NEURAL EPIGENOMICS

Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex

Chongyuan Luo,[1,2]* Christopher L. Keown,[3]* Laurie Kurihara,[4] Jingtian Zhou,[1,5] Yupeng He,[1,5] Junhao Li,[3] Rosa Castanon,[1] Jacinta Lucero,[6] Joseph R. Nery,[1] Justin P. Sandoval,[1] Brian Bui,[6] Terrence J. Sejnowski,[2,6,7] Timothy T. Harkins,[4] Eran A. Mukamel,[3]† M. Margarita Behrens,[6]† Joseph R. Ecker[1,2]†

The mammalian brain contains diverse neuronal types, yet we lack single-cell epigenomic assays that are able to identify and characterize them. DNA methylation is a stable epigenetic mark that distinguishes cell types and marks regulatory elements. We generated >6000 methylomes from single neuronal nuclei and used them to identify 16 mouse and 21 human neuronal subpopulations in the frontal cortex. CG and non-CG methylation exhibited cell type–specific distributions, and we identified regulatory elements with differential methylation across neuron types. Methylation signatures identified a layer 6 excitatory neuron subtype and a unique human parvalbumin-expressing inhibitory neuron subtype. We observed stronger cross-species conservation of regulatory elements in inhibitory neurons than in excitatory neurons. Single-nucleus methylomes expand the atlas of brain cell types and identify regulatory elements that drive conserved brain cell diversity.

Mammalian neuron types are identified by their structure, electrophysiology, and connectivity (*1*). The difficulty of scaling traditional cellular and molecular assays to whole neuronal populations has prevented comprehensive analysis of brain cell types. Sequencing mRNA transcripts from single cells or nuclei has identified cell types with unique transcriptional profiles in the mouse brain (*2*, *3*) and human brain (*4*). However, these methods are restricted to RNA signatures, which are influenced by the environment. Epigenomic marks, such as DNA methylation (mC), are cell type–specific and developmentally regulated, yet stable across individuals and over the life span (*5*–*7*). We theorized that epigenomic profiles using single-cell DNA methylomes could enable the identification of neuron subtypes in the mammalian brain.

During postnatal synaptogenesis, neurons accumulate substantial DNA methylation at non-CG sites (mCH) and reconfigure patterns of CG methylation (mCG) (*7*). Patterns of mCG and mCH at gene bodies, promoters, and enhancers are specific to neuronal types (*5*–*8*). Gene body mCH is more predictive of gene expression than mCG or chromatin accessibility (*5*). Because mCH is modulated over large domains, single-neuron methylomes with sparse coverage can be used to accurately estimate mCH levels for more than 90% of the genome by using coarse-grained bins (100 kb) (fig. S1). Whereas single-cell RNA sequencing mainly yields information about highly expressed transcripts, single-neuron methylome sequencing assays any gene or nongene region long enough to have sufficient coverage.

We developed a protocol for single-nucleus methylcytosine sequencing (snmC-seq) and applied it to neurons from young adult mouse (age 8 weeks) and human (age 25 years) frontal cortex (FC) (Fig. 1A) (*9*). snmC-seq provides a high rate of read mapping relative to published protocols (*10*–*12*) and allows multiplex reactions for large-scale cell type classification (fig. S2) (*9*). Like other bisulfite sequencing techniques (*13*), snmC-seq measures the sum of 5-methyl- and 5-hydroxymethylcytosines. Single neuronal nuclei labeled with antibody to NeuN were isolated by fluorescence-activated cell sorting (FACS) from human FC and from dissected superficial, middle, and deep layers of mouse FC. We generated methylomes from 3377 mouse neurons with an average of 1.4 million stringently filtered reads, covering 4.7% of the mouse genome per cell (Fig. 1, B and C, and table S1). We also

[1]Genomic Analysis Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [2]Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [3]Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92037, USA. [4]Swift Biosciences Inc., 58 Parkland Plaza, Suite 100, Ann Arbor, MI 48103, USA. [5]Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, CA 92093, USA. [6]Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [7]Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: ecker@salk.edu (J.R.E.); mbehrens@salk.edu (M.M.B.); emukamel@ucsd.edu (E.A.M.)

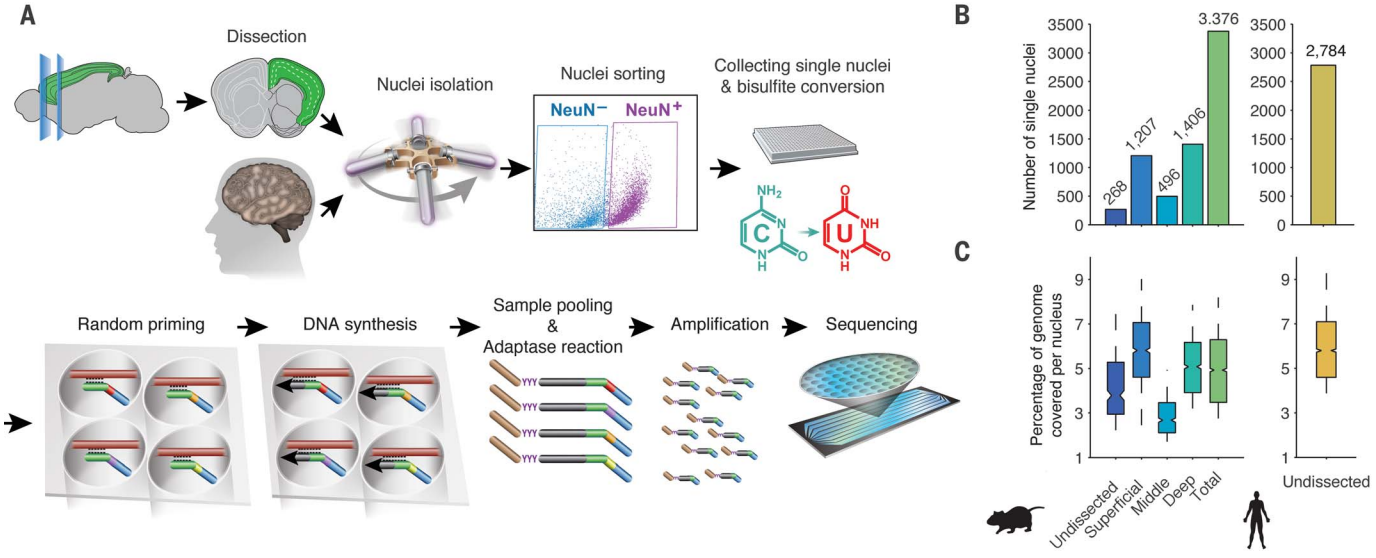

Fig. 1. High-throughput single-nucleus methylome sequencing (snmC-seq) of mouse and human frontal cortex (FC) neurons. (**A**) Workflow of snmC-seq. (**B** and **C**) Number of single-neuron methylomes (B) and distribution of genomic coverage per data set (C).

generated methylomes from 2784 human neurons with an average of 1.8 million stringently filtered reads, covering 5.7% of the human genome per cell (Fig. 1, B and C, and table S2).

We calculated the mCH level for each neuron in nonoverlapping 100-kb bins across the genome, followed by dimensionality reduction and visualization using t-distributed stochastic neighbor embedding [t-SNE (*14*)]. The two-dimensional tSNE representation was largely invariant over a wide range of experimental and analysis parameters (fig. S3). A substantially similar tSNE representation was obtained using CG methylation levels in 100-kb bins, which suggests that snmC-seq could be effective for cell type classification of nonbrain tissues without high levels of mCH (fig. S3F).

The mammalian cortex arises from a conserved developmental program that adds excitatory neuron classes in an inside-out fashion, progressing from deep layers (L5, L6) to middle (L4) and superficial layers (L2/3) (*1*). Inhibitory interneurons arise from distinct progenitors in the ganglionic eminences and migrate transversely to their cortical locations (*15*). We used mCH patterns to identify a conservative and unbiased clustering of nuclei for each species (*9*). Cluster robustness was validated by shuffling, downsampling, and comparison to density-based clustering (figs. S3 and S4) (*9*, *16*). In addition, clustering was not significantly associated with experimental factors (e.g., batches; false discovery rate > 0.1, χ^2 test; fig. S5).

We applied identical clustering parameters to mouse and human cortical neuron mCH data and identified 16 mouse and 21 human neuron clusters (Fig. 2, A to D). Assuming an inverse relationship between gene body mCH (average mCH across the annotated genic region) and gene expression (*7*), we annotated each cluster on the basis of depletion of mCH at known cortical glutamatergic or GABAergic neuron markers (e.g., *Satb2*, *Gad1*, *Slc6a1*), cortical layer markers (e.g., *Cux2*, *Rorb*, *Deptor*, *Tle4*), or inhibitory neuron subtype markers (e.g., *Pvalb*, *Lhx6*, *Adarb2*) (*1*, *15*, *17*) (Fig. 2, E and F, and figs. S6 and S7). For most clusters, mCH depletion at multiple marker genes (figs. S6 and S7) allowed us to assign cluster labels indicating the putative cell type. For example, we found a cluster of mouse neurons with ultralow mCH at *Rorb* (Fig. 2E and fig. S6), a known marker of L4 and L5a excitatory pyramidal cells (*17*). Combining this information with markers such as *Deptor* (fig. S6), which marks L5 but not L4 neurons, we labeled the cluster by species and layer (e.g., mL4 for mouse L4). Similarly, we used classical markers for inhibitory neurons such as *Pvalb* to label corresponding clusters (e.g., mPv for putative mouse Pvalb$^+$ fast-spiking interneurons) (*15*). We confirmed the accuracy of these classifications by comparison to layer-dissected cortical neurons (fig. S8, A and B) and coclustering with

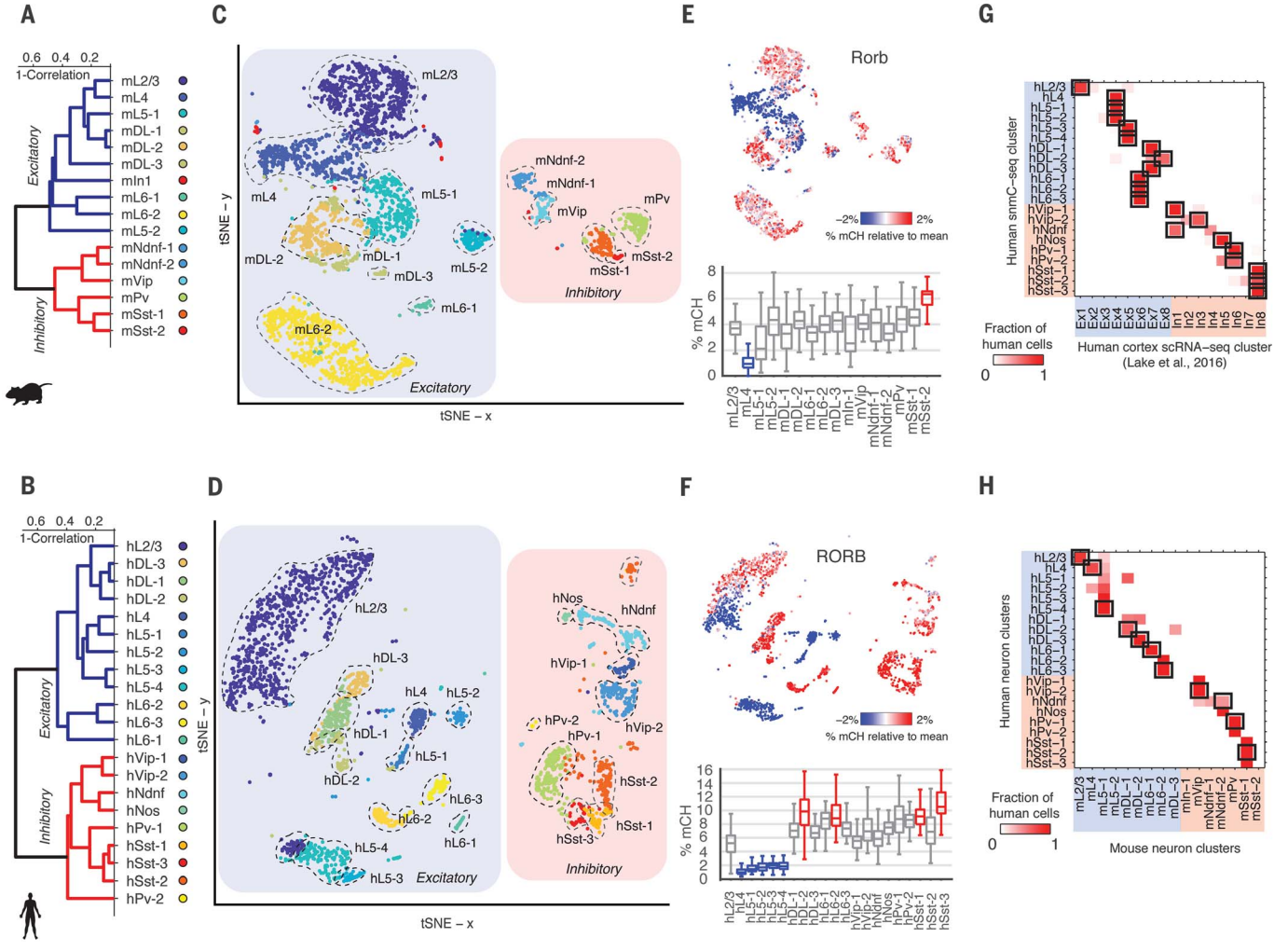

Fig. 2. Non-CG methylation (mCH) signatures identify distinct neuron populations in mouse and human FC. (**A** and **B**) Hierarchical clustering of neuron types according to gene body mCH level. (**C** and **D**) Two-dimensional visualization of single neuron clusters (tSNE) (*9*). Mouse and human homologous clusters are labeled with similar colors. (**E** and **F**) Gene body mCH at *Rorb* for each single neuron (top) and the distribution for each cluster (bottom); hyper- and hypomethylated clusters are highlighted in red and blue, respectively. (**G**) Comparison of human neuron clusters defined by mCH with clusters from single-nucleus RNA sequencing (*4*, *9*). (**H**) Fraction of cells in each human cluster assigned to each mouse cluster based on mCH correlation at orthologous genes (*9*). Mutual best matches are highlighted with black rectangles.

high-coverage methylC-seq data from purified populations of PV^+ and VIP^+ (*5*), as well as SST^+ inhibitory neurons (fig. S8C). Aggregated single-neuron methylomes showed consistent mCH and mCG profiles relative to bulk methylomes of matching cell populations (fig. S8, D and E, and fig. S15F). Neuronal cluster classification for each of the major cell subtypes in mouse and human cortex based on single-nuclei methylomes (Fig. 2G and fig. S9, A and B) was in good agreement with annotations based on single-cell RNA sequencing (*2*–*4*). Gene body mCH was anticorrelated with expression levels for corresponding clusters (fig. S9, C to E), validating our mCH marker gene–based annotation.

We found a greater diversity of excitatory neurons in deep layers than in superficial layers for both mouse and human (Fig. 2). In both species, we identified one neuronal cluster for cortical L2/3 (mL2/3, hL2/3) and L4 (mL4, hL4), whereas L5 and L6 contained seven clusters in mouse (mL5, mL6, and mDL, where DL denotes deep-layer neurons) and 10 clusters in human (hL5, hL6, and hDL). Mouse L5 excitatory clusters (mL5-1, mL5-2) were hypomethylated at *Deptor* and *Bcl6*, which mark cortical L5a and L5b, respectively (fig. S6) (*18*). L6 excitatory clusters included subtypes with low mCH at the L6 excitatory neuron marker *Tle4* [mL6-1, mL6-2; hL6-1, hL6-2, hL6-3 (*1*)]. Interestingly, several deep-layer neuron clusters (mDL-2, hDL-1, hDL-2, hDL-3) were not hypomethylated at *Tle4*. We identified marker genes for each neuron type on the basis of cell type–specific mCH depletion (table S3) (*9*). Although many marker genes were either classically established (*1*, *17*) or recently identified neuron type markers (fig. S10, A and B) (*2*–*4*), we identified a number of markers with no prior association to neuronal cell types (fig. S10, D and E, and table S3). mCH signature genes were hypomethylated in homologous clusters in mouse and human, with a few notable exceptions. For example, the mouse L5a marker *Deptor* showed no specificity for human L5 neurons (figs. S6, S7, and S10, A and B).

Most clusters were associated with classical cell type markers, but the identity of some clusters such as mDL-2 was less clear. We found that mDL-2 shares 24 marker genes with mL6-2, whereas 93 marker genes distinguish these clusters (table S3). To validate the distinction between the two cell types, we selected a shared marker (*Sulf1*) and one unique to mL6-2 (*Tle4*) and performed double in situ RNA hybridization experiments in mouse FC (fig. S11). The result confirmed the mCH-based prediction of a substantial proportion of L6 neurons expressing *Sulf1* but not *Tle4*; these neurons likely correspond to mDL-2 (fig. S11, A to D). The proportion of L6 neurons expressing both *Sulf1* and *Tle4* likely represents a subset of mL6-2 (fig. S11, A to D). *Tle4*-expressing neurons in somatosensory cortex project to the thalamus, whereas *Sulf1* is expressed by both corticothalamic and corticocortical projecting neurons (*18*); hence, the projection targets of neurons in cluster mDL-2 may be different from those of neurons in clusters showing hypomethylation of *Tle4* (e.g., mL6-2). We also observed extensive overlap of in situ hybridization signals when we used probes for a classical inhibitory neuron marker gene, *Pvalb*, and *Adgra3*, a predicted mCH signature of PV inhibitory neurons (fig. S11, E to G), further validating the specificity of marker prediction using mCH.

We paired homologous mouse and human neuron clusters by correlating mCH levels at homologous genes and found expanded neuronal diversity in human FC relative to mouse FC (Fig. 2H and fig. S12A) (*19*). Multiple human neuron clusters showed homology to mouse L5a excitatory neurons (mL5-1), L6a pyramidal neurons (mL6-2), or VIP, PV, and SST inhibitory neurons (Fig. 2H). We found a unique gene-specific mCH pattern and superenhancer-like mCG signatures in a potential human-specific inhibitory population (hPv-2; figs. S12B and S16J) (*9*).

Although we detected substantial mCH in all human and mouse neurons, cell types varied over a wide range in terms of their genome-wide mCH level (1.3 to 3.4% in mouse, 2.8 to 6.6% in human) (fig. S13, A to F). The sequence context of mCH was similar across all neuron types and consistent with previous reports (fig. S13, I and J) (*5*, *7*). Interestingly, global and gene-specific mCH differences were found in PV and SST inhibitory neurons located in different cortical layers (fig. S14) (*9*). Genes with low mCH in superficial-layer PV^+ neurons are enriched in functional annotations including neurogenesis, axon guidance functions, and synaptic component (fig. S14, F to H) (*9*), suggesting layer-specific epigenetic regulation of synaptic functions in inhibitory neurons.

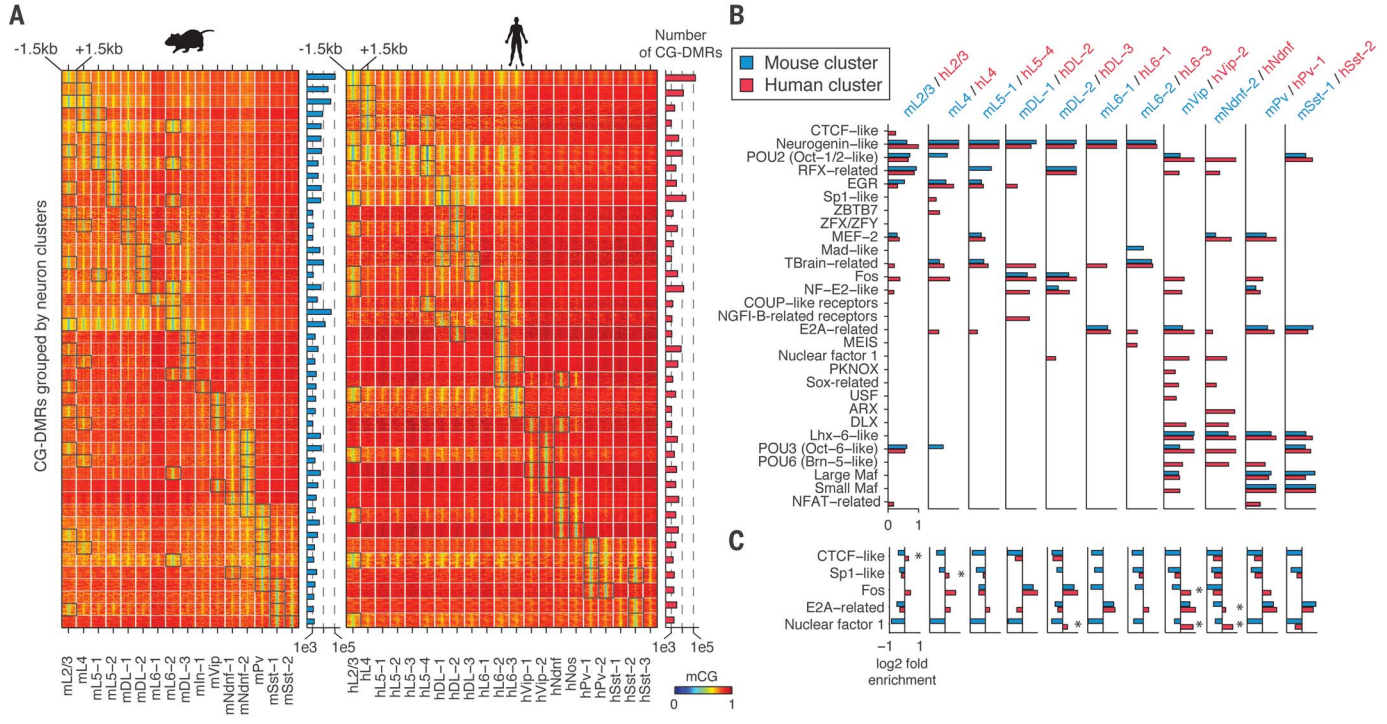

Fig. 3. Conserved and divergent neuron type–specific gene regulatory elements. (**A**) Heat map showing differentially methylated regions (CG-DMRs) hypomethylated in one or two neuron clusters; categories of DMRs containing >1000 regions are shown. (**B**) TF binding motif enrichment in CG-DMRs of homologous mouse and human clusters (false discovery rate < 10^{-10}). (**C**) Mouse- or human-specific enrichment and depletion of TF binding motifs. Asterisks indicate TF binding motifs that are significantly enriched in one species but depleted in the other.

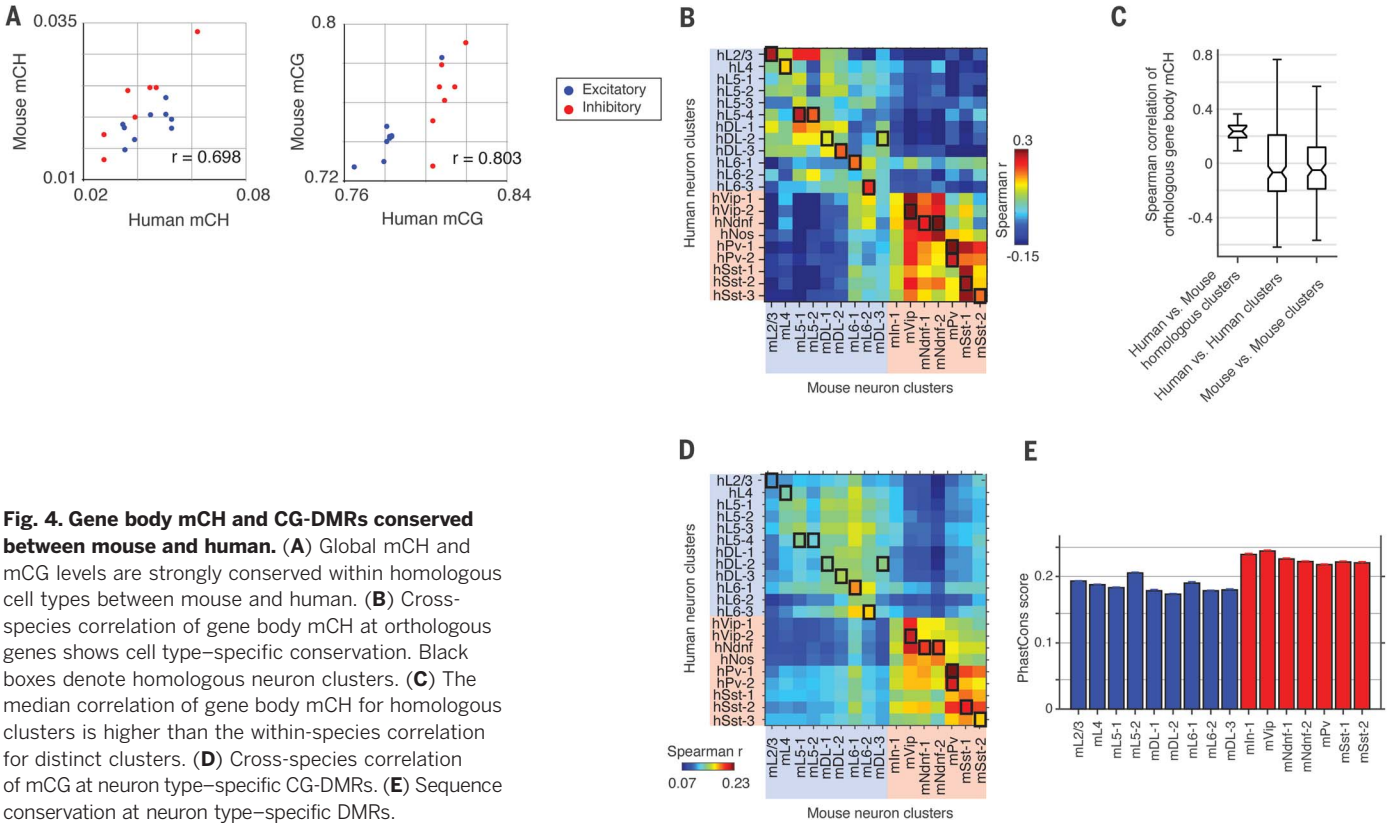

Fig. 4. Gene body mCH and CG-DMRs conserved between mouse and human. (**A**) Global mCH and mCG levels are strongly conserved within homologous cell types between mouse and human. (**B**) Cross-species correlation of gene body mCH at orthologous genes shows cell type–specific conservation. Black boxes denote homologous neuron clusters. (**C**) The median correlation of gene body mCH for homologous clusters is higher than the within-species correlation for distinct clusters. (**D**) Cross-species correlation of mCG at neuron type–specific CG-DMRs. (**E**) Sequence conservation at neuron type–specific DMRs.

A key advantage of single-cell methylome analysis is the ability to obtain regulatory information from the vast majority of the genome [>97% (*19*)] not directly assessed by RNA sequencing. By pooling reads from all neurons in each cluster, we could find statistically significant differentially methylated regions with low mCG in specific neuronal populations (CG-DMRs), which are reliable markers for regulatory elements (*5*). We found 575,524 mouse (498,432 human) CG-DMRs with average size of 263.6 bp (282.8 bp), covering 5.8% (5.0%) of the genome (Fig. 3A, fig. S15A, and tables S5 and S6). Most CG-DMRs (73.2% in mouse, 68.6% in human) are located >10 kb from the nearest annotated transcription start site (fig. S15, B to E). mPv and mVip CG-DMRs showed the strongest overlap with ATAC-seq peaks and putative enhancers identified from purified PV^+ and VIP^+ populations, respectively (fig. S15, G and H) (*9*, *20*). Hierarchical clustering of mCG levels at CG-DMRs grouped neuron types by cortical layer and inhibitory neuron subtypes (fig. S15, I and J). Thus, neuron type classification is supported by the epigenomic state of regulatory sequences.

We inferred transcription factors (TFs) that play roles in neuron type specification by identifying enriched TF-binding DNA sequence motifs in CG-DMRs (Fig. 3, B and C, and fig. S15K). We identified known transcriptional regulators and observed that several TF-binding motifs were enriched in human but depleted in mouse CG-DMRs in homologous clusters (Fig. 3C). The binding motif of NUCLEAR FACTOR 1 (NF1) was enriched in CG-DMRs for two human inhibitory neuron subtypes (hVip-2, hNdnf) but was depleted in homologous mouse clusters (mVip, mNdnf-2), suggesting a specific involvement of NF1 in human inhibitory neuron specification. Thus, although the TF regulatory circuits governing tissue types are conserved between mouse and human (*21*), fine-grained distinctions between neuronal cell types may be shaped by species-specific TF activity.

Superenhancers are clusters of regulatory elements, marked by large domains of mediator binding and/or the enhancer histone mark H3K27ac, that control genes with cell type–specific roles (*22*). Extended regions of depleted mCG (large CG-DMRs) are also reliable markers of superenhancers (fig. S16, A to C) (*9*, *23*). Therefore, we used our neuron type–specific methylomes to predict superenhancers for each mouse and human neuron type (fig. S16, D to I, and tables S7 and S8). For example, superenhancer activity was indicated by a large CG-DMRs at *Bcl11b* (*Ctip2*) in a subset of deep-layer neurons (fig. S16, F and G) and broad H3K27ac enrichment in mouse excitatory neurons (fig. S16F). Superenhancers overlap with key regulatory genes in the associated cell type, such as *Prox1* in VIP^+ and $NDNF^+$ neurons (fig. S16, H and I).

Global mCH and mCG levels were correlated between homologous clusters across mouse and human (Pearson $r = 0.698$ for mCH, $r = 0.803$ for mCG; $P < 0.005$), suggesting evolutionary conservation of cell type–specific regulation of mC (Fig. 4A and fig. S13, G and H). Examining 12,157 orthologous gene pairs, we found stronger correlation of gene body mCH between homologous clusters in mouse and human (median Spearman $r = 0.236$; Fig. 4, B and C) than between different cell types within the same species ($r = -0.050$, mouse; $r = -0.068$, human). For homologous clusters, we found shared and species-specific CG-DMRs based on sequence conservation (liftover; fig. S17, A and B). Cross-species correlation of mCG at CG-DMRs was significantly greater for inhibitory than for excitatory neurons ($P < 0.001$, Wilcoxon rank sum test; Fig. 4D and fig. S17C) (*9*). Greater sequence conservation at inhibitory neuron CG-DMRs could partly explain the greater regulatory conservation ($P < 0.001$, Wilcoxon rank sum test; Fig. 4E). Sequence conservation was observed only within 1 kb of the center of inhibitory neuron CG-DMRs and did not extend to the flanking regions (fig. S17G). These results support conservation of neuron type–specific DNA methylation, with greater conservation of inhibitory than of excitatory neuron regulatory elements.

Single-cell methylomes contain rich information enabling high-throughput neuron type classification, marker gene prediction, and identification of regulatory elements. Applying a uniform experimental and computational pipeline to mouse and human allowed unbiased comparison of neuronal epigenomic diversity in the two species. The expanded neuronal diversity in human, revealed by DNA methylation patterns, is consistent with more complex human neurogenesis, such as the presence of outer radial glia and the potential

dorsal origin of certain interneuron subtypes (*15*, *24*, *25*). Further anatomical, physiological, and functional experiments are needed to characterize the DNA methylation–based neuronal populations defined by our study. Single-neuron epigenomic profiling allowed the identification of regulatory elements with neuron type–specific activity outside of protein-coding regions of the genome. We expect that the single-nucleus methylome approach can be applied to studies of disease, drug exposure, or cognitive experience, thereby enabling examination of the role of cell type–specific epigenomic alterations in neurological or neuropsychiatric disorders.

REFERENCES AND NOTES

1. B. J. Molyneaux, P. Arlotta, J. R. L. Menezes, J. D. Macklis, *Nat. Rev. Neurosci.* **8**, 427–437 (2007).
2. A. Zeisel *et al.*, *Science* **347**, 1138–1142 (2015).
3. B. Tasic *et al.*, *Nat. Neurosci.* **19**, 335–346 (2016).
4. B. B. Lake *et al.*, *Science* **352**, 1586–1590 (2016).
5. A. Mo *et al.*, *Neuron* **86**, 1369–1384 (2015).
6. A. Kozlenkov *et al.*, *Nucleic Acids Res.* **44**, 2593–2612 (2016).
7. R. Lister *et al.*, *Science* **341**, 1237905 (2013).
8. A. Mo *et al.*, *eLife* **5**, e11613 (2016).
9. See supplementary materials.
10. S. A. Smallwood *et al.*, *Nat. Methods* **11**, 817–820 (2014).
11. M. Farlik *et al.*, *Cell Rep.* **10**, 1386–1397 (2015).
12. C. Angermueller *et al.*, *Nat. Methods* **13**, 229–232 (2016).
13. Y. Huang *et al.*, *PLOS ONE* **5**, e8888 (2010).
14. L. van der Maaten, G. Hinton, *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
15. C. P. Wonders, S. A. Anderson, *Nat. Rev. Neurosci.* **7**, 687–696 (2006).
16. M. Ester, H. Kriegel, J. Sander, X. Xu, *KDD'96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (Association for the Advancement of Artificial Intelligence, 1996), pp. 226–231; www.aaai.org/Papers/KDD/1996/KDD96-037.pdf.
17. E. S. Lein *et al.*, *Nature* **445**, 168–176 (2007).
18. S. A. Sorensen *et al.*, *Cereb. Cortex* **25**, 433–449 (2015).
19. F. Yue *et al.*, *Nature* **515**, 355–364 (2014).
20. Y. He *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1633–E1640 (2017).
21. A. B. Stergachis *et al.*, *Nature* **515**, 365–370 (2014).
22. W. A. Whyte *et al.*, *Cell* **153**, 307–319 (2013).
23. M. D. Schultz *et al.*, *Nature* **523**, 212–216 (2015).
24. D. V. Hansen, J. H. Lui, P. R. L. Parker, A. R. Kriegstein, *Nature* **464**, 554–561 (2010).
25. A. A. Pollen *et al.*, *Cell* **163**, 55–67 (2015).

ACKNOWLEDGMENTS

We thank C. O'Connor and C. Fitzpatrick at Salk Institute Flow Cytometry Core for sorting of nuclei; U. Manor and T. Zhang at Salk Institute Waitt Advanced Biophotonics Core for assisting with imaging; R. Loughnan for assistance of data analysis; and J. Simon for assisting with illustration. Supported by NIH BRAIN initiative grants 5U01MH105985 and 1R21MH112161 (J.R.E. and M.M.B.) and 1R21HG009274 (J.R.E.) and by NIH grant 2T32MH020002 (C.L.K.). T.J.S. and J.R.E. are investigators of the Howard Hughes Medical Institute. Data can be downloaded from NCBI GEO (GSE97179) and http://brainome.org. L.K. is an inventor on a patent application (US 14/384,113) submitted by Swift Biosciences Inc. that covers Adaptase. The source code for bioinformatic analyses is available at https://github.com/mukamel-lab/snmcseq and https://github.com/yupenghe/methylpy.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/357/6351/600/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S17
Tables S1 to S9
References (*26*–*44*)

31 March 2017; accepted 13 July 2017
10.1126/science.aan3351