32. J. Bollrath, F. M. Powrie, *Semin. Immunol.* **25**, 352–357 (2013).
33. M. S. Vacchio *et al.*, *Nat. Immunol.* **15**, 947–956 (2014).
34. B. Jabri, V. Abadie, *Nat. Rev. Immunol.* **15**, 771–783 (2015).

NEUROGENOMICS

Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain

Blue B. Lake,[1]* Rizi Ai,[2]* Gwendolyn E. Kaeser,[3,4]* Neeraj S. Salathia,[5]* Yun C. Yung,[3] Rui Liu,[1] Andre Wildberg,[2] Derek Gao,[1] Ho-Lim Fung,[1] Song Chen,[1] Raakhee Vijayaraghavan,[5] Julian Wong,[3] Allison Chen,[3] Xiaoyan Sheng,[3] Fiona Kaper,[5] Richard Shen,[5] Mostafa Ronaghi,[5] Jian-Bing Fan,[5]† Wei Wang,[2]† Jerold Chun,[3]† Kun Zhang[1]†

The human brain has enormously complex cellular diversity and connectivities fundamental to our neural functions, yet difficulties in interrogating individual neurons has impeded understanding of the underlying transcriptional landscape. We developed a scalable approach to sequence and quantify RNA molecules in isolated neuronal nuclei from a postmortem brain, generating 3227 sets of single-neuron data from six distinct regions of the cerebral cortex. Using an iterative clustering and classification approach, we identified 16 neuronal subtypes that were further annotated on the basis of known markers and cortical cytoarchitecture. These data demonstrate a robust and scalable method for identifying and categorizing single nuclear transcriptomes, revealing shared genes sufficient to distinguish previously unknown and orthologous neuronal subtypes as well as regional identity and transcriptomic heterogeneity within the human brain.

A lthough substantial progress has been achieved in mice (*1*–*3*), comprehensive classification of adult human brain neurons on the basis of their single-cell transcriptomes has yet to be realized. Examination of individual neuronal gene expression profiles for functional patterns could provide unbiased insights into subtypes from defined neuroanatomical regions, which are missed by gross anatomical studies that report limited transcriptomic differences across the neocortex (*4*–*7*). Previous analyses of single adult human neurons have been dependent on methods compatible with freshly isolated neurosurgical tissues (*8*), which can be difficult to obtain, with limited regional sampling and depth. In contrast, postmortem tissues provide a vastly more accessible source of both normal and diseased brain, in which challenges to interrogating single neuronal genomes can be overcome by using single nuclei (*9*, *10*) combined with RNA sequencing. Here, we report the development of a scalable pipeline from postmortem brain through nuclear transcriptome analyses that identifies both known and previously unknown neuronal subtypes across the cerebral cortex in humans.

With the goal of defining transcriptomic profiles of single neurons, neuronal nuclear antigen (NeuN) was used (*9*) to isolate neuronal nuclei (fig. S1) from the postmortem brain of a normal, 51-year-old female (Fig. 1A). We focused on six classically defined Brodmann Areas (BAs) with well-documented anatomical and electrophysiological properties that were derived from a single cortical hemisphere because interhemispheric and interindividual transcriptome differences were reported to be minimal (*4*–*7*). Isolation of nuclei was used to reduce transcriptomic contamination from other cells or degradation encountered with whole-neuron dis-

sociation or laser caption microdissection (fig. S2). Furthermore, sequencing of RNA from single nuclei on a limited scale has found gene expression values comparable with that of the whole cell (*11*, *12*). Therefore, we developed and implemented a highly scalable, single-nucleus RNA sequencing (SNS) pipeline (*13*) (Fig. 1A and figs. S1 and S3 to S8) that has broad applicability for postmortem brains derived from multiple brain banks or repositories (fig. S4F).

Using this pipeline, we processed 86 Fluidigm C1 chips and sequenced 4488 single nuclei to an average depth of 8.34 million reads (table S1 and fig. S5). Genomic mapping rates revealed a high proportion of reads that corresponded to intronic sequences (Fig. 1A and fig. S5A). The low percentage of intergenic reads argues against possible genomic contamination. Instead, the intronic reads likely captured an abundance of nascent RNA transcripts present in the nuclei. Intronic reads can be used to predict de novo expression (*14*), as well as whole-cell gene transcription levels (*15*). Additionally, our single-nuclei expression data inclusive of intronic reads accurately predicted cellular identity (fig. S7), providing initial validation for our SNS pipeline.

After quality filtering, including removal of doublets misclassified as single nuclei (Fig. 1A and fig. S6) (*13*), we achieved 3227 data sets across the six cortical regions (Fig. 1A and table S2). To identify neuronal subtypes, we developed a clustering and classification strategy that was capable of resolving 17 clusters (fig. S8A) (*13*) on the basis of differential expression of neuronally annotated marker genes (tables S3 and S4 and fig. S8B). These clusters showed distinct subgroup aggregation (Fig. 1B and fig. S9A) and specific gene expression profiles associated with neuronal ontologies (Fig. 1C, fig. S9B, and tables S5 and S6). With the exception of a single cluster (NoN, *n* = 44 data sets) deriving from one C1 chip having reduced mapping rates, 16 of these clusters were generated independent of detectable batch effects (table S2 and fig. S10). Differential expression of inhibitory markers associated with GABAergic interneurons (table S3) distinguished potential inhibitory (In) from excitatory (Ex) neuronal subtypes (Fig. 1B), which is consistent with mutually exclusive positivity of associated marker genes using a fraction of positive (FOP) thresholding method (Fig. 2A) (*2*). As such, our data set first differentiated two major classifications within the cerebral cortex: 972 inhibitory

[1]Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA. [2]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, USA. [3]Department of Molecular and Cellular Neuroscience, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA, USA. [4]Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA, USA. [5]Illumina, San Diego, CA, USA.
*These authors contributed equally to this work. †Corresponding author. Email: kzhang@bioeng.ucsd.edu (K.Z.); jchun@scripps.edu (J.C.); wei-wang@ucsd.edu (W.W.); jianbing_fan@yahoo.com (J.-B.F.)
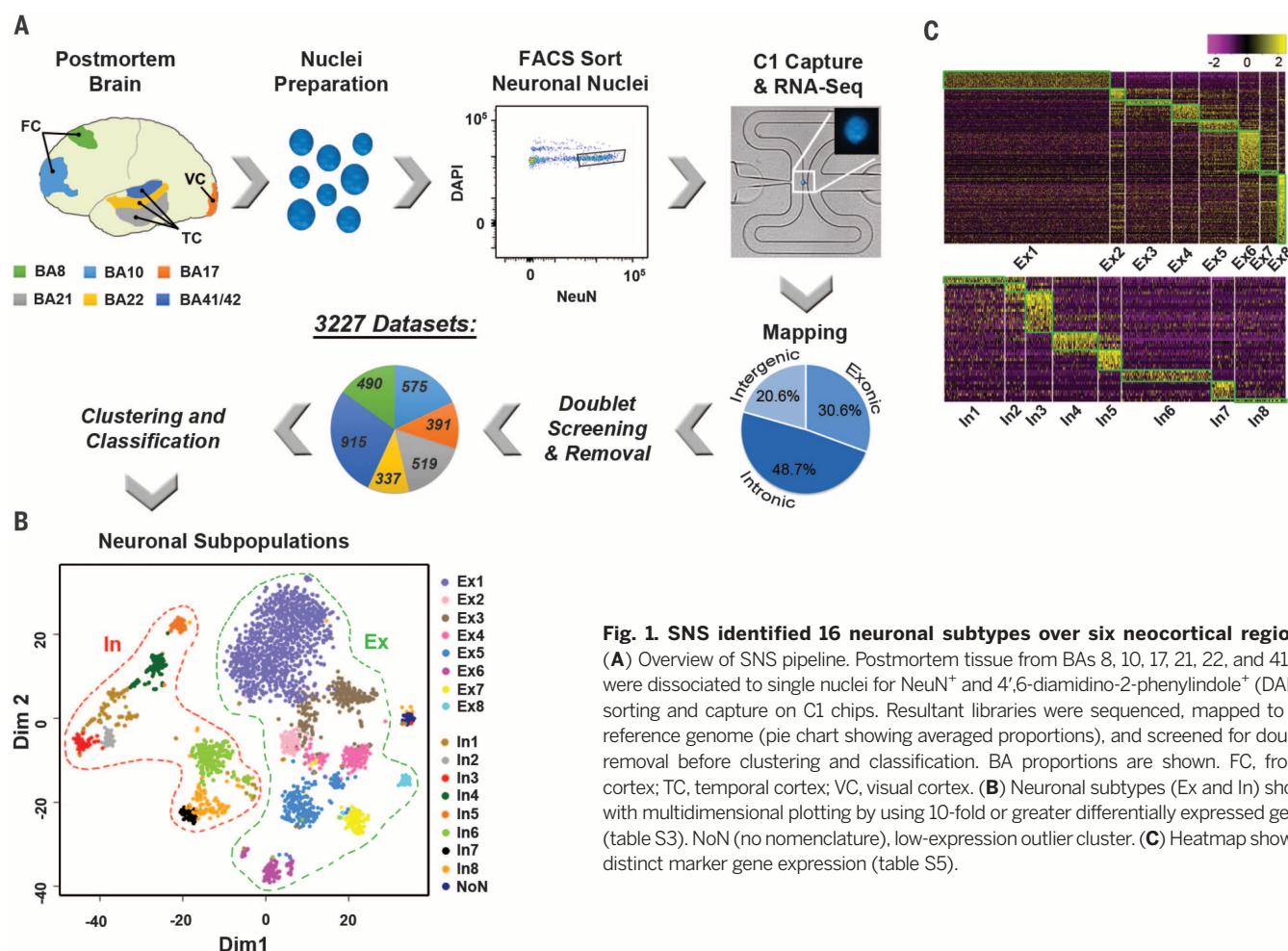
Fig. 1. SNS identified 16 neuronal subtypes over six neocortical regions. (**A**) Overview of SNS pipeline. Postmortem tissue from BAs 8, 10, 17, 21, 22, and 41/42 were dissociated to single nuclei for NeuN$^+$ and 4′,6-diamidino-2-phenylindole$^+$ (DAPI$^+$) sorting and capture on C1 chips. Resultant libraries were sequenced, mapped to the reference genome (pie chart showing averaged proportions), and screened for doublet removal before clustering and classification. BA proportions are shown. FC, frontal cortex; TC, temporal cortex; VC, visual cortex. (**B**) Neuronal subtypes (Ex and In) shown with multidimensional plotting by using 10-fold or greater differentially expressed genes (table S3). NoN (no nomenclature), low-expression outlier cluster. (**C**) Heatmap showing distinct marker gene expression (table S5).

neurons that generally encompass interneurons and 2253 excitatory neurons that generally encompass pyramidal or projection neurons (*16*). Furthermore, each subgroup within these classifications showed distinct contributions from each brain region (Fig. 2A and table S7), likely reflecting varied proportions of these neuronal subtypes across BAs, with most variability present in the visual cortex (BA17), which is known to have distinct cytoarchitecture and gene expression profiles (*7*, *17*).

In order to further annotate inhibitory neuron subtypes, we examined expression of known marker genes associated with cortical layers, developmental origin, and interneuron classification (Fig. 2B) (*13*). On the basis of in situ human brain expression data (fig. S11) (*17*), our inhibitory neuron subtypes were found to distribute spatially from the pial surface (most superficial boundary) to white matter (deepest boundary) of the neocortex and could be grouped by the developmental origin of interneurons from subcortical regions of the medial, lateral, or caudal ganglionic eminences (MGE, LGE, or CGE) (Fig. 2B) (*18*, *19*). Furthermore, distinct profiles of interneuron classification markers revealed subtypes that parallel those identified from the mouse somatosensory cortex (Fig. 2, B and C, and fig. S12A) (*3*). Cortical regional heterogeneity

within subtypes was also observed, as evident by a layer 3 population (In4) that showed a specific absence of *RELN*/*SST* expression in BA17 (Fig. 2C and fig. S11, B and D). As such, our data distinguished inhibitory neuron subtypes having heterogeneous distributions within the neocortex.

Most excitatory cortical projection or pyramidal neurons can be categorized by their layer position established during neocortical development (*17*) combined with their axonal projections (Fig. 3A) (*16*). Our excitatory neuron subgroups, which were also in high concordance with subtypes identified in mice (fig. S12B) (*3*), expressed known markers associated with a superficial-to-deep cortical distribution (Fig. 3, B to D, and fig. S13) (*13*), with more than one subtype occupying most layers. Our data set was able to resolve cortical region specificity, as seen for the *BHLHE22*-positive (Fig. 3C and fig. S13, A and D) layer 4 subtypes Ex2 and Ex3 (Fig. 4A), where Ex2 derived predominantly from rostral regions, BA8 and BA10, and Ex3 from caudal regions, BA17 and BA41/42 (Figs. 2A and 4B). Consistently, these subgroups showed distinct gene expression (Fig. 4C and table S8) associated with neuronal electrophysiology and connectivity (table S9). Furthermore, we were able to resolve intrasubtype heterogeneity, in terms of BA-

specific expression patterns, which was observed in all subtypes (Fig. 4B), such as within the Ex3 subtype between BA17 and BA41/42 regions (Fig. 4, B and D, and table S10). As such, regional neurophysiological differences in cortical regions may be attributed to not only variations in the proportions of interneuron and projection neuron subtypes, but also to cell-intrinsic transcriptomic differences among single neurons within a subtype. Consistent with this possibility, we found that genes having known variability between the visual and temporal cortices from in situ hybridization (ISH) studies (*17*) also had transcriptomic differences that could be attributed to subtypes defined by our data set (fig. S14A and table S11) (*13*). Therefore, our data highlight subtle yet region-defining gene expression signatures among specific neuronal subtypes that could not be detected from bulk analyses (fig. S14B).

To further understand the extent of heterogeneity that may exist within subtypes, we identified genes varying globally (table S12 and fig. S15A) or expressed differentially within each BA (table S13 and fig. S15B) for each subgroup. Although a subset of In and Ex subgroup-variable genes was associated with differential expression between brain regions, a large proportion were distinct (fig. S15C).
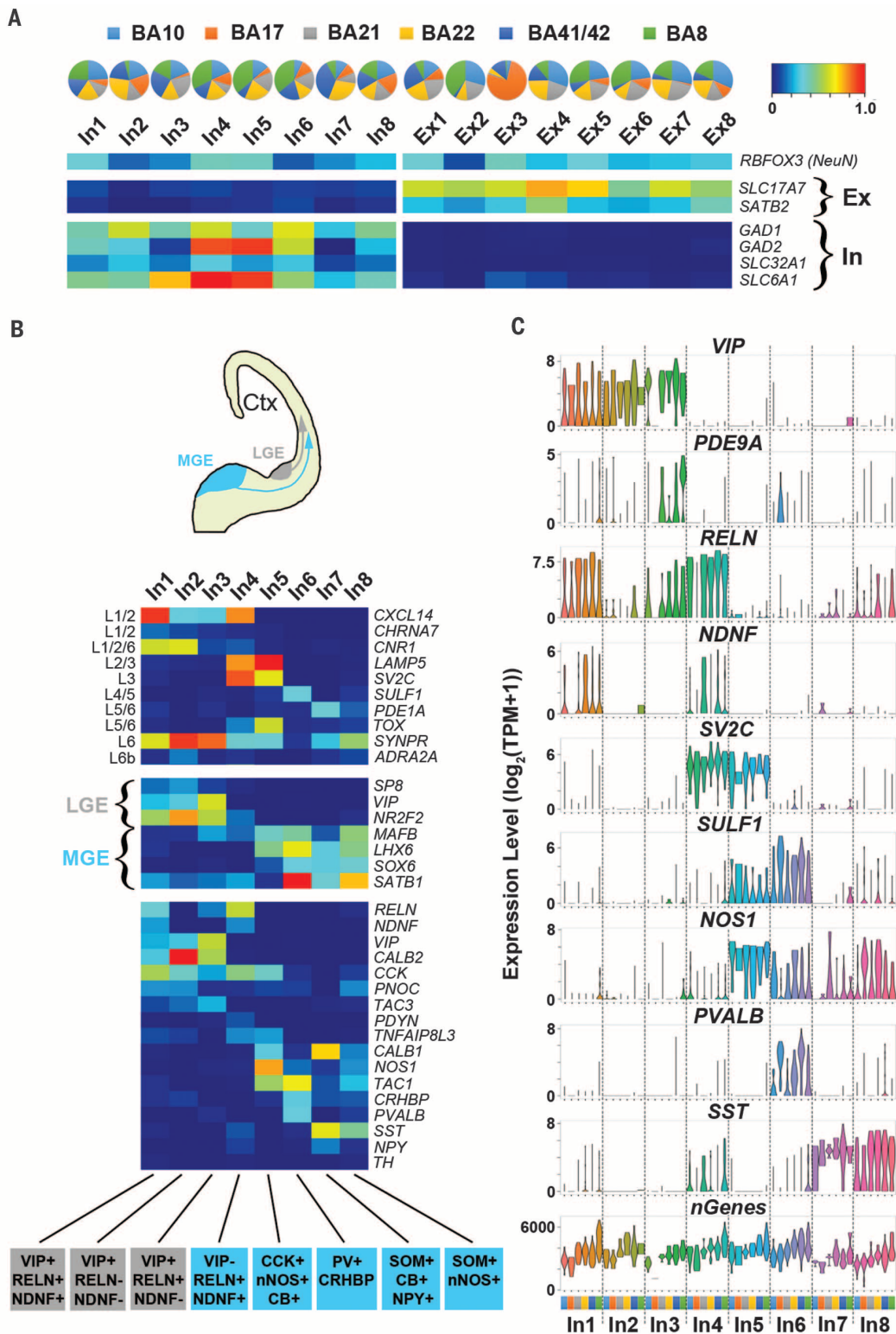
Fig. 2. SNS reveals distinct interneuron subtypes. (**A**) Pie charts display relative proportions of subtypes among BAs and FOP heatmaps for In and Ex marker genes. (**B**) Diagram of subpallial origins of interneurons from either the LGE or MGE with FOP heatmaps [scale as in (A)] for marker genes associated with cortical layer (L) (top), subpallial origin (middle), and interneuron classification (bottom). Potential interneuron subtypes are indicated below. SOM, *somatostatin* or *SST*; NPY, *neuropeptide Y*; CB, *calbindin-D-28k* or *CALB1*; VIP, *vasoactive intestinal peptide*; RELN, *reelin*; nNOS, *neuronal nitric oxide synthase* or *NOS1*; PV, *parvalbumin* or *PVALB*; CCK, *cholecystokinin;* NDNF, *neuron-derived neurotrophic factor*; CRHBP, *corticotropin releasing hormone binding protein*. (**C**) Violin plots showing select marker gene expression values by BA [colors as in (A)] for each inhibitory neuron subtype. *nGenes*, total number of genes identified.
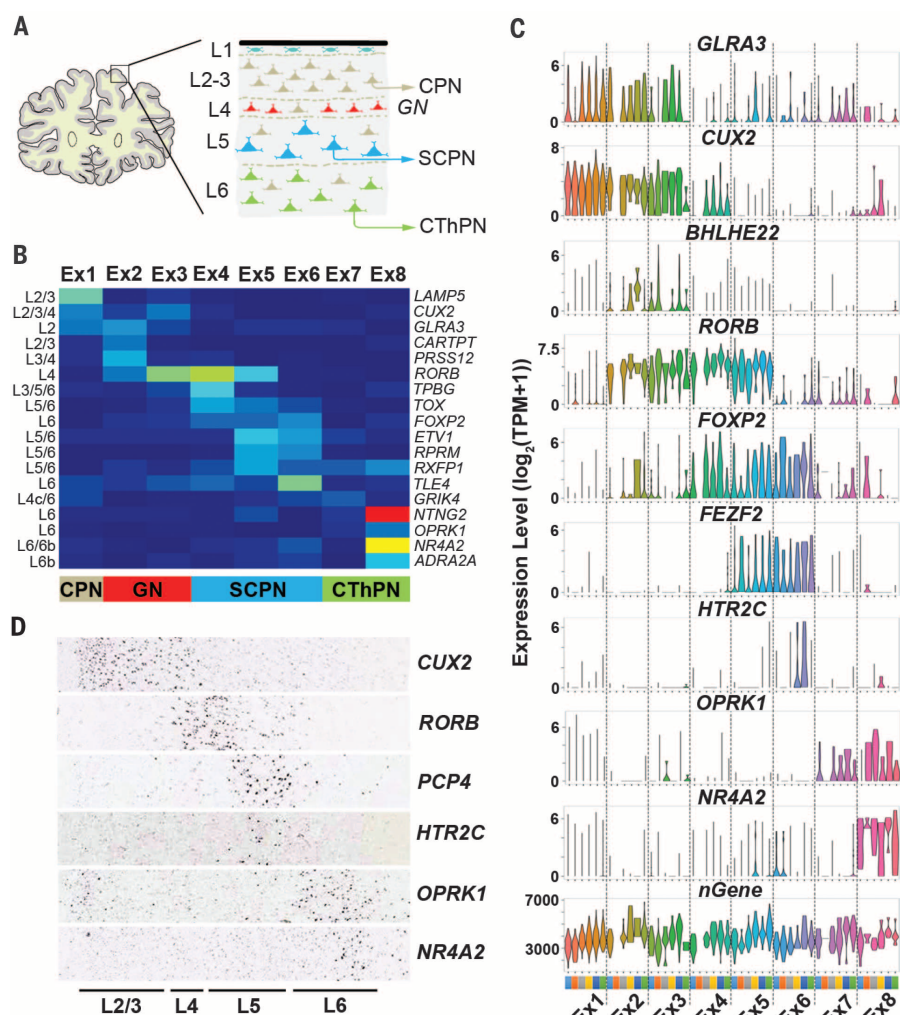
Fig. 3. Excitatory neuronal subtypes show distinct spatial organization. (**A**) Schematic of the prefrontal cortex showing projection neuron layers (L) and expected axonal projection destinations (layer 4 granule neurons typically receive outside inputs for distribution of signals locally). (**B**) FOP heatmap (scale as in Fig. 2A) for layer-specific marker genes showing expected cortical layer identity (L2-L6b) and excitatory neuron subclassification. CPN, cortical projection neuron; GN, granule neuron; SCPN, subcortical projection neuron; CThPN, corticothalamic projection neuron. (**C**) Violin plots showing selected marker gene expression values by Ex subtype and BA represented by colors (Fig. 2A). *nGenes*, total number of genes identified. (**D**) RNA ISH showing layer-specific expression of selected markers in the temporal cortex (Allen Human Brain Atlas, table S11).
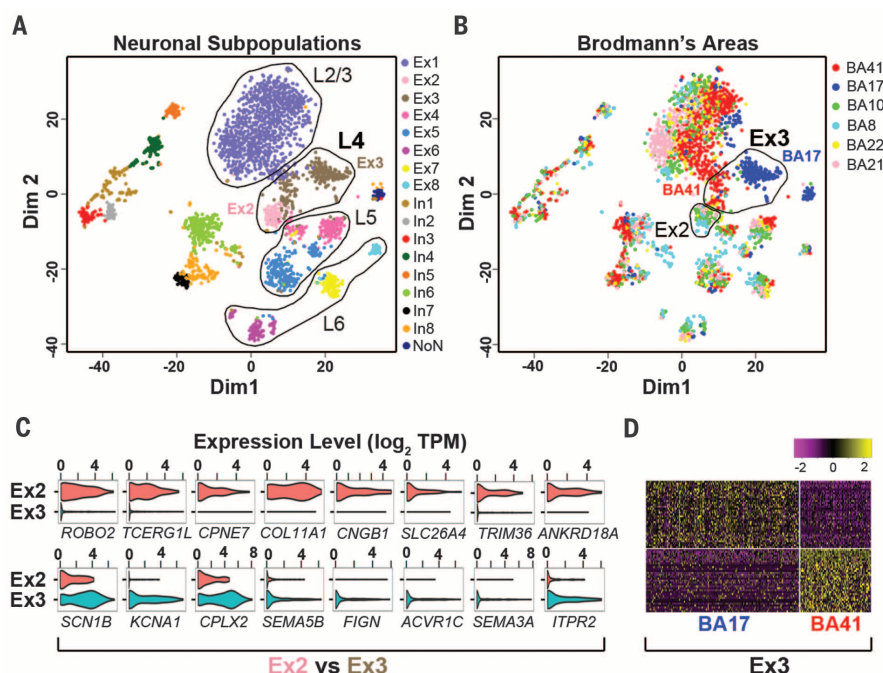
Fig. 4. Neuronal subtypes reveal heterogeneity among BAs. (**A**) Multidimensional plot showing projection neuron subtypes distributed according to their predicted cortical layer (L) identity. Layer 4 Ex2 and Ex3 subtypes are indicated. (**B**) Clusters shown in (A) colored by BA and with BA41/42 and BA17 subpopulations of Ex3 indicated. (**C**) Violin plots showing differentially expressed genes between Ex2 and Ex3 subtypes (table S8). (**D**) Heatmap showing genes differentially expressed between BA17 and BA41/42 within the Ex3 subtype (table S10).

Therefore, the potential exists for not only intra-regional cortical transcriptomic differences, but also further intrasubtype heterogeneity. This might reflect a technical need for increased sampling depth for further subtype resolution, yet may also indicate the potential for even more diversity within subtypes associated with a broader range of individualized neuronal activities. Consistent with these observations, proportions of subgroup-variable genes were associated with neuronal sub-type classification, postsynaptic function, and known regional expression variability (fig. S15C). These data support further local and regional functional heterogeneity existing among defined subtypes.

Our results demonstrate that postmortem SNS can identify expected and previously unidentified neuronal subtypes that provide insight into brain function through distinct profiles of activity-defining genes (fig. S16 and table S14). Furthermore, given that only a very small subset of layer-specific markers used in our analyses (*CARTPT*, *CHRNA7*, *PDYN*, and *RELN*) was found to have ISH differences between individual donors (*17*), our subtypes can be expected to be globally representative. Indeed, our subtypes remain highly conserved in mice (*3*), with differences highlighting evolutionary changes in potential orthologs (fig. S12). Our data sets reveal shared gene expression signatures that can distinguish subtypes and regional identity, supporting a transcriptional basis for well-known differences in cortical cytoarchitecture. Additional heterogeneity found within single neuronal transcriptomes may further reflect activities of complex neuronal networks that vary with function and time, as well as underlying genomic mosaicism that exists in human cortical neurons (*10*, *20–23*). Our study thus lays the groundwork for high-throughput global human brain transcriptome mapping using nuclei derived from readily available postmortem tissues for analyses of normal individuals, as assessed here, as well as myriad diseases of brain and mind.

REFERENCES AND NOTES

1. E. Z. Macosko et al., *Cell* **161**, 1202–1214 (2015).
2. D. Usoskin et al., *Nat. Neurosci.* **18**, 145–153 (2015).
3. A. Zeisel et al., *Science* **347**, 1138–1142 (2015).
4. M. J. Hawrylycz et al., *Nature* **489**, 391–399 (2012).
5. H. J. Kang et al., *Nature* **478**, 483–489 (2011).
6. M. B. Johnson et al., *Neuron* **62**, 494–509 (2009).
7. M. Hawrylycz et al., *Nat. Neurosci.* **18**, 1832–1844 (2015).
8. S. Darmanis et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7285–7290 (2015).
9. D. M. Bushman et al., *eLife* **4**, e05116 (2015).
10. J. Gole et al., *Nat. Biotechnol.* **31**, 1126–1132 (2013).
11. R. V. Grindberg et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19802–19807 (2013).
12. S. R. Krishnaswami et al., *Nat. Protoc.* **11**, 499–524 (2016).
13. Materials and methods are available as supplementary materials on *Science* Online.
14. A. Graf et al., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4139–4144 (2014).
15. D. Gaidatzis, L. Burger, M. Florescu, M. B. Stadler, *Nat. Biotechnol.* **33**, 722–729 (2015).
16. L. C. Greig, M. B. Woodworth, M. J. Galazo, H. Padmanabhan, J. D. Macklis, *Nat. Rev. Neurosci.* **14**, 755–769 (2013).
17. H. Zeng et al., *Cell* **149**, 483–496 (2012).
18. D. V. Hansen et al., *Nat. Neurosci.* **16**, 1576–1587 (2013).
19. T. Ma et al., *Nat. Neurosci.* **16**, 1588–1597 (2013).
20. M. A. Lodato et al., *Science* **350**, 94–98 (2015).
21. M. J. McConnell et al., *Science* **342**, 632–637 (2013).
22. S. K. Rehen et al., *J. Neurosci.* **25**, 2176–2180 (2005).
23. J. W. Westra et al., *J. Comp. Neurol.* **518**, 3981–4000 (2010).

BIOCHEMISTRY

Synthetic evolutionary origin of a proofreading reverse transcriptase

Jared W. Ellefson,* Jimmy Gollihar, Raghav Shroff, Haridha Shivram, Vishwanath R. Iyer, Andrew D. Ellington*

Most reverse transcriptase (RT) enzymes belong to a single protein family of ancient evolutionary origin. These polymerases are inherently error prone, owing to their lack of a proofreading (3′- 5′ exonuclease) domain. To determine if the lack of proofreading is a historical coincidence or a functional limitation of reverse transcription, we attempted to evolve a high-fidelity, thermostable DNA polymerase to use RNA templates efficiently. The evolutionarily distinct reverse transcription xenopolymerase (RTX) actively proofreads on DNA and RNA templates, which greatly improves RT fidelity. In addition, RTX enables applications such as single-enzyme reverse transcription–polymerase chain reaction and direct RNA sequencing without complementary DNA isolation. The creation of RTX confirms that proofreading is compatible with reverse transcription.

The molecular basis for life rests on the information flow between DNA, RNA, and proteins (*1*). Early notions of a unidirectional central dogma were amended after the discovery of the reverse transcriptase (RT) enzyme (*2*, *3*). The RT family has a single ancient evolutionary origin based on amino acid homology and the presence of RT across multiple domains of life (*4*). RTs are involved in processes such as telomere addition, mitochondrial plasmid replication, transposition, and the proliferation of retroviral genomes (*5*). It is also hypothesized to be the catalyst in the transition of the RNA to DNA world by providing an avenue to copy RNA into more stable DNA genomes (*6*).

The progenitor of RT is postulated to be an RNA-dependent RNA polymerase. Because RNA polymerases generally lack an error-checking 3′-5′ exonuclease domain, proofreading activity is also not present across the RT family, resulting in low-fidelity reverse transcription and characteristic quasispecies behavior in organisms that rely upon it for replication (*8*). In contrast to RTs, other DNA polymerase families have evolved exquisite proofreading mechanisms to increase DNA synthesis fidelity during genome replication (*9*).

To determine whether the evolutionary divide between RTs and DNA polymerases is a matter of history or function, we have attempted to directly evolve a reverse transcription xenopolymerase (RTX; Fig. 1A) from an error-correcting DNA polymerase using a modified directed evolution strategy (*10*), reverse transcription–compartmentalized self-replication (RT-CSR) (Fig. 1B). RT-CSR enables the simultaneous screening of up to 10^9 polymerase variants for RT activity.

We chose the Archaeal family-B DNA polymerases (*polB*) for directed evolution of the RTX as they are monomeric, hyperthermostable, highly processive, and contain proofreading domains. Attempts to rationally design these enzymes to use RNA templates have met with limited success (*11*, *12*), and initial experiments confirmed that two common *polB* enzymes from *Pyrococcus furiosus* and *Thermococcus kodakarensis* (KOD) (*13*, *14*) failed to polymerize across five template RNA bases (fig. S1). Modeling to identify mutations enabling RT activity was deemed impractical, given the extensive contacts these polymerases make with the template (>50 direct interactions). We initiated evolution using low-stringency RT-CSR (10 RNA residues) with a random library

Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, Department of Molecular Biosciences, University of Texas, 2500 Speedway, Austin, TX 78712, USA.

*Corresponding author. Email: jaredwellefson@gmail.com (J.W.E.); ellingtonlab@gmail.com (A.D.E.)

Science

AAAS

Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain

Blue B. Lake, Rizi Ai, Gwendolyn E. Kaeser, Neeraj S. Salathia, Yun C. Yung, Rui Liu, Andre Wildberg, Derek Gao, Ho-Lim Fung, Song Chen, Raakhee Vijayaraghavan, Julian Wong, Allison Chen, Xiaoyan Sheng, Fiona Kaper, Richard Shen, Mostafa Ronaghi, Jian-Bing Fan, Wei Wang, Jerold Chun and Kun Zhang (June 23, 2016)

Editor's Summary

Single-nucleus gene expression

Identifying the genes expressed at the level of a single cell nucleus can better help us understand the human brain. Blue *et al*. developed a single-nuclei sequencing technique, which they applied to cells in classically defined Brodmann areas from a postmortem brain. Clustering of gene expression showed concordance with the area of origin and defining 16 neuronal subtypes. Both excitatory and inhibitory neuronal subtypes show regional variations that define distinct cortical areas and exhibit how gene expression clusters may distinguish between distinct cortical areas. This method opens the door to widespread sampling of the genes expressed in a diseased brain and other tissues of interest.
Science, this issue p. 1586

Article Tools	Visit the online version of this article to access the personalization and article tools: http://science.sciencemag.org/content/352/6293/1586
Permissions	Obtain information about reproducing this article: http://www.sciencemag.org/about/permissions.dtl