Rizzardi et al. Genome Biology (2021) 22:116
https://doi.org/10.1186/s13059-021-02335-w

Genome Biology

RESEARCH Open Access

Human brain region-specific variably methylated regions are enriched for heritability of distinct neuropsychiatric traits

Lindsay F. Rizzardi[1,2], Peter F. Hickey[3,4,5], Adrian Idrizi[1], Rakel Tryggvadóttir[1], Colin M. Callahan[1], Kimberly E. Stephens[1,6,7], Sean D. Taverna[1,8], Hao Zhang[9], Sinan Ramazanoglu[1], GTEx Consortium, Kasper D. Hansen[1,3,10*] and Andrew P. Feinberg[1,11,12*]

* Correspondence: khansen@jhsph.edu; afeinberg@jhu.edu
[1]Center for Epigenetics, Johns Hopkins University School of Medicine, 855 N. Wolfe St., Baltimore, MD 21205, USA
Full list of author information is available at the end of the article

Abstract

Background: DNA methylation dynamics in the brain are associated with normal development and neuropsychiatric disease and differ across functionally distinct brain regions. Previous studies of genome-wide methylation differences among human brain regions focus on limited numbers of individuals and one to two brain regions.

Results: Using GTEx samples, we generate a resource of DNA methylation in purified neuronal nuclei from 8 brain regions as well as lung and thyroid tissues from 12 to 23 donors. We identify differentially methylated regions between brain regions among neuronal nuclei in both CpG (181,146) and non-CpG (264,868) contexts, few of which were unique to a single pairwise comparison. This significantly expands the knowledge of differential methylation across the brain by 10-fold. In addition, we present the first differential methylation analysis among neuronal nuclei from basal ganglia tissues and identify unique CpG differentially methylated regions, many associated with ion transport. We also identify 81,130 regions of variably CpG methylated regions, i.e., variable methylation among individuals in the same brain region, which are enriched in regulatory regions and in CpG differentially methylated regions. Many variably methylated regions are unique to a specific brain region, with only 202 common across all brain regions, as well as lung and thyroid. Variably methylated regions identified in the amygdala, anterior cingulate cortex, and hippocampus are enriched for heritability of schizophrenia.

Conclusions: These data suggest that epigenetic variation in these particular human brain regions could be associated with the risk for this neuropsychiatric disorder.

Keywords: Differentially methylated regions, Variably methylated regions, Neuropsychiatric disease, GTEx

© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

DNA methylation patterns are altered throughout development to establish distinct cell fates [1–3]. Analyses of methylation changes during cortical development have shown that, in neurons, these changes are often associated with synaptogenesis during the first 5 years of life [4]. These regions of dynamic methylation have been linked to neuropsychiatric disorders that are thought to have developmental origins [5–8]. In adulthood, the epigenome is intimately involved in neuronal plasticity in response to environmental exposures and synaptic activity [9, 10]. DNA methylation and chromatin features differ among neurons from functionally distinct adult brain regions and these regions are enriched for heritability of neuropsychiatric traits [11, 12]. The PsychENCODE consortium has made substantial contributions to our understanding of gene regulation in the human brain, particularly during development and disease [13, 14], and in distinct cell populations within the cortex [15]. However, these efforts have been somewhat limited either by examination of multiple cell populations within a single cortical region or by the use of bulk tissues that are strongly confounded by cellular heterogeneity. Thus, the extent of DNA methylation variation within neuronal populations from many adult human brain regions remains unknown and is the focus of this work.

The Genotype-Tissue Expression (GTEx) project [16–19] has enabled unprecedented analysis and understanding of tissue-specific expression and the genetic determinants of this expression. To complement the expression data generated by GTEx, the enhanced GTEx project [20] has profiled additional molecular traits across a subset of GTEx samples. We previously explored DNA methylation across four brain regions using non-GTEx samples from six individuals [12]. Here, we describe the results of the enhanced GTEx DNA methylation project that examined 182 samples representing 8 brain regions and 2 somatic tissues from 12 to 24 GTEx donors with extensive profiling of DNA methylation and the influence of genotype on methylation variability. Moreover, based on our earlier work, purification of neuronal rather than bulk tissue DNA prior to analysis was imperative to obtain high-quality methylation data due to the extensive neuron/glia heterogeneity even in adjacent sections from the same brain region. In contrast, GTEx expression data was obtained exclusively from bulk tissue samples confounded by cellular heterogeneity; therefore, we focused on how genotype influenced methylation rather than how methylation impacted gene expression. We identify differentially methylated regions among NeuN+ cells from 8 brain regions in both a CpG and non-CpG context, greatly extending our knowledge of functional epigenetic differences across the brain. Importantly, the larger number of brain regions and individuals enabled our identification of variably methylated regions (VMRs) that we initially defined as regions that are highly variable among individuals within a given tissue type [21, 22]. Surprisingly, many VMRs were brain region-specific, 60–95% involved two or more regions and were concordant for methylation, while only 202 were ubiquitous across all 8 brain regions and 2 somatic tissues. Along with CpG differentially methylated regions (CG-DMRs), VMRs from three brain regions were enriched for heritability of schizophrenia suggesting the importance of epigenetic variation in neuropsychiatric disease risk.

Results

Characterizing the neuronal DNA methylation landscape across 8 brain regions

Previous reports from us and others have shown distinct epigenetic landscapes among functionally diverse brain regions [11, 12]. Our previous study was limited to a small number of individuals and brain regions. However, we demonstrated that brain region-specific DNA methylation was primarily present in neuronal rather than non-neuronal nuclei. Further, the ratio of neurons to non-neurons even between adjacent sections from the same brain region differed greatly, severely confounding analysis of differential methylation in non-purified nuclei. Therefore, we applied a strategy of neuronal nuclei purification prior to whole genome bisulfite sequencing. Analyzing a much larger number of individuals and brain regions enabled us to address the potential existence of VMRs (regions of interindividual variation in methylation within a tissue), their relationship to each other, and the relationship to SNPs identified in these GTEx donors. Neuronal nuclei were isolated from brain tissues based on positive NeuN (*RBFOX3*) staining via fluorescence-activated nuclei sorting and NeuN+ nuclei are referred to as neuronal, while noting that this fraction is composed of multiple subpopulations (Additional file 1: Fig. S1a). We examined 8 brain regions collected from GTEx donors: amygdala ($n = 12$), anterior cingulate cortex (BA24) ($n = 15$), caudate ($n = 22$), frontal cortex (BA9) ($n = 24$), hippocampus ($n = 20$), hypothalamus ($n = 13$), nucleus accumbens ($n = 23$), and putamen ($n = 16$). In addition, we analyzed methylation of DNA isolated from two non-brain tissues from GTEx donors: lung ($n = 18$) and thyroid ($n = 19$) for a total of 182 samples (Additional file 2: Table S1). We generated > 30 billion uniquely mapped 150-bp paired-end reads with an average depth > 10X post-processing (Additional file 3: Table S2). Several samples were excluded due to genotype discordance, as the shipped sample genotype did not match the biobank records (Additional file 4: Table S3). Five additional samples were excluded after principal component analysis revealed sample mislabeling prior to receipt by our lab. We confirmed this by determining which tissues most closely matched the methylation of these samples (Additional file 1: Fig. S1b).

Principal component analysis of global neuronal DNA methylation levels revealed clear segregation of these brain regions in the first two principal components (Fig. 1a). We performed a differential analysis of CpG methylation identifying CG-DMRs, i.e., regions of differential CpG methylation among neuronal nuclei isolated from each brain region. Given that a single CG-DMR can represent a difference among multiple brain regions, rather than perform 28 pairwise comparisons, we used an *F*-test to identify 174,482 statistically significant autosomal neuronal CG-DMRs which are defined as regions of the genome where at least 2 of the 8 brain regions have different levels of CpG methylation (Additional file 5: Table S4). We control the family-wise error rate at 5% by permutation, and we use BSmooth to leverage information from nearby CpGs by smoothing. In a pilot study, we profiled NeuN+ cells from 4 brain regions using whole genome bisulfite sequencing on samples from 6 different individuals not part of GTEx [12]. We find that 99.5% of our previously identified neuronal CG-DMRs (13,019/13,074) overlap with CG-DMRs from our new analysis of GTEx samples (after correcting for multiple testing). To make a more precise comparison, we examined the correlation between the methylation differences between two tissues as measured separately in Rizzardi et al. [12] and this study (tissues were the nucleus accumbens and prefrontal

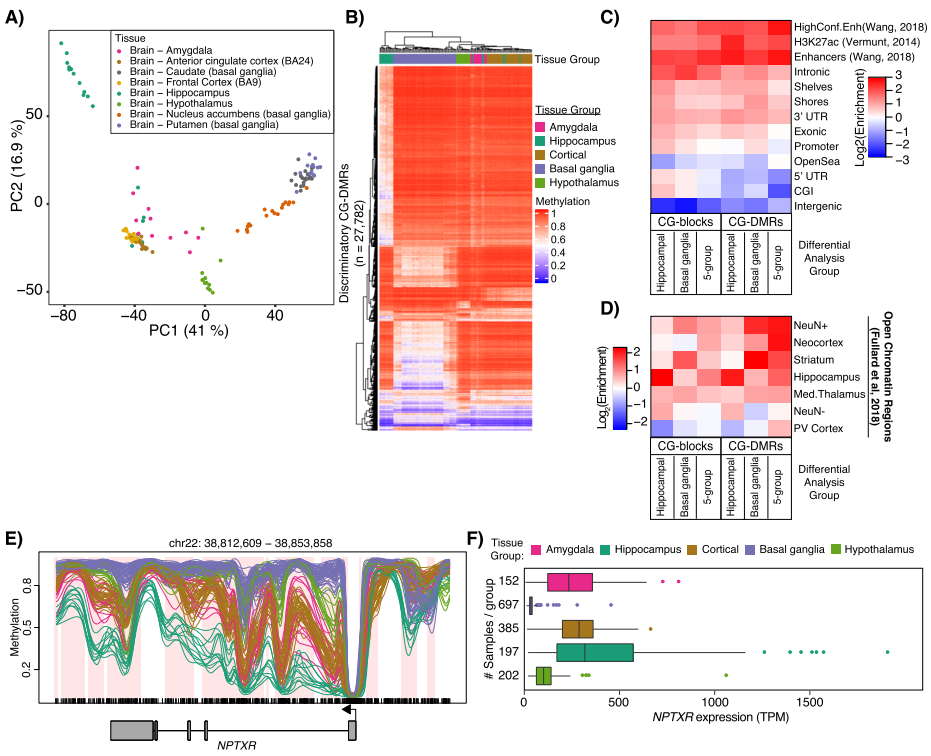

Fig. 1 Identification of CG-DMRs among neurons of functionally distinct brain regions. DNA methylation was assessed in neuronal nuclei isolated from 8 brain regions as indicated from 12 to 24 individuals. **a** Principal component analysis of distances derived from average autosomal CpG methylation in 1-kb bins. **b** Hierarchical clustering of samples based on the average methylation per sample in the most discriminatory CG-DMRs (see the "Methods" section). **c** Heatmap representing log2 enrichment of CpGs within CG-DMRs and blocks identified in each CG-DMR analysis compared to the rest of the genome for genomic features. Gene models from GENCODEv26 (promoters, intronic, exonic, 5′UTR, 3′UTR, intergenic), CpG islands (CGIs) and related features from UCSC (shores, shelves, OpenSea), putative enhancer regions (enhancers and high confidence enhancers from PsychENCODE [18] and H3K27ac [19]). 5-group = CG-DMRs or blocks between all 5 tissue groups. **d** As in **c**, showing enrichment in regions of open chromatin in NeuN− and NeuN+ nuclei and NeuN+ nuclei isolated from the indicated brain regions (PV cortex, primary visual cortex; Med. thalamus, mediodorsal thalamus) from [10]. **e** Example CG-DMRs covering the *NPTXR* gene showing average methylation values for NeuN+ nuclei from each tissue group color coded as in **b**. Regions of differential methylation are shaded in pink. **f** Expression of *NPTXR* from sample matched bulk brain tissues from GTEx v8 data release

cortex, selected because most of the CG-DMRs from Rizzardi et al. [12] were between those two tissues). We find a striking correlation of 0.97 as shown in Figure S1c highlighting the reproducibility of our experimental and analytical approaches across biobanks. This high level of reproducibility holds even when examining methylation differences among all 13,074 neuronal DMRs identified in [12] between two very similar cortical regions (Additional file 1: Figure S1d), which suggests that our approach is conservative, likely because we control the family-wise error rate.

To facilitate interpretation of our data, we conducted a simpler analysis. Specifically, we collapsed frontal cortex and anterior cingulate cortex samples into a "cortical" group and caudate, putamen, and nucleus accumbens into a "basal ganglia" group. The resulting 5 tissue groups are consistent with the developmental origins of the brain regions; the telencephalon gives rise to the cerebral cortex (which branches into the frontal cortex, the anterior cingulate cortex, and the hippocampal formation) and

cerebral nuclei (which branches into the amygdala and basal ganglia) while the diencephalon produces the hypothalamus [23]. We identified 181,146 autosomal neuronal CG-DMRs (196 Mb) among these 5 groups covering 11% of all CpGs. Further, the 5-group analysis captured 94% of the CG-DMRs identified in the 8-group analysis (Additional file 6: Table S5). Average DNA methylation levels of the most discriminatory CG-DMRs are, aside from several hippocampus samples, able to segregate samples into their tissue groups (Fig. 1b). We also identified 7671 large regions of differential CpG methylation (which we have previously termed "blocks" of differential CpG methylation; these are identified using a larger bandwidth for smoothing) among the 5 tissue groups (Additional file 7: Table S6). These CG-blocks covered 260 Mb and were on average 33.9 kb in size. CG-DMRs were enriched in enhancer regions identified by PsychENCODE [13], H3K27ac peaks found in the adult brain [24], and in regulatory chromHMM states from 4 brain regions [25] (Fig. 1c, Additional file 1: Fig. S1e). We also observed enrichment of our CG-DMRs in regions of open chromatin identified in NeuN+ nuclei from 14 brain regions [11] (Fig. 1d). Example CG-DMRs within the neuronal pentraxin 1 (NP1) gene (*NPTXR*) are shown (Fig. 1e) with hypomethylation in the hippocampal neurons associated with increased expression in bulk hippocampus tissue (Fig. 1f). NP1 is involved in glutamate receptor internalization and has been implicated in Alzheimer's disease as its upregulation in response to increased amyloid-beta promotes neuronal toxicity [26].

Though we grouped them together in our initial CG-DMR analysis, there is a clear distinction among the basal ganglia tissues. These regions are of particular interest due to their importance in addiction and reward pathways [27], yet no comprehensive analysis of methylation differences in the human brain has been performed to date. We performed an additional DMR analysis to assess the methylation differences among neurons from these tissues and identified 16,866 autosomal neuronal basal ganglia CG-DMRs (24 Mb) encompassing 1.7% of all CpGs (Additional file 8: Table S7). Consistent with their regional identity, these basal ganglia CG-DMRs were specifically enriched in open chromatin regions identified in neuronal nuclei from striatal tissues [11] (Fig. 1d). Over 13% (2295/16,866) of these basal ganglia, CG-DMRs were not identified in our 5-group CG-DMR analysis. We used the Genomic Regions Enrichment of Annotations Tool (rGREAT v4.0.0) [28] to identify enriched gene ontology terms associated with these unique basal ganglia CG-DMRs. Ten of the top 20 significantly enriched terms were related to ion transport or neuronal signaling (Additional file 9: Table S8). This result suggests that differential methylation near these genes (including Ca^{+2} and K^{+} voltage-gated channel subunit genes) could be involved in fine-tuning their expression in particular neuronal populations within the basal ganglia.

Differential methylation analysis identifies distinct neuronal subpopulations in the hippocampus

Interestingly, principal component analysis revealed two distinct clusters originating from the hippocampus that were not detected in our previous analysis of hippocampal tissues [12] (Figs. 1a and 2a). The hippocampus is composed of several subregions consisting of four "cornu ammonis" regions and the dentate gyrus. We hypothesized that our samples represented the specific pyramidal and granule neurons within these respective subregions. We tested this hypothesis by identifying autosomal hippocampal

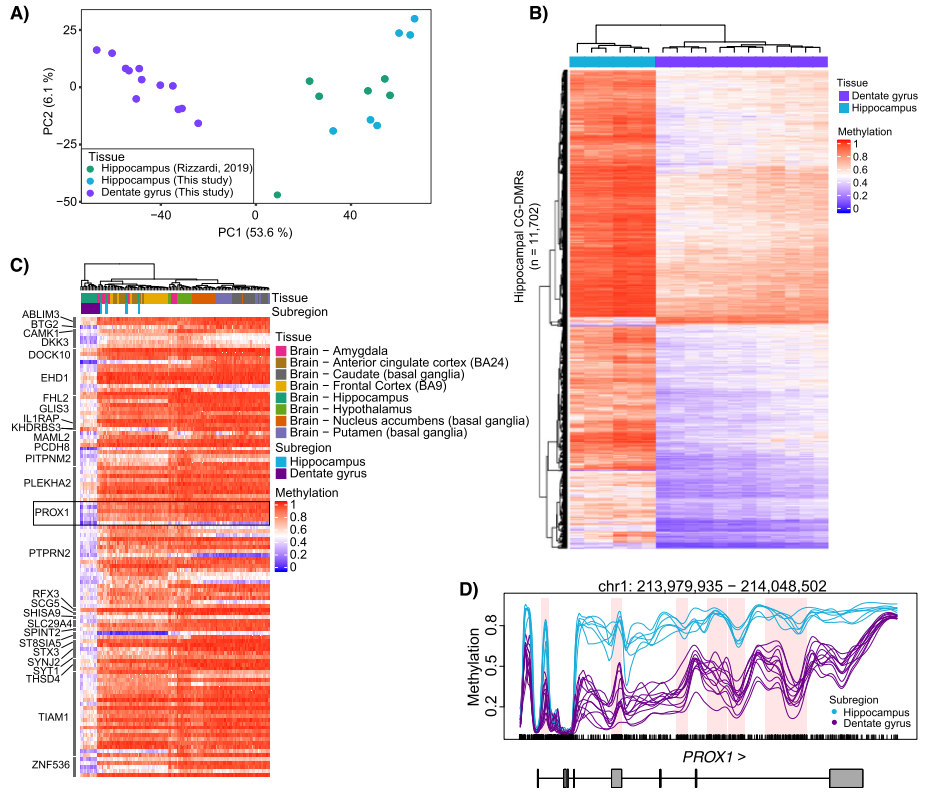

Fig. 2 Differential methylation reveals a subset of hippocampus samples originate from the dentate gyrus. **a** Principal component analysis of distances derived from average autosomal CpG methylation in 1-kb bins. Data shown are from this study and from [11] as indicated. **b** Hierarchical clustering of hippocampus samples based on the average methylation per sample in the CG-DMRs identified between the two hippocampus groups. **c** Hierarchical clustering of hippocampus samples based on the average methylation per sample of hippocampal CG-DMRs overlapping dentate gyrus marker genes. The primary marker of the dentate gyrus, *PROX1*, is boxed. **d** Example hippocampal CG-DMRs in the *PROX1* gene with average methylation values calculated from NeuN+ nuclei isolated from indicated tissue groups and regions of differential methylation shaded pink

CG-DMRs ($n = 11{,}702$) between these two clusters (Fig. 2b, Additional file 10: Table S9). GREAT analysis of the top 2000 hippocampal CG-DMRs showed enrichment in neurogenesis and generation of neurons (Additional file 9: Table S8). As adult neurogenesis occurs in the dentate gyrus, these data suggest that some of these samples originated from that particular subregion. Gene expression data from [29–31] were used to compile a list of 75 genes specifically expressed in dentate gyrus granule neurons (Additional file 11: Table S10) and we intersected hippocampal CG-DMRs with these genes and their promoters (TSS ± 4 kb). We identified 117 hippocampal CG-DMRs overlapping these genes and found that in 12 of the 18 hippocampus samples these marker genes are hypomethylated compared to the other 6 hippocampus samples and the other brain tissues examined (Fig. 2c). Specific examination of the *PROX1* gene, a marker of dentate gyrus granule neurons, reveals hypomethylation in the promoter and throughout the gene body in these 12 samples providing strong evidence that these samples were enriched for dentate gyrus neurons (Fig. 2d). This group of samples is referred to as dentate gyrus samples throughout the rest of the study bringing the total number of brain regions analyzed to nine.

mCH DMRs

Non-CpG methylation (mCH) is widespread in human neurons, and mCH over gene bodies and regulatory elements is generally associated with repression [1, 12, 32]. Interestingly, reduced mCH specifically at neuronal enhancers has recently been associated with Alzheimer's disease pathology [33]. Paradoxically, mCH has also been associated with lowly transcribed genes involved in neuronal development [34] as well as genes escaping X inactivation [35]. Given the importance of mCH in neuronal development and disease, we performed a differential analysis of mCH across our 5 tissue groups. We identified a total of 264,868 CH-DMRs across all contexts (CA, CT, CC) and strands (+, −) covering a third of the genome (1.0 Gb) (Additional file 12: Table S11). This result represents a > 10-fold increase in the number of CH-DMRs identified across the brain compared to our previous work [12]. In that study, we demonstrated high correlations among strands and contexts for mCH; therefore, we use mCA(+) to represent mCH in this study. Global analysis of mCA(+) by principal component analysis revealed segregation of samples based on tissue group though not to the same degree as CpG methylation (Fig. 3a). CH-DMRs were 3.5 times broader than CG-DMRs (3839 vs. 1086 bp, respectively) and were enriched in CG-DMRs with 67,979 (25%) CH-DMRs overlapping 118,621 CG-DMRs (65%). However, CH-DMRs showed little enrichment for genic or regulatory features and were depleted in CpG islands (Fig. 3b). CH-DMRs in the CA(+) context had a median methylation difference of 5.8% with 3195 having a methylation difference $\geq$ 10%. These highly divergent CH-DMRs (Fig. 3b; "> 10%") were particularly enriched in genic/intronic and enhancer regions. Results from our CH-DMR analysis among basal ganglia tissues (152,056 CH-DMRs) and between the two hippocampal clusters (100,757 CH-DMRs) were similar to the 5-group CH-DMR analysis (Additional file 13: Table S12). CH-DMRs also showed a slight enrichment in open chromatin across all brain regions analyzed in [11] (Fig. 3c). Interestingly, hippocampal and basal ganglia CH-DMRs did not show similar enrichments, but were actually depleted in regions of open chromatin in some tissues. Consistent with CG-DMRs, the highly divergent CH-DMRs were generally hypermethylated in basal ganglia tissues compared to the others (Fig. 3d). CH-DMRs exhibit a high degree of overlap among analyses performed using the 5 tissue groups, basal ganglia samples, and hippocampus samples; this is also true for CG-DMRs (Fig. 3e, top). Additionally, CH-DMRs show substantial overlap with CG-DMRs as we previously reported [12] (Fig. 3e, bottom). We can detect many additional CH- and CG-DMRs when looking only among basal ganglia tissues or between hippocampus groups. An example CH-DMR is shown within the gene body of *NRGN*, which encodes the brain-specific protein neurogranin, recently identified as a cerebral spinal fluid biomarker for Alzheimer's disease [36] (Fig. 3f).

Identification of VMRs in neurons isolated from human brain tissues

Interindividual variation in DNA methylation has been of interest to many groups and the GTEx sample collection allowed us to explore tissue-specific methylation variability at a genome-wide scale previously not possible. VMRs are loci that are highly variable among individuals within a given tissue type [21, 22]. As a matter of clarification, the word "variability" has been used in other work to refer to changes in DNA methylation between tissues [37], which is not the meaning of VMR used here. Prior studies of

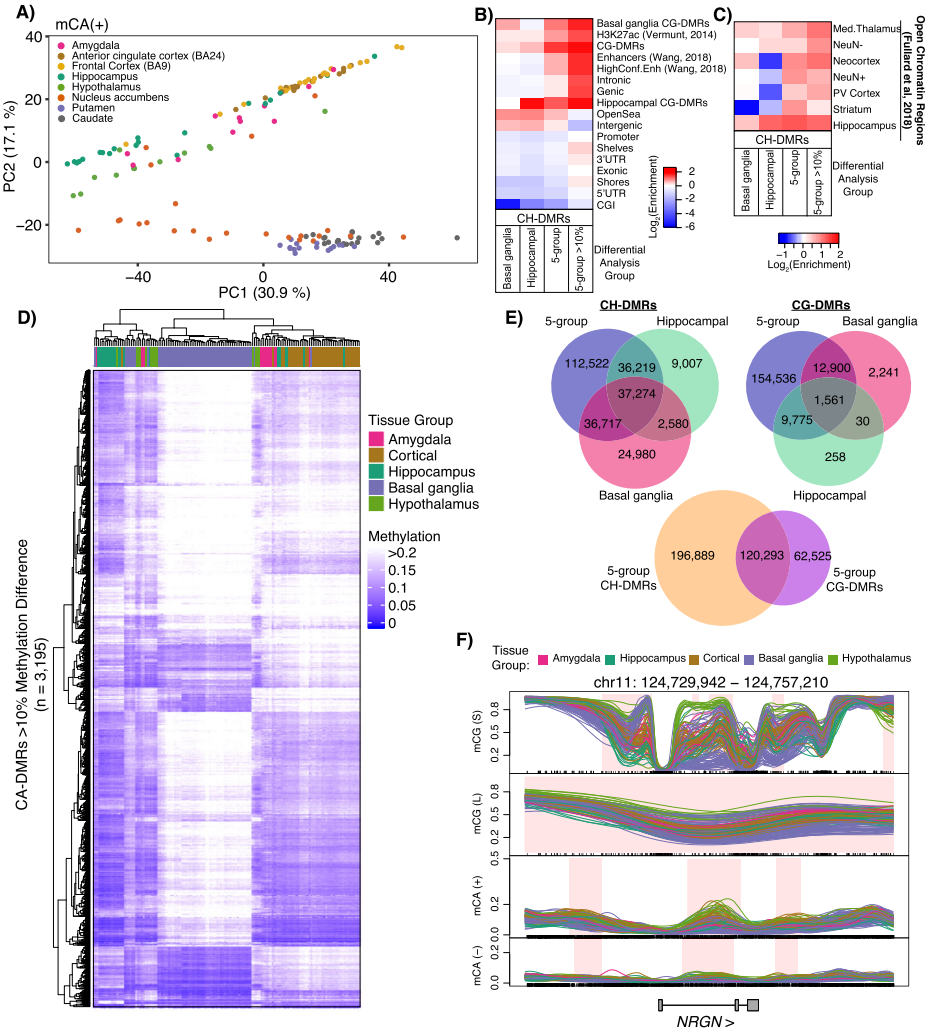

Fig. 3 Differential non-CpG methylation across functionally distinct brain regions. **a** Principal component analysis of distances derived from average autosomal, plus strand CA (mCA+) methylation in 1-kb bins. **b** Heatmap representing log2 enrichment of CA, CT, and CC within CH-DMRs compared to the rest of the genome for indicated features including CG-DMRs identified in this study. Gene models from GENCODEv26 (promoters, intronic, exonic, 5′UTR, 3′UTR, intergenic), CpG islands (CGIs) and related features from UCSC (shores, shelves, OpenSea), putative enhancer regions (enhancers and high confidence enhancers from PsychENCODE [18] and H3K27ac [19]). 5-group = CH-DMRs between all 5 groups; 5-group > 10% = 3195 CH-DMRs from 5-group comparison with mean CA-DMR methylation difference ≥ 10%. **c** As in **b**, showing enrichment in regions of open chromatin in NeuN− and NeuN+ nuclei and NeuN+ nuclei isolated from the indicated brain regions (PV cortex, primary visual cortex; Med. thalamus, mediodorsal thalamus) from [10]. **d** Hierarchical clustering of samples based on the average CA(+) methylation per sample in the CA-DMRs with > 10% methylation difference among the 5 tissue groups. **e** Venn diagrams illustrating intersections between CH- and CG-DMRs identified between different analyses. **f** Example CA-DMR over *NRGN* with both strands and CG-DMRs (mCG(S); obtained from small smoothing window) and blocks (mCG(L); obtained from large smoothing window). Average methylation values calculated from NeuN+ nuclei isolated from indicated tissue groups. Regions of differential methylation are shaded in pink

methylation variability in brain tissues have been limited to targeted genomic regions (Illumina 450k array) [38] or few individuals [39] using a single brain region. Systemic methylation variability can be driven by genetic effects (methylation QTLs in *cis* and/or *trans*), occur independently as metastable epialleles [40], be caused by environmental exposures, or be confounded by cell type heterogeneity. As we have previously shown

[12], much of the variability due to cell type heterogeneity is removed upon isolating NeuN+ nuclei from brain tissues. However, proportions of neuronal subpopulations between brain regions and among individuals could still contribute to variability measurements. Using EpiDISH [41], we estimated the NeuN+ proportions in our samples. For reference data, we used sites of differential methylation between NeuN+ and NeuN − nuclei isolated from orbitofrontal cortex identified in Kozlenkov et al. [42]. We filtered the 51,412 CpG sites they identified as "neuronal undermethylated" and "glial undermethylated" for $|\Delta\beta| > 0.7$ resulting in 426 sites. We then eliminated CpGs that overlapped DMRs identified in our 8-group and hippocampal analyses resulting in 201 reference CpGs. This step was critical to eliminate variation due to known region-specific neuronal methylation differences. We found that only four hippocampus samples had any evidence of glial (NeuN−) contamination thus providing independent validation of our sorting efficiency (Additional file 1: Fig. S2a). Only one of these was less than 97% neuronal with an estimate of 87%.

We identified VMRs by determining the 99th percentile of standard deviation of methylation values in each tissue and applying the lowest standard deviation value (SD = 0.095) as a single cutoff for all tissues (Additional file 1: Fig. S2b). Using the same SD cutoff allows different tissues to have different numbers of VMRs rather than taking the top most variable regions. This strategy allows for the possibility that some tissues are more variable than others. We identified a total of 81,130 unique VMRs containing > 10 CpGs and covering 159 Mb across all nine brain regions, lung, and thyroid (Fig. 4a, Table 1, Additional file 1: Fig. S2, Additional file 14: Table S13). The majority of VMRs are shared among two or more tissues (Fig. 4b, "Shared VMR") with 333 shared among all brain regions. Of those, 202 are "ubiquitous" VMRs, regions of variability shared among all tissues including lung and thyroid (Fig. 4c). Remarkably, an average of 24% of the VMRs identified in each tissue are unique to that tissue and we provide examples of tissue-specific VMRs (Fig. 4b). To quantify the effect size of the variability, we used the range of the per-sample, across-region average methylation. The median effect size is 35% with almost all VMRs having an effect size greater than 20% and some reaching 50% or higher (Additional file 1: Fig. S2c).

Almost all (97%) VMRs overlapped a CG-DMR with 35–50% of VMRs fully contained within a CG-DMR (Table 1). This percentage drops to 18–20% for VMRs identified in the lung and thyroid, which is expected as these two tissues were not included in CG-DMR analyses. This overlap is reflected in the enrichment of VMRs for CG-DMRs, which is less for the lung and thyroid for the reason stated above (Fig. 4d). This can be visualized in Fig. 4b (far right panels) where a VMR is present in the hypothalamus (bottom, green) and the mean methylation is significantly different than that in the amygdala (top, pink) or caudate (middle, gray) thus constituting a CG-DMR. These tissue-specific regions of methylation variability were enriched in putative regulatory regions including enhancer- and transcription-associated chromHMM states (Fig. 4d, Additional file 1: Fig. S3a). VMRs were found across all autosomes (Additional file 1: Fig. S3b) including the MHC region of chromosome 6 which is known to be highly variable. The MHC region, as well as the pericentromeric region of chromosome 20, harbored more ubiquitous VMRs than any other genomic region similar to previous results [39] (Additional file 1: Fig. S4). In contrast to most VMRs, ubiquitous VMRs were particularly enriched in CpG islands and shores (Fig. 4d).

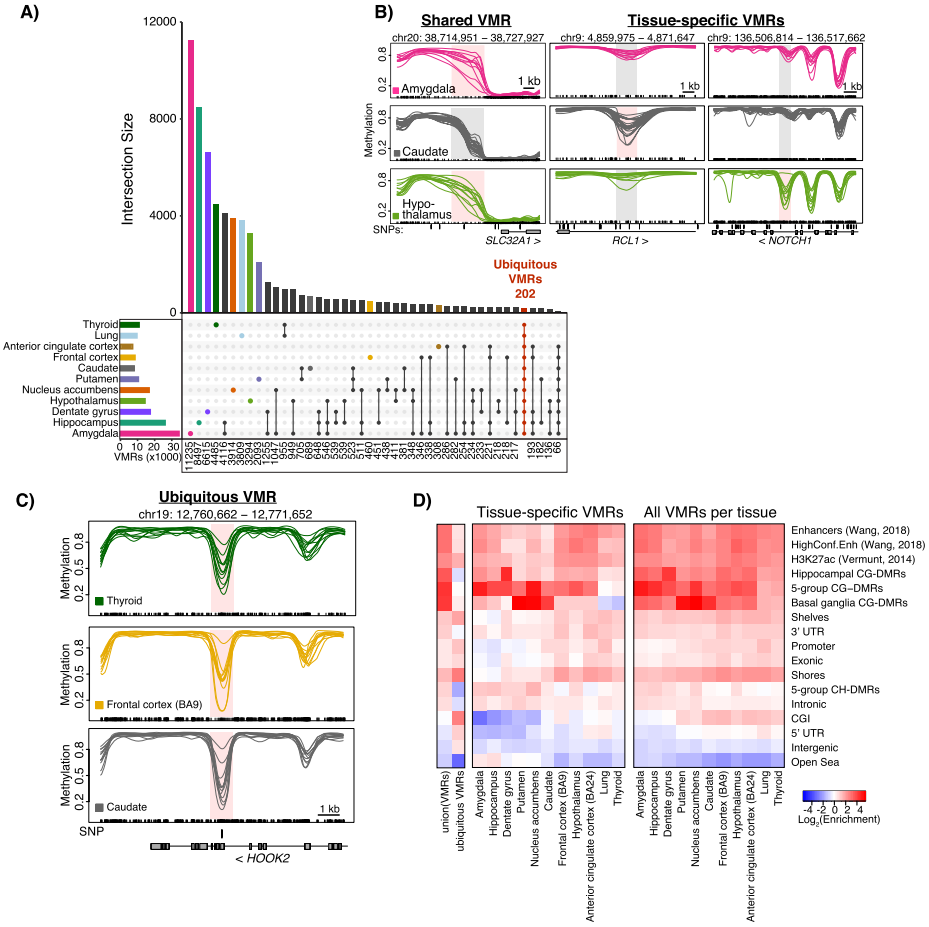

Fig. 4 Methylation variability across brain and non-brain tissues. **a** VMRs shared among two non-brain tissues and neuronal nuclei from distinct brain regions. The number of VMRs in each intersection is listed at the bottom of the plot. **b** Examples of shared and tissue-specific VMRs in the amygdala (pink), caudate (gray), and hypothalamus (green). VMRs are shaded pink in the tissue they were identified in and gray in tissues where they were not considered a VMR; SNPs are indicated when present. **c** Example of a ubiquitous VMR shared across brain and non-brain tissues as indicated. **d** Heatmap representing log2 enrichment of the union of all VMRs identified, ubiquitous VMRs, tissue-specific VMRs, and all VMRs identified in each tissue as indicated compared to the rest of the genome for genomic features. Gene models from GENCODEv26 (promoters, intronic, exonic, 5′UTR, 3′UTR, intergenic), CpG islands (CGIs) and related features from UCSC (shores, shelves, OpenSea), putative enhancer regions (enhancers and high confidence enhancers from PsychENCODE [18] and H3K27ac [19]). 5-group CG-DMRs = CG-DMRs identified among all 5 tissue groups

When considering tissue-specific VMRs, we first focused on the amygdala as 32% of VMRs identified in this tissue were unique to the amygdala. The amygdala displayed the highest interindividual variability among all tissues measured and this was evident both by principal component analysis (Fig. 1a) and in the increased number of total VMRs identified (Table 1). We hypothesized that, similar to the hippocampus, distinct subregions of the amygdala were isolated from these individuals resulting in increased neuronal heterogeneity. For tissue-specific VMRs, we cannot distinguish true variation from that due to cellular heterogeneity, which led us to investigate how large a contribution heterogeneity makes. Neuroimaging analyses of anatomical and functional connectivity have subdivided the amygdala into as many as 9 distinct subnuclei [43–45]. These subnuclei differ in strength of connectivity to other brain regions including the

Table 1 VMRs

Tissue	Total			Tissue-specific			Overlap CG-DMRs	Within CG-DMRs
	VMRs	Mb	CpGs	VMRs	Mb	CpGs		
Dentate gyrus	17,731	16.64	376,310	6,615	6.4	121,072	78.4%	42.3%
Hippocampus	26,536	34.81	609,008	8,497	12.63	152,296	84.3%	39.6%
Amygdala	34,876	44.98	824,870	11,235	15.77	209,042	87.9%	42.9%
Putamen	10,654	11.27	243,030	2,093	2.4	36,101	71.0%	35.4%
Hypothalamus	14,602	14.07	342,524	3,294	3.08	59,739	76.0%	42.8%
Anterior cingulate cortex	7,327	5.14	165,526	308	0.2	4,673	76.9%	50.4%
Nucleus accumbens	17,144	18.66	392,791	3,914	4.69	69,740	84.7%	46.1%
Caudate	8,137	7.19	180,073	689	0.61	10,838	69.0%	37.6%
Frontal cortex	8,719	6.3	205,346	460	0.25	7,804	74.5%	46.6%
Lung	9,770	8.59	233,724	3,809	3.29	77,481	44.1%	18.6%
Thyroid	11,016	10.07	258,613	4,485	3.92	86,891	44.8%	19.6%
Ubiquitous	202	0.49	14,870				25.7%	0.5%

hypothalamus, hippocampus, and cortical regions. For example, a recent neuroimaging study found that the basolateral nucleus displayed stronger connections to the hypothalamus and visual cortex than the centrocortical nucleus which showed stronger connections to the primary motor cortex [46]. As these categorizations are based primarily on neuroimaging data, molecular features of these subregions have yet to be elucidated in the human amygdala. However, single cell RNA-seq of the medial amygdala in mouse led to the identification of 16 distinct neuronal subtypes [47]. We examined methylation within 1 kb of the human homologs of the 262 genes used to cluster the neurons and found 649 VMRs, 116 of which were specific to the amygdala. No tissue-specific VMRs from any other brain region were detected near these genes. Hierarchical clustering of amygdala neuronal samples based on these 649 VMRs reveals three groups suggesting that these samples may have originated from distinct subnuclei within the amygdala (Additional file 1: Fig. S5a,b). VMRs within the *SLC17A7* gene, a marker of glutamatergic neurons, are shown as an example of variable methylation among these three sample groups (Additional file 1: Fig. S5c). These data strongly suggest that variability among amygdala samples is driven by neuronal subtype differences among the subregions sampled. We were unable to identify VMRs within these three distinct groups as only 3–4 individuals were in each subgroup.

When we consider those VMRs that are not tissue-specific, but are shared among at least one other brain region (as shown in Fig. 4a), it is unlikely those VMRs are due to neuronal heterogeneity, because they are shared between brain regions with distinct neuronal populations. Therefore, the variability of these regions must be shared among different cell types which suggests they have some common biological function. Further analysis of the 949 VMRs shared solely between the amygdala and hypothalamus revealed an enrichment for neurotransmitter transport genes, particularly in the SLC family (Additional file 15: Table S14). There is a VMR ~ 3 kb upstream of the TSS of *SLC32A1* (Fig. 4b, "shared VMR"), which is expressed in GABAergic neurons and mediates uptake of GABA and glycine to synaptic vesicles [48]. These shared VMRs are also found near *SLC6A1* (− 15 kb) and *SLC6A11* (+ 1.4 kb), two other GABA transporters for neurons and glia, respectively, as well as upstream of *SLC6A3* (− 52 kb), a dopamine transporter important in the pathogenesis of psychiatric disorders [49]. Among the VMRs identified in the other brain regions, 77–95% were shared among

two or more. These shared VMRs could be important regions for integrating signaling inputs from neuronal crosstalk.

DMRs and VMRs are enriched for heritability of brain-linked traits

We and others have shown a strong association between differential epigenetic features and neurological, neuropsychiatric, and behavioral-cognitive phenotypes [4, 11, 12]. Using stratified linkage disequilibrium score regression [50], we asked if the VMRs we identified in each brain region were also associated with brain-linked traits (Additional file 16: Table S15). First, we replicated our previous findings that regions of differential CpG methylation are enriched for heritability of brain-linked traits including schizophrenia, neuroticism, and depressive symptoms (Fig. 5a; Additional file 1: Fig. S6, Additional file 17: Table S16). In addition, CG-DMRs and those identified among basal ganglia regions showed significant enrichment for heritability of attention deficit hyperactivity disorder (ADHD). VMRs identified in the hypothalamus were also enriched for the heritability of ADHD, while VMRs identified in the amygdala, anterior cingulate cortex, and hippocampus were significantly enriched for the heritability of schizophrenia (Fig. 5b). Amygdala VMRs showed a greater enrichment than CG-DMRs (6.5 vs 4.6) though they cover ~ 75% less of the genome than CG-DMRs (Additional file 17: Table S16).

Genetic contributions to DNA methylation variability

The most likely explanation for a genetic contribution to DNA methylation variability is a genetic polymorphism inside the VMR region. Roughly 25% of VMRs do not overlap any SNPs with a minor allele frequency (MAF) > 0.1 across our samples (Additional file 18: Table S17).

Genotype data from all but two individuals were available from GTEx v8 [16] leaving 6–20 individuals per tissue, which is too few to conduct a rigorous methylation QTL analysis. However, we did identify several examples of VMRs that overlap one or more

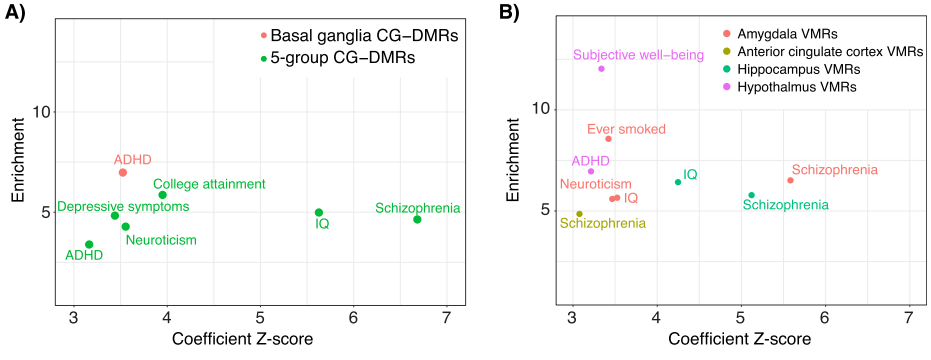

Fig. 5 Neuronal CG-DMRs and VMRs are highly enriched for explained heritability of multiple psychiatric, neurological, and behavioral-cognitive traits. Results from running stratified linkage disequilibrium score regression using 30 GWAS traits with 97 baseline features and either DMRs (**a**) or VMRs (**b**) identified in this study. Only feature-trait combinations with a coefficient *z*-score significantly larger than 0 (one-sided *z*-test with alpha = 0.05, *P*-values corrected within each trait using Holm's method) are shown. Enrichment score (*y*-axis) and coefficient *z*-score (*x*-axis) from running this analysis for each of the indicated methylation features combined with the baseline features are plotted

SNPs that are associated with altered methylation (Fig. 6a). One example is shown within the *MYO3A* gene which lies upstream of *GAD2*, a primary regulator of GABA synthesis that has been associated with schizophrenia [53], bipolar disorder [54], and major depression [55]. There is at least one *GAD2* enhancer located within the *MYO3A* gene, though it is ~ 150 kb away from this VMR [56]. To examine possible genetic contributions to our observed methylation variability, we asked if mQTLs previously identified from brain tissues were enriched in our VMRs (Fig. 6b). The mQTL data we chose were generated using 450k arrays on samples from the bulk hippocampus ($n = 110$)

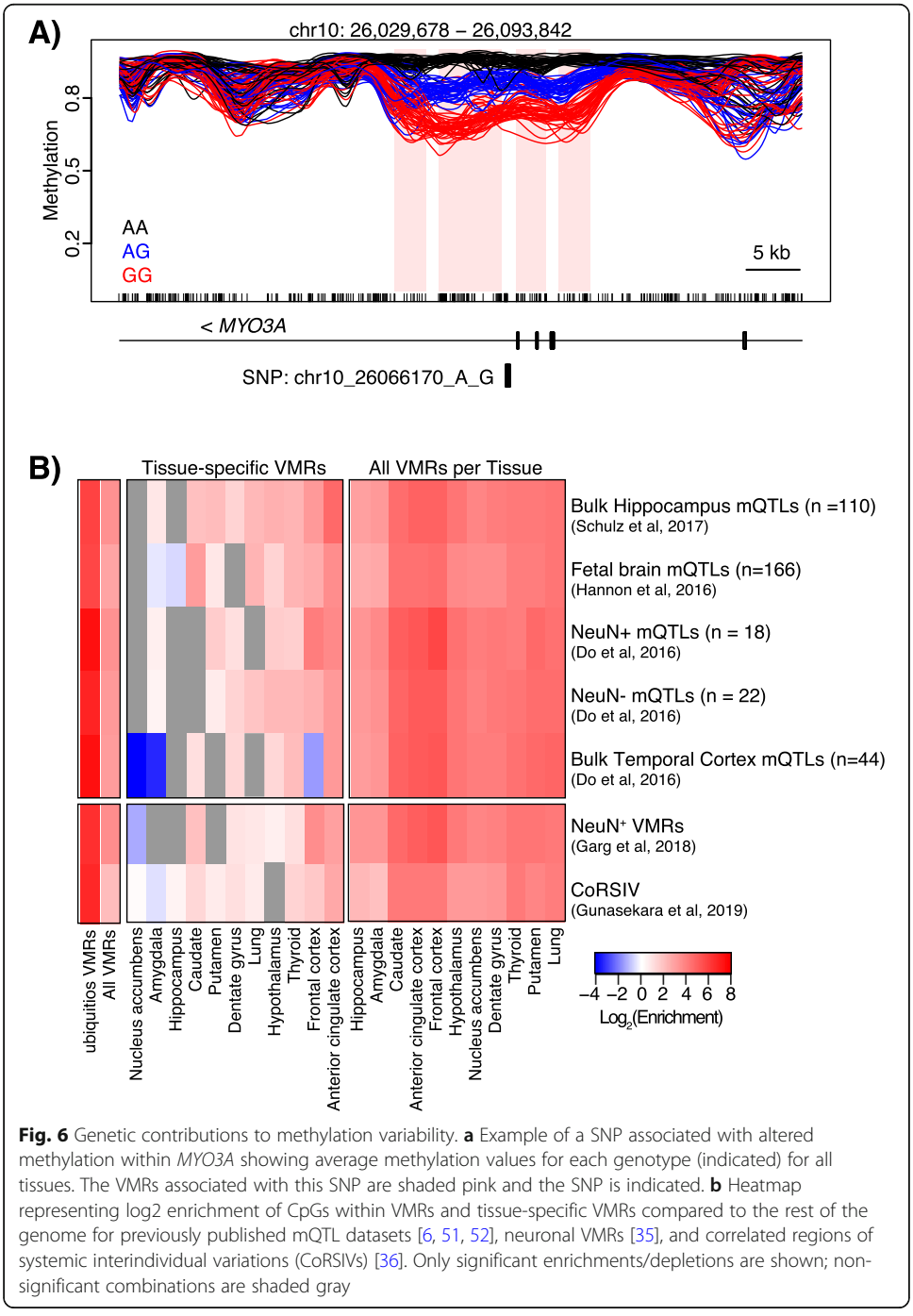

Fig. 6 Genetic contributions to methylation variability. **a** Example of a SNP associated with altered methylation within *MYO3A* showing average methylation values for each genotype (indicated) for all tissues. The VMRs associated with this SNP are shaded pink and the SNP is indicated. **b** Heatmap representing log2 enrichment of CpGs within VMRs and tissue-specific VMRs compared to the rest of the genome for previously published mQTL datasets [6, 51, 52], neuronal VMRs [35], and correlated regions of systemic interindividual variations (CoRSIVs) [36]. Only significant enrichments/depletions are shown; non-significant combinations are shaded gray

[51], fetal brain ($n = 166$) [6], and bulk temporal cortex ($n = 44$) that was also sorted into NeuN+ ($n = 18$) and NeuN− ($n = 22$) nuclei [52]. We found the greatest enrichments in all datasets within our ubiquitous VMRs consistent with our assertion that methylation variability in this class of VMRs is genetically driven. We also compared our VMRs to two existing datasets [38, 39] that examine genetic and environmental contributions to methylation variation. Garg et al. [38] profiled 58 NeuN-sorted frontal cortex samples using the Illumina 450k array, and identified 1136 neuronal VMRs, 996 of which were also detected in our analysis. In addition, they identified 149 VMRs in common among blood, brain, and fibroblast samples and while 142 of these were also identified in our analysis, only 18 were present in our set of ubiquitous VMRs. Gunasekara et al. [39] profiled DNA methylation in three tissues from 10 GTEx donors to identify regions where interindividual methylation variation was not tissue-specific. They identified 9926 correlated regions of systemic interindividual variation (CoRSIVs). We found that our VMRs were enriched for VMRs identified by Garg et al. [38] as well as for CoRSIVs [39] (Fig. 6c). We detected 16% (1588/9926) of CoRSIVs in our analysis. As CoRSIVs are, by definition, consistent across the three tissues sampled (heart, thyroid, and cerebellum), we were unsurprised to find that the enrichment in these regions was lower for tissue-specific VMRs.

Discussion

This resource constitutes the largest examination of DNA methylation differences across isolated neurons of the normal adult human brain encompassing 145 samples across 8 brain regions from 12 to 24 individuals (average of 18). Our focus on neuronal methylation makes this dataset particularly relevant to investigations of neurodegenerative and neuropsychiatric diseases that preferentially impact neuronal function. We have identified 181,146 CG-DMRs and 264,868 CH-DMRs among neurons isolated from the amygdala, hypothalamus, hippocampus, cortical and basal ganglia brain regions. This represents a > 10-fold increase in both CG- and CH-DMRs compared to our previous study of 4 brain regions in 6 individuals [12]. CG-DMRs were enriched for heritability of brain-related traits including schizophrenia, neuroticism, and depressive symptoms. Importantly, we performed the first analysis of differential methylation among basal ganglia regions (nucleus accumbens, caudate, and putamen) identifying 16,866 neuronal CG-DMRs that were enriched for heritability of ADHD. Interestingly, 2295 of these CG-DMRs were not identified in the full 5-group analysis and were enriched for genes encoding voltage-gated ion (Ca^{2+}, K^{+}) channel subunits responsible for maintaining neuronal activity and homeostasis.

Bisulfite sequencing does not distinguish between cytosine methylation and hydroxymethylation. The human brain has elevated levels of hydroxymethylation (~ 10–15%) compared to other tissues [57, 58]. In neuronal subpopulations isolated from the adult prefrontal cortex, 5hmC levels can be as high as 40% [15], and it will be important in future studies to accurately profile this mark across different brain regions.

Large methylation differences among hippocampus and amygdala neuronal samples enabled the identification of distinct anatomical substructures, particularly the dentate gyrus. We further investigated the differences between the hippocampus and dentate gyrus samples and identified CG-DMRs that were enriched for genes involved in

neurogenesis. Substructures within the amygdala are somewhat less well-defined, but we detected three distinct groups based on methylation data that correspond to known features of amygdala neuronal subpopulations identified in mouse [47]. Future studies are needed to fully characterize the substantial neuronal heterogeneity present in the human amygdala.

The large scale of this study enabled us to assess VMRs, regions of interindividual methylation variability within each brain region genome-wide, including identification of concurrent differences across brain regions in a given individual. Methylation variability can be driven by multiple biological and technical factors including genetic variants, environmental exposures, and cellular heterogeneity. We identified a total of 81,130 VMRs among 9 brain regions, lung, and thyroid. We expected genotype differences to drive variability at shared VMRs, and we found that of the 202 ubiquitous VMRs 96% overlapped a SNP (MAF > 0.019). We also identified a substantial number of VMRs that were tissue-specific. The amygdala had the highest number of tissue-specific VMRs and was the most heterogeneous tissue we analyzed suggesting many of these VMRs are driven by cell type differences. Across all tissues, the majority of VMRs (~ 75% on average) were shared among two or more brain regions with more overlap among functionally similar tissues (e.g., frontal cortex and anterior cingulate cortex). Methylation variability in this class of VMRs is unlikely to be due to cell type differences within a given tissue as they are shared across multiple distinct neuronal subtypes present in different regions. VMRs shared between the amygdala and hypothalamus were enriched near genes involved in neurotransmitter transport, particularly members of the SLC family of solute carrier transporters. These shared VMRs could represent loci whose methylation is coordinately regulated between interconnected neurons and thus are candidates for further investigation. Importantly, VMRs identified in the amygdala, hippocampus, and anterior cingulate cortex were enriched for heritability of schizophrenia suggesting their importance in neuropsychiatric disease. This vast resource of differential methylation and variability will be invaluable to future studies of diverse neuronal functions across the human brain.

Methods

Sample procurement

All tissue specimens and DNA samples were obtained from the GTEx Laboratory Data Analysis and Coordination Center at the Broad Institute. Complete descriptions of the donor enrollment and consent process and the biospecimen procurement methods, sample fixation, and histopathological review procedures were previously described [19, 59]. Flash-frozen tissues were obtained from the following brain regions: amygdala ($n = 12$), anterior cingulate cortex (BA24, $n = 15$), caudate (basal ganglia, $n = 22$), frontal cortex (BA9, $n = 24$), hippocampus ($n = 20$), hypothalamus ($n = 13$), nucleus accumbens (basal ganglia, $n = 23$), and putamen ($n = 16$). Genomic DNA was obtained from the lung ($n = 18$) and thyroid ($n = 19$). Demographic information for each donor is presented in Additional file 2: Table S1. Summary data and details on data production and processing are also available from the GTEx Portal (http://gtexportal.org).

Nuclei extraction, fluorescence-activated nuclei sorting, and DNA isolation

Total nuclei were isolated as previously described [12] with the following changes. Approximately 100–200 mg of frozen tissue per sample was homogenized in 5 mL of lysis buffer (0.32 M sucrose, 10 mM Tris pH 8.0, 5 mM $CaCl_2$, 3 mM Mg acetate, 1 mM DTT, 0.1 mM EDTA, 0.1% Triton X-100) by douncing 50 times in a 40-mL dounce homogenizer. Lysates were transferred to a 18-mL ultracentrifugation tube, and 9 mL of sucrose solution (1.8 M sucrose, 10 mM Tris pH 8.0, 3 mM Mg acetate, 1 mM DTT) was dispensed to the bottom of the tube. The samples were then centrifuged at 28,600 rpm for 2 h at 4 °C (Beckman Optima XE-90; SW32 Ti rotor). After centrifugation, the supernatant was removed by aspiration and the nuclear pellet was resuspended in 200 μL staining mix (2% normal goat serum, 0.1% BSA, 1:500 anti-NeuN conjugated to AlexaFluor488 (Millipore, cat#: MAB377X) in PBS) and incubated on ice. Unstained nuclei served as the negative control. The fluorescent nuclei were run through a Beckman Coulter MoFlo Cell Sorter with proper gate settings (Additional file 1: Fig. S1). A small portion of the $NeuN^+$ and $NeuN^-$ nuclei were re-run on the sorter to validate the purity which was greater than 95%. Only immuno-positive ($NeuN^+$) nuclei were collected. DNA was extracted directly from the sorted nuclei without centrifugation using the MasterPure DNA Extraction kit (Epicentre, Madison, WI, USA) following the manufacturer's instructions.

Whole genome bisulfite sequencing (WGBS)

WGBS single indexed libraries were generated using the NEBNext Ultra DNA library Prep kit for Illumina (New England BioLabs, Ipswich, MA, USA) on the Agilent Bravo automated liquid handling platform with a custom high-throughput protocol. Forty to 300 ng gDNA was quantified by Qubit dsDNA BR assay (Invitrogen, Carlsbad, CA, USA), and 1% unmethylated lambda DNA (cat#: D1521, Promega, Madison, WI, USA) was spiked in to measure bisulfite conversion efficiency. DNA was fragmented to an average insert size of 400–500 bp using the Covaris LE220 Focused-ultrasonicator in a 55-μl volume. The fragmented gDNA was converted to end-repaired, adenylated DNA using the NEBNext Ultra End Repair/dA-Tailing Module (cat#: 7442 L, New England BioLabs, Ipswich, MA, USA). Methylated adaptors (NEBNext Multiplex Oligos for Illumina; cat#: E7535L New England BioLabs, Ipswich, MA, USA) were ligated to the product from the preceding step using the NEBNext Ultra Ligation Module (cat#: 7445 L, New England BioLabs, Ipswich, MA, USA). Size selection was performed using AMPure XP beads and insert sizes of ~ 400 bp were isolated (0.37x and 0.20x ratios). Samples were bisulfite converted after size selection using the EZ-96 DNA Methylation-Gold Kit (cat#: D5008, Zymo, Irvine, CA, USA) following the manufacturer's instructions. Amplification was performed following bisulfite conversion using primers from the NEBNext Multiplex Oligos for Illumina module (cat#: E7535L, New England BioLabs, Ipswich, MA, USA) and the Kapa HiFi Uracil+ PCR system (cat#: KK2801, Kapa Biosystems, Boston, MA, USA) with the following cycling parameters: 98 °C 45 s/8 cycles: 98 °C 15 s, 65 °C 30 s, 72 °C 30 s/72 °C 1 min. The PCR enriched product was cleaned up using 0.9x AMPure XP beads (cat#: A63881, Beckman Coulter, Brea, CA, USA). Final libraries were run on 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) using the High-Sensitivity DNA assay; samples were also run on Bioanalyzer after

shearing and size selection for quality control purposes. Libraries were quantified by qPCR using the Library Quantification Kit for Illumina sequencing platforms (cat#: KK4824, KAPA Biosystems, Boston, MA, USA), using 7900HT Real-Time PCR System (Applied Biosystems). Libraries were sequenced with the Illumina HiSeq4000 using 151-bp paired-end run with a 27% PhiX spike-in.

Mapping and quality control of WGBS reads

We trimmed reads of their adapter sequences using Trim Galore! (v0.4.0) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and quality-trimmed using a cut-off of 30. We then aligned these trimmed reads to the hg38 build of the human genome [including autosomes, sex chromosomes, mitochondrial sequence (available from https://software.broadinstitute.org/gatk/download/bundle) plus lambda phage (accession NC_001416.1) but excluding non-chromosomal sequences] using Bismark [60] (v0.19.0) with the following alignment parameters: bismark --bowtie2 -1 ${READ1} -2 ${READ2}. Additional file 2: Table S2 summarizes the alignment results. Using the reads aligned to the lambda phage genome, we estimated that all libraries had a bisulfite conversion rate > 99%.

We then used bismark_methylation_extractor to summarize the number of reads supporting a methylated cytosine and the number of reads supported a unmethylated cytosine for every cytosine in the reference genome. Specifically, we first computed and visually inspected the M-bias [61] of our libraries. Based on these results, we decided to ignore the first 2 bp and last 1 bp of read1 and the first and last 3 bp of read2 in the subsequent call to bismark_methylation_extractor with parameters --ignore 2 --ignore_r2 3 --ignore_3prime 1 --ignore_3prime_r2 3. The final cytosine report file summarizes the methylation evidence at each cytosine in the reference genome.

To confirm the genotypes of our samples matched the correct individual GTEx donors, we first downloaded the GTEx v8 genotype data from dbGaP. We validated our genotypes with 58 SNPs used for quality control with the MethylationEPIC BeadChip microarray (Illumina) [62]. For each aligned bam file, we created a vcf of these SNPs using samtools mpileup -v -t DP,AD -f hg38.fasta –-positions 58_SNPs.bed -R -b list_of_bamfiles.txt | bcftools call -m > sample.vcf. The resulting vcf files were indexed and used as input to the bcftools gtcheck command. If the participant ID with the lowest discordance value did not match the sample participant ID, then the sample was removed from the analysis as failing genotype QC. Thirty samples failed this quality control measure.

EpiDISH analysis

To independently validate our sorting efficiency, we used the robust partial correlations (RPC) method from the EpiDISH R package v2.4.0 [41] to estimate the proportion of non-neuronal nuclei in our NeuN+ samples. The reference sites used were the 23,670 "neuronal undermethylated" and 27,742 "glial undermethylated" CpGs identified in [42]. We filtered for $|\Delta\beta| > 0.7$ resulting in 426 sites. To remove any known sites of region-specific methylation variation, we removed an addition 209 CpGs that overlapped our 8-group and hippocampal DMRs for a final reference set of 201 CpGs.

Neuronal proportions were estimated from smoothed BS β-values from all our NeuN+ samples.

Annotation and external data

The hg38 build of the human reference genome was used for all analyses. Genes, exons, introns, and UTRs were taken from GENCODE v26 (https://www.gencodegenes.org/human/release_26.html) [63]. Gene bodies were defined by taking the union over all transcripts (transcription start site to transcription end site) for each gene. Promoters were defined as 4 kb centered on the transcription start site. CpG islands were downloaded from UCSC (http://genome.ucsc.edu/) [64, 65]. CpG shores are defined as 2 kb flanking CpG islands and CpG shelves are defined as 2 kb flanking CpG islands. The 15-state chromHMM model for 7 adult brain tissues from the Roadmap Epigenomics Project [25] was downloaded using the R/Bioconductor AnnotationHub package (v2.6.4). The selected brain regions and their Roadmap Epigenomics codes were brain angular gyrus (E067), brain anterior caudate (E068), brain cingulate gyrus (E069), brain germinal matrix (E070), brain hippocampus middle (E071), brain inferior temporal lobe (E072), brain dorsolateral prefrontal cortex (E073), brain substantia nigra (E074), fetal brain male (E081) and fetal brain female (E082). Separate enrichments were calculated for each chromHMM annotation then averaged together for graphical representation. The PsychENCODE enhancer set for the prefrontal cortex was taken from [13] (http://resource.psychencode.org/). Brain enhancers denoted by H3K27ac were obtained from [24]. Regions of open chromatin in neuronal and non-neuronal nuclei from 14 brain regions were obtained from [11].

For the analysis of hippocampal subgroups, we used gene expression data from [29–31] to compile a list of 75 genes specifically expressed in dentate gyrus granule neurons (Additional file 11: Table S10). We examined DNA methylation of the hippocampal CG-DMRs that overlapped these genes or their promoters (TSS ± 4 kb). For the analysis of amygdala subregions, we profiled DNA methylation within 1 kb of the TSS of 216 genes used to cluster single cell expression data from mouse medial amygdala neurons (Supplementary Table S2 in [47]). We first converted the mouse genes to their human homologs and position in hg38 using the biomaRt v2.38.0 [66, 67] R package. We examined DNA methylation at VMRs identified in the amygdala that were within 1 kb of these genes. Any data originally mapped to hg19 were lifted over to hg38 using liftOver from the rtracklayer R package [68].

Gene expression data was downloaded from GTEx v8 (dbGaP Accession: phs000424.v8.p2). While these expression data were generated from the same donors and tissues used in our study, gene expression was measured in bulk tissues rather than in sorted neuronal nuclei. Therefore, no large-scale integrative analyses were performed with these data.

DMR analyses

Differentially methylated regions and blocks of differential methylation were identified and annotated as previously described [12] except a minimum of 70 CpGs were required in a smoothing window. Large blocks of differential methylation were identified after smoothing over windows of at least 20 kb containing at least 500 CpGs. Following

smoothing, we analyzed all CpGs that had a sequencing coverage of at least 1 in all samples for a total of 27,660,298 CpGs. Unlike CpGs, CpAs and CpTs are not palindromic, so were analyzed separately for each strand, for a total of 4 strand/dinucleotide combinations:

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

For each dinucleotide/strand combination, we ran a single "small-ish" smooth to identify DMRs (smoothing over windows of at least 3 kb containing at least 200 CpAs or CpTs). Following smoothing, we analyzed all CpAs and CpTs regardless of sequencing coverage. We performed five distinct CpG and non-CpG DMR analyses. The first analysis identified both CpG and non-CpG (CH) DMRs among neuronal nuclei isolated from 5 tissue groups: cortical (BA9, BA24), basal ganglia (putamen, nucleus accumbens, caudate), hippocampus, hypothalamus, and amygdala brain regions. To identify the most discriminatory CG-DMRs (in Fig. 1b), we used the annotations to identify those DMRs unique to a specific group. For example, DMRs identified only when the amygdala was compared to the other four groups, and not when those four groups were compared to each other. This resulted in a total of 27,782 discriminatory CG-DMRs:

1. Basal ganglia CG-DMRs = 10,000 (sorted by maxStat and took the top 10,000)
2. Cortical CG-DMRs = 1604
3. Amygdala CG-DMRs = 14
4. Hippocampus CG-DMRs = 14,158
5. Hypothalamus CG-DMRs = 2006

We also performed the same analyses across all 8 tissues individually. We next looked for DMRs among neuronal nuclei within the basal ganglia tissues. Finally, we identified DMRs among the two groups of hippocampus samples to determine why they were segregating into two distinct clusters identified by principal component analysis. We ordered these hippocampal CG-DMRs by absolute value of areaStat and used the top 2000 DMRs for analysis of gene ontology using rGREAT v4.0.0 [28].

VMR analyses

For variably methylated region (VMR) analysis, we considered 24,630,044 CpGs that had sequencing coverage of ≥ 5 reads in ≥ 100 samples. To VMRs, genome-wide standard deviations of DNA methylation were determined and a standard deviation of 0.095 was chosen as our cutoff for all tissues ensuring that at least 99% of the methylation distribution is within ± 2 standard deviations (equivalent to 38% methylation difference). VMRs were then filtered to include those with > 10 CpGs and with a Cook's distance > 0.7 to remove VMRs driven by sample outliers.

Enrichment of DMRs, blocks, and VMRs in genomic features

We formed a 2×2 contingency table of (n_{11}, n_{12}, n_{21}, n_{22}); specific values of (n_{11}, n_{12}, n_{21}, n_{22}) are described below. The enrichment log odds ratio was estimated by $\log_2(\mathrm{OR}) = \log_2(n_{11}) + \log_2(n_{22}) - \log_2(n_{12}) - \log_2(n_{21})$, its standard error was estimated by se (log2(OR)) = sqrt ($1 / n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22}$), and an approximate 95% confidence interval formed by [$\log_2(\mathrm{OR}) - 2 \times$ se (log2(OR)), $\log_2(\mathrm{OR}) + 2 \times$ se (log2(OR))]. As the odds ratio is equivalent to enrichment, for clarity, figures show $\log_2$(enrichment) rather than $\log_2$(OR). We also report the *P*-value obtained from performing Fisher's exact test for testing the null of independence of rows and columns in the 2×2 table (i.e., the null of no enrichment or depletion) using the fisher.test() function from the "stats" package in R [69].

We computed the enrichment of CpXs (CpGs, CpAs, or CpTs, as appropriate) within DMRs, blocks, or VMRs inside each genomic feature (e.g., exons, enhancers, etc.). Specifically, for each genomic feature, we constructed the 2×2 table (n_{11}, n_{12}, n_{21}, n_{22}), where:

- n_{11} = number of CpXs in DMRs/blocks/VMRs that were inside the feature
- n_{12} = number of CpXs in DMRs/blocks/VMRs that were outside the feature
- n_{21} = number of CpXs not in DMRs/blocks/VMRs that were inside the feature
- n_{22} = number of CpXs not in DMRs/blocks/VMRs that were outside the feature

The total number of CpXs, $n = n_{11} + n_{12} + n_{21} + n_{22}$, was the number of autosomal CpXs in the reference genome. We counted CpXs rather than the number of DMRs or bases because this accounts for the non-uniform distribution of CpXs along the genome and avoids double-counting DMRs that are both inside and outside the feature.

Genotype analysis

We utilized genotypes from GTEx Analysis Release v8 (dbGaP Accession: phs000424.v8.p2) to identify SNPs associated with methylation values at VMRs. We first removed SNPs in linkage disequilibrium, removed missing genotypes, and kept only autosomal SNPs using Plink v2.0 (www.cog-genomics.org/plink/2.0/) [70] with the following parameters: --geno 0.1 --indep-pairwise 100'kb' 1 0.2 --autosome. We then used Plink to filter the dataset to include only those 26 individuals for which we generated DNA methylation data and retained only SNPs with MAF $\geq$ 0.1 resulting in 149, 185 autosomal SNPs. A SNP matrix was created from the resulting vcf using the readVCF and genotypeToSnpMatrix functions in the VariantAnnotation (v1.28.13) [71] R package.

Stratified linkage disequilibrium score regression

We used stratified linkage disequilibrium score regression (SLDSR), implemented in the LDSC [72] software, to evaluate the enrichment of common genetic variants from genome-wide association study (GWAS) signals to partition trait heritability within functional categories represented by our DMRs and VMRs. SLDSR estimates the proportion of genome-wide single nucleotide polymorphism (SNP)-based heritability that can be attributed to SNPs within a given genomic feature by a regression model that

combines GWAS summary statistics with estimates of linkage disequilibrium from an ancestry-matched reference panel. Links to GWAS summary statistics are available in Additional file 16: Table S15. The GRCh38 "baseline-LD model v2.2" data files were downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE/ following instructions at https://github.com/bulik/ldsc/wiki.

We ran LDSC (v1.0.0; https://github.com/bulik/ldsc) to estimate the proportion of genome-wide SNP-based heritability of 30 traits (Additional file 16: Table S15) across the 97 "baseline" genomic features and our neuronal (NeuN+) DMRs and VMRs:

1. 5-group CG-DMRs: CG-DMRs among 5 brain tissue groups (196 Mb) (Additional file 6: Table S5)
2. 5-group CH-DMRs: union of CA-, CC-, and CT-DMRs (both strands) among 5 brain tissue groups (1010 Mb) (Additional file 12: Table S11)
3. Basal ganglia CG-DMRs: CG-DMRs among 3 basal ganglia tissues (24 Mb) (Additional file 8: Table S7)
4. Basal ganglia CH-DMRs: union of CA-, CC-, and CT-DMRs (both strands) among 3 basal ganglia tissues (284.4 Mb) (Additional file 13: Table S12)
5. Hippocampal CG-DMRs: CG-DMRs among the two hippocampal groups (24.4 Mb) (Additional file 10: Table S9)
6. Hippocampal CH-DMRs: union of CA-, CC-, and CT-DMRs (both strands) among the two hippocampal groups (596 Mb) (Additional file 13: Table S12)
7. VMRs identified in each brain tissue (Mb covered listed in Table 1) (Additional file 14: Table S13)

We performed a standard SLDSR analysis, as suggested by the method authors, whereby each of the brain-specific features was added one at a time to a "full baseline model" that included the 97 "baseline" categories that capture a broad set of genomic annotations. We used SLDSR to estimate a "coefficient *z*-score" and an "enrichment score" for each feature-trait combination. A brief description of their interpretation is given below; we refer the interested reader to the Online Methods of [50] for the complete mathematical derivation. A coefficient *z*-score statistically larger than zero indicates that adding the feature to the model increased the explained heritability of the trait, beyond the heritability explained by other features in the model. The enrichment score is defined as (proportion of heritability explained by the feature)/(proportion of SNPs in the feature). The enrichment score is unadjusted for the other features in the model, but is more readily interpretable as an effect size. Particularly interesting are those feature-trait combinations with statistically significant *z*-score coefficients and large enrichment scores. *z*-score coefficient *p*-values within each trait were post hoc adjusted for multiple testing using Holm's method [73] (Additional file 17: Table S16).

Supplementary software

All statistical analyses were performed using R [69] (v3.5.x) and made use of packages contributed to the Bioconductor project [66, 74]. In addition to those R/Bioconductor packages specifically referenced in the above, we made use of several other packages in preparing results for the manuscript:

- bsseq (v1.14)
- AnnotationHub (v2.6.4)
- biomaRt [66, 67] (v2.30.0)
- GenomicAlignments [75] (v1.10.0)
- GenomicFeatures [75] (v1.26.2)
- GenomicRanges [75] (v1.26.2)
- ggplot2 [76] (v2.2.1)
- Hmisc (v4.0-2)
- Matrix (v1.2-8)
- Rtracklayer [68] (v1.34.1)
- SummarizedExperiment (v1.4.0)
- EnrichedHeatmap (v1.4.0)
- Picard (v2.2.2)
- seqtk (v1.2-r94)
- tximport (v1.2.0)
- rGREAT (v4.0.0)

Supplementary Information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13059-021-02335-w.

Additional file 1. Supplementary Materials including Supplementary Figures and Description of Supplementary Tables.

Additional file 2: Table S1. Sample phenotype data.

Additional file 3: Table S2. WGBS sequencing statistics.

Additional file 4: Table S3. Samples failing genotype quality check.

Additional file 5: Table S4. CpG DMRs identified among neuronal samples isolated from all 8 brain regions.

Additional file 6: Table S5. CpG DMRs identified among the 5 brain region groups.

Additional file 7: Table S6. Large blocks of differential CpG methylation identified among the 5 brain region groups.

Additional file 8: Table S7. CpG DMRs identified among neuronal samples isolated from the basal ganglia tissues.

Additional file 9: Table S8. GREAT analysis of basal ganglia CpG DMRs and the top 2000 Hippocampal CpG DMRs.

Additional file 10: Table S9. CpG DMRs identified among neuronal samples isolated from the two hippocampus tissue groups.

Additional file 11: Table S10. List of 75 granule cell marker genes and hippocampal CpG DMRs that overlap granule cell marker genes or their promoters.

Additional file 12: Table S11. CH DMRs identified among the 5 brain regions groups.

Additional file 13: Table S12. CH DMRs identified among neuronal samples isolated from the basal ganglia tissues and between the two hippocampus tissue groups.

Additional file 14: Table S13. Lists of VMRs identified in each tissue.

Additional file 15: Table S14. GREAT analysis of VMRs shared between amygdala and hypothalamus samples.

Additional file 16: Table S15. Links and references to summary statistics for traits used in SLDSR analyses.

Additional file 17: Table S16. Results from SLDSR analyses.

Additional file 18: Table S17. Overlap between VMRs and SNPs at multiple allele frequencies.

Additional file 19. Review history.

Acknowledgements

We thank the donors and their families for their generous gifts of organ donation for transplantation, and tissue donations for the GTEx research project.

Review history

The review history is available as Additional file 19.

Peer review information
Anahita Bishop was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions
L.F.R, K.D.H, and A.P.F designed the study; L.F.R. and H.Z. performed nuclei sorting with the assistance of K.E.S. and S.D.T.; R.T., A.I., C.M.C., and L.F.R. performed DNA extractions; A.I., R.T., and C.M.C. performed WGBS library preparation and sequencing; S.R. and L.F.R. established the sequence processing pipeline, A.P.F. oversaw the experiments; K.D.H. oversaw the data analysis. L.F.R, P.F.H., and K.D.H performed the data analysis; L.F.R, P.F.H., K.D.H, and A.P.F. interpreted the results; L.F.R, K.D.H, and A.P.F wrote the manuscript. All authors read and approved the final manuscript.

Funding
The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The work performed here was supported by NIH Grant U01MH104393 (A.P.F.). The Flow Cytometry Cell Sorting Core Facility at Johns Hopkins School of Public Health was supported by CFAR: 5P30AI094189-04, 1S10OD016315-01, and 1S10RR13777001.

Availability of data and materials
Genotype-Tissue Expression (GTEx) project's raw whole genome bisulfite sequencing, whole transcriptome, and genome sequencing data are available via dbGaP accession number phs000424 [77] through AnVIL (https://app.terra.bio/#workspaces/anvil-datastorage/AnVIL_GTEx_V9_hg38). Processed GTEx data described in this manuscript are available via the GTEx portal (http://gtexportal.org/). All the code used for the reproducible analysis is available on Zenodo with the access code DOI https://doi.org/10.5281/zenodo.4635255 [78] and GitHub https://github.com/hansenlab/egtex_brain_wgbs [79]. Data are made available as a UCSC track hub that can be accessed by going to https://genome.ucsc.edu/cgi-bin/hgHubConnect#unlistedHubs and loading https://egtex-wgbs.s3.amazonaws.com/hub.txt. PsychENCODE prefrontal cortex enhancers from http://resource.psychencode.org/ [13]; brain enhancers denoted by H3K27ac from [24]; regions of open chromatin in neuronal and non-neuronal nuclei from 14 brain regions from [11]; 216 genes used to cluster single cell expression data from mouse medial amygdala neurons from [47]; GRCh38 "baseline-LD model v2.2" data files from https://data.broadinstitute.org/alkesgroup/LDSCORE/; 15-state chromHMM model for 7 adult brain tissues from the Roadmap Epigenomics Project [25] was downloaded using the R/Bioconductor AnnotationHub package (v2.6.4); Gencode, https://www.gencodegenes.org/human/release_26.html [63]; CpG islands, http://genome.ucsc.edu/; reference CpG sites used for EpiDISH deconvolution from [42]; non-GTEx CG-DMRs from [12].

Declarations

Ethics approval and consent to participate
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Center for Epigenetics, Johns Hopkins University School of Medicine, 855 N. Wolfe St., Baltimore, MD 21205, USA. [2]HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, AL 35806, USA. [3]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe St, Baltimore, MD 21205, USA. [4]Epigenetics and Development Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia. [5]Department of Medical Biology, University of Melbourne, Parkville, Victoria, Australia. [6]Department of Pediatrics, Division of Infectious Diseases, University of Arkansas for Medical Sciences, 13 Children's Way, Little Rock, AR 72202, USA. [7]Arkansas Children's Research Institute, Little Rock, AR 72202, USA. [8]Department of Pharmacology and Molecular Sciences, Johns Hopkins University, Baltimore, MD 21205, USA. [9]Department of Molecular Microbiology and Immunology, Johns Hopkins School of Public Health, 615 N. Wolfe St, Baltimore, MD 21205, USA. [10]Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [11]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [12]Departments of Biomedical Engineering and Mental Health, Johns Hopkins University Schools of Engineering and Public Health, Baltimore, MD, USA.

Received: 6 July 2020 Accepted: 30 March 2021
Published online: 22 April 2021

References
1. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghighi FG, Sejnowski TJ, Behrens MM, Ecker JR. Global epigenomic reconfiguration during mammalian brain development. Science. 2013;341(6146):1237905. https://doi.org/10.1126/science.1237905.
2. Zhu J, Adli M, Zou JY, Verstappen G, Coyne M, Zhang X, Durham T, Miri M, Deshpande V, De Jager PL, et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell. 2013;152(3):642–54. https://doi.org/10.1016/j.cell.2012.12.033.
3. Sun MA, Sun Z, Wu X, Rajaram V, Keimig D, Lim J, Zhu H, Xie H. Mammalian brain development is accompanied by a dramatic increase in bipolar DNA methylation. Sci Rep. 2016;6(1):32298. https://doi.org/10.1038/srep32298.

4. Price AJ, Collado-Torres L, Ivanov NA, Xia W, Burke EE, Shin JH, Tao R, Ma L, Jia Y, Hyde TM, Kleinman JE, Weinberger DR, Jaffe AE. Divergent neuronal DNA methylation patterns across human cortical development reveal critical periods and a unique role of CpH methylation. Genome Biol. 2019;20(1):196. https://doi.org/10.1186/s13059-019-1805-1.
5. Pidsley R, Viana J, Hannon E, Spiers H, Troakes C, Al-Saraj S, Mechawar N, Turecki G, Schalkwyk LC, Bray NJ, Mill J. Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. Genome Biol. 2014;15(10):483. https://doi.org/10.1186/s13059-014-0483-2.
6. Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, Troakes C, Turecki G, O'Donovan MC, Schalkwyk LC, Bray NJ, Mill J. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat Neurosci. 2016; 19(1):48–54. https://doi.org/10.1038/nn.4182.
7. Hoffmann A, Sportelli V, Ziller M, Spengler D. Epigenomics of major depressive disorders and schizophrenia: early life decides. Int J Mol Sci. 2017;18(8) https://doi.org/10.3390/ijms18081711.
8. Jaffe AE, Gao Y, Deep-Soboslay A, Tao R, Hyde TM, Weinberger DR, Kleinman JE. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. Nat Neurosci. 2016;19(1):40–7. https://doi.org/10.1038/nn.4181.
9. Karpova NN, Sales AJ, Joca SR. Epigenetic basis of neuronal and synaptic plasticity. Curr Top Med Chem. 2017;17(7):771–93. https://doi.org/10.2174/1568026616666160414124628.
10. Tognini P, Napoli D, Pizzorusso T. Dynamic DNA methylation in the brain: a new epigenetic mark for experience-dependent plasticity. Front Cell Neurosci. 2015;9:331.
11. Fullard JF, Hauberg ME, Bendl J, Egervari G, Cirnaru MD, Reach SM, Motl J, Ehrlich ME, Hurd YL, Roussos P. An atlas of chromatin accessibility in the adult human brain. Genome Res. 2018;28(8):1243–52. https://doi.org/10.1101/gr.232488.117.
12. Rizzardi LF, Hickey PF, Rodriguez DiBlasi V, Tryggvadottir R, Callahan CM, Idrizi A, Hansen KD, Feinberg AP. Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability. Nat Neurosci. 2019;22(2):307–16. https://doi.org/10.1038/s41593-018-0297-8.
13. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, Clarke D, Gu M, Emani P, Yang YT, Xu M, Gandal MJ, Lou S, Zhang J, Park JJ, Yan C, Rhie SK, Manakongtreecheep K, Zhou H, Nathan A, Peters M, Mattei E, Fitzgerald D, Brunetti T, Moore J, Jiang Y, Girdhar K, Hoffman GE, Kalayci S, Gümüş ZH, Crawford GE, PsychENCODE Consortium, Roussos P, Akbarian S, Jaffe AE, White KP, Weng Z, Sestan N, Geschwind DH, Knowles JA, Gerstein MB. Comprehensive functional genomic resource and integrative model for the human brain. Science. 2018;362(6420):eaat8464. https://doi.org/10.1126/science.aat8464.
14. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, Sunkin SM, Li Z, Shin Y, Zhu Y, Sousa AMM, Werling DM, Kitchen RR, Kang HJ, Pletikos M, Choi J, Muchnik S, Xu X, Wang D, Lorente-Galdos B, Liu S, Giusti-Rodríguez P, Won H, de Leeuw CA, Pardiñas AF, BrainSpan Consortium†, PsychENCODE Consortium†, PsychENCODE Developmental Subgroup†, Hu M, Jin F, Li Y, Owen MJ, O'Donovan MC, Walters JTR, Posthuma D, Reimers MA, Levitt P, Weinberger DR, Hyde TM, Kleinman JE, Geschwind DH, Hawrylycz MJ, State MW, Sanders SJ, Sullivan PF, Gerstein MB, Lein ES, Knowles JA, Sestan N. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science. 2018;362(6420):eaat7615. https://doi.org/10.1126/science.aat7615.
15. Kozlenkov A, Li J, Apontes P, Hurd YL, Byne WM, Koonin EV, Wegner M, Mukamel EA, Dracheva S. A unique role for DNA (hydroxy) methylation in epigenetic regulation of human inhibitory neurons. Sci Adv. 2018;4:eaau6190.
16. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369: 1318–30.
17. Kim-Hellmuth S, Aguet F, Oliva M, Muñoz-Aguirre M, Kasela S, Wucher V, et al. Cell type–specific genetic regulation of gene expression across human tissues. Science. 2020;369:eaaz8528.
18. Consortium G. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580–5.
19. Consortium G: Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348(6235):648–60. https://doi.org/10.1126/science.1262110.
20. eGTExProject. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. Nat Genet. 2017; 49(12):1664–70. https://doi.org/10.1038/ng.3969.
21. Jaffe AE, Feinberg AP, Irizarry RA, Leek JT. Significance analysis and statistical dissection of variably methylated regions. Biostatistics. 2012;13(1):166–78. https://doi.org/10.1093/biostatistics/kxr013.
22. Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, Eiriksdottir G, Harris TB, Launer L, Gudnason V, Fallin MD. Personalized epigenomic signatures that are stable over time and covary with body mass index. Sci Transl Med. 2010;2:49ra67.
23. Kirsch L, Chechik G. On expression patterns and developmental origin of human brain regions. Plos Comput Biol. 2016; 12(8):e1005064. https://doi.org/10.1371/journal.pcbi.1005064.
24. Vermunt MW, Reinink P, Korving J, de Bruijn E, Creyghton PM, Basak O, Geeven G, Toonen PW, Lansu N, Meunier C, van Heesch S, Netherlands Brain Bank, Clevers H, de Laat W, Cuppen E, Creyghton MP. Large-scale identification of coregulated enhancer networks in the adult human brain. Cell Rep. 2014;9(2):767–79. https://doi.org/10.1016/j.celrep.2014.09.023.
25. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317–30.
26. Abad MA, Enguita M, DeGregorio-Rocasolano N, Ferrer I, Trullas R. Neuronal pentraxin 1 contributes to the neuronal damage evoked by amyloid-beta and is overexpressed in dystrophic neurites in Alzheimer's brain. J Neurosci. 2006; 26(49):12735–47. https://doi.org/10.1523/JNEUROSCI.0575-06.2006.
27. Koob GF, Volkow ND. Neurobiology of addiction: a neurocircuitry analysis. Lancet Psychiatry. 2016;3(8):760–73. https://doi.org/10.1016/S2215-0366(16)00104-8.
28. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28(5):495–501. https://doi.org/10.1038/nbt.1630.
29. Cembrowski MS, Wang L, Sugino K, Shields BC, Spruston N. Hipposeq: a comprehensive RNA-seq database of gene expression in hippocampal principal neurons. Elife. 2016;5:e14997. https://doi.org/10.7554/eLife.14997.
30. Mancarci BO, Toker L, Tripathy SJ, Li B, Rocco B, Sibille E, Pavlidis P: Cross-laboratory analysis of brain cell type transcriptomes with applications to interpretation of bulk tissue data. eNeuro 2017, 4, 6, ENEURO.0212, ENEU17.2017, doi: https://doi.org/10.1523/ENEURO.0212-17.2017.

31. Jaffe AE, Hoeppner DJ, Saito T, Blanpain L, Ukaigwe J, Burke EE, Collado-Torres L, Tao R, Tajinda K, Maynard KR, Tran MN, Martinowich K, Deep-Soboslay A, Shin JH, Kleinman JE, Weinberger DR, Matsumoto M, Hyde TM. Profiling gene expression in the human dentate gyrus granule cell layer reveals insights into schizophrenia and its genetic risk. Nat Neurosci. 2020;23(4):510–9. https://doi.org/10.1038/s41593-020-0604-z.
32. Mo A, Mukamel EA, Davis FP, Luo C, Henry GL, Picard S, Urich MA, Nery JR, Sejnowski TJ, Lister R, Eddy SR, Ecker JR, Nathans J. Epigenomic signatures of neuronal diversity in the mammalian brain. Neuron. 2015;86(6):1369–84. https://doi.org/10.1016/j.neuron.2015.05.018.
33. Li P, Marshall L, Oh G, Jakubowski JL, Groot D, He Y, Wang T, Petronis A, Labrie V. Epigenetic dysregulation of enhancers in neurons is associated with Alzheimer's disease pathology and cognitive symptoms. Nat Commun. 2019;10(1):2246. https://doi.org/10.1038/s41467-019-10101-7.
34. Stroud H, Su SC, Hrvatin S, Greben AW, Renthal W, Boxer LD, Nagy MA, Hochbaum DR, Kinde B, Gabel HW, Greenberg ME. Early-life gene expression in neurons modulates lasting epigenetic states. Cell. 2017;171(5):1151–64 e1116. https://doi.org/10.1016/j.cell.2017.09.047.
35. Keown CL, Berletch JB, Castanon R, Nery JR, Disteche CM, Ecker JR, Mukamel EA. Allele-specific non-CG DNA methylation marks domains of active chromatin in female mouse brain. Proc Natl Acad Sci U S A. 2017;114(14):E2882–90. https://doi.org/10.1073/pnas.1611905114.
36. Kvartsberg H, Lashley T, Murray CE, Brinkmalm G, Cullen NC, Hoglund K, Zetterberg H, Blennow K, Portelius E. The intact postsynaptic protein neurogranin is reduced in brain tissue from patients with familial and sporadic Alzheimer's disease. Acta Neuropathol. 2019;137(1):89–102. https://doi.org/10.1007/s00401-018-1910-3.
37. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;500(7463):477–81. https://doi.org/10.1038/nature12433.
38. Garg P, Joshi RS, Watson C, Sharp AJ. A survey of inter-individual variation in DNA methylation identifies environmentally responsive co-regulated networks of epigenetic variation in the human genome. Plos Genet. 2018;14(10):e1007707. https://doi.org/10.1371/journal.pgen.1007707.
39. Gunasekara CJ, Scott CA, Laritsky E, Baker MS, MacKay H, Duryea JD, Kessler NJ, Hellenthal G, Wood AC, Hodges KR, Gandhi M, Hair AB, Silver MJ, Moore SE, Prentice AM, Li Y, Chen R, Coarfa C, Waterland RA. A genomic atlas of systemic interindividual epigenetic variation in humans. Genome Biol. 2019;20(1):105. https://doi.org/10.1186/s13059-019-1708-1.
40. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. Trends Genet. 2002;18(7):348–51. https://doi.org/10.1016/S0168-9525(02)02709-9.
41. Zheng SC, Breeze CE, Beck S, Dong D, Zhu T, Ma L, Ye W, Zhang G, Teschendorff AE. EpiDISH web server: epigenetic dissection of intra-sample-heterogeneity with online GUI. Bioinformatics. 2019; https://doi.org/10.1093/bioinformatics/btz833.
42. Kozlenkov A, Roussos P, Timashpolsky A, Barbu M, Rudchenko S, Bibikova M, Klotzle B, Byne W, Lyddon R, Di Narzo AF, et al. Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. Nucleic Acids Res. 2014;42(1):109–27. https://doi.org/10.1093/nar/gkt838.
43. Etkin A, Prater KE, Schatzberg AF, Menon V, Greicius MD. Disrupted amygdalar subregion functional connectivity and evidence of a compensatory network in generalized anxiety disorder. Arch Gen Psychiatry. 2009;66(12):1361–72. https://doi.org/10.1001/archgenpsychiatry.2009.104.
44. Bach DR, Behrens TE, Garrido L, Weiskopf N, Dolan RJ. Deep and superficial amygdala nuclei projections revealed in vivo by probabilistic tractography. J Neurosci. 2011;31(2):618–23. https://doi.org/10.1523/JNEUROSCI.2744-10.2011.
45. Saygin ZM, Kliemann D, Iglesias JE, van der Kouwe AJW, Boyd E, Reuter M, Stevens A, Van Leemput K, McKee A, Frosch MP, et al. High-resolution magnetic resonance imaging reveals nuclei of the human amygdala: manual segmentation to automatic atlas. Neuroimage. 2017;155:370–82. https://doi.org/10.1016/j.neuroimage.2017.04.046.
46. Abivardi A, Bach DR. Deconstructing white matter connectivity of human amygdala nuclei with thalamus and cortex subdivisions in vivo. Hum Brain Mapp. 2017;38(8):3927–40. https://doi.org/10.1002/hbm.23639.
47. Wu YE, Pan L, Zuo Y, Li X, Hong W. Detecting activated cell populations using single-cell RNA-Seq. Neuron. 2017;96(2):313–29 e316. https://doi.org/10.1016/j.neuron.2017.09.026.
48. Aubrey KR. Presynaptic control of inhibitory neurotransmitter content in VIAAT containing synaptic vesicles. Neurochem Int. 2016;98:94–102. https://doi.org/10.1016/j.neuint.2016.06.002.
49. Salatino-Oliveira A, Rohde LA, Hutz MH. The dopamine transporter role in psychiatric phenotypes. Am J Med Genet B Neuropsychiatr Genet. 2018;177(2):211–31. https://doi.org/10.1002/ajmg.b.32578.
50. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47(11):1228–35. https://doi.org/10.1038/ng.3404.
51. Schulz H, Ruppert AK, Herms S, Wolf C, Mirza-Schreiber N, Stegle O, Czamara D, Forstner AJ, Sivalingam S, Schoch S, Moebus S, Pütz B, Hillmer A, Fricker N, Vatter H, Müller-Myhsok B, Nöthen MM, Becker AJ, Hoffmann P, Sander T, Cichon S. Genome-wide mapping of genetic determinants influencing DNA methylation and gene expression in human hippocampus. Nat Commun. 2017;8(1):1511. https://doi.org/10.1038/s41467-017-01818-4.
52. Do C, Lang CF, Lin J, Darbary H, Krupska I, Gaba A, Petukhova L, Vonsattel JP, Gallagher MP, Goland RS, Clynes RA, Dwork A, Kral JG, Monk C, Christiano AM, Tycko B. Mechanisms and disease associations of haplotype-dependent allele-specific DNA methylation. Am J Hum Genet. 2016;98(5):934–55. https://doi.org/10.1016/j.ajhg.2016.03.027.
53. Fatemi SH, Stary JM, Earle JA, Araghi-Niknam M, Eagan E. GABAergic dysfunction in schizophrenia and mood disorders as reflected by decreased levels of glutamic acid decarboxylase 65 and 67 kDa and Reelin proteins in cerebellum. Schizophr Res. 2005;72(2-3):109–22. https://doi.org/10.1016/j.schres.2004.02.017.
54. Heckers S, Stone D, Walsh J, Shick J, Koul P, Benes FM. Differential hippocampal expression of glutamic acid decarboxylase 65 and 67 messenger RNA in bipolar disorder and schizophrenia. Arch Gen Psychiatry. 2002;59(6):521–9. https://doi.org/10.1001/archpsyc.59.6.521.
55. Moyer CE, Delevich KM, Fish KN, Asafu-Adjei JK, Sampson AR, Dorph-Petersen KA, Lewis DA, Sweet RA. Reduced glutamate decarboxylase 65 protein within primary auditory cortex inhibitory boutons in schizophrenia. Biol Psychiatry. 2012;72(9):734–43. https://doi.org/10.1016/j.biopsych.2012.04.010.

56. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res. 2007;35(Database):D88–92. https://doi.org/10.1093/nar/gkl822.
57. Wen L, Li X, Yan L, Tan Y, Li R, Zhao Y, Wang Y, Xie J, Zhang Y, Song C, Yu M, Liu X, Zhu P, Li X, Hou Y, Guo H, Wu X, He C, Li R, Tang F, Qiao J. Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. Genome Biol. 2014;15(3):R49. https://doi.org/10.1186/gb-2014-15-3-r49.
58. Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, Katz E, Dixon JM, Harrison DJ, Meehan RR. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Res. 2012;22(3):467–77. https://doi.org/10.1101/gr.126417.111.
59. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-Demchok J, Gelfand ET, et al. A novel approach to high-quality postmortem tissue procurement: the GTEx project. Biopreserv Biobank. 2015;13(5):311–9. https://doi.org/10.1089/bio.2015.0032.
60. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27(11):1571–2. https://doi.org/10.1093/bioinformatics/btr167.
61. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol. 2012;13(10):R83. https://doi.org/10.1186/gb-2012-13-10-r83.
62. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8(3):389–99. https://doi.org/10.2217/epi.15.114.
63. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigo R, Hubbard TJ. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012;22(9):1760–74. https://doi.org/10.1101/gr.135350.111.
64. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. Brief Bioinform. 2013;14(2):144–61. https://doi.org/10.1093/bib/bbs038.
65. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, Harte RA, Heitner S, Hickey G, Hinrichs AS, Hubley R, Karolchik D, Learned K, Lee BT, Li CH, Miga KH, Nguyen N, Paten B, Raney BJ, Smit AFA, Speir ML, Zweig AS, Haussler D, Kuhn RM, Kent WJ. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 2015;43(D1):D670–81. https://doi.org/10.1093/nar/gku1177.
66. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005;21(16):3439–40. https://doi.org/10.1093/bioinformatics/bti525.
67. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184–91. https://doi.org/10.1038/nprot.2009.97.
68. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics. 2009;25(14):1841–2. https://doi.org/10.1093/bioinformatics/btp328.
69. Team RC. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016. 2016
70. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1):7. https://doi.org/10.1186/s13742-015-0047-8.
71. Obenchain V, Lawrence M, Carey V, Gogarten S, Shannon P, Morgan M. VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. Bioinformatics. 2014;30(14):2076–8. https://doi.org/10.1093/bioinformatics/btu168.
72. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, Patterson N, Daly MJ, Price AL, Neale BM. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47(3):291–5. https://doi.org/10.1038/ng.3211.
73. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979;6:65–70.
74. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):R80. https://doi.org/10.1186/gb-2004-5-10-r80.
75. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. Plos Comput Biol. 2013;9(8):e1003118. https://doi.org/10.1371/journal.pcbi.1003118.
76. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. New York: Springer-Verlag; 2016. Available from: https://ggplot2.tidyverse.org.
77. GTEx Consortium: Genotype-Tissue Expression Project (GTEx). dbGaP 2020, Available from: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424. Accessed 13 Dec 2017.
78. Rizzardi LF, Hickey PF, Idrizi A, Tryggvadóttir R, Callahan CM, Stephens KE, et al. hansenlab/egtex_brain_wgbs: GenomeBiologyArchive (Version v1.0) Zenodo. 2021. Available from: https://doi.org/10.5281/zenodo.4635255.
79. Rizzardi LF, Hickey PF, Idrizi A, Tryggvadóttir R, Callahan CM, Stephens KE, et al. eGTEx brain WGBS analyses. GitHub. 2021. Available from: https://github.com/hansenlab/egtex_brain_wgbs.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.