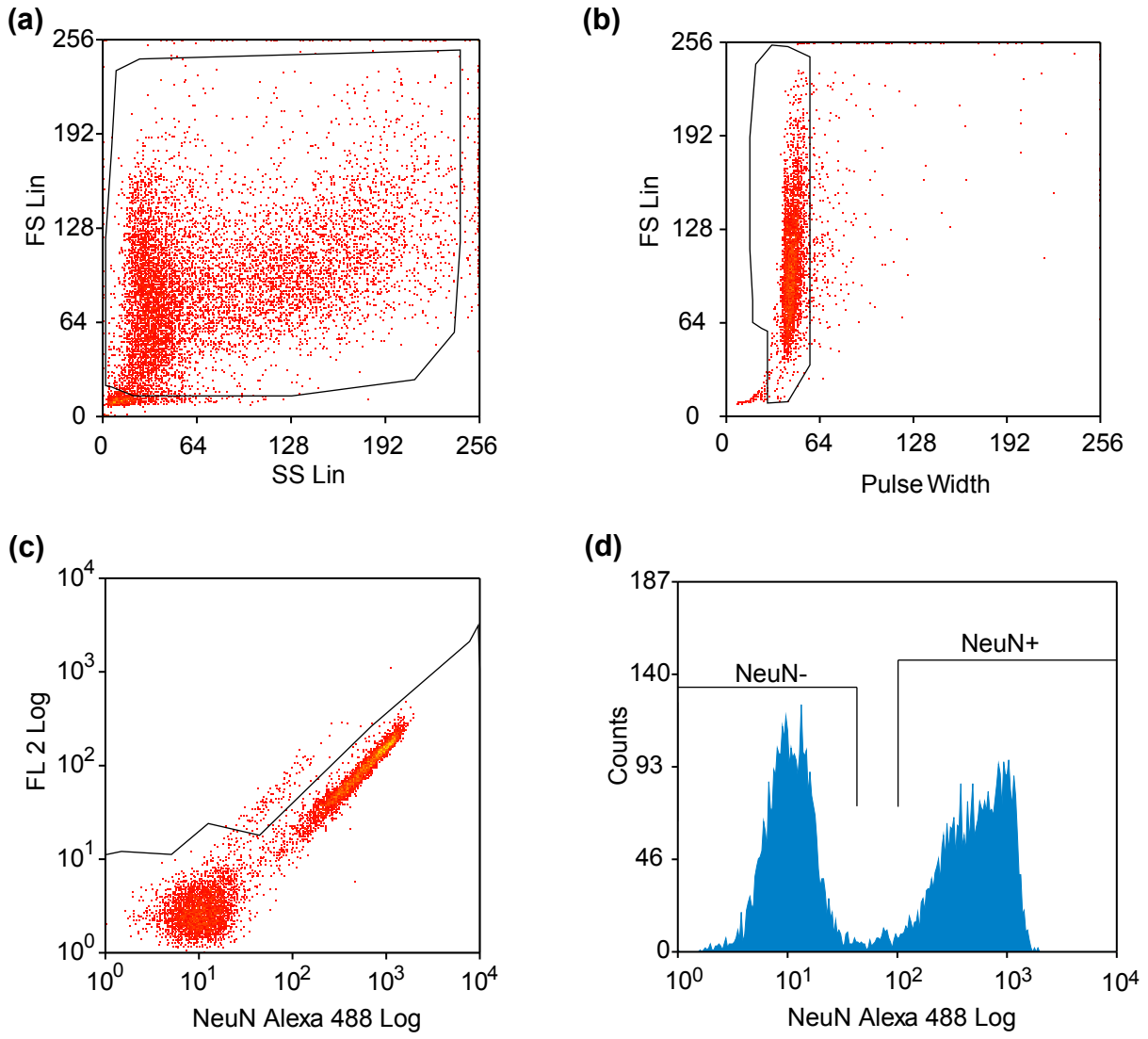
Supplementary Material

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

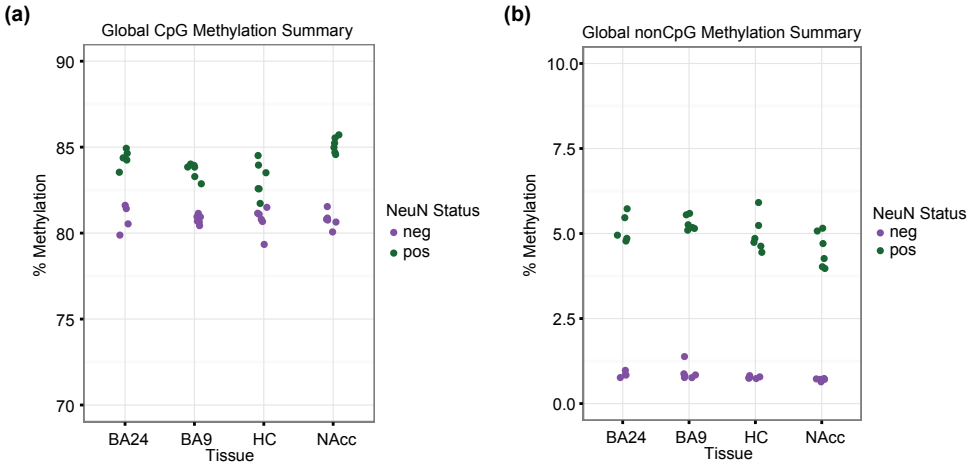
Lindsay F. Rizzardi[1,2,+], Peter F. Hickey[3,+], Varenka Rodriguez DiBlasi[1,2], Rakel Tryggvadóttir[1], Colin M. Callahan[1], Adrian Idrizi[1], Kasper D. Hansen[1,3,4,*], Andrew P. Feinberg[1,2,5,*]

[1]Center for Epigenetics, Johns Hopkins University School of Medicine, 855 N. Wolfe St., Baltimore, MD, 21205
[2]Department of Medicine, Johns Hopkins University School of Medicine
[3]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe St, Baltimore, MD 21205
[4]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Johns Hopkins University
[5]Departments of Biomedical Engineering and Mental Health, Johns Hopkins University Schools of Medicine, Engineering, and Public Health

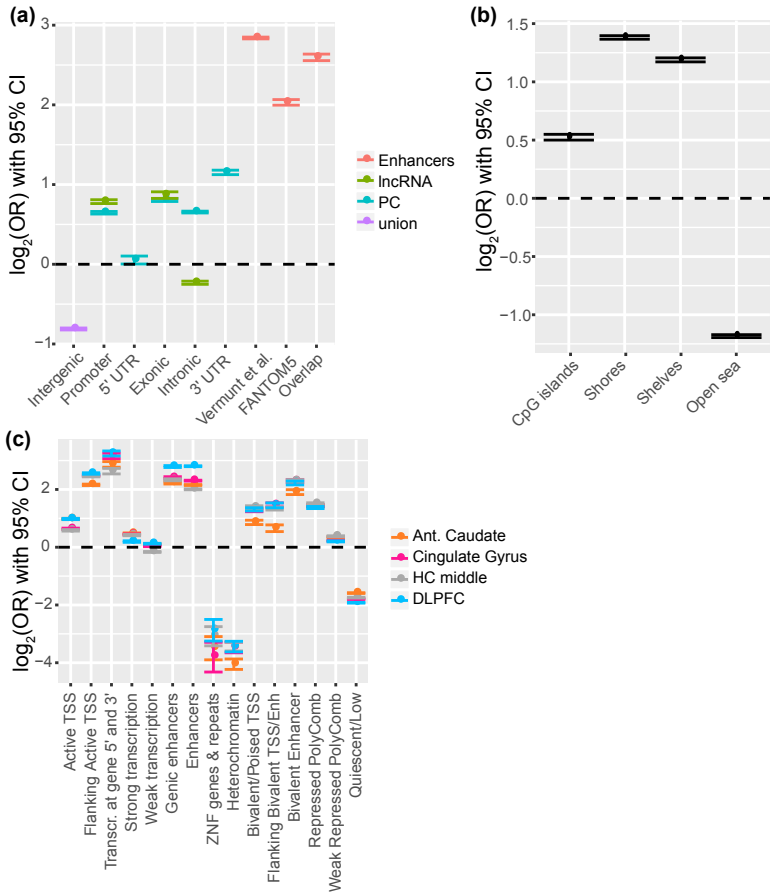
[+]both authors contributed equally to this work
*co-corresponding authors



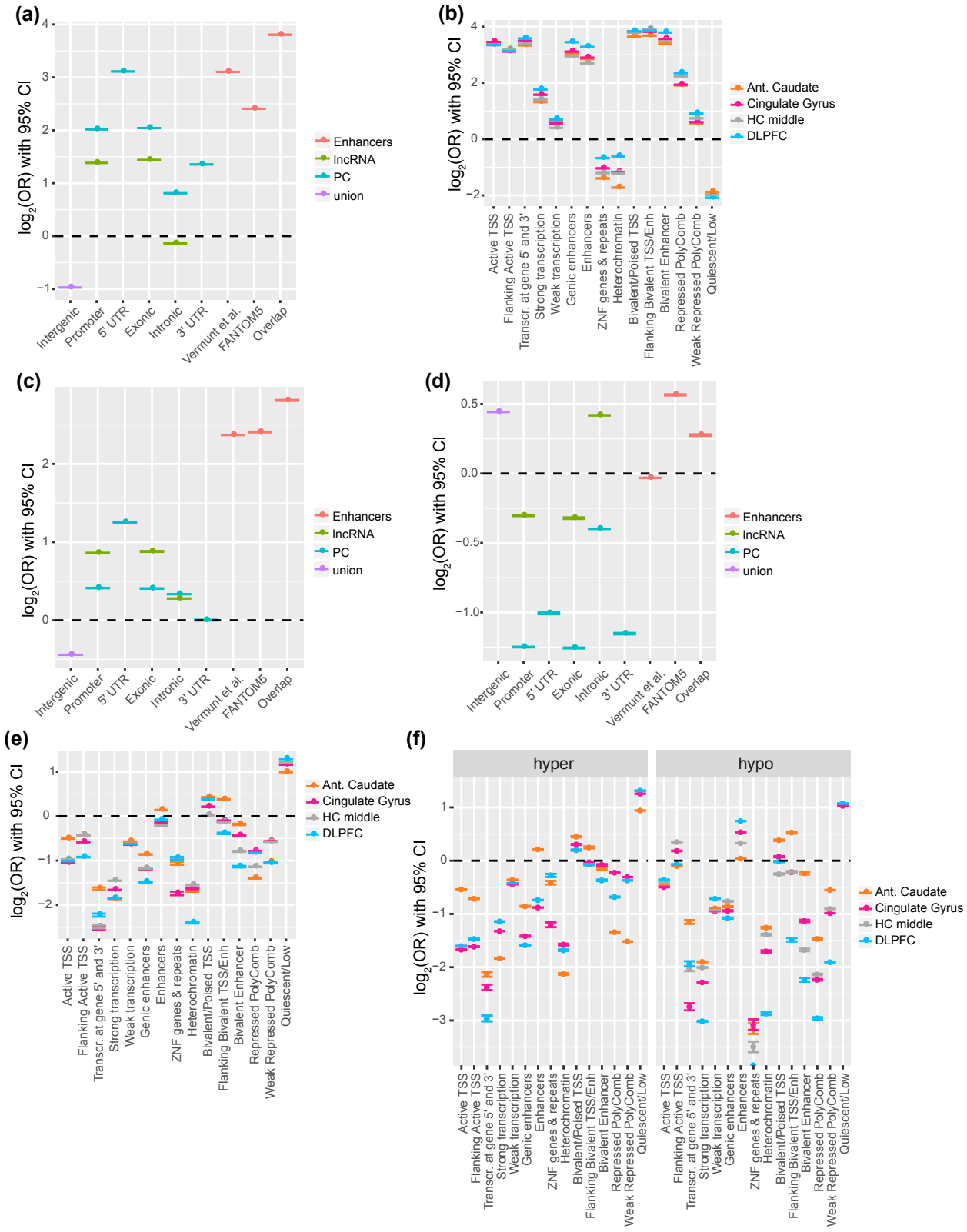
Supplementary Figure 1. Sorting neuronal nuclei from frozen brain tissues. Neuronal and non-neuronal nuclei were isolated by fluorescence activated nuclei sorting. Nuclei in this representative example were isolated from prefrontal cortex and (a) debris, (b) doublets, and (c) auto-fluorescent nuclei were gated out. (d) The remaining nuclei were separated based on detection of AlexaFluor 488-conjugated anti-NeuN antibody.



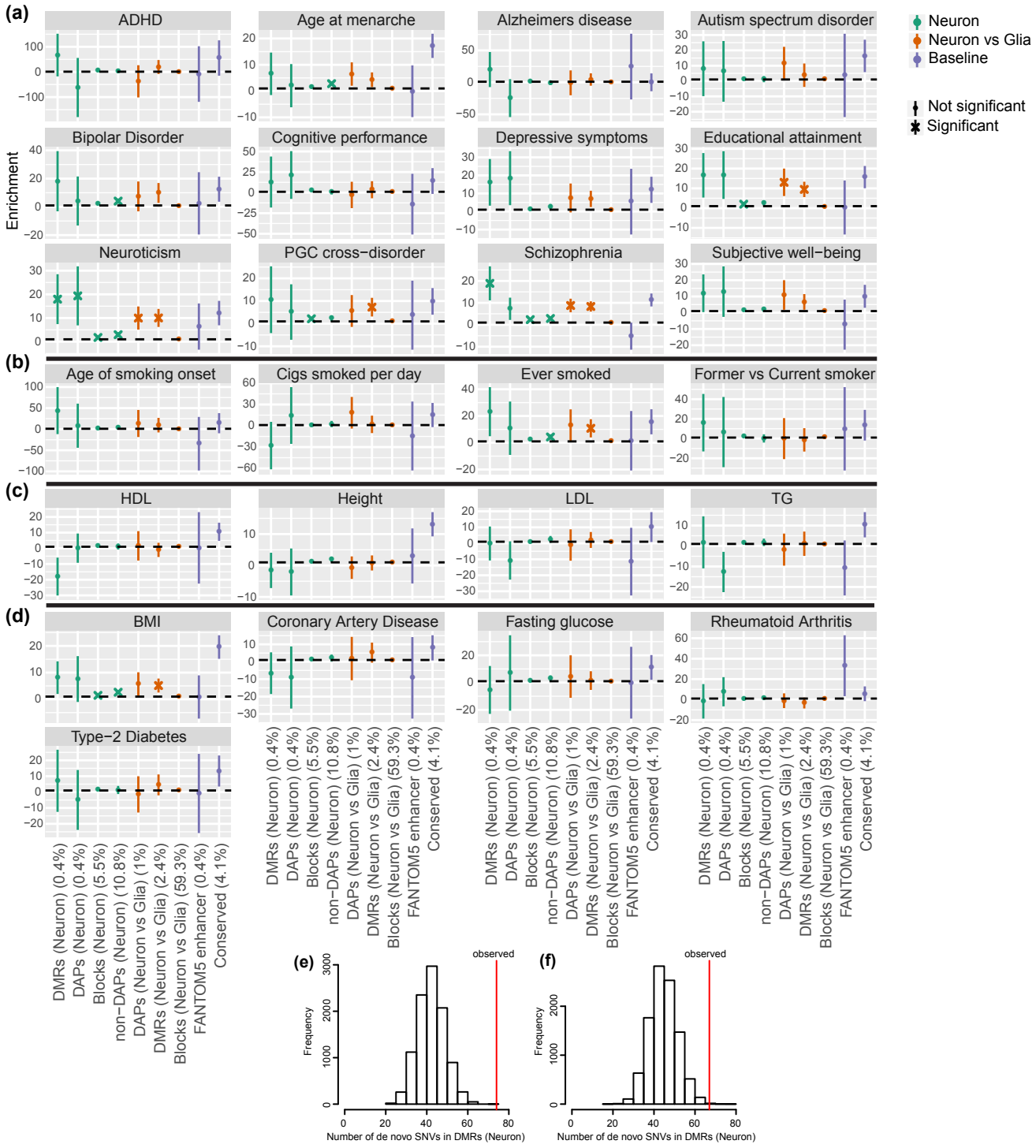
Supplementary Figure 2. Global methylation analysis of nuclei isolated from four different brain regions. Average autosomal (a) CpG and (b) non-CpG methylation of neuronal (NeuN pos; green) and glial (NeuN neg; purple) nuclei isolated from four brain regions: nucleus accumbens (NAcc; n_{pos} = 6, n_{neg} = 6), hippocampus (HC; n_{pos} = 6, n_{neg} = 6), anterior cingulate gyrus (BA24; n_{pos} = 5, n_{neg} = 4), and prefrontal cortex (BA9; n_{pos} = 6, n_{neg} = 6).



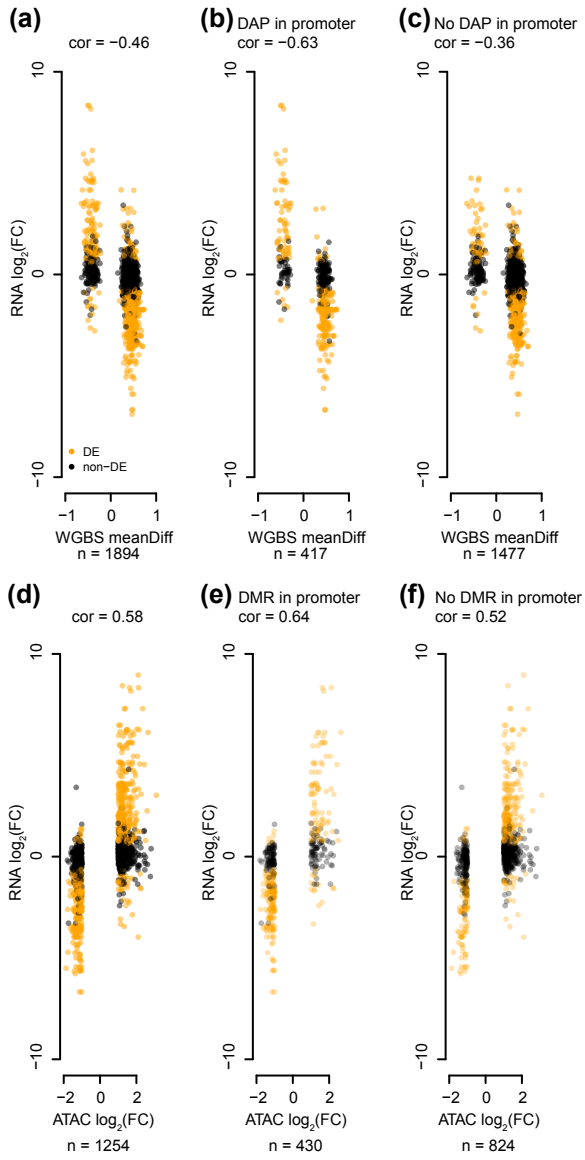
Supplementary Figure 3. Enrichment of neuronal DMRs between nucleus accumbens and prefrontal cortex over genomic features. Log odds ratios and 95% confidence intervals for the enrichment of CpGs within DMRs compared to the rest of the genome in different genomic features (see Methods). DMRs are those from the comparison of neuronal nuclei in the 4 brain regions. (a) Gene models from GENCODE and putative enhancer regions: PC = protein-coding genes; lncRNA = long non-coding RNA genes; union = protein-coding and long non-coding RNA genes. Putative enhancers abbreviated as: Vermunt et al.[40] = promoter-distal H3K27ac-enriched regions showing global characteristics of brain enhancers; FANTOM5[41] = permissive enhancer candidates; Overlap = FANTOM5 enhancers that are additionally overlapped by a H3K27ac-enriched region from Vermunt et al.[40]. (b) CpG islands and related features. (c) ChromHMM annotations (core 15-state model) for four brain regions from the Roadmap Epigenomics Project: Ant. Caudate = Anterior Caudate (similar to nucleus accumbens); Cingulate Gy. = Cingulate Gyrus (similar to BA24); HC Middle = Hippocampus Middle (same region as our HC), DLPFC = Dorsolateral Prefrontal Cortex (same region as BA9).



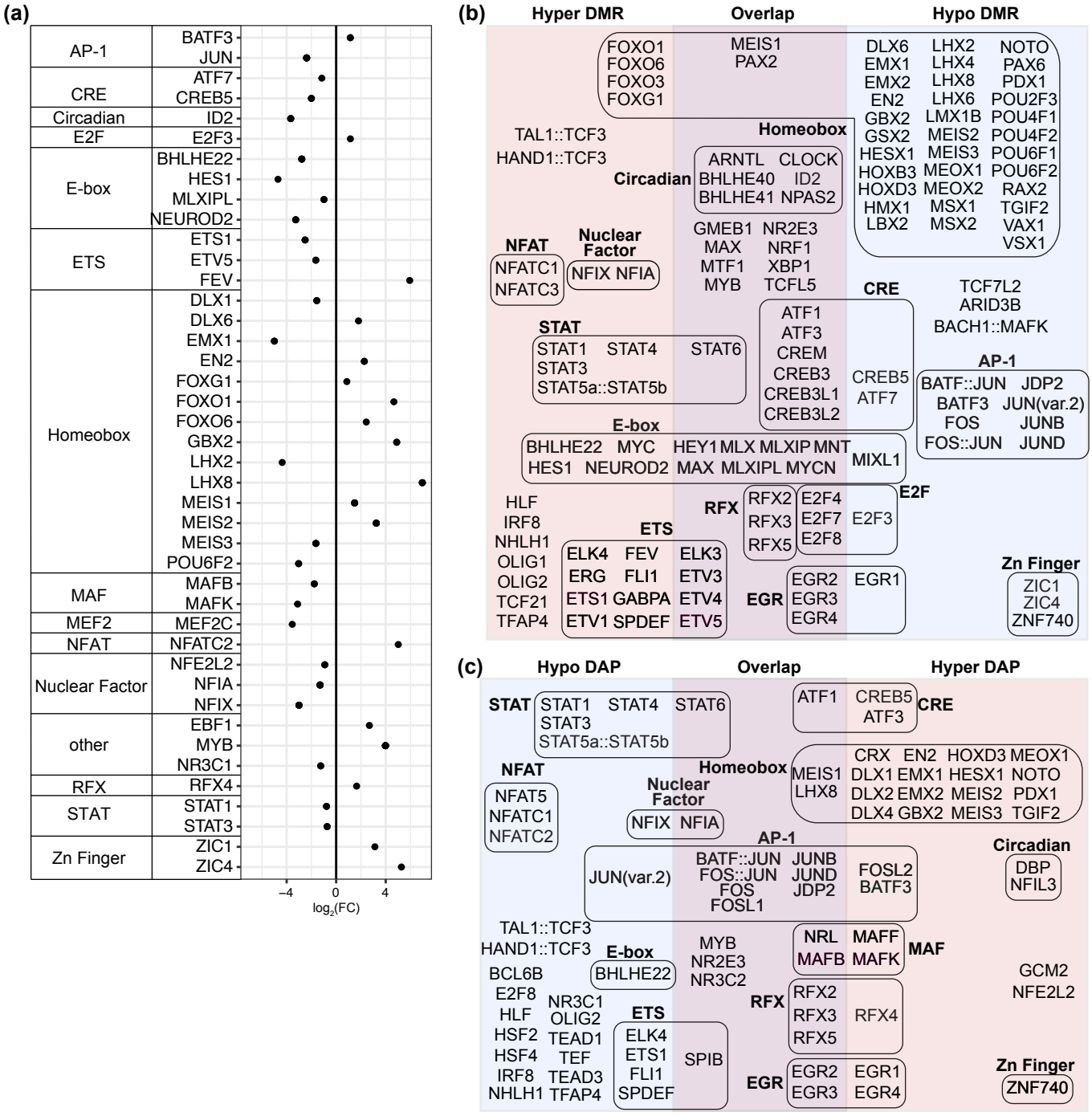
Supplementary Figure 4. Enrichment of ATAC-seq peaks and neuronal DAPs over genomic features. Log odds ratios and 95% confidence intervals for the enrichment over genomic features of (a, b) bases within ATAC-seq peaks compared to the rest of the genome; (c) bases within DAPs compared to the rest of the genome; (d-f) bases within DAPs compared to bases within 'Null-peaks', with the DAPs stratified in (f) by 'hyper' neuronal DAPs, which are more accessible in NAcc than BA9, and 'hypo' neuronal DAPs, which are less accessible in NAcc than BA9 (see Methods). Otherwise as Supplementary Figure 3.



Supplementary Figure 5. Enrichment of disease-associated variants and heritability within DMRs, DAPs, and ATAC peaks. (a-d) Estimates and 95% confidence intervals for the enrichment of explained heritability of GWAS traits using linkage disequilibrium score regression (see Methods). Enrichments within our features ('Neuron' and 'Neuron vs Glia') are contrasted with those enrichments within a 'Baseline' set of features, namely, permissive enhancers from the FANTOM5 project and regions of highly conserved sequence, which were previously shown to be highly enriched for explained heritability across a broad range of GWAS traits[53]. The size of each category is reported as a percentage of the size of the autosomal genome. (a) Neurological traits. (b) Addiction traits. (c) Negative control traits for which we expected no enrichment within our brain region-specific regions. (d) Other GWAS traits we investigated. (e,f) Observed number of germline or somatic de novo mutations within neuronal DMRs compared to that expected by chance in 10,000 simulations in (e) children with autism and (f) Dutch children without autism (see Methods).



Supplementary Figure 6. The relationship between differential expression, methylation, and accessibility in neurons over promoters. (a)-(c) Scatterplots showing Pearson correlation of differential expression with differential methylation in promoters (a), stratified by the gene also having a DAP in a promoter (b), or not having a DAP in a promoter (c). (d)-(f) Scatterplots showing correlation of differential expression with differential accessibility in a promoter (d), stratified by the gene also having a DMR in a promoter (e), or not having a DMR in a promoter (f). Note that the methylation measurements were obtained from samples distinct from accessibility and expression measurements. All panels show a comparison of neuronal nuclei from nucleus accumbens (NAcc) with neuronal nuclei from prefrontal cortex (BA9). Differentially expressed genes are shown in orange.



Supplementary Figure 7. Transcription factor motifs enriched in non-promoter DMRs and DAPs. (a) Gene expression log fold change in a comparison of neurons in nucleus accumbens (NAcc; n = 5) to prefrontal cortex (BA9; n = 6) for transcription factors whose motifs are enriched in DMRs and/or DAPs (regardless of overlap with promoters or not). Only motifs corresponding to transcription factors expressed in both NAcc and BA9 neuronal nuclei are shown. Transcription factor motifs enriched (> 1.25 enrichment over background) in (b) DMRs and (c) DAPs located outside promoter regions. “Hyper DMR” and “Hypo DMR” refer to regions of differential methylation when comparing NAcc to BA9. Similarly, “Hyper DAP” and “Hypo DAP” refer to ATAC peaks that are more, or less accessible, respectively, in NAcc as compared to BA9. Transcription factors are roughly grouped by family (or function).

Descriptions of Supplementary Tables

Miscellaneous notes

Tissue and cell abbreviations

Here and throughout the manuscript we use the following tissue abbreviations:

- HC: Hippocampus
- BA9: Dorsolateral prefrontal cortex
- BA24: Anterior cingulate cortex
- NAcc: Nucleus accumbens

For flow sorted data, we use the following suffixes:

- pos: $NeuN^+$ cells
- neg: $NeuN^-$ cells

For example, BA9_pos means $NeuN^+$ from the dorsolateral prefrontal cortex.

Genome build and coordinates

Here and throughout the manuscript we use:

- The GRCh37/hg19 build of the human reference genome is used in all analyses
    - We use 1-based genomic coordinates
    - We use closed genomic intervals. E.g., chr1:10-12 means positions 10, 11, 12 on chr1
- GENCODE v19 (protein-coding transcripts and long non-coding RNA transcripts) is used for all gene models (http://www.gencodegenes.org/releases/19.html)
    - The 'GENCODE v19 GTF' file was downloaded from ftp://ftp.sanger.ac.uk/pub/gencode/Gencode_human/release_19/gencode.v19.annotation.gtf.gz

Supplementary_Table_01.Demographics.xlsx

Demographic information for brain donor subjects.

- Individual: Donor ID
- PMI(h): Postmortem interval in hours
- Age(yr:day): Age in years and days at time of death
- Sex: M = Male or F = Female
- Ethnicity: Donor ethnicity
- Sorted Tissues: Semicolon delimited list of cell-sorted tissues available for whole-genome bisulfite-sequencing
- Bulk Tissues: Semicolon delimited list of bulk tissues available for whole-genome bisulfite-sequencing
- ATAC-seq Tissues: Semicolon delimited list of cell-sorted tissues available for ATAC-seq
- RNA-seq Tissues: Semicolon delimited list of cell-sorted tissues available for RNA-seq

Supplementary_Table_02.Summary_of_Unsorted_WGBS.csv

Summary of whole-genome bisulfite-sequencing data for unsorted samples (n = 27).

- Sample ID: Sample ID in the form <Donor>_<Region>
- Number sequenced PE reads: Number of sequenced paired-end reads
- Number aligned PE reads: Number of aligned paired-end reads
- Alignment rate (%): Percentage of sequenced reads that were aligned
- Number covered CpGs: Number of autosomal CpGs in the reference genome covered by at least one read
- Covered CpGs (%): Percentage of autosomal CpGs in the reference genome covered by at least one read
- Mean depth: Average sequencing depth of autosomal CpGs in the reference genome covered by at least one read
- Bisulfite conversion rate (%): Estimated bisulfite-conversion rate from spike-in unmethylated lambda DNA

Supplementary_Table_03.Summary_of_Sorted_Sequencing.csv

Summary of whole-genome bisulfite-sequencing data for flow sorted samples (n = 45).

- Sample ID: Sample ID in the form <Donor>_<Region>_<NeuNstatus>
- Number sequenced PE reads: Number of sequenced paired-end reads
- Number aligned PE reads: Number of aligned paired-end reads
- Alignment rate (%): Percentage of sequenced reads that were aligned
- Number covered CpGs: Number of autosomal CpGs in the reference genome covered by at least one read
- Covered CpGs (%): Percentage of autosomal CpGs in the reference genome covered by at least one read
- Mean depth: Average sequencing depth of autosomal CpGs in the reference genome covered by at least one read
- Bisulfite conversion rate (%): Estimated bisulfite-conversion rate from spike-in unmethylated lambda DNA

Supplementary_Table_04.Summary_of_DMRs_and_blocks_from_F-stat_Analysis.xlsx

Summary of BLOCKs and DMRs from F-stat analysis.

- Reference autosomal genome: The autosomal reference genome (hg19)
- Analyzed autosomal genome (unsorted): The set of autosomal CpGs used in the differential methylation analysis of the unsorted bulk tissue samples
- Analyzed autosomal genome (sorted): The set of autosomal CpGs used in the differential methylation analysis of the sorted tissue samples
- pos vs neg: Differential methylation analysis comparing $NeuN^+$ and $NeuN^-$ samples
- pos: Differential methylation analysis comparing $NeuN^+$ samples
- Non-NAcc pos: Differential methylation analysis comparing BA9_pos, BA24_pos, and HC_pos samples
- neg: Differential methylation comparing $NeuN^-$ samples
- Unsorted bulk tissue: Differential methylation analysis comparing unsorted bulk tissue samples

Columns:

- N: Number of features
- Number of CpGs: Number of CpGs in feature
- Genomic size (bp): Size of feature in base pairs
- Median size (bp): Median size of feature in base pairs

Supplementary_Table_05.DMRs_pos.csv

Coordinates and summary statistics of DMRs identified by F-stat analysis of $NeuN^+$ samples (n = 23).

- chromosome: Chromosome of DMR
- start: Start coordinate of DMR
- end: End coordinates of DMR

- n: Number of CpGs in DMR
- areaStat: The ‘area’ of the F-statistic; equal to the sum of the F-statistics for the individual CpGs
- BA24_pos: Average methylation in DMR of BA24_pos samples
- BA9_pos: Average methylation in DMR of BA9_pos samples
- HC_pos: Average methylation in DMR of HC_pos samples
- NAcc_pos: Average methylation in DMR of NAcc_pos samples
- NAcc_pos_vs_BA9_pos: Is this labeled a DMR between NAcc_pos and BA9_pos samples by our annotation pipeline (TRUE or FALSE)
- perm_P: P-value based on 1,000 permutations

Supplementary_Table_06.DMRs_neg.csv

Coordinates and summary statistics of DMRs identified by F-stat analysis of $NeuN^-$ samples (n = 22).

- chromosome: Chromosome of DMR
- start: Start coordinate of DMR
- end: End coordinates of DMR
- n: Number of CpGs in DMR
- areaStat: The ‘area’ of the F-statistic; equal to the sum of the F-statistics for the individual CpGs
- BA24_neg: Average methylation in DMR of BA24_neg samples
- BA9_neg: Average methylation in DMR of BA9_neg samples
- HC_neg: Average methylation in DMR of HC_neg samples
- NAcc_neg: Average methylation in DMR of NAcc_neg samples
- perm_P: P-value based on 1,000 permutations

Supplementary_Table_07.GREAT.xlsx

Results from GREAT analysis using the whole genome (hg19) as background. Settings used are: Basal+extension (constitutive 5.0 kb upstream and 1.0 kb downstream, up to 100.0 kb max extension) with curated regulatory domains included. Gene Ontology (GO) terms returned must be significant by both the binomial and hypergeometric tests using the multiple hypothesis correction false discovery rate (FDR) ≤ 0.05 whose binomial fold enrichment is at least 2.

Each sheet in this workbook is the GREAT results for a single set of regions:

- Hypo_DMRs_NAvALL: Hypomethylated NAcc neuronal DMRs
- Hyper_DMRs_NAvALL: Hypermethylated NAcc neuronal DMRs
- Non-NAcc_POS_DMRs: Neuronal DMRs in subset analysis of neurons from BA9, BA24, and HC
- DAPs: Neuronal DAPs
- Hyper_DAPs: Neuronal DAPs with increased accessibility in NAcc compared to BA9
- Hypo_DAPs: Neuronal DAPs with decreased accessibility in NAcc compared to BA9

- Hypo_NAvBA9_DMRs: Hypomethylated neuronal DMRs in NAcc vs. BA9 comparison
- Hyper_NAvBA9_DMRs: Hypermethylated neuronal DMRs in NAcc vs. BA9 comparison

The columns in each spreadsheet are:
- Ontology: Ontology source
- Term_Name: Term identifier from the ontology
- Binom_Rank: Ordinal rank of the binomial p-value compared to the p-values of other annotations
- Binom_Raw_P-Value: Uncorrected p-value from the binomial test over genomic regions
- Binom_FDR_Q-Val: False discovery rate q-value of the binomial test p-values
- Binom_Fold_Enrichment: (Observed / Expected)-fold enrichment of number of genomic regions in the test set with the annotation from the binomial test
- Binom_Observed_Region_Hits: Actual number of genomic regions in the binomial test set with the annotation
- Binom_Region_Set_Coverage: The proportion of all genomic regions in the binomial test set that lie in the regulatory domain of a gene with the annotation
- Hyper_Rank: Ordinal rank of the hypergeometric p-value compared to the p-values of other annotations
- Hyper_FDR_Q-Val: False discovery rate q-value of the hypergeometric test p-values
- Hyper_Fold_Enrichment: (Observed / Expected)-fold enrichment of number of genomic regions in the test set with the annotation from the hypergeometric test
- Hyper_Observed_Gene_Hits: Actual number of genomic regions in the hypergeometric test set with the annotation
- Hyper_Total_Genes: Number of genes in the genome with the annotation used in the hypergeometric test
- Hyper_Gene_Set_Coverage: Proportion of all genes with the annotation that are tagged by the test set

Supplementary_Table_08.Enrichr.xlsx

Results from Enrichr analysis of several genes sets against ‘GO Biological Process’ and ‘KEGG Pathway’ gene sets libraries. Only results with an adjusted P-value < 0.05 are reported.

Each sheet in this workbook is the Enrichr results for a single gene set:
- Genes_with_hypo_pDMR: Genes with a hypomethylated NAcc neuronal DMR in a promoter
- Genes_with_hyper_pDMR: Genes with a hypermethylated NAcc neuronal DMR in a promoter
- Genes_within_neuronal_blocks: Genes wholly within neuronal blocks of differential methylation
- Genes_downregulated: Genes downregulated in NAcc vs. BA9 in neurons

- Genes_upregulated: Genes upregulated in NAcc vs. BA9 in neurons

The columns in each spreadsheet are (as described in http://amp.pharm.mssm.edu/Enrichr/help#background):

- Analysis: The gene set library: 'GO Biological Process' or 'KEGG Pathway'
- Term: GO Biological Process or KEGG Pathway term
- Overlap: Input genes associated with term / all genes associated with term
- P-value: Computed using a standard statistical method used by most enrichment analysis tools: Fisher's exact test or the hypergeometric test. This is a binomial proportion test that assumes a binomial distribution and independence for probability of any gene belonging to any set
- Adjusted_P-value: An adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing
- Z-score: Computed using a modification to Fisher's exact test by computing a z-score for deviation from an expected rank. The rank based ranking is derived from running the Fisher exact test for many random gene sets in order to compute a mean rank and standard deviation from the expected rank for each term in the gene-set library and finally calculating a z-score to assess the deviation from the expected rank
- Combined_Score: Combination of the p-value and z-score calculated by multiplying the two scores [(c = log(p) * z). Where c is the combined score, p is the p-value computed using Fisher's exact test, and z is the z-score computed to assess the deviation from the expected rank.
- Genes: Genes from the input that were found to be associated with that term

Supplementary_Table_09.Non-NAcc_DMRs.csv

Coordinates and summary statistics of DMRs identified by F-stat analysis of NeuN-positive, non-NAcc samples (n = 17).

- chromosome: Chromosome of DMR
- start: Start coordinate of DMR
- end: End coordinates of DMR
- n: Number of CpGs in DMR
- areaStat: The 'area' of the F-statistic; equal to the sum of the F-statistics for the individual CpGs
- BA24_pos: Average methylation in DMR of BA24_pos samples
- BA9_pos: Average methylation in DMR of BA9_pos samples
- HC_pos: Average methylation in DMR of HC_pos samples
- perm_P: P-value based on 1,000 permutations

Supplementary_Table_10.blocks_pos.csv

Coordinates and summary statistics of blocks identified by F-stat analysis of NeuN-positive samples (n = 23).

- chromosome: Chromosome of block
- start: Start coordinate of block

- end: End coordinates of block
- n: Number of CpGs in block
- areaStat: The ‘area’ of the F-statistic; equal to the sum of the F-statistics for the individual CpGs
- BA24_pos: Average methylation in block of BA24_pos samples
- BA9_pos: Average methylation in block of BA9_pos samples
- HC_pos: Average methylation in block of HC_pos samples
- NAcc_pos: Average methylation in block of NAcc_pos samples
- perm_P: P-value based on 1,000 permutations
- coversPCGene: Does the block cover the entirety of a protein-coding gene(s)

Supplementary_Table_11.DMRs.csv

Coordinates and summary statistics of DMRs identified by F-stat analysis of NeuN-positive and NeuN-negative samples (n = 45).

- chromosome: Chromosome of DMR
- start: Start coordinate of DMR
- end: End coordinates of DMR
- n: Number of CpGs in DMR
- areaStat: The ‘area’ of the F-statistic; equal to the sum of the F-statistics for the individual CpGs
- BA24_neg: Average methylation in DMR of BA24_neg samples
- BA24_pos: Average methylation in DMR of BA24_pos samples
- BA9_neg: Average methylation in DMR of BA9_neg samples
- BA9_pos: Average methylation in DMR of BA9_pos samples
- HC_neg: Average methylation in DMR of HC_neg samples
- HC_pos: Average methylation in DMR of HC_pos samples
- NAcc_neg: Average methylation in DMR of NAcc_neg samples
- NAcc_pos: Average methylation in DMR of NAcc_pos samples
- pos_vs_neg: Is this labeled a DMR between pos and neg samples by our annotation pipeline (TRUE or FALSE)
- perm_P: P-value based on 1,000 permutations

Supplementary_Table_12.blocks.csv

Coordinates and summary statistics of blocks identified by F-stat analysis of NeuN-positive and NeuN-negative samples (n = 45).

- chromosome: Chromosome of block
- start: Start coordinate of block
- end: End coordinates of block
- n: Number of CpGs in block
- areaStat: The ‘area’ of the F-statistic; equal to the sum of the F-statistics for the individual CpGs
- BA24_neg: Average methylation in block of BA24_neg samples
- BA24_pos: Average methylation in block of BA24_pos samples
- BA9_neg: Average methylation in block of BA9_neg samples

- BA9_pos: Average methylation in block of BA9_pos samples
- HC_neg: Average methylation in block of HC_neg samples
- HC_pos: Average methylation in block of HC_pos samples
- NAcc_neg: Average methylation in block of NAcc_neg samples
- NAcc_pos: Average methylation in block of NAcc_pos samples
- pos_vs_neg: Is this labeled a block between pos and neg samples by our annotation pipeline (TRUE or FALSE)
- perm_P: P-value based on 1,000 permutations

Supplementary_Table_13.Novel_DMRs_Compared_to_Published_Data.xlsx

Summary of novel NeuN-positive vs. NeuN-negative DMRs from F-stat analysis of NeuN-positive and NeuN-negative samples (n = 45). See Methods for how novelty was determined.

- Reference autosomal genome: The autosomal reference genome (hg19)
- Analyzed autosomal genome (sorted): The set of autosomal CpGs used in the differential methylation analysis of the sorted tissue samples
- Novel small DMRs: Novel DMRs identified in our analysis of NeuN-positive and NeuN-negative samples
- Novel non-NAcc pos: Novel DMRs identiefied in our differential methylation analysis comparing BA9_pos, BA24_pos, and HC_pos samples

Columns:

- N: Number of features
- Number of CpGs: Number of CpGs in feature
- Genomic size (bp): Size of feature in base pairs
- Median size (bp): Median size of feature in base pairs

Supplementary_Table_14.RNA-seq.NAcc_posvsBA9_pos.csv

Results of RNA-seq differential gene analysis comparing NAcc_pos samples to BA9_pos samples (n = 11).

- gene_id: Ensembl gene ID in GENCODE v19 GTF file
- logFC: Estimate of the log2-fold-change corresponding to NAcc_pos vs. BA9_pos comparison
- BH_P: Benjamini-Hochberg adjusted P-value from the hypothesis test that logFC is zero

Supplementary_Table_15.RNA-seq.NAcc_negvsBA9_neg.csv

Results of RNA-seq differential gene analysis comparing NAcc_neg samples to BA9_neg samples (n = 9).

- gene_id: Ensembl gene ID in GENCODE v19 GTF file
- logFC: Estimate of the log2-fold-change corresponding to NAcc_neg vs. BA9_neg
- BH_P: Benjamini-Hochberg adjusted P-value for the hypothesis test that logFC is zero

Supplementary_Table_16.ATAC-seq.NAcc_pos_vs_BA9_pos.csv

Results of ATAC-seq differential accessibility analysis comparing NAcc_pos samples to BA9_pos samples (n = 11).

- chromosome: Chromosome of peak
- start: Start coordinate of peak
- end: End coordinate of peak
- logFC: Estimate of the log2-fold-change corresponding to NAcc_pos vs. BA9_pos
- BH_P: Benjamini-Hochberg adjusted P-value for the hypothesis test that logFC is zero

Supplementary_Table_17.ATAC-seq.NAcc_neg_vs_BA9_neg.csv

Results of ATAC-seq differential accessibility analysis comparing NAcc_neg samples to BA9_neg samples (n = 11).

- chromosome: Chromosome of peak
- start: Start coordinate of peak
- end: End coordinate of peak
- logFC: Estimate of the log2-fold-change corresponding to NAcc_pos vs. BA9_pos
- BH_P: Benjamini-Hochberg adjusted P-value for the hypothesis test that logFC is zero

Supplementary_Table_18.Summary_of_ATAC-seq.csv

Summary of ATAC-seq samples (n = 22).

- Sample ID: Sample ID in the form <Donor>_<Tissue>_<NeuNstatus>
- Number sequenced PE reads: Number of sequenced paired-end reads
- Alignment rate (%): Bowtie2 reported alignment rate
- Duplicate rate (%): Percentage of reads marked as potential PCR duplicates by Picard's MarkDuplicates tool
- Mitochondrial contamination (%): Percentage of reads aligned to mitochondrial chromosome

Supplementary_Table_19.Summary_of_RNA-seq.csv

Summary of RNA-seq samples (n = 20).

- Sample ID: Sample ID in the form <Donor>_<Tissue>_<NeuNstatus>
- Number sequenced PE reads: Number of sequenced paired-end reads
- Number of quasi-mapped PE reads: Salmon reported number of quasi-mapped reads
- Alignment rate (%): Salmon reported quasi-mapping rate

Supplementary_Table_20.Haystack_Promoter_Overlap.xlsx

Results from Haystack analysis of different subsets of DMRs and DAPs that overlap promoters (defined as +/- 2 kb from transcription start site). Only results with a q-value < 0.05 and an enrichment ratio > 1.25 are reported.

Each sheet in this workbook is the output for a particular input, where the input is the set of genomic coordinates defining the DMRs/DAPs:

- hyper_DMRs_overlap_promoters: Hypermethylated NAcc neuronal DMRs that overlap a gene promoter
- hypo_DMRs_overlap_promoters: Hypomethylated NAcc neuronal DMRs that overlap a gene promoter
- hypo_DAPs_overlap_promoters: Neuronal DAPs that are less accessible in NAcc than BA9 and overlap a gene promoter
- hyper_DAPs_overlap_promoters: Neuronal DAPs that are more accessible in NAcc than BA9 and overlap a gene promoter

The columns in each spreadsheet are:

- Motif ID: ID of motif in JASPAR database
- Motif Name: Name of motif in JASPAR database
- Presence in Target: Presence of motif in input
- Presence in BG: Presence of motif in background
- Ratio: Enrichment ratio of ‘Presence in Target’ to ‘Presence in BG’
- p-value: P-value from Fisher’s exact test of whether motif is enriched in target compared to background
- q-value: False discovery rate Q-value computed from ‘p-value’
- Expressed: “YES” if gene had at least 1 cpm in at least 4 libraries (the size of the smallest group of samples) (24,161 / 33,351 genes)
- DEG: “YES” if gene was found to be differentially expressed between NAcc and BA9

Supplementary_Table_21.Haystack_Not_Promoter_Overlap.xlsx

Results from Haystack analysis of different subsets of DMRs and DAPs that do not overlap promoter regions (defined as +/- 2kb from transcription start site).

Each sheet in this workbook is the Haystack results for a single set of regions:

- hyper_DMRs_not_overlap_promoter: Hypermethylated NAcc neuronal DMRs that do not overlap a gene promoter
- hypo_DMRs_not_overlap_promoter: Hypomethylated NAcc neuronal DMRs that do not overlap a gene promoter
- hypo_DAPs_not_overlap_promoters: Neuronal DAPs that are less accessible in NAcc than BA9 and do not overlap a gene promoter
- hyper_DAPs_not_overlap_promoters: Neuronal DAPs that are more accessible in NAcc than BA9 and do not overlap a gene promoter

The columns in each spreadsheet are:

- Motif ID: ID of motif in JASPAR database

- Motif Name: Name of motif in JASPAR database
- Presence in Target: Presence of motif in input
- Presence in BG: Presence of motif in background
- Ratio: Enrichment ratio of ‘Presence in Target’ to ‘Presence in BG’
- p-value: P-value from Fisher’s exact test of whether motif is enriched in target compared to background
- q-value: False discovery rate Q-value computed from ‘p-value’
- Expressed: “YES” if gene had at least 1 cpm in at least 4 libraries (the size of the smallest group of samples) (24,161 / 33,351 genes)

DEG: “YES” if gene was found to be differentially expressed between NAcc and BA9

Supplementary_Table_22.ATAC-seq.NAcc_neg_vs_BA9_neg.csv

Results of ATAC-seq differential accessibility analysis comparing $NeuN^+$ samples to $NeuN^-$ samples (n = 22).

- chromosome: Chromosome of peak
- start: Start coordinate of peak
- end: End coordinate of peak
- logFC: Estimate of the log2-fold-change corresponding to pos vs. neg
- BH_P: Benjamini-Hochberg adjusted P-value for the hypothesis test that logFC is zero