. *Cell Metab.* **13**, 195–204 (2011).

101. Gong, S. et al. Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27**, 9817–9823 (2007).

102. Taniguchi, H. et al. A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**, 995–1013 (2011).

103. Franco, S. J. et al. Fate-restricted neural progenitors in the mammalian cerebral cortex. *Science* **337**, 746–749 (2012).

104. Gerfen, C. R., Paletzki, R. & Heintz, N. GENSAT BAC cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron* **80**, 1368–1383 (2013).

105. Dhillon, H. et al. Leptin directly activates SF1 neurons in the VMH, and this action by leptin is required for normal body-weight homeostasis. *Neuron* **49**, 191–203 (2006).

106. Madisen, L. et al. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**, 942–958 (2015).

107. Hippenmeyer, S. et al. A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3**, e159 (2005).

108. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).

109. Vong, L. et al. Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron* **71**, 142–154 (2011).

110. Grimes, W. N., Seal, R. P., Oesch, N., Edwards, R. H. & Diamond, J. S. Genetic targeting and physiological features of VGLUT3+ amacrine cells. *Vis. Neurosci.* **28**, 381–392 (2011).

111. Tong, Q., Ye, C. P., Jones, J. E., Elmquist, J. K. & Lowell, B. B. Synaptic release of GABA by AgRP neurons is required for normal regulation of energy balance. *Nat. Neurosci.* **11**, 998–1000 (2008).

112. Kiyota, T. et al. AAV1/2-mediated CNS gene delivery of dominant-negative CCL2 mutant suppresses gliosis, β-amyloidosis, and learning impairment of APP/PS1 mice. *Mol. Ther.* **17**, 803–809 (2009).

113. Mao, T. et al. Long-range neuronal circuits underlying the interaction between sensory and motor cortex. *Neuron* **72**, 111–123 (2011).

114. Wertz, A. et al. Single-cell-initiated monosynaptic tracing reveals layer-specific cortical network modules. *Science* **349**, 70–74 (2015).