***Classification*:** Biological Sciences - Neuroscience

***Title*: Divergent neuronal DNA methylation patterns across human cortical development: Critical periods and a unique role of CpH methylation**

Authors

Price AJ[1,2*] Collado-Torres L[1,3*], Ivanov NA[1], Xia W[1], Burke EE[1], Shin JH[1], Tao R[1], Ma L[1], Jia Y[1], Hyde TM[1,4,5], Kleinman JE[1,5], Weinberger DR[1,2,4,5,6], Jaffe AE[1,2,3,7,8,+]

Affiliations

1. Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA

2. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

3. Center for Computational Biology, Johns Hopkins University, Baltimore, MD, USA

4. Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, USA

5. Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore, MD, USA

6. Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, USA

7. Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

8. Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

*equally contributing authors

Corresponding Author: Andrew E Jaffe, 855 N Wolfe St, Ste 300; Baltimore MD 21205. Phone: 1-443-287-6864; Email: andrew.jaffe@libd.org

***Keywords*:** DNA methylation, whole genome-bisulfite sequencing, brain development, neuropsychiatric disease, splicing, neurons

Abstract: We characterized the landscape of DNA methylation (DNAm) across the first two decades of human neocortical development in neurons and glia using whole-genome bisulfite sequencing. We show that the rate of DNAm changes more dramatically during the first five years of postnatal life than during the entire remaining period. We further refined global patterns of increasingly divergent neuronal CpG and CpH methylation (mCpG and mCpH) into six unique developmental trajectories, within which neighboring mC levels—independent of sequence context—were highly correlated, unlike across the genome, where mCpG levels were correlated but mCpH levels were not. We then integrated paired RNA-seq data and identified direct regulation of hundreds of transcripts and their splicing events exclusively by mCpH, independent of mCpG levels, across this period of human cortical development. In addition to expanding the relationship between mCpH and gene expression, these splicing-associated cytosines and developmentally dynamic DNAm regions were associated with neuropsychiatric disease risk sequence, providing insight into the cell type and timing of dynamic epigenomic remodeling in known disease risk genes and loci.

***Significance Statement*:** By studying DNA methylation levels within neurons and glia across human neocortical development, we show that the first five years of life is a critical period of epigenetic plasticity, identify a novel role of CpH methylation in gene expression and splicing during neuronal maturation and parlay a clearer picture of methylation remodeling into better understanding of the cell type and timing of neuropsychiatric disease risk sequence activity.

Introduction

Neurons are unique cells that persist throughout the lifespan, accumulating programmed developmental changes and environmental experience that fine tune neural circuitry in the brain. During development and maturation, neurons undergo precisely coordinated cascades of genetic regulation that combine with experience to shape the cellular output via progressive changes to the epigenome. DNA methylation (DNAm) is an integral facet of the epigenome that plays a role in establishing cell identity and developmental trajectories as well as adapting to experience via regulation of gene expression. Previous large-scale studies of DNAm across human brain development have been limited to homogenate tissue (1) or have used microarray technologies (2), creating ambiguity about the extent of cell type-specific developmental DNAm changes and effects on transcript isoforms across the genome (3).

To better characterize the DNAm landscape across human cortical development, we performed whole-genome bisulfite sequencing (WGBS, see Methods) on homogenate tissue and on neurons isolated from 24 human dorsolateral prefrontal cortex (DLPFC) samples aged 0-23 years using NeuN-based fluorescence-activated nuclear sorting (FANS, **Figure S1A**). To complement these data, we sequenced 8 FANS-derived glial (NeuN-) postnatal samples and 20 homogenate prenatal cortical samples, for a total of 75 samples after quality control (**Table S1**). We fully characterized the landscape of DNAm at both CpG and non-CpG (CpH) dinucleotides, allowing for a finer dissection of differential DNAm functional specificity. We also sequenced matched transcriptomes of homogenate cortical samples from these donors and a subset of three nuclear transcriptomes each from neurons and glia isolated by FANS to assess the functional consequences of epigenomic remodeling (53 total transcriptomes, **Table S2**). By exploring DNAm patterns in neurons across prenatal and postnatal human brain development, we show that the first five years of postnatal life are a critical period in epigenetic plasticity, and

we identify developmental shifts in neuronal DNAm in both the CpG and CpH context. We also clarify the relationship between CpG but particularly CpH methylation (mCpG and mCpH, respectively) and gene expression and splicing in neuronal development, and overlay the ramifications of these insights for neuropsychiatric disease.

Results

After data processing, quality control, and filtering, we analyzed 18.7 million cytosines in the CpG context at an average coverage of 15x. Comparable to previous reports (1, 4, 5), CpGs were overall highly methylated (71-76% CpGs with DNAm proportions, β>80%, **Table S3**). DNAm developmental changes identified in homogenate cortex were strongly confounded by developmental shifts in cell type proportions (OR=7.5, $p<10^{-100}$, **Figure S1B**) (2). While homogenate measurements were positively correlated with developmental changes that occurred in both neuronal and glial cell types (ρ=0.79, $p<10^{-100}$), cell type-specific developmental changes were less consistently observed (ρ=-0.26, $p<10^{-100}$, **Figure S1C-E**). Overall, roughly 40% of cell type-specific developmental DNAm changes could not be detected at all in homogenate cortex (**Figure S1F**), and many of the cell type-specific effects could not be accurately identified in homogenate tissue. These results highlight the importance of measuring DNAm in the appropriate cellular context for improved resolution in detecting true developmental changes.

DNAm as a map of putative functional genomic states

Local CpG methylation (mCpG) patterns are known to distinguish genomic states of DNA and chromatin. For instance, unmethylated regions (UMRs) are associated with promoters, with a subset of longer UMRs (DNAm valleys, DMVs) that overlap developmental genes often

encoding transcription factors (TFs) (6, 7); low-methylated regions (LMRs) often signify enhancer sequence (8); and partially methylated domains (PMDs) often associate with heterochromatin and late replicating DNA (9–11) . To better resolve the developing regulatory landscape in postnatal neurons and glia and in bulk prenatal cortex, we assessed the temporal dynamics of these selected DNAm patterns in the CpG context. Compared to prenatal homogenate cortex and postnatal glial cells, postnatal neurons showed a general accumulation of mCpG, at a rate 50% faster than the other cells; this was evident in the LMR and to a lesser extent the UMR landscape, since fewer and smaller LMRs were identified as neuronal development progressed (**Figure S2A-B**). As expected, UMRs and LMRs were highly enriched for transcription start sites (TSSs) and enhancers in DLPFC chromatin state data from the Roadmap Epigenomics Consortium (12) (**Figure S2C**). Interestingly, LMRs were similarly enriched in these states in both adult and fetal brain; this correspondence may reflect a shared regulatory landscape established early in development.

While PMDs are a common feature of most cell types, they have not been conclusively identified in neurons. Here we identified a range of 245 to 404 PMDs per neuronal sample (**Figure S3A**). PMDs were especially enriched for heterochromatin, and, interestingly, enhancers in our postnatal neuronal samples (**Figure S3B**). 65.4% of PMD base pairs were also identified as PMD in an independent WGBS dataset of sorted human neurons (**Figure S3C**). 40.3-61.0% of PMD bases per neuronal sample were identified as common PMD sequence, and 9.3-15.0% bases were additionally identified as PMD in at least one sample in a recent study profiling PMDs in multiple cell types and tissues (**Figure S3D**) (13). These data suggest that although the neuronal genome was overall highly methylated, a small but consistent portion displayed the characteristics of PMDs.

We further identified significant neuronal DMV changes through accumulation of mCpG that revealed regulators of cell identity and development and their temporal windows of expression change. Compared to bulk prenatal cortex, postnatal neurons and glia showed marked reduction in the size of DMVs (**Figure S4A**). Methylation shifts within DMVs led to inclusion and exclusion of transcription factor genes in an age-dependent manner, and on average transcription factor genes were higher expressed in the age group in which the gene was escaping the DMV state by accumulating DNAm (**Figure S4B-C**). Overall these results clarify the substantial DNAm landscape alterations that neurons and glia undergo during development in defined mCpG patterns, including previously unobserved PMDs.

Developmental shifts in neuronal mCpG highlight synaptic remodeling during first five years of postnatal life

We next quantified more localized changing mCpG levels by exploiting the correlation between neighboring mCpG levels to identify genomic regions with changing mCpG. We identified 11,179 differentially methylated regions (DMRs, FWER<5%, see Methods) in the CpG context between cell types (covering 31.1 Mb) that replicated in independent WGBS data (1) (98.4% concordant, ρ=0.925, **Figure S5A**, see Methods). Many of these DMRs overlapped genes involved in neuronal or glial-specific processes (**Figure S5B**). We found fewer DMRs for developmental mCpG changes compared with cell type differences, the majority being within rather than across cell types (2,178 versus 129 DMRs, respectively, at ~5% change in DNAm per decade of life, FWER<5%). Among the 2,178 cell type-specific developmental DMRs (cdDMRs, 3 Mb, **Table S4**), neuronal mCpG patterns seemed to diverge from an immature landscape shared by glia and prenatal cortex (**Figure 1A**), with the largest changes occuring in the first five years of life. Indeed, the magnitude of methylation changes in neurons and glia in

samples five years and younger was double that of the older samples (**Figure S5C-D**). These results provide epigenetic correlates to the known developmental processes occuring in the cortex in the first five postnatal years, including prolific synaptogenesis and gliogenesis.

We further parsed these cdDMRs using k-means clustering to partition the cdDMRs into six groups with unique DNAm characteristics (**Figure 1B**). 71.1% of cdDMRs fell within groups characterized by increasing neuronal and/or decreasing glial DNAm over postnatal development (**Figure 1B**, Groups 1, 2 and 6). A varying proportion of each cdDMR group corresponded to sequence differentially methylated by neuronal subtype from publicly available data (14) depending on the trajectory of neuronal methylation patterns in the group, suggesting that various neuronal subclasses contribute to these developmental patterns (**Figure S6A-B**, see Methods). Gene ontology enrichment in the six groups suggested that these groups are associated with related yet unique biological roles for the cell types implicated (**Figure 1C**). For example, **Figure 1D** shows a Group 6 cdDMR within *MBP*, a gene encoding a major component of the myelin sheath created by oligodendrocytes; glia but not neurons progressively lose DNAm at this site (1, 5), suggesting increased repression of a non-neuronal fate in maturing neurons not mirrored in glia over postnatal development in Group 6 cdDMRs. Likewise, the opposite pattern was observed in a Group 3 cdDMR within presynaptic neuronal gene *SNAP25* in which neurons uniquely and progressively lost DNAm over development (**Figure 1E**).

We lastly compared these cdDMR groups to a list of putative enhancers active in human brain development curated by evolutionary age (15) and found strong enrichment for these sequences across all six groups (**Figure S7A**). Human accelerated regions, or conserved sequences that have experienced rapid mutation in the human lineage (*14*), were also enriched for dynamic DNAm remodeling (**Figure S7B**), suggesting that our CpG-based cdDMRs may be enriched for sequences related to higher cognitive functions associated with the human DLPFC.

Overlapping cdDMRs with the mCpG features identified above provided additional insight to the potential functional genomic states underlying these regions. For instance, cdDMRs scarcely overlapped heterochromatic PMDs; cdDMRs losing neuronal mCpG were positively correlated with increasing LMR overlap, potentially reflecting enhancer element activation during cortical maturation in these groups (Groups 3 & 5 cdDMRs, t>3.8, ϱ>0.63, FDR<2.7e-03). Curiously, a high proportion of cdDMRs gaining DNAm in glia but not in neurons (Group 4 cdDMRs) overlapped DMVs early in development in glia which then steadily lose it with further development (t=-4.3, ϱ=-0.87, FDR=1.3e-02, **Figure S8A**). Assessing chromatin state from the homogenate Roadmap Epigenomics brain maps, in contrast, lacked the resolution to provide this nuance: all six cdDMR groups were similarly enriched for transcriptional (particularly TSS-flanking) and enhancer chromatin states, and depleted for heterochromatin and quiescent states (**Figure S8B**). These results confirm the role of dynamic DNAm in helping establish epigenomic states that guide cell lineage differentiation and emphasize the utility of creating genome-wide DNAm maps to better parse the functional diversity of cell type-specific developmental DNAm remodeling in the human cortex, a process that is particularly critical during the first five years of postnatal development.

Abundant neuronal CpH methylation is highly correlated with neighboring CpG methylation

Unlike many other somatic tissues and cell types, where mCpG is largely the only context of DNAm, mCpH is an abundant, conserved feature of the neuronal epigenome that distinguishes it from most other cell types and tissues (1, 4). We therefore analyzed 58.1 million cytosines in CpH contexts (H=A,T, or C) that were methylated across these samples (coverage ≥5, at least 5 samples with β>0, see Methods). As shown previously (1), mCpH sites were predominantly lowly methylated (92-99% CpHs with β<20%, **Table S3**). While mCpH was

distributed throughout the genome (**Figure S9A**), it was greater in neurons than glia (98.9% of 7,682,075 CpHs, FDR<5%) and mostly accumulated across postnatal development (99.3% of 3,194,618 CpHs, FDR<5%; **Table S5**). Most mCpH accumulated primarily in either the CAG or CAC context over the first five years of postnatal life—similarly to mCpG—followed by a tapered global increase into adulthood (**Figure S9B**).

While the majority of mCpH in embryonic stem cells (ESCs) occurs in the CAG context, previous work has shown that ESCs undergo loss of mCAG during neuronal differentiation followed by preferential accumulation of methylation in the CAC context (16). Here we further refine these patterns and find a cell type-specific relationship with trinucleotide context: overall, total mCAG increased 40% faster than mCAC in neurons, while in glia, mCAG accumulated 50% slower than mCAC (**Figure S9C**). However, taking into account relative genome-wide proportions of CAG and CAC, overall neuronal mCAG accumulated 30% slower than mCAC (**Figure 2A**). mCH that was greater in glia than neurons, or in younger than older neurons, was more likely to be in the CAG than CAC context (OR>4.13, p<2.2e-16). Interestingly, the 3,286 and 1,744 genes that contain significantly increasing and decreasing mCAC versus mCAG over development, respectively, were associated with different biological processes related to neuronal function and activity, particularly involving the synapse (**Figure S9D**).

Unlike the CpG context, where neighboring mCpG levels were highly correlated, neighboring CpH DNAm levels across the genome were not locally correlated. However, within the cdDMRs, while mCpH levels remained uncorrelated, together all methylated cytosines (i.e., mCpH+mCpG) showed similar autocorrelation as mCpG levels alone (**Figure 2B**). This was especially surprising given that there were about two times as many mCpHs than mCpGs within these regions and that the CpG and CpH were relatively interspersed, suggesting potential functional convergence in the developmentally regulated patterns identified by mCpG in these

regions. Indeed, unsupervised hierarchical clustering of CpH within the cdDMRs showed infant neuronal mCpH was even more similar to glia compared to older neurons than mCpG (**Figure 2C**). Examining the mean mCpH compared to mCpG within the k-means cdDMR clusters showed that the groups gaining mCpG were the most correlated with mCpH trajectories within the cdDMRs (ρ=0.97, t=17.6, p=2.0e-14), and that although mCpG is present at high levels prenatally, both mCpG and mCpH accumulate at similar rates over postnatal development in these groups, once again especially in the first five years of postnatal life where the majority of the methylation change takes place (t=-0.091, p=0.94; **Figure S10**). These results emphasize the potential regulatory importance of cdDMRs and the putative functional agreement between both contexts of DNAm in these regions.

mCpG and mCpH levels influence transcript isoform use

Previous studies have shown that both mCpG and mCpH in gene bodies but particularly in the promoter and first 2 kb of the gene are negatively associated with gene expression, and that genic mCpH is the most discriminating predictor of gene expression (1, 5). To anchor our DNAm patterns in transcriptional activity, we compared our WGBS data with NeuN-sorted nuclear RNA-seq data (see Methods) to correlate patterns of DNAm and gene expression. We took the average DNAm levels across six age groups—infant (ages 0-1), child (ages 1-10), and teen (age 10+) within both cell types individually (neuronal and glial)—and calculated associations between DNAm and expression (**Table S6**). Gene expression was negatively correlated with promoter mCpG regardless of age and cell type (-0.42<ρ<-0.36, p~0; 57,332 genes, p <10^{-100}). However, only neurons showed a negative correlation between gene expression and promoter mCpH, a relationship that became stronger over development (infant:

ϱ=-0.13 versus child: ϱ=-0.23 and teen: ϱ=-0.25, FDR<0.05). This pattern is consistent with the preferential accumulation of mCpH in neurons as the brain matures.

Because mCpG has previously been associated with alternative splicing (17) and mCpH is 15-20% greater in exons than in introns (9), we hypothesized that accumulating mCpH may contribute to the diversity of alternative splicing characteristic of the brain particularly during development. Leveraging our single-base resolution data, we were able to identify genome-wide functional correlates of mCpH, independent of nearby mCpG, by associating DNAm with nearby expression in the same cortical samples. Specifically, we tested whether methylation levels directly associated with gene or exon expression levels as well as the "percent spliced in" (PSI) of alternative splicing events using the 22 neuronal samples with matching homogenate polyA+ RNA-seq data (see Methods). We found 40,940 CpG and 40,303 CpH associations that explain changes in these three expression summarizations at FDR<5% with a genome-wide $p<5x10^{-4}$. We further identified 220,622 marginal (p<0.01) CpG associations with expression within 1 kb around the associated CpH. While an independent association of mCpH at the gene and PSI summarizations was rare, there were substantially more exons exclusively regulated by local mCpH, largely in the CHH context, in developing postnatal neurons (**Figure 3A**). Three examples of methylation-associated isoform changes are shown in **Figure 3B**.

Regardless of context specificity, these expression-associated cytosines were depleted in gene promoters and instead enriched in gene bodies and flanking regions (**Table 1**, see Methods). Both contexts were enriched for the high-GC 3' and 5' canonical splice site sequences (FDR<1.1e-04), although the associated cytosine could be either inside or outside the corresponding expression feature. Only 3.5-13.7% of expression-associated cytosines overlapped DMR sequence after stratifying by expressed feature and dinucleotide context,

indicating that these associations may arise from a more individualized mC effect than the DMRs.

Although the majority of these DNAm-expression associations were independent of development despite being identified in developing neurons, the mCpH changes at these sites were independently associated with age and expression. The genes including PSI events regulated by local DNAm levels in both CpG and CpH contexts were consistently enriched for neuronal components (**Figure S11**), while genes containing methylation-associated alternative exons were enriched for synaptic signaling and neurotransmitter transport (**Figure 3C** and **Figure S12**), suggesting that we are detecting true neuronal mC-expression associations despite measuring splicing in homogenate RNA-seq. Most expression-associated cytosines at the gene- and exon-level showed significant decreases in feature expression as methylation levels increased. Many of these genes also were differentially expressed between neuronal and glial nuclear RNA (FDR<0.05), further underscoring the limitations of homogenate tissue for splicing analysis (**Table 1**).

These associations between these putatively regulatory cytosines and nearby expression levels can be explored in a user-friendly webtool (https://jhubiostatistics.shinyapps.io/wgbsExprs/). Results can be interactively summarized such as in **Table 1** for user-selected subsets and visualized as in **Figure 3B** or via the UCSC genome browser (**Figure S13**). By integrating neuronal mCpG and mCpH levels with accompanying RNA-seq data in the same brains, we have identified, for the first time, direct regulation of hundreds of transcripts and their splicing events exclusively by mCpH, independent of mCpG levels, across the first two decades of human cortical development.

DNAm patterns shed light on the affected cell type and timing of neurodevelopmental disease risk

We finally examined the relationship between the CpG-based cdDMRs and dynamic mCpH sites with genetic risk for human brain disorders, hypothesizing that DNAm patterns may illuminate the affected cell types and potentially critical postnatal risk timeframes for genomic activity in these complex disorders. We first assessed regions of common disease risk identified by genome-wide association studies (GWAS) by examining the distribution of our DMRs within genomic loci associated with risk for Alzheimer's disease (18), Parkinson's disease (19), Type II Diabetes (20), and Schizophrenia (21). We found that GWAS risk loci for the different diseases were enriched for unique groups of cdDMRs (**Table S7; Figure 4A**). For instance, Type II Diabetes loci were highly enriched for Group 4 cdDMRs (characterized by increasing glial and and static neuronal DNAm; OR=32.3, FDR=4e-4), while both Parkinson's disease and schizophrenia GWAS loci were enriched for Group 3 cdDMRs (decreasing neuronal and static glial DNAm; OR>7.9, FDR<4.8E-06) and Group 6 cdDMRs (decreasing glial and static neuronal DNAm; OR>2.8, FDR<6.2e-3). Schizophrenia loci were also enriched for Group 5 cdDMRs (decreasing neuronal and increasing glial DNAm; OR=5.0, FDR=3.9e-3). These findings—loss of DNAm in neurons—are supported by cell type-specific expression-based enrichments of schizophrenia GWAS variants in neurons (22), given the observed negative correlation between DNAm and expression.

We further refined these relationships by focusing more specifically on genes containing *de novo* and inherited mutations associated with these and other disorders, including within rare copy number variants (CNVs) and syndromic variants. Using eight curated gene sets for neurodevelopmental and neurodegenerative brain disorders (21, 23), we found significant enrichment for genes implicated in neurodevelopmental disorders (e.g. autism spectrum

disorder (ASD), syndromal neurodevelopmental disorders, and intellectual disability) for many DMR groups including both increasing and decreasing neuronal and glial DNAm (**Table S8**). Cell type DMR enrichment confirmed a prominent role of neuronal functioning in most of the neurodevelopmental disorders, including risk genes for schizophrenia, ASD, bipolar affective disorder, intellectual disability and syndromal neurodevelopmental disorder using an orthogonal measurement approach as done previously (OR>2.0, FDR<0.027) (22). Genes hypomethylated in glia, however, also were also enriched in ASD genes from the SFARIgene database and rare and *de novo* single nucleotide variant schizophrenia genes (OR>1.9, FDR<4.3e-3). In contrast, a curated set of neurodegenerative disorder genes showed no enrichment for these DMRs generated within the first two decades of life, perhaps reflecting lack of relevant age ranges to the etiology of that disorder. These results expand the resolution of previous microarray-based reports, yielding more significant effect sizes despite the small number of subjects (2).

Although contrasting patterns of changing DNAm were often enriched within a single disease, these patterns highlighted the cell type and developmental timing of epigenetic regulation at individual risk genes. For instance, ASD gene sets were enriched for both cdDMRs progressively losing neuronal mCpG (Groups 3, 4 & 5) as well as decreasing glial mCpG in Group 6 (OR>2.8, FDR<0.05); similarly, genes within schizophrenia GWAS risk loci were enriched for both Group 1 (increasing neuronal, decreasing glial) and Group 3 (decreasing neuronal, static glial) cdDMRs (OR>3.12, FDR<2.9e-3), reflecting the complex etiology of these disorders.

However, by focusing on individual cdDMRs, specific effects underlying these enrichments can be emphasized. This is the case with *HDAC4* and *CACNA1B*, two genes associated with ASD in the SFARIgene database. We identified a 566 bp Group 3 (decreasing neuronal, static glial DNAm) cdDMR within an intron of *HDAC4*, a calcium-sensitive

transcriptional repressor. Early in development, this cdDMR is highly methylated, including through infancy; however, beginning in childhood, the cdDMR progressively lost mCpG selectively in neurons, suggesting the cell type (neurons) and timing (childhood) of risk for this gene (**Figure 4B)**. On the other hand, a 3.6 kb Group 5 (decreasing neuronal, increasing glial DNAm) cdDMR within *CACNA1B*, a gene encoding a calcium voltage-gated channel subunit, shows neurons selectively losing mCpG much earlier, with infant neuronal methylation levels already 50% decreased from that of bulk prenatal cortex and postnatal glia (**Figure 4C**). While *HDAC4* and *CACNA1B* could both be considered "neuronal" genes, given the general negative association between methylation and gene expression, the timing of cdDMR loss of mCpG suggests that *CACNA1B* activity occurs earlier in postnatal development than *HDAC4*. Given that onset is typically in early childhood, these risk genes may have differing implications in the etiology of ASD.

A third example is a 646 bp Group 6 (decreasing glial, static neuronal DNAm) cdDMR that overlaps the last intron and exon of *AKT3*, a serine/threonine-protein kinase gene. Although the AKT3/1q44 locus has been associated with schizophrenia risk, the mechanisms are not yet known given that *AKT3* is involved in many biological functions (Howell et al. 2017). Interestingly, this cdDMR selectively lost mCpG in glia beginning in infancy, suggesting that *AKT3* activity in human brain may be localized to glia beginning early in postnatal life (**Figure 4D**). These results highlight that profiling DNAm patterns in neurons and glia can better identify the timing and cellular context of epigenetic remodelling of known disease-associated genes.

We also found significant enrichments in the CpH context, as significant methylation changes between cell types and over development (FDR<0.05) were overall depleted in risk loci of the diseases tested (OR<0.94, FDR<3.4e-5). However, at the gene level, dynamic mCpH was depleted in intellectual disability genes (OR<0.36, FDR<3.4e-5), while syndromal

neurodevelopmental disorder genes, ASD genes from the SFARIgene database, and rare SNV schizophrenia genes were all enriched for dynamic mCpH (OR>2.0, FDR<0.036). These conflicting results are curious given the negative association between mCpH and gene expression and the implication of neurons in these disorders above; however, outside of the context of DMRs, individual mCpH could be associated both positively and negatively with expression. Indeed, many genes, exons, and PSI events both positively and negatively associated with both mCpH and mCpG were also enriched for the neurodevelopmental disorder gene sets (**Table S9**). Overall these results suggest that these examples of dynamic methylation and associated isoform switching may be playing a role in the development of higher cognitive functions during brain maturation associated with these diseases.

Discussion

Here we have created a single-base resolution map of the dynamic DNAm landscape across the first two decades of postnatal human neuronal and glial development. Using FANS-derived samples, we were able to identify 40% more true developmentally-regulated regions of changing DNAm than were identified in homogenate DNAm cortical data. We profiled specific features of the DNAm landscape including LMRs, UMRs, PMDs and DMVs and found that across features, neurons were typified by a general accumulation of mCpG. It is worth noting that in the absence of complementary cell type-specific chromatin data, characterizing known DNAm features provided a more granular view of the potential functional genomic state in these regions than the available predictions derived from a few homogenate cortical samples. Particularly in studies using human postmortem brain, where the tissue is often subjected to long postmortem intervals and low pH that degrades less stable epigenetic signatures, DNAm is a robust and durable marker that can be used to map the functional genomic terrain.

We further parsed the general accumulation of neuronal DNAm into six trajectories of cell type-specific developmental patterns and found that neuronal mCpG progressively diverged from a shared landscape with glia and bulk prenatal cortex as the brain matured. Importantly, these diverging patterns were most striking during infancy through the first five years of postnatal life. The human brain experiences an explosion of synaptic connections during this time, to nearly double the number found in the mature adult brain (24). Although previous work has underscored this timeframe in terms of rapid DNAm accumulation (1), this is the first work to refine DNAm patterns to reflect cell type-specific gain and loss of mCpG and mCpH within this critical window. By parsing these neuronal and glial DNAm patterns, we have highlighted epigenetically dynamic regions that may be contributing to the developmental processes such as synaptogenesis occuring during this timeframe that establish the foundation for fine-tuning connections throughout the remainder of brain maturation. These results provide a finely resolved depiction of epigenetic plasticity being greatest during this period of life, and supports other evidence that environmental experience during these years may have an especially enduring impact on brain function.

mCpH is unusually abundant in neurons compared to other cell types and appears to undergo trinucleotide-specific reprogramming during differentiation from ECSs (16). While most mCpH in ESCs occurs in the CAG context, neuronal mCpH predominantly accumulates in the CAC context (16). Here we elaborate on this relationship, showing that while both mCAG and mCAC aggregate in neurons as they mature and mCAG is gained faster than mCAC overall, mCAC accumulates proportionally faster in both neurons and glia over time. Interestingly, although neurons and glia contained mCpH in both trinucleotide contexts, mCAG was more likely to have higher levels in glia than neurons or be decreasing over development; indeed, genes containing decreasing mCAG but not mCAC were strongly associated with neuronal

biological processes. Trinucleotide context therefore may have as yet not well understood ramifications in brain development.

In terms of the relationship between mCpH and mCpG, we found that while neighboring mC (ie, mCpG+mCpH) was not correlated genome-wide, mC was highly correlated within the cdDMRs despite local mCpH not being correlated. In other words, there was a convergence of levels of all contexts of methylation within the cdDMRs that was not detected genome-wide. mCpH also recapitulated the pattern seen in mCpG of diverging from a shared DNAm landscape with glia. Given that mCpH and mCpG have previously been shown to work in concert to recruit MECP2 binding to fine-tune gene expression (25), it makes sense that levels of both contexts would perhaps reflect a shared functional role within putatively regulatory cdDMR sequence, since cdDMRs were also enriched for gene bodies and brain enhancer sequence. This work quantifies this correlation for the first time, a DNAm relationship unique to only a selection of cell types that includes neurons.

The identification of widespread association of mCpG and mCpH with expression and specific splicing events, particularly in neuronal genes enriched for neuropsychiatric diseases, highlights a potential novel role of mCpH and further expands the role of mCpG in the regulation of gene expression in neurons. Splicing is predominantly a co-transcriptional process influenced by changes in chromatin modifications and RNA binding proteins; the effects of DNAm on splicing decisions is not yet well studied currently (26). Here we found thousands of associations between mC levels and gene, exon and PSI expression in developing postnatal neurons, featuring particularly many exons that are exclusively associated with mCpH. Although it is not possible to establish a causal role for mC in this data, these analyses summarized in the provided website empower other researchers to explore the connection between DNAm and

alternative isoform use, a phenomenon particularly prevalent in the developing brain that is often associated with disease (27).

We finally explored the relationship between the dynamic DNAm identified above and known genomic risk for various diseases and found both expected and surprising associations. We confirmed neuropsychiatric disease gene enrichment in neurons and showed minimal enrichment of neurodegenerative genes in cdDMRs, for instance, but also found contrasting cdDMR groups were often enriched in the same disease. On one hand, these results make sense given the complex genetics and etiology underlying neuropsychiatric diseases. There is a large benefit however in being able to pinpoint the genomic boundaries, developmental timing and cellular context of epigenomic remodelling of regulatory elements or expressed features associated with known risk genes or loci. This work provides the first *ex vivo* look at the DNAm dynamics within human neurons and glia and thus allows for the first examination of those parameters within the relevant organ, the brain.

Despite these insights, this work must include several caveats. WGBS does not allow for the discrimination between mC and hydroxymethyl-cytosines (hmC), an intermediary in the demethylation pathway. Previous work however has shown (5) that only a fraction of CpGs, and no CpHs, have measurable levels of of hmC, suggesting that our results are not confounded. In the cited study, hmCpG signal from homogenate cortex represented 10% of the hypermethylation they found in excitatory neuron mCpG, suggesting that most of the mCpG signal in our neuronal data likely is true mCpG. Our data are also limited by not discriminating subcellular contexts more granular than neurons and glia. Recent work has shown that neuronal subtypes have distinct epigenomic profiles that are blended using NeuN-based FANS (5, 14, 28). While a percentage of the bases in the DMRs identified in this work have also been previously shown to be differentially methylated by cell type, the proportion of these subtypes

should be stable over postnatal development (29). Future epigenomics studies however can improve on the resolution of this study by isolating more specific neuronal subpopulations to refine the cellular specificity of these neuronal methylation changes largely occurring in the first few years of life.

By mapping the changing DNAm landscape over human postnatal neuronal and glial development, we have identified unique trajectories of DNAm change particularly dynamic during the first five years of life that show convergence between mCpG and mCpH, as well as associations between single mCpG and mCpH and alternative splicing.

Acknowledgements

Author Contributions

- A.J.P.: Conceptualization, Formal Analysis, Investigation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing.
- L.C.-T.: Conceptualization, Formal Analysis, Visualization, Software, Writing – Original Draft Preparation, Writing – Review & Editing.
- N.A.I.: Software, Data Curation.
- W.X.: Investigation.
- E.E.B.: Formal Analysis, Writing – Review & Editing.
- J.H.S.: Supervision.
- R.T.: Investigation.
- L.M.: Investigation.
- Y.J.: Investigation, Supervision.
- T.M.H.: Data Curation, Resources, Writing – Review & Editing.
- J.E.K.: Data Curation, Resources, Writing – Review & Editing.

- D.R.W.: Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing.
- A.E.J.: Conceptualization, Formal Analysis, Funding Acquisition, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing.

Funding

This project was supported by The Lieber Institute for Brain Development and by NIH grants R21MH102791 and R21MH105853. Data were generated as part of the PsychENCODE Consortium, supported by: U01MH103392, U01MH103365, U01MH103346, U01MH103340, U01MH103339, R21MH109956, R21MH105881, R21MH105853, R21MH103877, R21MH102791, R01MH111721, R01MH110928, R01MH110927, R01MH110926, R01MH110921, R01MH110920, R01MH110905, R01MH109715, R01MH109677, R01MH105898, R01MH105898, R01MH094714, P50MH106934 awarded to: Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke University), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (University of Southern California), Mark Gerstein (Yale University), Daniel Geschwind (University of California, Los Angeles), Fernando Goes (Johns Hopkins University), Thomas M. Hyde (Lieber Institute for Brain Development), Andrew E. Jaffe (Lieber Institute for Brain Development), James A. Knowles (University of Southern California), Chunyu Liu (SUNY Upstate Medical University), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Panos Roussos (Icahn School of Medicine at Mount Sinai), Stephan Sanders (University of California, San Francisco), Nenad Sestan (Yale University), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (University of California, San Francisco), Patrick Sullivan (University of North Carolina), Flora Vaccarino (Yale University), Daniel R. Weinberger (Lieber Institute for Brain

Development), Sherman Weissman (Yale University), Kevin White (University of Chicago), Jeremy Willsey (University of California, San Francisco), and Peter Zandi (Johns Hopkins University).

Competing Interest

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest: none declared.

Data availability and materials

Raw and processed data is available from Synapse.org at accession: syn5842535. Code is available through GitHub at: https://github.com/LieberInstitute/brain-epigenomics

Materials and Methods

Postmortem brain tissue acquisition and processing

As previously described in Jaffe *et al.* 2016 (2), post-mortem human brain tissue was obtained by autopsy primarily from the Offices of the Chief Medical Examiner of the District of Columbia, and of the Commonwealth of Virginia, Northern District, all with informed consent from the legal next of kin (protocol 90-M-0142 approved by the NIMH/NIH Institutional Review Board). Additional post-mortem prenatal, infant, child and adolescent brain tissue samples were provided by the National Institute of Child Health and Human Development Brain and Tissue Bank for Developmental Disorders (http://www.BTBank.org) under contracts NO1-HD-4-3368 and NO1-HD-4-3383. Postmortem human brain tissue was also provided by donation with

informed consent of next of kin from the Office of the Chief Medical Examiner for the State of Maryland (under Protocol No. 12-24 from the State of Maryland Department of Health and Mental Hygiene) and from the Office of the Medical Examiner, Department of Pathology, Homer Stryker, M.D. School of Medicine (under Protocol No. 20111080 from the Western Institute Review Board). The Institutional Review Board of the University of Maryland at Baltimore and the State of Maryland approved the protocol, and the tissue was donated to the Lieber Institute for Brain Development under the terms of a Material Transfer Agreement. Clinical characterization, diagnoses, and macro- and microscopic neuropathological examinations were performed on all samples using a standardized paradigm, and subjects with evidence of macro- or microscopic neuropathology were excluded, as were all subjects with any psychiatric diagnoses. Details of tissue acquisition, handling, processing, dissection, clinical characterization, diagnoses, neuropathological examinations, and quality control measures were further described previously (30). Postmortem tissue homogenates of the prefrontal cortex (dorsolateral prefrontal cortex, DLPFC, BA46/9) were obtained from all subjects.

Fluorescence activated nuclei sorting and NeuN staining

Nuclei were isolated from 100-300mg of pulverized DLPFC tissue using dounce homogenization followed by ultracentrifugation over a sucrose density gradient. Homogenization was performed on ice in 5mL lysis buffer [0.32M sucrose, 3mM magnesium acetate, 5mM calcium chloride, 5mM EDTA (ph 8.0), 10mM Tris-HCl (pH 8.0), 0.1% Triton X-100] and the resulting homogenate was layered over 38mL sucrose buffer [1.8M sucrose, 3mM magnesium acetate, 10mM Tris-HCl (pH 8.0)] and centrifuged at 139,800 x g for 2 hours at 4^{o}C. Cellular debris and lysis and sucrose buffers were removed and the pelleted nuclei were resuspended in 500uL PBS. Nuclei were then labeled in a solution of anti-NeuN antibody conjugated to Alexa

Fluor 488 (A60, Millipore, 1/1000) and 0.1% BSA, rocking for 30 minutes at 4°C, followed by the addition of DAPI. Nuclei sorting was performed at the the Johns Hopkins School of Public Health Flow Cytometry Core with a DakoCytomation MoFlo and data were analysed using FlowJo software (Tree Star).

Nucleic acid extraction and RNA-seq library preparations

RNA was extracted from homogenate and sorted samples using TRIzol LS Reagent (Thermo Fisher Scientific) followed by the RNeasy MinElute Cleanup Kit (Qiagen). Genomic DNA extraction was performed using the DNeasy Blood and Tissue Kit (Qiagen). Bisulfite conversion of 600 ng genomic DNA was performed with the EZ DNA methylation kit (Zymo Research). RNA sequencing libraries were made with the TruSeq RNA Library Prep Kit (Illumina) and the RiboGone Low-Input Ribosomal RNA Removal Kit (Clontech). Library concentrations were measured using a Qubit 2.0 and library fragment sizes were measured using a Caliper Life Sciences LabChip GX. One hundred base-pair paired-end sequencing was run on an Illumina HiSeq 2000.

Whole genome bisulfite library preparations and sequencing

Sequencing libraries were made with Illumina TruSeq DNA Methylation library preparation kits. Lambda (λ) DNA sequence was spiked in at 1% concentration to assess bisulfite conversion efficiency. Library concentrations were measured using a NanoDrop and library fragment sizes were measured using an Agilent Bioanalyzer 2100. Libraries were spiked in with 10% PhiX to improve base calibration calls and subsequently sequenced on an Illumina X-Ten Platform with paired-end reads (2x150bp), targeting 30x coverage and Q30 > 70% read quality.

Data processing/alignment

We sought to align the paired-end reads for each sample to the in-silico bisulfite-treated hg19 genome, which we created using the Bismark v0.15.0 (31) `bismark_genome_preparation` program. For each library of paired-end reads (one per sample), the following processing was performed (**Figure S14**):

- FastQC v0.11.4, to assess read quality, presence of adapter sequence, and overrepresented sequences.
- Trimmomatic v0.35 (32) to trim low quality and adapter-containing portions of the reads, with the following parameters: `PE -threads 12 -phred33 ILLUMINACLIP:/Trimmomatic-0.35/adapters/TruSeq3-PE.fa:2:30:10:1 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:75`. This resulted in three sub-libraries of reads per sample: one paired-end sub-library, and then two single-end sub-libraries where the other corresponding paired read was trimmed to a length below the defined threshold.
- FastQC v0.11.4, on each of the three sub-libraries, to assess the improvement in read quality and adapter content following trimming.
- FLASh v1.2.11 (33), to merge the paired-end sub-library reads into longer single-end reads, as reads that overlapped around CpGs and CpHs might bias or at the least double-count the DNAm estimates. Furthermore Bismark (31) could only be run on single- or paired-end reads and not a combination of both. This therefore further split the paired-end sub-library into three sub-libraries: the subset of paired-end reads that were merged into longer single-end reads, and then left and right single-end reads that could not be merged.

- Bismark v0.15.0 (31), to align each of the five now-single-end sub-libraries (left-trimmed, right-trimmed, FLASh-merged, FLASh-left-unmerged, and FLASh-right-unmerged) to the bisulfite-converted hg19 genome using bowtie2 (34) and the `--non-directional` argument.
- Resulting alignment (BAM) files across five sub-libraries were merged, sorted, and indexed using samtools v1.3 (35) to produce one large/merged BAM file per sample.
- Alignments with evidence of duplication were removed using the `MarkDuplicates` program in Picard tools v1.141, which systematically appeared to be localized to low complexity DNA sequence near centromeres.
- The Bismark (31) `bismark_methylation_extractor` program was run on each post-duplicate-removed BAM file per sample to extract CpG and CpH DNAm levels.

We additionally aligned reads from each sample to the PhiX and Lambda (λ) genomes to compute quality control metrics related to sequencing and bisulfite conversion quality. The average percentage of reads mapping back to the λ genome was 1.32% and the average bisulfite conversion efficiency was 98.64%. The average bisulfite conversion efficiency was not associated ($p>0.05$) with cell type, age, cell type (adjusting for age), age (adjusting for cell type), and the interaction of them for the NeuN- (glia) and NeuN+ (neuron) samples as well as for age in the homogenate samples.

Identifying methylation features

We identified PMDs, UMRs and LMRs using the bioconductor package MethylSeekR (version 1.20.0). We obtained coverage information from the cleaned set of ~18 million CpGs by extracting coverage and methylation using the `getCoverage()` function from the bsseq bioconductor package (36) v1.10.0. PMDs were called using `segmentPMDs()` and were

visually inspected using `plotPMDSegmentation()`. To create a more stringent cutoff for PMDs, we filtered PMDs to those longer than 100 Kbp. We calculated the FDRs using `calculateFDRs()`, while masking the >100 Kbp PMDs, setting the m parameter to 0.5 and the FDR cutoff to 10. PMDs were further filtered to exclude overlaps with the UCSC "gap" database table from hg19 except for the gaps labeled heterochromatin. We calculated UMRs and LMRs using `segmentUMRsLMRs()`. DMVs were defined as UMRs in which `pmeth` was less than or equal to 0.15 and the width was greater than or equal to 5 Kbp.

Identifying CpG differentially methylated regions (DMRs)

Using the bsseq (36) Bioconductor package v1.10.0 we loaded the Bismark (31) report files and filtered the CpG data to keep only the bases where all samples had a minimum coverage of 3 (18,664,892 number passed the filter). We smoothed the methylation values of the remaining CpGs using the `BSmooth()` function from bsseq with the `parallelBy="sample"` option. To identify the age and cell type DMRs we used a model that adjusted for both covariates while for the interaction model we included an additional interaction covariate. We identified the DMRs using the bumphunter (37) Bioconductor package v1.14.0 using the `maxGap=1000, B=250, nullMethod="permutation", smooth=FALSE` options, which tends to be conservative in DMR identification (37). For the `cutoff` option we used 0.1 for the cell type DMRs adjusting for age, 0.005 for the age DMRs adjusting for cell type, and 0.009 for the age and cell type interaction DMRs. These first two parameter cutoffs correspond to 10% minimum DNAm differences between neurons and glia and 5% change in DNAm per decade of life across cell types, and were chosen based on functionally-relevant change in DNAm. The cutoff for the interaction model (cdDMRs) was based on selecting the equivalent percentile of change from the overall age model (86th percentile) - this

percentile-based cutoff was in line with recommendations for selecting cutoffs for statistical models with less clear biological interpretations (37). We used a family wise error rate (FWER) threshold of 5% to determine the DMRs: `fwer` output from the `bumphunter()` function. A small subset of DMRs involved a single CpG, which arises from have a more significant area (length times effect size) than any DMRs identified in null permuted data.

Interaction DMRs processing

For the interaction DMRs with FWER<5% we extracted the methylation values from the glial and neuronal samples using bsseq (36) v1.13.9 and then computed a mean methylation value per DMR for each cell type. We also calculated the mean interaction coefficient for each DMR across all the cytosines in the DMR by cell type. Using the mean coefficients by cell type we clustered the interaction DMRs using the `kmeans()` function with `centers=6, nstart=100` options. We chose `centers=6` based on biological interpretability of the results and because k=6 results in an optimal AIC for clusters computed with mean centered and scaled data. For each cytosine in the interaction DMRs we calculated the t-statistic and coefficient for age explaining differences in methylation adjusted for cell type: `~ age + cell type`. For each DMR we computed the mean age coefficient by cell type and then calculated the median absolute coefficient across all interaction DMRs. The neuronal / glia ratio is 1.5 for such median absolute age effects across the interaction DMRs.

Comparing homogenate vs. cell type-specific WGBS

We first filtered CpGs to those with coverage in all 55 postnatal samples. For homogenate samples, we used `lmFit()` and `ebayes()` from the limma (38) Bioconductor package v3.34.5 to assess age-associated changes to DNAm levels with the linear model ~

`Age`. For the cell type-specific samples, we used a linear model `~ Age * Cell Type` to assess cell type, overall age, and age in a cell type changes to DNAm levels. We subset CpGs to those that were significantly differentially methylated by age in homogenate samples ($p<1x10^{-4}$) and plotted the coefficients for each CpG in **Figure S1B-E**. The relationship between each variable was quantified using Fisher's exact test.

Roadmap Epigenome enrichments

We computed the relative enrichments of different genomic regions using Epigenome Roadmap data (12) by computing the proportion of bases in each of the 15 ChromHMM states for each of the cells and tissues provided by the Consortium. We compared the proportion of bases in each state within each candidate region set to the overall genome and computed the corresponding log2 enrichments between the regions and this genomic background. We compared DMR and mCpG-based methylation feature regions to all profiled cell types in the Consortium for these analyses.

Assessing the contribution of neuronal subtypes

The percent of neuronal subtype-specific basses were calculated by reducing the total subtype-specific CpG-DMRs from Luo et al. (14), reducing the bases in each group of cdDMRs and calculating the percent of cdDMR bases that intersected the merged subtype-specific CpG-DMR bases.

Visualizing DMRs

Plots for the DMRs were made using bsseq (36) v1.14.0, EnsDb.Hsapiens.v75 v2.99.0, and RColorBrewer v1.1-2. Genes within 20 kb of a DMR were retained for the plots. ATAC

peaks (see ATAC-seq data processing) were included using the `addRegions` argument of the `plotManyRegions()` function from bsseq. Genes and exons were included using the `annoTrack` argument and we used `extend=2000` for making the plots.

CpH processing

Using Bismark v0.16.3 (31) we created report files using the methylation extractor program with the `CX_context` and `split_by_chromosome` options for the hg19 human genome in order to extract the methylation values for the CpHs. Then for each chromosome, using the bsseq (36) Bioconductor package v1.10.0 we loaded the Bismark (31) report files and added the c_context and trinucleotide_context information from Bismark using custom R code based on the bsseq internal code that uses the data.table package v1.10.4. After combining the results for each chromosome, we filtered the CpHs to keep only those where all samples had a minimum coverage of 5 (58,109,566 number passed the filter).

Lister, *et al.* data processing

We downloaded the WGBS data from Lister *et al.* data *(1)* with SRA accession SRP026048. We then processed and aligned the data following the same steps we used for our data. Using bsseq (36) as in the *CpH processing* section, we extracted the methylation values from the Bismark (31) report files and added the `c_context` and `trinucleotide_context` information per chromosome. We then merged the results for all the chromosomes retaining only the CpG and CpH positions we observed with a minimum coverage of 3 and 5 in our data, respectively. To assess the replication of our cell type DMR results we computed mean methylation differences across the CpG positions comparing neuron and non-neuron samples in the Lister *et al.* data (1) using the `rowttests()` function from the genefilter package version

1.56.0. We then computed the mean difference for each of the DMRs and compared this mean difference against the DMR mean methylation difference derived from our data to derive the concordance and correlation between them. To assess the replication of our age DMR results we modeled age as a continuous variable and calculated the mean methylation difference per year for every CpG contained in the age DMRs using `lmFit()` function from limma. Finally, we compared the mean methylation difference in the Lister *et al.* data (1) against the observed mean methylation difference for the age DMRs we derived.

Homogenate data processing

We processed the prenatal and postnatal homogenate brain samples using the same procedure described in the *Data processing/alignment* section to produce Bismark (31) report files. Then using bsseq (36) we extracted the methylation values for the CpG positions observed in our postnatal samples in order to make them comparable to each other.

Identification of differentially methylated positions:

With the set of CpGs and CpHs with a minimum coverage 3 and 5 in all samples, respectively, and the same models for identifying the DMRs, we identified the differentially methylated positions (DMPs) keeping the CpGs and CpHs separate. For the CpHs we further filtered to keep only those where at least 5 samples had a methylation value greater than 0 (40,818,742, 70.2%). We used the limma Bioconductor package v3.30.13 for determining the DMPs by running functions `lmFit()` and `eBayes()` with default parameters and FDR<5%.

Enrichments for genes related in brain disorders

We calculated the relative enrichments of genes overlapping different methylation features with the gene sets described by Birnbaum et al (23) and all genes overlapping the GWAS loci for the diseases identified above. For the genes identified for each disorder, we compared what fraction overlapped DMRs, and compared these ratio to the rest of the genome using a chi-squared test. We corrected for testing multiple disorder gene sets using the false discovery rate (FDR).

Enrichments for GWAS loci, HARs and enhancers

We calculated the relative enrichments of genomic segments overlapping cdDMRs and GWAS loci, human accelerated regions (HARs) and enhancers (15) using chi-squared tests. Here we used the set of CpG clusters used to identify the DMRs and calculated enrichments of clusters overlapping both cdDMRs and HARs, enhancers, or GWAS loci. We corrected for testing multiple disorder gene sets using the false discovery rate (FDR).

RNA-seq data processing:

Raw sequencing reads were mapped to the hg19/GRCh37 human reference genome with splice-aware aligner HISAT2 v2.0.4 (40). Feature-level quantification based on GENCODE release 25 (GRCh38.p7) annotation on hg19 coordinates was run on aligned reads using featureCounts (subread v1.5.0-p3) (41). Using custom R code we processed the different feature counts and created RangedSummarizedExperiment objects using the SummarizedExperiment Bioconductor package v1.4.0.

Percent spliced in (PSI) calculation:

We calculated the percent spliced in using the SGSeq (42) Bioconductor package v1.12.0 and the Gencode v25 annotation for the GRCh37 human reference genome (ftp.sanger.ac.uk/pub/gencode/Gencode_human/release_25/GRCh37_mapping/gencode.v25lift37.annotation.gtf.gz) from the BAM files generated by HISAT2. We used default arguments except for the function `analyzeVariants()` where we used a `min_denominator=10`.

Differential expression between cell types

We combined the gene counts for the polyA+ and RiboZero sequencing protocols for the cell-sorted RNA-seq data: 3 NeuN+ and 3 NeuN- samples for a total of 12 RNA-seq sequencing runs. We calculated the library size normalization factors using `calcNormFactors()` from edgeR (43) and identified the differentially expressed genes using `voom()`, `lmFit()` and `eBayes()` from limma (38, 44). We repeated this procedure for the exon counts.

Gene ontology analyses

Gene ontology enrichment analyses were performed using clusterProfiler (45) v3.6.0 using the options `pAdjustMethod="BH", pvalueCutoff=0.1, qvalueCutoff=0.05` on the ENSEMBL gene ids for each expression feature for the BP, CC and MF GO ontologies.

Methylation vs expression associations:

With the sorted RNA-seq data, we computed the average gene expression per age group (infant, child, teen) and cell type (six groups), and correlated these values to average DNAm levels in the same groups in both the CpG and CpH contexts at the gene promoter and body, exon (500bp window), and splice junction (50bp into each intron) levels.

With the RangedSummarizedExperiment objects with the RNA-seq polyA homogenate data and the bsseq objects with the CpG and CpH data we determined which CpG and CpH positions explained changes in expression (RPKM) at the gene or exon level as well as in percent spliced in (PSI). We retained only the expression and PSI data from the postnatal samples and matched them by brain identifier to the neuronal methylation data with a final sample size of 22. We filtered lowly expressed features using the `expression_cutoff()` function from the jaffelab package v0.99.18: mean RPKM>0.22 mean for genes, 0.26 for exons. For genes and exons we transformed the expression values to log2(RPKM + 1) and extracted the raw PSI values. Using MatrixEQTL (46) v2.2 (GitHub b9a9f01 patch) we then identified the methylation quantitative trait loci (QTL) for the CpG and CpH methylation data separately using the function `Matrix_eQTL_main()` function with options `pvOutputThreshold=0, pvOutputThreshold.cis=5e-4, useModel=modelLINEAR, cisDist=1000`. We identified marginal CpG associations near CpH associations by running MatrixEQTL again for the CpG in a +- 1kb window around the CpH positions with an association with expression at FDR <5% for each expression feature type using the same parameters as above except for `pvOutputThreshold.cis=0.01`. We filtered the associations to retain only those having at least 11 samples with non-zero methylation and 11 samples with non-one methylation values to remove extreme cases. We further restricted the results to protein coding genes and dropped any with infinite t-statistics. To assess whether age confounds the relationship between methylation and expression, we used a multiple linear regression model adjusting for age and checked if the methylation coefficient was still FDR<5%. Venn diagrams in **Figure 3** were made with the VennDiagram package v1.6.18. For the gene and PSI associations we used the gene id to check if it was present in the 3,473 differentially expressed genes from the sorted RNA-seq data (described above) with higher expression in neurons and the top 5,000 DE genes with

higher expression in glia at FDR<5%; similarly we did so for exons and the top 5,000 DE exons (FDR<5%) with higher expression in each cell type. The LIBD WBGS Expression explorer at https://jhubiostatistics.shinyapps.io/wgbsExprs/ was made using the bsseq (36) v1.14.0, DT v0.4, SGSeq (42) v1.12.0 and shiny v1.0.5 R packages.

Global autocorrelation:

Using CpG and CpH positions with a minimum coverage of 3 and 5 respectively for all samples we calculated the autocorrelation for the methylation levels for the CpGs, the CpHs, the CpHs with a CHG trinucleotide context, or the CpHs with a CHH trinucleotide context. For each of the sets, we grouped the positions using derfinder v1.12.0 into groups by a maximum distance of 1 kb. Only those groups with at least 5 Cs were further considered. For each sample we then calculated the autocorrelation using the `acf()` function with `lag.max=4` in parallel for each chromosome using BiocParallel v1.12.0. For each cluster of cytosines we calculated the mean across the neuronal (NeuN+) and the glia (NeuN-) samples at each autocorrelation lag. After combining and tidying the results, we visualized the global auto correlation using ggplot2 v2.2.1. We repeated this same analysis for the Lister *et al.* data (1).

Autocorrelation within DMRs:

Similar to the global autocorrelation, we extracted the methylation values at CpGs with a minimum coverage of 3 and the CpHs with a minimum coverage of 5 that were within each of the sets of DMRs (age, cell type or interaction). We then computed the autocorrelation for DMRs with a least 5 different cytosines using the `acf()` function with a `lag.max=4` and calculated the mean auto-correlation among the neuronal and glial samples.

Tables

Feature	Meth. Type	Direction	N	N Unique C	N Unique Features	Mean N Meth>0	Mean N Meth<1	Mean Expr Change	Prop. Age Confounded FDR >=5%	Prop. In Promoter	Prop. In Gene Body	Prop. In Gene Flanking	Prop. CHG	Prop. CHH	Prop. DE Glia	Prop. DE Neuron	Prop No Diff
gene	CpG	down	18458	18353	2074	21.9	21.9	1.39	0.0932	0.0671	0.966	0.906	0	0	0.126	0.329	0.546
		up	5217	5021	1743	21.8	21.8	1.69	0.0845	0.0571	0.979	0.92	0	0	0.304	0.193	0.503
	CpH	down	25812	25712	1611	21.1	21.1	1.4	0.115	0.0161	0.519	0.75	0.22	0.781	0.0912	0.413	0.496
		up	3842	3840	846	20.1	20.1	1.66	0.2	0.0219	0.532	0.807	0.26	0.744	0.123	0.27	0.607
exon	CpG	down	9505	4536	4168	21.9	21.9	1.23	0.104	0.19	0.934	0.991	0	0	0.0097	0.0256	0.965
		up	1328	638	687	21	21	1.93	0.191	0.175	0.928	0.988	0	0	0.0369	0.0136	0.95
	CpH	down	7874	4082	3458	21.1	21.1	1.72	0.074	0.0328	0.437	0.927	0.19	0.814	0.0086	0.0451	0.946
		up	1288	733	962	18.7	18.7	1.94	0.154	0.0435	0.481	0.918	0.28	0.722	0.0031	0.0311	0.966
psi	CpG	down	4286	2477	1315	21.7	21.7	0.124	0.0272	0.0434	0.994	0.936	0	0	0.242	0.212	0.547
		up	5782	3915	1646	21.8	21.8	0.121	0.0195	0.0398	0.991	0.919	0	0	0.261	0.222	0.517
	CpH	down	1296	945	489	16.8	16.8	0.447	0.0201	0.0274	0.537	0.676	0.36	0.644	0.187	0.286	0.526
		up	875	536	391	16.5	16.5	0.169	0.032	0.0279	0.559	0.775	0.36	0.637	0.194	0.33	0.477

Table 1: Summary of associations. Mean values are shown for the number of samples with methylation greater than zero, less than 1 and expression change variables: either ΔPSI or Δlog2(RPKM + 1) for genes and exons. Columns starting on “age confounded” are proportions. Proportion of cytosines overlapping different gene sections, cytosine context for CpH, and whether the gene/exon (gene for PSI) are in the up to top 5,000 differentially expressed genes between glia and neurons. See **Table S10** for a full description of the variables.

Figures

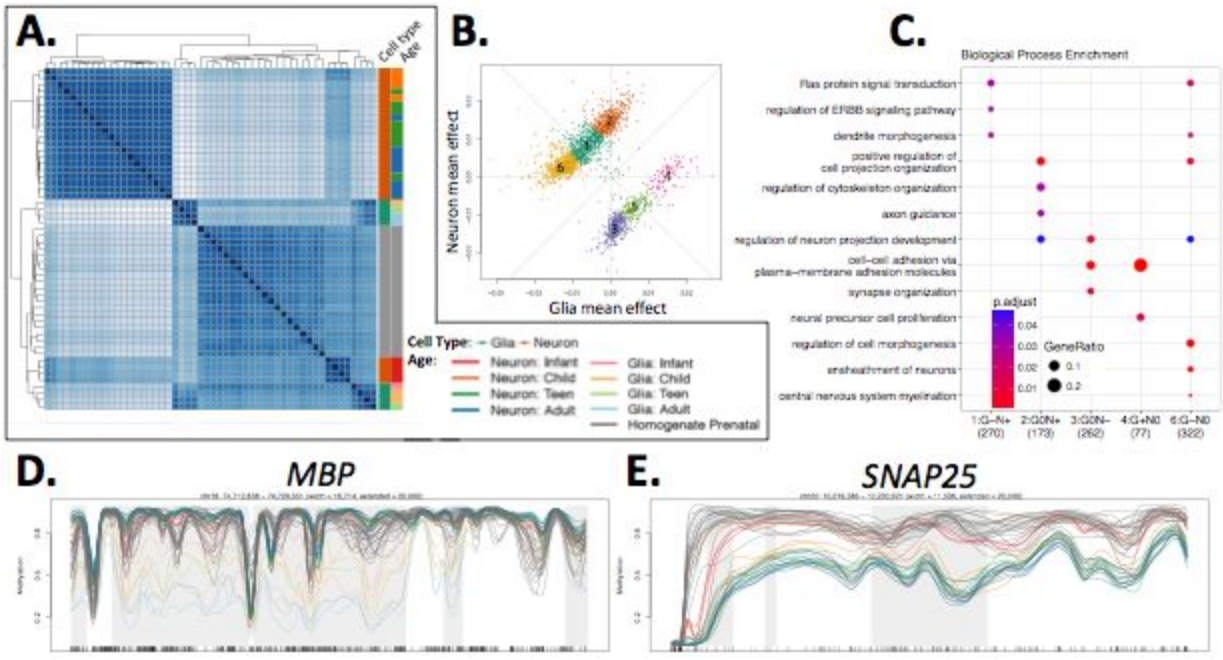

Figure 1: Regional cell type-specific developmental mCpG trajectories. (A) Euclidean distances between samples within cdDMRs shows that older neuronal samples cluster separately from infant neuronal samples, glial samples regardless of age, and prenatal homogenate cortex samples. **(B)** Decomposing cdDMRs patterns into into 6 clusters using *k-means* based on glia and neuron mean mCpG changes per year of life. **(C)** Gene ontology enrichment based on the six groups highlights diverse biological processes. No terms were enriched for Group 5. **(D)** Example cdDMR for Group 6 within *MBP* and **(E)** Example cDMR of Group 3 within *SNAP25.*

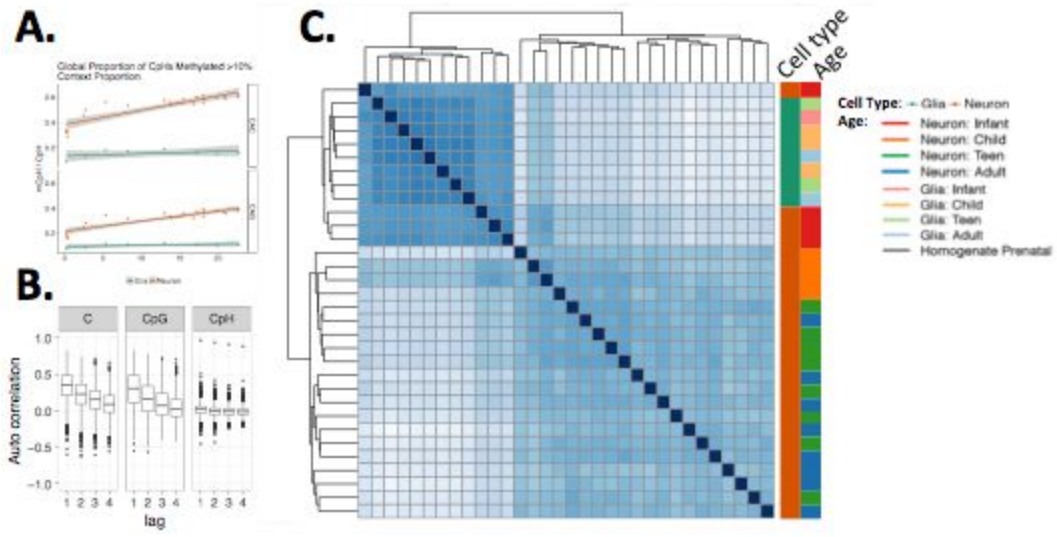

Figure 2: CpH methylation patterns across brain development. (A) Increasing mCpH in the CAC and CAG trinucleotide contexts in neurons across brain development. **(B)** Autocorrelation levels were similar for mCpG and all cytosines, with uncorrelated levels in the CpH context. mCpH levels also discriminated older neurons from both infant neurons and glia within mCpG-based cdDMRs. **(C)** Euclidean distances between samples based on mCpH within cdDMRs cluster into two groups in which infant neurons (orange cell type, dark red age bar) cluster with glia of all ages (green cell type bar), and older neurons form the second group.

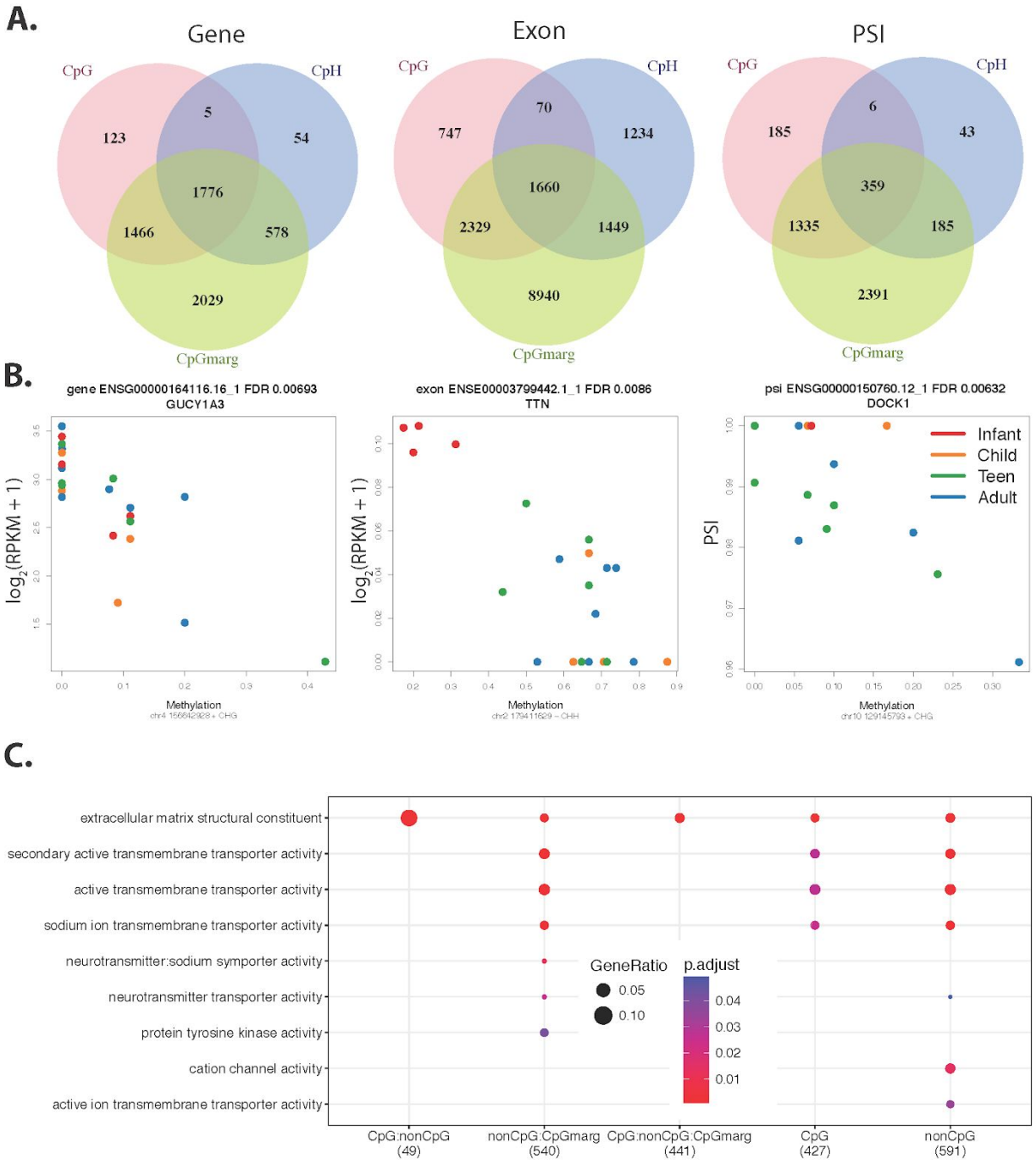

Figure 3: Methylation associations with expression. **(A)** Venn diagrams of the methylation associations by unique feature for the gene, exon and PSI features. The sets are determined by whether the association is FDR<5% genome-wide for CpG and CpH or if it is a CpG marginally significant within +/- 1 kb window of a CpH association. **(B)** Example associations between

methylation and expression at the gene level with colors by age group: red - infant, orange - child, green - teen, blue - adult. *GUCY1A3* is one of the top CpH that is differentially expressed between neurons and glia; Expression of an exon within *TTN*, a gene associated with autism, is negatively associated with mCpH. Lastely, DOCK1 PSI of an alternative end site is negatively associated with mCpH. **(C)** Enriched molecular function ontology terms for exons associated to methylation by the venn diagram groups from (A).

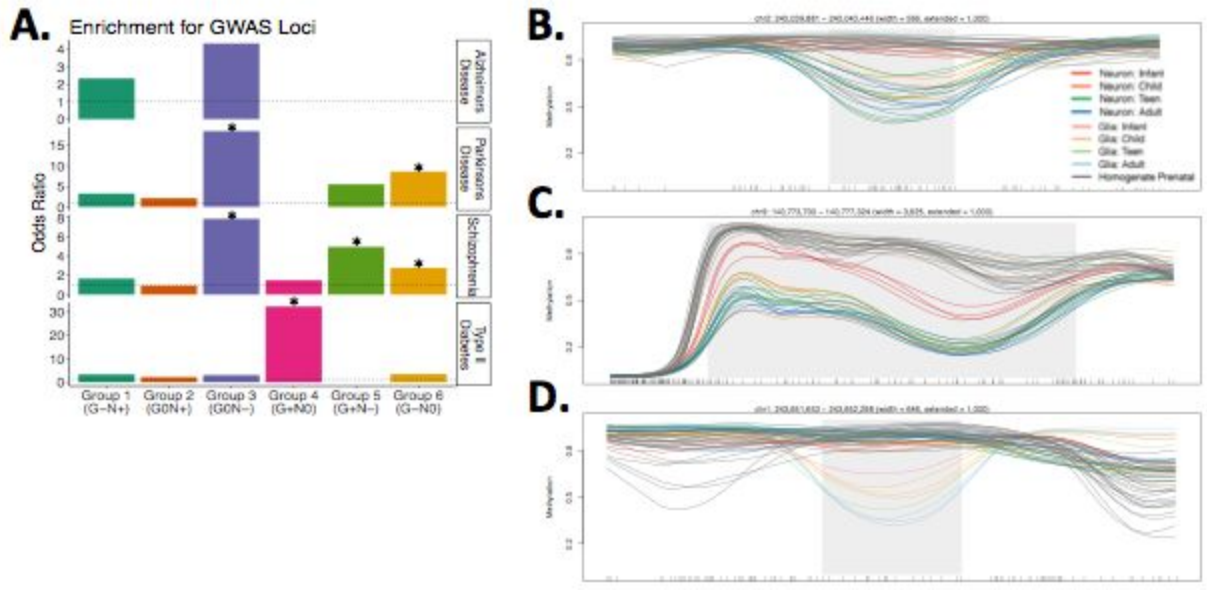

Figure 4: DNAm patterns and psychiatric disease risk. **(A)** Odds ratio for enrichment of cdDMR groups within genome-wide association study (GWAS) risk loci. Asterisks indicate FDR<0.05. **(B)** A cdDMR overlapping *HDAC4*, a gene associated with autism, shows the Group 3 pattern of decreasing neuronal and static glial DNAm. **(C)** A cdDMR overlapping *CACNA1B*, a gene associated with autism, shows the Group 5 pattern of decreasing neuronal and increasing glial DNAm. **(D)** A cdDMR overlapping *AKT3*, a gene associated with schizophrenia, shows the Group 6 pattern of decreasing glial and static neuronal DNAm. Tick marks in the region plots indicate individual CpGs.

Supplementary Figures

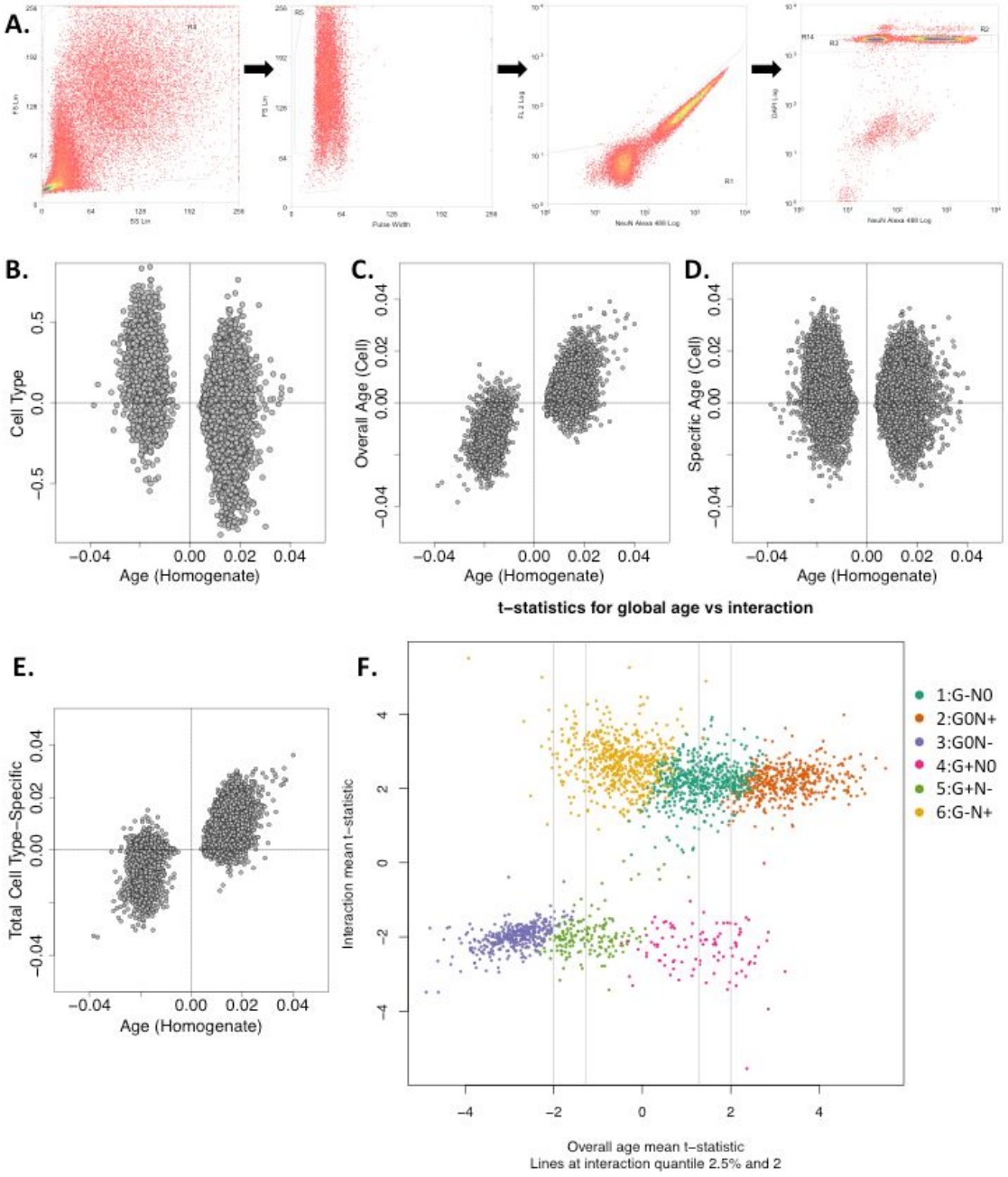

Figure S1: Detecting developmental changes in homogenate vs. cell type-specific DNAm data. (A) Representative gating strategy from fluorescence-activated nuclear sorting by NeuN antibody. Debris is first reduced by selecting events based on forward scatter and side scatter, then aggregates are reduced by measuring pulse width. Autofluorescent events are discarded

by measuring true Alexa fluor-488 signal compared to signal in the FL2 channel. Finally, singlets are determined by DAPI staining, and Alexa fluor-488 positive and negative events are collected. **(B)** Developmental age effect coefficients of individual CpGs as measured in homogenate samples compared against cell type effect coefficients in cell type-specific samples after adjusting for age. **(C)** Developmental age effect coefficients of individual CpGs as measured in homogenate samples compared against the developmental effect adjusting for cell type in cell type-specific samples. **(D).** Developmental age effect coefficients as measured in homogenate samples compared against the cumulative age and cell type interaction effect coefficients in cell type-specific samples at the CpG level. **(E)** Developmental age effect coefficients as measured in homogenate samples compared against the estimated cell type-specific age effects for cell type-specific samples at the CpG level. **(F)** Overall age mean t-statistic for the cdDMRs (x-axis) against the mean interaction t-statistic (y-axis) with the 2.5% quantile from the y-axis shown on the x-axis. Colors are the same as those from **Figure 1** B: teal=Group 1 (decreasing glial methylation, increasing neuronal methylation), orange=Group 2 (static glial methylation, increasing neuronal methylation), purple=Group 3 (static glial methylation, decreasing neuronal methylation), pink=Group 4 (increasing glial methylation, static neuronal methylation), green=Group 5 (increasing glial methylation, decreasing neuronal methylation), gold=Group 6 (decreasing glial methylation, static neuronal methylation).

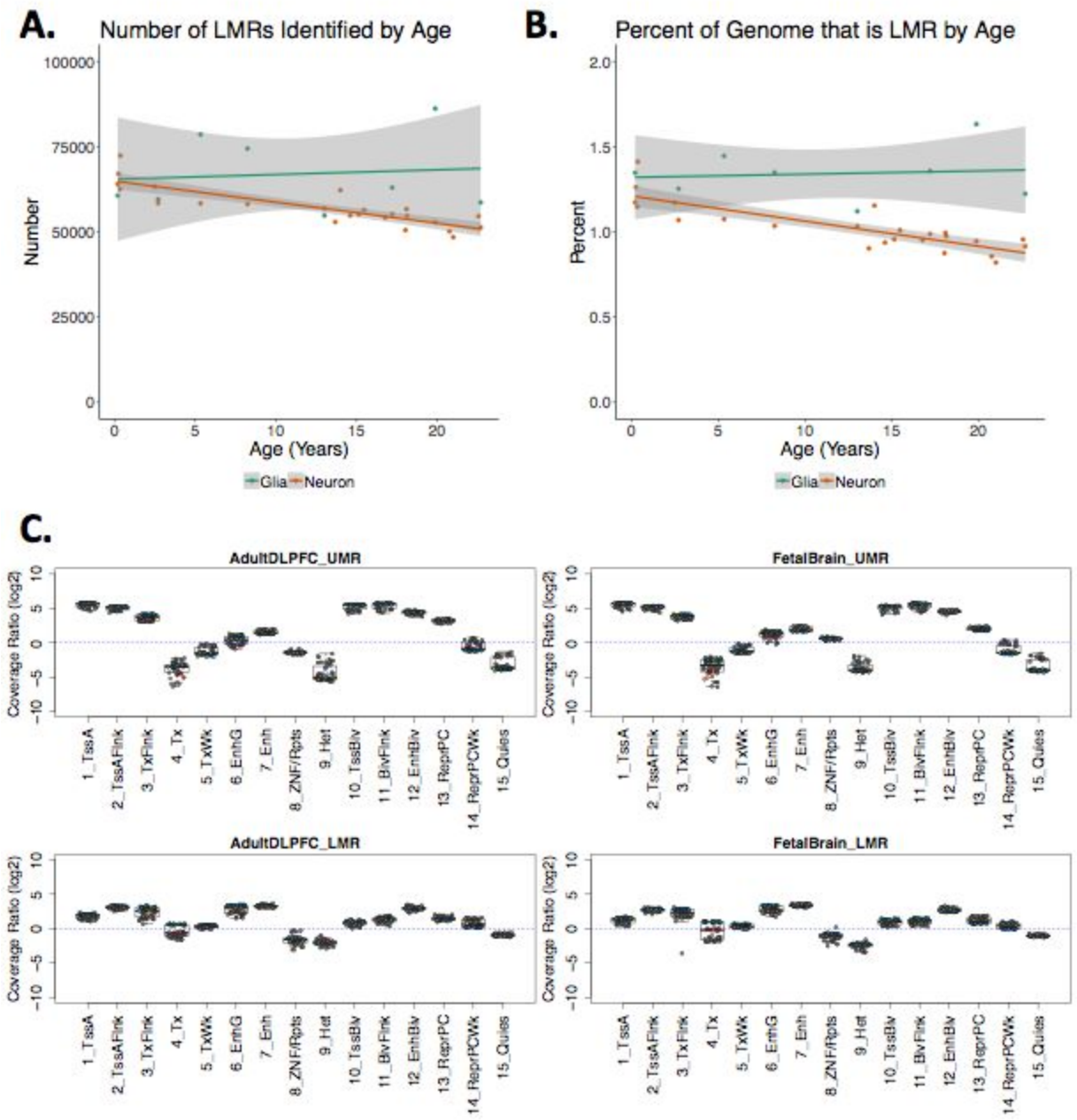

Figure S2: Unmethylated Regions (UMRs) and Low-methylated regions (LMRs). (A) Number of LMRs identified by age stratified by cell type. **(B)** Percent of the genome covered by LMRs by age stratified by cell type. **(C)** Roadmap Epigenomics Consortium chromatin enriched states for LMRs and UMRs as defined in adult DPLFC and fetal brain. log2(Coverage Ratio)

represents the enrichment of the proportion of bases within LMRs or UMRs in a chromatin state compared to the rest of the genome.

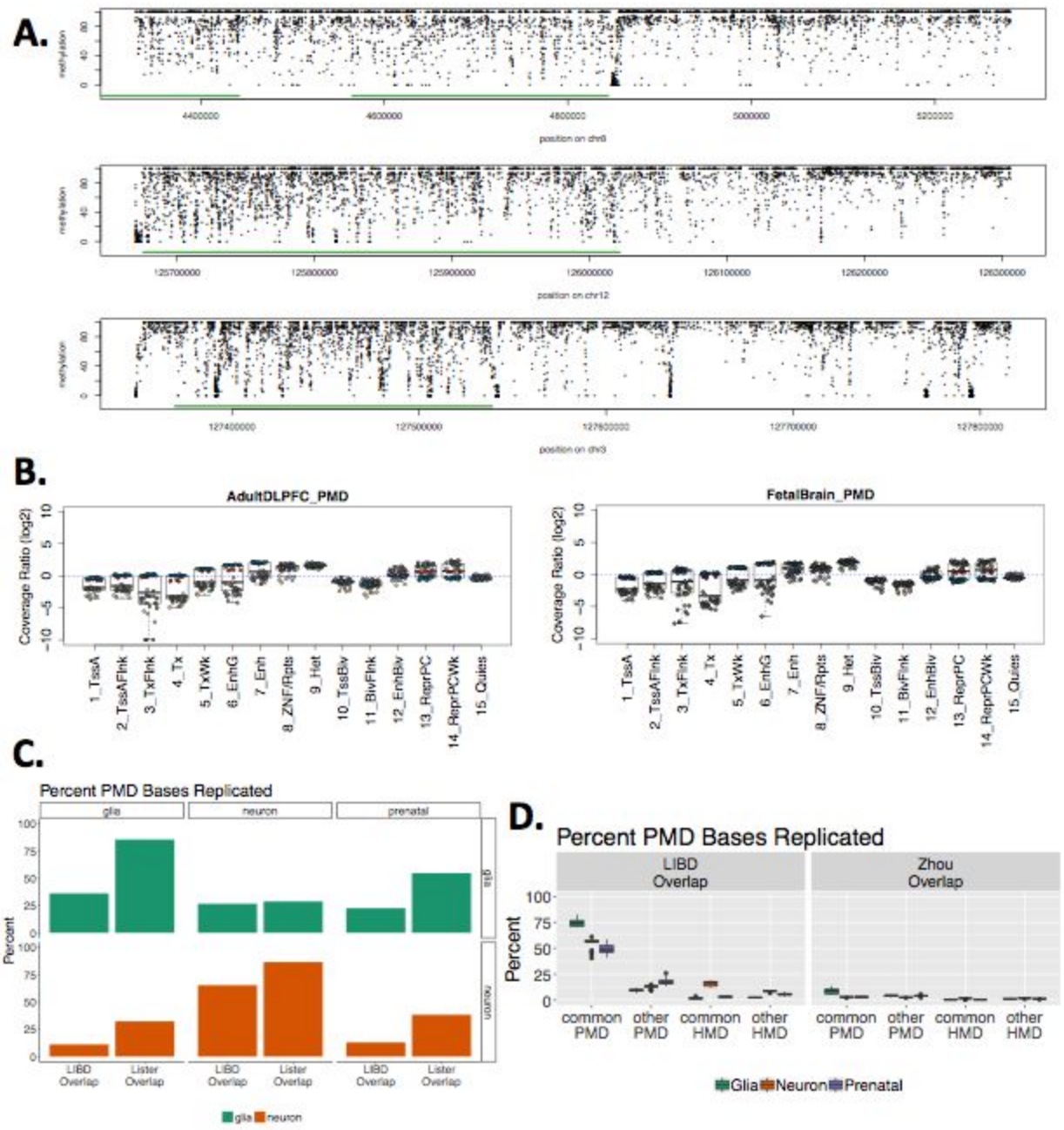

Figure S3: Partially methylated domains (PMDs). (A) Example PMDs on chromosomes 8, 12, and 3. PMDs are highlighted in green. **(B)** Roadmap Epigenomics Consortium chromatin state enrichment for PMDs. log2(Coverage Ratio) represents the enrichment of the proportion of bases within LMRs or UMRs in a state compared to the rest of the genome. **(C)** Percent of PMD base-pairs that are replicated in the Lister *et al.* dataset (1). Rows represent Lister glia and

neurons, and columns represent the glia, neurons and prenatal samples from this paper. Each bar represents the percent of PMD bases shared between that quadrant's cell types, using either total LIBD or Lister PMD bases as the denominator. **(D)** Percent of PMD base-pairs per sample that are replicated in either common or unique PMDs and high methylated domains (HMDs) identified in Zhou *et al.* Samples are colored based on whether the sample was postnatal neuron, postnatal glia, or bulk prenatal cortex.

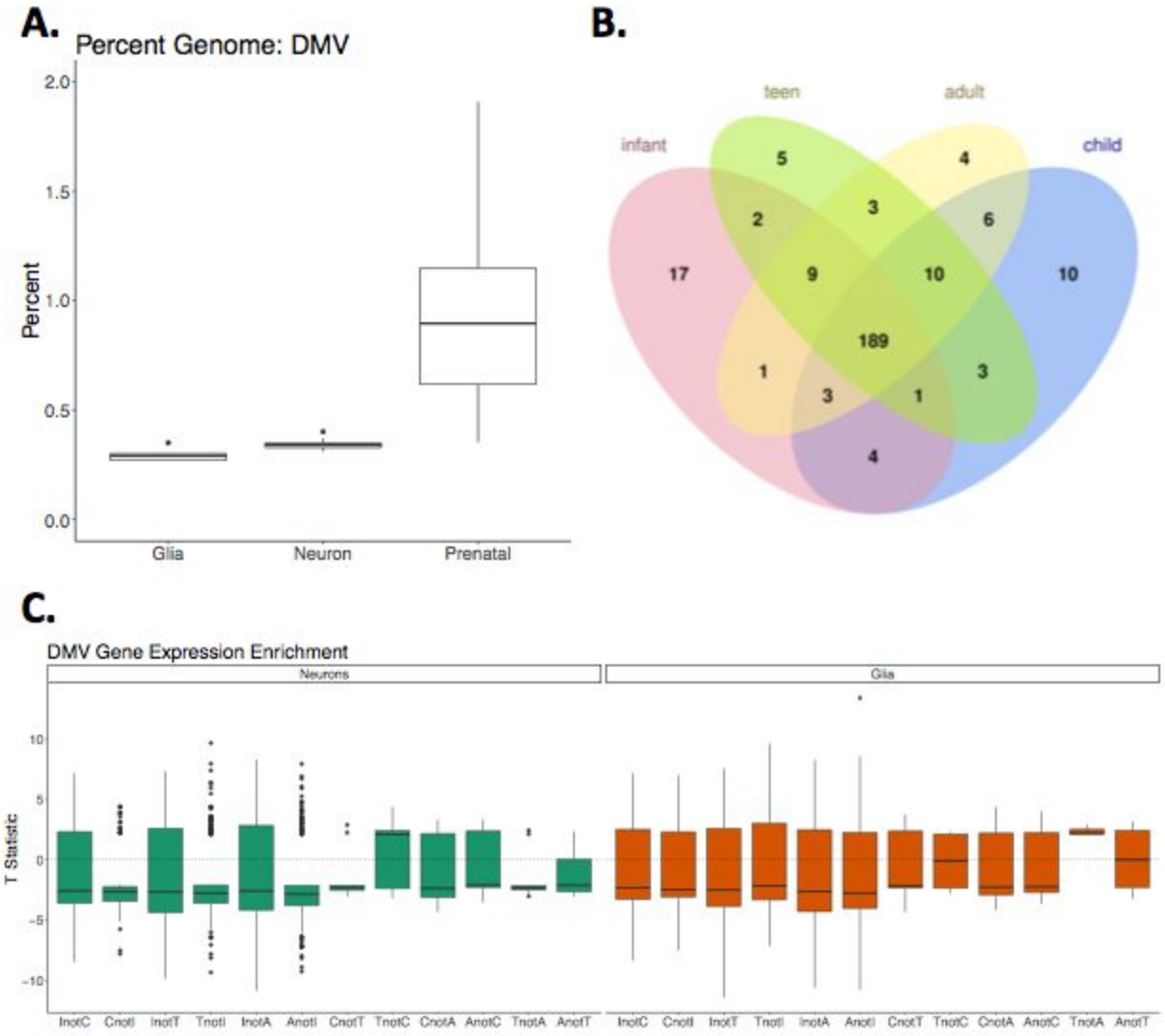

Figure S4: DNA methylation valleys (DMVs). (A) Percent of the genome covered by DMVs in postnatal glia and neurons and bulk prenatal cortex. **(B)** Overlap of transcription factor (TF) genes within DMVs by age in neurons. **(C)** Gene expression enrichment between age groups in TF genes excluded from DMVs in one group but not the other in neurons and glia. A negative T-statistic signifies greater expression in the age group in which the gene is not in a DMV. I=Infant (0-1 years); C=Child (1-10 years), T=Teen (11-17 years) and A=Adult (18+ years).

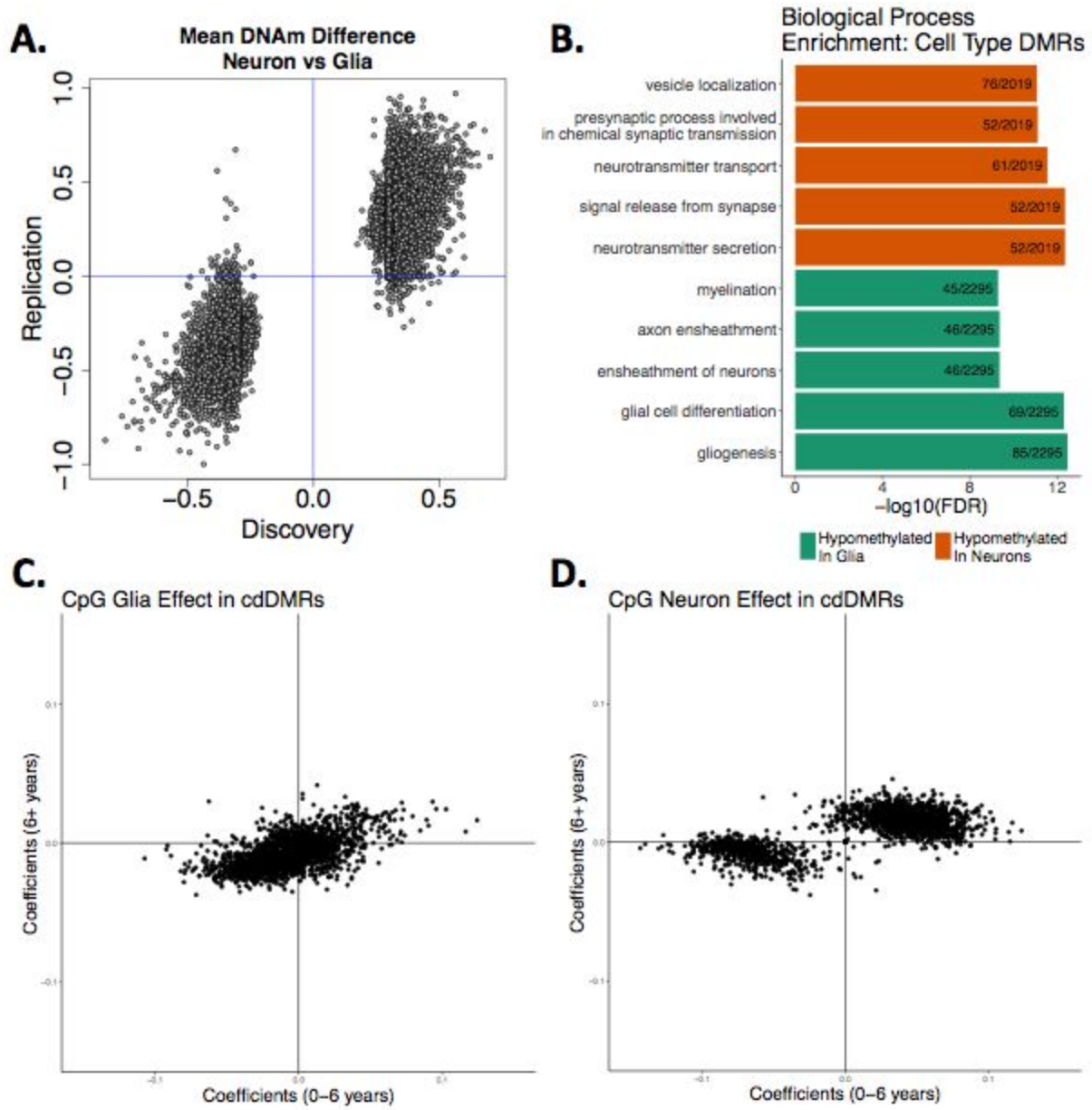

Figure S5: Differentially methylated regions (DMRs). (A) Replication of the DNAm differences at the CpG level between neurons and glia (NeuN+ and NeuN- samples) in our data and Lister *et al.* data *(1)*. (**B)** Enriched biological process ontology terms in the DMRs by cell type. (**C)** Coefficients of linear regression on the mean mCpG level per cdDMR in samples younger than 6 years and older. **(D)** Coefficients of linear regression on the mean mCpG level per cdDMR in samples younger than 6 years and older.

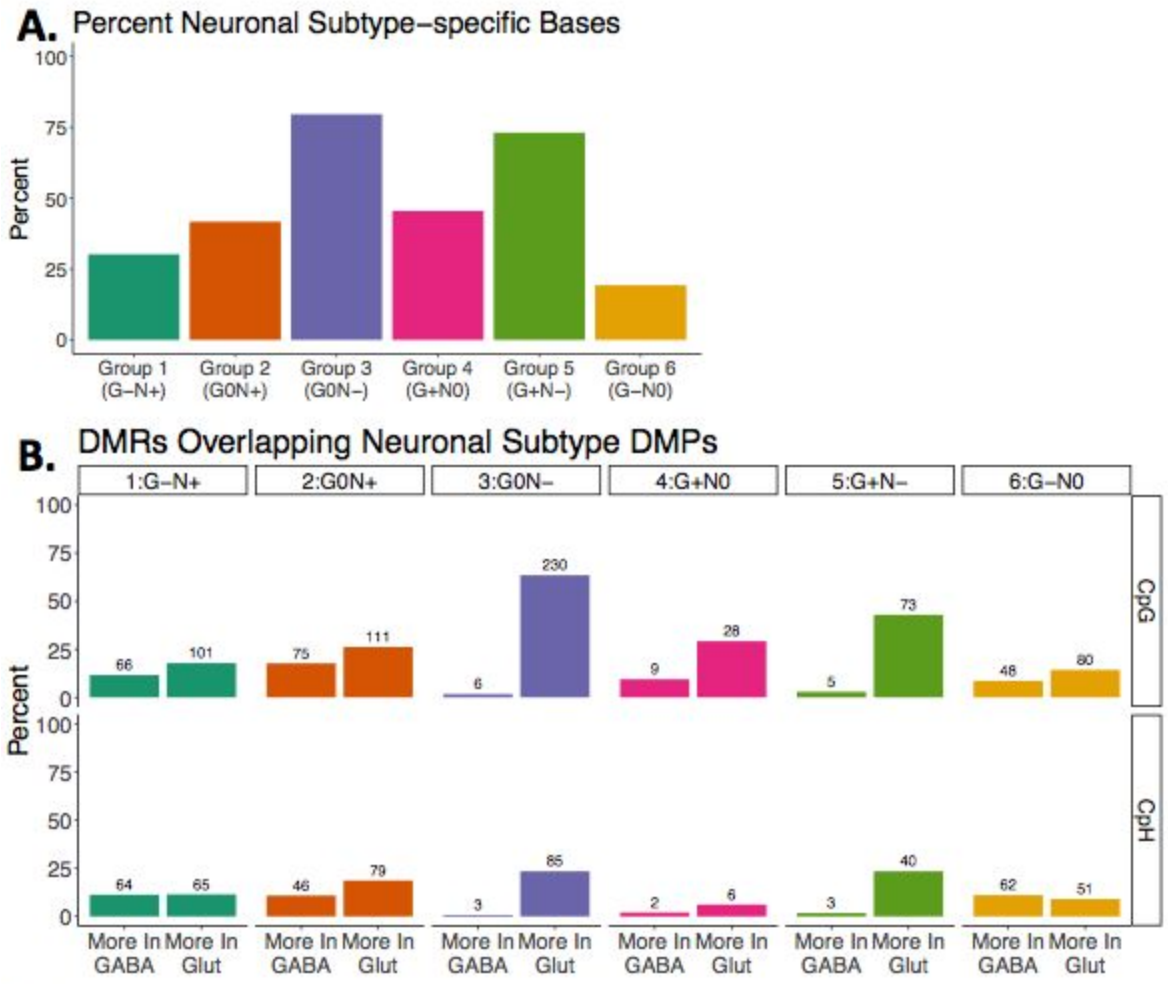

Figure S6: Cell type-specific developmental DMRs and neuronal subtypes. (A) Percent of bases within each group of cdDMRs as defined in **Figure 1B** that are differentially methylated by neuronal subtype in Luo *et al.* (14). **(B)** Percent of cdDMRs as defined in **Figure 1B** overlapping differentially methylated positions (DMPs) in the CpG and CpH contexts stratified by whether they are more highly methylated in GABAergic or glutamatergic neurons as defined by Kozlenkov *et al. (28)*. The number of cdDMRs are listed above the bar graph.

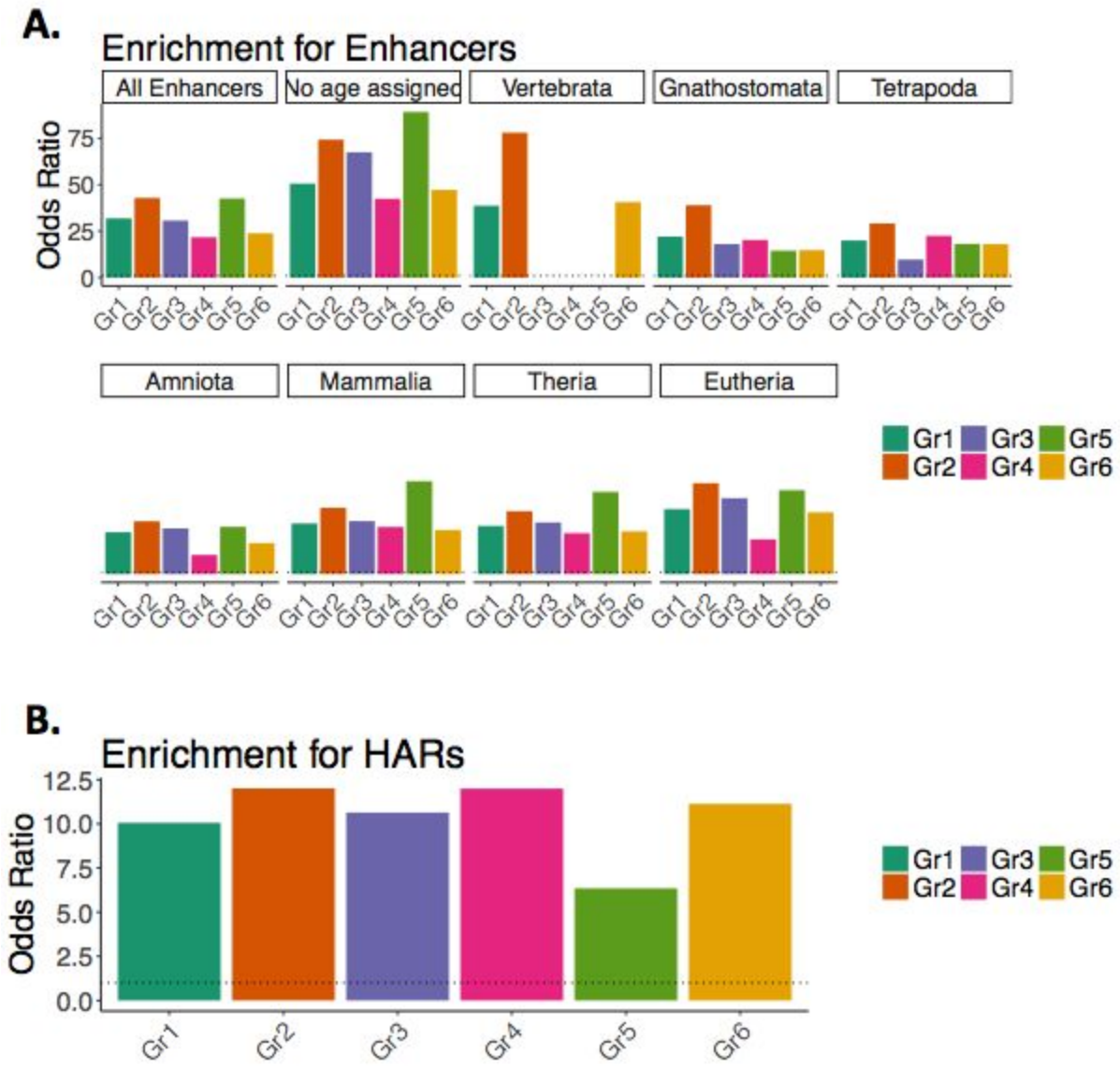

Figure S7: cdDMR overlap of human brain developmental enhancers and Human Accelerated Regions (HARs). (A) Enrichment for enhancers in the six clusters of cdDMRs from **Figure 1B** at the CpG-level. **(B)** Enrichment for human accelerated regions (HARs) in the six clusters of cdDMRs from **Figure 1B** at the CpG-level.

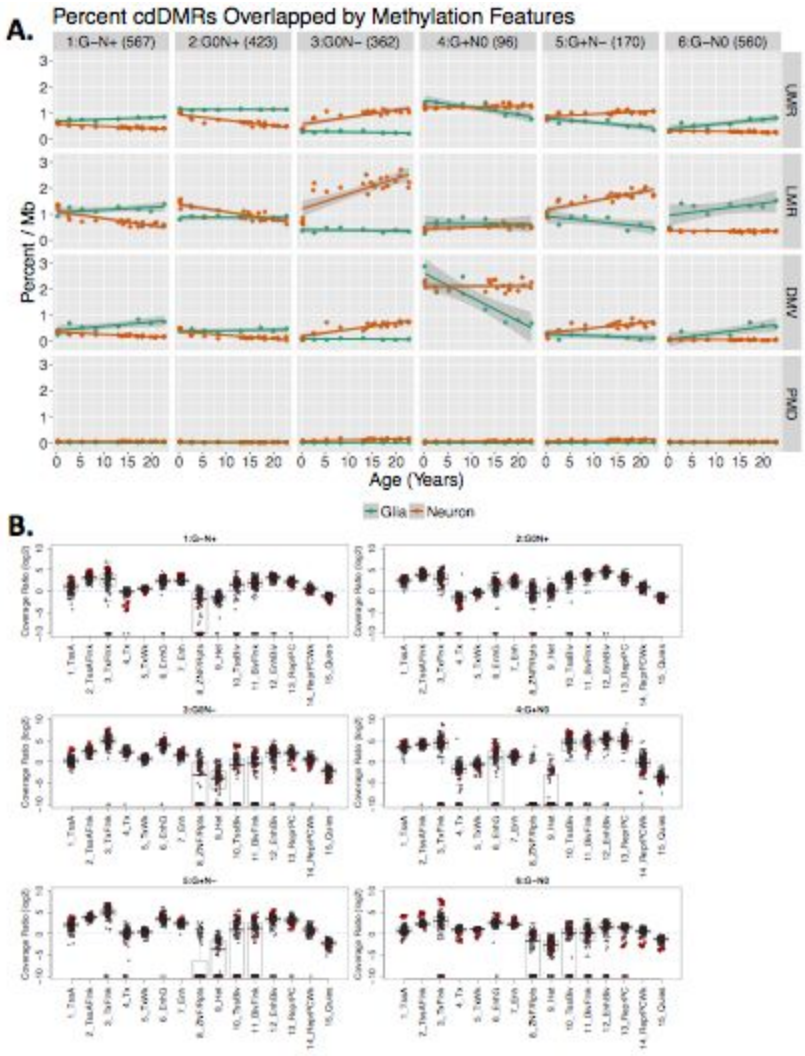

Figure S8: Cell type-specific, developmentally dynamic DMRs (cdDMRs) and epigenetic states. (A) Percent of cdDMRs by the k-mean groups from **Figure 1B** overlapped by UMR, LMR, DMV and PMD sequence across age. **(B)** Roadmap Epigenomics Consortium enriched chromatin states for the six clusters of cdDMRs from **Figure 1B** at the CpG-level.

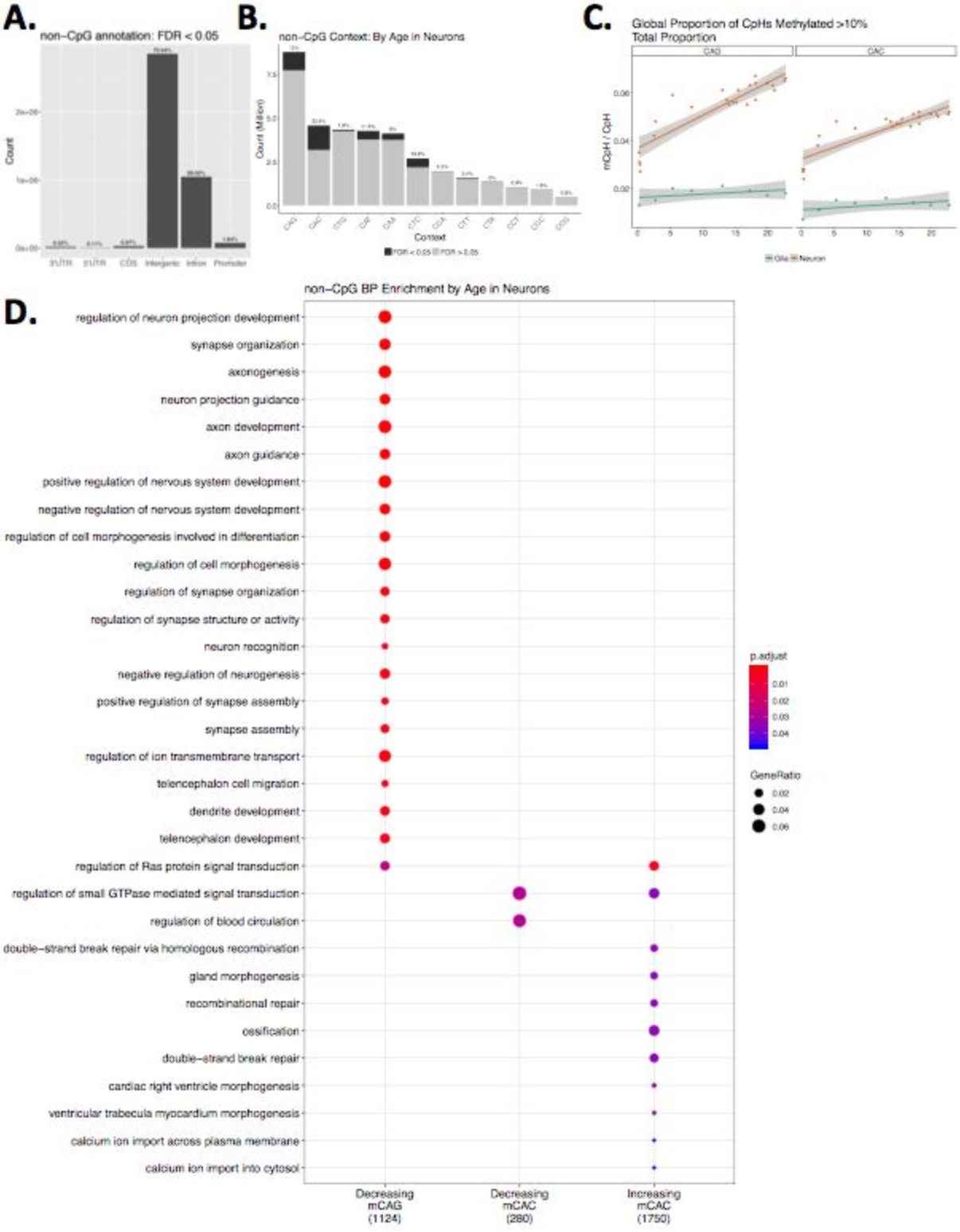

Figure S9: CpH methylation distribution, levels and context-specific biological process ontology. (A) Number of differentially methylated CpHs by cell type (FDR<5%) falling in different genomic annotations across the genome. Annotation was prioritized CDS > 5'UTR > 3'UTR > Intron > Promoter > Intergenic. **(B)** Breakdown of measured CpH by trinucleotide context. Differentially methylated mCpH by age in neurons (FDR<0.05) are colored in dark gray. **(C)** Accumulation of mCpH by trinucleotide context over development, stratified by cell type. **(D)** The top 20 biological process ontology terms enriched for genes exclusively overlapping CAG and CAC sites with significantly increasing or decreasing methylation levels in neurons over postnatal development (FDR<0.05).

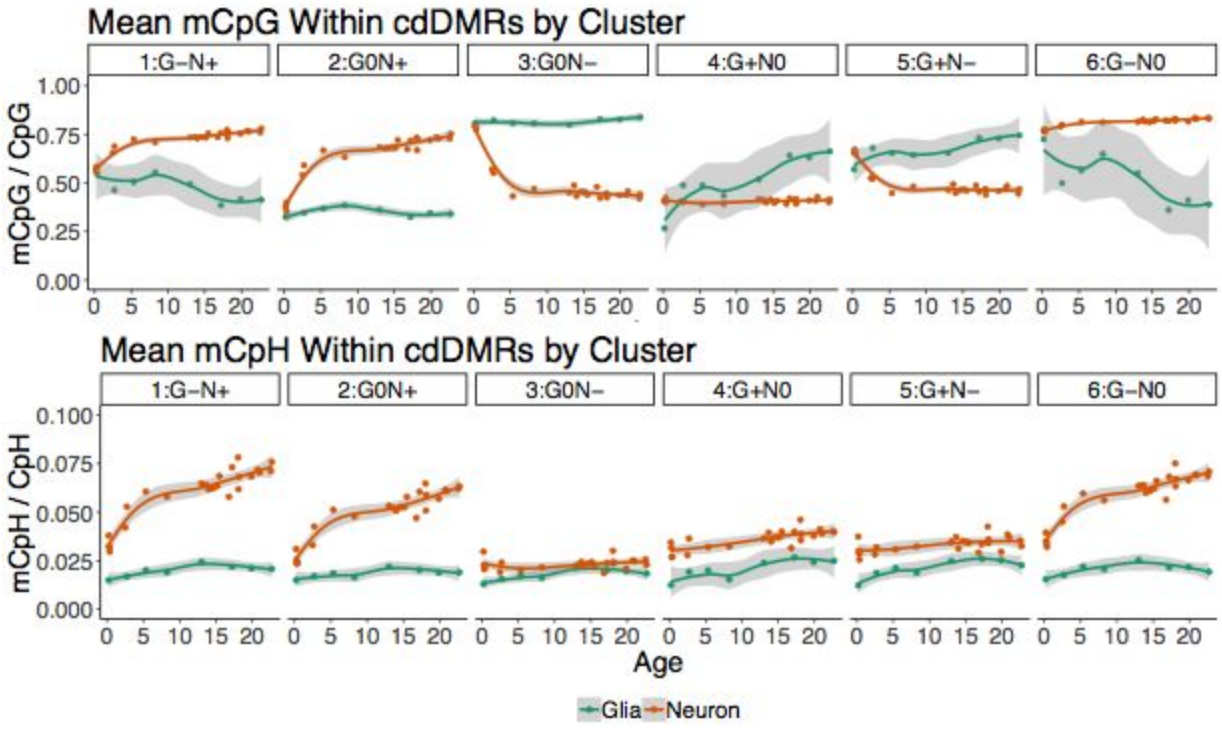

Figure S10: Trajectories of methylation accumulation in cdDMR groups. Mean mCpH and mCpG within the cdDMRs by cluster from **Figure 1B** across development.

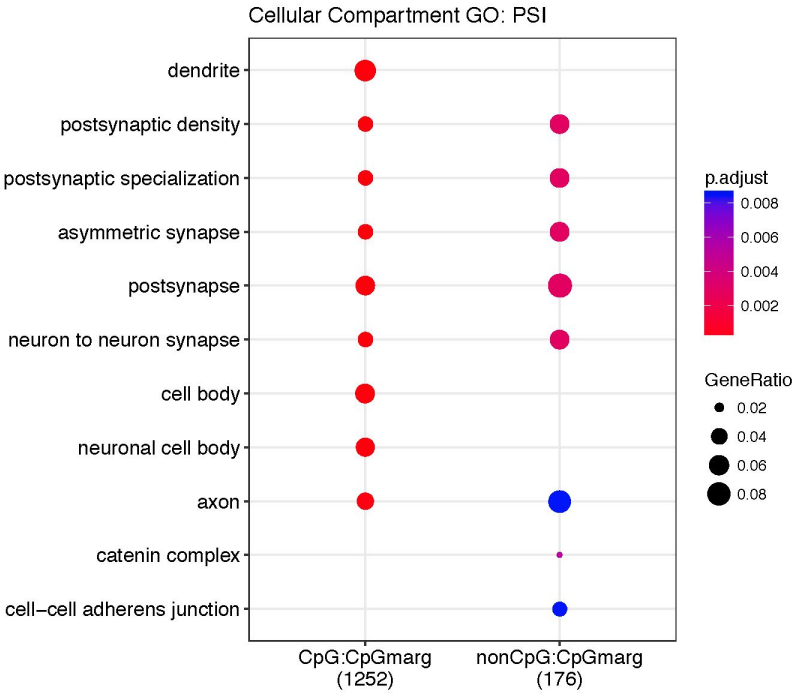

Figure S11: Cellular compartment ontology. Genes containing splicing events as measured by “percent spliced in” (PSI) that are associated with changing CpG and CpH methylation are enriched for cellular compartment gene ontology terms relating to neuronal features.

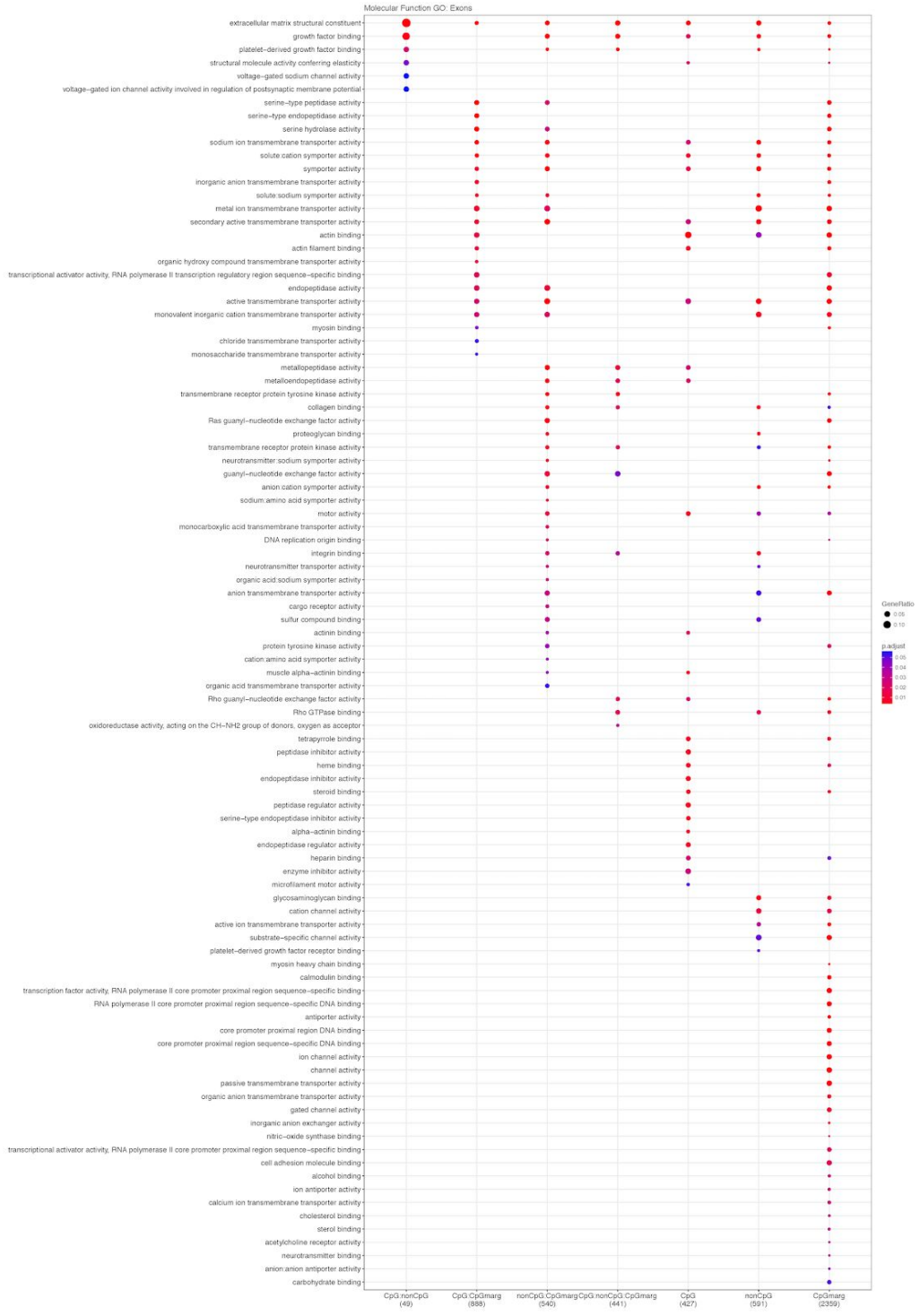

Figure S12: Molecular function ontology. Genes containing alternative exons that are associated with changing CpG and CpH methylation are enriched for molecular function ontology terms relating to neurons.

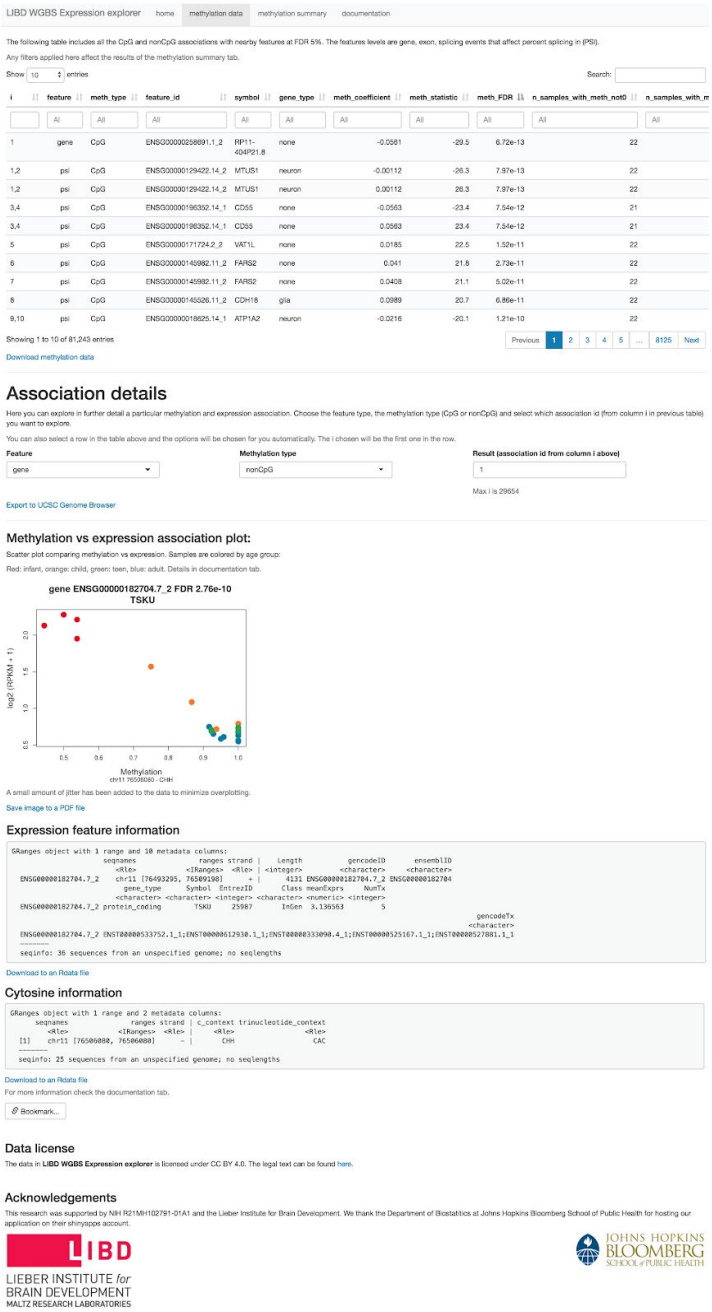

Figure S13: Web meQTL browser display. Interactive display of the CpHs and CpGs associated with expression at FDR<5% as shown at https://jhubiostatistics.shinyapps.io/wgbsExprs/. This screenshot shows the top nonCpG (mCpH) meQTL association at the gene expression level. Information about the gene is shown under *expression feature*, and information of the methylated C is shown under *cytosine*.

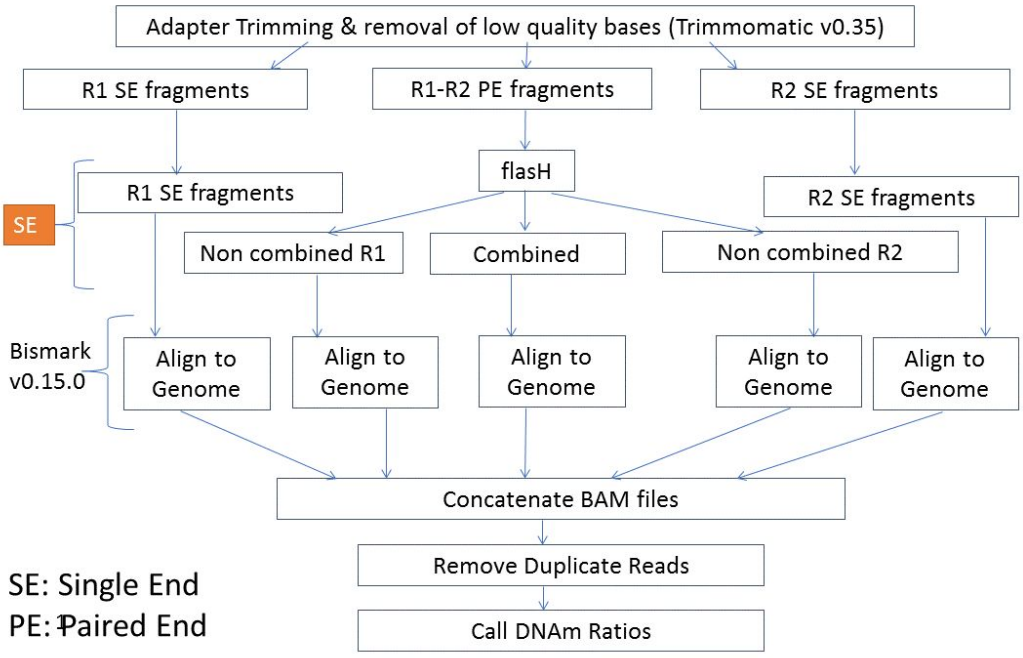

Figure S14: **WGBS data processing pipeline overview**. Diagram of the software and steps for processing the WGBS data. Trimmomatic was used to trim bases, then ... R1: read one from a fragment; R2: read two from a fragment; SE: single-end; PE: paired-end.

Supplementary Tables

Table S1: Phenotype and sequencing metrics for the WGBS samples. *See Table S1 WGBS phenotype and sequencing data.xlsx.*

Table S2: Phenotype and sequencing metrics for the RNA-seq samples. *See Table S2 RNA phenotype and sequencing data.xlsx.*

	N	CpG Context				CpH Context			
		# CpGs	Low (<20%)	Partial (20-80%)	High (>80%)	# CpHs	Low (<20%)	Partial (20-80%)	High (>80%)
Neuron	24	$18.7x10^6$	10.10%	13.90%	76.00%	$58.1x10^6$	91.60%	8.00%	0.40%
Glia	8	$18.7x10^6$	11.10%	17.80%	71.10%	$58.1x10^6$	99.40%	0.50%	0.10%
Fetal	20	$18.7x10^6$	12.10%	17.30%	70.60%	$58.1x10^6$	99.75%	0.17%	0.07%
Homogenate	23	$28.2x10^6$	8.50%	16.90%	73.70%	$58.1x10^6$	*NA*	*NA*	*NA*

Table S3: Number of cytosines measured and distribution of methylation by cytosine context. Sample sizes for each sample type, number of CpGs and CpHs and distribution of CpG and CpHs across low, partial and high levels of methylation.

Table S4: Cell type-specific, developmental differentially methylated regions (cdDMRs). *See Table S4 cdDMRs.xlsx.*

Model	**CpG Context**		**CpH Context**	
	FDR ≤ 0.05	**% Significant**	**FDR ≤ 0.05**	**% Significant**
Cell type	4824804	25.80%	7682075	18.80%
Age	536164	2.90%	3194618	7.80%
Cell type:Age	90227	0.50%	76	0%
Age in Neurons	*NA*	*NA*	4020371	9.80%

Table S5: mC association with cell type and age in postnatal cell type-specific samples. Number of differentially methylated CpGs and CpHs across each of the statistical models: cell

type effects adjusting for age, age effects adjust for cell type, cell type and age interaction effects, as well as age in neuron samples only.

Type	Location	Glia:Child	Glia:Infant	Glia:Teen	Neuron:Child	Neuron:Infant	Neuron:Teen
CpG	Promoter	-0.408	-0.417	-0.363	-0.413	-0.418	-0.357
CpH	Promoter	-0.0434	-0.0217	-0.0383	-0.23	-0.131	-0.249
CHG	Promoter	-0.046	-0.028	-0.048	-0.2	-0.12	-0.22
CHH	Promoter	-0.048	-0.015	-0.039	-0.22	-0.13	-0.24
CpG	Body	-0.205	-0.179	-0.181	-0.27	-0.223	-0.233
CpH	Body	-0.069	-0.0437	-0.0624	-0.294	-0.168	-0.322
CHG	Body	-0.054	-0.033	-0.055	-0.25	-0.14	-0.28
CHH	Body	-0.066	-0.041	-0.057	-0.28	-0.16	-0.31
CpG	Exon	0.0508	0.0938	-0.0112	-0.017	0.0369	-0.038
CpH	Exon	-0.0314	-0.00565	-0.0407	-0.213	-0.103	-0.252
CHG	Exon	-0.0236	-0.00553	-0.0372	-0.177	-0.0769	-0.219
CHH	Exon	-0.0447	-0.0147	-0.0484	-0.217	-0.124	-0.249
CpG	Junction	-0.0582	-0.087	-0.0461	-0.11	-0.124	-0.0501
CpH	Junction	-0.0324	-0.0311	-0.0265	-0.0866	-0.0609	-0.0804
CHG	Junction	-0.0218	-0.0116	-0.0213	-0.0764	-0.0579	-0.0836
CHH	Junction	-0.035	-0.0364	-0.0295	-0.0873	-0.063	-0.0767

Table S6: Feature expression association with DNAm. Correlation between expression of a feature (location) and DNA methylation by methylation context with CpH broken down by trinucleotide context stratified by age group and cell type.

Table S7: Enrichment of DMRs and mCpH for disease-associated genomic loci. *See sheet 1 of TableS7-9_Enrichment_for_disease-associated_gene_sets_and_loci.xlsx.*

Table S8: Enrichment of DMRs and mCpH for disease-associated gene sets. *See sheet 2 of TableS7-9_Enrichment_for_disease-associated_gene_sets_and_loci.xlsx.*

Table S9: Enrichment for disease gene sets in DNAm-splicing association features. *See sheet 3 of TableS7-9_Enrichment_for_disease-associated_gene_sets_and_loci.xlsx.*

Variable	Description
Feature	Either gene, exon or event affect the percent spliced in (PSI).
Methylation	Methylation type. Either CpG or CpH.
Direction	down.
N	Number of associations at FDR 5%.
N Unique C	Number of unique cytosines in these associations.
N Unique Features	Number of unique features in these associations.
N Meth>0	Mean number of samples that have non-zero methylation values. Total 22.
N Meth<1	Mean number of samples that have non-one methylation values.
Expr Change	Mean change in expression (ignoring NAs).
Age Confounded	Proportion of associations that are confounded by age, that is, proportion of associations where the methylation coefficient in a multiple linear regression adjusting for age has a FDR >= 5%.
Promoter	Proportion of cytosines in these associations that overlap gene promoters. A single cytosine can overlap multiple gene regions for different genes.
Gene Body	Proportion of cytosines in these associations that overlap gene bodies.
Gene Flanking	Proportion of cytosines in these associations that overlap gene flanking regions.
CHG	Proportion of cytosines in these associations that have a CHG context.
CHH	Proportion of cytosines in these associations that have a CHH context.
Glia	Proportion of features in these associations where the feature is differentially expressed between neurons and glia, with higher expression in glia. For genes and PSI, the gene list was used. Up to the top 5,000 features DE in each direction were used.
Neuron	Proportion of features in these associations where the feature is differentially expressed between neurons and glia, with higher expression in neurons.
No Diff	Proportion of features in outside the top differentially expressed genes between glia and neuron.

Table S10: Variable dictionary for Table 1. Variables used in **Table 1** with their full definition.

Bibliography

1. Lister R, et al. (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* 341(6146):1237905.
2. Jaffe AE, et al. (2016) Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat Neurosci* 19(1):40–47.
3. Jaffe AE, Irizarry RA (2014) Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 15(2):R31.
4. Schultz MD, et al. (2015) Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523(7559):212–216.
5. Mo A, et al. (2015) Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* 86(6):1369–1384.
6. Jeong M, et al. (2014) Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat Genet* 46(1):17–23.
7. Xie W, et al. (2013) Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* 153(5):1134–1148.
8. Stadler MB, et al. (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480(7378):490–495.
9. Lister R, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271):315–322.
10. Salhab A, et al. (2018) Partially methylated domains are hallmarks of a cell specific epigenome topology. *BioRxiv*.
11. Hovestadt V, et al. (2014) Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* 510(7506):537–541.
12. Roadmap Epigenomics Consortium, et al. (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539):317–330.
13. Zhou W, et al. (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* 50(4):591–602.
14. Luo C, et al. (2017) Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* 357(6351):600–604.
15. Emera D, Yin J, Reilly SK, Gockley J, Noonan JP (2016) Origin and evolution of developmental enhancers in the mammalian neocortex. *Proc Natl Acad Sci USA* 113(19):E2617-26.
16. He Y, Ecker JR (2015) Non-CG Methylation in the Human Genome. *Annu Rev Genomics Hum Genet* 16:55–77.

17. Shukla S, et al. (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479(7371):74–79.

18. Lambert JC, et al. (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 45(12):1452–1458.

19. Nalls MA, et al. (2014) Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* 46(9):989–993.

20. Morris AP, et al. (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 44(9):981–990.

21. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511(7510):421–427.

22. Skene NG, et al. (2018) Genetic identification of brain cell types underlying schizophrenia. *Nat Genet* 50(6):825–833.

23. Birnbaum R, Jaffe AE, Hyde TM, Kleinman JE, Weinberger DR (2014) Prenatal expression patterns of genes associated with neuropsychiatric disorders. *Am J Psychiatry* 171(7):758–767.

24. Stiles J, Jernigan TL (2010) The basics of brain development. *Neuropsychol Rev* 20(4):327–348.

25. Stroud H, et al. (2017) Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic States. *Cell* 171(5):1151–1164.e16.

26. Lev Maor G, Yearim A, Ast G (2015) The alternative role of DNA methylation in splicing regulation. *Trends Genet* 31(5):274–280.

27. Jaffe AE, et al. (2018) Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci* 21(8):1117–1125.

28. Kozlenkov A, et al. (2016) Substantial DNA methylation differences between two major neuronal subtypes in human brain. *Nucleic Acids Res* 44(6):2593–2612.

29. Sahara S, Yanagawa Y, O'Leary DDM, Stevens CF (2012) The fraction of cortical GABAergic neurons is constant from near the start of cortical neurogenesis to adulthood. *J Neurosci* 32(14):4755–4761.

30. Lipska BK, et al. (2006) Critical factors in gene expression in postmortem human brain: Focus on studies in schizophrenia. *Biol Psychiatry* 60(6):650–658.

31. Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27(11):1571–1572.

32. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

33. Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21):2957–2963.

34. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.

35. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

36. Hansen KD, Langmead B, Irizarry RA (2012) BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* 13(10):R83.

37. Jaffe AE, et al. (2012) Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 41(1):200–209.

38. Ritchie ME, et al. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43(7):e47.

39. Aryee MJ, et al. (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30(10):1363–1369.

40. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12(4):357–360.

41. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7):923–930.

42. Goldstein LD, et al. (2016) Prediction and Quantification of Splice Events from RNA-Seq Data. *PLoS ONE* 11(5):e0156132.

43. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–140.

44. Law CW, Chen Y, Shi W, Smyth GK (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15(2):R29.

45. Yu G, Wang L-G, Han Y, He Q-Y (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16(5):284–287.

46. Shabalin AA (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28(10):1353–1358.