

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

3 **Authors**

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

6 **Affiliations**

- 7 1. Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA  
8 2. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of  
9 Medicine, Baltimore, MD, USA  
10 3. Center for Computational Biology, Johns Hopkins University, Baltimore, MD, USA  
11 4. Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, USA  
12 5. Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore, MD, USA  
13 6. Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, USA  
14 7. Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore,  
15 MD, USA  
16 8. Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore,  
17 MD, USA  
18 \*equally contributing authors

19 **Corresponding Author:** Andrew E Jaffe, 855 N Wolfe St, Ste 300; Baltimore MD 21205.  
20 Phone: 1-443-287-6864; Email: [andrew.jaffe@libd.org](mailto:andrew.jaffe@libd.org)

21 **Abstract**

22 We have characterized the landscape of DNA methylation (DNAm) across the first two decades  
23 of human neocortical development in NeuN+ neurons using whole-genome bisulfite sequencing  
24 and compared them to non-neurons (primarily glia) and prenatal homogenate cortex. We show  
25 that DNAm changes more dramatically during the first five years of postnatal life than during the  
26 entire remaining period. We further refined global patterns of increasingly divergent neuronal  
27 CpG and CpH methylation (mCpG and mCpH) into six developmental trajectories and found that  
28 in contrast to genome-wide patterns, neighboring mCpG and mCpH levels within these regions  
29 were highly correlated. We then integrated paired RNA-seq data and identified direct regulation  
30 of hundreds of transcripts and their splicing events exclusively by mCpH levels, independently  
31 from mCpG levels, across this period. We finally explored the relationship between DNAm  
32 patterns and development of brain-related phenotypes and found enriched heritability for many  
33 phenotypes within DNAm features we identify.

## 34 **Introduction**

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

45           To better characterize the DNAm landscape across human cortical development, we  
46 performed whole-genome bisulfite sequencing (WGBS, see Methods) on homogenate tissue  
47 and on a neuron-enriched population isolated from 24 human dorsolateral prefrontal cortex  
48 (DLPFC) samples aged 0-23 years using NeuN-based fluorescence-activated nuclear sorting  
49 (FANS, **Figure S1A**). To complement these data, we sequenced eight FANS-derived NeuN-  
50 postnatal samples and 20 homogenate prenatal cortical samples, for a total of 75 samples after  
51 quality control (**Table S1**). We fully characterized the landscape of DNAm at both CpG and  
52 non-CpG (CpH) dinucleotides in these samples, allowing for a finer dissection of differential  
53 DNAm functional specificity. We also sequenced matched transcriptomes of homogenate  
54 cortical samples from these donors and a subset of three nuclear transcriptomes each from  
55 NeuN+ and NeuN- samples to assess the functional consequences of epigenomic remodeling  
56 (53 total transcriptomes, **Table S2**). By exploring DNAm patterns in neurons across prenatal and  
57 postnatal human brain development, we show that the first five years of postnatal life are a

58 critical period in epigenetic plasticity, and we identify developmental shifts in neuronal DNAm in  
59 both the CpG and CpH context. We also clarify the relationship between CpG but particularly  
60 CpH methylation (mCpG and mCpH, respectively) and gene expression and splicing in neuronal  
61 development, and explore the ramifications of these insights for neuropsychiatric disease.

## 62 **Results**

63 After data processing, quality control and filtering, we analyzed 18.7 million cytosines in  
64 the CpG context at an average coverage of 15x (see methods). Comparable to previous  
65 reports<sup>1,4,5</sup>, CpGs were overall highly methylated (71-76% CpGs with  $\beta > 80\%$ , **Table S3**).

66 While NeuN antibody labels most mature neuronal subtypes in human cortex, some  
67 neurons will not be labeled and will be captured in the NeuN- fraction amidst a diverse array of  
68 non-neuronal cell types, including oligodendrocytes, astrocytes, microglia, and epithelial cells.  
69 Gene expression differences between fractions confirmed, however, that NeuN+ and NeuN-  
70 samples are enriched for neuronal and glial-lineage cells, respectively (**Figure S1B-D**).  
71 Therefore in this work we refer to NeuN+ and NeuN- samples as “neurons” and “glia” although  
72 we acknowledge that these samples do not perfectly reflect these identities and mask more  
73 granular differences between subcellular identities contained within.

74 Developmental DNAm changes identified in homogenate cortex were strongly  
75 confounded by shifting cell type proportions (OR=7.5,  $p < 10^{-100}$ , **Figure S2A**)<sup>2</sup>. While  
76 homogenate measurements were positively correlated with developmental changes that  
77 occurred in both neuronal and glial cell types ( $\rho = 0.79$ ,  $p < 10^{-100}$ ), cell type-specific developmental  
78 changes were less consistently observed in homogenate preparations ( $\rho = -0.26$ ,  $p < 10^{-100}$ ,  
79 **Figure S2B-D**). Overall, ~40% of cell type-specific developmental DNAm changes could not be  
80 detected at all in homogenate cortex (**Figure S2E**), and many of the cell type-specific effects

81 could not be accurately identified in homogenate tissue. These results highlight the importance  
82 of measuring DNAm in the appropriate cellular context for improved resolution to detect true  
83 developmental changes.

#### 84 *DNAm as a map of putative functional genomic states*

85 Local CpG methylation (mCpG) patterns are known to distinguish genomic states of  
86 DNA and chromatin. For instance, unmethylated regions (UMRs) are associated with promoters,  
87 with a subset of longer UMRs (DNAm valleys, DMVs) that overlap developmental genes often  
88 encoding transcription factors (TFs)<sup>6,7</sup>; low-methylated regions (LMRs) often signify enhancer  
89 sequence<sup>8</sup>; and partially methylated domains (PMDs) are associated with heterochromatin and  
90 late replicating DNA<sup>9-11</sup>. To better resolve the developing regulatory landscape in postnatal  
91 neurons and glia and in bulk prenatal cortex, we assessed the temporal dynamics of these  
92 selected DNAm patterns in the CpG context. Compared to prenatal homogenate cortex and  
93 postnatal glial cells, postnatal neurons showed a general accumulation of mCpG, at a rate 50%  
94 faster than the other cells. This was evident in the LMR and to a lesser extent the UMR  
95 landscape, since fewer and smaller LMRs were identified as neuronal development progressed  
96 (**Figure S3A-B**). As expected, UMRs and LMRs were highly enriched for transcription start sites  
97 (TSSs) and enhancers in DLPFC chromatin state data from the Roadmap Epigenomics  
98 Consortium<sup>12</sup> (**Figure S3C**). Interestingly, LMRs were similarly enriched in these states in both  
99 adult and fetal brain; this correspondence may reflect a shared regulatory landscape established  
100 early in development.

101 While PMDs are a common feature of most cell types, they have not been conclusively  
102 identified in neurons. Here we identified a range of 245 to 404 PMDs per neuronal sample  
103 (**Figure S4A**). PMDs were especially enriched for heterochromatin and, interestingly, enhancers

104 in our postnatal neuronal samples (**Figure S4B**). 65.4% of PMD base pairs were also identified  
105 as PMD in an independent WGBS dataset of NeuN-sorted human neurons (**Figure S4C**).  
106 40.3-61.0% of PMD bases per neuronal sample were identified as common PMD sequence, and  
107 9.3-15.0% bases were additionally identified as PMD in at least one sample in a recent study  
108 profiling PMDs in multiple cell types and tissues<sup>13</sup> (**Figure S4D**). These data suggest that  
109 although the neuronal genome was overall highly methylated, a small but consistent portion  
110 displayed the characteristics of PMDs.

111 We further identified significant neuronal DMV changes through accumulation of mCpG  
112 that revealed regulators of cell identity and development and their temporal windows of  
113 expression change. Compared to bulk prenatal cortex, postnatal neurons and glia showed  
114 marked reduction in the size of DMVs (**Figure S5A**). Methylation shifts within DMVs led to  
115 inclusion and exclusion of transcription factor genes in an age-dependent manner, and on  
116 average transcription factor genes were higher expressed in the age group in which the gene  
117 was escaping the DMV state by accumulating DNAm (**Figure S5B-C**). These results underscore  
118 the substantial DNAm landscape alterations that neurons and glia undergo during development  
119 in defined mCpG patterns, including previously unobserved PMDs.

### 120 *Developmental shifts in neuronal mCpG highlight synaptic remodeling during first five years of* 121 *postnatal life*

122 We next quantified more localized changing mCpG levels by exploiting the correlation  
123 between neighboring mCpG levels to identify genomic regions with changing mCpG. We  
124 identified 11,179 differentially methylated regions (DMRs, FWER<5%, see Methods) in the CpG  
125 context between cell types (covering 31.1 Mb) that replicated in independent WGBS data<sup>1</sup>  
126 (98.4% concordant,  $\rho=0.925$ , **Figure S6A**). Many of these DMRs overlapped genes involved in

127 neuronal or glial-specific processes (**Figure S6B**). We found fewer DMRs for developmental  
128 mCpG changes compared with cell type differences, the majority being within rather than across  
129 cell types (2,178 versus 129 DMRs, at ~5% change in DNAm per decade of life, FWER<5%).  
130 Among the 2,178 cell type-specific developmental DMRs (cdDMRs, 3 Mb, **Table S4**), neuronal  
131 mCpG patterns seemed to diverge from an immature landscape shared by glia and prenatal  
132 cortex (**Figure 1A**), with the largest changes occurring in the first five years of life. Indeed, the  
133 magnitude of DNAm changes in neurons and glia in samples five years and younger was double  
134 that of older samples (**Figure S6C-D**). These results provide epigenetic correlates to the known  
135 developmental processes occurring in the cortex in the first five postnatal years, including prolific  
136 synaptogenesis and gliogenesis.

137 We further parsed these cdDMRs using k-means clustering to partition the cdDMRs into  
138 six groups with unique DNAm characteristics (**Figure 1B**). 71.1% of cdDMRs were in groups  
139 characterized by increasing neuronal and/or decreasing glial DNAm over postnatal development  
140 (**Figure 1B**; Groups 1, 2 & 6). A varying proportion of each cdDMR group corresponded to  
141 sequence differentially methylated by neuronal subtype from publicly available data<sup>14</sup> depending  
142 on the trajectory of neuronal methylation patterns in the group, suggesting that assorted  
143 neuronal subclasses contribute to these developmental patterns (**Figure S7A-B**). Gene  
144 ontology enrichment in the six groups suggested that these groups are associated with a  
145 continuum of biological roles, many relating to functions specific to the cell type with decreasing  
146 methylation (**Figure 1C**). For example, **Figure 1D** shows a Group 3 cdDMR within *SNAP25*, a  
147 presynaptic neuronal gene, in which neurons uniquely and progressively lost DNAm over  
148 development. This pattern suggests increased repression of a neuronal fate in maturing glia not  
149 mirrored in neurons over postnatal development in Group 3 cdDMRs. Likewise, the opposite  
150 pattern was observed in a Group 6 cdDMR within *MBP*, an oligodendrocyte gene encoding a

151 component of the myelin sheath, in which glia but not neurons progressively lost DNAm (**Figure**  
152 **1E**).

153 We lastly compared these cdDMR groups to a list of putative enhancers active in human  
154 brain development curated by evolutionary age<sup>15</sup> and found strong enrichment for these  
155 sequences across all six groups (**Figure S8A**). Human accelerated regions, or conserved  
156 sequences that have experienced rapid mutation in the human lineage<sup>16</sup>, were also enriched for  
157 dynamic DNAm remodeling (**Figure S8B**), suggesting that our CpG-based cdDMRs may be  
158 enriched for sequences related to higher cognitive functions associated with the human DLPFC.

159 Overlapping cdDMRs with the mCpG features identified above provided additional  
160 insight to the potential functional genomic states underlying these regions. For instance,  
161 cdDMRs scarcely overlapped heterochromatic PMDs; cdDMRs losing neuronal mCpG were  
162 positively correlated with increasing LMR overlap, potentially reflecting enhancer element  
163 activation during cortical maturation in these groups (Groups 3 & 5 cdDMRs; both with  $t > 3.8$ ,  
164  $\rho > 0.63$ , and  $FDR < 2.7e-03$ ). Curiously, a high proportion of cdDMRs gaining DNAm in glia but  
165 not in neurons (Group 4 cdDMRs) overlapped DMVs early in development in glia but steadily  
166 lost DMV status over time ( $t = -4.3$ ,  $\rho = -0.87$ ,  $FDR = 1.3e-02$ , **Figure S9A**). Assessing chromatin  
167 state from the homogenate Roadmap Epigenomics brain maps, in contrast, lacked the  
168 resolution to provide this nuance: all six cdDMR groups were similarly enriched for  
169 transcriptional (particularly TSS-flanking) and enhancer chromatin states, and depleted for  
170 heterochromatin and quiescent states (**Figure S9B**). These results confirm the role of dynamic  
171 DNAm in helping establish epigenomic states that guide cell lineage differentiation and  
172 emphasize the utility of creating genome-wide DNAm maps to better parse the functional  
173 diversity of cell type-specific developmental DNAm remodeling in the human cortex, a process  
174 that is particularly critical during the first five years of postnatal development.



175 *Abundant neuronal CpH methylation is highly correlated with neighboring CpG methylation*

176 Unlike in most other somatic tissues and cell types, mCpH is an abundant, conserved  
177 feature of the neuronal epigenome<sup>1,4</sup>. We therefore analyzed 58.1 million cytosines in CpH  
178 contexts (H=A,T, or C) that had evidence of methylation across the samples (coverage  $\geq 5$ , at  
179 least 5 samples with  $\beta > 0$ , see Methods). As shown previously<sup>1</sup>, mCpH sites were predominantly  
180 lowly methylated (92-99% CpHs with  $\beta < 20\%$ , **Table S3**). While mCpH was distributed  
181 throughout the genome (**Figure S10A**), it was greater in neurons than glia (98.9% of 7,682,075  
182 differentially methylated CpHs were in neurons, at FDR $< 5\%$ ) and mostly accumulated across  
183 postnatal development (99.3% of 3,194,618 CpHs, at FDR $< 5\%$ ; **Table S5**). Most mCpH  
184 accumulated primarily in either the CAG or CAC context over the first five years of postnatal  
185 life—similarly to mCpG—followed by a tapered global increase into adulthood (**Figure S10B**).

186 While the majority of mCpH in embryonic stem cells (ESCs) occurs in the CAG context,  
187 previous work has shown that ESCs undergo loss of mCAG during neuronal differentiation  
188 followed by preferential accumulation of mCAC<sup>17</sup>. Here we further refined these patterns and  
189 found a cell type-specific relationship with trinucleotide context: overall, total mCAG increased  
190 40% faster than mCAC in neurons, while in glia, mCAG accumulated 50% slower than mCAC  
191 (**Figure S10C**). Taking into account relative genome-wide proportions of CAG and CAC though,  
192 neuronal mCAG accumulated 30% slower than mCAC (**Figure 2A**). mCH that was greater in  
193 glia than neurons, or in younger than older neurons, was more likely to be in the CAG than CAC  
194 context (OR $> 4.13$ , p $< 2.2e-16$ ). Interestingly, the 3,286 and 1,744 genes that contained  
195 significantly increasing and decreasing mCAC versus mCAG over development, respectively,  
196 were associated with different biological processes related to neuronal function and activity,  
197 particularly involving the synapse (**Figure S10D**). These results reinforce that trinucleotide

198 context may be regulated by distinct mechanisms playing non-redundant biological roles in  
199 human brain development.

200 We next examined the relationship between neighboring mC levels by measuring  
201 autocorrelation, or how correlated the methylation level of a cytosine is with that of cytosines  
202 progressively further away. Unlike in the CpG context, where neighboring mCpG levels were  
203 highly correlated as previously described<sup>18</sup>, neighboring CpH DNAm levels across the genome  
204 were not autocorrelated. Within the cdDMRs, however, while mCpH levels separately remained  
205 uncorrelated, together all methylated cytosines (i.e., mCpH+mCpG) showed similar  
206 autocorrelation as mCpG levels alone (**Figure 2B**). This was especially surprising given that  
207 there were about two times as many CpHs than CpGs within these regions and that the CpG  
208 and CpH were relatively interspersed, suggesting potential functional convergence in the  
209 developmentally regulated patterns identified by mCpG in these regions. Indeed, unsupervised  
210 hierarchical clustering of CpH within the cdDMRs showed infant neuronal mCpH levels were  
211 even more similar to glia compared to older neurons than mCpG (**Figure 2C**). Examining the  
212 mean mCpH compared to mCpG within the k-means cdDMR clusters showed that the groups  
213 gaining mCpG were the most correlated with mCpH trajectories within the cdDMRs ( $\rho=0.97$ ,  
214  $t=17.6$ ,  $p=2.0e-14$ ), and that although mCpG (unlike mCpH) is present at high levels prenatally,  
215 both mCpG and mCpH accumulate at similar rates over postnatal development in these groups,  
216 once again especially in the first five years of postnatal life where the majority of the methylation  
217 change takes place ( $t=-0.091$ ,  $p=0.94$ ; **Figure S11**). These results emphasize the potential  
218 regulatory importance of cdDMRs and putative functional agreement between both contexts of  
219 DNAm in these regions.

220 *mCpG and mCpH levels influence transcript isoform use*

221 Previous studies show that both mCpG and mCpH in gene bodies but particularly in the  
222 promoter and first 2 kb of the gene are negatively associated with gene expression, and that  
223 genic mCpH is the most discriminating predictor of gene expression<sup>1,5</sup>. To anchor our DNAm  
224 patterns in transcriptional activity, we compared our WGBS data with NeuN-sorted nuclear  
225 RNA-seq data (see Methods). We took the average DNAm levels across six groups—infant  
226 (ages 0-1), child (ages 1-10), and teen (age 10+) within both cell types (neuronal and  
227 glial)—and calculated associations between DNAm and expression. Gene expression was  
228 negatively correlated with mCpG regardless of age and cell type in both promoter sequence and  
229 gene bodies ( $-0.42 < \rho < -0.22$ ,  $p \sim 0$ ; 57,332 genes,  $p < 10^{-100}$ ; **Figure S12A**). Interestingly, mCpG  
230 in exons was significantly but weakly positively correlated with exon expression in infancy,  
231 particularly in glial samples ( $\rho = 0.094$ ,  $p < 10^{-100}$ ), which may relate to the previously-identified  
232 positive relationship between mCpG and expression and higher methylation in exons than  
233 introns<sup>9</sup>.

234 Across promoters, gene bodies, and exons, only neurons showed a negative correlation  
235 between gene expression and mCpH, a relationship that became stronger over development  
236 (**Figure S12B**). This pattern was consistent with the preferential accumulation of mCpH in  
237 neurons as the brain matures. mCAC and mCAG showed similar patterns of increasingly strong  
238 negative correlation preferentially in neurons between methylation and expression across these  
239 features (**Figure S12C-D**). Both mCpG and mCpH surrounding exon-exon splice junctions were  
240 weakly negatively correlated with expression of the junction in neurons (**Figure S12E**).

241 Because mCpG has previously been associated with alternative splicing<sup>19</sup> and mCpH is  
242 15-20% greater in exons than in introns<sup>9</sup>, we hypothesized that accumulating mCpH may  
243 contribute to the diversity of alternative splicing characteristic of the brain particularly during  
244 development. Leveraging our single-base resolution data, we were able to identify genome-wide

245 functional correlates of mCpH, independent of nearby mCpG, by associating DNAm with nearby  
246 expression in the same cortical samples. Specifically, we tested whether methylation levels  
247 directly associated with gene or exon expression levels as well as the “percent spliced in” (PSI)  
248 of alternative splicing events using the 22 neuronal samples with matching homogenate polyA+  
249 RNA-seq data (see Methods). We found 40,940 CpG and 40,303 CpH associations that explain  
250 changes in these three expression summarizations at FDR<5% with a genome-wide  $p < 5 \times 10^{-4}$ .  
251 We further identified 220,622 marginal ( $p < 0.01$ ) CpG associations with expression within 1 kb  
252 around the associated CpH. While an independent association of mCpH at the gene and PSI  
253 summarizations was rare, there were substantially more exons exclusively regulated by local  
254 mCpH, largely in the CHH context, in developing postnatal neurons (**Figure 3A**). Three  
255 examples of methylation-associated isoform changes are shown in **Figure 3B**.

256       Regardless of context specificity, these expression-associated cytosines were depleted  
257 in gene promoters and instead enriched in gene bodies and flanking regions (**Table 1**, see  
258 Methods). Both contexts were enriched for the high-GC 3' and 5' canonical splice site  
259 sequences (FDR<1.1e-04), although the associated cytosine could be either inside or outside  
260 the corresponding expression feature. Only 3.5-13.7% of expression-associated cytosines  
261 overlapped DMR sequence after stratifying by expressed feature and dinucleotide context,  
262 indicating that these associations may arise from a more individualized mC effect than the  
263 DMRs.

264       Although the majority of these DNAm-expression associations were independent of  
265 development despite being identified in developing neurons, the mCpH changes at these sites  
266 were independently associated with age and expression. The genes including PSI events  
267 regulated by local DNAm levels in both CpG and CpH contexts were consistently enriched for  
268 neuronal components (**Figure S13**), while genes containing methylation-associated alternative

269 exons were enriched for synaptic signaling and neurotransmitter transport (**Figure 3C** and  
270 **Figure S14**), suggesting that we are detecting true neuronal mC-expression associations  
271 despite measuring splicing in homogenate RNA-seq. Many of these genes were also  
272 differentially expressed between neuronal and glial nuclear RNA (FDR<0.05, **Table 1**). Most, but  
273 not all, expression-associated cytosines at the gene- and exon-level showed significant  
274 decreases in expression as methylation levels increased.

275         The associations between these putatively regulatory cytosines and nearby expression  
276 levels can be explored in a webtool (<https://jhubiostatistics.shinyapps.io/wgbsExprs/>). Results  
277 can be interactively summarized such as in **Table 1** for user-selected subsets and visualized as  
278 in **Figure 3B** or via the UCSC genome browser (**Figure S15**). By integrating neuronal mCpG  
279 and mCpH levels with accompanying RNA-seq data in the same brains, we have identified, for  
280 the first time, direct regulation of hundreds of transcripts and their splicing events exclusively by  
281 mCpH, independent of mCpG levels, across the first two decades of human cortical  
282 development.

### 283 *DNAm patterns shed light on the active cell type and timing of neuropsychiatric phenotype* 284 *development*

285         Previous work has attributed a high proportion of neuropsychiatric trait heritability to  
286 neuron-specific DNA methylation patterns<sup>20</sup>. Given the role of dynamic DNAm in marking DNA  
287 sequence function over development, we examined the relationship between our methylation  
288 features with heritability for 30 human behavioral-cognitive traits, psychiatric and neurological  
289 disorders, and non-brain-related traits<sup>21</sup> (**Table S6**), hypothesizing that DNAm patterns may  
290 illuminate not only the active cell type but potential critical timeframes for genomic activity in  
291 these complex phenotypes. We used stratified linkage disequilibrium score regression (LDSC)<sup>22</sup>

292 to estimate the proportion of heritability measured in GWAS summary statistics for each  
293 phenotype that could be attributed to each of 16 genomic features, including 10 sets of DMRs,  
294 LMRs identified in the prenatal, glial or neuronal methylome, human brain regulatory sequence  
295 annotated by chromHMM or the LDSC package, or non-differential CpG clusters (**Figure 4A**). In  
296 agreement with previous findings<sup>20,23</sup>, human brain annotated regulatory sequence was broadly  
297 enriched for heritability of brain-specific traits (14 of 26 brain-associated phenotypes enriched in  
298 chromHMM or CNS (LDSC) regions at  $FDR \leq 0.05$ ), as were neuronal features (10 of 26 brain  
299 phenotypes enriched in neuronal hypomethylated regions,  $FDR \leq 0.05$ ; **Table S7**). Significantly  
300 differentially hypomethylated neuronal regions had on average 1.85 times higher enrichment  
301 scores than non-differential neuronal LMRs, meaning the DMRs explained 1.85x more  
302 heritability over regions containing a similar number of SNPs than the LMRs. Interestingly, body  
303 mass index (BMI) heritability was enriched in general brain regulatory sequence and  
304 hypomethylated neuronal DMRs ( $FDR \leq 0.05$ ), consistent with previous evidence linking this  
305 metabolic phenotype to regulatory sequence active in cells of the human central nervous  
306 system<sup>22</sup>.

307 In terms of developmental DNAm patterns, heritability of BMI, IQ, neuroticism, and major  
308 depressive disorder was enriched in both postnatal neuronal and prenatal LMRs, suggesting  
309 early action of genetic influence on the development of these phenotypes ( $FDR \leq 0.05$ ). Few  
310 developmental differential groups captured a significant proportion of heritability for the 30 traits  
311 tested, perhaps because of their small size compared to the cell type-specific or non-differential  
312 groups (cdDMRs covered ~31, 240 and 838 times less sequence and included ~32, 273 and  
313 904 times fewer SNPs than cell type DMRs, LMRs or general brain features, respectively).  
314 However, despite covering only 607 kilobases, group 3 cdDMRs (ie, static glial and decreasing  
315 neuronal DNAm) were significantly enriched for heritability of schizophrenia (coefficient

316 z-score=2.74, FDR=0.039). Group 4 cdDMRs (111.7 kilobases; increasing glial, static neuronal  
317 DNAm) were also enriched for heritability of PTSD (coefficient z-score=3.21, FDR=0.01).

318         Given the enrichment for psychiatric disease heritability measured in common SNPs in  
319 these cdDMR groups, we then expanded our analysis to include seven curated gene sets  
320 containing *de novo* and rare inherited variation--including rare copy number variants (CNVs) and  
321 syndromic variants--associated with psychiatric, neurodevelopmental and neurodegenerative  
322 disorders<sup>24,25</sup>. We again found significant enrichment of hypomethylated neuronal DMRs in  
323 genes implicated in psychiatric and neurodevelopmental disorders (i.e., schizophrenia, autism  
324 spectrum disorder (ASD), syndromal neurodevelopmental disorders, and intellectual disability;  
325 all with OR>2.04 and FDR<1.9e-02; [Table S8](#)). In this analysis, we also found enrichment for  
326 hypermethylated neuronal DMRs in ASD genes from the SFARI Gene database and  
327 schizophrenia genes containing *de novo* mutations (both with OR>1.92 and FDR<5.0e-03).  
328 These results confirmed a prominent role of neuronal functioning in most of the  
329 neurodevelopmental disorders using an orthogonal measurement approach as done  
330 previously<sup>23</sup>. Over postnatal development, Group 5 cdDMRs (increasing glial, decreasing  
331 neuronal DNAm) were enriched in ASD genes from the SFARI Gene database (OR=5.7,  
332 FDR=4.1e-03), while Group 6 cdDMRs (decreasing glial, static neuronal DNAm) were enriched  
333 in ASD, syndromal neurodevelopmental disorder, and intellectual disability genes (all with  
334 OR>3.1 and FDR<1.9e-02). In contrast, a curated set of neurodegenerative disorder genes  
335 showed no enrichment for cdDMRs, perhaps reflecting lesser relevance of the first two decades  
336 of postnatal epigenomic remodeling to the etiology of those disorders.

337         In the non-CpG context, we found significant enrichment of both increasing and  
338 decreasing mCpH levels in genes associated with schizophrenia, ASD, and syndromal  
339 neurodevelopmental disorders (all with OR>2.1 and FDR<2.0e-02). CpH hypomethylation in

340 neurons was also enriched in the neurodegenerative disease gene set (OR=2.6, FDR=3.7e-03).  
341 Finally, significantly increasing mCpH was depleted in genes associated with intellectual  
342 disability (OR=0.34, FDR=3.7e-06). While enrichment for conflicting mCpH patterns are at first  
343 curious given the overall negative association between mCpH and gene expression, outside of  
344 the context of DMRs, individual mCpH could be associated both positively and negatively with  
345 expression. Indeed, many genes, exons, and PSI events whose expression both positively and  
346 negatively associated with both mCpH and mCpG were also enriched in genes associated with  
347 schizophrenia, ASD, and syndromal neurodevelopmental disorders (all with OR>2.1 and  
348 FDR<2.5e-02; [Table S9](#)). Overall these results suggest that these examples of dynamic  
349 methylation and associated isoform switching may play a role in the development of higher  
350 cognitive functions during brain maturation associated with these diseases.

## 351 **Discussion**

352 Here we have created a single-base resolution map of the dynamic DNAm landscape  
353 across the first two decades of postnatal human brain development in two cell type-enriched  
354 populations. Using FANS-derived samples, we were able to identify 40% more  
355 developmentally-regulated regions of changing DNAm than were identified in homogenate  
356 DNAm cortical data. We profiled specific features of the DNAm landscape including LMRs,  
357 UMRs, PMDs and DMVs and found that across features, neurons were typified by a general  
358 accumulation of mCpG. In the absence of complementary cell type-specific chromatin data,  
359 characterizing known DNAm features provided a more granular view of the potential functional  
360 genomic state in these regions than the available predictions derived from a few homogenate  
361 cortical samples. Particularly in studies using human postmortem brain, where tissue is often  
362 subjected to long postmortem intervals and low pH that degrades less stable epigenetic



363 signatures, DNAm is a robust and durable marker that can be used to map the functional  
364 genomic terrain. These DNAm maps complement recently available epigenomic maps of  
365 different modalities generated on FANS-derived samples in the psychENCODE Consortium<sup>27</sup>.

366 We further parsed the general accumulation of neuronal DNAm into six trajectories of  
367 cell type-specific developmental patterns and found that neuronal mCpG progressively diverged  
368 from a shared landscape with glia and bulk prenatal cortex as the brain matured. Importantly,  
369 these diverging patterns were most striking during infancy through the first five years of  
370 postnatal life. The human brain experiences an explosion of synaptic connections during this  
371 time period, to nearly double the number found in the mature adult brain<sup>28</sup>. Although previous  
372 work has underscored this timeframe in terms of rapid DNAm accumulation<sup>1</sup>, this is the first  
373 work to refine DNAm patterns to reflect cell type-specific gain and loss of mCpG and mCpH  
374 within this critical window. By parsing these neuronal and glial DNAm patterns, we have  
375 highlighted epigenetically dynamic regions that may be contributing to the developmental  
376 processes such as synaptogenesis occurring during this timeframe that establish the foundation  
377 for fine-tuning connections throughout the remainder of brain maturation. These results provide  
378 a finely resolved depiction of epigenetic plasticity being greatest during this period of life, and  
379 support other evidence that environmental experience during these years may have an  
380 especially enduring impact on brain function<sup>29</sup>.

381 mCpH is unusually abundant in neurons compared to other cell types and appears to  
382 undergo trinucleotide-specific reprogramming during differentiation from ESCs<sup>17</sup>. While most  
383 mCpH in ESCs occurs in the CAG context, neuronal mCpH predominantly accumulates in the  
384 CAC context<sup>17</sup>. Here we elaborate on this relationship, showing that while both mCAG and  
385 mCAC aggregate in neurons as they mature and mCAG is gained faster than mCAC overall,  
386 mCAC accumulates proportionally faster in both neurons and glia over time. Interestingly,

387 although neurons and glia contained mCpH in both trinucleotide contexts, mCAG was more  
388 likely to have higher levels in glia than neurons or be decreasing over development; indeed,  
389 genes containing decreasing mCAG but not mCAC were strongly associated with neuronal  
390 biological processes. mCpH trinucleotide context, therefore, may have as yet not well  
391 understood ramifications in brain development.

392 In terms of the relationship between mCpH and mCpG, we found that while neighboring  
393 mC (i.e., mCpG+mCpH) was not correlated genome-wide, mC was highly correlated within the  
394 cdDMRs despite local mCpH not being correlated. In other words, there was a convergence of  
395 levels of all contexts of methylation within the cdDMRs that was not detected genome-wide.  
396 mCpH also recapitulated the pattern seen in mCpG of diverging from a shared DNAm  
397 landscape with glia. Given that mCpH and mCpG have previously been shown to work in  
398 concert to recruit MECP2 binding to fine-tune gene expression<sup>30</sup>, it is sensible that levels of both  
399 contexts would perhaps reflect a shared functional role within putatively regulatory cdDMR  
400 sequence, since cdDMRs were also enriched for gene bodies and brain enhancer sequence.  
401 This work quantifies this correlation for the first time, a DNAm relationship unique to only a  
402 selection of cell types that includes neurons.

403 The identification of widespread association of mCpG and mCpH with expression and  
404 specific splicing events, particularly in neuronal genes enriched for neuropsychiatric diseases,  
405 highlights a potential novel role of mCpH and further expands the role of mCpG in the regulation  
406 of gene expression in neurons. Splicing is predominantly a co-transcriptional process influenced  
407 by changes in chromatin modifications and RNA binding proteins; the effects of DNAm on  
408 splicing decisions is not yet well studied<sup>31</sup>. Here we found thousands of associations between  
409 mC levels and gene, exon and PSI expression in developing postnatal neurons, particularly  
410 featuring many exons that are exclusively associated with mCpH. Although it is not possible to

411 establish a causal role for mC in these data, these analyses, which are summarized in the  
412 provided website, can empower other researchers to explore the connection between DNAm  
413 and alternative isoform use, a phenomenon particularly prevalent in the developing brain that is  
414 often associated with disease<sup>32</sup>.

415 We also explored the relationship between DNAm patterns and genetic associations with  
416 various phenotypes and found both expected and surprising associations. We confirmed  
417 enrichment for heritability of brain traits generally in neurons, and heritability for schizophrenia, a  
418 disorder with strong neurodevelopmental underpinnings, specifically in genomic regions losing  
419 DNAm preferentially in neurons over early postnatal development (ie, Group 3 cdDMRs). This  
420 result emphasizes the critical nature of neuronal development and maturation in early  
421 establishment of pathological connectivity and function for this adult-onset disorder, as most  
422 DNAm loss--generally associated with increased activity of a gene or regulatory  
423 element--occurred within the first five postnatal years. Indeed, Group 3 cdDMRs were present in  
424 genes such as *GRIN1*, *SYN1* and *CAMK2A*, and others involved in establishing synapse  
425 organization and function, a hallmark of early postnatal brain development, implicating abnormal  
426 genetic regulation of neuronal connectivity in schizophrenia development. Interestingly,  
427 heritability for PTSD was significantly enriched in regions preferentially gaining DNAm in the  
428 non-neurons over development (Group 4 cdDMRs), regions associated with neural precursor  
429 cell proliferation and cell-cell connective properties (Figure 1C). This association, particularly  
430 given the small amount of sequence covered by Group 4 cdDMRs, could lead to fruitful insights  
431 into susceptibility for PTSD and warrants further study.

432 We also found enrichment of genes associated with rare variants implicated in  
433 psychiatric and neurodevelopmental disease in a variety of cell types and developmental  
434 trajectories, highlighting the the genomic boundaries, developmental timing and cellular context

435 of epigenomic remodelling of regulatory elements or expressed features associated with known  
436 risk genes. Two examples of this are *HDAC4* and *CACNA1B*, genes associated with ASD in the  
437 SFARI Gene database. We identified a 566 bp Group 3 (decreasing neuronal, static glial DNAm)  
438 cdDMR within an intron of *HDAC4*, a calcium-sensitive transcriptional repressor, and a 3.6 kb  
439 Group 5 (decreasing neuronal, increasing glial DNAm) cdDMR within *CACNA1B*, a gene  
440 encoding a voltage-gated calcium channel subunit (**Figure 4B-C**). Even though both genes are  
441 implicated in ASD and both cdDMRs are hypomethylated in neurons, the timing of loss of mCpG  
442 suggests that *CACNA1B* activity occurs earlier in postnatal development than *HDAC4*. Given  
443 that ASD onset is typically in early childhood, these risk genes may therefore have differing  
444 implications in the etiology of ASD. Another example is a 646 bp Group 6 (decreasing glial,  
445 static neuronal DNAm) cdDMR that overlaps the last intron and exon of *AKT3*, a  
446 serine/threonine-protein kinase gene. Although the *AKT3/1q44* locus has been associated with  
447 schizophrenia risk, the mechanisms are not yet known given that *AKT3* is involved in many  
448 biological functions<sup>25</sup>. Interestingly, this cdDMR selectively lost mCpG in glia beginning in  
449 infancy, suggesting that *AKT3* activity in human DLPFC may be localized to glia beginning in  
450 infancy or earlier (**Figure 4D**). This work provides the first *ex vivo* look at the DNAm dynamics  
451 within human neurons and glia and thus allows for the first examination of those parameters  
452 within the relevant organ, the brain.

453         Despite these insights, our data invoke several caveats. While NeuN based FANS  
454 greatly improves identifying developmental DNAm changes over homogenate data, designating  
455 NeuN+ and NeuN- samples as “neurons” and “glia” is not completely accurate in that NeuN-  
456 samples will include signal from unlabeled neurons and mask non-neuronal diversity. However,  
457 recent work assessing brain regional DNAm differences between NeuN+ and NeuN- found that  
458 NeuN- contributed comparatively marginal variability in DNAm compared to NeuN+, suggesting

459 that neuronal methylomes are much more dynamic than non-neurons<sup>20</sup>. Likewise, while a  
460 percentage of the bases in the DMRs identified in this work have also been previously shown to  
461 be differentially methylated by neuronal subtypes whose unique methylomes are masked using  
462 NeuN-based FANS<sup>5,14,33</sup>, the proportion of these subtypes should be stable over postnatal  
463 development<sup>34</sup>. Future epigenomics studies however can improve on the resolution of our study  
464 by isolating more specific neuronal subpopulations to refine the cellular specificity of these  
465 neuronal methylation changes largely occurring in the first few years of life.

466 Another caveat is that WGBS does not allow for the discrimination between mC and  
467 hydroxymethyl-cytosines (hmC), an intermediary in the demethylation pathway. Previous work  
468 has shown<sup>5</sup> that only a fraction of CpGs have measurable levels of hmC, suggesting that our  
469 results are not confounded. In the cited study<sup>5</sup>, hmCpG signal from homogenate cortex  
470 represented 10% of the hypermethylation found in excitatory neuron mCpG, suggesting that  
471 most of the mCpG signal in our neuronal data likely is true mCpG. The level of hmC in the  
472 non-CpG context remains controversial, with some studies not identifying hmCpH<sup>5</sup> and others  
473 detecting low amounts (1% in FANS-derived human glutamatergic and 0.47% in GABAergic  
474 neurons)<sup>35</sup>. Further, FANS-derived human oligodendrocytes showed little hmC in the same  
475 study<sup>35</sup>. Overlap of our cdDMRs with DMRs between neurons and oligodendrocytes in a study of  
476 mC in FANS-derived human PFC samples<sup>32</sup> showed that cell type differences primarily reflected  
477 true mC rather than hmC contamination in WGBS, while hmC DMRs between neuronal  
478 subtypes primarily were reflected in hypomethylated neuronal cdDMR groups and not the  
479 hypermethylated neuronal groups that would potentially include hmC signal contamination  
480 **(Figure S7B)**. Future work, however, should more closely examine the contribution of hmC to  
481 the global hypermethylation seen during neurodevelopment.

482 By mapping the changing DNAm landscape over human postnatal neuronal and glial  
483 development, we have identified unique trajectories of DNAm change particularly dynamic  
484 during the first five years of life that show convergence between mCpG and mCpH, as well as  
485 associations between single mCpG and mCpH and alternative splicing. These patterns may  
486 also help illuminate the mechanisms through which psychological, neurological and psychiatric  
487 traits are developed by placing known genetic contribution in an epigenomic context.

## 488 **Acknowledgements**

### 489 *Author Contributions*

- 490 ● A.J.P.: Conceptualization, Formal Analysis, Investigation, Visualization, Writing – Original  
491 Draft Preparation, Writing – Review & Editing.
- 492 ● L.C.-T.: Conceptualization, Formal Analysis, Visualization, Software, Writing – Original  
493 Draft Preparation, Writing – Review & Editing.
- 494 ● N.A.I.: Software, Data Curation.
- 495 ● W.X.: Investigation.
- 496 ● E.E.B.: Formal Analysis, Writing – Review & Editing.
- 497 ● J.H.S.: Supervision.
- 498 ● R.T.: Investigation.
- 499 ● L.M.: Investigation.
- 500 ● Y.J.: Investigation, Supervision.
- 501 ● T.M.H.: Data Curation, Resources, Writing – Review & Editing.
- 502 ● J.E.K.: Data Curation, Resources, Writing – Review & Editing.
- 503 ● D.R.W.: Conceptualization, Funding Acquisition, Supervision, Writing – Review &  
504 Editing.

- 505       • A.E.J.: Conceptualization, Formal Analysis, Funding Acquisition, Supervision, Writing –  
506       Original Draft Preparation, Writing – Review & Editing.

507    *Funding*

508    This project was supported by The Lieber Institute for Brain Development and by NIH grants  
509    R21MH102791, R21MH105853, and R01MH112751. Data were generated as part of the  
510    PsychENCODE Consortium, supported by: U01MH103392, U01MH103365, U01MH103346,  
511    U01MH103340, U01MH103339, R21MH109956, R21MH105881, R21MH105853,  
512    R21MH103877, R21MH102791, R01MH111721, R01MH110928, R01MH110927,  
513    R01MH110926, R01MH110921, R01MH110920, R01MH110905, R01MH109715,  
514    R01MH109677, R01MH105898, R01MH105898, R01MH094714, P50MH106934 awarded to:  
515    Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke  
516    University), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham  
517    (University of Southern California), Mark Gerstein (Yale University), Daniel Geschwind  
518    (University of California, Los Angeles), Fernando Goes (Johns Hopkins University), Thomas M.  
519    Hyde (Lieber Institute for Brain Development), Andrew E. Jaffe (Lieber Institute for Brain  
520    Development), James A. Knowles (University of Southern California), Chunyu Liu (SUNY  
521    Upstate Medical University), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Panos  
522    Roussos (Icahn School of Medicine at Mount Sinai), Stephan Sanders (University of California,  
523    San Francisco), Nenad Sestan (Yale University), Pamela Sklar (Icahn School of Medicine at  
524    Mount Sinai), Matthew State (University of California, San Francisco), Patrick Sullivan  
525    (University of North Carolina), Flora Vaccarino (Yale University), Daniel R. Weinberger (Lieber  
526    Institute for Brain Development), Sherman Weissman (Yale University), Kevin White (University

527 of Chicago), Jeremy Willsey (University of California, San Francisco), and Peter Zandi (Johns  
528 Hopkins University).

### 529 *Competing Interest*

530 The funders had no role in study design, data collection and analysis, decision to publish, or  
531 preparation of the manuscript. Conflict of Interest: none declared.

### 532 *Data availability*

533 Raw and processed sequence data that support the findings of this study have been deposited  
534 at [www.Synapse.org](http://www.Synapse.org) with the accession code: syn5842535.

### 535 *Code availability*

536 Code is available through GitHub at: <https://github.com/LieberInstitute/brain-epigenomics>.

## 537 **Methods**

### 538 *Postmortem brain tissue acquisition and processing*

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



546 NO1-HD-4-3383. Postmortem human brain tissue was also provided by donation with informed  
547 consent of next of kin from the Office of the Chief Medical Examiner for the State of Maryland  
548 (under Protocol No. 12-24 from the State of Maryland Department of Health and Mental  
549 Hygiene) and from the Office of the Medical Examiner, Department of Pathology, Homer Stryker,  
550 M.D. School of Medicine (under Protocol No. 20111080 from the Western Institute Review  
551 Board). The Institutional Review Board of the University of Maryland at Baltimore and the State  
552 of Maryland approved the protocol, and the tissue was donated to the Lieber Institute for Brain  
553 Development under the terms of a Material Transfer Agreement. Clinical characterization,  
554 diagnoses, and macro- and microscopic neuropathological examinations were performed on all  
555 samples using a standardized paradigm, and subjects with evidence of macro- or microscopic  
556 neuropathology were excluded, as were all subjects with any psychiatric diagnoses. Details of  
557 tissue acquisition, handling, processing, dissection, clinical characterization, diagnoses,  
558 neuropathological examinations, and quality control measures have also been further described  
559 previously<sup>36</sup>. Homogenate postmortem tissue of the prefrontal cortex (dorsolateral prefrontal  
560 cortex, DLPFC, BA46/9) was obtained from all subjects.

#### 561 *Fluorescence activated nuclei sorting*

562 Nuclei were isolated from 100-300mg of pulverized DLPFC tissue using dounce homogenization  
563 followed by ultracentrifugation over a sucrose density gradient. Homogenization was performed  
564 on ice in 5mL lysis buffer [0.32M sucrose, 3mM magnesium acetate, 5mM calcium chloride,  
565 5mM EDTA (pH 8.0), 10mM Tris-HCl (pH 8.0), 0.1% Triton X-100] and the resulting homogenate  
566 was layered over 38mL sucrose buffer [1.8M sucrose, 3mM magnesium acetate, 10mM Tris-HCl  
567 (pH 8.0)] and centrifuged at 139,800 x g for 2 hours at 4°C. Cellular debris and lysis and sucrose  
568 buffers were removed and the pelleted nuclei were resuspended in 500uL PBS. Nuclei were

569 then labeled in a solution of anti-NeuN antibody conjugated to Alexa Fluor 488 (A60, Millipore,  
570 1/1000) and 0.1% BSA, rocking for 30 minutes at 4°C, followed by the addition of DAPI. Nuclei  
571 sorting was performed at the Johns Hopkins School of Public Health Flow Cytometry Core with  
572 a MoFlo Legacy (Beckman Coulter) using Summit (version 4.3) software. The purity of sorted  
573 populations was determined to be >99% based on resorting NeuN+ and NeuN- populations  
574 through the same gates.

575         The identity of the NeuN+ and NeuN- populations as neuron-enriched and glia-enriched  
576 was confirmed by sequencing nuclear RNA from each population and determining that neuronal  
577 and glial biological processes were enriched in genes differentially expressed between the two  
578 groups (FDR<0.05; **Figure S1B**). Likewise, cell type marker gene expression patterns also  
579 corroborated the neuronal- and glial-enriched identities of NeuN+ and NeuN- samples (**Figure**  
580 **S1C**). The estimated proportion of neurons in homogenate DNA methylation data based on  
581 deconvolution using differentially methylated sites between NeuN+ and NeuN- samples was  
582 highly correlated with the empirical proportion of neurons (**Figure S1D**). Raw sorting data is  
583 shown in **Figure S16**.

#### 584 *Nucleic acid extraction and RNA-seq library preparations*

585 RNA was extracted from homogenate and sorted samples using TRIzol LS Reagent (Thermo  
586 Fisher Scientific) followed by the RNeasy MinElute Cleanup Kit (Qiagen). Genomic DNA  
587 extraction was performed using the DNeasy Blood and Tissue Kit (Qiagen). Bisulfite conversion  
588 of 600 ng genomic DNA was performed with the EZ DNA methylation kit (Zymo Research). RNA  
589 sequencing libraries were made with the TruSeq RNA Library Prep Kit (Illumina) and the  
590 RiboGone Low-Input Ribosomal RNA Removal Kit (Clontech). Library concentrations were

591 measured using a Qubit 2.0 and library fragment sizes were measured using a Caliper Life  
592 Sciences LabChip GX. One hundred base-pair paired-end sequencing was run on an Illumina  
593 HiSeq 2000.

#### 594 *Whole genome bisulfite library preparations and sequencing*

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

#### 601 *Data processing/alignment*

602 We sought to align the paired-end reads for each sample to the in-silico bisulfite-treated hg19  
603 genome, which we created using the Bismark v0.15.0<sup>37</sup> `bismark_genome_preparation`  
604 program. For each library of paired-end reads (one per sample), the following processing was  
605 performed (**Figure S17**):

606 · FastQC v0.11.4, to assess read quality, presence of adapter sequence, and  
607 overrepresented sequences.

608 · Trimmomatic v0.35<sup>38</sup> to trim low quality and adapter-containing portions of the reads, with  
609 the following parameters: `PE -threads 12 -phred33`  
610 `ILLUMINACLIP:/Trimmomatic-0.35/adapters/TruSeq3-PE.fa:2:30:10:1`

611 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:75. This resulted in three  
612 sub-libraries of reads per sample: one paired-end sub-library, and then two single-end  
613 sub-libraries where the other corresponding paired read was trimmed to a length below the  
614 defined threshold.

615 · FastQC v0.11.4, on each of the three sub-libraries, to assess the improvement in read  
616 quality and adapter content following trimming.

617 · FLASH v1.2.11<sup>39</sup>, to merge the paired-end sub-library reads into longer single-end reads,  
618 as reads that overlapped around CpGs and CpHs might bias or at the least double-count  
619 the DNAm estimates. Furthermore Bismark<sup>37</sup> could only be run on single- or paired-end  
620 reads and not a combination of both. This therefore further split the paired-end sub-library  
621 into three sub-libraries: the subset of paired-end reads that were merged into longer  
622 single-end reads, and then left and right single-end reads that could not be merged.

623 · Bismark v0.15.0<sup>37</sup>, to align each of the five now-single-end sub-libraries (left-trimmed,  
624 right-trimmed, FLASH-merged, FLASH-left-unmerged, and FLASH-right-unmerged) to the  
625 bisulfite-converted hg19 genome using bowtie2<sup>40</sup> and the --non-directional argument.

626 · Resulting alignment (BAM) files across five sub-libraries were merged, sorted, and  
627 indexed using samtools v1.3<sup>41</sup> to produce one large/merged BAM file per sample.

628 · Alignments with evidence of duplication were removed using the MarkDuplicates program  
629 in Picard tools v1.141, which systematically appeared to be localized to low complexity  
630 DNA sequence near centromeres.

631 · The Bismark<sup>37</sup> bismark\_methylation\_extractor program was run on each  
632 post-duplicate-removed BAM file per sample to extract CpG and CpH DNAm levels.

633 We additionally aligned reads from each sample to the PhiX and Lambda ( $\lambda$ ) genomes to  
634 compute quality control metrics related to sequencing and bisulfite conversion quality. The  
635 average percentage of reads mapping back to the  $\lambda$  genome was 1.32% and the average  
636 bisulfite conversion efficiency was 98.64%. The average bisulfite conversion efficiency was not  
637 associated ( $p>0.05$ ) with cell type, age, cell type (adjusting for age), age (adjusting for cell type),  
638 and the interaction of them in the NeuN- (glia) and NeuN+ (neuron) samples as well as for age  
639 in the homogenate samples. The genome coverage decreased from an initial average of 43x to  
640 10x across the processing stages as shown in **Figure S18 (coverage)**. For each of the  
641 processing stages, there was no significant difference between cell types (adjusting for age),  
642 age (adjusting for cell type) as well as the interaction between age and cell type ( $p$   
643 Bonferroni $>0.05$ ). The genome coverage was extracted from the `FASTQC` reports and by using  
644 `bamcount` (<https://github.com/BenLangmead/bamcount>) v0.2.6.

#### 645 *Identifying methylation features*

646 We identified PMDs, UMRs and LMRs using the bioconductor package `MethylSeekR` (version  
647 1.20.0). We obtained coverage information from the cleaned set of ~18 million CpGs by  
648 extracting coverage and methylation using the `getCoverage()` function from the `bsseq`  
649 bioconductor package<sup>42</sup> v1.10.0. PMDs were called using `segmentPMDs()` and were visually  
650 inspected using `plotPMDSegmentation()`. To create a more stringent cutoff for PMDs, we  
651 filtered PMDs to those longer than 100 Kbp. We calculated the FDRs using  
652 `calculateFDRs()`, while masking the >100 Kbp PMDs, setting the `m` parameter to 0.5 and  
653 the FDR cutoff to 10. PMDs were further filtered to exclude overlaps with the UCSC "gap"  
654 database table from hg19 except for the gaps labeled heterochromatin. We calculated UMRs

655 and LMRs using `segmentUMRsLMRs()`. DMVs were defined as UMRs in which `pmeth` was less  
656 than or equal to 0.15 and the width was greater than or equal to 5 Kbp.

### 657 *Identifying CpG differentially methylated regions (DMRs)*

658 Using the `bsseq`<sup>42</sup> Bioconductor package v1.10.0 we loaded the Bismark<sup>37</sup> report files and  
659 filtered the CpG data to keep only the bases where all samples had a minimum coverage of 3  
660 (18,664,892 number passed the filter). We smoothed the methylation values of the remaining  
661 CpGs using the `BSmooth()` function from `bsseq` with the `parallelBy="sample"` option. To  
662 identify the age and cell type DMRs we used a model that adjusted for both covariates while for  
663 the interaction model we included an additional interaction covariate. We identified the DMRs  
664 using the `bumphunter`<sup>43</sup> Bioconductor package v1.14.0 using the `maxGap=1000`, `B=250`,  
665 `nullMethod="permutation"`, `smooth=FALSE` options, which tends to be conservative in  
666 DMR identification<sup>43</sup>. For the `cutoff` option we used 0.1 for the cell type DMRs adjusting for  
667 age, 0.005 for the age DMRs adjusting for cell type, and 0.009 for the age and cell type  
668 interaction DMRs. These first two parameter cutoffs correspond to 10% minimum DNAm  
669 differences between neurons and glia and 5% change in DNAm per decade of life across cell  
670 types, and were chosen based on functionally-relevant change in DNAm. The cutoff for the  
671 interaction model (cdDMRs) was based on selecting the equivalent percentile of change from  
672 the overall age model (86th percentile) - this percentile-based cutoff was in line with  
673 recommendations for selecting cutoffs for statistical models with less clear biological  
674 interpretations<sup>43</sup>. We used a family wise error rate (FWER) threshold of 5% to determine the  
675 DMRs: `fwcr` output from the `bumphunter()` function. A small subset of DMRs involved a  
676 single CpG, which arises from having a more significant area (length times effect size) than any

677 DMRs identified in null permuted data. All DMRs showed less than 10% median percent  
678 absolute bias to technical and biological covariates as shown in **Figure S19 (sensitivity)**.

### 679 *cdDMR processing*

680 For the “interaction” DMRs with FWER<5% (cdDMRs) we extracted the methylation values from  
681 the glial and neuronal samples using bsseq<sup>42</sup> v1.13.9 and then computed a mean methylation  
682 value per DMR for each cell type. We also calculated the mean interaction coefficient for each  
683 DMR across all the cytosines in the DMR by cell type. Using the mean coefficients by cell type  
684 we clustered the interaction DMRs using the `kmeans()` function with `centers=6`,  
685 `nstart=100` options. We chose `centers=6` based on biological interpretability of the results  
686 and because `k=6` results in an optimal AIC for clusters computed with mean centered and  
687 scaled data. For each cytosine in the cdDMRs we calculated the t-statistic and coefficient for  
688 age explaining differences in methylation adjusted for cell type:  $\sim \text{age} + \text{cell type}$ . For  
689 each DMR we computed the mean age coefficient by cell type and then calculated the median  
690 absolute coefficient across all cdDMRs. The neuronal / glia ratio is 1.5 for such median absolute  
691 age effects across the cdDMRs.

### 692 *Comparing homogenate vs. cell type-specific WGBS*

693 We first filtered CpGs to those with coverage in all 55 postnatal samples. For homogenate  
694 samples, we used `lmFit()` and `ebayes()` from the limma<sup>44</sup> Bioconductor package v3.34.5 to  
695 assess age-associated changes to DNAm levels with the linear model  $\sim \text{Age}$ . For the cell  
696 type-specific samples, we used a linear model  $\sim \text{Age} * \text{Cell Type}$  to assess cell type,  
697 overall age, and age in a cell type changes to DNAm levels. We subset CpGs to those that were

698 significantly differentially methylated by age in homogenate samples ( $p < 1 \times 10^{-4}$ ) and plotted the  
699 coefficients for each CpG in **Figure S2A-D**. The relationship between each variable was  
700 quantified using Fisher's exact test.

#### 701 *Roadmap Epigenome enrichments*

702 We computed the relative enrichments of different genomic regions using Epigenome Roadmap  
703 data<sup>12</sup> by computing the proportion of bases in each of the 15 ChromHMM states for each of the  
704 cells and tissues provided by the Consortium. We compared the proportion of bases in each  
705 state within each candidate region set to the overall genome and computed the corresponding  
706 log<sub>2</sub> enrichments between the regions and this genomic background. We compared DMR and  
707 mCpG-based methylation feature regions to all profiled cell types in the Consortium for these  
708 analyses.

#### 709 *Assessing the contribution of neuronal subtypes*

710 The percent of neuronal subtype-specific bases were calculated by reducing the total  
711 subtype-specific CpG-DMRs from Luo *et al.*<sup>14</sup>, reducing the bases in each group of cdDMRs and  
712 calculating the percent of cdDMR bases that intersected the merged subtype-specific CpG-DMR  
713 bases.

#### 714 *Visualizing DMRs*

715 Plots for the DMRs were made using bsseq<sup>42</sup> v1.14.0, EnsDb.Hsapiens.v75 v2.99.0, and  
716 RColorBrewer v1.1-2. Genes within 20 kb of a DMR were retained for the plots. Genes and



717 exons were included using the `annoTrack` argument and we used `extend=2000` for making the  
718 plots.

### 719 *CpH processing*

720 Using Bismark v0.16.3<sup>37</sup> we created report files using the methylation extractor program with the  
721 `CX_context` and `split_by_chromosome` options for the hg19 human genome in order to  
722 extract the methylation values for the CpHs. Then for each chromosome, using the `bsseq`<sup>42</sup>  
723 Bioconductor package v1.10.0 we loaded the Bismark<sup>37</sup> report files and added the `c_context`  
724 and `trinucleotide_context` information from Bismark using custom R code based on the `bsseq`  
725 internal code that uses the `data.table` package v1.10.4. After combining the results for each  
726 chromosome, we filtered the CpHs to keep only those where all samples had a minimum  
727 coverage of 5 (58,109,566 number passed the filter).

### 728 *Lister, et al. (2013) data processing*

729 We downloaded the WGBS data from Lister *et al.* data<sup>1</sup> with SRA accession SRP026048. We  
730 then processed and aligned the data following the same steps we used for our data. Using  
731 `bsseq`<sup>42</sup> as in the *CpH processing* section, we extracted the methylation values from the  
732 Bismark<sup>37</sup> report files and added the `c_context` and `trinucleotide_context` information  
733 per chromosome. We then merged the results for all the chromosomes retaining only the CpG  
734 and CpH positions we observed with a minimum coverage of 3 and 5 in our data, respectively.  
735 To assess the replication of our cell type DMR results we computed mean methylation  
736 differences across the CpG positions comparing neuron and non-neuron samples in the Lister *et*  
737 *al.* data<sup>1</sup> using the `rowttests()` function from the `genefilter` package version 1.56.0. We then

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

#### 745 *Homogenate data processing*

746 We processed the prenatal and postnatal homogenate brain samples using the same procedure  
747 described in the *Data processing/alignment* section to produce Bismark<sup>37</sup> report files. Then  
748 using `bsseq`<sup>42</sup> we extracted the methylation values for the CpG positions observed in our  
749 postnatal samples in order to make them comparable to each other.

#### 750 *Identification of differentially methylated positions*

751 With the set of CpGs and CpGs with a minimum coverage 3 and 5 in all samples, respectively,  
752 and the same models for identifying the DMRs, we identified the differentially methylated  
753 positions (DMPs), keeping the CpGs and CpGs separate. For the CpGs we further filtered to  
754 keep only those where at least 5 samples had a methylation value greater than 0 (40,818,742,  
755 70.2%). We used the limma Bioconductor package v3.30.13 for determining the DMPs by  
756 running functions `lmFit()` and `eBayes()` with default parameters and FDR<5%.

757 *Enrichments for HARs and enhancers*

758 We calculated the enrichment of genomic segments overlapping cdDMRs and human  
759 accelerated regions (HARs) and enhancers<sup>15</sup> using Fisher's exact test. We calculated the  
760 overlap of methylation features (DMRs, mCpH, expression-associated cytosines) and HARs or  
761 enhancers with the entire set of CpG clusters used to identify the DMRs as background. We  
762 corrected for multiple testing using the false discovery rate (FDR).

763 *Stratified linkage disequilibrium score regression*

764 GWAS summary statistics for 30 phenotypes (described in Anttila et al. and Rizzardi et  
765 al.)<sup>20,21</sup> were downloaded from the sources listed in Table S6. We used LDSC (LD Score) v1.0.0  
766 to estimate the proportion of heritability captured in 16 sets of genomic DNAm features for each  
767 GWAS phenotype, including the DMRs and LMRs defined above in this work, non-differentially  
768 methylated CpG clusters (called above before the DMR analysis, excluding DMR sequence),  
769 central nervous system annotations included in the LDSC package (referred as CNS (LDSC)),  
770 and regions annotated as putatively regulatory in human brain using chromHMM (i.e., the union  
771 of regions annotated as "Bivalent Enhancer", "Bivalent/Poised TSS", "Genic enhancers",  
772 "Flanking Active TSS", "Active TSS", "Strong transcription", and "Enhancers" in the following  
773 tracks accessed using the AnnotationHub Bioconductor package (v2.14.5)<sup>45</sup>: AH46920,  
774 AH46921, AH46922, AH46923, AH46924, AH46925, AH46926, AH46927, AH46934, and  
775 AH46935).

776 We first converted the GWAS summary statistics into the .sumstats format using  
777 `munge_sumstats.py` and keeping only HapMap 3 SNPs (downloaded from  
778 [https://data.broadinstitute.org/alkesgroup/LDSCORE/w\\_hm3.snplist.bz2](https://data.broadinstitute.org/alkesgroup/LDSCORE/w_hm3.snplist.bz2)) as described in the

779 Partitioned Heritability LDSC tutorial. We made .annot files for each custom feature set based  
780 on the list of SNPs in the CNS cell type annotations provided in the LDSC package and  
781 estimated partitioned LD scores for each feature using 1000 Genomes plink files (downloaded  
782 from [https://data.broadinstitute.org/alkesgroup/LDSCORE/1000G\\_Phase3\\_plinkfiles.tgz](https://data.broadinstitute.org/alkesgroup/LDSCORE/1000G_Phase3_plinkfiles.tgz)) using  
783 `ldsc.py`. We finally estimated the partitioned heritability for each feature-phenotype  
784 combination by adding each feature individually to the “baseline model” including 53 baseline  
785 annotations described in Finucane et al.<sup>22</sup>

#### 786 *Enrichments for genes associated with brain disorders*

787 We calculated the enrichment of genes overlapping different methylation features in gene sets  
788 described by Birnbaum *et al.*<sup>24</sup> We measured what fraction of the genes in each set overlapped  
789 methylation features (ie, DMRs, mCpH, and expression-associated cytosines) with Fisher’s  
790 exact test using all expressed genes with Entrez IDs as background. We corrected for testing  
791 multiple disorder gene sets using the false discovery rate (FDR).

#### 792 *RNA-seq data processing*

793 Raw sequencing reads were mapped to the hg19/GRCh37 human reference genome with  
794 splice-aware aligner HISAT2 v2.0.4<sup>46</sup>. Feature-level quantification based on GENCODE release  
795 25 (GRCh38.p7) annotation on hg19 coordinates was run on aligned reads using featureCounts  
796 (subread v1.5.0-p3)<sup>47</sup>. Using custom R code we processed the different feature counts and  
797 created RangedSummarizedExperiment objects using the SummarizedExperiment  
798 Bioconductor package v1.4.0.

799 *Percent spliced in (PSI) calculation*

800 We calculated the percent spliced in using the SGSeq<sup>48</sup> Bioconductor package v1.12.0 and the  
801 Gencode v25 annotation for the GRCh37 human reference genome  
802 ([ftp.sanger.ac.uk/pub/gencode/Gencode\\_human/release\\_25/GRCh37\\_mapping/gencode.v25lift](ftp.sanger.ac.uk/pub/gencode/Gencode_human/release_25/GRCh37_mapping/gencode.v25lift)  
803 [37.annotation.gtf.gz](ftp.sanger.ac.uk/pub/gencode/Gencode_human/release_25/GRCh37_mapping/gencode.v25lift)) from the BAM files generated by HISAT2. We used default arguments  
804 except for the function `analyzeVariants()` where we used a `min_denominator=10`.

805 *Differential expression between cell types*

806 We combined the gene counts for the polyA+ and RiboZero sequencing protocols for the  
807 cell-sorted RNA-seq data: 3 NeuN+ and 3 NeuN- samples for a total of 12 RNA-seq sequencing  
808 runs. We calculated the library size normalization factors using `calcNormFactors()` from  
809 `edgeR`<sup>49</sup> (v3.22.3) and identified the differentially expressed genes using `voom()`, `lmFit()` and  
810 `eBayes()` from `limma`<sup>44,50</sup>. We repeated this procedure for the exon counts.

811 *Gene ontology analyses*

812 Gene ontology enrichment analyses were performed using `clusterProfiler`<sup>51</sup> v3.6.0 using the  
813 options `pAdjustMethod="BH"`, `pvalueCutoff=0.1`, and `qvalueCutoff=0.05` on the  
814 Entrez IDs for each expression feature for the BP, CC and MF GO ontologies. Only cytosines or  
815 DMRs overlapping genes were included.

816 *Methylation vs expression associations*

817 With the sorted RNA-seq data, we computed the average gene expression per age group  
818 (infant, child, teen) and cell type (six groups), and correlated these values to average DNAm  
819 levels in the same groups in both the CpG and CpH contexts at the gene promoter and body,  
820 exon (500bp window), and splice junction (50bp into each intron) levels. With the  
821 RangedSummarizedExperiment objects with the RNA-seq polyA homogenate data and the  
822 bsseq objects with the CpG and CpH data we determined which CpG and CpH positions  
823 explained changes in expression (RPKM) at the gene or exon level as well as in percent spliced  
824 in (PSI). We retained only the expression and PSI data from the postnatal samples and matched  
825 them by brain identifier to the neuronal methylation data with a final sample size of 22. We  
826 filtered lowly expressed features using the `expression_cutoff()` function from the `jaffelab`  
827 package v0.99.18: mean RPKM>0.22 mean for genes, 0.26 for exons. For genes and exons we  
828 transformed the expression values to  $\log_2(\text{RPKM} + 1)$  and extracted the raw PSI values. Using  
829 MatrixEQTL<sup>52</sup> v2.2 (GitHub b9a9f01 patch) we then identified the methylation quantitative trait  
830 loci (QTL) for the CpG and CpH methylation data separately using the function  
831 `Matrix_eQTL_main()` function with options `pvOutputThreshold=0,`  
832 `pvOutputThreshold.cis=5e-4, useModel=modelLINEAR, cisDist=1000`. We identified  
833 marginal CpG associations near CpH associations by running MatrixEQTL again for the CpG in  
834 a +- 1kb window around the CpH positions with an association with expression at FDR <5% for  
835 each expression feature type using the same parameters as above except for  
836 `pvOutputThreshold.cis=0.01`. We filtered the associations to retain only those having at  
837 least 11 samples with non-zero methylation and 11 samples with non-one methylation values to  
838 remove extreme cases. We further restricted the results to protein coding genes and dropped  
839 any with infinite t-statistics. To assess whether age confounds the relationship between  
840 methylation and expression, we used a multiple linear regression model adjusting for age and

841 checked if the methylation coefficient was still  $FDR < 5\%$ . Venn diagrams in **Figure 3** were made  
842 with the VennDiagram package v1.6.18. For the gene and PSI associations we used the gene  
843 ID to check if it was present in the 3,473 differentially expressed genes from the sorted RNA-seq  
844 data (described above) with higher expression in neurons and the top 5,000 DE genes with  
845 higher expression in glia at  $FDR < 5\%$ ; similarly we did so for exons and the top 5,000 DE exons  
846 ( $FDR < 5\%$ ) with higher expression in each cell type. The LIBD WBS Expression explorer at  
847 <https://jhubiostatistics.shinyapps.io/wgbsExprs/> was made using the bsseq<sup>42</sup> v1.14.0, DT v0.4,  
848 SGSeq<sup>48</sup> v1.12.0 and shiny v1.0.5 R packages.

#### 849 *Global autocorrelation*

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

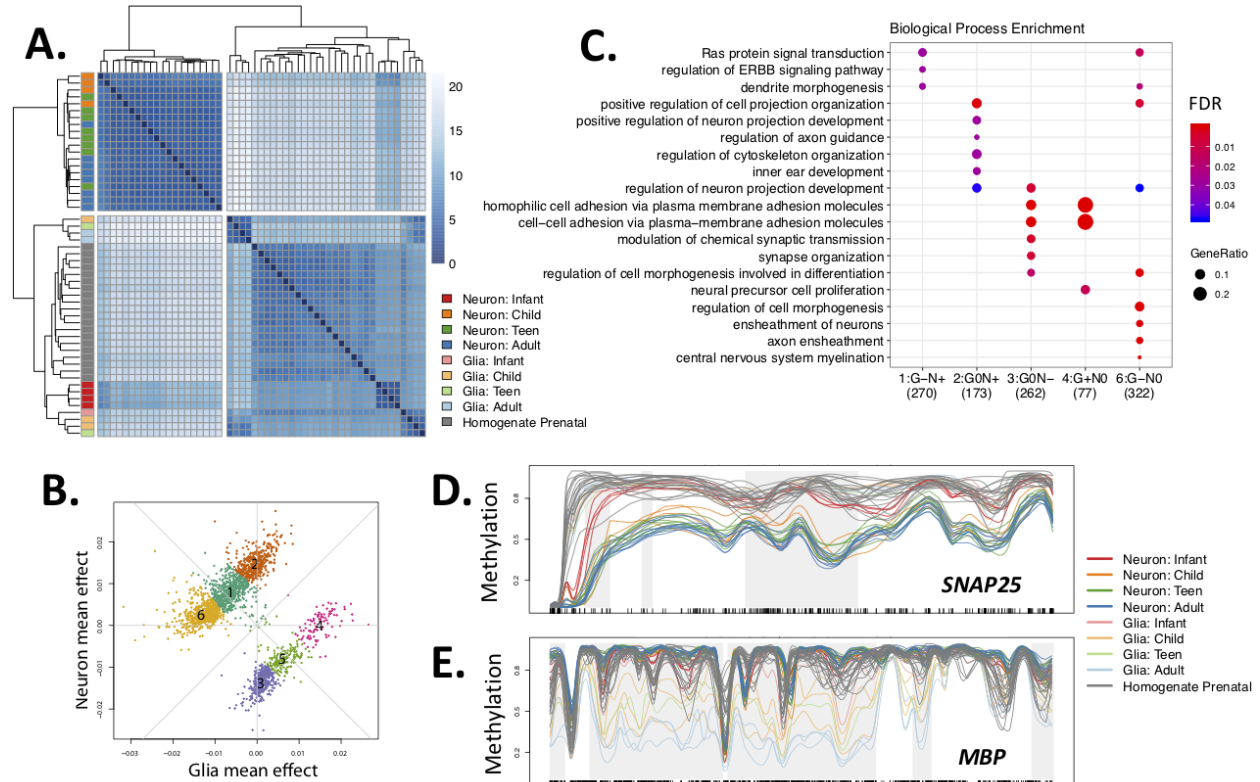
#### 860 *Autocorrelation within DMRs*

861 Similar to the global autocorrelation, we extracted the methylation values at CpGs with a  
862 minimum coverage of 3 and the CpGs with a minimum coverage of 5 that were within each of  
863 the sets of DMRs (age, cell type or interaction). We then computed the autocorrelation for DMRs  
864 with a least 5 different cytosines using the `acf()` function with a `lag.max=4` and calculated the  
865 mean auto-correlation among the neuronal and glial samples. The lag is proportional to the  
866 genomic distance as shown in **Figure S20 (lag and distance)**.

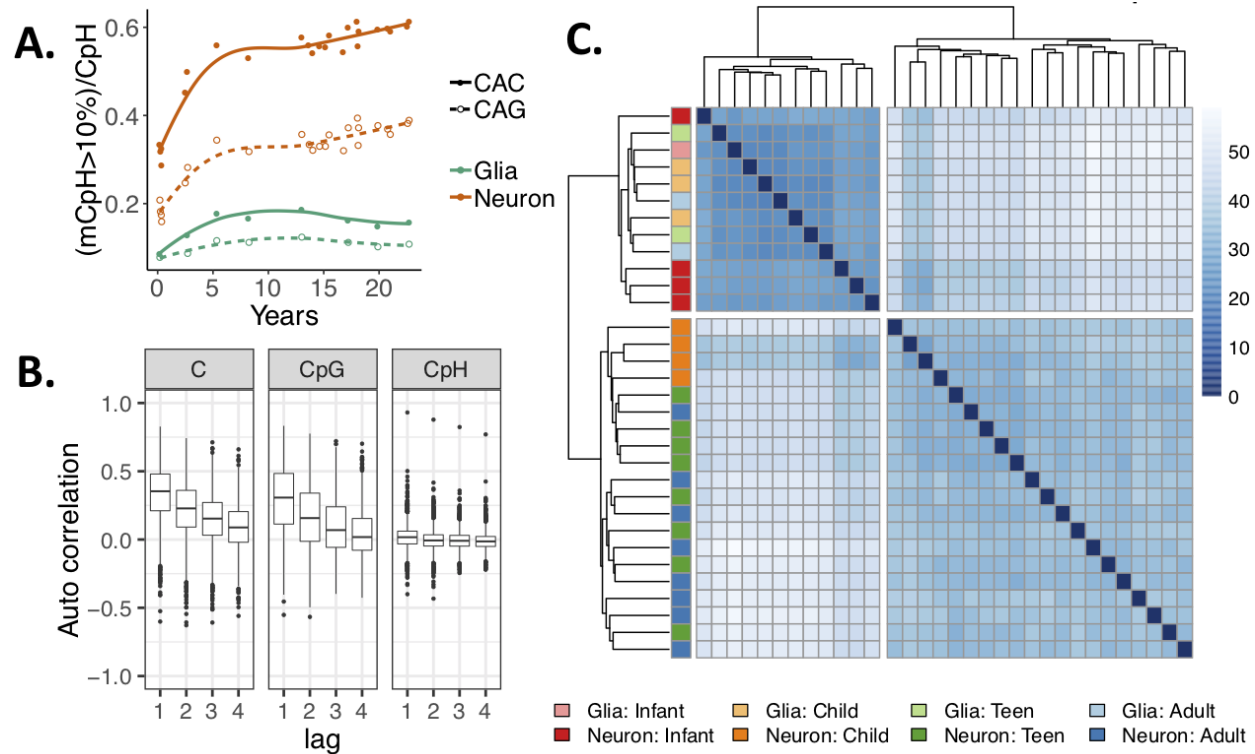


867 **Table 1: Summary of methylation-expression associations.** Mean values are shown for the  
 868 number of samples with  $\beta > 0$ ,  $\beta < 1$ , and expression change variables (either  $\Delta\text{PSI}$  or  
 869  $\Delta\log_2(\text{RPKM} + 1)$  for genes and exons). Columns from “Age Confounded” are proportions. See  
 870 **Table S10** for a full description of the variables.

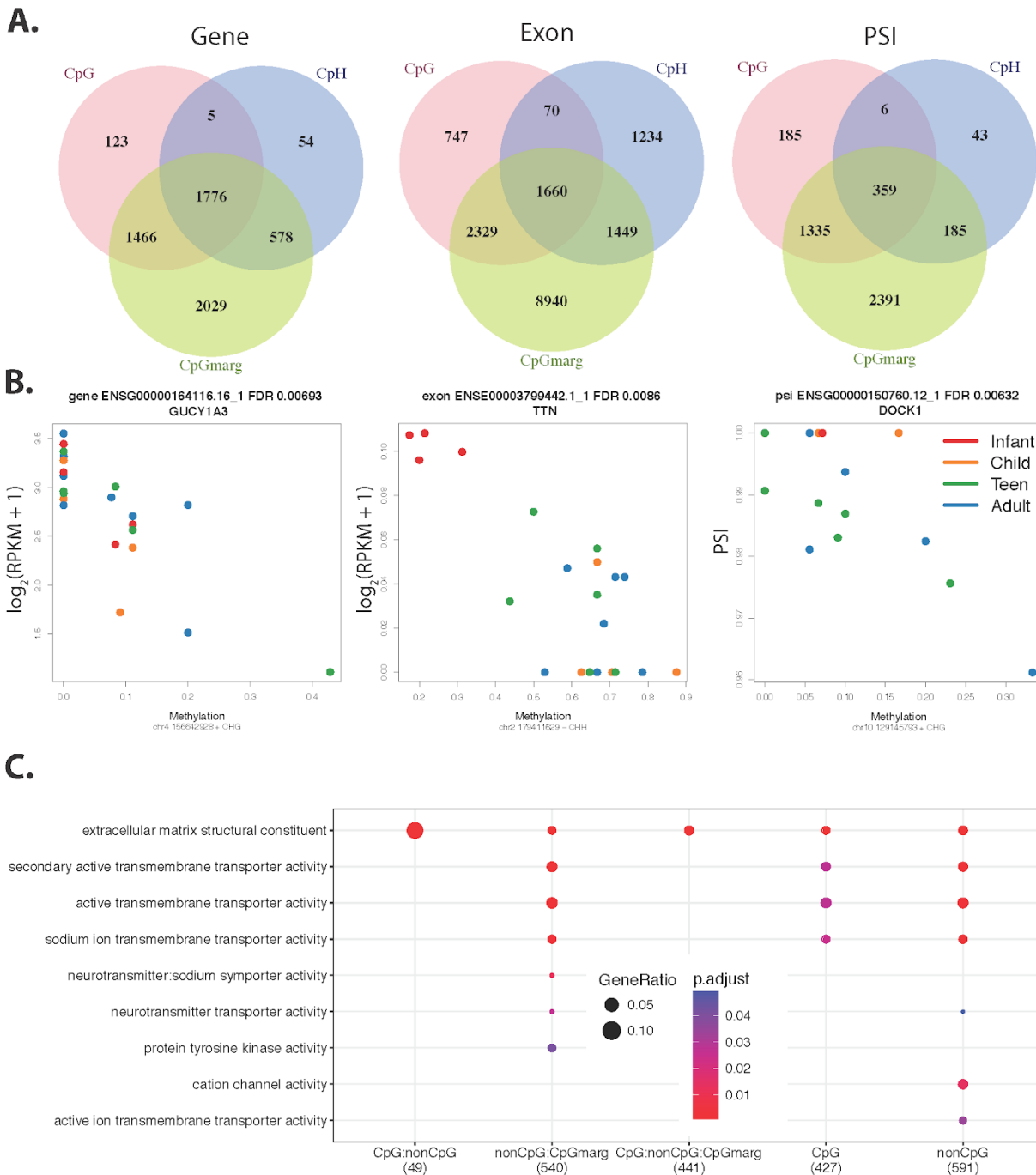
Feature	Meth. Type	Direction	N	N Unique C	N Unique Features	Mean N Meth>0	Mean N Meth<1	Mean Expr Change	Prop. Age Confounded FDR >=5%	Prop. In Promoter	Prop. In Gene Body	Prop. In Gene Flanking	Prop. CHG	Prop. CHH	Prop. DE Glia	Prop. DE Neuron	Prop No Diff
gene	CpG	down	18458	18353	2074	21.9	21.9	1.39	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



871 **Figure 1: Regional cell type-specific developmental mCpG trajectories. (A)** Euclidean  
 872 distances between samples within cdDMRs shows that older neuronal samples cluster  
 873 separately from infant neuronal samples, glia regardless of age, and bulk prenatal cortex. **(B)**  
 874 Decomposing cdDMRs patterns into 6 clusters using *k-means* based on glia and neuron mean  
 875 mCpG changes per year of life. **(C)** The top five most enriched gene ontology terms for each of  
 876 the six groups in **(B)** highlights diverse biological processes amongst the groups. No terms were  
 877 enriched for Group 5. **(D)** Example Group 3 cdDMR within *SNAP25*. **(E)** Example Group 6  
 878 cdDMR within *MBP*. Gray shading indicates the boundaries of the cdDMR, and black tick marks  
 879 on the x-axis indicate the position of CpGs. Key: Neuron (NeuN+); Glia (NeuN-); Infant (0-1  
 880 years); Child (1-10 years), Teen (11-17 years) and Adult (18+ years).

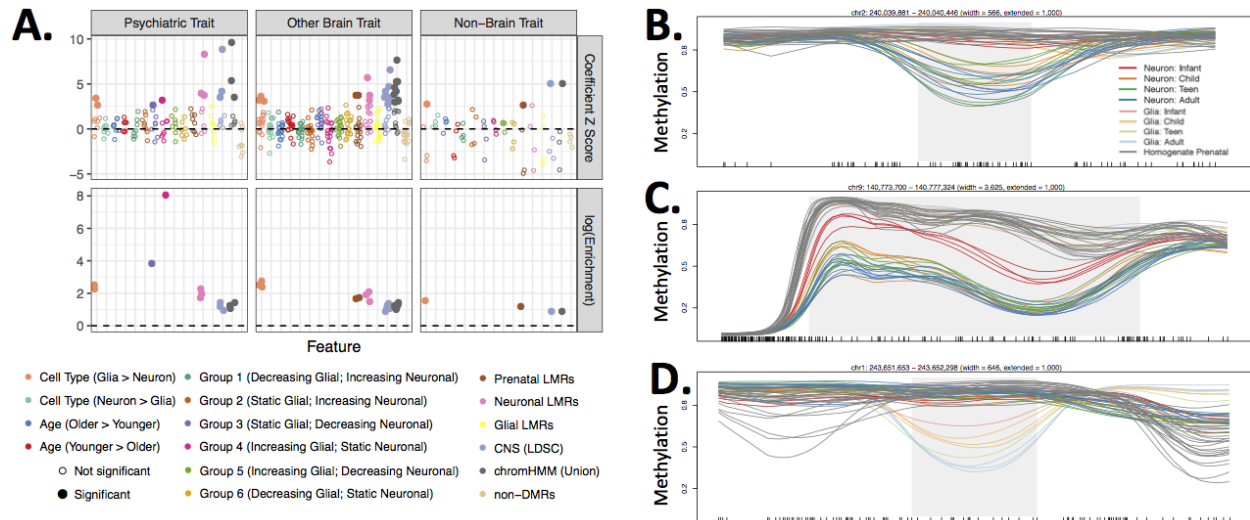


881 **Figure 2: CpH methylation patterns across brain development. (A)** The proportion of CAC  
 882 and CAG sites that are greater than 10% methylated in neurons and non-neurons (glia) across  
 883 brain development. **(B)** Autocorrelation levels for different cytosine contexts in neurons.  
 884 Autocorrelation levels were similar for mCpG and all cytosines, with uncorrelated levels in the  
 885 CpH context. **(C)** Euclidean distances between samples based on mCpH within cdDMRs again  
 886 cluster infant neurons dark red) with glia of all ages (light colors) rather than with older neurons.



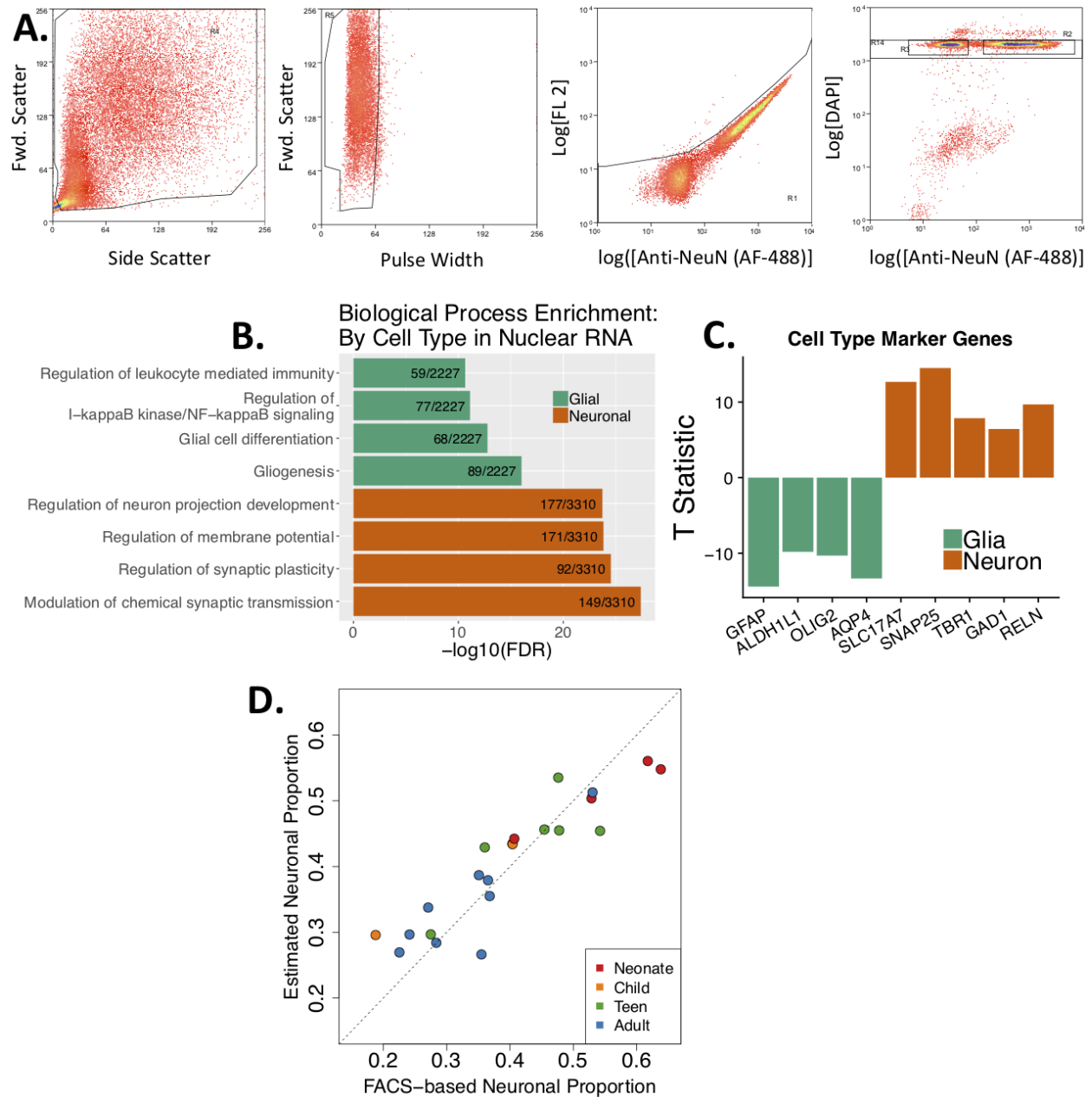
887 **Figure 3: Methylation associations with expression. (A)** Venn diagrams of the methylation  
 888 associations by unique feature for the gene, exon and PSI. The sets are determined by if the  
 889 association is FDR<5% genome-wide for CpG and CpH or if it is a CpG marginally significant  
 890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

891 and expression at the gene level colored by age: red - infant, orange - child, green - teen, blue -  
892 adult. *GUCY1A3* contains one of the top CpH differentially expressed between neurons and glia.  
893 Expression of an exon of *TTN*, an autism-associated gene, is negatively associated with mCpH.  
894 *DOCK1* PSI of an alternative end site is negatively associated with mCpH. **(C)** Enriched  
895 molecular function ontology terms for methylation-associated exons by the venn diagram groups  
896 from (A).



897 **Figure 4: DNAm patterns and brain trait heritability.** (A) Results assessing enrichment for  
 898 heritability of 30 phenotypes within 16 groups of DNAm features using Stratified Linkage  
 899 Disequilibrium Score Regression (SLDSR). Each dot represents results for a single  
 900 phenotype:DNAm feature pair. The color indicates the DNAm feature, and the phenotypes are  
 901 stratified by column into psychiatric phenotypes, other brain-related phenotypes (ie, neurological  
 902 or behavioral-cognitive), or non-brain-related traits. The upper row shows the coefficient Z score  
 903 for each tested phenotype:DNAm pair, or the amount of additional heritability explained by the  
 904 DNAm feature over 53 baseline features in the SLDSR model. The lower row shows the  
 905 enrichment score, or the proportion of heritability attributed to the feature divided by the  
 906 proportion of SNPs in the feature. For clarity enrichment scores of only the significant  
 907 feature-trait combination are depicted. Filled in circles indicate significantly enriched heritability  
 908 for a phenotype in a feature (coefficient p-value corrected using Holms method  $\leq 0.05$ ). (B) A  
 909 cdDMR overlapping *HDAC4*, a gene associated with autism spectrum disorder (ASD), shows  
 910 the Group 3 pattern of decreasing neuronal and static glial DNAm. (C) A cdDMR overlapping  
 911 *CACNA1B*, a gene associated with ASD, shows the Group 5 pattern of decreasing neuronal and  
 912 increasing glial DNAm. (D) A cdDMR overlapping *AKT3*, a gene associated with schizophrenia,

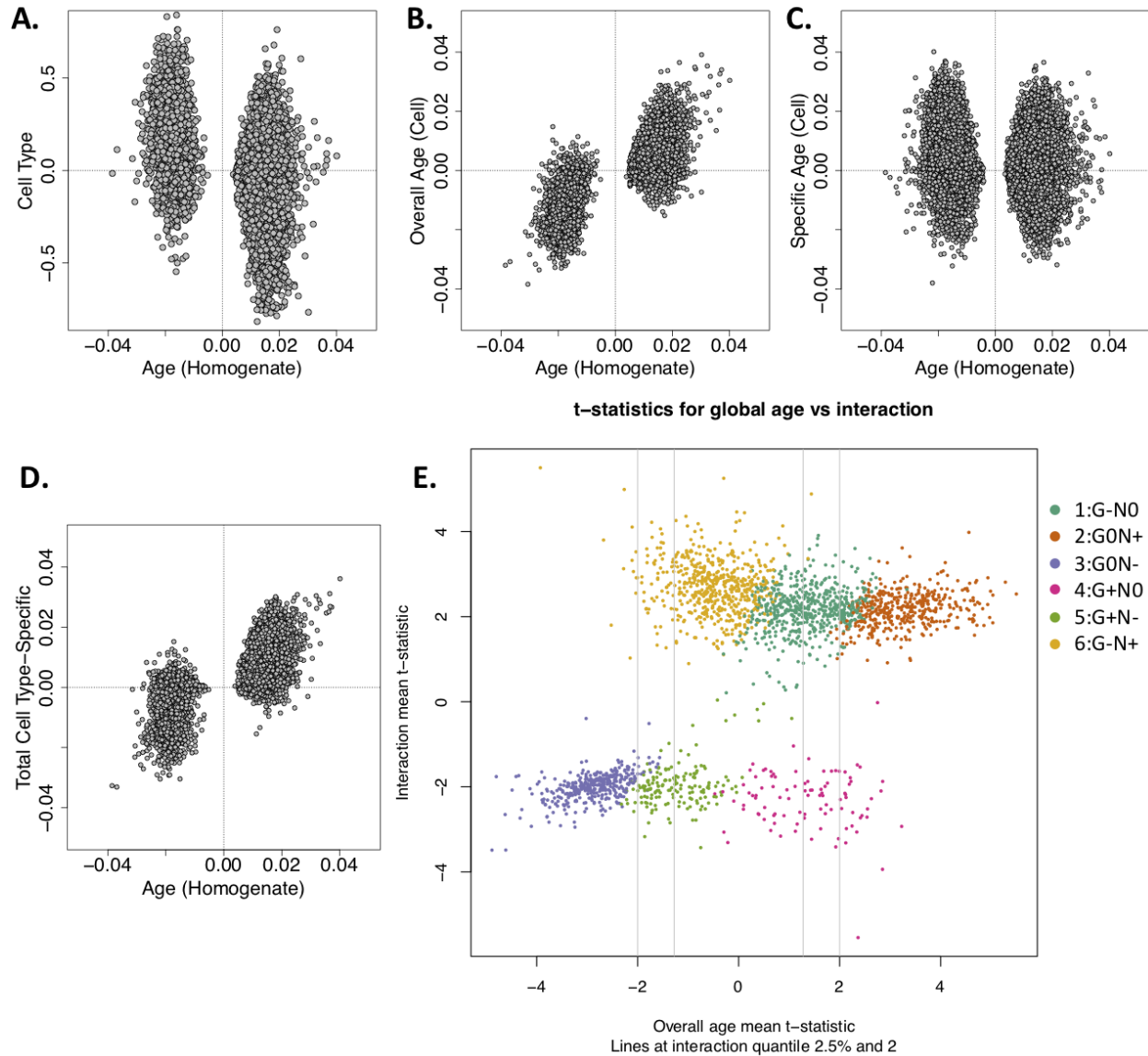
913 shows the Group 6 pattern of decreasing glial and static neuronal DNAm. Gray shading  
914 indicates the boundaries of the cdDMR, and black tick marks on the x-axis indicate the position  
915 of CpGs.



916 **Figure S1: Confirmation of neuronal- and glial-enriched identity of NeuN+ and NeuN-**  
 917 **samples. (A)** Representative gating strategy from fluorescence-activated nuclear sorting by  
 918 NeuN antibody. Debris is first reduced by selecting events based on forward scatter and side  
 919 scatter, then aggregates are reduced by measuring pulse width. Autofluorescent events are  
 920 discarded by measuring true Alexa fluor-488 signal compared to signal in the FL2 channel.

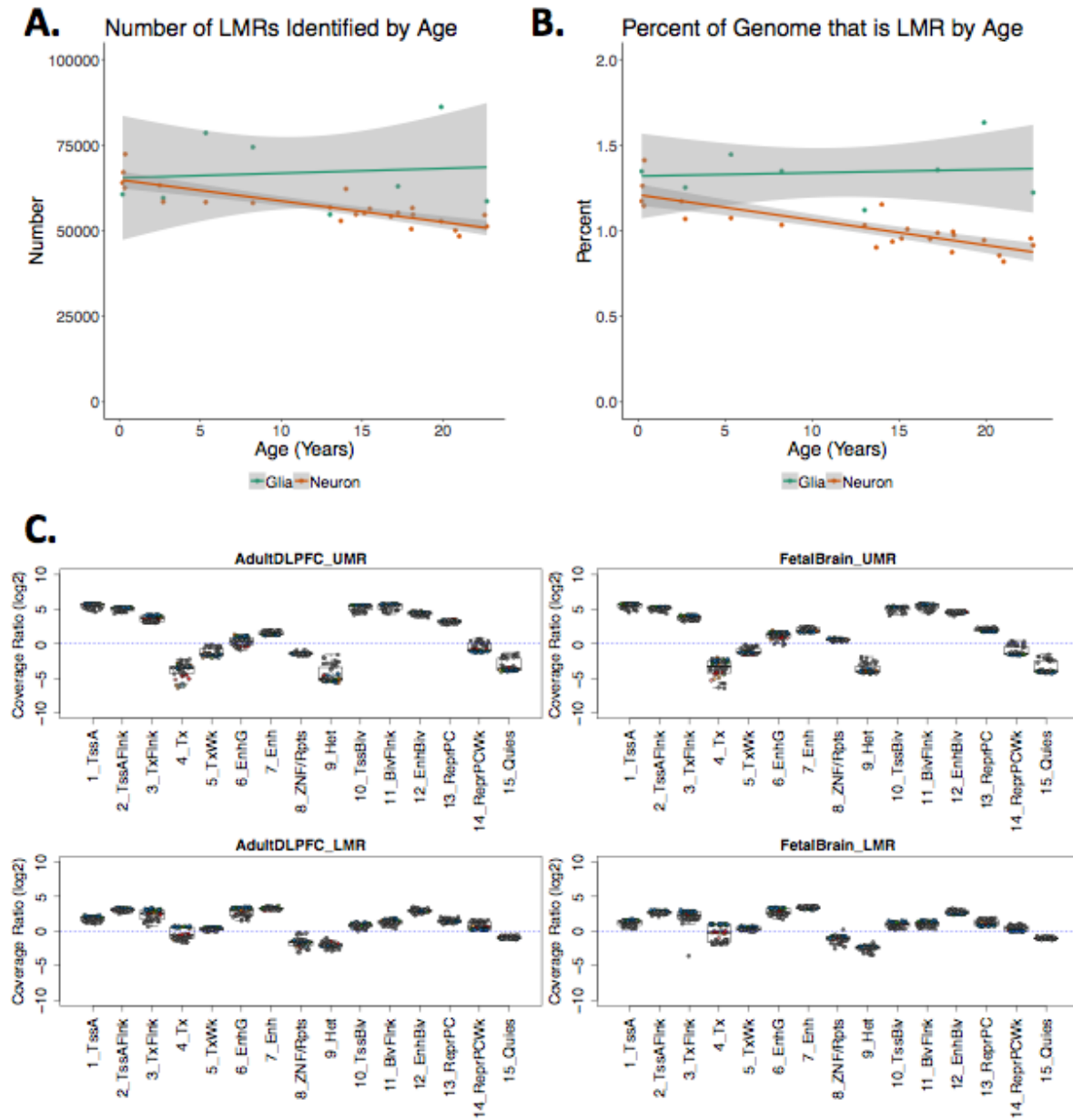


921 Finally, singlets are determined by DAPI staining, and Alexa fluor-488 positive and negative  
922 events are collected. **(B)** Top biological processes enriched in genes significantly differentially  
923 expressed in nuclear RNA from NeuN+ (labeled “Neurons”) and NeuN- (labeled “Glia”) at  
924 FDR<0.05. **(C)** T statistics for marker genes for neuronal and non-neuronal identity show that  
925 the genes are differentially expressed in NeuN+ and NeuN- nuclear RNA (FDR<0.05). **(D)** The  
926 estimated proportion of neurons in homogenate WGBS samples based on statistical  
927 deconvolution using differentially methylated cytosines between NeuN+ and NeuN- samples is  
928 highly correlated with the empirically-derived proportion of NeuN+ events.

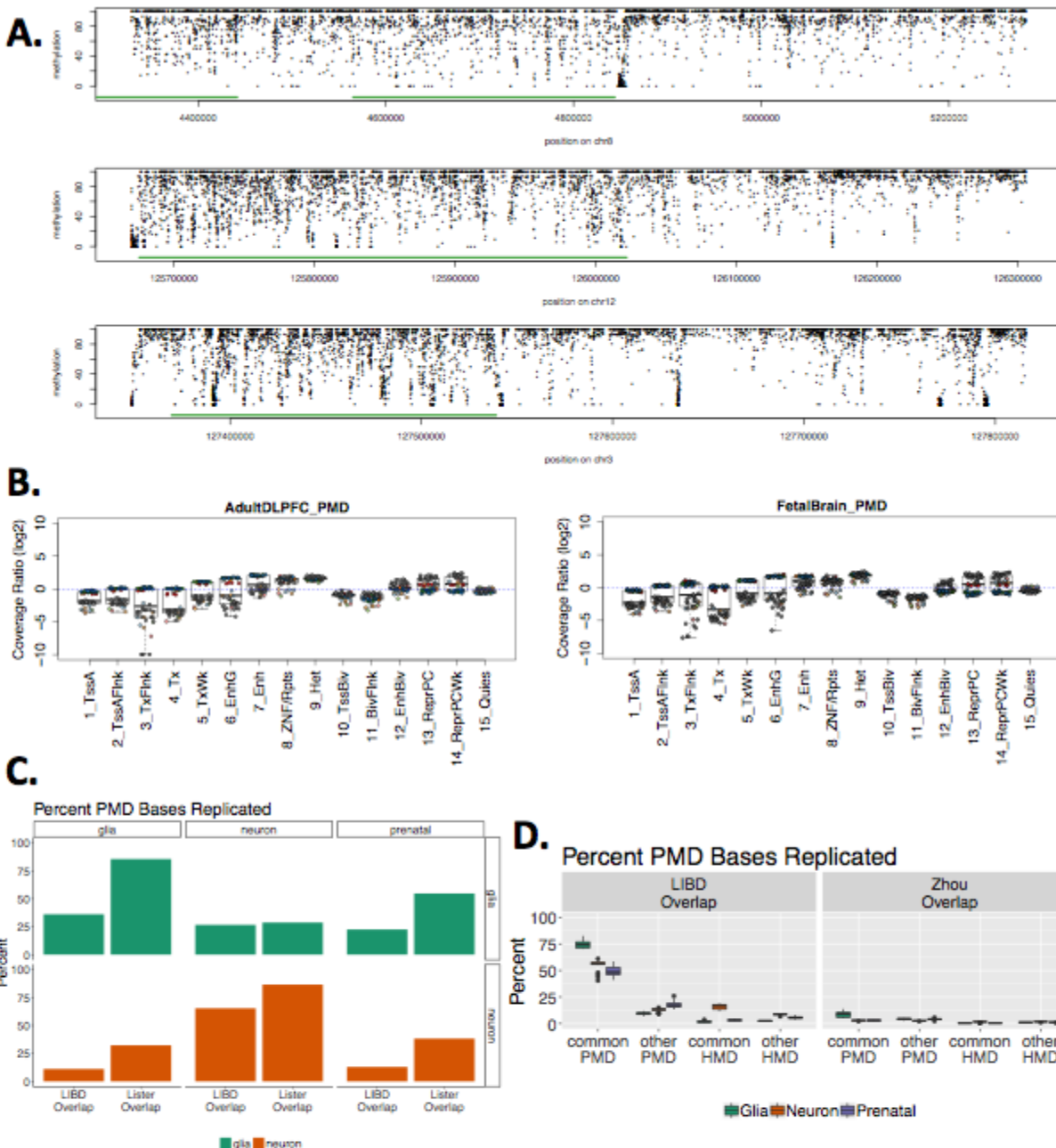


929 **Figure S2: Detecting developmental changes in homogenate vs. cell type-specific DNAm**  
 930 **data. (A)** Developmental age effect coefficients of individual CpGs as measured in homogenate  
 931 samples compared against cell type effect coefficients in cell type-specific samples after  
 932 adjusting for age. **(B)** Developmental age effect coefficients of individual CpGs as measured in  
 933 homogenate samples compared against the developmental effect adjusting for cell type in cell  
 934 type-specific samples. **(C)** Developmental age effect coefficients as measured in homogenate  
 935 samples compared against the cumulative age and cell type interaction effect coefficients in cell

936 type-specific samples at the CpG level. **(D)** Developmental age effect coefficients as measured  
937 in homogenate samples compared against the estimated cell type-specific age effects for cell  
938 type-specific samples at the CpG level. **(E)** Overall age mean t-statistic for the cdDMRs (x-axis)  
939 against the mean interaction t-statistic (y-axis) with the 2.5% quantile from the y-axis shown on  
940 the x-axis. Colors are the same as those from **Figure 1B**: teal=Group 1 (decreasing glial  
941 methylation, increasing neuronal methylation), orange=Group 2 (static glial methylation,  
942 increasing neuronal methylation), purple=Group 3 (static glial methylation, decreasing neuronal  
943 methylation), pink=Group 4 (increasing glial methylation, static neuronal methylation),  
944 green=Group 5 (increasing glial methylation, decreasing neuronal methylation), gold=Group 6  
945 (decreasing glial methylation, static neuronal methylation).

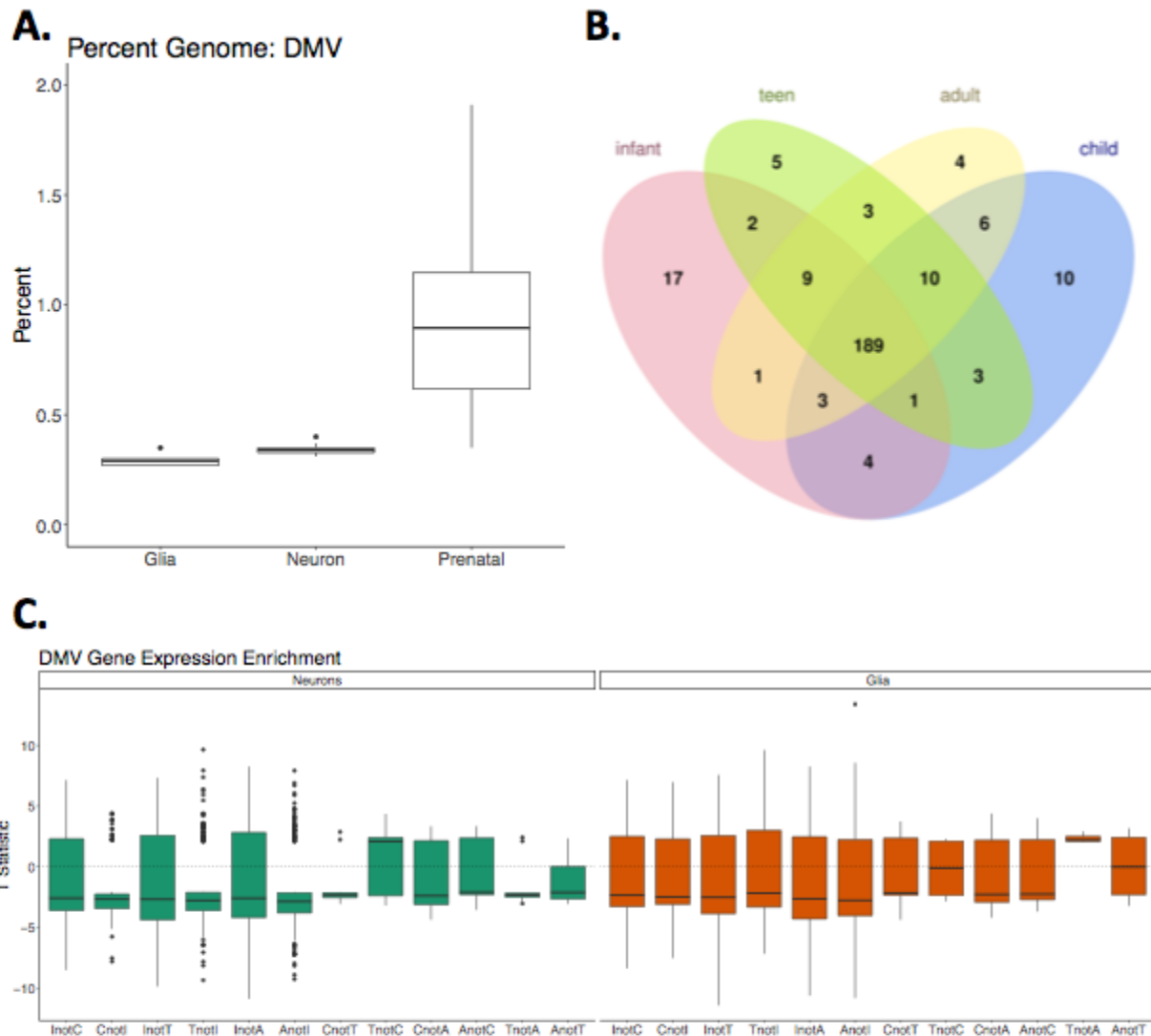


946 **Figure S3: Unmethylated Regions (UMRs) and Low-methylated regions (LMRs). (A)**  
 947 **Number of LMRs and (B) percent of the genome covered by LMRs by age stratified by cell type.**  
 948 **Shading indicates the standard error of the linear model. (C) Roadmap Epigenomics Consortium**  
 949 **predicted chromatin-enriched states for LMRs and UMRs as defined in adult DPLFC and fetal**  
 950 **brain.  $\log_2(\text{Coverage Ratio})$  represents the enrichment of the proportion of bases within LMRs**  
 951 **or UMRs in a chromatin state compared to the rest of the genome.**

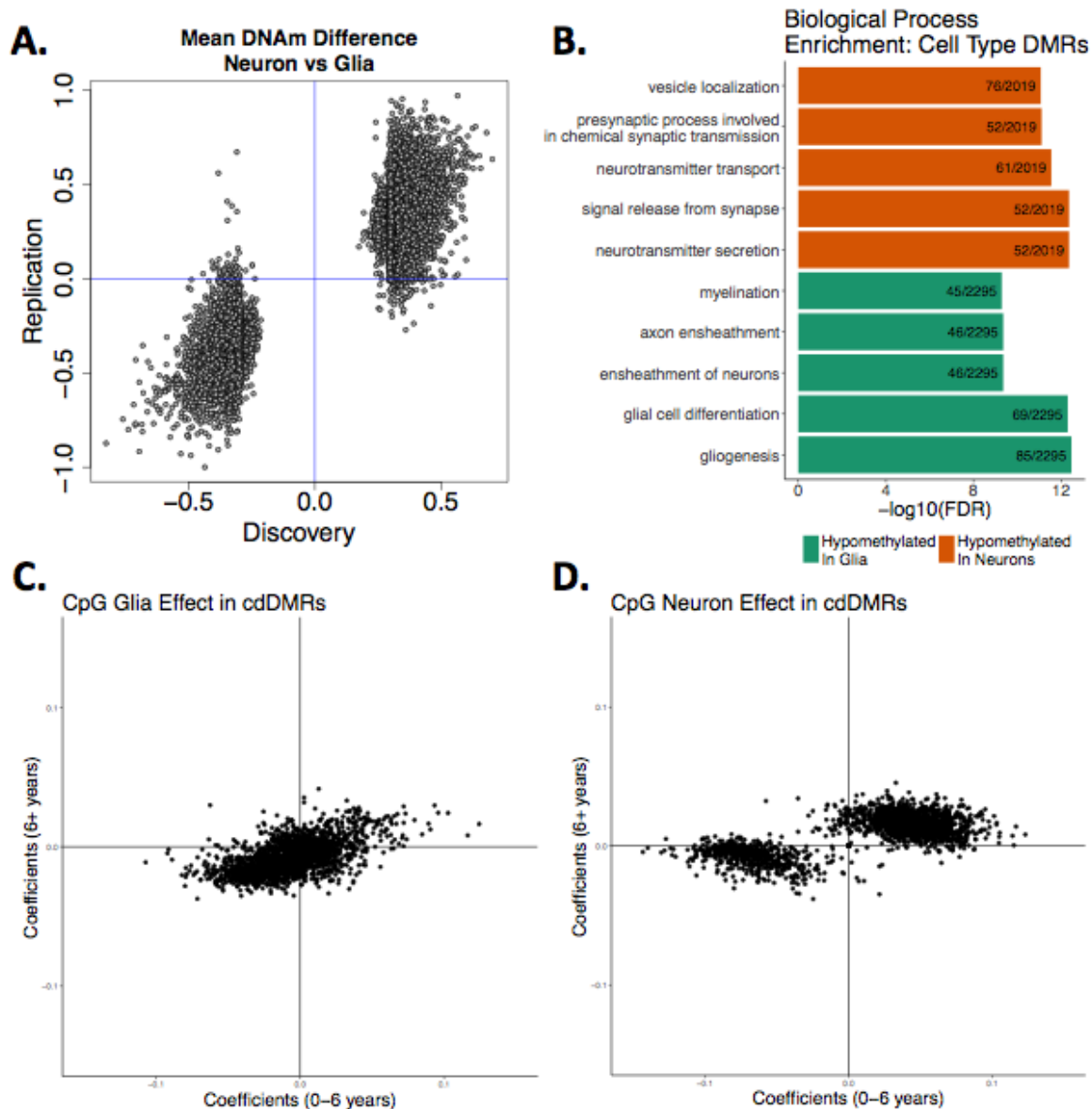


952 **Figure S4: Partially methylated domains (PMDs).** (A) Example PMDs on chromosomes 8, 12,  
 953 and 3. PMDs are highlighted in green. (B) Roadmap Epigenomics Consortium chromatin state  
 954 enrichment for PMDs.  $\log_2(\text{Coverage Ratio})$  represents the enrichment of the proportion of  
 955 bases within PMDs in a state compared to the rest of the genome. (C) Percent of PMD

956 base-pairs that are replicated in the Lister *et al.* (2013) FANS samples<sup>1</sup>. Rows represent Lister  
957 *et al.* (2013) glia and neurons, and columns represent the glia, neurons and prenatal samples  
958 from this paper (LIBD). Each bar represents the percent of PMD bases shared between that  
959 quadrant's cell types, using either total LIBD or total Lister PMD bases as the denominator. **(D)**  
960 Percent of PMD base pairs per sample that are replicated in either common or unique PMDs  
961 and high methylated domains (HMDs) identified in Zhou *et al.*<sup>13</sup> Samples are colored based on  
962 whether the sample was postnatal neuron, postnatal glia, or bulk prenatal cortex.

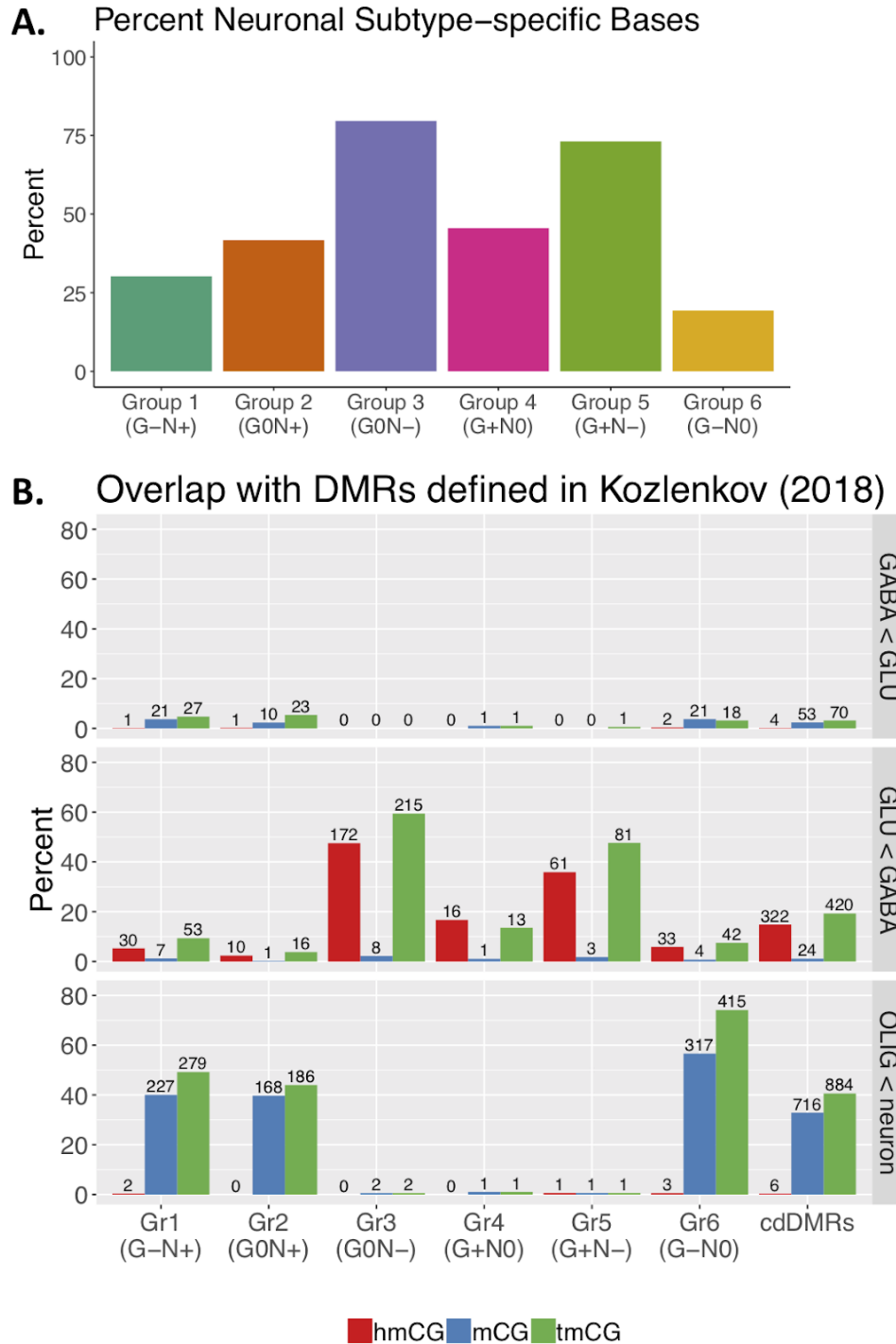


963 **Figure S5: DNA methylation valleys (DMVs).** (A) Percent of the genome covered by DMVs in  
 964 postnatal glia and neurons and bulk prenatal cortex. (B) Overlap of transcription factor (TF)  
 965 genes within DMVs by age in neurons. (C) Expression enrichment between age groups in TF  
 966 genes excluded from DMVs in one group but not the other in neurons and glia. A negative  
 967 T-statistic signifies greater expression in the age group in which the gene is not in a DMV.  
 968 I=Infant (0-1 years); C=Child (1-10 years), T=Teen (11-17 years) and A=Adult (18+ years).



969 **Figure S6: Differentially methylated regions (DMRs).** (A) Replication of cell type DNAm  
 970 differences at the CpG level between neurons and glia (NeuN+ and NeuN- samples) in our and  
 971 Lister *et al.* data<sup>1</sup>. X and Y axes are linear model coefficients. (B) Enriched biological process  
 972 ontology terms in the DMRs by cell type. The number of genes within each ontology group that  
 973 overlaps a cell type DMR is listed. (C) Coefficients of linear regression on the mean mCpG level  
 974 per cdDMR in samples younger than 6 years and older. (D) Coefficients of linear regression on  
 975 the mean mCpG level per cdDMR in samples younger than 6 years and older.

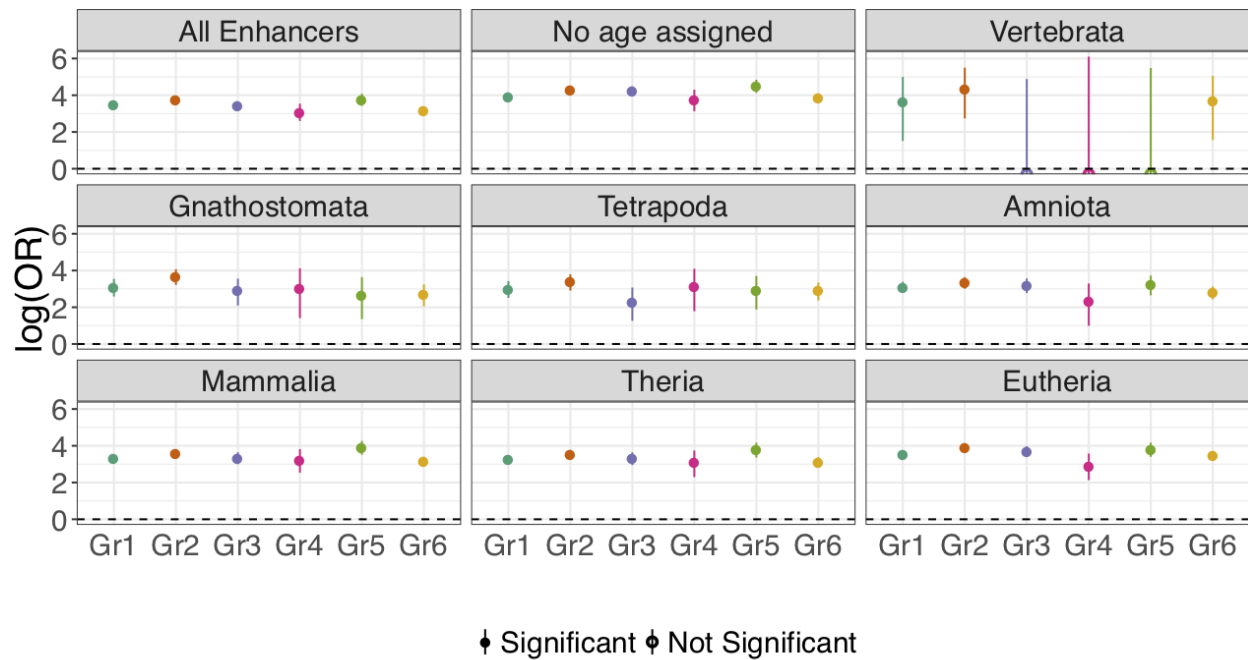




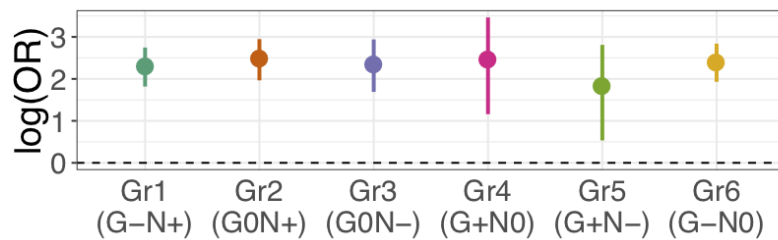
976 **Figure S7: Cell type-specific developmental DMRs and neuronal subtype methylation and**  
 977 **hydroxymethylation. (A) Percent of bases within each group of cdDMRs as defined in Figure**  
 978 **1B that are differentially methylated by neuronal subtypes in Luo *et al.*<sup>14</sup>. (B) Percent of cdDMRs**

979 within each of the 6 groups as defined in **Figure 1B** that overlap DMRs of methylcytosine (mC  
980 captured using oxidative bisulfite sequencing), hydroxymethylcytosine (hmC), or total cytosine  
981 methylation (tmC or mC+hmC, captured using standard bisulfite sequencing) as defined by  
982 Kozlenkov *et al.*<sup>50</sup>. The plot is stratified by whether the Kozlenkov *et al.* (2018) DMRs are more  
983 highly methylated in GABAergic or glutamatergic neurons, or in overall neurons compared to  
984 oligodendrocytes. The number of overlapping cdDMRs are listed above each bar.

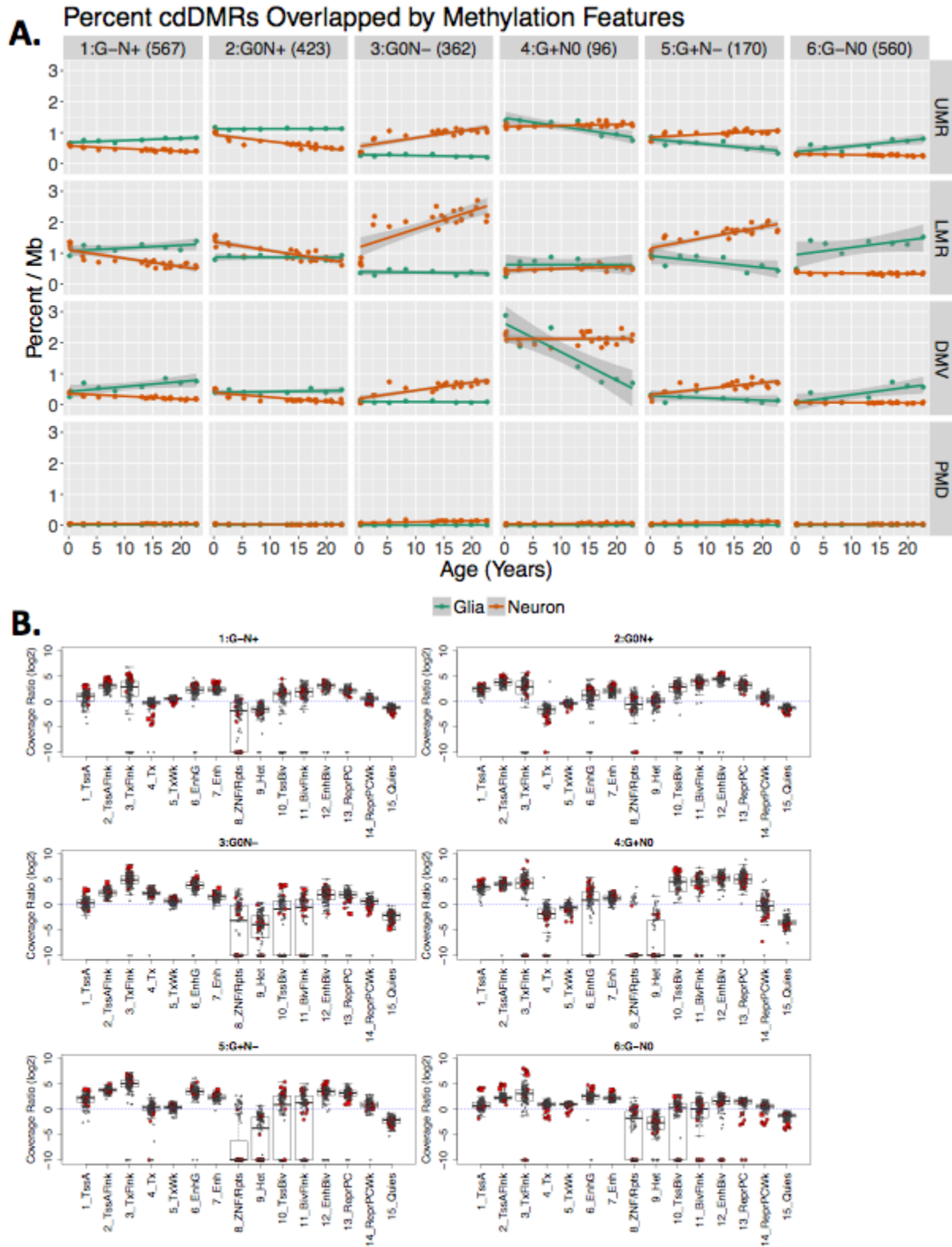
## A. Enrichment for Enhancers



## B. Enrichment for HARs

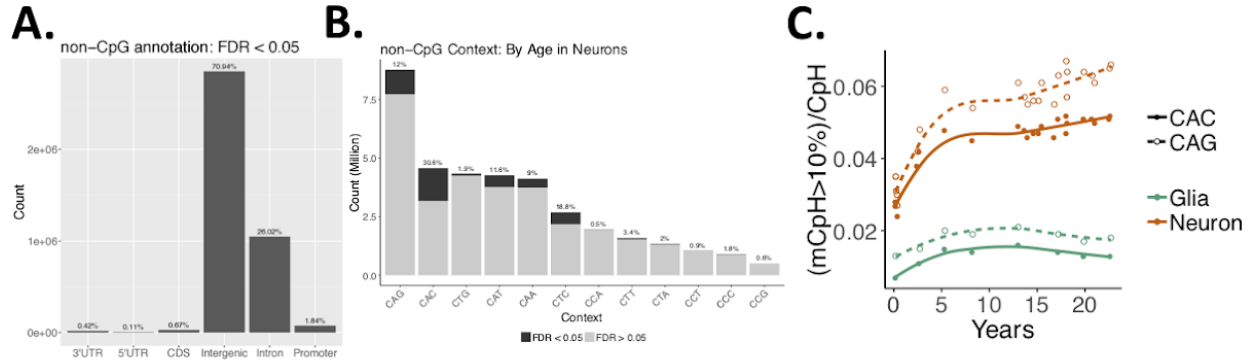


985 **Figure S8: cdDMR overlap of human brain developmental enhancers and Human**  
 986 **Accelerated Regions (HARs).** (A) Enrichment for enhancers in the six clusters of cdDMRs  
 987 from **Figure 1B** at the CpG-level. (B) Enrichment for human accelerated regions (HARs) in the  
 988 six clusters of cdDMRs from **Figure 1B** at the CpG-level. Filled in circles indicate FDR<0.05,  
 989 and error bars are the 95% confidence interval for the log(odds ratio).

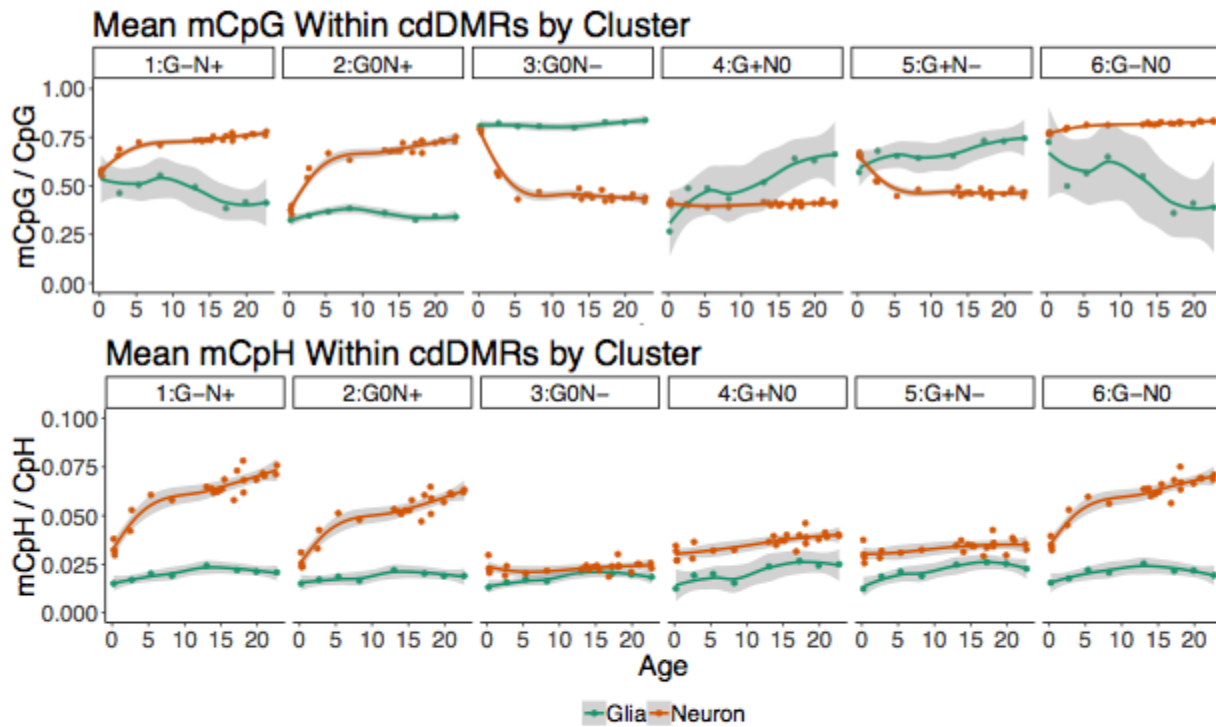


990 **Figure S9: Cell type-specific, developmentally dynamic DMRs (cdDMRs) and epigenetic**

991 **states. (A)** Percent of cdDMRs by the k-mean groups from **Figure 1B** overlapped by UMR,  
992 LMR, DMV and PMD sequence across age. Lines show linear regression and shading indicates  
993 the standard error. **(B)** Roadmap Epigenomics Consortium enriched chromatin states for the six  
994 clusters of cdDMRs from **Figure 1B** at the CpG-level.  $\log_2(\text{Coverage Ratio})$  represents the  
995 enrichment of the proportion of bases within each cdDMR group in a chromatin state compared  
996 to the rest of the genome.



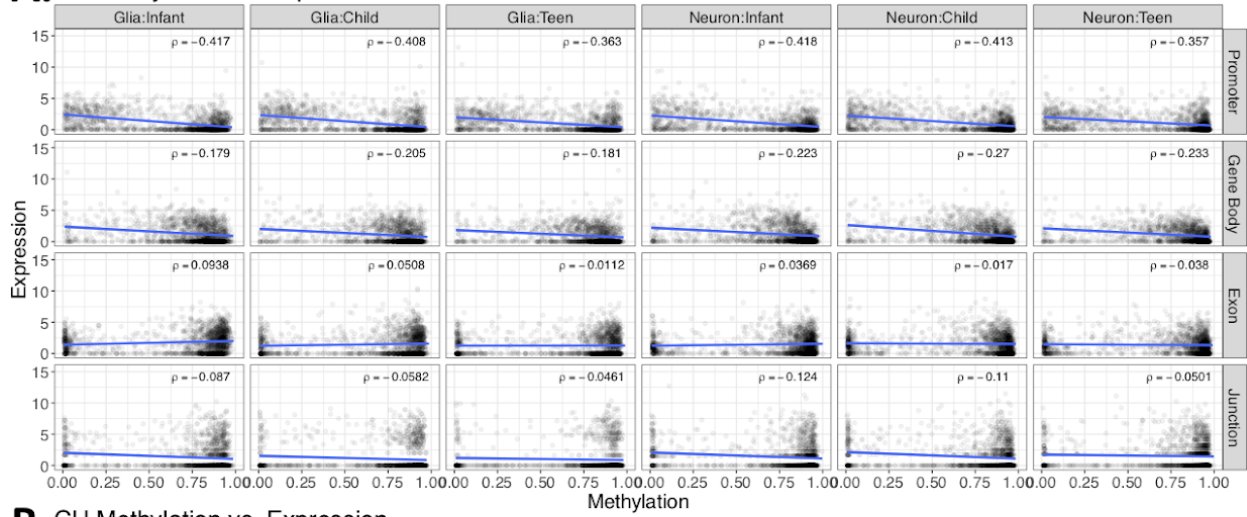
997 **Figure S10: CpH methylation distribution, levels and context-specific biological process**  
998 **ontology. (A)** Number of differentially methylated CpHs by cell type (FDR<5%) falling in  
999 different genomic annotations across the genome. Annotation was prioritized CDS > 5'UTR >  
1000 3'UTR > Intron > Promoter > Intergenic. **(B)** Breakdown of measured CpH by trinucleotide  
1001 context. Differentially methylated mCpH by age in neurons (FDR<0.05) are colored in dark gray.  
1002 **(C)** Accumulation of mCpH by trinucleotide context over development, stratified by cell type. **(D)**  
1003 The top 20 biological process ontology terms enriched for genes exclusively overlapping CAG  
1004 and CAC sites with significantly increasing or decreasing methylation levels in neurons over  
1005 postnatal development (FDR<0.05).



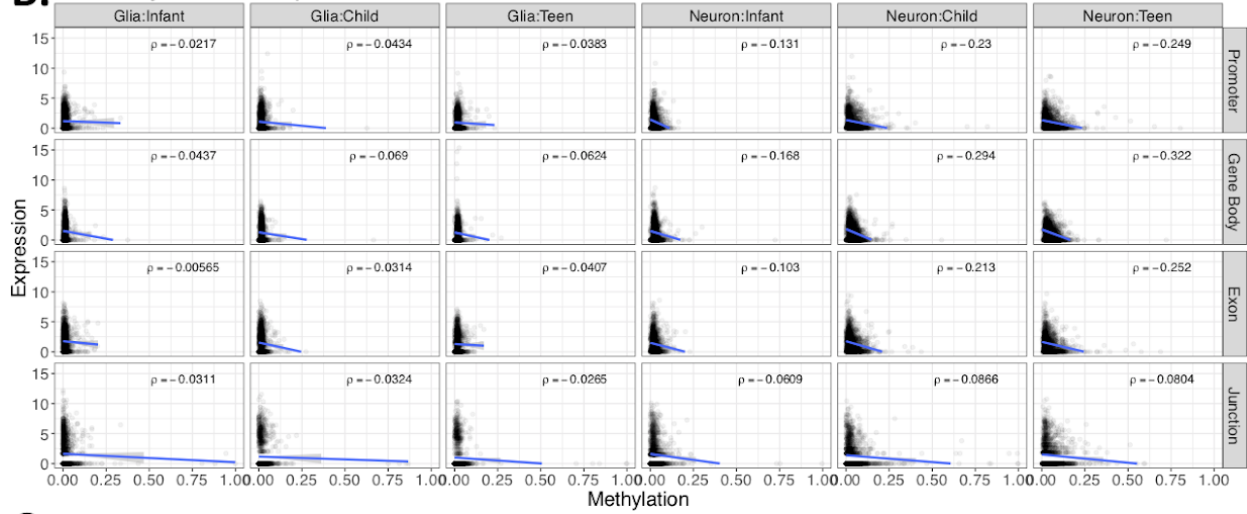
1006 **Figure S11: Trajectories of methylation accumulation in cdDMR groups.** Mean mCpH and  
1007 mCpG within the cdDMRs by cluster from **Figure 1B** across development. Loess line and  
1008 standard error shading are depicted.



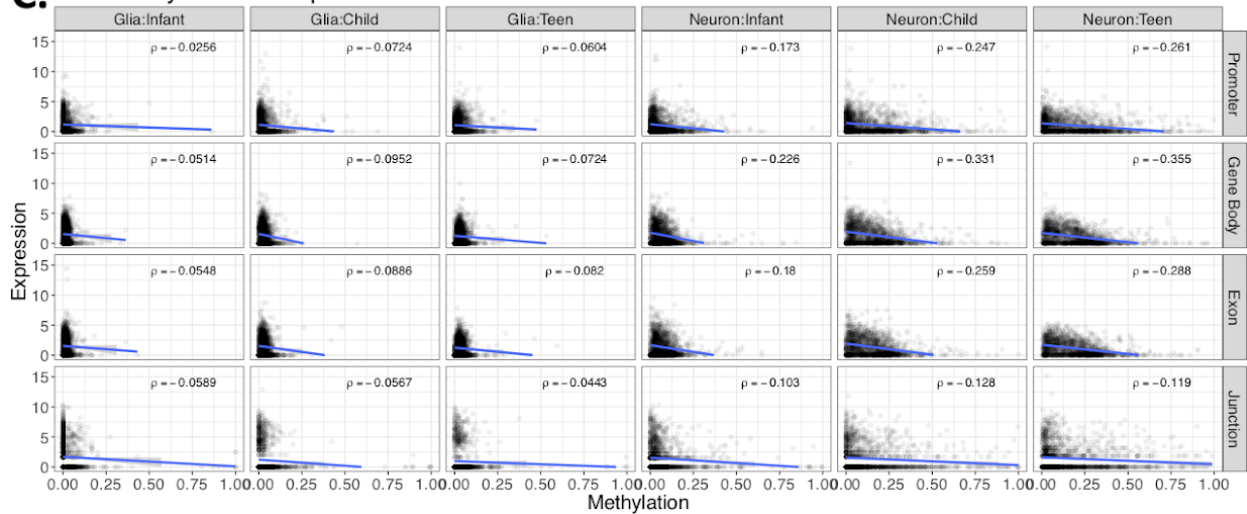
### A. CG Methylation vs. Expression



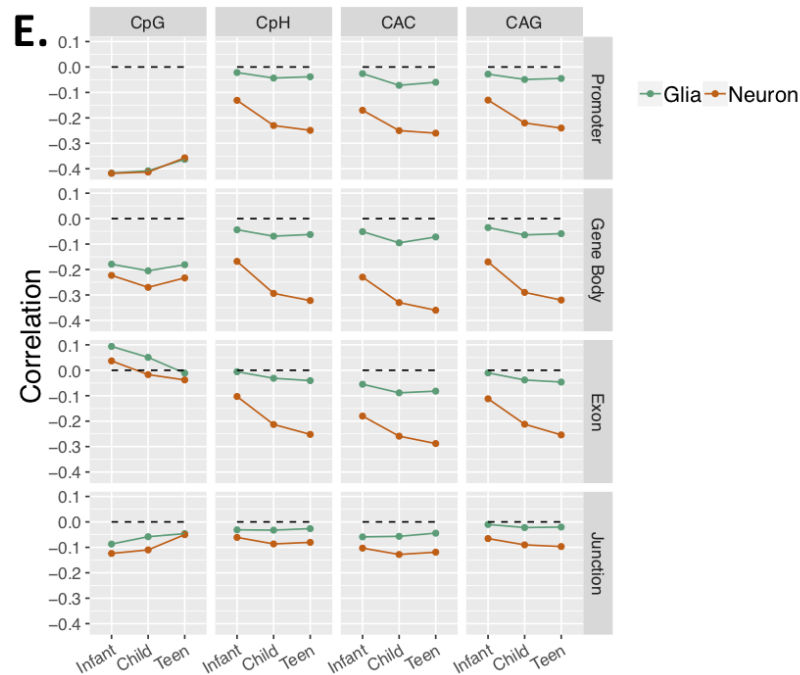
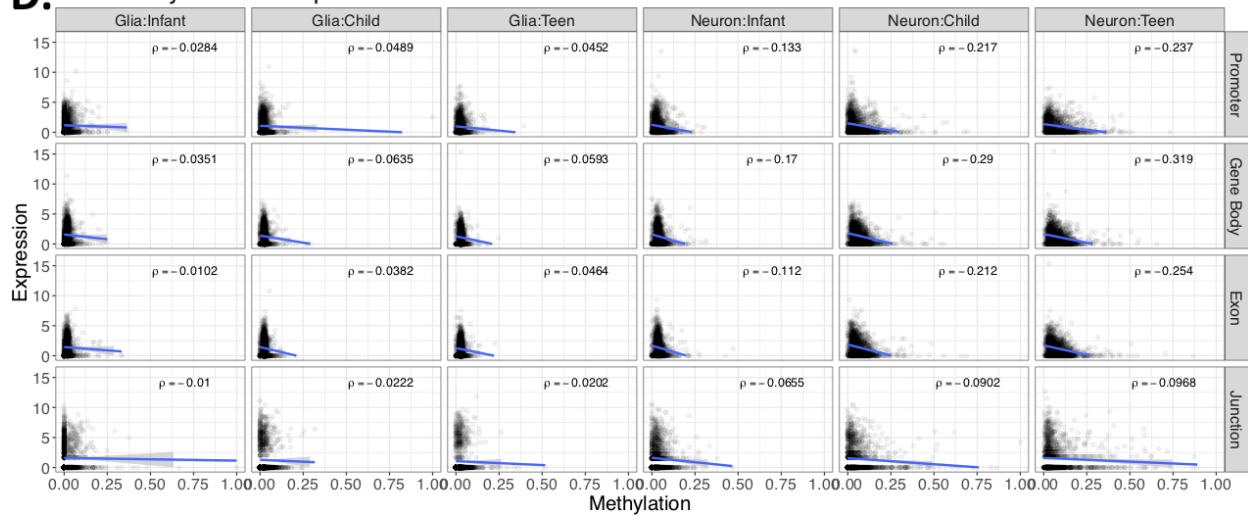
### B. CH Methylation vs. Expression



### C. CAC Methylation vs. Expression

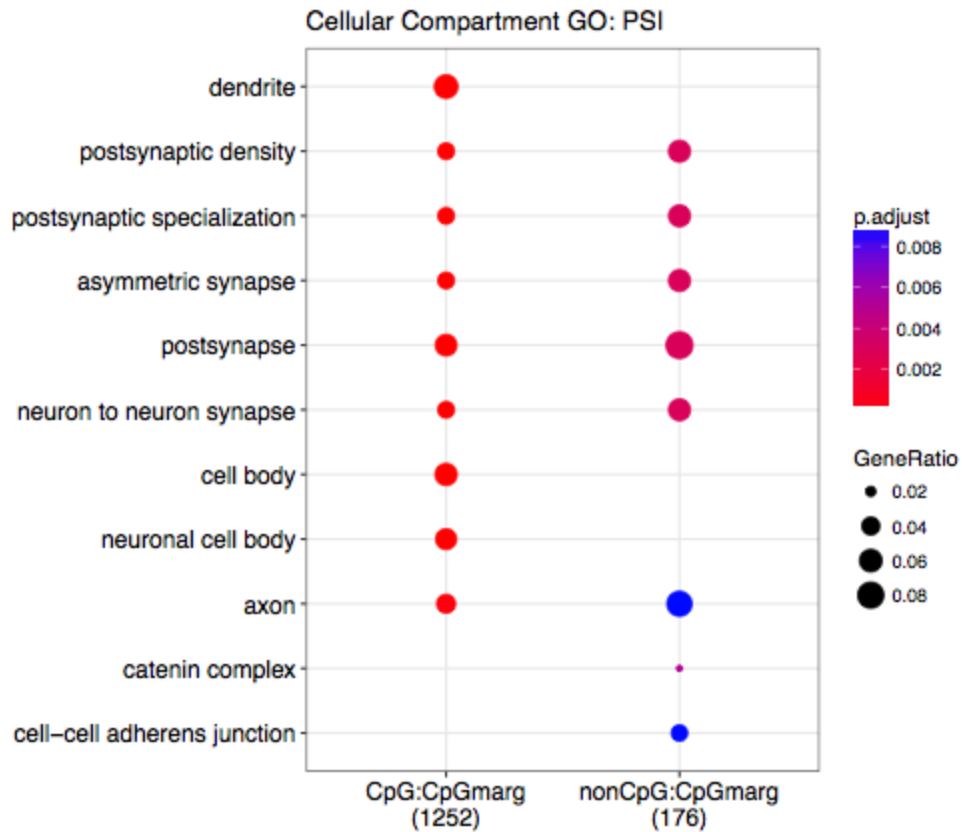


#### D. CAG Methylation vs. Expression



1009 **Figure S12: Relationship between methylation and expression. (A-D)** A random sample of  
 1010 10,000 expression feature:methylation pairs stratified by age and cell type in columns and  
 1011 feature type in rows. Each dot represents the mean methylation level of **(A)** CpGs, **(B)** CpHs,  
 1012 **(C)** CACs and **(D)** CAGs within the feature across all samples of that age and cell type on the  
 1013 x-axis and the  $\log_2([\text{mean Rpkm}] + 1)$  for the promoters, gene bodies, and exons and the

1014  $\log_2([\text{mean junction-overlapping reads per 10 million mapped reads}] + 1)$  for the junctions on the  
1015 y-axis. 10,000 pairs were chosen to reduce overplotting. Linear regression with shaded standard  
1016 error and the rho for each correlation is listed for each plot. **(E)** Correlation of feature expression  
1017 and methylation stratified by cytosine context (columns) and feature type (rows).



1018 **Figure S13: Cellular compartment ontology.** Genes containing splicing events as measured  
1019 by “percent spliced in” (PSI) that are associated with changing CpG and CpH methylation are  
1020 enriched for cellular compartment gene ontology terms relating to neuronal features.



1021 **Figure S14: Molecular function ontology.** Genes containing alternative exons that are  
 1022 associated with changing CpG and CpH methylation are enriched for molecular function  
 1023 ontology terms relating to neurons.

LIBD WGBS Expression explorer home methylation data methylation summary documentation

The following table includes all the CpG and nonCpG associations with nearby features at FDR 5%. The feature levels are gene, exon, splicing events that affect percent splicing in (PSI). Any filters applied here affect the results of the methylation summary tab.

Show 10 2 entries Search

id	feature	meth_type	feature_id	symbol	gene_type	meth_coefficient	meth_statistic	meth_FDR	n_samples_with_meth_not0	n_samples_with_m
1	gene	CpG	ENSG00000286891.1.2	RPL14AP19.9	none	-0.0561	-29.9	8.72e-13	22	
1.2	pep	CpG	ENSG00000128422.14.2	MTUR1	neuron	-0.00112	-26.3	7.87e-13	22	
1.2	pep	CpG	ENSG00000128422.14.2	MTUR1	neuron	0.00112	26.3	7.87e-13	22	
3.4	pep	CpG	ENSG00000196332.14.1	CD55	none	-0.0563	-29.4	7.54e-12	21	
3.4	pep	CpG	ENSG00000196332.14.1	CD55	none	0.0563	29.4	7.54e-12	21	
5	pep	CpG	ENSG00000171724.2.2	VAT1L	none	0.0185	22.5	1.52e-11	22	
6	pep	CpG	ENSG00000145882.11.2	FAR1B2	none	0.041	21.8	2.75e-11	22	
7	pep	CpG	ENSG00000145882.11.2	FAR1B2	none	0.0408	21.1	5.02e-11	22	
8	pep	CpG	ENSG00000145526.11.2	CDH18	glia	0.0089	20.7	6.86e-11	22	
8.10	pep	CpG	ENSG0000018825.14.1	ATP1A2	neuron	-0.0216	-20.1	1.21e-10	22	

Showing 1 to 10 of 81242 entries Previous 1 2 3 4 5 ... 8125 Next

Download methylation data

### Association details

Here you can explore in further detail a particular methylation and expression association. Choose the feature type, the methylation type (CpG or nonCpG) and select which association id (from column 1 in previous table) you want to explore.

You can also select a row in the table above and the options will be chosen for you automatically. The 1 chosen will be the first one in the row.

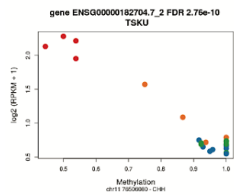
Feature:  Methylation type:  Result (association id from column 1 above):

Max is 29554

Export to UCSC Genome Browser

### Methylation vs expression association plot:

Scatter plot comparing methylation vs expression. Samples are colored by age group: Red: infant, orange: child, green: teen, blue: adult. Details in documentation tab.



A small amount of jitter has been added to the data to minimize overplotting.

Save image to a PDF file

### Expression feature information

```

@Changes object with 1 range and 18 metadata columns:
sequences      ranges strand | Length  geneID      ensemblID
<list>         <list> <list> | <integer> <character> <character>
ENSG00000182704.2 chr11 [74692295, 76580190] + 4131 ENSG00000182704.2 ENSG00000182704
gene_type      Symbol  EnsemblID | Class  nameScore  RankX
ENSG00000182704.2 protein_coding TSKU      25987      | Index  3,136063  9
-----
ENSG00000182704.2 [ENST0000033752.1,ENST00000312938.1,ENST00000333998.4,ENST00000325167.1,ENST00000327881.1]
-----
seqInfo: 36 sequences from an unspecified genome; no seqLengths

```

Download to an RStudio file

### Cytosine information

```

@Changes object with 1 range and 2 metadata columns:
sequences      ranges strand | C_context trinucleotide_context
<list>         <list> <list> | <list> <list>
[1] chr11 [70566808, 70566808] - | C=H CAC
-----
seqInfo: 25 sequences from an unspecified genome; no seqLengths

```

Download to an RStudio file

For more information check the documentation tab.

Bookmark

### Data license

The data in LIBD WGBS Expression explorer is licensed under CC BY 4.0. The legal text can be found here.

### Acknowledgements

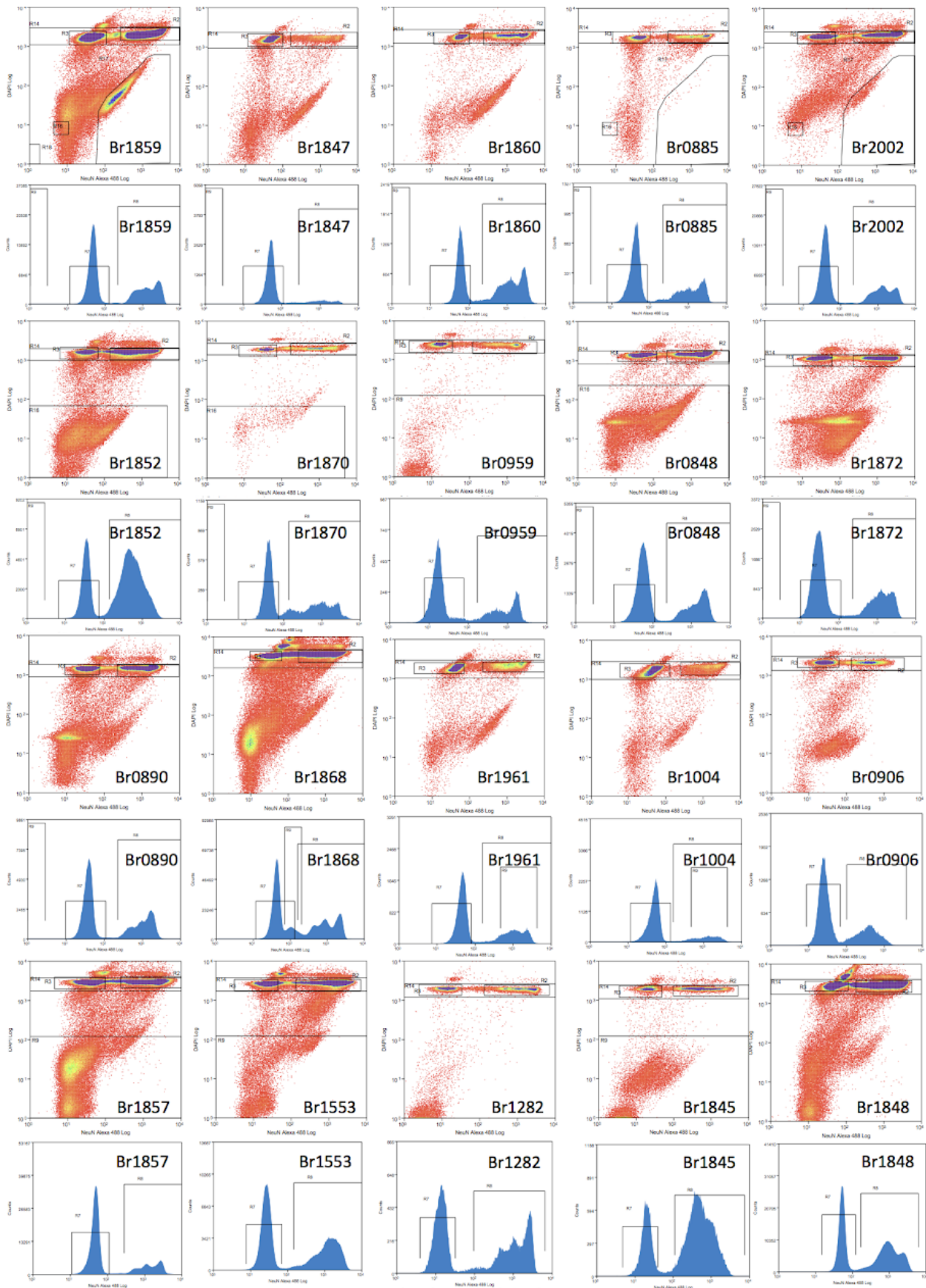
This research was supported by NIH R21MH102791-01A1 and the Lieber Institute for Brain Development. We thank the Department of Biostatistics at Johns Hopkins Bloomberg School of Public Health for hosting our application on their shinyapps account.

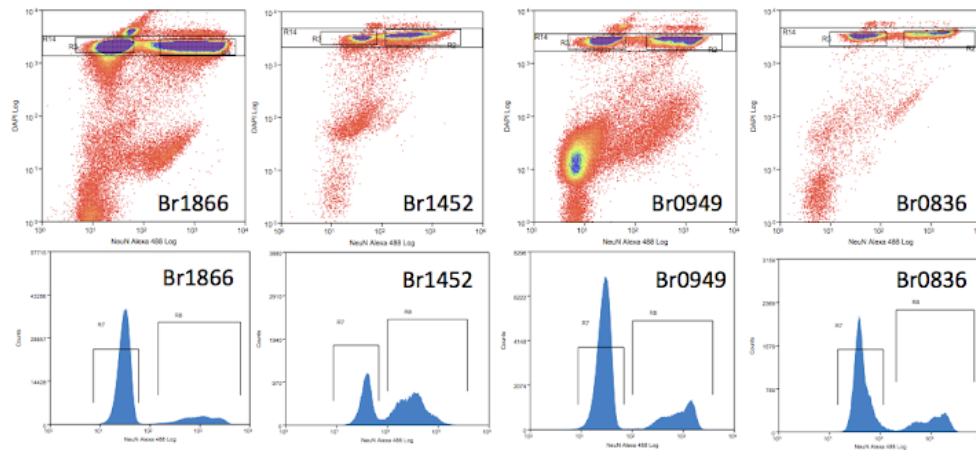


LIEBER INSTITUTE for  
BRAIN DEVELOPMENT  
MALTZ RESEARCH LABORATORIES



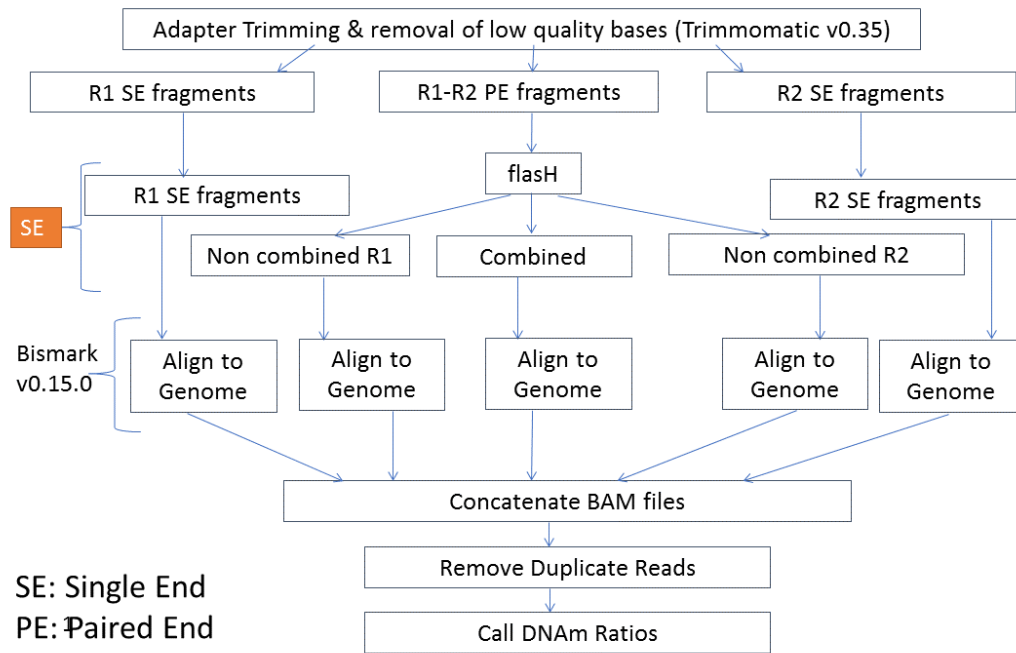
1024 **Figure S15: Web meQTL browser display.** Interactive display of the CpGs and CpGs  
 1025 associated with expression at FDR<5% as shown at  
 1026 <https://jhubiostatistics.shinyapps.io/wgbsExprs/>. This screenshot shows the top nonCpG  
 1027 (mCpH) meQTL association at the gene expression level. Information about the gene is shown  
 1028 under *expression feature*, and information of the methylated C is shown under *cytosine*.



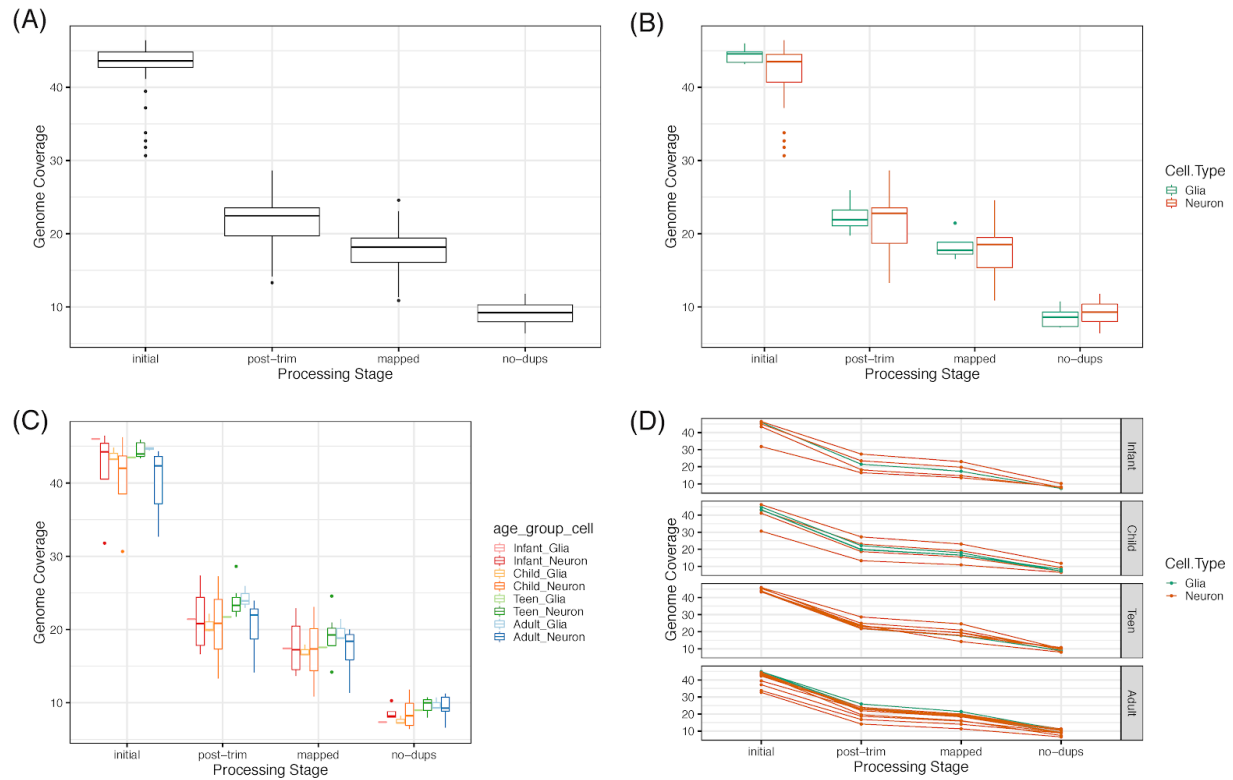


1029 **Figure S16: Raw sort data.** Raw data collected from the MoFlo Legacy (Beckman Coulter)  
1030 using Summit (version 4.3) software. The density plots show DAPI signal on the y-axis and  
1031 Alexa Fluor 488-NeuN signal on the x-axis for events filtered via gates 1-3 as depicted in Figure  
1032 S1A. The histograms show the distribution of Alexa Fluor 488-NeuN signal after including the  
1033 gate for singlets based on DAPI signal (gate R14 in the raw data).

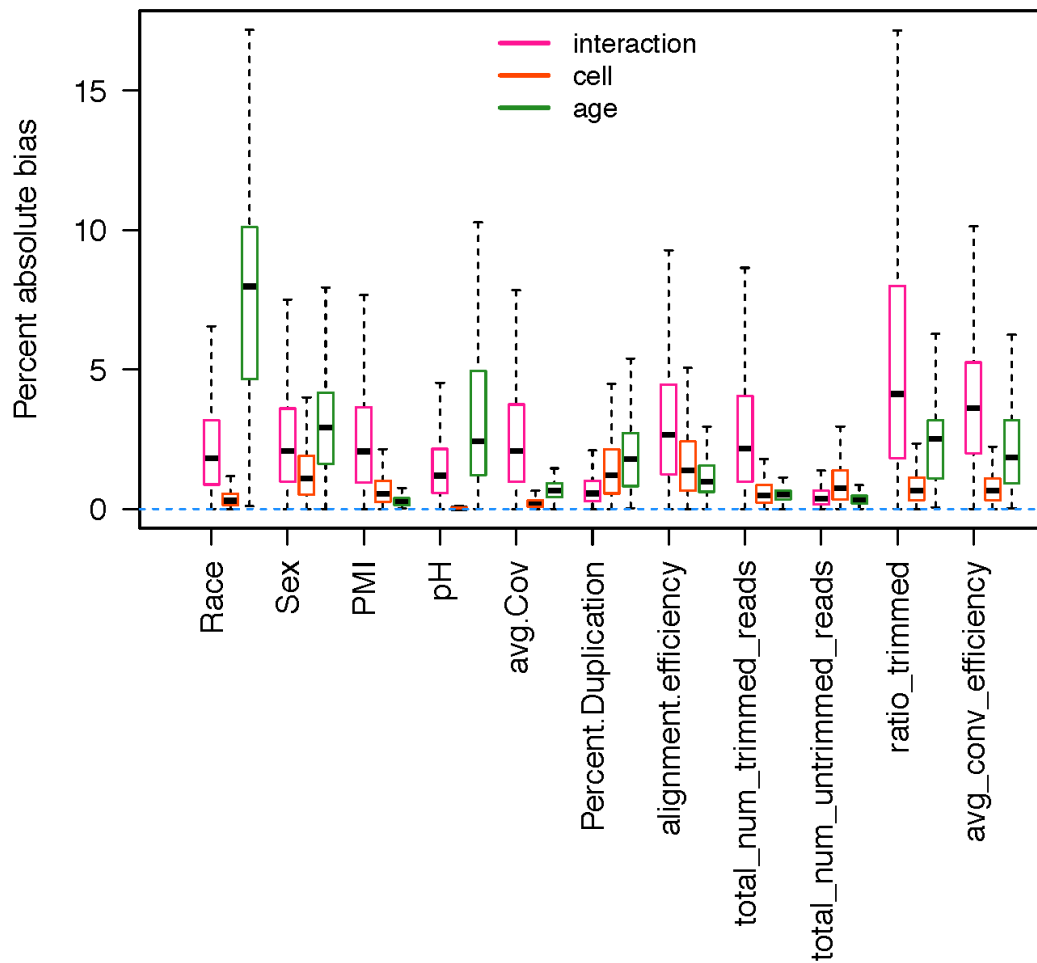




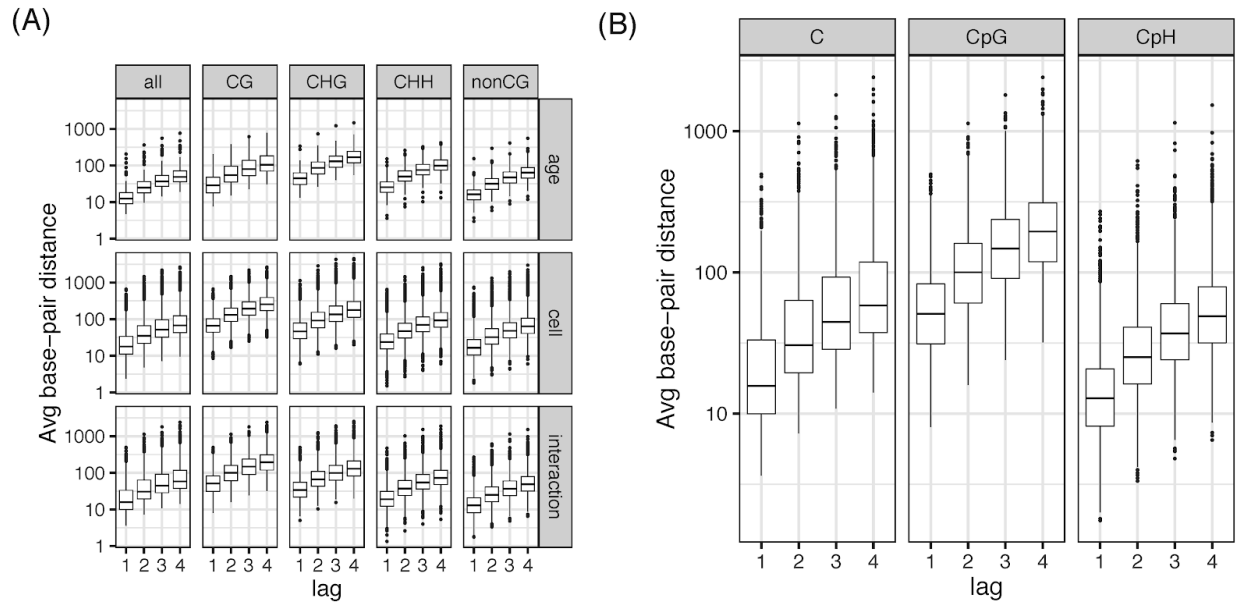
1034 **Figure S17: Data processing/alignment pipeline.** Overview of the processing steps taken to  
1035 prepare the WGBS data.



1036 **Figure S18: Genome coverage across processing stages. (A)** Genome coverage across the  
1037 four main processing stages: initial FASTQ files, after trimming, after mapping to the genome,  
1038 and after removing duplicated reads. **(B)** Same as (A) but with samples separated by cell type.  
1039 **(C)** Genome coverage across cell types and four age categories: infant, child, teen and adult.  
1040 **(D)** Genome coverage trajectories for each sample separated by age category and colored by  
1041 cell type.



1042 **Figure S19: DMR sensitivity analyses.** This figure shows the percent absolute bias for the  
1043 three coefficients of interest used to define the DMRs: age (adjusting for cell type), cell type  
1044 (adjusting for age), and the interaction between age and cell type. For each DMR type we  
1045 computed the coefficients after adjusting for each of the covariates in the X-axis ( $\beta_{adj}$ ) and  
1046 computed the percent absolute bias:  $100 * |\beta_{adj} - \beta_{original}| / |\beta_{original}|$ . All covariates considered  
1047 show less than 10% absolute bias. The covariates considered are: race, sex, PMI, pH, average  
1048 coverage, duplication percent, alignment efficiency, total number of trimmed reads, total number  
1049 of untrimmed reads, ratio of trimmed reads, and average conversion efficiency (lambda).



1050 **Figure S20: genome distance versus autocorrelation lag.** For each of the cytosines  
1051 considered for the autocorrelation analysis in **Figure 2B**, we computed the genome base-pair  
1052 distance and averaged it for each group of cytosines. **(A)** Average genomic distance for each of  
1053 the three DMR models and for various contexts of cytosines. **(B)** Average genome distance for  
1054 the interaction DMRs (cdDMRs) for all cytosines, CpG and CpH. The genome distance is  
1055 proportional to the lag. This panel is the exact complement to **Figure 2B**.

1056 **Supplementary Tables**

1057 **Table S1: Phenotype and sequencing metrics for the WGBS samples.**

1058 *See sheet 1 of Supplementary\_tables.xlsx.*

1059 **Table S2: Phenotype and sequencing metrics for the RNA-seq samples.**

1060 *See sheet 2 of Supplementary\_tables.xlsx.*

1061 **Table S3: Number of cytosines measured and distribution of methylation by cytosine**

1062 **context.** Sample sizes for each sample type, number of CpGs and CpHs and distribution of

1063 CpG and CpHs across low, partial and high levels of methylation.

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

1064 **Table S4: Cell type-specific, developmental differentially methylated regions (cdDMRs).**

1065 *See sheet 4 of Supplementary\_tables.xlsx.*

1066 **Table S5: mC association with cell type and age in postnatal cell type-specific samples.**

1067 Number of differentially methylated CpGs and CpHs across each of the statistical models: cell  
1068 type effects adjusting for age, age effects adjust for cell type, cell type and age interaction  
1069 effects, as well as age in neuron samples only.

Model	CpG Context		CpH Context	
	FDR $\leq$ 0.05	% Significant	FDR $\leq$ 0.05	% Significant
<b>Cell type</b>	4824804	25.80%	7682075	18.80%
<b>Age</b>	536164	2.90%	3194618	7.80%
<b>Cell type:Age</b>	90227	0.50%	76	0%
<b>Age in Neurons</b>	NA	NA	4020371	9.80%

1070 **Table S6: GWAS traits assessed using LDSC.**

1071 *See sheet 6 of Supplementary\_tables.xlsx.*

1072 **Table S7: Stratified linkage disequilibrium score regression results.**

1073 *See sheet 7 of Supplementary\_tables.xlsx.*

1074 **Table S8: Enrichment of DMRs and mCpH for disease-associated gene sets.**

1075 *See sheet 8 of Supplementary\_tables.xlsx.*

1076 **Table S9: Enrichment for disease gene sets in DNAm-splicing association features.**

1077 *See sheet 9 of Supplementary\_tables.xlsx.*

1078 **Table S10: Variable dictionary for Table 1.** Variables used in **Table 1** with their full definition.

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 $\geq 5\%$ .
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.

1079 **Bibliography**

- 1080 1. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development.  
1081 *Science* **341**, 1237905 (2013).
- 1082 2. Jaffe, A. E. *et al.* Mapping DNA methylation across development, genotype and  
1083 schizophrenia in the human frontal cortex. *Nat. Neurosci.* **19**, 40–47 (2016).
- 1084 3. Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in  
1085 epigenome-wide association studies. *Genome Biol.* **15**, R31 (2014).
- 1086 4. Schultz, M. D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation  
1087 variation. *Nature* **523**, 212–216 (2015).
- 1088 5. Mo, A. *et al.* Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron*  
1089 **86**, 1369–1384 (2015).
- 1090 6. Jeong, M. *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a.  
1091 *Nat. Genet.* **46**, 17–23 (2014).
- 1092 7. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem  
1093 cells. *Cell* **153**, 1134–1148 (2013).
- 1094 8. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory  
1095 regions. *Nature* **480**, 490–495 (2011).
- 1096 9. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic  
1097 differences. *Nature* **462**, 315–322 (2009).
- 1098 10. Salhab, A. *et al.* Partially methylated domains are hallmarks of a cell specific epigenome  
1099 topology. *BioRxiv* (2018). doi:10.1101/249334
- 1100 11. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA  
1101 methylation sequencing. *Nature* **510**, 537–541 (2014).
- 1102 12. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human



- 1103 epigenomes. *Nature* **518**, 317–330 (2015).
- 1104 13. Zhou, W. *et al.* DNA methylation loss in late-replicating domains is linked to mitotic cell  
1105 division. *Nat. Genet.* **50**, 591–602 (2018).
- 1106 14. Luo, C. *et al.* Single-cell methylomes identify neuronal subtypes and regulatory elements in  
1107 mammalian cortex. *Science* **357**, 600–604 (2017).
- 1108 15. Emera, D., Yin, J., Reilly, S. K., Gockley, J. & Noonan, J. P. Origin and evolution of  
1109 developmental enhancers in the mammalian neocortex. *Proc Natl Acad Sci USA* **113**,  
1110 E2617-26 (2016).
- 1111 16. Doan, R. N. *et al.* Mutations in human accelerated regions disrupt cognition and social  
1112 behavior. *Cell* **167**, 341–354.e12 (2016).
- 1113 17. He, Y. & Ecker, J. R. Non-CG Methylation in the Human Genome. *Annu. Rev. Genomics*  
1114 *Hum. Genet.* **16**, 55–77 (2015).
- 1115 18. Eckhardt, F. *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat.*  
1116 *Genet.* **38**, 1378–1385 (2006).
- 1117 19. Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to  
1118 splicing. *Nature* **479**, 74–79 (2011).
- 1119 20. Rizzardi, L. F. *et al.* Neuronal brain-region-specific DNA methylation and chromatin  
1120 accessibility are associated with neuropsychiatric trait heritability. *Nat. Neurosci.* **22**,  
1121 307–316 (2019).
- 1122 21. Brainstorm Consortium *et al.* Analysis of shared heritability in common disorders of the  
1123 brain. *Science* **360**, (2018).
- 1124 22. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide  
1125 association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- 1126 23. Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat.*

- 1127        *Genet.* **50**, 825–833 (2018).
- 1128    24. Birnbaum, R., Jaffe, A. E., Hyde, T. M., Kleinman, J. E. & Weinberger, D. R. Prenatal  
1129        expression patterns of genes associated with neuropsychiatric disorders. *Am. J. Psychiatry*  
1130        **171**, 758–767 (2014).
- 1131    25. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights  
1132        from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
- 1133    26. Howell, K. R., Floyd, K. & Law, A. J. PKBy/AKT3 loss-of-function causes learning and  
1134        memory deficits and deregulation of AKT/mTORC2 signaling: Relevance for schizophrenia.  
1135        *PLoS ONE* **12**, e0175993 (2017).
- 1136    27. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the  
1137        human brain. *Science* **362**, (2018).
- 1138    28. Stiles, J. & Jernigan, T. L. The basics of brain development. *Neuropsychol. Rev.* **20**,  
1139        327–348 (2010).
- 1140    29. Tierney, A. L. & Nelson, C. A. Brain development and the role of experience in the early  
1141        years. *Zero Three* **30**, 9–13 (2009).
- 1142    30. Stroud, H. *et al.* Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic  
1143        States. *Cell* **171**, 1151–1164.e16 (2017).
- 1144    31. Lev Maor, G., Yearim, A. & Ast, G. The alternative role of DNA methylation in splicing  
1145        regulation. *Trends Genet.* **31**, 274–280 (2015).
- 1146    32. Jaffe, A. E. *et al.* Developmental and genetic regulation of the human cortex transcriptome  
1147        illuminate schizophrenia pathogenesis. *Nat. Neurosci.* **21**, 1117–1125 (2018).
- 1148    33. Kozlenkov, A. *et al.* Substantial DNA methylation differences between two major neuronal  
1149        subtypes in human brain. *Nucleic Acids Res.* **44**, 2593–2612 (2016).
- 1150    34. Sahara, S., Yanagawa, Y., O’Leary, D. D. M. & Stevens, C. F. The fraction of cortical

- 1151 GABAergic neurons is constant from near the start of cortical neurogenesis to adulthood. *J.*  
1152 *Neurosci.* **32**, 4755–4761 (2012).
- 1153 35. Kozlenkov, A. *et al.* A unique role for DNA (hydroxy)methylation in epigenetic regulation of  
1154 human inhibitory neurons. *Sci. Adv.* **4**, eaau6190 (2018).
- 1155 36. Lipska, B. K. *et al.* Critical factors in gene expression in postmortem human brain: Focus on  
1156 studies in schizophrenia. *Biol. Psychiatry* **60**, 650–658 (2006).
- 1157 37. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for  
1158 Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
- 1159 38. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence  
1160 data. *Bioinformatics* **30**, 2114–2120 (2014).
- 1161 39. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve  
1162 genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
- 1163 40. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**,  
1164 357–359 (2012).
- 1165 41. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,  
1166 2078–2079 (2009).
- 1167 42. Hansen, K. D., Langmead, B. & Irizarry, R. A. BSmooth: from whole genome bisulfite  
1168 sequencing reads to differentially methylated regions. *Genome Biol.* **13**, R83 (2012).
- 1169 43. Jaffe, A. E. *et al.* Bump hunting to identify differentially methylated regions in epigenetic  
1170 epidemiology studies. *Int. J. Epidemiol.* **41**, 200–209 (2012).
- 1171 44. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and  
1172 microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
- 1173 45. Morgan, M. *AnnotationHub: Client to access AnnotationHub resources.* (Bioconductor,  
1174 2019). at <<http://bioconductor.org/packages/release/bioc/html/AnnotationHub.html>>

- 1175 46. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory  
1176 requirements. *Nat. Methods* **12**, 357–360 (2015).
- 1177 47. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for  
1178 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
- 1179 48. Goldstein, L. D. *et al.* Prediction and Quantification of Splice Events from RNA-Seq Data.  
1180 *PLoS ONE* **11**, e0156132 (2016).
- 1181 49. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for  
1182 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140  
1183 (2010).
- 1184 50. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model  
1185 analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
- 1186 51. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing  
1187 biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
- 1188 52. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations.  
1189 *Bioinformatics* **28**, 1353–1358 (2012).