

Epigenomic complexity of the human brain revealed by single-cell DNA methylomes and 3D genome structures

Wei Tian^{1,†}, Jingtian Zhou^{1,2,†}, Anna Bartlett¹, Qiurui Zeng^{1,3}, Hanqing Liu¹, Rosa G. Castanon¹, Mia Kenworthy¹, Jordan Altshul¹, Cynthia Valadon¹, Andrew Aldridge¹, Joseph R. Nery¹, Huaming Chen¹, Jiaying Xu¹, Nicholas D. Johnson⁴, Jacinta Lucero⁴, Julia K. Osteen⁴, Nora Emerson⁴, Jon Rink⁴, Jasper Lee⁴, Yang Li⁵, Kimberly Siletti⁶, Michelle Liem⁷, Naomi Claffey⁷, Caz O'Connor⁷, Anna Marie Yanny⁸, Julie Nyhus⁸, Nick Dee⁸, Tamara Casper⁸, Nadiya Shapovalova⁸, Daniel Hirschstein⁸, Rebecca Hodge⁸, Boaz P. Levi⁸, C. Dirk Keene⁹, Sten Linnarsson⁶, Ed Lein⁸, Bing Ren^{5,10,11,12,13}, M. Margarita Behrens⁴ and Joseph R. Ecker^{1,14,*}

¹Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA. ²Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, CA, USA. ³Division of Biological Sciences, University of California, San Diego, La Jolla, CA, USA. ⁴Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA. ⁵Ludwig Institute for Cancer Research, La Jolla, CA, USA. ⁶Department of Medical Biochemistry and Biophysics, Karolinska Institutet; Stockholm, Sweden. ⁷Flow Cytometry Core Facility, The Salk Institute for Biological Studies, La Jolla, CA, USA. ⁸Allen Institute for Brain Science; Seattle, WA, USA. ⁹Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA. ¹⁰Center for Epigenomics, University of California, San Diego School of Medicine, La Jolla, CA, USA. ¹¹Department of Cellular and Molecular Medicine, University of California, San Diego School of Medicine, La Jolla, CA, USA. ¹²Institute of Genomic Medicine, University of California, San Diego School of Medicine, La Jolla, CA, USA. ¹³Moore's Cancer Center, University of California, San Diego School of Medicine, La Jolla, CA, USA. ¹⁴Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA, USA.

†These authors contributed equally to this work

*Corresponding author. Email: ecker@salk.edu

Delineating the gene regulatory programs underlying complex cell types is fundamental for understanding brain functions in health and disease. Here, we comprehensively examine human brain cell epigenomes by probing DNA methylation and chromatin conformation at single-cell resolution in over 500,000 cells from 46 brain regions. We identified 188 cell types and characterized their molecular signatures. Integrative analyses revealed concordant changes in DNA methylation, chromatin accessibility, chromatin organization, and gene expression across cell types, cortical areas, and basal ganglia structures. With these resources, we developed scMCodes that reliably predict brain cell types using their methylation status at select genomic sites. This multimodal epigenomic brain cell atlas provides new insights into the complexity of cell type-specific gene regulation in the adult human brain.

The brain is the most complex organ in the human body consisting of billions of neuronal and non-neuronal cells, characterized by extensive diversity in gene expression, anatomy and functions. High throughput epigenomic profiling has been used to elucidate the gene regulatory programs underlying such cellular complexity (1–3), in both normal and disease brain states. Considered the fifth base of DNA, 5'-methylcytosines (5mCs) are the most common modified bases in mammalian genomes, providing an important epigenetic mechanism for regulating gene expression. Most 5mCs in vertebrate genomes occur at cytosine-guanine dinucleotides (CpGs). In vertebrate neuronal systems, however, 5mCs are also abundantly detected in non-CG (or CH, H=A, C, or T) contexts (4). Both CG- and CH-methylation (mCG and mCH, respectively) are highly dynamic during brain development and show remarkable cell type specificity (1, 5, 6). mCG and mCH are both essential for gene regulation and brain functions (7). In addition to DNA methylation, gene expression also requires proper conformation of chromatin folding (3D chromatin conformation), which is organized into active or repressive chromatin compartments, topologically associating domains

(TADs) and chromatin loops (8). Such spatial organization facilitates the interaction between gene promoters and their regulatory elements (e.g., enhancers), providing additional critical layers of regulatory mechanisms. DNA methylation and chromatin conformation interplay and coordinate in regulating gene expression and these processes are highly correlated (3).

Here, we profiled both DNA methylation and chromatin conformation in a large number of adult human brain cells from cortical and subcortical regions using single-nucleus single and multiomic epigenomic sequencing technologies. These epigenomic cell maps complement single cell/nucleus transcriptome-based brain cell census approaches by providing a genome-wide view of DNA methylation and chromatin organization in distinct brain cell types, allowing an assessment of the cell-type specific function of non-coding regions in the human genome (9, 10). Moreover, such epigenomic information not only allows annotation of cell type-specific regulatory elements but also provides a comprehensive description of the unique and dynamic nature of 3D chromatin structures found in cell types across brain regions (1, 2).

Epigenome-based brain cell type taxonomies

We dissected 46 brain regions covering brain structures of the cerebral cortex (CX, 22 regions), basal forebrain (BF, 2), basal nuclei (BN, 11), hippocampus (HIP, 5), thalamus (THM, 2), midbrain (MB, 1), pons (PN, 1) and cerebellum (CB, 2) (Fig. 1A, Fig. S1A, and tables S1 and S2). Most regions had independent biological replicates from three adult male donors (table S3) except for two regions in the amygdala (BM and CEN; two donors for each) due to the dissection difficulty and tissue availability (Fig. S1A and table S1). Fluorescence-activated nuclei sorting (FANS) was used to isolate 90% NeuN-positive and 10% NeuN-negative cells in each sample to enrich neurons (Fig. S1A). We profiled genome-wide DNA methylation in each cell using snmC-seq3 (“mC”)(11) with samples from all 46 brain regions. Additionally, we also simultaneously profiled DNA methylation and chromatin conformation from individual cells using snm3C-seq3 (“m3C”)(3) with samples from 17 brain regions from CX, BF and BN (Fig 1B, and Fig. S1A). In total, 378,940 mC and 145,070 m3C nuclei passed rigorous quality control measures (Fig. S1B; see Methods). On average, 0.94 million stringently filtered reads were produced per mC cell and 2.20 million reads per m3C cell, covering 3.3% and 5.4% of the cytosines of the human genome, respectively. On average, 406,000 chromatin contacts were detected per m3C cell. These levels of genome coverage enabled reliable quantification of the DNA methylation levels of genomic features (Fig. S1C), identification of methylation variable regions, and determination of TADs and chromatin loops in distinct cell types throughout the brain.

We performed iterative clustering analysis on the mC dataset using CG- and CH-methylation levels on 100 kb bins throughout the human genome (Methods). The nuclei were first classified into three cell classes: telencephalic excitatory neurons, inhibitory/non-telencephalic neurons, and non-neuronal cells (Fig. 1, C and D). Each cell class was further divided into major cell types (referred to as “major types” through this paper) and annotated based on CH-hypomethylated gene markers for neuronal cell types and CG-hypomethylated markers for non-neuronal cell types. Where possible, cell types were annotated using the nomenclature for known brain cell types previously described in the literature (e.g. (12)); otherwise, cell clusters were annotated according to either the regional composition or distinct marker genes of the cell type. One caveat of the former approach is that cell types annotated using marker genes defined in rodents or non-human primates might not reflect the corresponding gene activity in human cell types. For example, the gene SNCG is lowly expressed in the human major type corresponding to mouse Sncg cells. Nevertheless, using common nomenclature aids in cross-species comparison and existing knowledge transfer. We

identified 40 major types from the three cell classes, which were further divided into 188 cell subtypes (Fig. 1D, Fig. S2, A to C, and tables S4 and S5). We performed subtype clustering using stringent parameters, requiring each subtype to have at least five unique marker genes compared with all other subtypes (Methods). Despite some cell proportion variation, all major types and subtypes were conserved across donors (Fig. 1D and Fig. S2C).

While non-neuronal major types have a relatively even distribution across brain structures, major neuronal types show considerable spatial specificity (Fig. 1, D to F). Most major telencephalic excitatory neurons are specifically grouped by spatial location. Excitatory neurons from the hippocampus are divided into separate clusters according to their origins in the sub-structures of the hippocampus (CA1, CA3, & DG). Cortical excitatory neurons cluster together according to both their cortical layers (L2/3, L4, L5, L5/6, & L6; L=layer) and projection types (IT = intratelencephalic, CT = corticothalamic, ET = extratelencephalic, and NP = near-projecting). The basal nuclei excitatory neurons mainly originate from the amygdala and form one specific group. The major type L6-IT-Car3 consists of neurons originating from the deep layers of the cortex and claustrum of the basal nuclei. Although they have similar DNA methylation patterns at the major type level, a recent study in mice showed that L6-IT-Car3 cells have divergent projection patterns (13). Telencephalic inhibitory neurons formed eleven major types, seven from mostly cortical regions (Pvalb, Pvalb-ChC, Sst, Lamp5, Lamp5-Lhx6, Sncg, and Vip) and four from basal nuclei and basal forebrain (MSN-D1, D2, Foxp2, and Chd7). In the thalamus, we identified one excitatory and two inhibitory major types. One inhibitory major type, THM-MB, shares similar DNA methylation profiles with a small population of cells in the midbrain. The other inhibitory major type, THM-Inh, is very rare; we detected only 361 cells in the entire dataset (0.07% of brain cells profiled). As THM-Inh features hypomethylation in genes EPHB1 and EPHA1 (Fig. S2D), this major type could originate from the habenular nuclei of the thalamus, possibly due to contamination during dissection. Neurons from Pons formed one specific major type (PN). The cerebellum contains two highly specific major types: Purkinje cells (PKJ), a rare cell type in the brain (867 cells, 0.17%), and cerebellar granular cells (CB). Some populations of cells from the basal nuclei and the midbrain form a conspicuous major type SubCtx-Cplx, which shows striking heterogeneity among its subtypes. For example, SubCtx-Cplx consists of both excitatory and inhibitory cells (Fig. 1E), while all other major neuronal types contain only one of them. SubCtx-Cplx also features highly variable DNA methylation among subtypes in the genes of neurotransmitter receptors, transporters, and neuropeptides (Fig. S2E).

The cell types determined from single nucleus DNA methylation profiles were corroborated using two orthogonal data types: single-cell transcriptome (scRNA-seq) data (see companion manuscript Siletti et al. (14)) and

single-cell chromatin accessibility data (snATAC-seq; see companion manuscript Li et al.) produced using nuclei prepared from dissection regions from the same three human brains (Methods). Integrative analysis between these

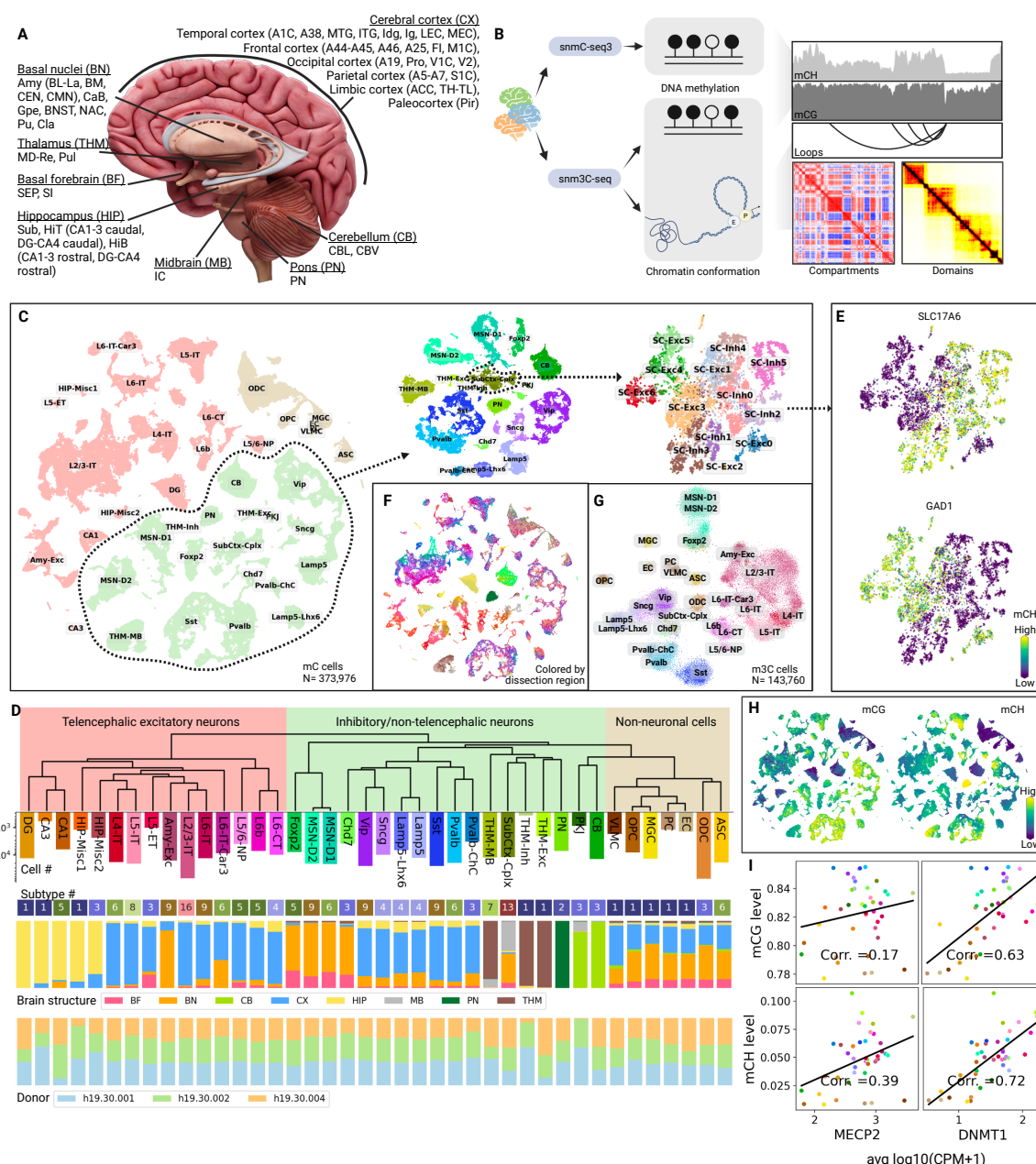


Figure 1. Epigenomic profiling of human brain cells with snmC-seq3 and snm3C-seq. (A) Human brain structures and regions covered in this study. (B) Schematics of profiling aspects of snmC-seq3 and snm3C-seq. (C) Iterative clustering and annotation of human brain nuclei. Cells from the whole mC dataset, from the inhibitory/non-telencephalic neuron cell class, and from the SubCtx-Cplx major type are visualized successively using t-distributed stochastic neighbor embedding (t-SNE), colored by the cell groups annotated in the corresponding iterations. (D) The robust dendrogram of the major types and the meta info of subtype numbers, brain structure, and donor origins. (E) The SubCtx-Cplx major type has both excitatory and inhibitory subtypes, marked by CH-hypomethylated genes SLC17A1 and GAD1, respectively. (F) Human brain cells are colored by the dissection regions. (G) 2D visualization of brain nuclei profiled by snm3C-seq. The visualization was generated from chromatin conformation data in 100kb resolution. (H) Global CG- and CH-methylation levels vary across brain cell types. (I) Global methylation levels of major types correlate with DNA methylation reader MECP2 and writer DNMT1.

datasets reveals the strong correspondence between cell types determined using different molecular modalities (Fig. S3A). Notably, all epigenome-based cell subtypes correspond well with transcriptome-based clusters (Fig. S3B), though they were derived from ~10 times more cells and from ~2 times more brain regions. The DNA methylation-derived cell type labels were transferred to the matching scRNA and snATAC cells to facilitate further multi-omic joint analysis (Methods).

The robust dendrograms constructed from the single nuclei DNA methylation profiles show similarities between major types and subtypes (Fig. 1D and Fig. S2C; Methods). Telencephalic excitatory and inhibitory/non-telencephalic neurons are well-separated from non-neuronal cells, each type forming a specific clade (except CB and PKJ). CB and PKJ were grouped with the non-neuronal cell types in the dendrogram, likely owing to their similar global levels of CG- and CH-methylation (Fig. 1H and Fig. S4A).

Global methylation levels varied across major cell types ranging from 77.7% to 85.5% for CG-methylation and from 0.8% to 10.7% for CH-methylation. Non-neuronal and granular major types (DG and CB) have the lowest global levels in CG- and CH-methylation, consistent with the previous study in mice (*1*). Cortical inhibitory neurons have the highest CG-methylation levels, while several non-telencephalic neurons, particularly from the thalamus, midbrain, and pons, show the highest CH-methylation levels (Fig. 1H and Fig. S4A). Interestingly, cell-type global methylation levels correlate with the gene expression levels of DNA methylation readers and modifiers (Fig. 1I and Fig. S4B). The expression level of MECP2 and DNMT3A, the major CH-methylation reader and writer, show positive correlations ($\rho=0.39, 0.35$) with the global CH-methylation levels in contrast to a weaker correlation with CG-methylation ($\rho=0.17, 0.08$; Fig. 1I and Fig. S4B). The CG DNA methyltransferase gene DNMT1 showed a positive correlation ($\rho=0.63$) between its expression levels and global mCG across cell types (Fig. 1I), matching its role as the major mCG maintainer in mature neurons (*15*). On the other hand, we unexpectedly observed an even higher correlation between DNMT1 expression levels and mCH levels ($\rho=0.72$, Fig. 1I) compared with mCG; though DNMT1 is thought to have little effect on mCH (*16*). These findings suggest an unknown relationship between DNMT1 and mCH or some yet-to-be-discovered factor influencing both DNMT1 expression and mCH.

To obtain consistent cell type annotation for the cells profiled by m3C, we iteratively co-clustered the m3C nuclei with mC nuclei and transferred cell type labels from annotated mC nuclei to m3C nuclei (table S6; Methods). With the improved scHiCluster framework (*17*), we not only readily separated nearly all major cell types (exceptions being MSN-D1 and MSN-D2) using the chromatin contact

only (Fig. 1G) but also revealed, for the first time, considerable regional diversity on the single-cell embedding (Fig. S2F). These results suggest considerable cell type and regional specificity of 3D chromatin structures.

Chromatin organization in human brain cells

Chromatin is organized into structures at different scales. The subchromosomal-level compartment brings together regions that are tens to hundreds of megabases (Mb) away, while TADs and chromatin loops are driven by interactions within several megabases (*18*). To examine the cell-type specificity of genome folding at different lengths scales, we first analyzed the proportion of contacts detected at different genome distances within each single cell. Intriguingly, we observed that neurons are highly enriched for shorter-range interactions (<2Mb), while in most non-neuronal cells, longer-range contacts (>20Mb) are enriched (Fig. 2, A to C). Astrocyte and oligodendrocyte progenitor cells show intermediate interaction ranges, while mature oligodendrocytes are more similar to the non-neuronal cells (Fig. 2, A and B). In addition, cortical and subcortical excitatory neurons are more enriched for shorter-range interactions than cortical inhibitory cells (Fig. 2, A and B). The ratio between shorter and longer interactions is highly correlated with the number of RNA molecules detected in the cell type ($\rho=0.87$, Fig. S5A), indicating the contact distance could be associated with gene expression activity. These findings suggest that the contact distance spectrum, which is usually thought to be a signature of cell-cycle phases (*19*), can also be highly cell-type dependent.

To investigate whether the increase in long-range chromatin interactions in non-neuronal cells leads to stronger compartmentalization, we identified chromosome compartments within each major type at 100 kb resolution (Fig. 2D). This analysis revealed that the enrichment of longer-range interactions showed little correlation with the compartment strength (Fig. S5, B and C), indicating that the enriched long-range interactions include both intra- and inter-compartment interactions. We next hypothesized that the methylation status of two genome loci is more likely to change together if they are physically closer and have stronger 3C interactions. Indeed, the co-methylation coefficient matrices (correlation of methylation levels between two genomic bins across single cells) form plaid patterns that resemble the compartment structures defined by chromatin contacts (Fig. 2E and Fig. S5D), suggesting the genome is segregated into local co-methylation domains as well as two sets of co-methylation regions across the chromosome. Notably, this coregulation structure was also observed with single-cell ATAC-seq data for chromatin accessibility (*20*), which, together with our data, provides independent evidence for genome compartmentalization. To further test the association between DNA methylation and 3D genome architecture, we compute the correlation between the chromatin interaction strengths and the average

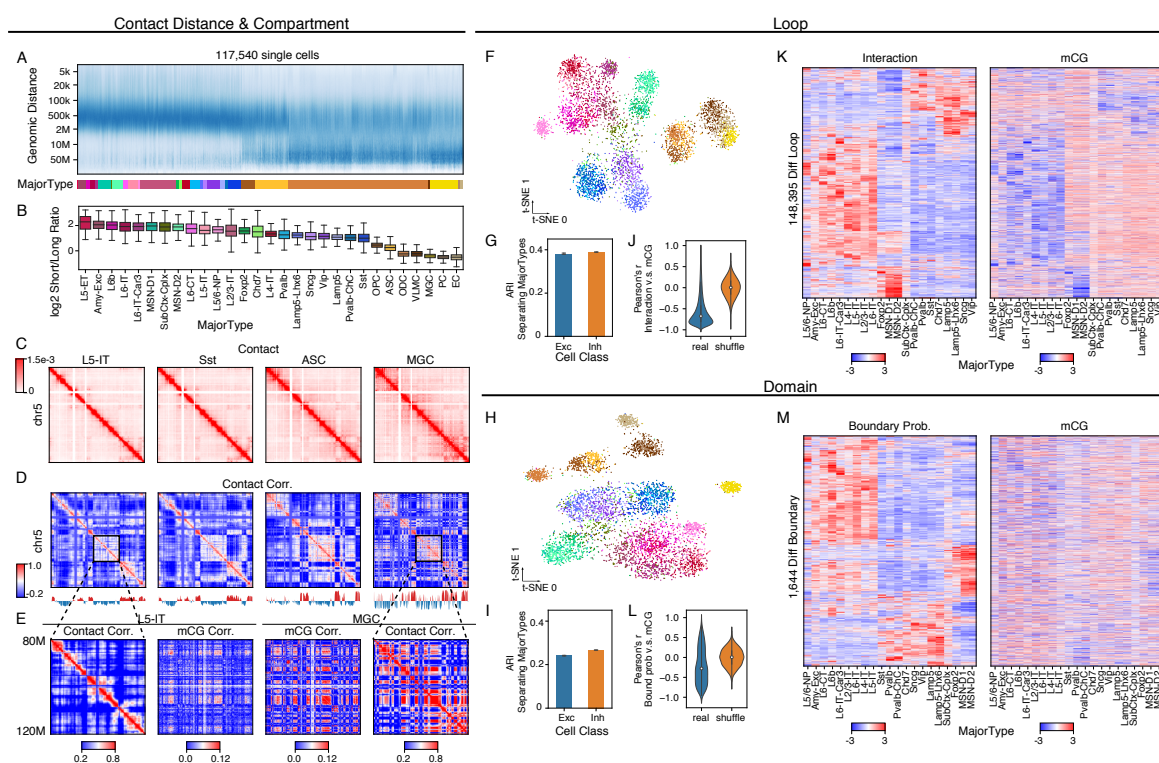


Figure 2. Diversity of 3D genome structures across major types. (A) Contact frequency against genomic distances across single cells. The bottom color bar shows which major type the cell belongs to. (B) Log2 ratio between shorter range and longer range contact frequency in each major type, ordered by the median ratio over cells. (C) The imputed contact maps of four major types. (D) The heatmaps show the correlation matrices of distance normalized contact maps in (C), and the line plots show the first principal component of the correlation matrices, with 0 as the middle line to separate red and blue. (E) Zoom in view of two matrices in (D) and the corresponding correlation matrices of CG methylation across cells. (F, H) t-SNE plot of cells (n=5,707) using loops (F) or domains (H) as features, colored by major types. (G, I) The adjusted rand index (ARI) between cortical excitatory major types or cortical inhibitory major types and clusters identified with loop (G) or domain (I) based cell embedding. Bars represent mean $\pm$ s.e.m of K-Means with 10 random seeds. (J) The imputed contact strength of differential loops and the average mCG level of the two loop anchors. (K) The distribution of Pearson's correlation between two matrices in (J) across rows. (L) The boundary probability of differential domains and the mCG level at the boundaries. (M) The distribution of Pearson's correlation between two matrices in (L) across rows. For all box plots, the center line denotes the median; box limits denote the first and third quartiles, and whiskers denote $1.5 \times$ the interquartile range.

methylation levels of the interacting anchors across single cells at 100 kb resolution. We noticed that these correlations are also connected with chromosome compartments (Fig. S5E), where the negative correlations are more likely to be observed in the active compartment (A Compartment; p-values $<1e-300$, Wilcoxon Rank-Sum test; Fig. S5F). Next, we identify domains in single cells using the imputed contact matrices at 25 kb resolution. More domains are identified in neurons (median 4,813) than in non-neurons (median 4,308, p-values $<1e-300$, Wilcoxon Rank-Sum test) (Fig. S5G). We defined the boundary probability of a genomic bin as the proportion of cells in which the bin is determined as a domain boundary. The recurrent domain boundaries across single cells correspond to the domain boundaries identified with cell-type pseudo-bulk contact maps (Fig. S5H). We also identified chromatin loops at 10 kb resolution in each of the 119 cell subtypes with ≥ 100 m3C cells (median 541,551 loop pixels, with 59,905 loop

summits). In total, 24.3% of loops represent interactions between distal DMRs and annotated gene promoters (TSS ± 2 kb), 38.1% are distal DMR-DMR loops, and 5.8% are promoter-promoter loops (Fig. S5I). Using either domain boundary probabilities or loop strengths was able to recapitulate the hierarchical structure of cell type similarities (Fig. S5, J and K) and separate single cells by cell types in the embedding spaces (Methods; Fig. 2, F and H), suggesting the cell type specificity of these 3D structural features. Chromatin domains showed relatively weaker power to distinguish finer scale major types within excitatory or inhibitory cell class (Fig. 2, G and I). This finding indicates that different scales of 3D features may play a role in the gene regulation specification at different granularities of cell types, where the DNA loops we identified are more major type-specific than domains.

To investigate the cell-type-specific 3D structural features more closely, we identified 1,644 differential domain

boundaries across the major cell types. Chromatin domains had been considered conserved across cell types (21–24). However, in agreement with more recent studies showing domain dynamics across cell types and development (3, 25–27), our data suggest that domains may differ even between similar cell types (Fig. S5H). We also identified 148,395 differential loops across major types. Promoter-DMR loops show higher cell-type specificity than promoter-promoter or DMR-DMR loops (Fig. S5L). The strength of loops is strongly anticorrelated with DNA CG methylation (Fig. 2, J and K), and this anticorrelation tends to be stronger at loops with higher variability across cell types (Fig. S5M). The boundary probability is also anticorrelated with CG methylation; however, the correlation is weaker compared with loops (Fig. 2, L and M). The anticorrelation could have resulted from the effect of DNA methylation on the binding of factors driving genome folding (e.g., CTCF) or the recruitment or exclusion of methylation writer/eraser (e.g., DNMT, TET) through loop formation. Further developmental or mechanistic studies are needed to resolve a causality relationship and its possible interplay with gene expression. Together, these analyses reveal the cell-type specificity of chromatin architecture and its relationship with DNA methylation at an unprecedented cell-type resolution in the human brain.

Cell-type specific DNA methylation patterns and associated gene regulatory landscapes

To delineate the cell type-specific methylation profiles, we identified 24,455 CH- and 13,096 CG- pairwise differentially methylated genes (DMGs) among the subtypes (Fig. S6A; Methods). We also identified 2,059,466 CG- differentially methylated regions (DMRs) in the human genome from 188 brain cell subtypes (Fig. 3A; Methods). These cell type-specific DMG and DMR patterns depict distinct DNA methylation signatures for brain cell identities. In addition, they provide critical information to understand gene regulatory programs in brain cells. Both CG- and CH-methylation levels of DMGs negatively correlate with their gene expression levels (5, 28, 29), while DMRs predict candidate cis-regulatory elements (cCREs) (6, 28), and motifs enriched in their sequence indicate the cell type-specific transcription factors (TFs; (29)).

To determine putative TFs in each cell type, we performed a motif enrichment analysis for cell type-specific hypomethylated DMRs (hypo-DMRs; see Methods). Combining with the DMG analysis results, we assigned TFs to specific cell types if a TF has its gene body hypomethylated in the cell type and its binding motif enriched in hypo-DMRs of the cell type (Fig. S6B; Methods). In total, 612 TFs were assigned to major neuronal types and subtypes, where they potentially play important roles in shaping and maintaining cell identities. For example, the TF TBR1 which plays a fate-determining role in the

development of corticofugal projection neurons (30) was assigned to deep-layer excitatory neurons, particularly L6-CT and L6b (Fig. 2B). The TFs ZNF423 and EBF2 were assigned to Purkinje cells. ZNF423 was additionally assigned to cerebellar granular cells (Fig. 2B). Both TFs are crucial in the proper development of the cerebellum, and EBF2 particularly directs the migration of Purkinje cells (31–33).

In addition to major types, analysis of subtypes can further provide insights into subtle differences in TF utilization among subtypes. For instance, the TF PBX3, assigned to the MSN-D1 major type, is only hypomethylated in the subtypes showing hypo-methylation at MEIS2 and FOXP1 genes, two marker genes of striatal striosome compartment (Fig. S6C). This suggests that PBX3 may also be expressed only in the striosome but not in the matrix compartment (Fig. 3B). This finding agrees with the previous observation that PBX3 is preferentially expressed in striosomes, and could be crucial for striatum development (34, 35). Furthermore, examination of the genome-wide potential binding sites (overlapped regions between hypo-DMRs and TF motifs) of PBX3 also shows that the average methylation levels are lower in striosome subtypes compared to the matrix (Fig. 3B), suggesting a spatial-specific regulating role of this TF in the striatum.

We further integrated information on DMGs, DMRs, and different loops to identify putative enhancers in each cell type (Fig. 3C). DMRs were associated with genes if the TSS of the gene was within 5 Mb of the DMR. Pearson correlations between mCG levels of DMRs and mCH levels of genes body across cell subtypes were calculated to assess the association. The associations are slightly stronger between DMRs and DMGs than between DMRs and other genes (Fig. S6D). We further filtered the DMR-DMG pairs and kept only the ones that overlapped with both anchors of a loop or a differential loop (DL). Both loop- and DL-filtered DMR-DMG pairs show stronger associations (Fig. 3D and Fig. S6D). DL-filtered DMRs have the strongest association with paired genes (Fig. 3D) and an increasing overlapping ratio with open chromatin regions (Fig. 3E), indicating their putative regulating roles in adult brain cells. Putative regulating DMR/gene pairs (3.2M) were identified between 1,122,919 DMRs and 12,327 genes. These DMRs, their paired genes, and the corresponding loops were (anti-) correlated (Fig. 3F) and together may collectively orchestrate cell type-specific gene regulatory programs. For example, the gene SYT1 encodes Synaptotagmin-1, an integral membrane protein of synaptic vesicles, which serves as the major switch allowing neurons to release neurotransmitters at synapses when sensing calcium ions (36). Distal DMRs hypomethylated in L2/3-IT are connected to the promoter region of SYT1 through DNA loops which are either depleted or weakened in MSN-D1 cells.

associations between 20 brain diseases or complex traits (Methods) and DMRs or loop-overlapping DMRs in human brain cells (Fig. 3H and Fig. S6F; Methods). Risk variants related to schizophrenia, bipolar disorder, and neuroticism show strong enrichment in hypomethylated DMRs of excitatory neurons located in the cortices and hippocampus. Alzheimer's disease (AD) is associated with microglia (MGC; Fig. 3H; (39)). Risk variants of tobacco usage disorder are enriched in a dopamine neuron cell type *Foxp2* from basal ganglia (Fig. 3H), a brain region related to tobacco addiction (40).

Regional heterogeneity in cortices and basal ganglia

In addition to the distinct cell type specificity for the brain anatomical structures discussed above, we also observed that cells of the same type from different regions within one brain structure also manifest considerable DNA methylation diversity. Prominent functional and gene expression heterogeneity across the cortex has been described in previous studies, particularly in the neocortex and hippocampus. Gradients of morphogens in the developing mouse neocortex play crucial roles in early cortical region specialization (41). Microarray profiling showed that the transcriptomes of two neocortical regions are more similar if the two regions are closer spatially (42). Recent single-cell studies also revealed that the regional heterogeneity is cell type dependent in the mouse brain, which was particularly prominent in neocortical excitatory neurons in gene expression and gene body DNA methylation (1, 43). In addition to the neocortex, our dataset encompasses a broader range of brain regions, including the allocortex (e.g. Pir) and periallocortex (e.g. LEC and MEC), allowing investigation of the gene regulatory elements underlying the transcriptome diversity across these brain regions.

To distinguish regional diversity from other confounding sources of heterogeneity within one cell type, we developed a workflow to extract the regional landscape hidden within single-nuclei DNA methylation profiles (Fig. 4A). By combining single nuclei DNA methylation profiles and the corresponding brain region information, we transformed the cells from the DNA methylation space to a “joint methylation-brain region space” (Fig. 4A; Methods), representing from which brain regions cells sharing similar methylation profiles were dissected. Specifically, if two cells are closer in this space, their methylation neighbors (neighboring cells in the methylation space) are more likely to come from similar brain regions. Therefore, a cell's trajectory in this space represents a transition across brain regions accompanied with changes in their methylomes. By emphasizing the regional methylation similarity rather than simply capturing the major axes of methylation variation, the transformed embedding of single cells facilitates further analyses examining regional effects on DNA methylation in brain cells.

Cortical excitatory neurons showed remarkable regional diversity in their DNA methylation profiles, especially the intratelencephalic-projecting neurons (LX-IT; Fig. 4B). The regional diversity of cortical inhibitory neurons (44) was less studied due to their inconspicuous regional patterns in transcriptome and epigenome (1, 43, 45). Our analysis uncovered their regional diversity: despite being less profound, cortical inhibitory neurons showed considerable regional differences (Fig. 4B). The embeddings in the regional spaces showed a shared gradient among cortical neuronal cell types, both excitatory and inhibitory (Fig. 4B). We further derived regional axes for each cell type in the aforementioned joint methylation-brain region spaces using a single-cell trajectory analysis tool STREAM (46) to quantitatively analyze the gradient molecular diversity across regions. Although the regional axes may have certain discrepancies among cell types, they followed a common Posterior–Lateroanterior–Medioanterior (P–LA–MA) trend: from the posterior part of the brain (e.g., the primary visual cortex VIC) to the anterior lateral part (e.g., the prefrontal cortex A46 & the middle temporal gyrus MTG) and then to the anterior medial part (e.g., the anterior cingulate cortex ACC & the lateral entorhinal cortex LEC; Fig. 4, C and D). Only one cell type, L6-CT, shows an exceptional regional pattern (Fig. 4, B and C) from the P–LA–MA axis. Nevertheless, the shared pattern allows us to merge the axes of different cell types into a consensus regional axis for cortical neurons (Fig. 4C; Methods).

Epigenetic features vary along the regional axis, which could involve the regional specification of brain cortices. For example, the transcription factor NR2F1, also known as COUP-TFI, is a crucial morphogen in brain development, which establishes the caudal-rostral neocortical regional specialization (41) and the boundary between the neocortex and the entorhinal cortex (47). NR2F1 has relatively high expression levels in VIC (P) and LEC (MA) while low in A46 (LA; Fig. S7A). The gene body CG- and CH-methylation levels show a negatively correlated pattern with its expression (Fig. 4G and Fig. S7B). Interestingly, NR2F1 is associated with two chromatin domains whose strengths change in opposite directions from VIC to A46 to LEC (Fig. 4F). The upstream domain shows stronger interactions in VIC, and DMRs within this domain are hypo-methylated in VIC compared to LEC. Contrarily, the downstream gene domain shows stronger interaction and harbors DMRs hypo-methylated in LEC (Fig. 4G, and Fig. S7, C and D). These coherent regional variations in epigenetic features and transcription activity imply a regulatory domain switching for NR2F1 from VIC to A46 to LEC, which needs further investigation. Systematic examination of the regionally

differential epigenetic features determined in total 14,606 (average 2.9k within each major type) from cortical neuronal

cell types, 885.4k (63.2k) regional DMRs (rDMRs; Fig. S7G; Methods), 773k (71.2k) regional differential loops and

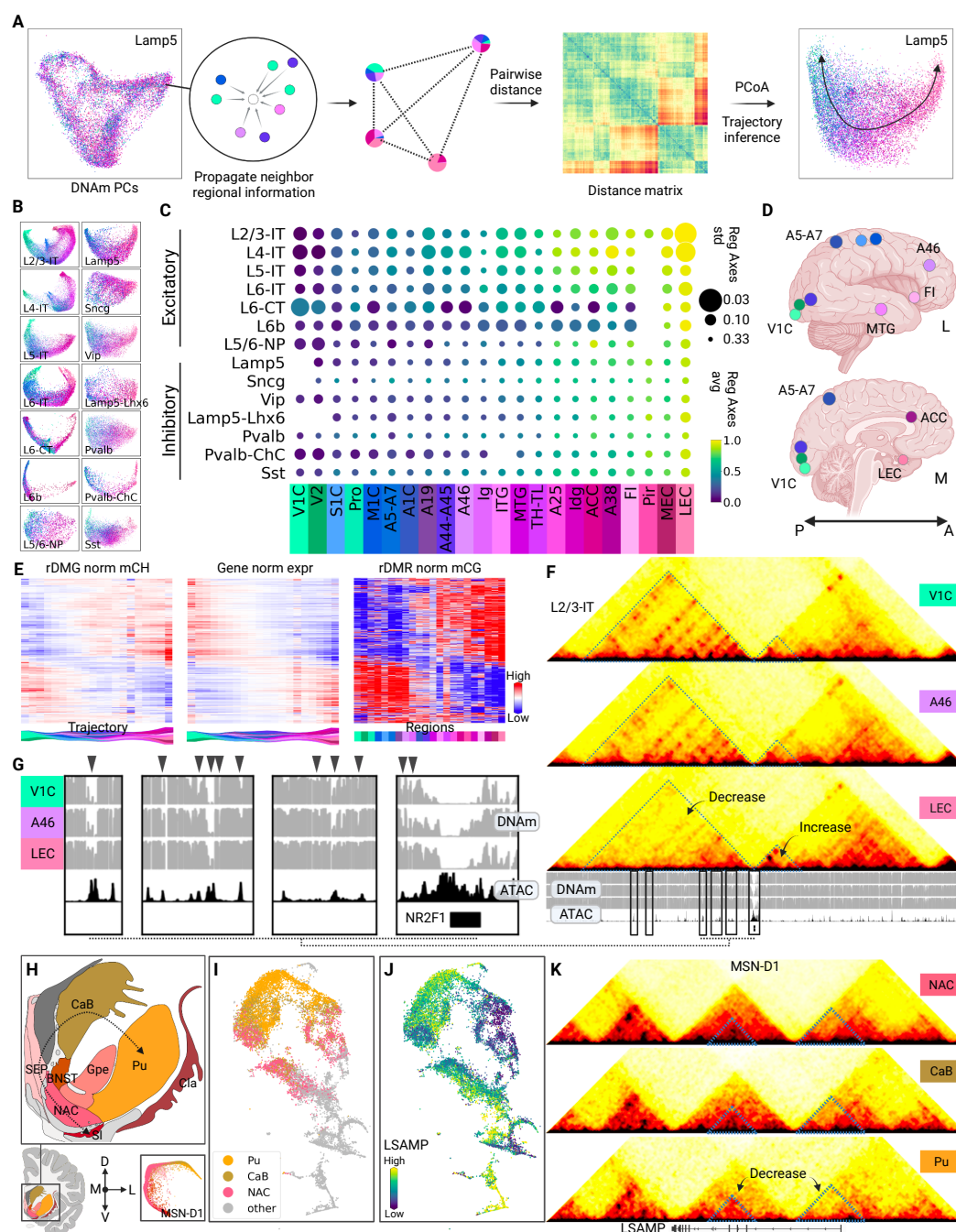


Figure 4. Regional axes of cortical and subcortical cells. (A) Workflow of determining regional axis from single nucleus DNA methylome. (B) 2D visualization of cortical neurons in regional spaces. The cortical neurons were colored by their dissection locations. (C) Trajectory analysis shows a common axis among cortical neurons. The scatter plot shows how regional indices vary in each cortical region. (D) Schematic of example cortical dissection locations. (E) Example of rDMGs in L2/3-IT cells. The rDMGs show gradient methylation levels along the axis we detected, which (anti-) correlate with the gene expression and the methylation levels of nearby rDMRs. (F) Chromatin conformation around the gene NR2F1 shows gradient changes in domain and loop strength. The two associated chromatin domains change in opposite directions. (G) Zoom-in view of example differential-loop-overlapping rDMRs from the "decreasing" domain. The methylation levels increase from V1C to A46 to LEC. (H) Inhibitory neurons in basal ganglia show an L-D-V axis in DNA methylation (I) 2D t-SNE visualization of MSN-D1. Cells from NAC, CaB and Pu are highlighted. (J) Gene body methylation of LSAMP in MSN-D1 shows the difference among cells from NAC, CaB and Pu. (K) Chromatin conformation around the gene LSAMP in MSN-D1. Cells from NAC and CaB show stronger chromatin interactions in LSAMP compared to Pu.

1,495 (136) regional differential domains (Fig. S7E). Many rDMGs and rDMRs show monotonic methylation gradients along the P–LA–MA axis (Fig. 4E, and Fig. S7, F, and G), while more complex patterns (such as NR2F1) also exist. These multiomic analyses provide an epigenetic basis for the regional diversity of gene expression in the brain cortex and allow the identification of key enhancers and regulatory programs in the specialization of cortical regions.

The inhibitory neurons in basal ganglia showed remarkable regional diversity as well. Four major cell types were identified in the regions of human basal ganglia, including D1- and D2-type medium spiny neurons (MSN-D1 and MSN-D2), and two cell types named *Foxp2* and *Chd7*, featuring CH-hypomethylation of the genes *FOXP2* and *CHD7* respectively. Our analysis reveals a profound molecular regional axis from lateral to dorsal to ventral parts (L–D–V) of the basal ganglia (Fig. 4H). We also found regional differences in gene regulation along this regional axis of basal ganglia. For example, the gene of Limbic System Associated Membrane Protein (*LSAMP*) shows a methylation gradient: CH methylation levels elevated from NAC to CaB to Pu (Fig. 4, I and J), accompanied by the decrease of strengths of chromatin domains and loops around this gene (Fig. 4K). In total, we determined 6,371 (2.7k) rDMGs and 398.8k (99.7k) rDMRs from all four cell types (Fig. S7E). Two cell types, MSN-D1 and MSN-D2, had enough m3C cells for regional differential analysis on chromatin conformation, in which we identified 98,276 (50,271) rDLs and 193 (99) rDDs (Fig. S7E). Most rDMGs and rDMRs (anti-) correlate strongly with the L–D–V axis (Fig. S7, H and I), indicating that regional diversity is a critical source of within-cell-type heterogeneity in the basal ganglia. As the major component of the basal ganglia, the striatum comprises the ventral part, including the nucleus accumbens (NAC), and the dorsal part, including caudate (CaB) and putamen (Pu). The two parts were known to differ in both functions and neural connections (48, 49). Our data and analysis provide the epigenetic basis of the D–V difference. In addition, we extended the D–V axis and revealed regional differences in a finer resolution within the dorsal part of the striatum (Fig. 4H).

Conservation of brain cell types and DMRs between humans and mice.

Overall conservation of cortical cell types between primates and rodents was noted in comparative single-cell transcriptomic studies of several cortical regions (50, 51). Whether the conservation of taxonomies holds in broader brain regions remains less studied. To address this question, we performed an integration analysis comparing the single nuclei DNA methylation profiles between human and mouse brains (1). Corresponding brain regions between the two species were used, including the cerebral cortex, basal forebrain, basal nuclei, and hippocampus (Methods). Two-

dimensional embedding of the integration shows a general matching between cell types of the two species (Fig. 5A and Fig. S8A). However, three major types defined from human brains are noticeably discrepant with mouse brain cells. Mouse L4-IT neurons aligned only to subpopulations of their human counterparts (Fig. 5B), suggesting a larger heterogeneity in human L4-IT neurons compared to mice and confirming findings from a previous comparative study between human MTG and mouse V1 (50). The human hippocampal major type HIP-Misc1 was integrated with a population of mouse cortical IT neurons instead of hippocampal neurons. Another human hippocampal major type HIP-Misc2 did not match any mouse brain cell type. Integration analysis against the parallel scRNA study (14) validates these two human hippocampal major types, which both have the signatures of DNA hypomethylation and RNA expression of the transcription factor *TSHZ2* and its overlapped long non-coding RNA (lncRNA) *AL109930.1* (Fig. 5C and Fig. S8B). Although two unmatched hippocampal major types will need further investigation, taxonomies at the major type level are generally conserved across broader brain regions between humans and mice. These matched cell types show consistently higher global levels in both CG- and CH-methylation in humans than in mice (Fig. 5D and Fig. S8C).

To compare the gene regulation between human and mouse brain cells, we first identified major type-specific hypodMRs within each species and used liftOver to match DMRs in corresponding major types between species (Fig. 5E). Then, we defined three hierarchical DMR categories based on direct and reciprocal matching: 1) OrthSeqs, which have ortholog sequences in the other species; 2) OrthDMRs, the OrthSeq whose ortholog sequences are also hypomethylated DMRs in the other species; and 3) CnsvDMRs, the orthDMRs that can be liftOvered reciprocally between the two species (Fig. 5E). In general, 40~60% DMRs are in the category of OrthSeqs across cell types, around half of which are OrthDMRs (Fig. 5F and Fig. S8D), and most (95%) of OrthDMRs are also CnsvDMRs (Fig. 5F and Fig. S8D). Furthermore, the methylation levels of CnsvDMRs showed a significant correlation between humans and mice across cell types (Fig. 5, G and H). We further compared the observed correlations against the cell-type-wise randomly shuffled background to select CnsvDMRs with significant correlation across species, which we called hcCnsvDMRs (Fig. 5G). Functional enrichment analysis of hcCnsvDMRs shows that they are highly enriched in biological processes related to forebrain development and in cellular components related to dendrites and synapses (Figs. S8, F and G; Methods). Comparison to histone modifications in mouse forebrains (6) demonstrated these DMRs are depleted from heterochromatic regions (H3K9me3) as well as enriched in regions of enhancers (H3K27ac & H3K4me1), promoters (H3K4me3), and poised enhancers (H3K27m3; Fig. S8E;

Methods). Categorizing the DMRs further into open or closed status based on their chromatin accessibility (2) shows that open DMRs are enriched in enhancers and promoters. In contrast, closed DMRs are particularly enriched in the poised enhancers (Fig. 5I), which were probably active during development.

The high conservation of DMR methylation levels between species points to a strategy of enhancer discovery through comparative epigenetics. For example, INPP5J, a gene expressed specifically in Pvalb cell type, has a number of distal and proximal hypomethylated hcCnsvDMRs, which overlap with matched chromatin-accessible regions as well (Fig. 5J). Two of these CREs were previously validated as

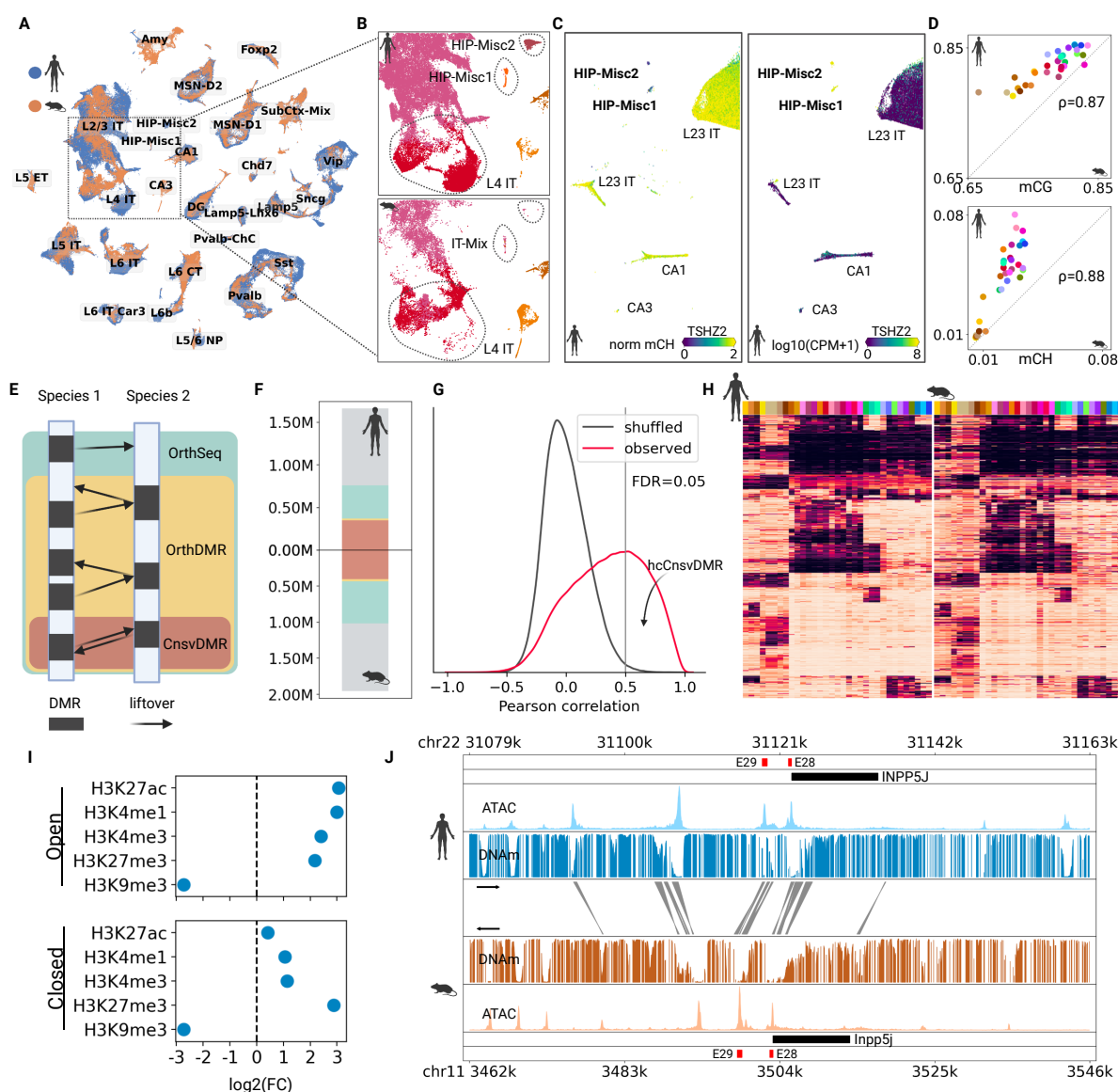


Figure 5. Cross-species comparison between human and mouse brain cell methylomes. (A) Integration of single-cell methylomes between human and mouse brains shows general cell type conservation across species. Neuronal cell types are visualized using 2D t-SNE. (B) Discrepancy between cell types of human and mouse brains in cell types L4-IT, HIP-Misc1, and HIP-Misc2. (C) Integration between human brain DNA methylation and transcriptomic datasets validates the cell types of HIP-Misc1 and HIP-Misc2, which both feature CH-hypomethylation and gene expression of TF TSHZ2. (D) Conserved cell types between human and mouse show correlated global CH- and CG-methylation levels. Methylation levels in human cells are higher than in mouse cells. (E) Schematic of cross-species matching of cell type DMRs. (F) Overall, ~50% of DMRs have orthologous sequences in the other species, among which ~25% are reciprocal DMRs. (G) Distribution of cross-species correlation of DMR methylations. DMR methylation in humans and in mice is highly correlated compared to the randomly shuffled background. (H) Examples of methylation levels of hcCnsvDMRs. (I) The hcCnsvDMRs show enrichment in the histone modification marks. The open hcCnsvDMRs are likely to be active enhancers and promoters while the closed ones are likely to be vestigial enhancers. (H) Browser view of hcCnsvDMRs around gene INPP5J in major type Pvalb. The regions colored by red are the cell type-specific distal enhancers validated in Ref (29).

specific enhancers that can be used in viral tools to target and manipulate Pvalb cells (Fig. 5J) (52).

Single-cell methylation barcodes (scMCodes) reliably predict human brain cell type identification.

Cytosine methylation variation in the genomes of cells contains molecular “engrams” in the form of hyper- and hypo-methylated regions that may represent past and present gene regulatory events (53). Methylation variation at single-nucleotide resolution has successfully been utilized for a wide array of applications, such as predicting epigenetic ages (54), tracing cell lineage (55, 56), and diagnosing life-threatening diseases like cancer from circulating free DNA (or cell-free DNA, cfDNA) (57, 58). We have shown that DNA methylation levels of gene bodies or 100kb genome bins can effectively distinguish brain cell types (1). In addition to these relatively long features, we observed many CpG sites with methylation patterns highly specific to certain cell types in the human brain. For example, the CpG site at chr1:46203766 is fully methylated in all cells other than THM-MB, while the CpG site at chr18:36273621 is unmethylated in all DG cells (Fig. S9B). With such high specificity at both single-cell and single CpG site levels, we derived single cell Methylation barCodes (scMCodes) for human brains, which determine brain cell types using the methylation states of a number of CpG sites at single-cell resolution (Fig. 6A and Fig. S9A; Methods). We iteratively selected cell type-specific and high-confidence CpG sites that distinguish brain cell types (Methods). The CpG sites were further clustered into 39k groups according to their across-cell-type methylation patterns and filtered by their cell-type predicting power through a Random Forest model (Fig. S9A; Methods). We selected 800 groups with a total of 12k CpG sites as the scMCode (Fig. 6B, and tables S7 and S8) to achieve high enough cell type predicting power (Fig. 6C) while maintaining a relatively small feature number (Fig. S9C). This scMCode achieved ~93% cell type prediction accuracy (Fig. 6D; Methods). To ensure the scMCode approach is robust, we conducted cross-donor tests by deriving scMCodes and training the predicting model using methylation data from one donor while testing cell type predictions using the others. These cross-donor predictions achieved high accuracies (92~93%) among all training-test pairs (Fig. 6E), demonstrating that the scMCode method is a robust and reliable predictor of cell type identity across individuals. Because single-cell sequencing has limited genomic coverage, not all CpG sites are expected to be covered in each single cell. Indeed, on average only ~200 CpG sites from different feature groups are actually used in each cell to achieve such high prediction accuracy (Fig. 6F), indicating the effectiveness of scMCode in determining human brain cell types from a few hundred select methylation sites.

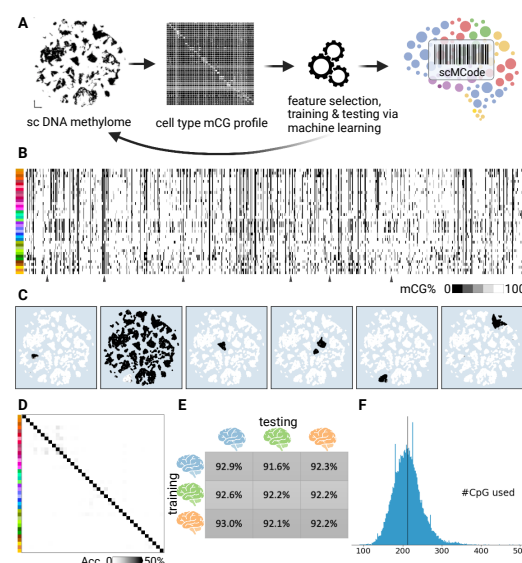


Figure 6. scMCodes for brain cell types. (A) Workflow of deriving scMCodes. (B) scMCodes derived from all three donors. (C) The methylation status of scMCode features is highly specific to cell types. (D) Heatmap of confusion matrix shows scMCodes can accurately predict cell types. (E) scMCodes are robust to individual differences in determining brain cell types. (F) scMCodes predict human cell types with a limited number of CpG sites at single-cell resolution.

Discussion

Knowledge of cell type diversity in the human brain and their specific gene regulatory programs is essential for understanding brain functions and developing therapeutics for brain disorders. Here we report comprehensive profiling of DNA methylation and chromatin conformation at single-cell resolution in the human brain, comprising 524,010 deep coverage nuclei from 46 cortical and subcortical brain regions. This dataset allows the identification of 188 epigenetically distinct cell types and the characterization of their molecular features, including marker genes, differentially methylated regions, signature transcription factors and DNA loops. Furthermore, integrative DNA methylation and chromatin conformation analysis can effectively distinguish the cell-type-specific candidate regulatory elements and their interaction with target genes. Moreover, the diversity of 3D genome structures was revealed at different scales and with unprecedented cell type resolution.

Taking advantage of the broad dissection coverage in brain sampling and in-depth epigenome sequencing coverage for each nucleus, we could delineate regional axes in both cortex and basal ganglia, which provides a novel approach for investigating brain region heterogeneity. Cross-species comparison between human and mouse brain cell methylomes reveals conservation in cell types and

regulatory elements, suggesting an enhancer discovery strategy via comparative epigenomics.

The rich and precise regulatory information stored in cytosine methylation allowed the derivation of scMCodes for human brain cells, providing a novel approach for reliable cell type identification using the methylation status of a small number of select CpG sites. Analysis of circulating-free DNA (cfDNA) methylation represents a non-invasive method for cancer diagnosis (58). cfDNA methylation has also been implicated as promising biomarkers for brain disorders (59). Our scMCode approach could be directly applied in profiling and analyzing cfDNA, which provides potential means for brain disorder diagnosis, pathological region identification, and treatment selection.

Overall, this multimodal human brain cell atlas complements transcriptomic studies of cell type diversity, regional heterogeneity, and evolution conservation with an underlying epigenetic basis, providing both a valuable resource for the neuroscience community to study gene regulation in brain cells and the raw material (enhancers) to develop new genetic tools for targeting specific cell types.

References and Notes

1. H. Liu, J. Zhou, W. Tian, C. Luo, A. Bartlett, A. Aldridge, J. Lucero, J. K. Osteen, J. R. Nery, H. Chen, A. Rivkin, R. G. Castanon, B. Clock, Y. E. Li, X. Hou, O. B. Poirion, S. Preissl, A. Pinto-Duarte, C. O'Connor, L. Boggeman, C. Fitzpatrick, M. Nunn, E. A. Mukamel, Z. Zhang, E. M. Callaway, B. Ren, J. R. Dixon, M. M. Behrens, J. R. Ecker, DNA methylation atlas of the mouse brain at single-cell resolution. *Nature*. **598**, 120–128 (2021).
2. Y. E. Li, S. Preissl, X. Hou, Z. Zhang, K. Zhang, Y. Qiu, O. B. Poirion, B. Li, J. Chiou, H. Liu, A. Pinto-Duarte, N. Kubo, X. Yang, R. Fang, X. Wang, J. Y. Han, J. Lucero, Y. Yan, M. Miller, S. Kuan, D. Gorkin, K. J. Gaulton, Y. Shen, M. Nunn, E. A. Mukamel, M. M. Behrens, J. R. Ecker, B. Ren, An atlas of gene regulatory elements in adult mouse cerebrum. *Nature*. **598**, 129–136 (2021).
3. D.-S. Lee, C. Luo, J. Zhou, S. Chandran, A. Rivkin, A. Bartlett, J. R. Nery, C. Fitzpatrick, C. O'Connor, J. R. Dixon, J. R. Ecker, Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods*. **16**, 999–1006 (2019).
4. A. de Mendoza, D. Poppe, S. Buckberry, J. Pflueger, C. B. Albertin, T. Daish, S. Bertrand, E. de la Calle-Mustienes, J. L. Gómez-Skarmeta, J. R. Nery, J. R. Ecker, B. Baer, C. W. Ragsdale, F. Grützner, H. Escriva, B. Venkatesh, O. Bogdanovic, R. Lister, The emergence of the brain non-CpG methylation system in vertebrates. *Nat Ecol Evol*. **5**, 369–378 (2021).
5. R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens, J. R. Ecker, Global epigenomic reconfiguration during mammalian brain development. *Science*. **341**, 1237905 (2013).
6. Y. He, M. Hariharan, D. U. Gorkin, D. E. Dickel, C. Luo, R. G. Castanon, J. R. Nery, A. Y. Lee, Y. Zhao, H. Huang, B. A. Williams, D. Trout, H. Amrhein, R. Fang, H. Chen, B. Li, A. Visel, L. A. Pennacchio, B. Ren, J. R. Ecker, Spatiotemporal DNA methylome dynamics of the developing mouse fetus. *Nature*. **583**, 752–759 (2020).
7. R. Tillotson, J. Cholewa-Waclaw, K. Chhatbar, J. C. Connelly, S. A. Kirschner, S. Webb, M. V. Koerner, J. Selfridge, D. A. Kelly, D. De Sousa, K. Brown, M. J. Lyst, S. Kriaucionis, A. Bird, Neuronal non-CG methylation is an essential target for MeCP2 function. *Mol. Cell*. **81**, 1260–1275.e12 (2021).
8. J. Dekker, L. Mirny, The 3D Genome as Moderator of Chromosomal Communication. *Cell*. **164**, 1110–1121 (2016).
9. BRAIN Initiative Cell Census Network (BICCN), A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature*. **598**, 86–102 (2021).
10. Z. Yao, H. Liu, F. Xie, S. Fischer, R. S. Adkins, A. I. Aldridge, S. A. Ament, A. Bartlett, M. M. Behrens, K. Van den Berge, D. Bertagnolli, H. R. de Bézieux, T. Biancalani, A. S. Boeshaghi, H. C. Bravo, T. Casper, C. Colantuoni, J. Crabtree, H. Creasy, K. Crichton, M. Crow, N. Dee, E. L. Dougherty, W. I. Doyle, S. Dudoit, R. Fang, V. Felix, O. Fong, M. Giglio, J. Goldy, M. Hawrylycz, B. R. Herb, R. Hertzano, X. Hou, Q. Hu, J. Kancharla, M. Kroll, K. Lathia, Y. E. Li, J. D. Lucero, C. Luo, A. Mahurkar, D. McMillen, N. M. Nadaf, J. R. Nery, T. N. Nguyen, S.-Y. Niu, V. Ntranos, J. Orvis, J. K. Osteen, T. Pham, A. Pinto-Duarte, O. Poirion, S. Preissl, E. Purdom, C. Rimorin, D. Risso, A. C. Rivkin, K. Smith, K. Street, J. Sulc, V. Svensson, M. Tieu, A. Torkelson, H. Tung, E. D. Vaishnav, C. R. Vanderburg, C. van Velthoven, X. Wang, O. R. White, Z. J. Huang, P. V. Kharchenko, L. Pachter, J.

- Ngai, A. Regev, B. Tasic, J. D. Welch, J. Gillis, E. Z. Macosko, B. Ren, J. R. Ecker, H. Zeng, E. A. Mukamel, A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature*. **598**, 103–110 (2021).
11. C. Luo, A. Rivkin, J. Zhou, J. P. Sandoval, L. Kurihara, J. Lucero, R. Castanon, J. R. Nery, A. Pinto-Duarte, B. Bui, C. Fitzpatrick, C. O'Connor, S. Ruga, M. E. Van Eden, D. A. Davis, D. C. Mash, M. M. Behrens, J. R. Ecker, Robust single-cell DNA methylome profiling with snmC-seq2. *Nat. Commun.* **9**, 3824 (2018).
12. J. A. Miller, N. W. Gouwens, B. Tasic, F. Collman, C. T. van Velthoven, T. E. Bakken, M. J. Hawrylycz, H. Zeng, E. S. Lein, A. Bernard, Common cell type nomenclature for the mammalian brain. *Elife*. **9** (2020), doi:10.7554/eLife.59928.
13. H. Peng, P. Xie, L. Liu, X. Kuang, Y. Wang, L. Qu, H. Gong, S. Jiang, A. Li, Z. Ruan, L. Ding, Z. Yao, C. Chen, M. Chen, T. L. Daigle, R. Dalley, Z. Ding, Y. Duan, A. Feiner, P. He, C. Hill, K. E. Hirokawa, G. Hong, L. Huang, S. Kebede, H.-C. Kuo, R. Larsen, P. Lesnar, L. Li, Q. Li, X. Li, Y. Li, Y. Li, A. Liu, D. Lu, S. Mok, L. Ng, T. N. Nguyen, Q. Ouyang, J. Pan, E. Shen, Y. Song, S. M. Sunkin, B. Tasic, M. B. Veldman, W. Wakeman, W. Wan, P. Wang, Q. Wang, T. Wang, Y. Wang, F. Xiong, W. Xiong, W. Xu, M. Ye, L. Yin, Y. Yu, J. Yuan, J. Yuan, Z. Yun, S. Zeng, S. Zhang, S. Zhao, Z. Zhao, Z. Zhou, Z. J. Huang, L. Esposito, M. J. Hawrylycz, S. A. Sorensen, X. W. Yang, Y. Zheng, Z. Gu, W. Xie, C. Koch, Q. Luo, J. A. Harris, Y. Wang, H. Zeng, Morphological diversity of single neurons in molecularly defined cell types. *Nature*. **598**, 174–181 (2021).
14. K. Siletti, R. Hodge, A. M. Albiach, L. Hu, K. W. Lee, P. Lönnerberg, T. Bakken, S.-L. Ding, M. Clark, T. Casper, N. Dee, J. Gloe, C. Dirk Keene, J. Nyhus, H. Tung, A. M. Yanny, E. Arenas, E. S. Lein, S. Linnarsson, Transcriptomic diversity of cell types across the adult human brain. *bioRxiv* (2022), p. 2022.10.12.511898.
15. J. Feng, Y. Zhou, S. L. Campbell, T. Le, E. Li, J. D. Sweatt, A. J. Silva, G. Fan, Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.* **13**, 423–430 (2010).
16. Y. He, J. R. Ecker, Non-CG Methylation in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **16**, 55–77 (2015).
17. J. Zhou, J. Ma, Y. Chen, C. Cheng, B. Bao, J. Peng, T. J. Sejnowski, J. R. Dixon, J. R. Ecker, Robust single-cell Hi-C clustering by convolution- and random-walk-based imputation. *Proceedings of the National Academy of Sciences*. **116**, 14011–14018 (2019).
18. M. Yu, B. Ren, The Three-Dimensional Organization of Mammalian Genomes. *Annu. Rev. Cell Dev. Biol.* **33**, 265–289 (2017).
19. T. Nagano, Y. Lubling, C. Várnai, C. Dudley, W. Leung, Y. Baran, N. Mendelson Cohen, S. Wingett, P. Fraser, A. Tanay, Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature*. **547**, 61–67 (2017).
20. J. D. Buenrostro, B. Wu, U. M. Litzenburger, D. Ruff, M. L. Gonzales, M. P. Snyder, H. Y. Chang, W. J. Greenleaf, Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*. **523**, 486–490 (2015).
21. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. **485**, 376–380 (2012).
22. S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. **159**, 1665–1680 (2014).
23. J. R. Dixon, I. Jung, S. Selvaraj, Y. Shen, J. E. Antosiewicz-Bourget, A. Y. Lee, Z. Ye, A. Kim, N. Rajagopal, W. Xie, Y. Diao, J. Liang, H. Zhao, V. V. Lobanenko, J. R. Ecker, J. A. Thomson, B. Ren, Chromatin architecture reorganization during stem cell differentiation. *Nature*. **518**, 331–336 (2015).
24. A. D. Schmitt, M. Hu, I. Jung, Z. Xu, Y. Qiu, C. L. Tan, Y. Li, S. Lin, Y. Lin, C. L. Barr, B. Ren, A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **17**, 2042–2059 (2016).
25. B. Bonev, N. Mendelson Cohen, Q. Szabo, L. Fritsch, G. L. Papadopoulos, Y. Lubling, X. Xu, X. Lv, J.-P. Hugnot, A. Tanay, G. Cavalli, Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*. **171**, 557–572.e24 (2017).

26. Y. Zhang, T. Li, S. Preissl, M. L. Amaral, J. D. Grinstein, E. N. Farah, E. Destici, Y. Qiu, R. Hu, A. Y. Lee, S. Chee, K. Ma, Z. Ye, Q. Zhu, H. Huang, R. Fang, L. Yu, J. C. Izpisua Belmonte, J. Wu, S. M. Evans, N. C. Chi, B. Ren, Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **51**, 1380–1388 (2019).
27. W. Winick-Ng, A. Kukalev, I. Harabula, L. Zea-Redondo, D. Szabó, M. Meijer, L. Serebreni, Y. Zhang, S. Bianco, A. M. Chiariello, I. Irastorza-Azcarate, C. J. Thieme, T. M. Sparks, S. Carvalho, L. Fiorillo, F. Musella, E. Irani, E. Torlai Triglia, A. A. Kolodziejczyk, A. Abentung, G. Apostolova, E. J. Paul, V. Franke, R. Kempfer, A. Akalin, S. A. Teichmann, G. Dechant, M. A. Ungless, M. Nicodemi, L. Welch, G. Castelo-Branco, A. Pombo, Cell-type specialization is encoded by specific chromatin topologies. *Nature*. **599**, 684–691 (2021).
28. C. Luo, C. L. Keown, L. Kurihara, J. Zhou, Y. He, J. Li, R. Castanon, J. Lucero, J. R. Nery, J. P. Sandoval, B. Bui, T. J. Sejnowski, T. T. Harkins, E. A. Mukamel, M. M. Behrens, J. R. Ecker, Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science*. **357**, 600–604 (2017).
29. A. Mo, E. A. Mukamel, F. P. Davis, C. Luo, G. L. Henry, S. Picard, M. A. Urich, J. R. Nery, T. J. Sejnowski, R. Lister, S. R. Eddy, J. R. Ecker, J. Nathans, Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron*. **86**, 1369–1384 (2015).
30. W. E. Heavner, S. Ji, J. H. Notwell, E. S. Dyer, A. M. Tseng, J. Birgmeier, B. Yoo, G. Bejerano, S. K. McConnell, Transcription factor expression defines subclasses of developing projection neurons highly similar to single-cell RNA-seq subtypes. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 25074–25084 (2020).
31. S. Warming, R. A. Rachel, N. A. Jenkins, N. G. Copeland, Zfp423 is required for normal cerebellar development. *Mol. Cell. Biol.* **26**, 6913–6922 (2006).
32. F. Casoni, L. Croci, C. Bosone, R. D'Ambrosio, A. Badaloni, D. Gaudesi, V. Barili, J. R. Sarna, L. Tessarollo, O. Cremona, R. Hawkes, S. Warming, G. Consalez, Zfp423/ZNF423 regulates cell cycle progression, the mode of cell division and the DNA-damage response in Purkinje neuron progenitors. *Development*. **144**, 3686–3697 (2017).
33. L. Croci, S.-H. Chung, G. Masserdotti, S. Gianola, A. Bizzoca, G. Gennarini, A. Corradi, F. Rossi, R. Hawkes, G. G. Consalez, A key role for the HLH transcription factor EBF2COE2, O/E-3 in Purkinje neuron migration and cerebellar cortical topography. *Development*. **133**, 2719–2729 (2006).
34. K. Takahashi, F.-C. Liu, T. Oishi, T. Mori, N. Higo, M. Hayashi, K. Hirokawa, H. Takahashi, Expression of FOXP2 in the developing monkey forebrain: comparison with the expression of the genes FOXP1, PBX3, and MEIS2. *J. Comp. Neurol.* **509**, 180–189 (2008).
35. T. Kaoru, F.-C. Liu, M. Ishida, T. Oishi, M. Hayashi, M. Kitagawa, K. Shimoda, H. Takahashi, Molecular characterization of the intercalated cell masses of the amygdala: implications for the relationship with the striatum. *Neuroscience*. **166**, 220–230 (2010).
36. R. Fernández-Chacón, A. Königstorfer, S. H. Gerber, J. García, M. F. Matos, C. F. Stevens, N. Brose, J. Rizo, C. Rosenmund, T. C. Südhof, Synaptotagmin I functions as a calcium regulator of release probability. *Nature*. **410**, 41–49 (2001).
37. J. Nasser, D. T. Bergman, C. P. Fulco, P. Guckelberger, B. R. Doughty, T. A. Patwardhan, T. R. Jones, T. H. Nguyen, J. C. Ulirsch, F. Lekschas, K. Mualim, H. M. Natri, E. M. Weeks, G. Munson, M. Kane, H. Y. Kang, A. Cui, J. P. Ray, T. M. Eisenhaure, R. L. Collins, K. Dey, H. Pfister, A. L. Price, C. B. Epstein, A. Kundaje, R. J. Xavier, M. J. Daly, H. Huang, H. K. Finucane, N. Hacohen, E. S. Lander, J. M. Engreitz, Genome-wide enhancer maps link risk variants to disease genes. *Nature*. **593**, 238–243 (2021).
38. B. K. Bulik-Sullivan, P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, N. Patterson, M. J. Daly, A. L. Price, B. M. Neale, LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
39. D. V. Hansen, J. E. Hanson, M. Sheng, Microglia in Alzheimer's disease. *J. Cell Biol.* **217**, 459–472 (2018).
40. M. Liu, Y. Jiang, R. Wedow, Y. Li, D. M. Brazel, F. Chen, G. Datta, J. Davila-Velderrain, D. McGuire, C. Tian, X. Zhan, 23andMe Research Team, HUNT All-In Psychiatry, H. Choquet, A. R. Docherty, J. D. Faul, J. R. Foerster, L. G. Fritsche, M. E. Gabrielsen,

- S. D. Gordon, J. Haessler, J.-J. Hottenga, H. Huang, S.-K. Jang, P. R. Jansen, Y. Ling, R. Mägi, N. Matoba, G. McMahon, A. Mulas, V. Orrù, T. Palviainen, A. Pandit, G. W. Reginsson, A. H. Skogholt, J. A. Smith, A. E. Taylor, C. Turman, G. Willemsen, H. Young, K. A. Young, G. J. M. Zajac, W. Zhao, W. Zhou, G. Bjornsdottir, J. D. Boardman, M. Boehnke, D. I. Boomsma, C. Chen, F. Cucca, G. E. Davies, C. B. Eaton, M. A. Ehringer, T. Esko, E. Fiorillo, N. A. Gillespie, D. F. Gudbjartsson, T. Haller, K. M. Harris, A. C. Heath, J. K. Hewitt, I. B. Hickie, J. E. Hokanson, C. J. Hopfer, D. J. Hunter, W. G. Iacono, E. O. Johnson, Y. Kamatani, S. L. R. Kardina, M. C. Keller, M. Kellis, C. Kooperberg, P. Kraft, K. S. Krauter, M. Laakso, P. A. Lind, A. Loukola, S. M. Lutz, P. A. F. Madden, N. G. Martin, M. McGue, M. B. McQueen, S. E. Medland, A. Metspalu, K. L. Mohlke, J. B. Nielsen, Y. Okada, U. Peters, T. J. C. Polderman, D. Posthuma, A. P. Reiner, J. P. Rice, E. Rimm, R. J. Rose, V. Runarsdottir, M. C. Stallings, A. Stančáková, H. Stefansson, K. K. Thai, H. A. Tindle, T. Tyrfinsson, T. L. Wall, D. R. Weir, C. Weisner, J. B. Whitfield, B. S. Winsvold, J. Yin, L. Zuccolo, L. J. Bierut, K. Hveem, J. J. Lee, M. R. Munafò, N. L. Saccone, C. J. Willer, M. C. Cornelis, S. P. David, D. A. Hinds, E. Jorgenson, J. Kaprio, J. A. Stitzel, K. Stefansson, T. E. Thorgeirsson, G. Abecasis, D. J. Liu, S. Vrieze, Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).
41. S. N. Sansom, F. J. Livesey, Gradients in the brain: the control of the development of form and function in the cerebral cortex. *Cold Spring Harb. Perspect. Biol.* **1**, a002519 (2009).
42. M. J. Hawrylycz, E. S. Lein, A. L. Guillozet-Bongaarts, E. H. Shen, L. Ng, J. A. Miller, L. N. van de Lagemaat, K. A. Smith, A. Ebbert, Z. L. Riley, C. Abajian, C. F. Beckmann, A. Bernard, D. Bertagnolli, A. F. Boe, P. M. Cartagena, M. M. Chakravarty, M. Chapin, J. Chong, R. A. Dalley, B. David Daly, C. Dang, S. Datta, N. Dee, T. A. Dolbeare, V. Faber, D. Feng, D. R. Fowler, J. Goldy, B. W. Gregor, Z. Haradon, D. R. Haynor, J. G. Hohmann, S. Horvath, R. E. Howard, A. Jeromin, J. M. Jochim, M. Kinnunen, C. Lau, E. T. Lazarz, C. Lee, T. A. Lemon, L. Li, Y. Li, J. A. Morris, C. C. Overly, P. D. Parker, S. E. Parry, M. Reding, J. J. Royall, J. Schulkin, P. A. Sequeira, C. R. Slaughterbeck, S. C. Smith, A. J. Sodt, S. M. Sunkin, B. E. Swanson, M. P. Wawter, D. Williams, P. Wohnoutka, H. R. Zielke, D. H. Geschwind, P. R. Hof, S. M. Smith, C. Koch, S. G. N. Grant, A. R. Jones, An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. **489**, 391–399 (2012).
43. Z. Yao, C. T. J. van Velthoven, T. N. Nguyen, J. Goldy, A. E. Sedeno-Cortes, F. Baftizadeh, D. Bertagnolli, T. Casper, M. Chiang, K. Crichton, S.-L. Ding, O. Fong, E. Garren, A. Glandon, N. W. Gouwens, J. Gray, L. T. Graybuck, M. J. Hawrylycz, D. Hirschstein, M. Kroll, K. Lathia, C. Lee, B. Levi, D. McMillen, S. Mok, T. Pham, Q. Ren, C. Rimorin, N. Shapovalova, J. Sulc, S. M. Sunkin, M. Tieu, A. Torkelson, H. Tung, K. Ward, N. Dee, K. A. Smith, B. Tasic, H. Zeng, A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell*. **184**, 3222–3241.e26 (2021).
44. F. M. Krienen, M. Goldman, Q. Zhang, R. C. H. Del Rosario, M. Florio, R. Machold, A. Saunders, K. Levandowski, H. Zaniewski, B. Schuman, C. Wu, A. Lutservitz, C. D. Mullally, N. Reed, E. Bien, L. Bortolin, M. Fernandez-Otero, J. D. Lin, A. Wysoker, J. Nemesh, D. Kulp, M. Burns, V. Tkachev, R. Smith, C. A. Walsh, J. Dimidschstein, B. Rudy, L. S. Kean, S. Berretta, G. Fishell, G. Feng, S. A. McCarroll, Innovations present in the primate interneuron repertoire. *Nature*. **586**, 262–269 (2020).
45. B. Tasic, Z. Yao, L. T. Graybuck, K. A. Smith, T. N. Nguyen, D. Bertagnolli, J. Goldy, E. Garren, M. N. Economo, S. Viswanathan, O. Penn, T. Bakken, V. Menon, J. Miller, O. Fong, K. E. Hirokawa, K. Lathia, C. Rimorin, M. Tieu, R. Larsen, T. Casper, E. Barkan, M. Kroll, S. Parry, N. V. Shapovalova, D. Hirschstein, J. Pendergraft, H. A. Sullivan, T. K. Kim, A. Szafer, N. Dee, P. Groblewski, I. Wickersham, A. Cetin, J. A. Harris, B. P. Levi, S. M. Sunkin, L. Madisen, T. L. Daigle, L. Looger, A. Bernard, J. Phillips, E. Lein, M. Hawrylycz, K. Svoboda, A. R. Jones, C. Koch, H. Zeng, Shared and distinct transcriptomic cell types across neocortical areas. *Nature*. **563**, 72–78 (2018).
46. H. Chen, L. Albergante, J. Y. Hsu, C. A. Lareau, G. Lo Bosco, J. Guan, S. Zhou, A. N. Gorban, D. E. Bauer, M. J. Aryee, D. M. Langenau, A. Zinovyev, J. D. Buenrostro, G.-C. Yuan, L. Pinello, Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nature Communications*. **10** (2019), doi:10.1038/s41467-019-09670-4.

47. J. Feng, W.-H. Hsu, D. Patterson, C.-S. Tseng, H.-W. Hsing, Z.-H. Zhuang, Y.-T. Huang, A. Faedo, J. L. Rubenstein, J. Touboul, S.-J. Chou, COUP-TFI specifies the medial entorhinal cortex identity and induces differential cell adhesion to determine the integrity of its boundary with neocortex. *Sci Adv.* **7** (2021), doi:10.1126/sciadv.abf6808.
48. G. E. Alexander, M. R. DeLong, P. L. Strick, Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* **9**, 357–381 (1986).
49. F. A. Middleton, P. L. Strick, Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Res. Brain Res. Rev.* **31**, 236–250 (2000).
50. R. D. Hodge, T. E. Bakken, J. A. Miller, K. A. Smith, E. R. Barkan, L. T. Graybuck, J. L. Close, B. Long, N. Johansen, O. Penn, Z. Yao, J. Eggermont, T. Höllt, B. P. Levi, S. I. Shehata, B. Aeversmann, A. Beller, D. Bertagnolli, K. Brouner, T. Casper, C. Cobbs, R. Dalley, N. Dee, S.-L. Ding, R. G. Ellenbogen, O. Fong, E. Garren, J. Goldy, R. P. Gwinn, D. Hirschstein, C. D. Keene, M. Keshk, A. L. Ko, K. Lathia, A. Mahfouz, Z. Maltzer, M. McGraw, T. N. Nguyen, J. Nyhus, J. G. Ojemann, A. Oldre, S. Parry, S. Reynolds, C. Rimorin, N. V. Shapovalova, S. Somasundaram, A. Szafer, E. R. Thomsen, M. Tieu, G. Quon, R. H. Scheuermann, R. Yuste, S. M. Sunkin, B. Lelieveldt, D. Feng, L. Ng, A. Bernard, M. Hawrylycz, J. W. Phillips, B. Tasic, H. Zeng, A. R. Jones, C. Koch, E. S. Lein, Conserved cell types with divergent features in human versus mouse cortex. *Nature.* **573**, 61–68 (2019).
51. T. E. Bakken, N. L. Jorstad, Q. Hu, B. B. Lake, W. Tian, B. E. Kalmbach, M. Crow, R. D. Hodge, F. M. Krienen, S. A. Sorensen, J. Eggermont, Z. Yao, B. D. Aeversmann, A. I. Aldridge, A. Bartlett, D. Bertagnolli, T. Casper, R. G. Castanon, K. Crichton, T. L. Daigle, R. Dalley, N. Dee, N. Dembrow, D. Diep, S.-L. Ding, W. Dong, R. Fang, S. Fischer, M. Goldman, J. Goldy, L. T. Graybuck, B. R. Herb, X. Hou, J. Kancherla, M. Kroll, K. Lathia, B. van Lew, Y. E. Li, C. S. Liu, H. Liu, J. D. Lucero, A. Mahurkar, D. McMillen, J. A. Miller, M. Moussa, J. R. Nery, P. R. Nicovich, S.-Y. Niu, J. Orvis, J. K. Osteen, S. Owen, C. R. Palmer, T. Pham, N. Plongthongkum, O. Poirion, N. M. Reed, C. Rimorin, A. Rivkin, W. J. Romanow, A. E. Sedeño-Cortés, K. Siletti, S. Somasundaram, J. Sulc, M. Tieu, A. Torkelson, H. Tung, X. Wang, F. Xie, A. M. Yanny, R. Zhang, S. A. Ament, M. M. Behrens, H. C. Bravo, J. Chun, A. Dobin, J. Gillis, R. Hertzano, P. R. Hof, T. Höllt, G. D. Horwitz, C. D. Keene, P. V. Kharchenko, A. L. Ko, B. P. Lelieveldt, C. Luo, E. A. Mukamel, A. Pinto-Duarte, S. Preissl, A. Regev, B. Ren, R. H. Scheuermann, K. Smith, W. J. Spain, O. R. White, C. Koch, M. Hawrylycz, B. Tasic, E. Z. Macosko, S. A. McCarroll, J. T. Ting, H. Zeng, K. Zhang, G. Feng, J. R. Ecker, S. Linnarsson, E. S. Lein, Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature.* **598**, 111–119 (2021).
52. D. Vormstein-Schneider, J. D. Lin, K. A. Pelkey, R. Chittajallu, B. Guo, M. A. Arias-Garcia, K. Allaway, S. Sakopoulos, G. Schneider, O. Stevenson, J. Vergara, J. Sharma, Q. Zhang, T. P. Franken, J. Smith, L. A. Ibrahim, K. J. Mastro, E. Sabri, S. Huang, E. Favuzzi, T. Burbridge, Q. Xu, L. Guo, I. Vogel, V. Sanchez, G. A. Saldi, B. L. Gorissen, X. Yuan, K. A. Zaghoul, O. Devinsky, B. L. Sabatini, R. Batista-Brito, J. Reynolds, G. Feng, Z. Fu, C. J. McBain, G. Fishell, J. Dimidschstein, Viral manipulation of functionally distinct interneurons in mice, non-human primates and humans. *Nat. Neurosci.* **23**, 1629–1636 (2020).
53. G. C. Hon, N. Rajagopal, Y. Shen, D. F. McCleary, F. Yue, M. D. Dang, B. Ren, Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* **45**, 1198–1206 (2013).
54. S. Horvath, Y. Zhang, P. Langfelder, R. S. Kahn, M. P. M. Boks, K. van Eijk, L. H. van den Berg, R. A. Ophoff, Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* **13**, R97 (2012).
55. C. Gabbutt, R. O. Schenck, D. J. Weisenberger, C. Kimberley, A. Berner, J. Househam, E. Lakatos, M. Robertson-Tessi, I. Martin, R. Patel, S. K. Clark, A. Latchford, C. P. Barnes, S. J. Leedham, A. R. A. Anderson, T. A. Graham, D. Shibata, Fluctuating methylation clocks for cell lineage tracing at high temporal resolution in human tissues. *Nat. Biotechnol.* **40**, 720–730 (2022).
56. L. A. Salas, J. K. Wiencke, D. C. Koestler, Z. Zhang, B. C. Christensen, K. T. Kelsey, Tracing human stem cell lineage during development using DNA methylation. *Genome Res.* **28**, 1285–1295 (2018).
57. E. Gormally, E. Caboux, P. Vineis, P. Hainaut, Circulating free DNA in plasma or serum as biomarker of carcinogenesis: practical aspects and

- biological significance. *Mutat. Res.* **635**, 105–117 (2007).
58. M. Cisneros-Villanueva, L. Hidalgo-Pérez, M. Rios-Romero, A. Cedro-Tanda, C. A. Ruiz-Villavicencio, K. Page, R. Hastings, D. Fernandez-Garcia, R. Allsopp, M. A. Fonseca-Montañó, S. Jimenez-Morales, V. Padilla-Palma, J. A. Shaw, A. Hidalgo-Miranda, Cell-free DNA analysis in current cancer clinical trials: a review. *Br. J. Cancer.* **126**, 391–400 (2022).
59. Y. Jin, E. G. Allen, P. Jin, Cell-free DNA methylation as a potential biomarker in brain disorders. *Epigenomics.* **14**, 369–374 (2022).
60. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* **34**, i884–i890 (2018).
61. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013), (available at <http://arxiv.org/abs/1303.3997>).
62. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *Gigascience.* **10** (2021), doi:10.1093/gigascience/giab008.
63. R. Poplin, V. Ruano-Rubio, M. A. DePristo, T. J. Fennell, M. O. Carneiro, G. A. Van der Auwera, D. E. Kling, L. D. Gauthier, A. Levy-Moonshine, D. Roazen, K. Shakir, J. Thibault, S. Chandran, C. Whelan, M. Lek, S. Gabriel, M. J. Daly, B. Neale, D. G. MacArthur, E. Banks, Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* (2018), p. 201178.
64. H. M. Amemiya, A. Kundaje, A. P. Boyle, The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).
65. A. Zeisel, H. Hochgerner, P. Lönnerberg, A. Johnsson, F. Memic, J. van der Zwan, M. Häring, E. Braun, L. E. Borm, G. La Manno, S. Codeluppi, A. Furlan, K. Lee, N. Skene, K. D. Harris, J. Hjerling-Leffler, E. Arenas, P. Ernfors, U. Marklund, S. Linnarsson, Molecular Architecture of the Mouse Nervous System. *Cell.* **174**, 999–1014.e22 (2018).
66. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods.* **16**, 1289–1296 (2019).
67. B. Hie, B. Bryson, B. Berger, Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691 (2019).
68. M. D. Schultz, Y. He, J. W. Whitaker, M. Hariharan, E. A. Mukamel, D. Leung, N. Rajagopal, J. R. Nery, M. A. Urich, H. Chen, S. Lin, Y. Lin, I. Jung, A. D. Schmitt, S. Selvaraj, B. Ren, T. J. Sejnowski, W. Wang, J. R. Ecker, Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature.* **523**, 212–216 (2015).
69. O. Fornes, J. A. Castro-Mondragon, A. Khan, R. van der Lee, X. Zhang, P. A. Richmond, B. P. Modi, S. Correard, M. Gheorghe, D. Baranašić, W. Santana-Garcia, G. Tan, J. Chèneby, B. Ballester, F. Parcy, A. Sandelin, B. Lenhard, W. W. Wasserman, A. Mathelier, JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
70. D. Kobak, P. Berens, The art of using t-SNE for single-cell transcriptomics. *Nat. Commun.* **10**, 1–14 (2019).
71. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* **326**, 289–293 (2009).
72. E. P. Nora, A. Goloborodko, A.-L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, B. G. Bruneau, Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell.* **169**, 930–944.e22 (2017).
73. M. Yu, A. Abnoui, Y. Zhang, G. Li, L. Lee, Z. Chen, R. Fang, T. M. Lagler, Y. Yang, J. Wen, Q. Sun, Y. Li, B. Ren, M. Hu, SnapHiC: a computational pipeline to identify chromatin loops from single-cell Hi-C data. *Nat. Methods.* **18**, 1056–1059 (2021).
74. G. Sabarís, I. Laiker, E. Preger-Ben Noon, N. Frankel, Actors with Multiple Roles: Pleiotropic Enhancers and the Paradigm of Enhancer Modularity. *Trends Genet.* **35**, 423–433 (2019).

75. J. E. Savage, P. R. Jansen, S. Stringer, K. Watanabe, J. Bryois, C. A. de Leeuw, M. Nagel, S. Awasthi, P. B. Barr, J. R. I. Coleman, K. L. Grasby, A. R. Hammerschlag, J. A. Kaminski, R. Karlsson, E. Krapohl, M. Lam, M. Nygaard, C. A. Reynolds, J. W. Trampush, H. Young, D. Zabaneh, S. Hägg, N. K. Hansell, I. K. Karlsson, S. Linnarsson, G. W. Montgomery, A. B. Muñoz-Manchado, E. B. Quinlan, G. Schumann, N. G. Skene, B. T. Webb, T. White, D. E. Arking, D. Avramopoulos, R. M. Bilder, P. Bitsios, K. E. Burdick, T. D. Cannon, O. Chiba-Falek, A. Christoforou, E. T. Cirulli, E. Congdon, A. Corvin, G. Davies, I. J. Deary, P. DeRosse, D. Dickinson, S. Djurovic, G. Donohoe, E. D. Conley, J. G. Eriksson, T. Espeseth, N. A. Freimer, S. Giakoumaki, I. Giegling, M. Gill, D. C. Glahn, A. R. Hariri, A. Hatzimanolis, M. C. Keller, E. Knowles, D. Koltai, B. Konte, J. Lahti, S. Le Hellard, T. Lencz, D. C. Liewald, E. London, A. J. Lundervold, A. K. Malhotra, I. Melle, D. Morris, A. C. Need, W. Ollier, A. Palotie, A. Payton, N. Pendleton, R. A. Poldrack, K. Räikkönen, I. Reinvang, P. Roussos, D. Rujescu, F. W. Sabb, M. A. Scult, O. B. Smeland, N. Smyrnis, J. M. Starr, V. M. Steen, N. C. Stefanis, R. E. Straub, K. Sundet, H. Tiemeier, A. N. Voineskos, D. R. Weinberger, E. Widen, J. Yu, G. Abecasis, O. A. Andreassen, G. Breen, L. Christiansen, B. Debrabant, D. M. Dick, A. Heinz, J. Hjerling-Leffler, M. A. Ikram, K. S. Kendler, N. G. Martin, S. E. Medland, N. L. Pedersen, R. Plomin, T. J. C. Polderman, S. Ripke, S. van der Sluis, P. F. Sullivan, S. I. Vrieze, M. J. Wright, D. Posthuma, Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).
76. A. Okbay, J. P. Beauchamp, M. A. Fontana, J. J. Lee, T. H. Pers, C. A. Rietveld, P. Turley, G.-B. Chen, V. Emilsson, S. F. W. Meddens, S. Oskarsson, J. K. Pickrell, K. Thom, P. Timshel, R. de Vlaming, A. Abdellaoui, T. S. Ahluwalia, J. Bacelis, C. Baumbach, G. Bjornsdottir, J. H. Brandsma, M. Pina Concas, J. Derringer, N. A. Furlotte, T. E. Galesloot, G. Girotto, R. Gupta, L. M. Hall, S. E. Harris, E. Hofer, M. Horikoshi, J. E. Huffman, K. Kaasik, I. P. Kalafati, R. Karlsson, A. Kong, J. Lahti, S. J. van der Lee, C. deLeeuw, P. A. Lind, K.-O. Lindgren, T. Liu, M. Mangino, J. Marten, E. Mihailov, M. B. Miller, P. J. van der Most, C. Oldmeadow, A. Payton, N. Pervjakova, W. J. Peyrot, Y. Qian, O. Raitakari, R. Rueedi, E. Salvi, B. Schmidt, K. E. Schraut, J. Shi, A. V. Smith, R. A. Poot, B. St Pourcain, A. Teumer, G. Thorleifsson, N. Verweij, D. Vuckovic, J. Wellmann, H.-J. Westra, J. Yang, W. Zhao, Z. Zhu, B. Z. Alizadeh, N. Amin, A. Bakshi, S. E. Baumeister, G. Biino, K. Bønnelykke, P. A. Boyle, H. Campbell, F. P. Cappuccio, G. Davies, J.-E. De Neve, P. Deloukas, I. Demuth, J. Ding, P. Eibich, L. Eisele, N. Eklund, D. M. Evans, J. D. Faul, M. F. Feitosa, A. J. Forstner, I. Gandin, B. Gunnarsson, B. V. Halldórsson, T. B. Harris, A. C. Heath, L. J. Hocking, E. G. Holliday, G. Homuth, M. A. Horan, J.-J. Hottenga, P. L. de Jager, P. K. Joshi, A. Jugessur, M. A. Kaakinen, M. Kähönen, S. Kanoni, L. Keltigangas-Järvinen, L. A. L. M. Kiemeny, I. Kolcic, S. Koskinen, A. T. Kraja, M. Kroh, Z. Kutalik, A. Latvala, L. J. Launer, M. P. Lebreton, D. F. Levinson, P. Lichtenstein, P. Lichtner, D. C. M. Liewald, LifeLines Cohort Study, A. Loukola, P. A. Madden, R. Mägi, T. Mäki-Opas, R. E. Marioni, P. Marques-Vidal, G. A. Meddens, G. McMahon, C. Meisinger, T. Meitinger, Y. Milaneschi, L. Milani, G. W. Montgomery, R. Myhre, C. P. Nelson, D. R. Nyholt, W. E. R. Ollier, A. Palotie, L. Paternoster, N. L. Pedersen, K. E. Petrovic, D. J. Porteous, K. Räikkönen, S. M. Ring, A. Robino, O. Rostapshova, I. Rudan, A. Rustichini, V. Salomaa, A. R. Sanders, A.-P. Sarin, H. Schmidt, R. J. Scott, B. H. Smith, J. A. Smith, J. A. Staessen, E. Steinhausen-Thiessen, K. Strauch, A. Terracciano, M. D. Tobin, S. Ulivi, S. Vaccargiu, L. Quaye, F. J. A. van Rooij, C. Venturini, A. A. E. Vinkhuyzen, U. Völker, H. Völzke, J. M. Vonk, D. Vozzi, J. Waage, E. B. Ware, G. Willemsen, J. R. Attia, D. A. Bennett, K. Berger, L. Bertram, H. Bisgaard, D. I. Boomsma, I. B. Borecki, U. Bültmann, C. F. Chabris, F. Cucca, D. Cusi, I. J. Deary, G. V. Dedoussis, C. M. van Duijn, J. G. Eriksson, B. Franke, L. Franke, P. Gasparini, P. V. Gejman, C. Gieger, H.-J. Grabe, J. Gratten, P. J. F. Groenen, V. Gudnason, P. van der Harst, C. Hayward, D. A. Hinds, W. Hoffmann, E. Hyppönen, W. G. Iacono, B. Jacobsson, M.-R. Järvelin, K.-H. Jöckel, J. Kaprio, S. L. R. Kardia, T. Lehtimäki, S. F. Lehrer, P. K. E. Magnusson, N. G. Martin, M. McGue, A. Metspalu, N. Pendleton, B. W. J. H. Penninx, M. Perola, N. Pirastu, M. Pirastu, O. Polasek, D. Posthuma, C. Power, M. A. Province, N. J. Samani, D. Schlessinger, R. Schmidt, T. I. A. Sørensen, T. D. Spector, K. Stefansson, U. Thorsteinsdottir, A. R. Thurik, N. J. Timpson, H. Tiemeier, J. Y. Tung, A. G. Uitterlinden, V. Vitart, P. Vollenweider, D. R. Weir, J. F. Wilson, A. F. Wright, D. C. Conley, R. F. Krueger, G. Davey Smith, A. Hofman, D. I. Laibson, S. E. Medland, M. N. Meyer, J. Yang, M. Johannesson, P. M. Visscher, T. Esko, P. D. Koellinger, D. Cesarini, D. J. Benjamin, Genome-wide association study identifies 74 loci associated

- with educational attainment. *Nature*. **533**, 539–542 (2016).
77. G. Schumann, C. Liu, P. O'Reilly, H. Gao, P. Song, B. Xu, B. Ruggeri, N. Amin, T. Jia, S. Preis, M. Segura Lepe, S. Akira, C. Barbieri, S. Baumeister, S. Cauchi, T.-K. Clarke, S. Enroth, K. Fischer, J. Hällfors, S. E. Harris, S. Hieber, E. Hofer, J.-J. Hottenga, Å. Johansson, P. K. Joshi, N. Kaartinen, J. Laitinen, R. Lemaitre, A. Loukola, J. 'an Luan, L.-P. Lyytikäinen, M. Mangino, A. Manichaikul, H. Mbarek, Y. Milaneschi, A. Moayyeri, K. Mukamal, C. Nelson, J. Nettleton, E. Partinen, R. Rawal, A. Robino, L. Rose, C. Sala, T. Satoh, R. Schmidt, K. Schraut, R. Scott, A. V. Smith, J. M. Starr, A. Teumer, S. Trompet, A. G. Uitterlinden, C. Venturini, A.-C. Vergnaud, N. Verweij, V. Vitart, D. Vuckovic, J. Wedenoja, L. Yengo, B. Yu, W. Zhang, J. H. Zhao, D. I. Boomsma, J. Chambers, D. I. Chasman, T. Daniela, E. de Geus, I. Deary, J. G. Eriksson, T. Esko, V. Eulenburg, O. H. Franco, P. Froguel, C. Gieger, H. J. Grabe, V. Gudnason, U. Gyllenstein, T. B. Harris, A.-L. Hartikainen, A. C. Heath, L. Hocking, A. Hofman, C. Huth, M.-R. Jarvelin, J. W. Jukema, J. Kaprio, J. S. Kooner, Z. Kutalik, J. Lahti, C. Langenberg, T. Lehtimäki, Y. Liu, P. A. F. Madden, N. Martin, A. Morrison, B. Penninx, N. Pirastu, B. Psaty, O. Raitakari, P. Ridker, R. Rose, J. I. Rotter, N. J. Samani, H. Schmidt, T. D. Spector, D. Stott, D. Strachan, I. Tzoulaki, P. van der Harst, C. M. van Duijn, P. Marques-Vidal, P. Vollenweider, N. J. Wareham, J. B. Whitfield, J. Wilson, B. Wolffenbuttel, G. Bakalkin, E. Evangelou, Y. Liu, K. M. Rice, S. Desrivieres, S. A. Kliewer, D. J. Mangelsdorf, C. P. Müller, D. Levy, P. Elliott, KLB is associated with alcohol drinking, and its gene product β -Klotho is necessary for FGF21 regulation of alcohol preference. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14372–14377 (2016).
78. I. E. Jansen, J. E. Savage, K. Watanabe, J. Bryois, D. M. Williams, S. Steinberg, J. Sealock, I. K. Karlsson, S. Hägg, L. Athanasiu, N. Voyle, P. Proitsi, A. Witoelar, S. Stringer, D. Aarsland, I. S. Almdahl, F. Andersen, S. Bergh, F. Bettella, S. Bjornsson, A. Brækhus, G. Bråthen, C. de Leeuw, R. S. Desikan, S. Djurovic, L. Dumitrescu, T. Fladby, T. J. Hohman, P. V. Jonsson, S. J. Kiddle, A. Rongve, I. Saltvedt, S. B. Sando, G. Selbæk, M. Shuai, N. G. Skene, J. Snaedal, E. Stordal, I. D. Ulstein, Y. Wang, L. R. White, J. Hardy, J. Hjerling-Leffler, P. F. Sullivan, W. M. van der Flier, R. Dobson, L. K. Davis, H. Stefansson, K. Stefansson, N. L. Pedersen, S. Ripke, O. A. Andreassen, D. Posthuma, Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
79. E. A. Stahl, G. Breen, A. J. Forstner, A. McQuillin, S. Ripke, V. Trubetskoy, M. Mattheisen, Y. Wang, J. R. I. Coleman, H. A. Gaspar, C. A. de Leeuw, S. Steinberg, J. M. W. Pavlides, M. Trzaskowski, E. M. Byrne, T. H. Pers, P. A. Holmans, A. L. Richards, L. Abbott, E. Agerbo, H. Akil, D. Albani, N. Alliey-Rodriguez, T. D. Als, A. Anjorin, V. Antilla, S. Awasthi, J. A. Badner, M. Bækvad-Hansen, J. D. Barchas, N. Bass, M. Bauer, R. Belliveau, S. E. Bergen, C. B. Pedersen, E. Bøen, M. P. Boks, J. Boocock, M. Budde, W. Bunney, M. Burmeister, J. Bybjerg-Grauholm, W. Byerley, M. Casas, F. Cerrato, P. Cervantes, K. Chambert, A. W. Charney, D. Chen, C. Churchhouse, T.-K. Clarke, W. Coryell, D. W. Craig, C. Cruceanu, D. Curtis, P. M. Czerski, A. M. Dale, S. de Jong, F. Degenhardt, J. Del-Favero, J. R. DePaulo, S. Djurovic, A. L. Dobbyn, A. Dumont, T. Elvsåshagen, V. Escott-Price, C. C. Fan, S. B. Fischer, M. Flickinger, T. M. Foroud, L. Forty, J. Frank, C. Fraser, N. B. Freimer, L. Frisén, K. Gade, D. Gage, J. Garnham, C. Giambartolomei, M. G. Pedersen, J. Goldstein, S. D. Gordon, K. Gordon-Smith, E. K. Green, M. J. Green, T. A. Greenwood, J. Grove, W. Guan, J. Guzman-Parra, M. L. Hamshere, M. Hautzinger, U. Heilbronner, S. Herms, M. Hipolito, P. Hoffmann, D. Holland, L. Huckins, S. Jamain, J. S. Johnson, A. Juréus, R. Kandaswamy, R. Karlsson, J. L. Kennedy, S. Kittel-Schneider, J. A. Knowles, M. Kogevinas, A. C. Koller, R. Kupka, C. Lavebratt, J. Lawrence, W. B. Lawson, M. Leber, P. H. Lee, S. E. Levy, J. Z. Li, C. Liu, S. Lucae, A. Maaser, D. J. MacIntyre, P. B. Mahon, W. Maier, L. Martinsson, S. McCarroll, P. McGuffin, M. G. McInnis, J. D. McKay, H. Medeiros, S. E. Medland, F. Meng, L. Milani, G. W. Montgomery, D. W. Morris, T. W. Mühleisen, N. Mullins, H. Nguyen, C. M. Nievergelt, A. N. Adolfsson, E. A. Nwulia, C. O'Donovan, L. M. O. Loohuis, A. P. S. Ori, L. Oruc, U. Ösby, R. H. Perlis, A. Perry, A. Pfennig, J. B. Potash, S. M. Purcell, E. J. Regeer, A. Reif, C. S. Reinbold, J. P. Rice, F. Rivas, M. Rivera, P. Roussos, D. M. Ruderfer, E. Ryu, C. Sánchez-Mora, A. F. Schatzberg, W. A. Scheftner, N. J. Schork, C. Shannon Weickert, T. Shekhtman, P. D. Shilling, E. Sigurdsson, C. Slaney, O. B. Smeland, J. L. Sobell, C. Söholm Hansen, A. T. Spijker, D. St Clair, M. Steffens, J. S. Strauss, F. Streit, J. Strohmaier, S. Szelinger, R. C. Thompson, T. E. Thorgeirsson, J. Treutlein, H. Vedder, W. Wang, S. J. Watson, T. W. Weickert, S. H. Witt, S. Xi, W. Xu, A. H. Young, P.

- Zandi, P. Zhang, S. Zöllner, eQTLGen Consortium, BIOS Consortium, R. Adolfsson, I. Agartz, M. Alda, L. Backlund, B. T. Baune, F. Bellivier, W. H. Berrettini, J. M. Biernacka, D. H. R. Blackwood, M. Boehnke, A. D. Børglum, A. Corvin, N. Craddock, M. J. Daly, U. Dannlowski, T. Esko, B. Etain, M. Frye, J. M. Fullerton, E. S. Gershon, M. Gill, F. Goes, M. Grigoriou-Serbanescu, J. Hauser, D. M. Hougaard, C. M. Hultman, I. Jones, L. A. Jones, R. S. Kahn, G. Kirov, M. Landén, M. Leboyer, C. M. Lewis, Q. S. Li, J. Lissowska, N. G. Martin, F. Mayoral, S. L. McElroy, A. M. McIntosh, F. J. McMahon, I. Melle, A. Metspalu, P. B. Mitchell, G. Morken, O. Mors, P. B. Mortensen, B. Müller-Myhsok, R. M. Myers, B. M. Neale, V. Nimgaonkar, M. Nordentoft, M. M. Nöthen, M. C. O'Donovan, K. J. Oedegaard, M. J. Owen, S. A. Paciga, C. Pato, M. T. Pato, D. Posthuma, J. A. Ramos-Quiroga, M. Ribasés, M. Rietschel, G. A. Rouleau, M. Schalling, P. R. Schofield, T. G. Schulze, A. Serretti, J. W. Smoller, H. Stefansson, K. Stefansson, E. Stordal, P. F. Sullivan, G. Turecki, A. E. Vaaler, E. Vieta, J. B. Vincent, T. Werge, J. I. Nurnberger, N. R. Wray, A. Di Florio, H. J. Edenberg, S. Cichon, R. A. Ophoff, L. J. Scott, O. A. Andreassen, J. Kelsoe, P. Sklar, Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
80. D. Demontis, R. K. Walters, J. Martin, M. Mattheisen, T. D. Als, E. Agerbo, G. Baldursson, R. Belliveau, J. Bybjerg-Grauholm, M. Bækvad-Hansen, F. Cerrato, K. Chambert, C. Churchhouse, A. Dumont, N. Eriksson, M. Gandal, J. I. Goldstein, K. L. Grasby, J. Grove, O. O. Gudmundsson, C. S. Hansen, M. E. Hauberg, M. V. Hollegaard, D. P. Howrigan, H. Huang, J. B. Maller, A. R. Martin, N. G. Martin, J. Moran, J. Pallesen, D. S. Palmer, C. B. Pedersen, M. G. Pedersen, T. Poterba, J. B. Poulsen, S. Ripke, E. B. Robinson, F. K. Satterstrom, H. Stefansson, C. Stevens, P. Turley, G. B. Walters, H. Won, M. J. Wright, ADHD Working Group of the Psychiatric Genomics Consortium (PGC), Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium, 23andMe Research Team, O. A. Andreassen, P. Asherson, C. L. Burton, D. I. Boomsma, B. Cormand, S. Dalsgaard, B. Franke, J. Gelernter, D. Geschwind, H. Hakonarson, J. Haavik, H. R. Kranzler, J. Kuntsi, K. Langley, K.-P. Lesch, C. Middeldorp, A. Reif, L. A. Rohde, P. Roussos, R. Schachar, P. Sklar, E. J. S. Sonuga-Barke, P. F. Sullivan, A. Thapar, J. Y. Tung, I. D. Waldman, S. E. Medland, K. Stefansson, M. Nordentoft, D. M. Hougaard, T. Werge, O. Mors, P. B. Mortensen, M. J. Daly, S. V. Faraone, A. D. Børglum, B. M. Neale, Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
81. M. Luciano, S. P. Hagenaars, G. Davies, W. D. Hill, T.-K. Clarke, M. Shiri, S. E. Harris, R. E. Marioni, D. C. Liewald, C. Fawns-Ritchie, M. J. Adams, D. M. Howard, C. M. Lewis, C. R. Gale, A. M. McIntosh, I. J. Deary, Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat. Genet.* **50**, 6–11 (2018).
82. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. **511**, 421–427 (2014).
83. W. van Rheenen, A. Shatunov, A. M. Dekker, R. L. McLaughlin, F. P. Diekstra, S. L. Pulit, R. A. A. van der Spek, U. Vösa, S. de Jong, M. R. Robinson, J. Yang, I. Fogh, P. T. van Doormaal, G. H. P. Tazelaar, M. Koppers, A. M. Blokhuis, W. Sproviero, A. R. Jones, K. P. Kenna, K. R. van Eijk, O. Harschnitz, R. D. Schellevis, W. J. Brands, J. Medic, A. Menelaou, A. Vajda, N. Ticozzi, K. Lin, B. Rogelj, K. Vrabec, M. Ravník-Glavač, B. Koritnik, J. Zidar, L. Leonardi, L. D. Grošelj, S. Millecamps, F. Salachas, V. Meininger, M. de Carvalho, S. Pinto, J. S. Mora, R. Rojas-García, M. Polak, S. Chandran, S. Colville, R. Swigler, K. E. Morrison, P. J. Shaw, J. Hardy, R. W. Orrell, A. Pittman, K. Sidle, P. Fratta, A. Malaspina, S. Topp, S. Petri, S. Abdulla, C. Drepper, M. Sendtner, T. Meyer, R. A. Ophoff, K. A. Staats, M. Wiedau-Pazos, C. Lomen-Hoerth, V. M. Van Deerlin, J. Q. Trojanowski, L. Elman, L. McCluskey, A. N. Basak, C. Tunca, H. Hamzeiy, Y. Parman, T. Meitinger, P. Lichtner, M. Radivojkovic-Blagojevic, C. R. Andres, C. Maurel, G. Bensimon, B. Landwehrmeyer, A. Brice, C. A. M. Payan, S. Saker-Delye, A. Dürr, N. W. Wood, L. Tittmann, W. Lieb, A. Franke, M. Rietschel, S. Cichon, M. M. Nöthen, P. Amouyel, C. Tzourio, J.-F. Dartigues, A. G. Uitterlinden, F. Rivadeneira, K. Estrada, A. Hofman, C. Curtis, H. M. Blauw, A. J. van der Kooij, M. de Visser, A. Goris, M. Weber, C. E. Shaw, B. N. Smith, O. Pansarasa, C. Cereda, R. Del Bo, G. P. Comi, S. D'Alfonso, C. Bertolin, G. Sorarù, L. Mazzini, V. Pensato, C. Gellera, C. Tiloca, A. Ratti, A. Calvo, C. Moglia, M. Brunetti, S. Arcuti, R. Capozzo, C. Zecca, C. Lunetta, S. Penco, N. Riva, A. Padovani, M. Filosto,

- B. Muller, R. J. Stuit, PARALS Registry, SLALOM Group, SLAP Registry, FALS Sequencing Consortium, SLAGEN Consortium, NNIPPS Study Group, I. Blair, K. Zhang, E. P. McCann, J. A. Fifita, G. A. Nicholson, D. B. Rowe, R. Pamphlett, M. C. Kiernan, J. Grosskreutz, O. W. Witte, T. Ringer, T. Prell, B. Stubendorff, I. Kurth, C. A. Hübner, P. N. Leigh, F. Casale, A. Chio, E. Beghi, E. Pupillo, R. Tortelli, G. Logroscino, J. Powell, A. C. Ludolph, J. H. Weishaupt, W. Robberecht, P. Van Damme, L. Franke, T. H. Pers, R. H. Brown, J. D. Glass, J. E. Landers, O. Hardiman, P. M. Andersen, P. Corcia, P. Vourc'h, V. Silani, N. R. Wray, P. M. Visscher, P. I. W. de Bakker, M. A. van Es, R. J. Pasterkamp, C. M. Lewis, G. Breen, A. Al-Chalabi, L. H. van den Berg, J. H. Veldink, Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat. Genet.* **48**, 1043–1048 (2016).
84. W. Zhou, J. B. Nielsen, L. G. Fritsche, R. Dey, M. E. Gabrielsen, B. N. Wolford, J. LeFaive, P. VandeHaar, S. A. Gagliano, A. Gifford, L. A. Bastarache, W.-Q. Wei, J. C. Denny, M. Lin, K. Hveem, H. M. Kang, G. R. Abecasis, C. J. Willer, S. Lee, Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
85. P. R. Jansen, K. Watanabe, S. Stringer, N. Skene, J. Bryois, A. R. Hammerschlag, C. A. de Leeuw, J. S. Benjamins, A. B. Muñoz-Manchado, M. Nagel, J. E. Savage, H. Tiemeier, T. White, 23andMe Research Team, J. Y. Tung, D. A. Hinds, V. Vacic, X. Wang, P. F. Sullivan, S. van der Sluis, T. J. C. Polderman, A. B. Smit, J. Hjerling-Leffler, E. J. W. Van Someren, D. Posthuma, Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).
86. C. P. Nelson, A. Goel, A. S. Butterworth, S. Kanoni, T. R. Webb, E. Marouli, L. Zeng, I. Ntalla, F. Y. Lai, J. C. Hopewell, O. Giannakopoulou, T. Jiang, S. E. Hamby, E. Di Angelantonio, T. L. Assimes, E. P. Bottinger, J. C. Chambers, R. Clarke, C. N. A. Palmer, R. M. Cubbon, P. Ellinor, R. Erme, E. Evangelou, P. W. Franks, C. Grace, D. Gu, A. D. Hingorani, J. M. M. Howson, E. Ingelsson, A. Kastrati, T. Kessler, T. Kyriakou, T. Lehtimäki, X. Lu, Y. Lu, W. März, R. McPherson, A. Metspalu, M. Pujades-Rodriguez, A. Ruusalepp, E. E. Schadt, A. F. Schmidt, M. J. Sweeting, P. A. Zalloua, K. AlGhalayini, B. D. Keavney, J. S. Kooner, R. J. F. Loos, R. S. Patel, M. K. Rutter, M. Tomaszewski, I. Tzoulaki, E. Zeggini, J. Erdmann, G. Dedoussis, J. L. M. Björkegren, EPIC-CVD Consortium, CARDIoGRAMplusC4D, UK Biobank CardioMetabolic Consortium CHD working group, H. Schunkert, M. Farrall, J. Danesh, N. J. Samani, H. Watkins, P. Deloukas, Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391 (2017).
87. V. Deary, S. P. Hagenaars, S. E. Harris, W. D. Hill, G. Davies, D. C. M. Liewald, International Consortium for Blood Pressure GWAS, CHARGE consortium Aging and Longevity Group, CHARGE consortium Inflammation Group, A. M. McIntosh, C. R. Gale, I. J. Deary, Genetic contributions to self-reported tiredness. *Mol. Psychiatry.* **23**, 789–790 (2018).
88. J. Chiou, R. J. Geusz, M.-L. Okino, J. Y. Han, M. Miller, R. Melton, E. Beebe, P. Benaglio, S. Huang, K. Korgaonkar, S. Heller, A. Kleger, S. Preissl, D. U. Gorkin, M. Sander, K. J. Gaulton, Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature.* **594**, 398–402 (2021).
89. A. Mahajan, D. Taliun, M. Thurner, N. R. Robertson, J. M. Torres, N. W. Rayner, A. J. Payne, V. Steinthorsdottir, R. A. Scott, N. Grarup, J. P. Cook, E. M. Schmidt, M. Wuttke, C. Sarnowski, R. Mägi, J. Nano, C. Gieger, S. Trompet, C. Lecocoeur, M. H. Preuss, B. P. Prins, X. Guo, L. F. Bielak, J. E. Below, D. W. Bowden, J. C. Chambers, Y. J. Kim, M. C. Y. Ng, L. E. Petty, X. Sim, W. Zhang, A. J. Bennett, J. Bork-Jensen, C. M. Brummett, M. Canouil, K.-U. Eckardt, K. Fischer, S. L. R. Kardia, F. Kronenberg, K. Läll, C.-T. Liu, A. E. Locke, J. 'an Luan, I. Ntalla, V. Nylander, S. Schönherr, C. Schurmann, L. Yengo, E. P. Bottinger, I. Brandslund, C. Christensen, G. Dedoussis, J. C. Florez, I. Ford, O. H. Franco, T. M. Frayling, V. Giedraitis, S. Hackinger, A. T. Hattersley, C. Herder, M. A. Ikram, M. Ingelsson, M. E. Jørgensen, T. Jørgensen, J. Kriebel, J. Kuusisto, S. Ligthart, C. M. Lindgren, A. Linneberg, V. Lyssenko, V. Mamakou, T. Meitinger, K. L. Mohlke, A. D. Morris, G. Nadkarni, J. S. Pankow, A. Peters, N. Sattar, A. Stančáková, K. Strauch, K. D. Taylor, B. Thorand, G. Thorleifsson, U. Thorsteinsdottir, J. Tuomilehto, D. R. Witte, J. Dupuis, P. A. Peyser, E. Zeggini, R. J. F. Loos, P. Froguel, E. Ingelsson, L. Lind, L. Groop, M. Laakso, F. S. Collins, J. W. Jukema, C. N. A. Palmer, H. Grallert, A. Metspalu, A. Dehghan, A. Köttgen, G. R. Abecasis, J. B.

- Meigs, J. I. Rotter, J. Marchini, O. Pedersen, T. Hansen, C. Langenberg, N. J. Wareham, K. Stefansson, A. L. Gloyn, A. P. Morris, M. Boehnke, M. I. McCarthy, Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
90. M. A. Ferreira, J. M. Vonk, H. Baurecht, I. Marenholz, C. Tian, J. D. Hoffman, Q. Helmer, A. Tillander, V. Ullemar, J. van Dongen, Y. Lu, F. Rüschenhoff, J. Esparza-Gordillo, C. W. Medway, E. Mountjoy, K. Burrows, O. Hummel, S. Grosche, B. M. Brumpton, J. S. Witte, J.-J. Hottenga, G. Willemsen, J. Zheng, E. Rodriguez, M. Hotze, A. Franke, J. A. Revez, J. Beesley, M. C. Matheson, S. C. Dharmage, L. M. Bain, L. G. Fritsche, M. E. Gabrielsen, B. Balliu, 23andMe Research Team, AAGC collaborators, BIOS consortium, LifeLines Cohort Study, J. B. Nielsen, W. Zhou, K. Hveem, A. Langhammer, O. L. Holmen, M. Løset, G. R. Abecasis, C. J. Willer, A. Arnold, G. Homuth, C. O. Schmidt, P. J. Thompson, N. G. Martin, D. L. Duffy, N. Novak, H. Schulz, S. Karrasch, C. Gieger, K. Strauch, R. B. Melles, D. A. Hinds, N. Hübner, S. Weidinger, P. K. E. Magnusson, R. Jansen, E. Jorgenson, Y.-A. Lee, D. I. Boomsma, C. Almqvist, R. Karlsson, G. H. Koppelman, L. Paternoster, Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* **49**, 1752–1757 (2017).
 91. R. J. P. van der Valk, E. Kreiner-Møller, M. N. Kooijman, M. Guxens, E. Stergiakouli, A. Sääf, J. P. Bradfield, F. Geller, M. G. Hayes, D. L. Cousminer, A. Körner, E. Thiering, J. A. Curtin, R. Myhre, V. Huikari, R. Joro, M. Kerkhof, N. M. Warrington, N. Pitkänen, I. Ntalla, M. Horikoshi, R. Veijola, R. M. Freathy, Y.-Y. Teo, S. J. Barton, D. M. Evans, J. P. Kemp, B. St Pourcain, S. M. Ring, G. Davey Smith, A. Bergström, I. Kull, H. Hakonarson, F. D. Mentch, H. Bisgaard, B. Chawes, J. Stokholm, J. Waage, P. Eriksen, A. Sevelsted, M. Melbye, Early Genetics and Lifecourse Epidemiology (EAGLE) Consortium, C. M. van Duijn, C. Medina-Gomez, A. Hofman, J. C. de Jongste, H. R. Taal, A. G. Uitterlinden, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, L. L. Armstrong, J. Eriksson, A. Palotie, M. Bustamante, X. Estivill, J. R. Gonzalez, S. Llop, W. Kiess, A. Mahajan, C. Flexeder, C. M. T. Tiesler, C. S. Murray, A. Simpson, P. Magnus, V. Sengpiel, A.-L. Hartikainen, S. Keinänen-Kiukaanniemi, A. Lewin, A. Da Silva Couto Alves, A. I. Blakemore, J. L. Buxton, M. Kaakinen, A. Rodriguez, S. Sebert, M. Vaarasmäki, T. Lakka, V. Lindi, U. Gehring, D. S. Postma, W. Ang, J. P. Newnham, L.-P. Lyytikäinen, K. Pakkala, O. T. Raitakari, K. Panoutsopoulou, E. Zeggini, D. I. Boomsma, M. Groen-Blokhuis, J. Ilonen, L. Franke, J. N. Hirschhorn, T. H. Pers, L. Liang, J. Huang, B. Hoche, M. Knip, S.-M. Saw, J. W. Holloway, E. Melén, S. F. A. Grant, B. Feenstra, W. L. Lowe, E. Widén, E. Sergeyev, H. Grallert, A. Custovic, B. Jacobsson, M.-R. Jarvelin, M. Atalay, G. H. Koppelman, C. E. Pennell, H. Niinikoski, G. V. Dedoussis, M. I. McCarthy, T. M. Frayling, J. Sunyer, N. J. Timpson, F. Rivadeneira, K. Bønnelykke, V. W. V. Jaddoe, Early Growth Genetics (EGG) Consortium, A novel common variant in DCST2 is associated with length in early life and height in adulthood. *Hum. Mol. Genet.* **24**, 1155–1168 (2015).
 92. M. Horikoshi, R. N. Beaumont, F. R. Day, N. M. Warrington, M. N. Kooijman, J. Fernandez-Tajes, B. Feenstra, N. R. van Zuydam, K. J. Gaulton, N. Grarup, J. P. Bradfield, D. P. Strachan, R. Li-Gao, T. S. Ahluwalia, E. Kreiner, R. Rueedi, L.-P. Lyytikäinen, D. L. Cousminer, Y. Wu, E. Thiering, C. A. Wang, C. T. Have, J.-J. Hottenga, N. Vilor-Tejedor, P. K. Joshi, E. T. H. Boh, I. Ntalla, N. Pitkänen, A. Mahajan, E. M. van Leeuwen, R. Joro, V. Lagou, M. Nodzenski, L. A. Diver, K. T. Zondervan, M. Bustamante, P. Marques-Vidal, J. M. Mercader, A. J. Bennett, N. Rahmioglu, D. R. Nyholt, R. C. W. Ma, C. H. T. Tam, W. H. Tam, CHARGE Consortium Hematology Working Group, S. K. Ganesh, F. J. van Rooij, S. E. Jones, P.-R. Loh, K. S. Ruth, M. A. Tuke, J. Tyrrell, A. R. Wood, H. Yaghootkar, D. M. Scholtens, L. Paternoster, I. Prokopenko, P. Kovacs, M. Atalay, S. M. Willems, K. Panoutsopoulou, X. Wang, L. Carstensen, F. Geller, K. E. Schraut, M. Murcia, C. E. van Beijsterveldt, G. Willemsen, E. V. R. Appel, C. E. Fonvig, C. Trier, C. M. Tiesler, M. Standl, Z. Kutalik, S. Bonas-Guarch, D. M. Hougaard, F. Sánchez, D. Torrents, J. Waage, M. V. Hollegaard, H. G. de Haan, F. R. Rosendaal, C. Medina-Gomez, S. M. Ring, G. Hemani, G. McMahon, N. R. Robertson, C. J. Groves, C. Langenberg, J. 'an Luan, R. A. Scott, J. H. Zhao, F. D. Mentch, S. M. MacKenzie, R. M. Reynolds, Early Growth Genetics (EGG) Consortium, W. L. Lowe Jr, A. Tönjes, M. Stumvoll, V. Lindi, T. A. Lakka, C. M. van Duijn, W. Kiess, A. Körner, T. I. Sørensen, H. Niinikoski, K. Pakkala, O. T. Raitakari, E. Zeggini, G. V. Dedoussis, Y.-Y. Teo, S.-M. Saw, M. Melbye, H. Campbell, J. F. Wilson, M. Vrijheid, E. J. de Geus, D. I. Boomsma, H. N. Kadamideen, J.-C. Holm, T. P. Hansen, S. Sebert, A. T. Hattersley, L. J. Beilin, J. P.

- Newnham, C. E. Pennell, J. Heinrich, L. S. Adair, J. B. Borja, K. L. Mohlke, J. G. Eriksson, E. E. Widén, M. Kähönen, J. S. Viikari, T. Lehtimäki, P. Vollenweider, K. Bønnelykke, H. Bisgaard, D. O. Mook-Kanamori, A. Hofman, F. Rivadeneira, A. G. Uitterlinden, C. Pisinger, O. Pedersen, C. Power, E. Hyppönen, N. J. Wareham, H. Hakonarson, E. Davies, B. R. Walker, V. W. Jaddoe, M.-R. Jarvelin, S. F. Grant, A. A. Vaag, D. A. Lawlor, T. M. Frayling, G. Davey Smith, A. P. Morris, K. K. Ong, J. F. Felix, N. J. Timpson, J. R. Perry, D. M. Evans, M. I. McCarthy, R. M. Freathy, Genome-wide associations for birth weight and correlations with adult disease. *Nature*. **538**, 248–252 (2016).
93. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global reference for human genetic variation. *Nature*. **526**, 68–74 (2015).
94. A. N. Gorban, A. Y. Zinovyev, "Principal Graphs and Manifolds" in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques* (IGI Global, 2010), pp. 28–59.
95. C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
96. A. Heger, C. Webber, M. Goodson, C. P. Ponting, G. Lunter, GAT: a simulation framework for testing the association of genomic intervals. *Bioinformatics*. **29**, 2046–2048 (2013).

Acknowledgments: We thank Jesse R Dixon, Trygve E Bakken, Nikolas L Jorstad, and Song-Lin Ding for the scientific discussion and suggestive comments. We are very grateful to members of the Ecker group for feedback and discussion. We thank Dr. Erica Melief and Aimee Schantz for outstanding administrative assistance, Lisa Keene, Katelyn Kern, and Amanda Keen for support in tissue collection, and most importantly the brain tissue donors and their loved ones, without whom this work would be impossible.

Funding: This work was supported by grants from NIMH U01MH121282 to B.R., M.M.B and J.R.E., UM1 MH130994 to J.R.E., M.M.B. and B.R., NIMH U01MH114812 to E.L and S.L, and in part by the Nancy and

Buster Alvord Endowment to C.D.K. The Flow Cytometry Core Facility of the Salk Institute is supported by funding from NIH-NCI CCSG: P30 014195 and Shared Instrumentation Grant S10-OD023689. J.R.E is an investigator of the Howard Hughes Medical Institute.

Author contributions: Study supervision: J.R.E. Contribution to data generation: W.T., J.Z., A.B, R.G.C., M.K., J.A., A.A., J.R.N., H.C., N.D.J., J.L., J.K.O., A.P., N.E., J.R., J.L., M.L., N.C., C.O., C.V., A.M.Y., J.N., N.D., T.C., N.S., D.H., R.H., B.P.L., C.D.K. Contribution to data analysis: W.T., J.Z. Q.Z., J.X. Contribution to data interpretation: W.T., J.Z., K.S., E.L., B.R., M.M.B., J.R.E. Contribution to writing the manuscript: W.T, J.Z., J.R.E. All authors edited and approved the manuscript.

Competing interests: J.R.E is a member of the scientific advisor for Zymo Research and Ionis. B.R. is a co-founder and consultant of Arima Genomics Inc. and co-founder of Epigenome Technologies.

Data availability: The data analyzed in this study can be downloaded from <http://neomorph.salk.edu/wtian/hba-data/>. The data were produced through the Brain Initiative Cell Census Network (BICCN:RRID:SCR_015820) and deposited in the NEMO Archive (RRID:SCR_002001) under identifier nemo:dat-jx4eu3g accessible at <https://assets.nemoarchive.org/dat-jx4eu3g>. Raw and processed data were also deposited to NCBI GEO/SRA with accession number GSE215353. A browser of major type DNA methylome can be found here <http://neomorph.salk.edu/hba/hba-majoritytype.php>.

Supplementary materials

Materials and Methods

Figs. S1 to S9

Tables S1 to S8