Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability

Lindsay F. Rizzardi1,2,7, Peter F. Hickey3,7, Varena Rodriguez DiBlasi1,2, Rakel Tryggvadóttir1, Colin M. Callahan1, Adrian Idrizi1, Kasper D. Hansen1,3,4, and Andrew P. Feinberg1,2,5,6*

Distinct cognitive functions and behaviors are often correlated with particular regions and/or cell types in the brain; most disease-based brain research is focused on identifying the anatomical structures that mediate normal function, for example, the hippocampus in memory, the prefrontal cortex in cognition, and the nucleus accumbens in addictive behavior. Further, many neuropsychiatric diseases preferentially affect individual neuronal subpopulations present in particular brain regions. The epigenome is particularly important in maintaining cellular identity and responding to environmental perturbations; therefore, mapping the cell type– and brain region–specific transcriptional and epigenetic landscapes is necessary for identifying functional genomic differences that contribute to disease phenotypes. Recent transcriptome analyses have revealed extensive differences between non-cerebellar human brain regions and between distinct neuronal and glial subpopulations in mice. Substantial transcriptional heterogeneity also exists among single cells profiled in mammalian cortical regions. Single-cell DNA methylation within a single brain region has only recently been measured, and the specific locations and functional consequences of methylation changes among cell types remain underexplored.

In contrast to gene expression, few, if any, DNA methylation differences among phenotypically normal non-cerebellar human brain regions have been reported. DNA methylation is known to be altered in patients with neuropsychiatric disease, including schizophrenia, Alzheimer’s disease, and major depressive disorder. The only whole-genome bisulfite sequencing (WGBS) analyses across multiple human brain regions found few differences among cortical tissues. The apparent lack of epigenetic diversity among brain regions is surprising, given the known transcriptional diversity. Importantly, large-scale –omics studies and case-control studies are predominately conducted using bulk tissues comprised of neuronal and non-neuronal cell populations at variable proportions. Although the authors of these studies acknowledge the confounding potential due to cellular heterogeneity, a robust reference of sorted populations from multiple brain regions would greatly improve current computational deconvolution strategies.

In contrast to studies in mice, few studies have examined DNA methylation between cellular subpopulations isolated from human brain tissue. Neurons are easily distinguished from non-neurons (astrocytes, oligodendrocytes, microglia, and epithelial cells) using the nuclear neuronal marker NeuN. DNA methylation differences between these two broadly defined populations have been widely reported and have been assessed genome-wide within a single brain region across development and in the context of several neurodegenerative diseases. Still, a comprehensive analysis of brain region–specific DNA methylation has not been performed using sorted nuclei isolated from human tissues.

Here, we addressed this knowledge gap by analyzing the DNA methylation landscape (both CpG and non-CpG) using WGBS in fractionated neuronal and non-neuronal nuclei and bulk tissues isolated from four post-mortem brain regions: dorso-lateral prefrontal cortex (BA9), anterior cingulate gyrus (BA24), hippocampus (HC), and nucleus accumbens (NAcc). We also examined both gene expression and chromatin accessibility (n = 22) in NAcc and BA9 nuclei. Importantly, we find that regions of differential methylation, specifically within the neuronal populations, are highly enriched for heritability of schizophrenia, addictive behavior, and neuroticism.
Results

Cell type heterogeneity obscures epigenetic differences between brain regions. We mapped the DNA methylation landscape using WGBS of four human post-mortem brain regions: dorsolateral prefrontal cortex (BA9), anterior cingulate cortex (BA24), hippocampus (HC), and nucleus accumbens (NAcc) (Supplementary Table 1). We isolated neuronal and non-neuronal nuclei based on the neuronal marker NeuN using fluorescence-activated nuclear sorting (FANS), then WGBS in a total of 45 samples from six donors (Supplementary Table 2; Supplementary Fig. 1a,b) and generated data from 27 bulk tissue samples. We observed substantial variation in the proportion of NeuN+ nuclei among brain regions, within the same brain region between individuals and even among samplings from the same tissue specimen (Supplementary Fig. 1c). Several factors contribute to this variability, including nuclei loss or damage during sample processing, unequal subsampling of tissue specimens, and differences in post-mortem intervals, though only within HC and NAcc (Supplementary Fig. 1d; \( P = 0.025 \) from linear mixed model). This finding emphasizes the inherent challenge in accounting for cellular heterogeneity in complex solid tissues that, unlike blood and other peripheral tissues, cannot be repeatedly sampled.

Neuronal DNA methylation (CpG (mCG) and non-CpG (mCH)) in the NAcc was distinct from BA9 and HC (Supplementary Fig. 1e). We also confirmed previous observations\(^\text{30-32}\) of higher global mCG and mCH (\( P < 2.2 \times 10^{-16} \), for both) in neuronal compared with non-neuronal nuclei (Supplementary Fig. 1f). Interestingly, NAcc neuronal nuclei had higher global levels of mCG, but lower levels of mCH than the other brain regions. Principal component analysis (PCA) of autosomal DNA methylation reveals clear segregation of cell types (neuronal from non-neuronal nuclei) and brain regions within the neuronal population (Fig. 1a) that is not observed in analysis of bulk tissue (Supplementary Fig. 2a, Supplementary Table 3). Segregation by brain region becomes even more apparent upon separate analysis of neuronal nuclei (Supplementary Fig. 2b), and joint analysis of bulk tissues and sorted nuclei allows better resolution of brain region–specific differences among the bulk samples (compare Supplementary Fig. 2a and c). In contrast, separate analysis of non-neuronal nuclei reveals segregation of samples by donor rather than tissue (Supplementary Fig. 2d), similar to previous findings\(^\text{33}\).

Consistent with the PCA results, a high correlation among bulk tissues and non-neuronal nuclei (Supplementary Fig. 2e) was observed. Further, correlation within an individual is much lower than within sample type (NeuN+, NeuN−, or bulk tissue) (Supplementary Fig. 3a). Although the non-neuronal population is also composed of several distinct cell types, we found that these samples had the least between-sample variability and the smallest effect size (1%, proportion of CpGs with absolute mean difference > 10%), with bulk tissue samples having the greatest variability and the neuronal population having the greatest effect size (3–12%) (Supplementary Fig. 3b,c). Although our results cannot exclude the possibility of brain region–specific methylation in rare non-neuronal populations, they indicate that sample-to-sample variability in cell type composition is not confounding our assertion that non-neuronal samples show little brain region–specific methylation. This finding contrasts with the observations of regional transcriptional differences in astrocytes in mice\(^\text{34}\), but is consistent with the relatively consistent composition of glial subpopulations across multiple brain regions\(^\text{35}\). Our data indicate that the neuronal populations contain the relevant functional variability in DNA methylation between brain regions and are consistent with the recent characterizations of cell type–specific gene expression and methylation signatures of sorted neuronal nuclei and among distinct neuronal populations in both mouse and humans\(^\text{36-38}\).

Neuronal nuclei display brain region–specific DNA methylation profiles. To identify differentially methylated regions (DMRs), we extended our previously published statistical method\(^\text{39}\) to accommodate multigroup comparisons while accounting for the variation between biological replicates. This new method allows us to simultaneously compare all 45 samples and identify regions of differential methylation between any two groups.

mCG. We found substantial (>10%) mean methylation differences among all four brain regions in the NeuN+ population (Fig. 1b) in addition to the previously reported\(^\text{21,22}\) ‘cell type’ differences between neuronal and non-neuronal nuclei (Supplementary Fig. 4a). We identified 97,924 autosomal cell type CG-DMRs, of which 21,802 are novel\(^\text{21,22}\), and 19,072 large blocks of differential mCG (family-wise error rates (FWER) ≤ 5%) with neurons being primarily hypermethylated, consistent with their global hypermethylated status\(^\text{21,22}\) (Supplementary Fig. 1e, Supplementary Tables 4–7). These cell type CG-DMRs clearly distinguish NeuN+ from NeuN− nuclei and are present at the promoters of many cell type–specific genes (Supplementary Fig. 4a,b).

Given the large methylation differences among NeuN+ nuclei, we repeated our differential methylation analysis on the NeuN+ and NeuN− samples separately and compared the results to those generated from bulk tissues. We identified 13,074 autosomal neuronal (NeuN+) CG-DMRs between brain regions containing >1% of all CpGs analyzed (FWER < 5%) (Fig. 1b). These neuronal CG-DMRs, as for the cell type CG-DMRs, were enriched in regulatory regions that have been (1) defined by H3K27ac in human brain regions\(^\text{26}\), (2) identified as permissive enhancers across many cell types and tissues\(^\text{37}\), and (3) identified using a map of chromatin states in four brain regions\(^\text{41}\) and in regions of open chromatin (Fig. 1c). In contrast, we found few autosomal CG-DMRs among NeuN− nuclei (114 CG-DMRs) or bulk tissues (71 CG-DMRs) from these brain regions (Supplementary Tables 4, 8–9). These findings demonstrate that methylation differences among neurons from distinct brain regions are masked by the substantial variation in cellular heterogeneity across samples.

Several patterns emerged upon hierarchical clustering of methylation levels over neuronal CG-DMRs, similar to those seen in cancer-to-normal comparisons\(^\text{40}\) (Fig. 1d). The largest clusters (groups 1 and 2) consisted of symmetric methylation differences enriched for regulatory regions (Fig. 1e). The other group consisted primarily of CG-DMRs representing shifts in methylation boundaries and was enriched in promoters, CpG islands (CGIs), and shores. These CG-DMRs (group 3) represent the smallest fraction of our neuronal CG-DMRs, but their enrichment is similar to CG-DMRs we previously identified as a consequence of oncogenic transformation of B cells using the Epstein–Barr virus\(^\text{41}\). These data suggest that methylation is altered in specific ways depending on the genomic feature, with focal increases or decreases in methylation at regulatory regions and more subtle spreading or shrinking of methylation boundaries at promoter-associated features.

Recent single-cell analyses have clearly shown that epigenetic and transcriptional profiles reflect distinct neuronal subpopulations with distinct functions\(^\text{42,43}\). Therefore, it is unsurprising that the neuronal compositions of the brain regions we analyzed are reflected in their DNA methylation profiles (Supplementary Fig. 2b). Unlike NAcc (composed primarily of inhibitory GABAergic medium spiny neurons (MSNs)), HC, BA9, and BA24 are heterogeneous, containing both excitatory and inhibitory subpopulations\(^\text{44,45}\). The majority of neuronal CG-DMRs (11,895) distinguish the NAcc from the other brain regions. Using the Genomic Regions Enrichment of Annotations Tool (GREAT)\(^\text{24}\), we found that regions of hypomethylation in NAcc were enriched in categories highly relevant to MSN function: dopamine receptor signaling pathway (\( q \text{ value} = 2.59 \times 10^{-23}\)), adenylyl cyclase–activating dopamine receptor signaling pathway (\( q \text{ value} = 3.55 \times 10^{-16}\)), and synaptic transmission, dopaminergic (\( q \text{ value} = 5.21 \times 10^{-5}\)) (Supplementary Table 10).
Fig. 1 | Neuronal nuclei isolated from different brain regions display widespread differences in CpG methylation. DNA methylation was assessed in samples from four tissues: anterior cingulate cortex (BA24, pink), prefrontal cortex (BA9, blue), nucleus accumbens (NAcc; orange), and hippocampus (HC; gray). a, PCA of distances derived from average autosomal CpG methylation (n 1 kb intervals) in NeuN + and NeuN − nuclei from each brain region; each point is a sample (n = 6 for BA9, HC, NAcc in both NeuN + and NeuN −; for BA24, n = 5 for NeuN + and n = 4 for NeuN −). b, Hierarchical clustering of samples based on the average methylation per sample of the neuronal CG-DMRs (n = 13,074). c, Log odds ratios (ORs) for the enrichment of CpGs within cell type (NeuN + versus NeuN −) and neuronal (NeuN +) CG-DMRs and blocks compared with the rest of the genome for genomic features. Gene models from GENCODE (promoters, intronic, exonic, 5′ UTR, 3′ UTR, intergenic), CpG islands (CGIs) and related features from UCSC (shores, shelves, OpenSea), putative enhancer regions (H3K27ac, FANTOM5), open chromatin regions (OCRs) from this study, and brain-specific ChromHMM annotations (core 15-state model). d, Average methylation for NeuN + nuclei over a 3 kb window centered on the neuronal CG-DMRs. CG-DMRs were grouped by k-means clustering based on their methylation patterns. Metagene plots for each group are shown to the right, with the number of CG-DMRs in each group indicated. e, Heatmap as in c, showing enrichment within neuronal CG-DMRs from groups 1–3 and CG-DMRs from Epstein–Barr-transformed B cells. f, Example CG-DMR (left) and CG-block (right) showing average methylation values for NeuN + nuclei from each tissue, as indicated. Regions of differential methylation are shaded in pink.

Examples of hypomethylated genes include those encoding markers of GABAergic neurons (BCL11B and DARPP-32) (Fig. 1f) and dopamine receptors (DRD1 and DRD3) (Supplementary Fig. 4c). In contrast, regions of hypermethylation were enriched in multiple brain development categories (q values < 1 × 10−11) (Supplementary Table 10).
Given the overwhelming differences between the NAcc and the other brain regions, we hypothesized that neuronal CG-DMRs among BA9, BA24, and HC could be obscured. Therefore, we repeated our analysis using only the neuronal samples from these regions, resulting in the identification of 208 autosomal neuronal CG-DMRs including 25 CG-DMRs unique to this analysis (Supplementary Fig. 4d and Supplementary Tables 4 and 11). Again, we find clusters of CG-DMRs representing boundary shifts and focal changes, such as that shown in the promoter of SATB2 (Supplementary Fig. 4c).

Previous studies have identified large blocks of differential methylation during brain development and cancer—normal comparisons36,37. We also identify 1,808 blocks of differential mCG among neurons from our four brain regions; the majority (964) of which have a mean methylation difference ≥10% (Supplementary Tables 4 and 12). Interestingly, 23% of these blocks cover the entirety of a protein-coding gene, including BCL11B (Fig. 1f) and GABRB2 (Supplementary Fig. 4f and Supplementary Table 12).

mCH. In addition to mCG, neurons and embryonic stem cells also have extensive non-CpG methylation (mCH). While mCH and mCG are spatially correlated38,39, there is evidence to suggest that they can have independent functions in the brain40,41. In agreement with previous reports, we detected the highest methylation levels (~10%) in the CA context of our NeuN with previous reports, we detected the highest methylation levels (via ATAC-seq35) in neuronal and non-neuronal populations measured gene expression (via RNA-seq) and chromatin accessibility (via open chromatin regions (OCRs) did not distinguish brain regions within the neuronal cell type (compare Supplementary Fig. 2e with Supplementary Fig. 7a). Similar to previous reports42,163,026 of the 283,812 OCRs tested were differentially accessible (FDR < 5%) between NeuN+ and NeuN− nuclei, termed cell type differentially accessible regions (DARs) (Supplementary Fig. 7b and Supplementary Table 18). We further identified 68,021 ‘neuronal’ DARs (only 13 identified among non-neuronal nuclei) between NAcc and BA9 (FDR < 5%) (Fig. 3d,e and Supplementary Tables 19–21). OCRs were enriched over multiple features, including gene-centric features such as promoters and exons, and depleted in intergenic and ‘open sea’ regions (Fig. 3f). In contrast, neuronal DARs lacked strong enrichment for any specific genomic feature, with the exception of neuronal CG-DMRs.

Given this enrichment of neuronal CG-DMRs in DARs, we further investigated the overlap between these features using the 12,895 CG-DMRs identified between NAcc and BA9 neurons. Only 10% of DARs overlapped CG-DMRs, and conversely, 55% of CG-DMRs overlapped DARs. This was true even when restricted to the most divergent DARs (absolute fold change > 2, 12% overlap CG-DMRs). The direction of change in DARs and CG-DMRs is highly concordant with 99.9% of neuronal CG-DMRs having higher methylation when the region is less accessible, consistent with earlier reports in mice43. Further, we observed consistent methylation differences near all DARs (absolute fold change > 2, 12% overlap CG-DMRs). This analysis indicates that strand and context are paired that which of these modalities exhibit the strongest change can vary, probably due to genome location.

Differential mCH is more strongly correlated with differential gene expression than either mCG or chromatin accessibility. We next investigated the relationship between differential expression of protein-coding genes, differential methylation, and differential accessibility. An inverse relationship between global genic cytosine methylation and gene expression in neurons is well-established (Supplementary Fig. 8a), and this relationship is maintained when correlating differential methylation and differential expression between NAcc and BA9 (Supplementary Fig. 8b). Likewise, differential accessibility and expression are positively correlated (Supplementary Fig. 8c). CG-DMRs that overlap DARs revealed a stronger correlation with gene expression than CG-DMRs alone, which was not the case with CA-DMRs (Supplementary Fig. 8d,e).

Both DARs and DMRs preferentially occur within DEGs, with CG-DMRs and DARs near the TSS and CA-DMRs and CG-blocks distributed across the gene body (Fig. 4a). Additionally, CA-DMRs have the strongest correlation with expression over gene bodies, followed by CG-blocks (Fig. 4b). Interestingly, as a larger portion of the gene body is covered by a CA-DMR, that gene is more likely to be differentially expressed (Fig. 4c). This is not the case for DARs, CG-DMRs, or CG-blocks. We also find that protein coding genes with CA-DMRs tend to have a lower average expression across all neuronal samples (0.58 versus 1.83 reads per kilobase of transcript,
per million mapped reads (RPKM), respectively; \( P \) value < 2.2 × 10\(^{-16} \); Mann–Whitney) even when differentially expressed (0.78 versus 2.03 RPKM, respectively; \( P \) value < 2.2 × 10\(^{-16} \); Mann–Whitney). This finding is consistent with a previous report\(^{19} \), and our own data, showing that lowly expressed genes have higher levels of mCA than highly expressed genes (Supplementary Fig. 8f). Together, these data support and extend previous assertions that among accessibility, mCG and mCH, mCH is the best predictor of gene expression\(^{3} \) and further, that differential mCH specifically occurs over the gene bodies of DEGs, consistent with the recent finding that mCA serves to fine-tune gene expression of lowly expressed genes\(^{19} \).

**Differential epigenomics identify transcription factors driving tissue-specific gene expression.** We hypothesized that neuronal DARs and CG-DMRs could identify transcription factor (TF) binding motifs actively involved in regulating brain region–specific neuronal function. Using Haystack\(^{37} \), we find that these regions were enriched in immediate-early genes (IEGs), a class of TFs influenced by synaptic activity with important roles in regulating neuronal function\(^{38} \) (Fig. 4d and Supplementary Table 22). Several of these IEGs have been implicated in schizophrenia\(^{39} \) and bipolar disorder\(^{40} \) or play known roles in addiction\(^{41} \). DNA methylation influences binding of a diverse set of TFs\(^{42} \), many of which were enriched in regions of differential methylation and accessibility (Fig. 4d). Additionally, many TFs whose motifs were enriched were also differentially expressed between NAcc and BA9 neurons (Fig. 4d). When we restrict our analysis to those neuronal DARs and CG-DMRs that overlap promoters, we again detect enrichment for IEGs, specifically those encoding AP-1 family members (\( JUN, FOS, JDP2 \)) (Fig. 4e).
Using our neuronal DARs, we employed a method that uses the accessibility of known TF binding sites in gene promoters and enhancers to calculate the importance of each TF for the observed gene expression profile. A positive regression coefficient indicates that the presence of a binding site for a particular TF predicts the target gene to be upregulated in NAcc versus BA9, and vice versa (Fig. 4f; and Supplementary Table 23). MECP2, in particular, has a negative correlation coefficient (Fig. 4f) and is recruited to gene bodies by mCA to restrain transcription of lowly expressed genes. Several TFs identified are known regulators of neuronal function (NFIX, MECP2, ARX, DLXI, ZBTB33/KAISO), whereas others have not been previously implicated in adult neuronal regulation (TEAD2, ARID5A/B, ZNF354C/hKID3) (Fig. 4f). Eight of these TFs were also differentially expressed between NAcc and BA9 (Fig. 4). Taken together, our data show that differential methylation and accessibility mark regions of the genome that
Fig. 4 | Differential gene expression is strongly associated with neuronal CA-DMRs. Regions where neuronal CG-DMRs and DARs intersect are enriched in binding sites for TFs associated with synaptic activity. a. Proportion of protein-coding genes (n = 19,823) with a differential epigenetic feature in each bin around gene bodies. Each gene length is split into 100 bins, and data extend 2 gene-lengths upstream and downstream. DEGs are indicated. b, Absolute Pearson correlation values of differential epigenetic features with protein-coding gene expression across scaled gene bodies, with each gene length split into 100 bins (0–100%); data extend 1Mb up- and downstream in 1kb bins. The sample for the Pearson correlation in each bin is the subset of features that overlap that bin, and, therefore, the sample size for which Pearson’s correlations were determined varies for each bin and feature. c, Estimated probability (with 95% CI) that a gene is differentially expressed in terms of the proportion of the gene covered by differential epigenetic features (Methods). Example DEGs are annotated. d, Enrichment of TFs whose motifs were enriched in DARs that overlap hypo- (n = 11,734) or hypermethylated CG-DMRs (n = 14,463) in NAcc compared with BA9. Expression (log2(fold change)) of each TF is shown, and differentially expressed TFs in NAcc versus BA9 are shown by red bars (see Methods for description of RNA-seq analysis). TFs whose binding is influenced by DNA methylation (as determined in ref. 14) are indicated with blue bars (MS, methyl-sensitive). e, As in d, showing TFs with motifs enriched where DARs/CG-DMRs overlap promoters (hyper, n = 2,618; hypo, n = 1,435). Significance determined using one-sided Fisher’s test in d and e, f, Bar plot showing the predicted impact of TFs on gene expression (normalized coefficient) for each TF with a binding site within 5 kb of a gene. Large values denote a higher impact of the TF on differential gene expression. Boldface type indicates that the TF is differentially expressed between NAcc and BA9.

are regulated by synaptic activity–responsive TFs and suggest that these TFs regulate different targets in neurons from distinct brain regions.

Differential epigenomics identify regions of genetic importance for psychiatric disorders and behavioral-cognitive traits. The regions we identified using differential epigenomics, particularly
the CG-DMRs and DARs, have characteristics typical of regulatory elements, as expected\textsuperscript{21,22,36}. Multiple studies have provided evidence that disease-associated genetic variation is enriched in regulatory elements active in physiologically relevant cell types\textsuperscript{45,46}. We therefore asked whether our differential epigenetic features were linked with neurological, psychiatric, and behavioral-cognitive phenotypes (brain-linked traits).

We used stratified linkage disequilibrium score regression (SLDSR\textsuperscript{19}) to identify genomic features strongly associated with brain-linked traits. SLDSR partitions the heritability of a trait across a set of overlapping genomic features using summary statistics from a genome-wide association study (GWAS). We considered eight brain-derived genomic features: five ‘differential’ features identified in our analyses described above and three previously published ‘non-differential’ features (Methods). For each feature–trait combination, SLDSR reports a ‘z score’, indicating whether the feature explains heritability beyond that explained by other features in the model, and an ‘enrichment score’ relating the heritability explained by the feature relative to that expected for a feature containing an equal number of SNPs. At least one of the eight brain-derived features explained a significant proportion of heritability for 13 of 27 brain-linked traits, whereas none explained a significant proportion of heritability for any negative control trait (Fig. 5b, Supplementary Fig. 9, Supplementary Tables 24 and 25). The neuronal CG-DMRs have much higher enrichment scores than any of the non-differential features (Fig. 5a), explaining additional heritability in six out of 13 traits, including schizophrenia and neuroticism, although their size (20–40\times smaller) results in more uncertain estimates (Supplementary Fig. 9b).

We then performed a more stringent analysis of the five differential features by testing them against a baseline model that included the three non-differential brain-specific features. We found that when using this stringent approach, only CG-DMRs still contributed significantly (with a 14-fold enrichment) to the explained heritability of a brain-linked trait (schizophrenia; z-score adjusted \(P\) value = 0.013; Fig. 5c,d; Supplementary Fig. 10a,b).

Finally, we removed regions found by our differential approach from the non-differential features and repeated the analysis. This ensured that the heritability associated with a region common to two sets of brain-specific features is exclusively assigned to the differential feature. This approach highlights the trade-offs of using the highly specific differential features compared with the more general maps of brain regulatory regions, as evidenced by the much larger size of the non-differential features. On the basis of this analysis, the neuronal CG-DMRs are significantly associated with the heritability of schizophrenia, ADHD, BMI, IQ, and neuroticism (Fig. 5f and Supplementary Fig. 11a), with a 10- to 16-fold enrichment (Fig. 5e and Supplementary Fig. 11b).

Together, these data support the hypothesis that the genetic signal associated with neuropsychiatric traits is mediated through epigenetically distinct regions among neurons from diverse brain regions, particularly CG-DMRs. These analyses further demonstrate the power of using differential methods to precisely identify key regulatory regions, while also highlighting that,
from a genetic perspective, CH-DMRs may be less interesting than CG-DMRs.

Discussion

This work represents the most comprehensive data set to date of WGBS across neuronal and non-neuronal populations and bulk tissues from multiple human brain regions. We have identified 12 Mb of differential mCG and 40 Mb of differential mCH among neuronal nuclei isolated from four brain regions implicated in neuropsychiatric disorders. We have further correlated these methylation changes to changes in both chromatin accessibility (118 Mb of differential accessibility) and gene expression (2,952 DEGs) in two of the tissues examined (NAcc and BA9). These data are made available as the ‘BrainEpigenomeHub’ UCSC track hub (see Data Availability) that can be used as a resource of normal epigenetic states and variation in the human brain to advance neuroepigenetics.

Our findings have four main implications important for the field of neuroepigenetics. First, differential methylation among brain regions appears to be driven by the neuronal cell population. While the non-neuronal fraction can be divided into distinct subpopulations, we show through our analysis of sample-to-sample variability that this heterogeneity cannot account for the remarkable consistency of non-neuronal methylation across the human brain. Importantly, we have demonstrated that the ratio of neuronal to non-neuronal nuclei is highly variable, both among brain regions and between samples taken from a single brain region, even within a single individual. While the need to account for cellular heterogeneity is widely appreciated, these results should be of immense value in improving and benchmarking current computational deconvolution strategies applied to studies of human brain tissues.

Second, this study illustrates the power and specificity of differential epigenetics across brain regions, particularly the ability of differential mCG to identify regulatory elements and the ability of differential mCH to identify differential transcription. Whereas differential chromatin accessibility alone showed little enrichment in any particular genome feature, the specificity for regulatory elements (particularly for regions marked by brain-specific H3K27ac) increased nearly fivefold when combined with differential mCG.

Third, we find that regions of differential mCG and chromatin accessibility in neurons are enriched in binding sites for transcription factors regulated by synaptic activity (for example, FOS, JUN, MEF2C) (reviewed in43), several possessing methylation sensitivity44. These data are consistent with a link between synaptic activity and epigenetic modification of transcription factor binding sites. Additionally, some transcription factors we identified play known roles in neurodevelopment, but have not previously been implicated in adult neuronal function (for example, HMGAI). These findings further illustrate the importance of differential epigenetic analysis in understanding normal neuronal function and provide novel targets for further investigation.

Finally, we show that tissue-specific differential mCG in the neuronal population has a fivefold greater enrichment for the heritability of neuropsychiatric diseases than the more generally defined regulatory genomic fraction identified in the human brain by ChromHMM and is contained within only 12 Mb of the human genome. While the majority of neuronal CG-DMRs (9,033/13,074; 69%) are found within genes, they primarily occur in intronic regions (7,669; 58.6%) that frequently harbor regulatory elements. Importantly, many of these differentially methylated genes have been repeatedly implicated in schizophrenia and other neuropsychiatric disorders. For example, we identified differential methylation in 19 of 36 schizophrenia GWAS-derived genes analyzed in a previous expression study45 (CACNA1C, CACNB2, CACNAl1, GPM6A, GRAMD1B, SATB2, MEF2C, GRIN2A, MAD1L1, BCL11B, TCF4, TLE1, TLE3, PODXL, ZNF536, KCNV1, MMP16, MAN2A1, and GALNT10). Further, CG-DMRs are present in 56 (Supplementary Table 26) of the 237 genes recently shown to have differentially expressed features in prefrontal cortex schizophrenia versus control samples46. Clearly, these 12 Mb of regional differential methylation constitute critical sequences for future epigenetic analyses of neuropsychiatric diseases. It will be interesting to determine whether a subset of these regions are particularly associated with individual disorders and whether these associations are brain region-specific. These observations are consistent with the idea47 that tissue-specific epigenetic patterning is frequently disrupted in human disease.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41593-018-0297-8.

Received: 7 May 2018; Accepted: 13 November 2018; Published online: 14 January 2019

References

1. Hoffmann, A., Sportelli, V., Ziller, M. & Spengler, D. Epigenomics of major depressive disorders and schizophrenia: Early life decides. Int. J. Mol. Sci. 18, E1711 (2017).
2. Negi, S. K. & Guda, C. Global gene expression profiling of healthy human brain and its application in studying neurological disorders. Sci. Rep. 7, 897 (2017).
3. Mo, A. et al. Epigenomic signatures of neuronal diversity in thalamocortical nuclei. Neuron 86, 1369–1384 (2015).
4. Marques, S. et al. Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. Science 352, 1326–1329 (2016).
5. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science 352, 1586–1590 (2016).
6. Tasic, B. et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. 19, 335–346 (2016).
7. Zeisel, A. et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science 347, 1388–1442 (2015).
8. Luo, C. et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. Science 357, 600–604 (2017).
9. Davies, M. N. et al. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genom. Biol. 13, R43 (2012).
10. Illingworth, R. S. et al. Inter-individual variability contrasts with regional homogeneity in the human brain DNA methylome. Nucleic Acids Res. 43, 732–744 (2015).
11. Ladd-Acosta, C. et al. DNA methylation signatures within the human brain. Am. J. Hum. Genet. 81, 1304–1315 (2007).
12. Viana, I. et al. Schizophrenia-associated methylation variation: molecular signatures of disease and polygenic risk burden across multiple brain regions. Hum. Mol. Genet. 26, 210–225 (2017).
13. Watson, C. T. et al. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer’s disease. Genome Biol. 8, 5 (2016).
14. Kaut, O. et al. Aberrant NMDA receptor DNA methylation detected by epigenome-wide analysis of hippocampus and prefrontal cortex in major depression. Eur. Arch. Psychiatry Clin. Neurosci. 265, 331–341 (2015).
15. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015).
16. Jaffe, A. E. et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. Nat. Neurosci. 19, 40–47 (2016).
17. Ellis, S. E., Gupta, S., Moes, A., West, A. B. & Arkling, D. E. Exaggerated CpH methylation in the autism-affected brain. Mol. Autism 8, 6 (2017).
18. Mo, A. et al. Epigenomic landscapes of retinal rods and cones. eLife 5, e11613 (2016).
19. Stroud, H. et al. Early-life gene expression in neurons modulates lasting epigenetic states. Cell 171, 1151–1164.e16 (2017).
20. Kozenkov, A. et al. Substance DNA methylation differences between two major neuronal subtypes in human brain. Nucleic Acids Res. 44, 22593–22612 (2016).
21. Kozenkov, A. et al. Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. Nucleic Acids Res. 42, 109–127 (2014).
22. Lister, R. et al. Global epigenomic reconfiguration during mammalian brain development. Science 341, 1237905 (2013).
23. Sanchez-Mat, J. V. et al. Human DNA methylomes of neurodegenerative diseases show common epigenomic patterns. Transl. Psychiatry 6, e718 (2016).
24. Morel, L. et al. Molecular and functional properties of regional astrocytes in the adult brain. *J. Neurosci.* **37**, 8706–8717 (2017).
25. von Bartheld, C. S., Balbey, J. & Herculano-Houzel, S. The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *J. Comp. Neurol.* **524**, 3865–3895 (2016).
26. Hansen, K. D., Langmead, B. & Irizarry, R. A. BSsmooth: from whole genome bisulfitesequencing reads to differentially methylated regions. *Genome Biol.* **13**, R83 (2012).
27. Montaño, C. M. et al. Measuring cell-type specific differential methylation in human brain tissue. *Genome Biol.* **14**, R94 (2013).
28. Vermunt, M. W. et al. Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Rep.* **9**, 767–779 (2014).
29. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
30. Hansen, K. D. et al. Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* **43**, 768–775 (2011).
31. Hansen, K. D. et al. Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. *Genome Res.* **24**, 177–184 (2014).
32. McLean, C. Y. et al. GREAT greatly improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
33. Guo, J. U. et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* **17**, 215–222 (2014).
34. Ziller, M. J. et al. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet.* **7**, e1002389 (2011).
35. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 1–9 (2015).
36. Fullard, J. F. et al. Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci. *Hum. Mol. Genet.* **26**, 1942–1951 (2017).
37. Pinello, L., Faroumi, R. & Yuan, G. C. Haystack: systematic analysis of the variation of epigenetic states and cell-type specific regulatory elements. *Bioinformatics* **34**, 1930–1933 (2018).
38. Fukuchi, M. & Tsuda, M. Convergence of neurotransmissions at synapse on IEG regulation in nucleus. *Front. Biosci. (LandmarkEd.)* **1052–1072 (2017).
39. Hu, T. M., Chen, C. H., Chuang, Y. A., Hsu, S. H. & Cheng, M. C. Resequencing of early growth response 2 (EGR2) gene revealed a recurrent patient-specific mutation in schizophrenia. *Hum. Mol. Genet.* **22**, 958–960 (2015).
40. Pfaffenseller, B. et al. Differential expression of transcriptional regulatory units in the prefrontal cortex of patients with bipolar disorder: potential role of early growth response gene 3. *Transl. Psychiatry* **6**, e805 (2016).
41. Larson, E. B. et al. Striatal regulation of FosB, FosB, and cFos during cocaine self-administration and withdrawal. *J. Neurochem.* **115**, 112–122 (2010).
42. Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).
43. Durek, P. et al. Epigenomic profiling of human CD4+ T cells supports a linear differentiation model and highlights molecular regulators of memory development. *Immunity* **45**, 1148–1161 (2016).
44. Schmidt, F. et al. Combining transcription factor binding affinities with open-chromatin data for accurate gene expression prediction. *Nucleic Acids Res.* **45**, 54–66 (2017).
45. Forteza, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
46. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
47. Okuno, H. Regulation and function of immediate-early genes in the brain: beyond neuronal activity markers. *Neurosci. Res.* **69**, 175–186 (2011).
48. Hertzberg, L., Katsel, P., Roussos, P., Haroutunian, V. & Domaney, E. Integration of gene expression and GWAS results supports involvement of calcium signaling in schizophrenia. *Schizophr. Res.* **164**, 92–99 (2015).
49. Jaffe, A. E. et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat. Neurosci.* **21**, 1117–1125 (2018).
50. Feinberg, A. P. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**, 433–440 (2007).
51. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).

**Acknowledgements**

This work was supported by funding awarded to A.P.F. (U01MH104393) through the enhanced Genotype-Tissue Expression (eGTEx) project supported by the Common Fund of the Office of the Director of NIH. This work was supported by National Cancer Institute under U24CA180996. We would like to thank H. Zhang from the Flow Cytometry Cell Sorting Core Facility at Johns Hopkins School of Public Health for flow sorting. The core facility is supported by CFAR: 5P30AI094189–04, 1S10OD016315–01, and 1S10RR13777001. Brain tissues were received from the NIH NeuroBioBank at the University of Maryland and University of Pittsburgh.

**Author Contributions**

L.F.R, K.D.H., and A.P.F. designed the study; L.F.R. performed nuclei sorting, DNA and RNA extractions; V.R.D. performed ATAC-seq; R.T., A.I., C.M.C. performed WGBS and RNA-seq library preparation and sequencing; A.P.F. oversaw the experiments; K.D.H. oversaw the data analysis. L.F.R., P.F.H., K.D.H., and A.P.F. performed data analysis and interpreted the results; L.F.R, P.F.H., K.D.H, and A.P.F. wrote manuscript.

**Competing Interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-018-0297-4.

Reprints and permissions information is available at www.nature.com/reprints.
Methods
Experimental Methods. Human postmortem brain samples. Fluorescence-activated nuclear sorting (FANS) was performed on flash-frozen postmortem dorsolateral prefrontal cortex (BA9), hippocampus (HC), nucleus accumbens (NAcc), and anterior cingulate gyrus (BA24) from six individuals not affected with neurological or psychiatric disease. These samples underwent nuclei extraction and sorting as described below for subsequent DNA methylation analysis. Additionally, neuronal nuclei were isolated from the NAcc and BA9 of six individuals for RNA-seq and ATAC-seq analysis. To underscore the importance of cell sorting, we also prepared DNA from bulk tissue from the four brain regions (BA9, n = 9; HC, n = 7; NAcc, n = 7; BA24, n = 5). The majority of individuals were matched between sorted and bulk tissues, but not all. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications.

All samples were obtained from the University of Maryland Brain and Tissue Bank, which is a Brain and Tissue Repository of the NIH NeuroBioBank (Supplementary Table S1). We have followed all relevant ethical regulations. The description here was classified as not human subjects research (NHSCR) by the Johns Hopkins Institutional Review Board (IRB00061004). Data collection and analyses were not performed blind to tissue of origin and randomization was performed at the library preparation stage.

Nuclei extraction, FANS, and DNA isolation. Total nuclei were extracted via sucrose gradient centrifugation as previously described\(^2\) with the following changes. For WGBS analysis, a total of 2 x 250 mg of frozen tissue per sample was homogenized in 5 mL of lysis buffer (0.32 M sucrose, 10 mM Tris, pH 8.0, 0.5 mM CaCl\(_2\), 3 mM Mg acetate, 1 mM DT T, 0.1 mM EDTA, 0.1% Triton X-100) by bouncing 50 times in a 40 mL sonication homogenizer. Lysates were centrifuged (10,000 rpm, 15 minutes) with a 100 mL ultracentrifugation tube, and 18 mL of sucrose solution (1.8 M sucrose, 10 mM Tris, pH 8.0, 0.3 mM Mg acetate, 1 mM DT T) was dispensed to the bottom of the tube. The samples were then centrifuged at 28,600 r.p.m. for 2 h at 4 °C (Beckman Optima XE-90; SW32 Ti rotor). After centrifugation, the supernatant was removed by aspiration, and the nuclear pellet was resuspended in 500 uL of 1x staining mix (2% normal goat serum, 0.1% BSA, 1:500 anti-NeuN conjugated to AlexaFluor488 (Millipore, cat#: MAB377X) in PBS) and incubated on ice. Unstained nuclei and nuclei stained with only secondary antibody served as negative controls.

The fluorescent nuclei were run through a Beckman Coulter MoFlo Cell Sorter with proper gate settings using Summit v4.3.02 software (Supplementary Fig. 1). A small portion of the NeuN\(^+\) and NeuN\(^-\) nuclei were rerun on the sorter to validate the purity, which was greater than 95%. Immunonegative (NeuN\(^-\)) and immunopositive (NeuN\(^+\)) nuclei were collected in parallel. For DNA extraction, sorted nuclei were pelleted by adding 2 mL of sucrose solution, 50 uL of 1 M CaCl\(_2\), and 30 uL of Mg acetate to 10 mL of nuclei in PBS. This solution was incubated on ice for 15 minutes, then the supernatant was removed at 3000 uL for 20 min. The nuclear pellets were flash frozen in liquid nitrogen and stored at −80 °C. DNA was extracted from the frozen nuclear pellets using the MasterPure DNA Extraction kit (cat#: MCRS5200, Epicentre, Madison, Wisconsin, USA) following the manufacturer's instructions. For ATAC-seq and RNA-seq, nuclei were processed as described below.

Whole-genome bisulfite sequencing (WGBS). WGBS single indexed libraries were generated using NEBNext Ultra DNA library Prep kit for Illumina (cat#: E7370L, New England BioLabs, Ipswich, MA, USA) according to the manufacturer's instructions with modifications. 400 ng gDNA was quantified by Qubit dsDNA BR assay (cat#: Q32853, Invitrogen, Carlsbad, CA, USA), and 1% unmethylated lambda DNA (cat#: D1521, Promega, Madison, WI, USA) was spiked in to measure bisulfite conversion efficiency. Samples were fragmented to an average insert size of 350 bp using a Covaris S2 sonicator. Size selection was performed using a New England BioLabs, Ipswich, MA, USA) according to the manufacturer's instructions. Following the manufacturer's instructions, amplification was performed after the bisulfite conversion using Kapa HiFi Uracil (+) (cat#: KK282, Kapa Biosystems, Boston, USA) polymerase using the following cycling conditions: 98 °C 45 s/8 cycles; 98 °C 15 s, 65 °C 30 s, 72 °C 30 s/72°C 1 min. Final libraries were run on a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) High-Sensitivity DNA assay; samples were also run on a Bioanalyzer after shearing and size selection for quality control purposes. Libraries were quantified by qPCR using the Library Quantification Kit for Illumina sequencing platforms (cat#: KK4824, KAPA Biosystems, Boston, USA), using 7900HT Real Time PCR System (Applied Biosystems). Libraries were sequenced with the Illumina HiSeq4000 using 75 bp paired-end single indexed run with a 5% PhiX spike-in.

RNA sequencing. RNA isolated from bulk tissue was assessed, and only tissues with an RNA integrity number (RIN) ≥ 2 were used for nuclei isolation. NeuN\(^+\) and NeuN\(^-\) nuclei were isolated as previously described, with the addition of 20 u/mL RNase inhibitors (cat#: N8080119, Applied Biosystems) to the lysis buffer, sucrose solution, and antibody solution, and protease inhibitor cocktail (cat#: 50–751–7359, Amresco). This was added to the lysis buffer only. Approximately 200,000 nuclei were sorted directly into RLT buffer+ 150 mM 2-mercaptoethanol, and RNA isolated using the Qiagen RNeasy Kit (cat #74106, Qiagen, Valencia, CA, USA). Nuclear RNA quality was assessed by running samples on a Total RNA Pico Chip on a 2100 BioAnalyzer (Agilent, Santa Clara, CA, USA). RNA-seq libraries were created using 2.5 ng input RNA with the SMARTer Stranded Total RNA-Seq Kit - Pico Input Mammalian (cat#: E53605, Takara Bio, Mountain View, CA, USA) following the manufacturer's instructions for degraded RNA samples. Libraries were sequenced with the Illumina HiSeq4000 using 75 bp paired-end single indexed run with 5% PhiX spike-in.

Computational Methods. Annotation. The hg19 build of the human reference genome and UCSC Genome Browser were used for all analyses of annotation. Genes, exons, introns, and UTRs were taken from GENCODE v19 (https://www.gencodegenes.org/human/release_19.html). Gene bodies were defined by taking the union over all transcripts (transcription start site to transcription end site) for each gene. Promoters were defined as 4 kb centered on the transcription start site. CpG islands data were downloaded from UCSC (http://genome.ucsc.edu)\(^3\). A Granges object was defined creating CpG shores as 2 kb flanking CpG islands and CpG shelves as 2 kb flanking CpG island shores. 'Open sea regions' are parts of the genome that are not CpG islands, shores or shelves.

The 15-state ChromHMM model for seven adult brain tissues (E071, E074, E068, E069, E072, E067, E073) from the Roadmap Epigenomics Project\(^4\) was downloaded using the R/Bioconductor AnnotationHub package (v2.6.4).

Mapping and quality control of whole genome bisulfite sequencing (WGBS). We trimmed reads of their adaptor sequences using Trim Galore! (v0.4.0) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and quality-trimmed the following the following parameters: "trim_galore –q 25 –paired |$\text{READ1}$ |$\text{READ2}$"). We then aligned these trimmed reads to the hg19 build of the human genome (including autosomes, sex chromosomes, mitochondrial sequence, and lambda phage (accession NC_001416.1) but excluding non-chromosomal sequences) using Bismark\(^5\) with the following alignment parameters: "bismark–bowtie2 -X 1000–1 |$\text{READ1}$ –2 |$\text{READ2}$". Supplementary Tables 2 and 3 summarize the alignment results. Using the reads aligned to the lambda phage genome, we estimated that all libraries had a bisulfite conversion rate > 99%.

We then used bismark_methylation_extractor to summarize the number of reads supporting a methylated cytosine and the number of reads supporting an unmethylated cytosine for every cytosine in the reference genome. Specifically, we first counted and visually inspected the M-bias\(^6\) of our libraries. Based on these results, we decided to ignore the first 5 bp of read1 and the first 10 bp of read2 in the subsequent call to bismark_methylation_extractor with parameters: "--ignore 5 –ignore_r2 10". The final cytosine report file summarizes the methylation evidence at each cytosine in the reference genome.

Smoothing WGBS. We used BSmooth to estimate Cpg methylation levels as previously described\(^7\). Specifically, we ran a 'small' smooth to identify small DMRs (smoothing over windows of at least 1 kb containing at least 20 Cpgs) and a 'large' smooth to identify large-scale blocks (smoothing over windows of at least 20 kb containing at least 500 Cpgs). Following smoothing, we analyzed all Cpgs that passed a sequencing coverage of at least 1 in all samples (n=45 for sorted data, n=7 for bulk data).

We also adapted BSmooth to estimate CpA and CpT methylation levels in NeuN\(^-\) samples. Unlike CpGs, CpAs and CpTs are not palindromic, so the analysis was performed separately for each strand, for a total of 4 strand/dinucleotide combinations:

- mCA (forward strand)
- mA (reverse strand)
- mCT (forward strand)
- mTC (reverse strand)

For each dinucleotide/strand combination we ran a single 'smallish' smooth to identify DMRs (smoothing over windows of at least 3 kb containing at least 20 Cpas or CtTs). Following smoothing, we analyzed all Cpas and CpTs regardless of sequencing coverage.

Identification of small DMRs and large-scale blocks. We ran separate analyses to identify 6 types of differentially methylated regions:
1. CG-DMRs: Using data from the ‘small’ smooth of CpG methylation levels
2. CG-blocks: Using data from the ‘large’ smooth of CpG methylation levels
3. CA-DMRs (forward strand): Using data from the ‘small-ish’ smooth of CpA methylation levels on the forward strand
4. CA-DMRs (reverse strand): Using data from the ‘small-ish’ smooth of CpA methylation levels on the reverse strand
5. CT-DMRs (forward strand): Using data from the ‘small-ish’ smooth of CpT methylation levels on the forward strand
6. CT-DMRs (reverse strand): Using data from the ‘small-ish’ smooth of CpT methylation levels on the reverse strand

Previously we have used BSsmooth to perform pairwise (two-group) comparisons. In the present study, we used 8 groups to compare: 4 brain regions (BA9, BA24, HC, NAcc) and, for the sorted data, 2 cell types (NeuN+ and NeuN−). However, it does not tell us which group(s) are hypomethylated or hypermethylated.

Annotation of small CG-DMRs and CG-blocks.

The F-statistic approach allows us to perform a pairwise comparison for each candidate DMR/block, in how many permutations did we see a null DMR/block? We performed 1000 such permutations. We then asked, for each cell type (NeuN+ and NeuN−), we repeated the F-statistic analysis using just the NeuN+ cells.

To perform the differential analysis, we first took the union of condition-specific OCRs on the autosomes to construct an ‘overall’ set of OCRs. This ‘overall’ set of OCRs contained 853,053 regions (630 Mb). For each sample, we then counted the number of fragments (fragment = start of read to end of read) overlapping each of the ‘overall’ OCRs using the summarizeOverlaps() function in the GenomicAlignments R/Bioconductor package. We took the ‘narrowPeaks’ produced by MACS and filtered out those regions overlapping the ENCODE mappability consensus blacklist regions.

Identifying differentially accessible ATAC-seq regions (DARS). Peaks were called in each condition (NAcc_pos, NAcc_neg, BA9_pos, and BA9_neg) using MACS (v2.1.0).

To perform the differential analysis, we first took the union of condition-specific OCRs on the autosomes to construct an ‘overall’ set of OCRs. This ‘overall’ set of OCRs contained 853,053 regions (630 Mb). For each sample, we then counted the number of fragments (fragment = start of read to end of read) overlapping each of the ‘overall’ OCRs using the summarizeOverlaps() function in the GenomicAlignments R/Bioconductor package. We took the ‘narrowPeaks’ produced by MACS and filtered out those regions overlapping the ENCODE mappability consensus blacklist regions.

identifying differentially expressed ATAC-seq regions (DARS). Peaks were called in each condition (NAcc_pos, NAcc_neg, BA9_pos, and BA9_neg) using MACS (v2.1.0). Specifically, data from each condition were combined into a metasample formed by using all non-duplicate-marked reads with a mapping quality > 30 and no PCR/nuclease–peak overlaps. All summits - t $(BAMS)$ were used for comparison with our CG-DMRs.

To perform the differential analysis, we first took the union of condition-specific OCRs on the autosomes to construct an ‘overall’ set of OCRs. This ‘overall’ set of OCRs contained 853,053 regions (630 Mb). For each sample, we then counted the number of fragments (fragment = start of read to end of read) overlapping each of the ‘overall’ OCRs using the summarizeOverlaps() function in the GenomicAlignments R/Bioconductor package (v1.10.0). Specifically, we only counted those fragments whose reads had a mapping-quality score > 30, reads not marked as potential PCR duplicates, and those whose position on the fragment overlapped exactly one peak.

We then analyzed these data using the voom method, originally designed for differential expression analysis of RNA-seq data. Briefly, the read counts were transformed to counts per million (cpm) and only those 283,812 / 853,053 peaks with at least 1 cpm for at least 5 samples (the size of the smallest group of samples) were retained. These 283,812 peaks were used in all downstream analyses of differential accessibility described in the main text. We normalized these counts using tximport (v1.2.0), then used edgeR (v3.16.5) and limma (v3.30.7) to transform these counts to log-cpm, estimate the mean-variance relationship, and compute appropriate observation-level weights ready for linear modelling.

In our design matrix, we blocked on donor (donor1, …, donor6) and included a term for each group (for example, BA9_neg for NeuN+ cells from BA9, BA9_pos for NeuN− cells from BA9, etc.). We can use linear models to transform these counts to log-cpm, estimate the mean-variance relationship, and compute appropriate observation-level weights ready for linear modelling.

Identifying differentially expressed genes (DEGs). We used tximport to compute normalized gene-level counts from the transcript-level abundance estimates (scaling these using the average transcript length over samples and the library size) for autosomal genes (33,351 genes). Only autosomal genes with at least 1 cpm in at least 4 libraries (the size of the smallest group of samples) were retained for downstream statistical analysis (24,161 / 33,351 genes). We normalized these counts using TMM (v1.10.1) and edgeR (v3.16.5) and limma (v3.30.7) to transform these counts to log-cpm, estimate the mean-variance relationship, and compute appropriate observation-level weights ready for linear modelling.
In our design matrix, we blocked on donor (donor1, …, donor6) and included a term for each group (for example, BA9_neg for NeuN+ cells from BA9, BA9_pos for NeuN- cells from BA9, etc.). We ran surrogate variable analyses1 using the sva (v3.22.0) R/Bioconductor package and identified 5 surrogate variables, some of which correlated with the date on which these samples were flow-sorted. We ultimately decided to include all 5 surrogate variables in the linear model. Using the empirical Bayes shrinkage method implemented in limma, we tested for differential expression of genes in three comparisons: (1) NAcc vs. BA9 in NeuN+ cells; (2) NAcc vs. BA9 in NeuN- cells; (3) NeuN+ cells vs NeuN- cells. For a gene to be called a differentially expressed gene (DEG), it had to have a Benjamini–Hochberg adjusted P-value < 0.05 with no minimum log fold change cutoff.

**Enrichment odds ratios and P-values.** We formed a 2 × 2 contingency table of (n11, n12, n21, n22), specific values of (n11, n12, n21, n22) are described below. The enrichment log odds ratio was estimated by log(OR) = log(n11) - log(n21) - log(n12) - log(n22), its standard error was estimated by se(\log(OR)) = sqrt(1/n11 + 1/n12 + 1/n21 + 1/n22), and an approximate 95% confidence interval formed by [\log(OR) - 2 \times se(\log(OR)), \log(OR) + 2 \times se(\log(OR))]. We also report the P-value obtained from performing Fisher’s exact test for testing the null of independence of rows and columns in the 2 × 2 table (that is the null of no enrichment or depletion) using the fisher.test() function from the ‘stats’ package in R.

**Enrichment of DMRs and blocks in genomic features.** For DMRs and blocks, we computed the enrichment of CpXs (CpGs, CpAs, or CpTs) as appropriate) inside each genomic feature (for example, exons, FANTOM5 enhancers, etc.). Specifically, for each genomic feature, we constructed the 2 × 2 table (n11, n12, n21, n22), where:

- n11 = Number of CpXs in DMRs/blocks that were inside the feature
- n12 = Number of CpXs in DMRs/blocks that were outside the feature
- n21 = Number of CpXs not in DMRs/blocks that were inside the feature
- n22 = Number of CpXs not in DMRs/blocks that were outside the feature

The total number of CpXs, \( n = n_1 + n_2 + n_1 + n_2 \), was the number of autosomal CpXs in the reference genome. We counted CpXs rather than the number of DMRs or bases because this accounts for the non-uniform distribution of CpXs along the genome and avoids double-counting DMRs that are both inside and outside the feature.

**Enrichment of open chromatin regions (OCRs) in genomic features.** For OCRs, we computed the enrichment of bases within OCRs inside each genomic feature. Specifically, for each genomic feature, we constructed the 2 × 2 table (n11, n12, n21, n22), where:

- n11 = Number of bases in OCRs that were inside the feature
- n12 = Number of bases in OCRs that were outside the feature
- n21 = Number of bases in the rest of the genome that were inside the feature
- n22 = Number of bases in the rest of the genome that were outside the feature

The total number of bases, \( n = n_1 + n_2 + n_1 + n_2 \), was the number of autosomal bases in the reference genome. We counted bases rather than number of OCRs to account for the variation in OCR width, the large variation in width for the ‘rest of the genome’ features, and to avoid double-counting OCRs that were both inside and outside the feature.

**Enrichment of differentially accessible regions (DARs) in genomic features.** For DARs, we computed the enrichment of bases within DARs inside each genomic feature in two ways. First, for each genomic feature, we constructed the 2 × 2 table (n11, n12, n21, n22), where:

- n11 = Number of bases in DARs that were inside the feature
- n12 = Number of bases in DARs that were outside the feature
- n21 = Number of bases in the rest of the genome that were inside the feature
- n22 = Number of bases in the rest of the genome that were outside the feature

Second, for each genomic feature, we constructed the 2 × 2 table (n11, n12, n21, n22), where:

- n11 = Number of bases in DARs that were inside the feature
- n12 = Number of bases in DARs that were outside the feature
- n21 = Number of bases in null-regions that were inside the feature
- n22 = Number of bases in null-regions that were outside the feature

‘Null-regions’ were those OCRs that were not differentially accessible between the relevant condition (NAcc and BA9 in NeuN+ cells) based on the peak having a Benjamini–Hochberg adjusted P-value > 0.05 in the analysis of differential accessibility. By comparing to null-regions rather than the rest of the genome, we account for the non-uniform distribution of OCRs along the genome.

We counted the number of bases rather than the number of DARs to account for the variation in DAR width and to avoid double-counting DARs that were both inside and outside the feature.

**Association of gene body methylation and chromatin accessibility with gene expression.** We analyzed the relationship between gene body methylation and chromatin accessibility with gene expression using protein coding genes (n = 19,823). We focused on data from NAcc (NeuN+) and BA9 (NeuN-) samples because for these brain regions we had WGBS, ATAC-seq, and RNA-seq, as well as a substantial number of ‘DiffEpi’ marks (DiffEpi being a collective abbreviation for DMRs, blocks, and DARs).

To examine the spatial distribution of DiffEpi marks around gene bodies, we took 100 bins across each gene and recorded whether that bin overlapped a DiffEpi mark. We extended this for 2 gene body equivalents upstream of the transcription start site (TSS) and downstream of the transcription end site (TES) (that is if a gene was 1 kb long then we extended it 2 kb upstream of TSS and 2 kb downstream of TES) and similarly recorded whether each bin overlapped a DiffEpi mark. Fig. 4a plots the proportion of genes with a DiffEpi mark in each bin as we move from upstream of the TSS, across the gene body, and downstream of the TES, stratified by whether the gene was differentially expressed between NAcc (NeuN+) and BA9 (NeuN-) samples.

We performed various analyses of the relationship between mCA (taking estimates from the same strand as the gene), mCG (aggregated over strand and so ‘unstranded’), and average chromatin accessibility (ATAC-seq reads-per-kilobase mapped (RPKM)) with gene expression (RNA-seq RPKM) for all protein coding genes for NAcc (NeuN+) and BA9 (NeuN-) samples. We examined these features over gene bodies and gene promoters, as well as in bins along and surrounding each gene. Different types of bins were used for different analyses, as noted in figure captions. For some analyses we used a fixed number of bins for each gene (so that bin width varied according to gene width for example, 100 bins covering the gene body in Fig. 4a and Supplementary Fig. 8a). For the rest of the analyses, we used a fixed bin size (for example, 1 kb bins upstream of TSS and downstream of TES in Fig. 4b). Furthermore, some analyses used scaled distances upstream of the TSS and downstream of the TES (for example, Fig. 4a and Supplementary Fig. 8a) while others used fixed distances (for example, Fig. 4b).

We then focused on correlating DiffEpi marks with changes in gene expression. We analyzed all protein coding genes and created 100 bins across each gene body. We identified all DiffEpi marks that overlapped each bin and correlated the methylation (mCG from small smooth for CG-DMRs; mCG from large smooth for CG-blocks; mCA on the same strand as the gene for CA-DMRs) or change in chromatin accessibility (log_FC using RPKM for DARs) with the change in expression of the gene (log_FC). To emphasize, only genes with a DiffEpi mark in that bin contribute to the correlation estimate for that bin. We performed a similar procedure upstream of the TSS and downstream of the TES, but here using a fixed bin size (1 kb).

Finally, we used a binomial generalized additive model to estimate the probability that a protein coding gene differentially expressed given that 8% of the gene body is covered by a DiffEpi mark. The ‘gam’ function in the ‘mgcv’ package (v1.8–23) was used to fit the generalized additive model with the formula \( \logit(z) = \beta_{0} + \beta_{1} \times \text{DiffEpi mark} \) and additional argument family = “binomial”. This estimated probability and its standard error are shown in Fig. 4c, annotated with example genes.

**Stratified linkage disequilibrium score regression (SLDSR).** We used SLDSR, implemented in the LDSC68 software, to evaluate the enrichment of common genetic variants from GWAS signals to partition trait heritability within functional categories represented by our DMRs, OCRs, and DARs1. SLDSR estimates the proportion of genome-wide single nucleotide polymorphism (SNP)-based heritability that can be attributed to SNPs within a given genomic feature by a regression model that combines GWAS summary statistics with estimates of linkage disequilibrium from an ancestry-matched reference panel. Links to GWAS summary statistics are available in Supplementary Table 24. Additional files needed for the SLDSR analysis were downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE/ following instructions at https://github.com/bulik/lodscore/wiki.

We ran LDSC (v1.0.0; https://github.com/bulik/lodscore) to estimate the proportion of genome-wide SNP-based heritability of 30 traits (Supplementary Table 24) across 53 ‘baseline’ genomic features (24 main annotations, 500 bp windows around each of the 24 main annotations, and 100 bp windows around 5 sets of ChIP-seq peaks; described in ref. 3) and eight brain-specific genomic features:

1. CG-DMRs (NeuN+): CG-DMRs between brain regions in NeuN+ samples (11.8 Mb) (Supplementary Table 8)
2. CH-DMRs (NeuN+): Union of CA-DMRs and CT-DMRs between brain regions in NeuN+ samples (39.6 Mb) (Supplementary Table 13)
3. CG-DMRs (NeuN+ vs. NeuN–): CG-DMRs between NeuN+ and NeuN– samples (70.0 Mb) (Supplementary Table 5)
4. DARs (NeuN+): DARs between brain regions in NeuN+ samples (118.1 Mb) (Supplementary Table 19)
5. DARs (NeuN+ vs. NeuN–): DARs NeuN+ and NeuN– samples (275.8 Mb) (Supplementary Table 18)
6. Brain H3K27ac: A set of regions marked by H3K27ac in human brain (415.4 Mb)
7. CNS (LDSC): A union of regulatory regions active in brain, previously considered by the authors of LDSC (338.8 Mb)
8. ChromHMM (union): A union of regulatory regions found by Chrom-
HMM using Roadmap Epigenomics\textsuperscript{39} data (498.1 Mb). The selected brain regions and their Roadmap Epigenomics codes were: Brain Angular Gyrus (E067); Brain Anterior Caudate (E068); Brain Cingulate Gyrus (E069); Brain Germinat Matrix (E070); Brain Hippocampus Middle (E071); Brain Inferior Temporal Lobe (E072); Brain Dorsolateral Prefrontal Cortex (E073); Brain Substantia Nigra (E074); Fetal Brain Male (E081); Fetal Brain Female (E082). The ChromHMM states selected as ‘regulatory regions’ were: Bivalent Enhancer; Bivalent/Positve TSS; Genic enhancers; Flanking Active TSS; Active TSS; Strong transcription; Enhancers. Processed ChromHMM tracks were downloaded using the AnnotationHub R/Bioconductor package (v1.36.2).

Collectively, we refer to features 1–5 as ‘differential features’, being the product of differential analyses, and features 6–8 as ‘non-differential feature’, each being the union of regions marked by various active histone modifications present in different brain regions.

We performed three rounds of analysis:

1. ‘Baseline’: A standard SLDSR analysis, as suggested by the LDSC authors, whereby each of the eight brain-specific features was added one at a time to a full baseline model that included the 33 ‘baseline’ categories that capture a proportion of heritability explained by the feature that overlaps it is added to the model (for example, the enrichment scores “proportion of heritability explained by the feature” will decrease when another feature of interest overlaps the other features in the set. In particular, the size. It should be noted, however, that the enrichment score depends on the terms for the other features in the model, but is more readily interpretable as an effect size. In this way, we were post-hoc adjusted for multiple testing using Holm’s method\textsuperscript{69}.

2. ‘Stringent’: Add each of the 5 differential features one at a time to a model that included the 53 baseline features and 3 brain-specific non-differential features.

3. ‘Adjusting for non-differential features (ndf)’ excluding differential features (df). Add each of the 5 differential features on at a time to a model that included the 3 brain-specific non-differential features (after having excluded any regions shared with the brain-specific differential feature) and the 53 baseline features.

For each round of analysis, we used SLDSR to estimate a coefficient $z$-score and an ‘enrichment score’ for each feature-trait combination. A brief description of their interpretation is given below; we refer the interested reader to the Online Methods of Finucane et al.\textsuperscript{70} for the complete mathematical derivation.

A coefficient $z$-score statistically larger than zero indicates that adding the feature to the model increased the explained heritability of the trait, beyond the heritability explained by other features in the model.

The enrichment score is defined as (proportion of heritability explained by the feature) / (proportion of SNPs in the feature). The enrichment score is unadjusted for the other features in the model, but is statistically independent of the other features.

For the complete list of neuronal DAR–CG-DMRs, Haystack selected a random, CG-content-matched subset of the input background to use for enrichment calculations. Significance was determined using a one-sided Fisher’s test. For the accession number GSE96615, Processed data is available through a UCSC hub, at http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&hubUrl=https://s3.us-east-2.amazonaws.com/brainepigenome/hub.txt.

Transcription factor motif enrichment.

We used the Haystack (v0.6.4)\textsuperscript{68} haystack.motifs module to scan for vertebrate JASPAR\textsuperscript{29} (2014) transcription factor binding motifs enriched in our data sets. A list of neuronal DARs that overlapped hypo- or hypermethylated neuronal CG-DMRs identified between NACC and BA9 in Neun\textsuperscript{2} nuclei were input into Haystack as a BED file (hyper CG-DMR/DARs, $n = 2,618$; hypo CG-DMR/DARs, $n = 1,435$). All autorosomal promoters were input as background for the neuronal DAR–CG-DMRs that overlapped promoters. For the complete list of neuronal DAR–CG-DMRs, Haystack selected a random, CG-content-matched subset of the input background to use for enrichment calculations. Significance was determined using a one-sided Fisher’s test.

To identify novel transcriptional regulators for the differentially expressed genes found between Neun\textsuperscript{2} populations from NACC and BA9, we first generated transcription factor gene scores using TEPIC\textsuperscript{40}. This software utilizes epigenetic information along with transcription factor binding sites to generate these scores, which can then be used by DYNAMITE\textsuperscript{41} to infer potentially important transcriptional regulators for predicting up- or downregulation for differentially expressed genes. Using the combined TEPIC/DYNAMITE pipeline, transcription factor affinity scores were computed within the DARs identified between Neun\textsuperscript{2} nuclei from NACC and BA9 ($n = 68,021$) using the provided human_jaspar_hoc_kellis. PSEM position weight matrix. The affinities per gene were calculated over a 5 kb window around a gene’s TSS, incorporating the signal abundance (cpm) within a peak into the transcription factor annotation. The auROC step size was 0.01.

We reported only those transcription factors that were expressed in our Neun\textsuperscript{2} nuclear and had a correlation coefficient $> 0.04$ in Fig. 4f. The full output (including transcription factors not expressed in our samples) is reported in Supplementary Table 23.

Gene ontology annotation. We used the Genomic Regions Enrichment of Annotations Tool (GREAT; v3.0.0)\textsuperscript{42} to assess nearest gene enrichment for hyper- and hypermethylated CG-DMRs. We used the hg19 assembly, reduced the default input parameter for max extension to 100 kb, and kept all other default parameters the same (set of available at the GREAT website: http://great.stanford.edu/). Gene Ontology (GO) terms returned must be significant by both the binomial and hypergeometric tests using the multiple hypothesis correction false discovery rate (FDR) $\leq 0.05$ whose binomial fold enrichment is at least 2.0.

Metascape\textsuperscript{43} was used to perform GO, Reactome, and KEGG pathway analyses using lists of gene symbols for differentially expressed genes (either up- or downregulated in NACC as compared to BA9 Neun\textsuperscript{2} nuclei) as input. The gene lists were generated by matching GENCODE gene IDs to gene symbols (“external gene id”) using biomaR\textsuperscript{44} (v2.30.0).

Supplementary software. All statistical analyses were performed using R\textsuperscript{3.3.3} and made use of packages contributed to the Bioconductor project\textsuperscript{45}. In addition to those R/Bioconductor packages specifically referenced in the above, we made use of several other packages in preparing results for the manuscript:

- biocodex (v1.14)
- AnnotationHub (v2.6.4)
- biomaR\textsuperscript{44} (v2.30.0)
- GenomicAlignments\textsuperscript{3}(v1.10.0)
- GenomicFeatures\textsuperscript{3}(v1.26.2)
- GenomicRanges\textsuperscript{3}(v1.26.2)
- ggplot2\textsuperscript{46} (v2.2.1)
- Hmisc (v4-0 2)
- Matrix (v1-2-8)
- rtracklayer\textsuperscript{47} (v1.34.1)
- SummarizedExperiment (v1.4.0)
- EnrichedHeatmap (v1.4.0)
- mgcv (v1.8-23)
- sva (v3.22.0)
- EnrichedHeatmap (v1.4.0)
- Picard (v2.2.2)
- bitqik (v1.2-94)
- Salmon (v6.7.2)
- tximport (v1.2.0)
- Metascape (http://metascape.org)
- GREAT (v3.0.0; http://great.stanford.edu)

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Code Availability.** The code used in this manuscript is available through Github (hansenlab/BrainEpigenomicsNN) and is archived at Zenodo (https://doi.org/10.5281/zenodo.1469577).

**Accession Codes.** All data generated in this study are deposited in NCBI GEO under accession code GSE96615.

**Data Availability**

Raw and processed data generated are available through NCBI GEO under accession number GSE96615. Processed data is available through a UCSC hub, at http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&hubUrl=https://s3.us-east-2.amazonaws.com/brainepigenome/hub.txt.

**References**

- Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. J. Mol. Biol. 196, 261–282 (1987).
- Rosenblloom, K. R. et al. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 43, D670–D681 (2015).
- Krüger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 31, 357–359 (2015).
- Huminizer, M. et al. Software for computing and annotating genomic ranges. Nature Protocols 9, 11, 739 (2014).
- Lawrence, M. et al. Software for computing and annotating genomic ranges. Nature Protocols 9, 144–161 (2013).
- Metascape (http://metascape.org): Analysis of functional enrichment across networks, pathways and GO terms. Bioinformatics 31, 9, 137–143 (2015).
- Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS): Genome Biol. 9, R137 (2008).
- Lawrence, M. et al. Software for computing and annotating genomic ranges. Bioinformatics 31, 11, 137–143 (2015).
- Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, R29 (2014).
60. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25 (2010).
61. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010).
62. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47 (2015).
63. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PloS Genet. 3, 1724–1735 (2007).
64. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference. Nat. Methods 14, 417–419 (2017).
65. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 4, 1521 (2015).
66. R. Core Team. A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, Austria, 2016).
67. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. B 73, 3–36 (2011).
68. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295 (2015).
69. Holm, S. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, 65–70 (1979).
70. Mathelier, A. et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 44(D1), D110–D115 (2016).
71. Tripathi, S. et al. Meta- and orthogonal integration of influenza “OMICs” data defines a role for UBR4 in virus budding. Cell Host Microbe 18, 723–735 (2015).
72. Durinck, S. et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 21, 3439–3440 (2005).
73. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191 (2009).
74. Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. Genome. Biol. 5, R80 (2004).
75. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods 12, 115–121 (2015).
76. Wickham, H. ggplot2: Elegant Graphics for Data Analysis, (Springer, New York, NY, USA, 2016).
77. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics 25, 1841–1842 (2009).
Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

**Statistical parameters**

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| n/a | Confirmed |
|-----|-----------|
| ☑  | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☑  | An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☑  | The statistical test(s) used AND whether they are one- or two-sided |
| ☑  | Only common tests should be described solely by name; describe more complex techniques in the Methods section. |
| ☑  | A description of all covariates tested |
| ☑  | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☑  | A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☑  | For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted |
| ☑  | Give P values as exact values whenever suitable. |
| ☑  | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☑  | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☑  | Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated |
| ☑  | Clearly defined error bars |
| ☑  | State explicitly what error bars represent (e.g. SD, SE, CI) |

Our web collection on statistics for biologists may be useful.

**Software and code**

Policy information about availability of computer code

**Data collection**

Summit v4.3.02 Build 2451 was used for flow cytometry.

**Data analysis**

Alignments and methylation calling was performed using Bismark (v1.4.3). Transcription factor binding motif analysis was performed with Haystack (v0.4). Differential analysis was performed using TEPIC (v2.0)/DYNAMITE (v1.0) to identify novel transcriptional regulators for differentially expressed genes. LDSC (v1.0.0) was used to estimate the proportion of genome-wide SNP-based heritability. ATAC-seq reads were trimmed with trimadap (v0.1), aligned with Bowtie2 (v2.2.5), and peaks were called using MACS (v2.1.0). All statistical analyses were performed using R (v3.3.x) and made use of packages contributed to the Bioconductor project. In addition to those R/Bioconductor packages specifically referenced in the above, we made use of several other packages and online programs in preparing results for the manuscript:

bsseq (v1.14), AnnotationHub (v2.6.4), biomaRt (v2.30.0), GenomicAlignments (v1.10.0), GenomicFeatures (v1.26.2), GenomicRanges (v1.26.2), ggplot2 (v2.2.1), Hmisc (v4.0-2), Matrix (v1.2-8), rtracklayer (v1.34.1), SummarizedExperiment (v1.4.0), edgeR (v3.16.5), limma (v3.30.7), sva (v3.22.0), EnrichedHeatmap (v1.4.0), mgcv (v1.8-23), Picard (v2.2.2), seqtk (v1.2-r94), Salmon (v0.7.2), tximport (v1.2.0), Metascape (http://metascape.org), GREAT (v3.0.0; http://great.stanford.edu)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw and processed data generated are available through NCBI GEO under accession number GSE96615. [Reviewer link for GEO: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE96615.]

Processed data is available through a UCSC hub, at http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&hubUrl=https://s3.us-east-2.amazonaws.com/brainepigenome/hub.txt.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No sample size calculation was performed; we chose 3 males and 3 females to allow the possibility of a segregation by sex.

Data exclusions

No data were excluded.

Replication

We did not replicate. We do have limited within sample replication as some of the same samples were analyzed using WGBS, RNA-seq and/or ATAC seq.

Randomization

Randomization was performed at the level of library preparation. There was no intervention so randomization was not employed in that sense.

Blinding

The analysts were not blinded.

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| | | Unique biological materials |
| | | Antibodies |
| | | Eukaryotic cell lines |
| | | Palaeontology |
| | | Animals and other organisms |
| | | Human research participants |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| | | ChiP-seq |
| | | Flow cytometry |
| | | MRI-based neuroimaging |

Unique biological materials

Policy information about availability of materials

Obtaining unique materials

The human brain tissues used in this study were obtained from the NIH Neurobiobank and while these exact tissues no longer exist, similar tissues can be made available by request to the tissue bank.

Antibodies

Antibodies used

For FANS we used anti-NeuN conjugated to AlexaFluor488 (cat. no. MAB377X, Millipore).
Validation

From the Manufacturer: Vertebrate neuron-specific nuclear protein called NeuN (Neuronal Nuclei). MAB377X reacts with most neuronal cell types throughout the nervous system of mice including cerebellum, cerebral cortex, hippocampus, thalamus, spinal cord and neurons in the peripheral nervous system including dorsal root ganglia, sympathetic chain ganglia and enteric ganglia. The immunohistochemical staining is primarily in the nucleus of the neurons with lighter staining in the cytoplasm. The few cell types not reactive with MAB377X include Purkinje, mitral and photoreceptor cells. Developmentally, immunoreactivity is first observed shortly after neurons have become postmitotic, no staining has been observed in proliferative zones. The antibody is an excellent marker for neurons in primary cultures and in retinoic acid-stimulated P19 cells. It is also useful for identifying neurons in transplants. Rat and mouse. It is expected that the Alexa Fluor 488 conjugated antibody will also react with human, ferret, chick and salamander. Reference using on human and mouse tissues: Lister et al. Science (2013) 9;341 PMID: 23828890

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
☒ All plots are contour plots with outliers or pseudocolor plots.
☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

A total of 2 x 250 mg of frozen tissue per sample was homogenized in 5 mL of lysis buffer (0.32 M sucrose, 10 mM Tris pH 8.0, 5 mM CaCl2, 3 mM Mg acetate, 1 mM DTT, 0.1 mM EDTA, 0.1% Triton X-100) by douncing 50 times in a 40 mL dounce homogenizer. Lysates were combined and transferred to a 38 mL ultracentrifugation tube and 18 mL of sucrose solution (1.8 M sucrose, 10 mM Tris pH 8.0, 3 mM Mg acetate, 1 mM DTT) was dispensed to the bottom of the tube. The samples were then centrifuged at 28,600 rpm for 2 h at 4C (Beckman Optima XE-90; SW32 Ti rotor). After centrifugation, the supernatant was removed by aspiration and the nuclear pellet was resuspended in 500 uL staining mix (2% normal goat serum, 0.1% BSA, 1:500 anti-NeuN conjugated to AlexaFluor488 (Millipore, cat#: MAB377X) in PBS) and incubated on ice. Unstained nuclei and nuclei stained with only secondary antibody served as negative controls.

Instrument

Beckman Coulter MoFlo Legacy Cell Sorter FL2 = PE

Software

Summit v4.3.02 Build 2451

Cell population abundance

The abundance of NeuN+ and NeuN- nuclei varied substantially from sample to sample and is in fact a main finding of our study. This is detailed in Supplementary Figure 1c. A small portion of the NeuN+ and NeuN- nuclei were re-run on the sorter to validate the purity which was greater than 95%.

Gating strategy

From the scatterplots in SuppFig1a, FS/SS events represent morphology of nuclei gated with debris excluded, doublets were excluded based on pulse width measurement. Gates for NeuN+ and NeuN- nuclei were set based on NeuN-488 staining intensity visualized by a histogram depicted in the representative example.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.