Parkinson-Associated *SNCA* Enhancer Variants Revealed by Open Chromatin in Mouse Dopamine Neurons

Sarah A. McClymont,[1] Paul W. Hook,[1] Alexandra I. Soto,[2] Xylena Reed,[1] William D. Law,[1] Samuel J. Kerans,[1] Eric L. Waite,[1] Nicole J. Briceno,[1] Joey F. Thole,[1] Michael G. Heckman,[3] Nancy N. Diehl,[3] Zbigniew K. Wszolek,[4] Cedric D. Moore,[5] Heng Zhu,[5] Jennifer A. Akiyama,[6] Diane E. Dickel,[6] Axel Visel,[6,7,8] Len A. Pennacchio,[6,7,9] Owen A. Ross,[2,10,11] Michael A. Beer,[1,12] and Andrew S. McCallion[1,13,14,*]

The progressive loss of midbrain (MB) dopaminergic (DA) neurons defines the motor features of Parkinson disease (PD), and modulation of risk by common variants in PD has been well established through genome-wide association studies (GWASs). We acquired open chromatin signatures of purified embryonic mouse MB DA neurons because we anticipated that a fraction of PD-associated genetic variation might mediate the variants' effects within this neuronal population. Correlation with >2,300 putative enhancers assayed in mice revealed enrichment for MB cis-regulatory elements (CREs), and these data were reinforced by transgenic analyses of six additional sequences in zebrafish and mice. One CRE, within intron 4 of the familial PD gene *SNCA*, directed reporter expression in catecholaminergic neurons from transgenic mice and zebrafish. Sequencing of this CRE in 986 individuals with PD and 992 controls revealed two common variants associated with elevated PD risk. To assess potential mechanisms of action, we screened >16,000 proteins for DNA binding capacity and identified a subset whose binding is impacted by these enhancer variants. Additional genotyping across the *SNCA* locus identified a single PD-associated haplotype, containing the minor alleles of both of the aforementioned PD-risk variants. Our work posits a model for how common variation at *SNCA* might modulate PD risk and highlights the value of cell-context-dependent guided searches for functional non-coding variation.

Introduction

Parkinson disease (PD) is a common progressive neurodegenerative disorder characterized by preferential and extensive degeneration of dopaminergic (DA) neurons in the *substantia nigra*.[1,2] This loss of midbrain (MB) DA neurons disrupts the nigrostriatal pathway and results in the movement phenotypes observed in PD. Although this disorder affects approximately 1% of people over 70 years old worldwide,[3] the mechanisms underlying genetic risk of sporadic PD in the population remain largely unknown. Familial cases of PD with known pathogenic mutations are better understood but account for ≤10% of PD cases.[4]

The α-synuclein gene (*SNCA*) is commonly disrupted in familial PD through missense mutations predicted to promote misfolding[5–7] or genomic multiplications, resulting in an overexpression paradigm.[8] The *SNCA* locus has also been shown by genome-wide association studies (GWASs) to harbor common variants, modulating risk of sporadic PD.[9] In the same way, common variants at more than 40 additional loci have been implicated in PD,[10] but the causal variants which are responsible for elevating risk, and the genes they modulate in doing so, remain largely undetermined.

That most GWAS-implicated variants are non-coding[11] is a major source of this uncertainty, obstructing the identification of: (1) the causative variant at a locus; (2) the context in which a variant is acting; and (3) the mechanism by which a variant asserts its effect on disease risk.

GWASs are inherently biologically agnostic, and their exploitation of linkage disequilibrium (LD) structure frequently results in the implication of many variants at a particular locus, but no one variant is prioritized over those in LD. One method to prioritize non-coding variants is to examine the chromatin status at that locus.[11–13] Accessible chromatin is more likely to be functional, and variants therein might impact that activity more so than those variants residing in inaccessible chromatin. Recent studies have prioritized neuropsychiatric variants through examination of the chromatin status of iPSC-derived neurons or post-mortem whole brain tissues.[14,15] However, chromatin accessibility is dynamic and often varies across cell types and developmental time; therefore, understanding and isolating the *in vivo* cellular context in which

[1]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [2]Department of Neuroscience, Mayo Clinic, Jacksonville, FL 32224, USA; [3]Division of Biomedical Statistics and Informatics, Mayo Clinic, Jacksonville, FL 32224, USA; [4]Department of Neurology, Mayo Clinic, Jacksonville, FL 32224, USA; [5]Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [6]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; [7]US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA; [8]School of Natural Sciences, University of California, Merced, CA 95343, USA; [9]Comparative Biochemistry Program, University of California, Berkeley, CA 94720, USA; [10]Neuroscience Track, Mayo Graduate School, Jacksonville, FL 32224, USA; [11]Department of Clinical Genomics, Mayo Clinic, Jacksonville, FL 32224, USA; [12]Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [13]Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [14]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA
*Correspondence: andy@jhmi.edu

https://doi.org/10.1016/j.ajhg.2018.10.018.

variation acts is critical to increasing our ability to prioritize variants and query their methods of action.[11,16–18]

By exploiting the preferential vulnerability of MB DA neurons in PD, we have prioritized DA neurons as the biological context in which a fraction of PD-associated variants are likely to act. DA neurons in other brain regions, such as the forebrain (FB), provide a related substrate that is less vulnerable to loss in PD. We sought to use chromatin data from *ex vivo* populations of DA neurons to investigate the contributions of non-coding variation to PD risk. To maximize the specificity of the biological context, we generated chromatin signatures for purified mouse MB and FB DA neurons. We examined the resulting regulatory regions for their ability to direct *in vivo* reporter expression and developed a regulatory sequence vocabulary specific to DA neurons. In doing so, we identified a novel MB DA regulatory element that falls within intron 4 of *SNCA* and demonstrated its ability to direct reporter expression in catecholaminergic neurons from transgenic mice and zebrafish, confirming it to be an enhancer. Furthermore, this enhancer harbors two common variants, falling within a haplotype that we determine to be associated with PD risk. We demonstrate that these enhancer variants impact protein binding, and we propose a model for how the variants and the haplotype at large contribute to *SNCA* regulatory control. This work illustrates the power of cell-context-dependent guided searches for the identification of disease-associated and functional non-coding variation.

Material and Methods

Animal Husbandry

Tg(Th-EGFP)DJ76Gsat mice (Th-EGFP) were generated by the GENSAT project[19] and purchased through the Mutant Mouse Resource and Research Centers Repository. Colony-maintenance matings were between hemizygous male Th-EGFP mice and female Swiss Webster (SW) mice, obtained from Charles River Laboratories. This same mating scheme was used for establishing timed matings and thus generating litters for assay; the day on which the vaginal plug was observed was E0.5. Adult AB zebrafish lines were maintained in system water according to standard methods.[20] All work involving mice and zebrafish (husbandry, colony maintenance, procedures, and euthanasia) were reviewed and preapproved by the institutional care and use committee.

Neural Dissociation and Fluorescence-Activated Cell Sorting

Pregnant SW mice were euthanized at E15.5, and the embryos were removed and immediately placed in chilled Eagle's Minimum Essential Medium (EMEM) on ice. The embryos were decapitated, and the brains were removed into Hank's Balanced Salt Solution without Mg^{2+} and Ca^{2+} (HBSS w/o) on ice. Under a fluorescent microscope, EGFP+ brains were identified and microdissected to yield the desired MB and FB regions. Microdissected regions were placed in fresh HBSS w/o on ice and pooled per litter for dissociation.

Pooled brain regions were dissociated via the Papain Dissociation System (Worthington Biochemical Corporation). The tissue was dissociated in the papain solution for 30 min at 37°C, and gentle trituration was performed every 10 min with a sterile Pasteur pipette. After dissociation, cells were passed through a 40 μm cell strainer into a 50 mL conical tube, centrifuged for 5 min at 300 × g, resuspended in albumin-inhibitor solution containing DNase, applied to a discontinuous density gradient, and centrifuged for 6 min at 70 × g. The resulting cell pellet was resuspended in HBSS with Mg^{2+} and Ca^{2+} and submitted to fluorescence-activated cell sorting (FACS). Aliquots of 50,000 EGFP+ cells were sorted directly into 300 μL HBSS with Mg^{2+} and Ca^{2+} and 10% FBS for ATAC-seq. Aliquots containing ≥50,000 EGFP+ cells were sorted into kit-provided lysis buffer for RNA-seq. This procedure was repeated such that a single aliquot of cells from each region per litter was submitted to either ATAC-seq or bulk RNA-seq, repeated three (ATAC-seq) or four (RNA-seq) times for each region.

ATAC-seq Library Preparation and Quantification

ATAC-seq library preparation generally followed the steps as set out in the original ATAC-seq paper,[21] with minor modifications. Aliquots of 50,000 EGFP+ cells were centrifuged for 5 min at 4°C and 500 × g, washed with 50 μL of chilled PBS, and centrifuged again for 5 mins at 4°C and 500 × g. Next, the cell pellet was resuspended in lysis buffer, as set out in the protocol, and cells were left to lyse for 5 min at 4°C before being centrifuged for 10 min at 4°C at 500 × g. The resulting nuclei pellet was tagmented, as written, using the transposase from the Nextera DNA Library Preparation Kit. After transposition, DNA was purified with the MinElute Reaction Clean-up Kit (QIAGEN) and eluted in 10 μL elution buffer.

The libraries were amplified according to the original ATAC-seq protocol.[21] The qPCR surveillance steps were modified such that the additional number of cycles of amplification were calculated as $^1/_4$ maximum intensity, so as to limit PCR duplication rates in the final libraries. The amplified libraries were purified with Ampure XP beads (Beckman Coulter) according to the Nextera DNA Library Prep Protocol Guide. The libraries were quantified with the Qubit dsDNA High Sensitivity Assay (Invitrogen) in combination with the High Sensitivity DNA Assay (Agilent) on the Agilent 2100 Bioanalyzer.

ATAC-seq Sequencing, Alignment, and Peak Calling

Individual ATAC-seq libraries were sequenced on the Illumina MiSeq to a minimum depth of 15 million, 2 × 75 bp reads per library. A single MB ATAC-seq library was sequenced on the Illumina HiSeq in rapid run mode with 2 × 100 bp reads to a depth of ≥350 million paired-end reads.

The quality of sequencing was evaluated with FastQC (v0.11.2). Reads were aligned to mm9 with Bowtie2[22] (v2.2.5) under –local mode. Reads aligning to the mitochondrial genome; unknown and random chromosomes; and PCR duplicates were removed prior to peak calling (SAMtools[23]). Peaks were called on individual libraries and on a concatenated file combining all MB or all FB ("joint") libraries via MACS2[24] (v2.1.1.20160309) "callpeak" with the following options: –nomodel –nolambda -B -f BAMPE –gsize mm –keep-dup all. Peaks overlapping blacklisted regions that were called by ENCODE and that were in the original ATAC-seq paper were removed.[21,25] Peaks were examined for their genomic distribution via CEAS in the Cistrome pipeline.[26,27] The fragment lengths were extracted from the SAM files and plotted with a custom script. Mouse (mm9) transcriptional start site

(TSS) coordinates were extracted from the UCSC Genome Browser,[28] and deepTools[29] was used for quantifying the pileup of reads over TSSs.

RNA-seq Library Preparation and Quantification

Total RNA was extracted with the Purelink RNA Micro Kit (Invitrogen). After FACS isolation into kit-provided lysis buffer, samples were homogenized, and RNA extraction proceeded according to the manufacturer's recommendations. Total RNA integrity was determined with the RNA Pico Kit (Agilent) on the Agilent 2100 Bioanalyzer. RNA samples were sent to the Sidney Kimmel Comprehensive Cancer Center Next Generation Sequencing Core at Johns Hopkins for library preparation with the Ovation RNA-Seq System V2 (Nugen) and for sequencing.

RNA-seq Sequencing, Alignment, and Transcript Quantification

The libraries were pooled and sequenced on Illumina's HiSeq 2500 in rapid-run mode with 2 × 100 bp reads to an average depth of >90 million reads per library. The quality of sequencing was evaluated via FastQC. FASTQ files were aligned to mm9 with HISAT2[30] (v2.0.1-beta) with –dta specified.

Aligned reads from individual samples were quantified against a reference transcriptome with the Rsubread package[31–33] (v1.22.3) function "featureCounts" with the following options: isPairedEnd = TRUE, requireBothEndsMapped = TRUE, isGTFAnnotationFile = TRUE, and useMetaFeature = TRUE. The GENCODE vM9 GTF was downloaded[34] (date: March 30, 2016) and lifted over from the mm10 genome to the mm9 genome with CrossMap (v0.2.2) under default parameters.[35] This was used for quantification, in which gene-level raw counts were converted to RPKM (reads per kilobase of transcript per million) values and means for each region were calculated.

Relationship between RNA-seq and ATAC-seq

The 1,000 most highly expressed genes and the 1,000 least highly expressed genes (RPKM $\geq$ 1) in both the MB and FB were identified, and their transcriptional start sites (Ensembl) were extracted from the UCSC Table Browser.[28] Intervals of 1, 10, and 100 kb surrounding these TSSs were intersected with the ATAC-seq libraries, and the overlap quantified[36] and plotted. These same TSSs were provided to deepTools,[29] and the ATAC-seq signal over these most highly- and least highly-expressed genes was quantified and plotted. Additionally, the 1,000 highest and lowest ATAC-seq peaks (by q value) were extracted, and the expression of the nearest gene was quantified and plotted as a final metric to relate the RNA-seq and ATAC-seq datasets.

cDNA Synthesis and RT-qPCR for DA Neuron Markers

RNA was extracted with the RNeasy Mini Kit (QIAGEN) after 50,000 cells were sorted directly into Buffer RLT. Aliquots of 50,000 non-fluorescing cells were also collected and processed in parallel. 100 ng of each RNA sample was submitted to first-strand cDNA synthesis with the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen) according to the oligo(dT) method.

Primers (Table S1) were designed with Primer-BLAST[37] under default parameters, except the requirement for exon-exon junction spanning was specified. qPCR was performed with Power SYBR Green Master Mix (Applied Biosystems). Reactions were run in triplicate under default SYBR Green Standard cycle specifications on the Viia7 Real-Time PCR System (Applied Biosystems). Relative quantification followed the $2^{-\Delta\Delta CT}$ method, and results were normalized to *Actb* in the EGFP− aliquot of cells for each region.

Correlation Analysis between Regions and within Replicates

Peaks from all six ATAC-seq libraries and the two "joint" ATAC-seq libraries were concatenated together, sorted on the basis of chromosomal location, merged into a unified peak set,[36] and converted to Simplified Annotation Format. Reads from each BAM file overlapping this unified peak set were quantified with the Rsubread package's "featureCounts" command, with the following options: isPairedEnd = TRUE and requireBothEndsMapped = FALSE. Read counts were normalized for each library via conditional quantile normalization;[38] library size, peak length, and peak GC content were accounted for. Pearson correlation coefficients were calculated from this normalized count matrix and visualized with corrplot, RColorBrewer, and LSD (see Web Resources).

Sequence-Constraint Analysis

Average phastCons[39] were calculated for the "joint" peak file for both the MB and FB libraries with Cistrome.[27] Beforehand, peaks with overlap of exons or promoters (defined here as ± 2,000 bp from the transcriptional start site) were removed. The exon and promoter BED files were downloaded from the UCSC Table Browser[28] (Mouse genome; mm9 assembly; Genes and Gene Predictions; RefSeq Genes track using the table refGene).

Gene Ontology of Nearest Expressed Gene

The Genomic Regions Enrichment of Annotations Tool[40] (GREAT; v3.0.0) predicted the gene ontology (GO) term enrichment in the catalogs. Beforehand, peaks were processed so that (1) peaks overlapping commonly open regions would be removed; (2) the top 20,000 peaks would be selected; (3) peaks would overlap the nearest expressed gene's transcriptional start site (TSS).

First, regions that are commonly open were defined as those regions of the genome that are open in >30% of ENCODE DNase hypersensitivity site (DHS) assays in mouse tissues. These ubiquitously open regions were removed from the peak files. Next, so that the binomial distribution for calculating enrichment was still valid, we limited the number of regions considered by GREAT by only submitting the top 20,000 peaks on the basis of q value.

Finally, in order to limit ourselves to the nearest expressed gene, we supplied a list of the TSSs of the nearest expressed genes, considering only those genes that are used by GREAT.

Only genes that are in this list and for which RPKM >1 were considered to be expressed. The expressed gene nearest to each of the top 20,000 peaks was identified. Each peak is associated with its nearest expressed gene, and to ensure that GREAT only considered these nearest genes for analysis, we submitted these nearest expressed genes' TSSs as a proxy for each peak. We submitted these proxy peaks to GREAT by using the NCBI build 37 (mm9) assembly under the "whole-genome background regions" setting; we used the single nearest gene as the association rule and included curated regulatory domains.

Quantification of Overlap between CRE Catalogs and the VISTA Enhancer Browser

All elements tested *in vivo* were downloaded from the VISTA Enhancer Browser on September 4, 2016. These regions were stratified into those annotated as positive and those annotated as negative. BED coordinates of these regions were extracted and intersected with the ATAC-seq catalogs. Positive regions were further stratified into those with annotations for only forebrain, only midbrain, or only hindbrain; combinations of regions ("multiple regions"); or all three regions ("whole brain"). These categories comprised the "neuronal" category. Other regions that were annotated as positive but not driving expression in any of those three regions were placed in the "non-neuronal" category.

Testing Five Putative CREs for *In Vivo* Reporter Activity

Prioritized regions were amplified (Table S1) from human genomic DNA via PCR and cloned into either pENTR for mouse *lacZ* assays (Invitrogen) or pDONR221 for zebrafish assays (Invitrogen). Sequences were validated, and regions were cloned via LR cloning (Invitrogen) into either an *hsp68-lacZ* vector or pXIG vector, with a TdTomato cassette in place of GFP.

Generation of transgenic mice and E11.5 embryo-staining were performed as previously described[41–43] with FVB strain mice. Embryos expressing the *lacZ* reporter gene were scored and annotated for their expression patterns by multiple curators. For a construct to be considered positive, a minimum of three embryos per construct were required to demonstrate reporter activity in the same tissue. Mouse transient transgenic assays were approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee.

Generation of transgenic zebrafish was performed as previously described[44] in AB zebrafish. Reporter expression patterns were evaluated at both 3 and 5 days post-fertilization (dpf). For a construct to be considered as positive, ≥25% of mosaic embryos had to display reporter activity in one or more anatomical structures. Positive zebrafish were quantified for reporter activity in five anatomical regions (forebrain, midbrain, hindbrain, amacrine cells, and the spinal cord).

Development of Regulatory Vocabulary

We applied the machine-learning algorithm gkm-SVM[45] to our MB and FB catalogs under default settings. We trained gkm-SVM on the sequences underlying the summits ±250 bp of non-ubiquitously open, top 10,000 peaks by q value versus five negative sets, which were matched for GC content, length, and repeat content. Weights across all five tests were averaged for all 10-mers.

All 10-mers with weight ≥1.50 were clustered on the basis of sequence similarity via Starcode;[46] sphere clustering was specified, and the distance was set to 3. clustalOmega[47] aligned the sequences within these clusters, and MEME,[48] under default parameters except for -dna -maxw 12, generated position weight matrices (PWMs) for these aligned clusters. Tomtom,[49] querying the Jolma 2013, JASPAR Core 2014, and Uniprobe mouse databases, identified the top transcription factors corresponding to these PWMs, under default parameters except for -no-ssc -min-overlap 5 -evalue -thresh 10.0.

We used the same procedure to identify transcription factors specifically conveying regulatory potential in the MB library relative to the FB library; except during gkm-SVM training, the positive set was specified as the top 10,000 non-ubiquitously open MB summits, and the negative set was specified to be the top 10,000 non-ubiquitously open FB summits, both ±250 bp.

Transcription-Factor Footprinting

CENTIPEDE[50] was used for footprint identification. Sequences underlying the deeply sequenced MB library peaks, less those that were ubiquitously open, were extracted. FIMO,[51] with options –text –parse-genomic-coord, identified all locations underlying ATAC-seq peaks of the motifs identified above. Additionally, conservation data from 30-way vertebrate phastCons were considered in the CENTIPEDE calculations; for each PWM site, those with a mean conservation score greater than 0.9 were considered. Finally, the BAM file read-end coordinates were adjusted in response to the shift resulting from the transposase insertion.[52] As such, reads were adjusted +4 bp on the positive strand and −5 bp on the negative strand as performed in the original ATAC-seq method.[21]

Genome-wide Read Pileup over Predicted Motif Sites

FIMO, as above, was used for genome-wide identification of all coordinates of the identified motifs. The deepTools'[29] "bamCoverage" tool was run, under default conditions, for conversion of the deeply sequenced MB library BAM to bigwig format. After this, a matrix file was generated with "computeMatrix;" specified options were –referencePoint center -b 1000 -a 1000 -bs 50. Finally, "plotHeatmap" was used for generating plots indicating ATAC-seq read pileup over predicted motif sites.

Intersection of CRE Catalogs and PD-Associated GWAS Variants

Lead SNPs from two of the most recent meta-analyses[9,10] were submitted to rAggr (see Web Resources), and SNPs in LD were identified (1000 Genomes, phase 3, EUR populations; minimum minor-allele frequency [MAF] = 0.05, $r^2 \geq 0.8$; maximum distance = 5,000 kb). These variants were intersected[36] with the CRE catalogs after they were lifted over to hg19 coordinates, and the overlap was extracted and quantified.

In Vivo Validation of the MB-Specific Enhancer

The MB-specific peak was amplified (Table S1) from human genomic DNA via PCR and cloned into pCR8 via TA cloning (Invitrogen). Sequences were validated, and regions were cloned into either an *hsp68-lacZ* vector or a modified pXIG vector, with a TdTomato cassette in place of GFP, via LR cloning (Invitrogen).

For zebrafish transgenesis, the modified pXIG vector was injected into 1- to 2-cell-stage embryos as previously described[44] in AB zebrafish. TdTomato reporter expression was assayed at 72 hr post-fertilization (hpf) and 5 dpf; mosaic embryos positive for TdTomato expression were selected and raised to adulthood, and founders were identified. The progeny of founders were screened at 72 hpf for reporter activity. For mouse transgenesis, the generated *hsp68-lacZ* vector was purified in a double CsCl gradient (Lofstrand Labs), and stable mouse transgenesis was performed in C57BL/6 mice by Cyagen Biosciences. Multiple founder lines were generated. For *lacZ* staining, embryos were collected at E12.5, and mouse brains were isolated at E15.5, P7, P30, and P574. Brains were roughly sectioned in 1 mm sections at P7 and P30. The animals were perfused at P574, and fixed brains were sectioned (200 μm) with a vibratome. Specimens were subsequently fixed for 2 hr on ice in 1% formaldehyde, 0.2% glutaraldehyde, and 0.02% Igepal CA-630 in PBS. After fixation, tissues were permeabilized over the course of three 15 min washes in 2 mM $MgCl_2$ and 0.02% Igepal CA-630 in PBS at room temperature. Embryos and/or tissues were incubated overnight at 37°C in a staining solution containing 320 μg/mL X-Gal in N,N-dimethyl formamide, 12 mM K-ferricyanide, 12 mM K-ferrocyanide, 0.002% Igepal CA-630, and 4 mM $MgCl_2$ in PBS. Specimens were washed twice for 30 min each time in 0.2% Igepal CA-630 in PBS and finally stored in 4% formaldehyde, 100 mM sodium phosphate, and 10% methanol.

Individuals with and without PD-Sequencing and Genotyping at *SNCA*

986 individuals with PD and 992 controls who were all seen at the Mayo Clinic in Jacksonville, FL were sequenced across the putative enhancer and genotyped for 25 variants across the *SNCA* locus. The variants chosen for genotyping were those identified during the sequencing of the enhancer and a subset of those identified by Guella et al.[53] For PD-affected individuals, median age at blood draw was 69 years (range: 28–97 years), median age at PD onset was 67 years (range: 28–97 years), and 631 subjects (64.0%) were male. Median age at blood draw for control individuals was 67 years (range: 18–92 years), and 415 subjects (41.8%) were male. Individuals with PD were diagnosed according to standard clinical criteria.[54] All subjects are unrelated, non-Hispanic Caucasians of European descent. The Mayo Clinic Institutional Review Board approved the study, and all subjects provided written informed consent.

Genomic DNA was extracted from whole blood with the Autogen FlexStar. Sanger sequencing of the enhancer region was performed bidirectionally with the ABI 3730xl DNA analyzer (Applied Biosystems) according to standard protocols. Sequence data were analyzed with SeqScape (v2.5; Applied Biosystems). Statistical analyses were performed with both SAS and R (see Web Resources). Of the variants identified within the enhancer, only those with a MAF greater than 5% were evaluated for association with PD in single-variant analysis. Associations between individual variants and PD were evaluated with logistic regression models, adjusted for both sex and age at blood draw, and, when variants were considered, under an additive model (i.e., the effect of each additional minor allele was evaluated). Odds ratios and 95% confidence intervals were estimated. A Bonferroni correction for multiple testing was utilized in single-variant analysis because of the four common variants that were evaluated for association with PD, after which p values ≤ 0.0125 were considered as statistically significant.

Genotyping of the 25 SNPs across the *SNCA* locus was performed with the iPLEX Gold protocol on the MassARRAY System and analyzed in TYPER 4.0 software (Agena Bioscience). For the 25 SNPs genotyped across the *SNCA* locus, all genotype call rates were >95%, and there was no evidence for departure from Hardy-Weinberg equilibrium (all χ^2 p values >0.05 after Bonferroni correction). Haplotype frequencies in cases and controls were estimated with the haplo.stats package function "haplo.group" (see Web Resources). The haplo.score function was used to perform score tests for association evaluating the relationships between haplotypes and the risk of PD.[55] Tests were adjusted for both sex and age at blood draw, haplotypes occurring in less than 1% of subjects were excluded, and only individuals with no missing genotype calls for any variants were included. A Bonferroni correction for multiple testing was applied as a result of the 12 different common haplotypes that were observed and tested for association with PD risk, after which p values ≤ 0.0042 were considered as statistically significant.

LD structure and r^2 values at the *SNCA* locus in the 1000 Genomes EUR population were extracted from LDlink[56] with the LDmatrix tool and plotted with R. The chromatin structure at *SNCA* was extracted from the 3D Genome Browser, and POLR2A binding in MCF-7 cells was examined at the SNCA promoter.

Protein Array Testing Differential Binding

HuProt v3.1 human proteome microarrays (Grace Bio-Labs) containing >16,000 unique proteins representing 12,586 genes (CDI laboratories)[58] were blocked with 25 mM HEPES (pH 8.0), 50 mM potassium glutamate, 8 mM $MgCl_2$, 3 mM dithiothreitol (DTT), 10% glycerol, 0.1% Triton X-100, and 3% BSA on an orbital shaker at 4°C for ≥ 3 hr. Allele-specific protein-DNA binding interactions were identified through dye-swap competition of major and minor alleles labeled with either Cy3 or Cy5. DNA fragments for dbSNP: rs2737024 and dbSNP: rs2583959 were synthesized such that each allele was flanked by 15 nucleotides of the upstream and downstream sequence and there was a common priming site at the 3′ end (Table S1).

We created the dsDNA fragments by separately annealing a primer containing a Cy3 or Cy5 label and adding Klenow (New England Biolabs) with dNTP to fill in the complementary strand for each allele.[59] Cy3-labeled major allele was mixed with Cy5-labeled minor allele (each at 40 nM) in 1× hybridization buffer (10 mM Tris-Cl [pH 8.0], 50 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 5% glycerol, 10 μM $ZnCl_2$, and 3 mg/mL BSA) and added to an array; dyes were then swapped for each allele, and the mixture was then added to a second array. DNA was allowed to bind overnight at 4°C on an orbital shaker with protection from light. Chips were washed once with cold 1× Tris-buffered saline-Triton solution (0.1% Triton X-100) for 5 min at 4°C, rinsed, and dried in the centrifuge. Cy5 and Cy3 images were taken separately on a Genepix 4000B scanner and, after alignment to the GAL file, individual spot intensities were extracted with Genepix Pro software.

Allele-specific interactions were identified through dye-swap analysis. The ratio of major to minor allele binding was calculated from the duplicate spot-average, median-foreground signal for each protein according to the following equation:

$$log_2\sqrt{\frac{Cy3_{major} * Cy5_{major}}{Cy3_{minor} * Cy5_{minor}}}$$

Mean intensity was calculated from the average foreground signal for the Cy3 and Cy5 channels of the major and minor alleles. MA plots were made for each allele using the calculated mean intensity and the log ratio of the major to the minor allele.

Results

ATAC-seq Identifies Open Chromatin in MB and FB DA Neurons

To identify open chromatin regions (OCRs) in DA neurons, we performed ATAC-seq[21] on ~50,000 FACS-isolated cells (per replicate) from microdissected regions of the MB and FB of embryonic day 15.5 (E15.5) Tg(Th-EGFP)DJ76Gsat BAC transgenic mice[60] (Figure 1A). EGFP, expressed under the control of the tyrosine hydroxylase (Th) locus, labels catecholaminergic neurons (i.e.: DA, noradrenergic, and adrenergic neurons) in this mouse line. To confirm capture of the corresponding catecholaminergic neurons, we performed RT-qPCR on the isolated reporter-labeled cells and established them to be enriched for DA neuronal markers in comparison to unlabeled populations from the same dissected tissues (Figure S1).

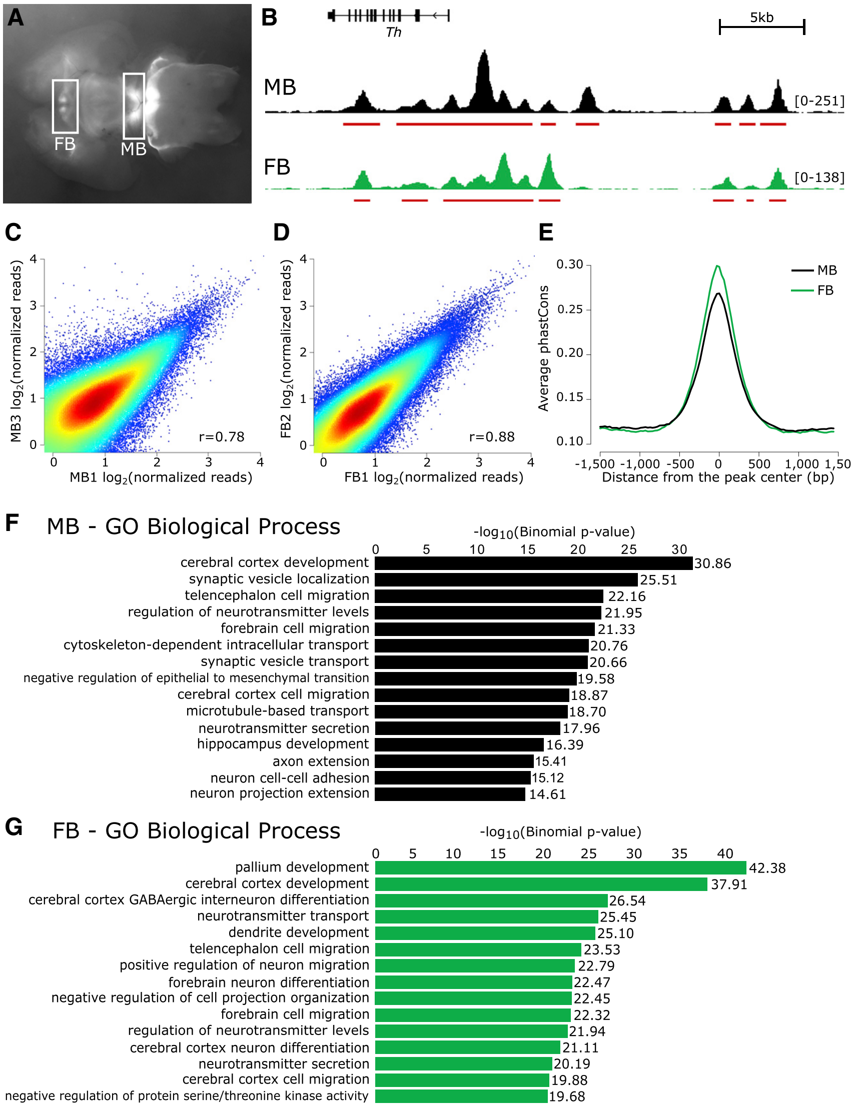


Figure 1. Preliminary Validation of ATAC-seq Catalogs Generated from *Ex Vivo* DA Neurons

(A) The midbrain (MB) and forebrain (FB) of E15.5 brains from Tg(Th-EGFP) DJ76Gsat mice are microdissected, dissociated, and isolated by FACS.

(B) Read pileup and called peaks for the MB and FB libraries at the *Th* locus.

(C and D) Chromatin accessibility, genome-wide, is correlated between replicates.

(E) The sequences underlying MB and FB peaks display a high degree of evolutionary sequence constraint as measured by phastCons scores.

(F and G) For both the MB and FB, gene ontology terms of the expressed genes nearest to each peak reflect the neuronal origin and function of these catalogs.

To evaluate the ATAC-seq libraries, we examined *in silico* quality-control measures (Figure S2); evaluated the called peaks and read pileups with the Integrative Genomics Viewer (IGV);[61,62] and quantified the correlation between brain regions and within replicates. A representative browser trace at the *Th* locus in both MB and FB libraries is presented in Figure 1B. Replicates are well correlated: MB library replicates have an average correlation of 0.72 (Figure 1C), and FB replicates are more highly correlated, at r = 0.86 (Figure 1D). Given the robust correlation between replicates, we pooled all reads from the same brain region and called peaks on this unified set to increase our power to detect regions of open chromatin. As a result, we identified 104,217 regions of open chromatin in the MB DA neurons and 87,862 regions in the FB. MB and FB libraries are moderately well correlated (average r = 0.64; Figure S3), and approximately 60% of MB OCRs are also represented in the FB libraries.

To assess these catalogs for characteristics of functionality, we examined the sequence constraint underlying the called regions of open chromatin but excluded peaks that overlap promoters. Promoters are typically accessible,[63] and thus we aimed to reduce the inflation that affects sequence conservation as a result of highly conserved promoter-overlapping ATAC-seq peaks. Despite removal of these highly conserved peaks, we found that the degree of sequence constraint underlying open-chromatin peaks was high in comparison to background constraint (Figure 1E). The fact that elements in these libraries of putative cis-regulatory elements (CREs) are constrained highlights their probable functional significance.

To further examine the OCR catalogs for biological relevance, we explored the GO terms of nearby genes. Although CREs are not restricted to acting solely on the nearest gene, this restriction is often used as a proxy in the absence of other information. To bolster our predictions, we have also generated bulk RNA-seq data on these same populations of sorted cells (Figure S4) and used these data to examine the GO terms of the nearest expressed gene (RPKM ≥1). Although still imperfect, implementing this method as a proxy for function results in GO terms that are enriched for neuronal functions in both MB and FB OCR catalogs (Figures 1F and 1G). Thus, we establish that these OCR catalogs are enriched for putative CREs, probably directing the expression of genes with key roles in neuronal biology.

Candidate Regulatory Regions Are Capable of Directing Expression *In Vivo*

Although sequence conservation and GO suggest that our OCR catalogs are enriched for functional elements, both of these metrics are indirect surrogates for true measures of

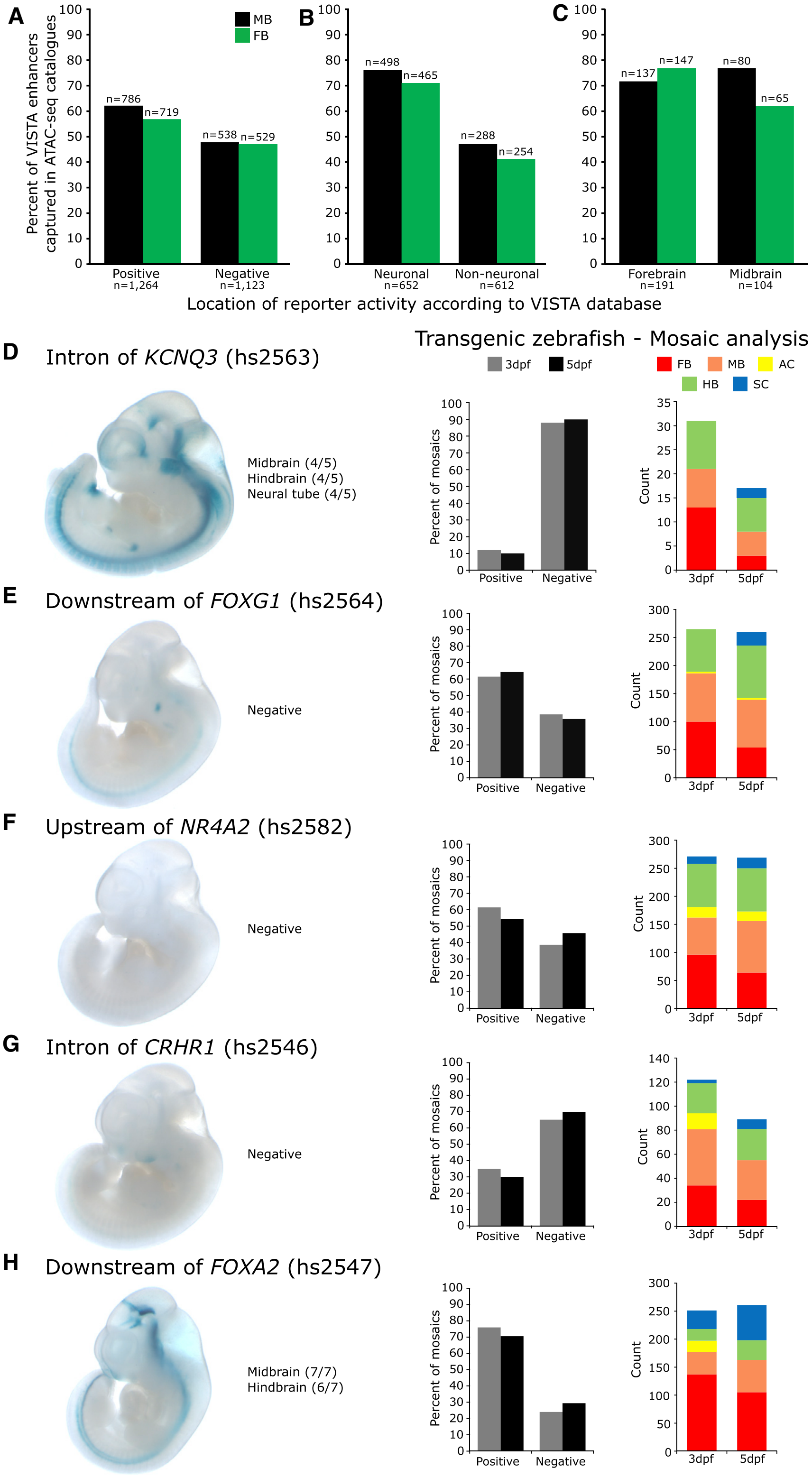


Figure 2. Validation of the Putative CRE Catalogs *In Vivo*

(A) Of the elements annotated in VISTA as having enhancer activity, 62% and 56% of these are represented in the MB and FB catalogs, respectively.

(B) An abundance of open-chromatin regions in the MB and FB catalogs overlap confirmed neuronal enhancers ($\geq 70\%$).

(C) Neuronal enhancers were stratified by the anatomical domains in which they are active; those that are reported active in the MB and FB are enriched in our MB and FB catalogs, respectively.

(D–H) Testing five prioritized putative CREs *in vivo* identifies five neuronal enhancers.

(D) A putative CRE in intron 1 of *KCNQ3* directs expression in the midbrain, hindbrain, and neural tube of E11.5 *lacZ* reporter mice. It fails to direct expression in a transgenic zebrafish assay at either 3 or 5 days post fertilization (dpf); reporter expression is present in $\leq 25\%$ of mosaics.

(E, F, and G) Putative CREs downstream of *FOXG1*, upstream of *NR4A2*, and in an intron of *CRHR1* fail to direct expression in transgenic mice; however, they direct robust neuronal appropriate expression in transgenic zebrafish reporter assays (scored for expression in MB, FB, amacrine cells [ACs], hindbrain [HB], and spinal cord [SC]).

(H) A putative CRE downstream of *FOXA2* directs neuronal expression in both transgenic mice and zebrafish assays. n mosaic zebrafish scored: ≥ 141 for 3 dpf, ≥ 119 for 5 dpf. All constructs have since been deposited in the VISTA database under the supplied hs numbers.

function. To more directly measure the biological relevance of the catalogs and to identify enhancers, we assessed the capability of the candidate CREs to direct expression *in vivo*.

We took advantage of the large repository of elements that have already been tested in *lacZ* reporter assays *in vivo* and that have been cataloged in the VISTA Enhancer Browser[64] (accessed September 4, 2016). Overlap between our catalogs and all 2,387 VISTA Enhancer Browser elements, which were scored for their ability to direct *lacZ* reporter expression in E11.5 mice, was quantified (Table S2). Of the 1,264 VISTA elements identified as enhancers, 786 were present in the MB catalog, and 719 were present in the FB catalog (Figure 2A). We examined the overlap of the FB and MB catalogs with enhancers that have been demonstrated to direct expression in either non-neuronal or neuronal tissues, and we observed that 42%–47% of enhancers reported to direct expression in non-neuronal tissues are present in the catalogs. In contrast, 71%–76% of enhancers that direct expression in one or more regions of the brain overlapped the FB and MB catalogs (Figure 2B), confirming an abundance of brain enhancers in our catalogs. Stratifying these confirmed neuronal enhancers on the basis of their expression patterns in VISTA, we observed an abundance of MB-specific enhancers in our MB catalog and an abundance of FB-specific enhancers in our FB catalog; 77% of MB- and FB-specific enhancers in VISTA were captured in our MB and FB catalogs, respectively (Figure 2C). Collectively, these data establish that our region-specific OCR catalogs capture region-specific, active CREs with high efficiency.

To extend our assessment of the biological activity of sequences within these OCR catalogs, we focused on an additional five candidate CREs not already tested in the VISTA browser and evaluated their ability to act as enhancers in *lacZ* reporter mice and in transgenic zebrafish TdTomato reporter assays. All five regions were represented by robust peaks in both the MB and FB catalogs (Figure S5). Two regions, one in the first intron of *Kcnq3* and the other downstream of *Foxg1*, were additionally prioritized via H3K27ac ChIP-seq from a variety of tissues taken from E11.5 and E15.5 embryonic mice because we sought to limit our selection to candidate enhancers predicted to have neuronal-specific activity. The remaining three candidate CREs were selected on the basis of their proximity to genes important in DA neuron biology. We selected sequences at *Foxa2* and *Nr4a2* because both are key transcription factors (TFs) in the development and maintenance of DA neurons.[65–68] We selected the final region, located in an intron of *Crhr1*, because this locus has been implicated in PD by GWAS,[9] and our group has recently prioritized this gene as a candidate for PD risk.[69] All selected sequences were lifted over to hg19, a process which facilitated the identification and assay of their corresponding human sequence intervals.

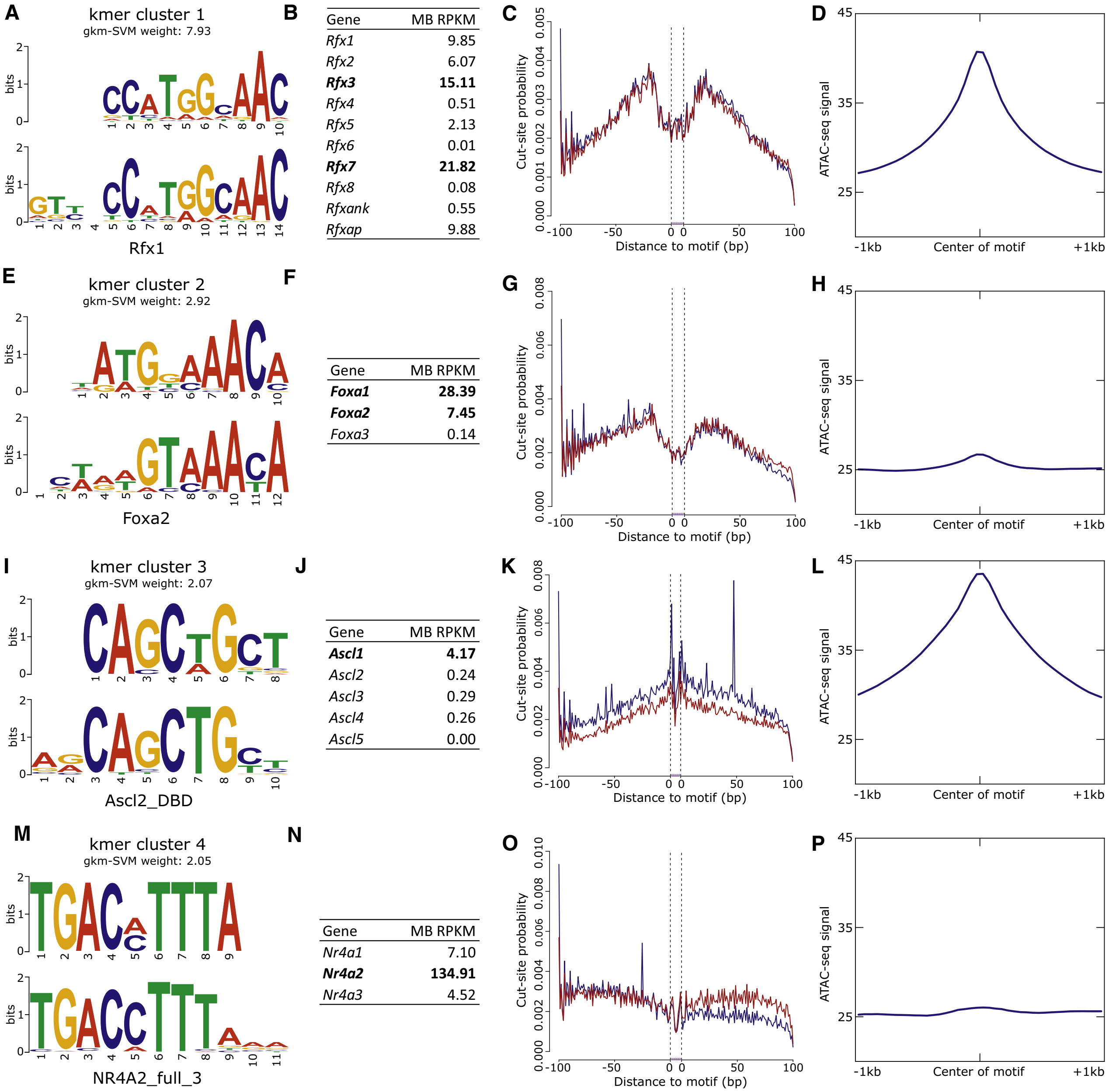
When tested in transgenic reporter mice at E11.5 (Figure S5), two of the five regions (those near *KCNQ3* and *FOXA2*) were validated as enhancers (Figures 2D and 2H). Recognizing that a disparity exists between the developmental time at which we generated the catalogs (E15.5) and the time at which the mice were assayed (E11.5), and that this disparity might compromise validation rates, we also assayed each sequence across multiple time points in zebrafish. All assayed regions except the region at *KCNQ3* directed reporter expression in mosaic transgenic zebrafish (Figures 2E–2H). All five regions displayed enhancer activity *in vivo* in neuronal tissues in one or both transgenic assays. Our transgenic animal experiments corroborate the results of the retrospective VISTA Enhancer Browser intersection, implying that our OCR catalogs are biologically active and enriched for sequences capable of driving neural expression *in vivo*.

Candidate CREs Are Enriched for TF Motifs Active in DA Neurons

To identify sequence modules (kmers) predicted to contribute regulatory activity of putative CREs in our catalogs, we applied the machine learning algorithm gkm-SVM.[45] The resulting regulatory vocabularies of kmers had high predictive power ($\text{auROC}_{\text{MB}} = 0.915$, $\text{auROC}_{\text{FB}} = 0.927$). We rank-ordered and collapsed related kmers to reveal motifs enriched in the OCRs and their corresponding TFs (Figures 3A, 3E, 3I, and 3M). In the MB, the four most enriched motifs correspond to RFX1, FOXA2, ASCL2, and NR4A2. Given the degeneracy of binding motifs within TF families, we consulted the bulk RNA-seq data for each of the implicated TF families and examined the relative expression levels to prioritize which TFs are most likely producing the observed motif enrichments (Figures 3B, 3F, 3J, and 3N). For example, the reported DNA binding domain is highly conserved between RFX family members, and as a result, the predicted sequence motif for each is highly similar;[70,71] thus, we must use other means to identify which family member is likely to be acting in these cells. Although no member of the RFX family has been canonically associated with MB DA neurons, we expect *Rfx3* and *Rfx7*, as the two most highly-expressed *Rfx* genes, to probably be active in MB DA neurons and be driving this motif enrichment (Figure 3B). FOXA1 and, especially, FOXA2 are both known to DA neuron biology[65,72] and both are highly expressed in the MB DA neurons (Figure 3F). Regarding enrichment for the ASCL family, ASCL1 is known to be involved in DA neuron biogenesis[73] and is more highly expressed than any other TF in the family (Figure 3J). Finally, NR4A2 is both canonically associated with DA neurons and required for their development;[68] we observe it to be highly expressed in MB DA neurons (Figure 3N). Examining the sequences underlying the OCR catalogs, we identified TF families known and unknown to DA neuron biology and further refined the TF associations by using expression data.

We also examined the qualities that differentiate MB CREs from FB CREs by examining the sequences underlying MB-specific and FB-specific regions. We developed a vocabulary that discriminates MB and FB regions with high predictive power (auROC = 0.926) and identified kmers enriched in MB-specific peaks where the top corresponding TFs are FOXA1 and/or FOXA2 and NR4A2 (Figure S6). We confirmed this MB bias by again considering the bulk RNA-seq for genes encoding these proteins. As expected, these TFs are more highly expressed in the MB where *Nr4a2* is present at 12-fold higher levels than in the FB (135 RPKM in the MB versus 11 RPKM in the FB) and *Foxa1* and *Foxa2* are not expressed in the FB but are present in the MB (*Foxa1*: 28 RPKM, *Foxa2*: 7 RPKM). Not only do we identify FOXA1 and/or FOXA2 and NR4A2 as more active in MB DA neurons than in the FB, but we also did so solely by comparing their role in the vocabulary of MB-specific OCRs versus FB-specific OCRs.

In a parallel strategy to identify TFs actively engaging the DNA in MB DA neurons, we performed TF footprinting in a single, deeply sequenced MB ATAC-seq library. In doing so, we confirm that two of the TFs prioritized by gkm-SVM leave robust footprints. The motif corresponding to RFX-binding results in a dearth of cuts directly over predicted binding sites (Figure 3C). The same can be seen to a lesser extent for the motif corresponding to FOXA1 and/or FOXA2 (Figure 3G). By contrast, motifs corresponding to



B

Gene	MB RPKM
Rfx1	9.85
Rfx2	6.07
Rfx3	**15.11**
Rfx4	0.51
Rfx5	2.13
Rfx6	0.01
Rfx7	**21.82**
Rfx8	0.08
Rfxank	0.55
Rfxap	9.88

F

Gene	MB RPKM
Foxa1	**28.39**
Foxa2	**7.45**
Foxa3	0.14

J

Gene	MB RPKM
Ascl1	**4.17**
Ascl2	0.24
Ascl3	0.29
Ascl4	0.26
Ascl5	0.00

N

Gene	MB RPKM
Nr4a1	7.10
Nr4a2	**134.91**
Nr4a3	4.52

Figure 3. Identification of Transcription Factors (TFs) Important to DA Neurons

(A) The kmer predicted to have the greatest regulatory potential underlying MB ATAC-seq peaks corresponds to the RFX family of TFs.
(B, C, and D) RNA-seq quantification in these same cells indicates this enrichment is likely due to RFX3 or RFX7 activity. Examining the ATAC-seq signal over predicted binding sites reveals a robust TF footprint (C) and a general enrichment of reads overlapping RFX sites genome-wide (D).
(E–H) Similarly, a kmer corresponding to FOXA1 and/or FOXA2 has similar evidence for the activity of one or both of these TFs.
(I–L) The third-ranked motif most likely corresponds to ASCL1, and although it fails to leave a robust TF footprint (K), there is clear enrichment of ATAC-seq signal overlapping genome-wide predicted ASCL1 binding sites (L).
(M–P) NR4A2, canonically associated with DA neuron biology, is identified as a highly expressed TF that probably contributes to the regulatory potential of the putative CREs; however, it fails to leave a TF footprint in the cut-site patterns around predicted motif sites (O) and is only mildly enriched for ATAC-seq reads over its predicted binding sites (P).

ASCL1 or NR4A2 fail to leave a robust mark on the chromatin availability (Figures 3K and 3O). It has been noted that nuclear receptors, such as NR4A2, only transiently interact with DNA,[74] and as a result, it could be that the short DNA residence time fails to result in a robust footprint detectable by transposition. These footprinting data substantiate the claim that the RFX family of TFs and FOXA1 and/or FOXA2 are active in MB DA neuron CREs.

We confirmed that these sequences are indeed enriched in the catalogs by examining the pileup of reads overlapping all genome-wide predicted motif binding sites for each motif identified by gkm-SVM. We see an abundance of reads over predicted binding sites of all four motifs (Figures 3D, 3H, 3L, and 3P); the strongest enrichment overlaps RFX and ASCL1 motif sites (Figures 3D and 3L). Despite the less robust footprint generated at the ASCL1 sites, this TF clearly underlies a larger-than-expected proportion of OCRs in the MB catalog. Integrating a support vector machine-learning algorithm, as applied to the sequences underlying OCRs, with footprinting analysis in the same chromatin substrate both powerfully identifies TFs that are important for DA neuron biology and suggests the RFX family of TFs, FOXA1 and FOXA2, ASCL1, and NR4A2 are actively influencing gene expression in the MB DA neurons.

A Candidate CRE in Intron 4 of *SNCA* Is Associated with PD Risk

Having established the biological robustness of the OCR catalog, we moved to exploit these data to investigate how non-coding variation therein might be contributing to PD risk. We established two complementary strategies. First, we sought to globally examine the overlap of PD GWAS SNPs[9,10] and those in LD ($r^2 > 0.8$) with our DA OCR catalogs. In doing so, we identified 129 unique PD-associated variants that occur at 20 GWAS-associated loci and that are present in one or both of our OCR catalogs (34 specifically overlap the MB catalog, 14 specifically overlap the FB catalog, and 81 overlap both; Table S3).

Second, we examined the chromatin landscape surrounding familial PD genes by focusing on genes with no obvious overlaps in the first strategy. In doing this, we turned our attention to the *SNCA* locus. Despite the fact that this locus is the most significant hit in PD GWASs,[9,10] the LD structure surrounding the lead SNP (dbSNP: rs356182) is such that no variants in LD are apparent at our r^2 cut-off, and the lead SNP itself is not overlapped by either our MB or FB catalog. Given α-synuclein's established role in PD pathogenesis and the strength of GWAS signal at *SNCA*, we prioritized this locus for a closer, more targeted, inspection.

We first noted that *Snca* expression differs significantly between the MB and FB DA neurons in our bulk RNA-seq analysis (Figure 4B). Examining the chromatin accessibility at the *Snca* locus, we found that the MB and FB were largely the same with the exception of one robust peak in intron 4 (mm9: chr6: 60,742,503–60,744,726) that is present in the MB but completely absent in the FB (Figure 4A). DNase hypersensitivity site (DHS) linkage[57,63] suggests that this putative CRE interacts with the *SNCA* promoter. Given the MB specificity of this putative CRE and indications that it interacts with the *SNCA* promoter, we suspected that this region might be a driving force behind the MB-specific expression of *Snca*.

To test this hypothesis, we assayed whether, when lifted over to hg19, the central portion of this putative CRE (chr4: 90,721,063–90,722,122), is capable of directing appropriate reporter expression in transgenic zebrafish and mouse reporter assays. Stable transgenesis of zebrafish indicates that this CRE directs reporter expression at 72 hpf in the locus coeruleus, a key population of catecholaminergic neurons preferentially degenerated in PD,[75] and along the catecholaminergic tract through the hindbrain, which is largely composed of DA neurons[76] (Figure 4C). Additionally, we observe reporter expression throughout the diencephalic catecholaminergic cluster with projections to the subpallium, which is analogous to mammalian dopaminergic projections from the ventral midbrain to the striatum.[77] Reporter expression in these transgenic zebrafish is largely consistent with an enhancer active in catecholaminergic populations.

To further evaluate this CRE in a mammalian system, we generated *lacZ* reporter mice and examined reporter activity across developmental time. Whole-mount E12.5 reporter mice indicate this enhancer directs exquisitely restricted expression in Th+ populations, including the dorsal root ganglia, extending into the sympathetic chain and throughout the cranial nerves (particularly the trigeminal). Additional diffuse staining is noted throughout the MB and FB (Figure 4D). Specifically examining the brains of *lacZ* animals at E15.5, we identified reporter expression in the MB and hypothalamus, as well as strong expression through the amygdala and piriform cortex and along the anterior portion of the sympathetic chain (Figure 4E); we see similar reporter patterns at P7 (Figure 4F). At P30, we detect reporter activity in the amygdala, hypothalamus, thalamus, periaqueductal gray area, brain stem, and importantly, in the *substantia nigra* and ventral tegmental area (Figure 4G). By contrast, in aged *lacZ* reporter mice (574 days old [~19 months]), we detect strong reporter expression only in the brain stem and observe weak reporter expression in the amygdala (Figure 4H). Collectively, the regions in which we detect reporter activity reflect those compromised in PD; Lewy bodies (aggregates of α-synuclein) have been detected in the locus coeruleus, sympathetic chain, amygdala, hypothalamus, ventral tegmental area, and periaqueductal gray area of PD-affected individuals.[78–82] Critically, the preferential degradation of the *substantia nigra* is the pathological hallmark of PD progression.[2] This enhancer directs region-specific appropriate expression throughout development in key locations in concordance with SNCA activity in PD pathogenesis.

After confirming this CRE's regulatory activity in brain regions associated with PD, we next inspected this sequence for PD-associated variation. We sequenced across this interval in 986 individuals with PD and 992 controls and identified 14 variants (Table S4); four of these variants had an MAF greater than 5% and were common and present in both affected individuals and controls. Of these, two tightly linked variants ($r^2 = 0.934$; Table S5), dbSNP: rs2737024 (OR = 1.25, 95% CI = 1.09–1.44, p value = 0.002) and dbSNP: rs2583959 (OR = 1.22, 95% CI = 1.06–1.40, p value = 0.005), were significantly associated with PD (Table 1). These data support a role for variation within the enhancer in conferring PD risk.

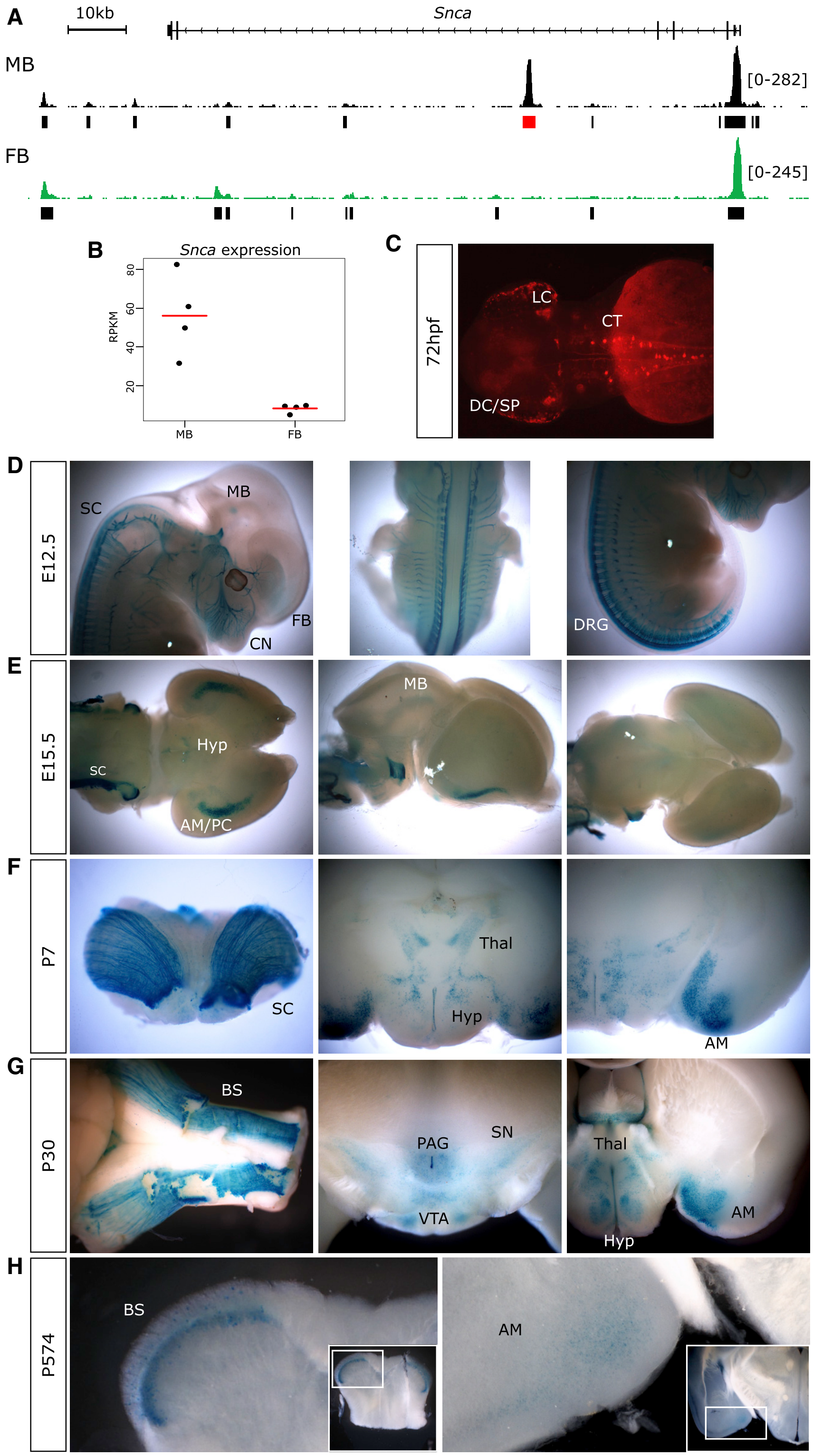


Figure 4. A MB-Specific Enhancer Directs Expression in Catecholaminergic Populations of Neurons Known to Parkinson Disease Biology

(A) An IGV track indicating the location of the MB-specific region of open chromatin located in intron 4 of *Snca*.

(B) *Snca* is differentially expressed between the MB and FB DA neurons. The red bar is the mean expression of the four replicates (black dots).

(C) At 72 hpf, stable transgenic zebrafish reporter assays indicate this putative CRE is capable of directing reporter expression in key catecholaminergic neuronal populations, including the locus coeruleus (LC), the catecholaminergic tract (CT) of the hindbrain, the diencephalic cluster (DC), and the subpallium (SP), into which the DC projects.

(D–H) Further studies in *lacZ* reporter assays in embryonic (E) and post-natal (P) mice indicate dynamic enhancer usage across developmental time.

(D) This enhancer directs expression throughout the MB, FB, dorsal root ganglia (DRG), sympathetic chain (SC), and cranial nerves (CN) of E12.5 mice.

(E) By E15.5, reporter expression is observed in the amygdala and/or piriform cortex (AM/PC), sympathetic chain, MB, and hypothalamus (Hyp).

(F) Patterns of reporter expression at P7 reflect those seen at E15.5.

(G) Reporter activity is observed at P30 in the amygdala; hypothalamus and thalamus (Thal); brain stem (BS); *substantia nigra* (SN); ventral tegmental area (VTA); and the periaqueductal gray area (PAG).

(H) In aged mice (P574), reporter expression is detected robustly in the brain stem and faintly in the amygdala.

To assess how these variants might impact enhancer function and thus PD risk, we assayed differential protein binding at these variants for >16,000 proteins.[58] In doing so, we identified five proteins whose binding is robustly impacted by these implicated variants: NOVA1, APOBEC3C, PEG10, SNRPA, and CHMP5 (Figures 5A–5C). Of these, all except APOBEC3C (RPKM ≤1) are expressed at appreciable levels in both MB and FB DA neurons (Figure 5D). Of the remaining four proteins, three (PEG10, SNRPA, and CHMP5) demonstrate an increased binding affinity for the minor risk allele over the major allele; this direction of effect is consistent with the overexpression paradigm by which *SNCA* confers PD risk.[8] Interestingly, out of those proteins we identified, CHMP5 is the only one whose binding affinity is impacted by variant dbSNP: rs2583959, and our group has recently implicated one of its family members, CHMP7, in conferring PD risk,[69] perhaps indicating a role for this family of proteins in PD. Although no single protein stands out, the increased affinity of proteins expressed in DA neurons for the risk alleles of the identified enhancer variants is consistent with a potential mechanistic contribution to *SNCA* expression and, therefore, to PD risk.

Table 1. Two Tightly Linked SNPs within the Enhancer are Significantly Associated with PD Risk

				Association with PD	
Variant	**MA**	**MAF in PD-Affected Individuals (n = 986)**	**MAF in Controls (n = 992)**	**OR (95% CI)**	**p Value**
rs7684892	A	0.063	0.069	0.93 (0.72–1.20)	0.562
rs17016188	C	0.082	0.061	1.35 (1.04–1.75)	0.023
rs2583959	G	0.317	0.271	1.22 (1.06–1.40)	0.005*
rs2737024	G	0.319	0.270	1.25 (1.09–1.44)	0.002*

Abbreviations are as follows: MA = minor allele; MAF = minor-allele frequency; OR = odds ratio; and CI = confidence interval.
Only variants with MAF > 0.05 were considered.
ORs, 95% CIs, and p values result from additive logistic regression models adjusted for sex and age at blood draw. p values ≤ 0.0125 were considered as statistically significant after a Bonferroni correction for multiple testing (*).

Finally, we set out to refine the haplotype structure and understand how this identified variation might be interacting with other variants at this locus. A panel of common variants had previously been genotyped across *SNCA*, and PD-associated haplotypes were identified.[53] After genotyping our PD-affected individuals and controls for a subset of this panel of variants (in addition to all enhancer-associated variants identified by sequencing [Table S6]), we identified a single haplotype that was significantly associated with PD (p value = 0.003) and that had a higher observed frequency in PD-affected individuals (28.3%) than in controls (23.4%; Table 2). This haplotype implicates some of the same variants as those in Guella et al.[53] (dbSNP: rs356220, dbSNP: rs737029) but also implicates dbSNP: rs356225 and dbSNP: rs356168, as well as the two enhancer-associated variants. Additionally, within the 1000 Genomes data, we observe that moderate LD structure exists between the lead GWAS variant (dbSNP: rs356182) and the enhancer variants (r^2 = 0.418, D′ = 0.745) in the general European population. Despite the moderate LD, the risk allele of dbSNP: rs356182 falls in our identified PD-associated haplotype 94% of the time. Thus, it is likely that at least part of the risk captured by dbSNP: rs356182 can be attributed to these enhancer variants and the implicated haplotype reported here. Furthermore, this does not preclude additional variants from being present and contributing to the risk captured by the lead SNP because the dbSNP: rs356182 risk allele can occur in the absence of the enhancer-associated variants (i.e.: ~31% of EUR individuals with the dbSNP: rs356182 risk allele do not carry the risk alleles of the PD-associated enhancer variants). A schematic depiction of the variants, open-chromatin regions, chromatin interactions,[57,63] and LD structure at this locus is presented in Figure 6. Of the variants, including dbSNP: rs356182, whose minor alleles define this PD-associated haplotype, only the two enhancer-associated variants and dbSNP: rs2737029 are identified as eQTLs for *SNCA* expression in any tissue in the GTEx database (Figure S7). Collectively, these data identify a catecholaminergic enhancer harboring common variation that is part of a larger haplotype associated with PD risk, likely by modulating *SNCA* activity.

Discussion

The identification and prioritization of biologically pertinent non-coding variation associated with disease remains challenging. Recent studies by our and other groups have emphasized the importance of cellular context in the identification of sequences harboring biologically pertinent variation and the genes they regulate. To this end, we used chromatin signatures from *ex vivo* isolated DA neurons to reveal biologically active sequences that harbor non-coding variation contributing to PD risk. We generated robust OCR catalogs for both MB and FB DA neurons, confirmed their capacity to act as enhancers, identified motifs that confer their regulatory potential, and notably, identified two variants that are located within a MB-specific enhancer and are associated with an increase in PD risk.

In contrast to strategies predicated solely on dissection of post-mortem tissues or on the differentiation of cultured cells, we leveraged the use of transgenic reporter mice to specifically isolate Th-expressing neurons from discrete neuroanatomical (FB and MB) domains. Although our approach assays a more refined population of DA neurons than would be achieved via gross dissection, recent single-cell RNA-seq analyses of these same cells make clear that even within these highly restricted MB and FB populations there exist two primary cellular phenotypes.[69] The "homogenous" MB and FB populations each are comprised of an immature neuroblast population and a more mature, domain-specific, post-mitotic population of DA neurons. As a result, our OCR catalogs capture the chromatin accessibility from both of these states. These catalogs are demonstrably biologically relevant for our purposes, but future studies requiring even greater homogeneity might wish to consider single-cell ATAC-seq to refine these domains further.[83]

Through *in silico* validation of the catalogs, we established that they are enriched for both sequence constraint and biological relevance in a manner consistent with function and their FB or MB origin. Furthermore, these sequences are frequently domain-appropriate enhancers, and each catalog captures a large fraction (77%) of previously validated MB and FB enhancers. Although more regions are shown to direct neuronal expression than to direct negative or non-neuronal expression, it is

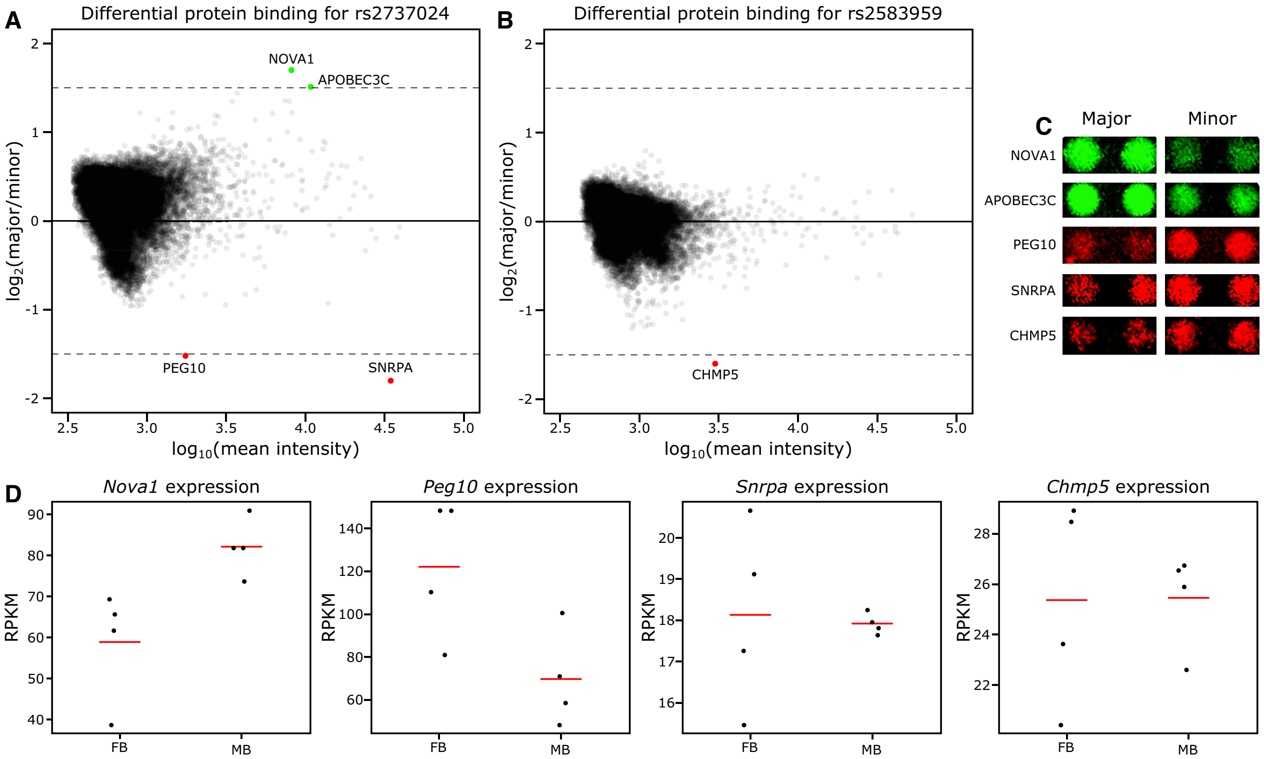


Figure 5. Identification of Proteins Whose Binding Is Impacted by the Implicated PD-risk SNPs

(A and B) MA plots for both dbSNP: rs2737024 and dbSNP: rs2583959 indicate the magnitude of the effect of the minor and major allele on binding. Cut-off for differential binding: $\log_2(\text{major/minor}) \geq 1.5$ or ≤ -1.5.

(A) NOVA1 and APOBEC3C (green circles) bind at dbSNP: rs2737024 with greater affinity for the major allele, but PEG10 and SNRPA (red circles) have a greater affinity for the minor allele.

(B) CHMP5 (red circle) has a greater affinity for the minor allele of dbSNP: rs2583959.

(C) Representative images of the protein binding for each of the differentially bound proteins.

(D) Expression analysis in the MB and FB DA neurons for each of the differentially bound proteins indicates *Nova1*, *Peg10*, *Snrpa*, and *Chmp5* to be highly expressed in these populations, yet none of the *Apobec* family member genes are expressed (RPKMs $\leq$1, data not shown). The red bar is the mean expression of the four replicates (black dots).

interesting to note that almost half of the sequences previously documented as not directing expression *in vivo* are also represented in one or both of our catalogs.

Given the frequently dynamic nature of CRE activity, this overlap with negative regions most likely results from temporal differences in these assays. Our data indicate that these regions are accessible at E15.5, but the *lacZ* reporter assays were carried out at E11.5; regions that have been annotated as negative at E11.5 might be active at later time points and, as such, appear in our catalogs. As we moved from these unbiased functional comparisons to more highly selected ones, the potential impact of temporal differences became more pronounced. In mouse transgenic reporter assays, two of five assayed putative CREs directed detectable expression of *lacZ* in neuronal populations. Consistent with the temporally dynamic nature of CREs, when we tested these same regions in zebrafish across multiple developmental time points, we observed that four of the five sequences acted as neuronal enhancers.

By examining the sequence composition underlying the ATAC-seq peaks, we illuminate powerful vocabularies for both FB and MB DA neuron transcriptional regulatory control. The machine-learning algorithm gkm-SVM prioritized four transcription factor families (RFX, FOXA1/2, NR4A2, ASCL1/2) as conveying significant regulatory potential in the CRE catalogs. Of these, the RFX family had not previously been implicated in DA neuron biology. Although several of the RFX family members have been annotated as having expression in the cerebellum or fetal brain,[70] a role specifically in MB DA neurons has not previously been appreciated. By contrast, NR4A2 is canonically associated with MB DA neurons,[67,68] is highly expressed in this population (139 RPKM), and was prioritized as having TF-conferring regulatory potential in these cells; however, TF footprinting fails to provide evidence supporting its activity. We postulate that this lack of a footprint might reflect the transient DNA-binding dynamics of NR4A2. Transcription factors with short DNA residence times often fail to reveal footprints, and nuclear receptors, such as NR4A2, have markedly transient DNA interactions.[74]

Taken collectively, these data establish a robust biological platform from which PD-associated variation can be evaluated. To this end, an obvious candidate to interrogate was an apparent MB-specific open-chromatin domain within

Table 2. A Single Haplotype, Containing the Minor Alleles of the Implicated SNPs, Is Significantly Associated with PD Risk

Haplotypes spanning the *SNCA* locus	rs356220	rs356225	rs3857057	rs356168	rs10018362	rs2737029	rs62306323	rs2737024	rs2583959	rs17016188	rs7684892	rs7689942	Frequency in PD-Affected Individuals	Frequency in Controls	p Value
1	C	**C**	A	**G**	T	T	C	A	C	T	G	C	0.015	0.027	0.012
2	C	G	A	A	T	T	**T**	A	C	T	G	C	0.092	0.110	0.029
3	C	**C**	A	**G**	**C**	**C**	C	A	C	T	**A**	**T**	0.039	0.048	0.184
4	**T**	**C**	A	**G**	T	T	C	A	C	T	G	C	0.037	0.040	0.480
5	C	G	A	A	T	T	C	A	C	T	G	C	0.380	0.397	0.593
6	**T**	**C**	A	**G**	T	T	**T**	A	C	T	G	C	0.010	0.011	0.698
7	C	G	A	A	T	**C**	C	**G**	**G**	T	G	C	0.009	0.012	0.944
8	C	**C**	A	**G**	T	**C**	C	**G**	**G**	T	G	C	0.016	0.015	0.768
9	**T**	**C**	**G**	**G**	**C**	**C**	C	A	C	T	**A**	**T**	0.021	0.017	0.360
10	**T**	**C**	**G**	**G**	T	**C**	C	A	C	**C**	G	C	0.014	0.009	0.189
11	**T**	**C**	**G**	**G**	**C**	**C**	C	A	C	**C**	G	C	0.057	0.044	0.124
12	**T**	**C**	A	**G**	T	**C**	C	**G**	**G**	T	G	C	0.283	0.234	0.003*

Only haplotypes with frequency ≥0.01 were considered. Black boxes indicate the minor allele in Europeans.
p values result from score tests for association that were performed under an additive model and adjusted for sex and age at blood draw.
p values ≤0.0042 were considered as statistically significant after application of a Bonferroni correction for multiple testing (*).

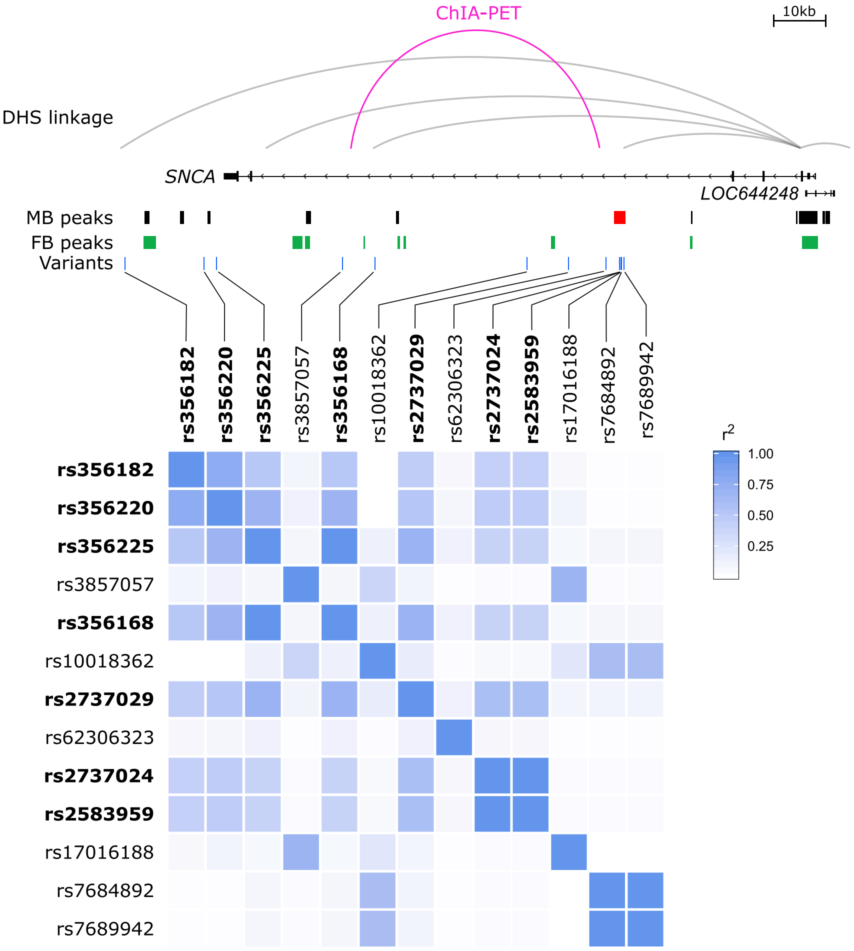


Figure 6. A Schematic of the Chromatin Interactions, LD Structure, Variation, and Open Chromatin at the *SNCA* Locus
Publicly available DNase hypersensitivity site (DHS) linkage analysis suggests that the promoter of *SNCA* possibly interacts with the identified MB-specific enhancer, the lead GWAS variant (dbSNP: rs356182), and a previously functionally validated variant (dbSNP: rs356168). ChIA-PET data suggest that the MB-specific enhancer might interact with variant dbSNP: rs356168. Open-chromatin data from DA neurons do not overlap with any variants at this locus or haplotype other than at the MB-specific enhancer. LD analysis at this locus indicates that despite the low LD structure between the lead GWAS variant (dbSNP: rs356182) and the enhancer-associated variants (dbSNP: rs2737024 and dbSNP: rs2583959), the variants are in the same haplotype. Therefore, the GWAS signal might, at least in part, be flagging the identified enhancer-associated variants.

intron 4 of the known PD-associated gene *SNCA*. We assayed the activity of this putative CRE in zebrafish and across the life course of mice and found it to be active in key catecholaminergic structures (e.g.: the *substantia nigra* and locus coeruleus) injured in PD, from mid-gestation until at least P30. Thereafter, the utilization of this enhancer in the brain is diminished and by late life appears restricted to the brainstem and amygdala. By the time of clinical presentation, individuals diagnosed with PD have already lost a significant proportion (≥30%) of their nigral DA neurons;[2,84] the observed biology of this CRE is consistent with a progressive pathogenic influence that begins early in life and renders these populations preferentially vulnerable to loss over an extensive period of time.

Sequencing this interval in PD-affected individuals and controls revealed two common variants (dbSNP: rs2737024 and dbSNP: rs2583959), individually associated with an increased risk of PD. After testing these variants for their effect on protein binding, we identified five proteins whose binding is affected, three of which (PEG10, SNRPA, and CHMP5) display greater affinity for the risk allele. Furthermore, we identified a larger haplotype containing these variants; this haplotype is also significantly associated with PD risk. Although none of the other SNPs in this haplotype overlap with CREs identified in the DA neuron catalogs, there is significant functional evidence of the activity and contribution to PD risk by variant dbSNP: rs356168.[85] The same DHS correlation analysis[57,63] that suggests an interaction between the *SNCA* promoter and our identified CRE also suggests an interaction between the *SNCA* promoter and the dbSNP: rs356168 variant. Additionally, ChIA-PET data[25,57] indicate that sequences encompassing this variant might interact with our enhancer, suggesting a potential cooperative mode of action; Gupta and colleagues[86] recently proposed that such a paradigm takes place at the *EDN1* locus. We propose that the variants within the enhancer, independently or in concert with other variation within the identified haplotype, might act throughout an individual's lifespan to render key populations of catecholaminergic neurons vulnerable and thus increase PD risk in individuals harboring this variation.

This work emphasizes the value of biologically informed, cell-context-dependent guided searches for the identification of disease-associated and functional non-coding variation. Given the extent of non-coding GWAS-identified variation, the need for strategies to prioritize variants for functional follow-up is greater than ever. Here, we generate chromatin accessibility data from purified populations of DA neurons to generate catalogs of putative CREs. We have demonstrated how these data can be used to reveal non-coding variation contributing to PD risk; focusing on a single region of open chromatin at the *SNCA* locus, we uncover PD-associated variation therein and propose a model through which this sequence can contribute to normal DA neuronal biology and PD risk. There remains a plethora of information still to be explored in these catalogs, either through further single-locus investigations or through massively parallel assays. For example, our MB DA-neuron OCR catalog overlaps variants at 20 of 49 (41%) PD-associated loci,[9,10] all of which can be investigated further for the mechanisms by which they impact PD risk. Our work establishes a powerful paradigm, leveraging transgenic model systems to systematically generate cell-type-specific chromatin accessibility data and reveal disease-associated variation, in a manner that can be progressively guided by improved biological understanding.

Accession Numbers

ATAC-sequencing, RNA-sequencing and related data will be available at the Gene Expression Omnibus (GEO) under the accession number GSE122450.

Supplemental Data

Supplemental Data include seven figures and six tables and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.10.018.

Acknowledgments

This research undertaken at Johns Hopkins University School of Medicine was supported in part by awards from the NIH (NS62972 and MH106522 to A.S.M.; GM111514 to H.Z.; HG007348 to M.A.B.). The Mayo Clinic collection was supported in part by Morris K. Udall Center of Excellence in Parkinson's Disease Research (P50 NS072187), American Parkinson Disease Association Center, and The Mangurian Foundation for Lewy Body Research. O.A.R. is supported by NS078086 and NS10069 (NIH), W81XWH-17-1-0249 (Department of Defense), The Michael J. Fox Foundation, and The Little Family Foundation. Z.K.W. is supported by the Mayo Clinic Center for Regenerative Medicine; the Mayo Clinic Center for Individualized Medicine; the Mayo Clinic Neuroscience Focused Research Team (Cecilia and Dan Carmichael Family Foundation; and the James C. and Sarah K. Kennedy Fund for Neurodegenerative Disease Research at Mayo Clinic in Florida); a gift from Carl Edward Bolch, Jr., and Susan Bass Bolch; The Sol Goldman Charitable Trust; and Donald G. and Jodi P. Heeringa. Research conducted at the E.O. Lawrence Berkeley National Laboratory was performed under U.S. Department of Energy contract DE-AC02-05CH11231, University of California, and was supported by HG003988 (NIH) to L.A.P.

Declaration of Interests

The authors declare no competing interests.

Received: July 11, 2018
Accepted: October 17, 2018
Published: November 29, 2018

Web Resources

The 3D Genome Browser, http://promoter.bx.psu.edu/hi-c/index.html

Cistrome, http://cistrome.org/ap/

FastQC, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

The Genomic Regions Enrichment of Annotations Tool (GREAT), http://great.stanford.edu

LDlink, https://ldlink.nci.nih.gov/

The List of Genes and Their TSSs Used by GREAT: http://bejerano.stanford.edu/help/download/attachments/2752609/mm9.great3.0.genes.txt

The UCSC Table Browser, http://genome.ucsc.edu

R: A Language and Environment for Statistical Computing, https://www.R-project.org/

R *Corrplot* package, https://github.com/taiyun/corrplot

R *haplo.stats* package, https://CRAN.R-project.org/package=haplo.stats

R *LSD* package, https://CRAN.R-project.org/package=LSD

R *RColorBrewer* package, https://CRAN.R-project.org/package=RColorBrewer

rAggr, http://raggr.usc.edu/

VISTA Enhancer Browser, https://enhancer.lbl.gov

References

1. Ma, S.Y., Röyttä, M., Rinne, J.O., Collan, Y., and Rinne, U.K. (1997). Correlation between neuromorphometry in the substantia nigra and clinical features in Parkinson's disease using disector counts. J. Neurol. Sci. *151*, 83–87.
2. Fearnley, J.M., and Lees, A.J. (1991). Ageing and Parkinson's disease: Substantia nigra regional selectivity. Brain *114*, 2283–2301.
3. Pringsheim, T., Jette, N., Frolkis, A., and Steeves, T.D.L. (2014). The prevalence of Parkinson's disease: A systematic review and meta-analysis. Mov. Disord. *29*, 1583–1590.
4. Thomas, B., and Beal, M.F. (2007). Parkinson's disease. Hum. Mol. Genet. *16 Spec No. 2*, R183–R194.
5. Zarranz, J.J., Alegre, J., Gómez-Esteban, J.C., Lezcano, E., Ros, R., Ampuero, I., Vidal, L., Hoenicka, J., Rodriguez, O., Atarés, B., et al. (2004). The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. Ann. Neurol. *55*, 164–173.
6. Krüger, R., Kuhn, W., Müller, T., Woitalla, D., Graeber, M., Kösel, S., Przuntek, H., Epplen, J.T., Schöls, L., and Riess, O. (1998). Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nat. Genet. *18*, 106–108.
7. Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., et al. (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science *276*, 2045–2047.
8. Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R., et al. (2003). alpha-Synuclein locus triplication causes Parkinson's disease. Science *302*, 841.
9. Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., et al.; International Parkinson's Disease Genomics Consortium (IPDGC); Parkinson's Study Group (PSG) Parkinson's Research: The Organized GENetics Initiative (PROGENI); 23andMe; GenePD; NeuroGenetics Research Consortium (NGRC); Hussman Institute of Human Genomics (HIHG); Ashkenazi Jewish Dataset Investigator; Cohorts for Health and Aging Research in Genetic Epidemiology (CHARGE); North American Brain Expression Consortium (NABEC); United Kingdom Brain Expression Consortium (UKBEC); Greek Parkinson's Disease Consortium; and Alzheimer Genetic Analysis Group (2014). Large-scale meta-analysis of

genome-wide association data identifies six new risk loci for Parkinson's disease. Nat. Genet. *46*, 989–993.

10. Chang, D., Nalls, M.A., Hallgrímsdóttir, I.B., Hunkapiller, J., van der Brug, M., Cai, F., Kerchner, G.A., Ayalon, G., Bingol, B., Sheng, M., et al.; International Parkinson's Disease Genomics Consortium; and 23andMe Research Team (2017). A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat. Genet. *49*, 1511–1516.
11. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.
12. Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., and Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. Genome Res. *22*, 1748–1759.
13. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.
14. Forrest, M.P., Zhang, H., Moy, W., McGowan, H., Leites, C., Dionisio, L.E., Xu, Z., Shi, J., Sanders, A.R., Greenleaf, W.J., et al. (2017). Open chromatin profiling in hiPSC-derived neurons prioritizes functional noncoding psychiatric risk variants and highlights neurodevelopmental loci. Cell Stem Cell *21*, 305–318.e8.
15. Fullard, J.F., Giambartolomei, C., Hauberg, M.E., Xu, K., Voloudakis, G., Shao, Z., Bare, C., Dudley, J.T., Mattheisen, M., Robakis, N.K., et al. (2017). Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci. Hum. Mol. Genet. *26*, 1942–1951.
16. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An expanded view of complex traits: From polygenic to omnigenic. Cell *169*, 1177–1186.
17. Lee, D., Gorkin, D.U., Baker, M., Strober, B.J., Asoni, A.L., McCallion, A.S., and Beer, M.A. (2015). A method to predict the impact of regulatory variants from DNA sequence. Nat. Genet. *47*, 955–961.
18. Praetorius, C., Grill, C., Stacey, S.N., Metcalf, A.M., Gorkin, D.U., Robinson, K.C., Van Otterloo, E., Kim, R.S.Q., Bergsteinsdottir, K., Ogmundsdottir, M.H., et al. (2013). A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. Cell *155*, 1022–1033.
19. Gong, S., Zheng, C., Doughty, M.L., Losos, K., Didkovsky, N., Schambra, U.B., Nowak, N.J., Joyner, A., Leblanc, G., Hatten, M.E., and Heintz, N. (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature *425*, 917–925.
20. Westerfeld, M. (2007). The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio) (Eugene: Univ. Oregon Press).
21. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.
22. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
23. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
24. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.
25. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
26. Shin, H., Liu, T., Manrai, A.K., and Liu, X.S. (2009). CEAS: cis-regulatory element annotation system. Bioinformatics *25*, 2605–2606.
27. Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y., et al. (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol. *12*, R83.
28. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.
29. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44* (W1), W160-5.
30. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: A fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.
31. Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108.
32. Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods *12*, 115–121.
33. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.
34. Mudge, J.M., and Harrow, J. (2015). Creating reference gene annotation for the mouse C57BL6/J genome assembly. Mamm. Genome *26*, 366–378.
35. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., and Wang, L. (2014). CrossMap: A versatile tool for coordinate conversion between genome assemblies. Bioinformatics *30*, 1006–1007.
36. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.
37. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134.
38. Hansen, K.D., Irizarry, R.A., and Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics *13*, 204–216.
39. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. *15*, 1034–1050.

40. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.
41. Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature *444*, 499–502.
42. Poulin, F., Nobrega, M.A., Plajzer-Frick, I., Holt, A., Afzal, V., Rubin, E.M., and Pennacchio, L.A. (2005). In vivo characterization of a vertebrate ultraconserved enhancer. Genomics *85*, 774–781.
43. Kothary, R., Clapoff, S., Darling, S., Perry, M.D., Moran, L.A., and Rossant, J. (1989). Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. Development *105*, 707–714.
44. Fisher, S., Grice, E.A., Vinton, R.M., Bessling, S.L., Urasaki, A., Kawakami, K., and McCallion, A.S. (2006). Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat. Protoc. *1*, 1297–1305.
45. Ghandi, M., Lee, D., Mohammad-Noori, M., and Beer, M.A. (2014). Enhanced regulatory sequence prediction using gapped k-mer features. PLoS Comput. Biol. *10*, e1003711.
46. Zorita, E., Cuscó, P., and Filion, G.J. (2015). Starcode: Sequence clustering based on all-pairs search. Bioinformatics *31*, 1913–1919.
47. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.
48. Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. *2*, 28–36.
49. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol. *8*, R24.
50. Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. *21*, 447–455.
51. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: Scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.
52. Adey, A., Morrison, H.G., Asan, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., and Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. *11*, R119.
53. Guella, I., Evans, D.M., Szu-Tu, C., Nosova, E., Bortnick, S.F., Goldman, J.G., Dalrymple-Alford, J.C., Geurtsen, G.J., Litvan, I., Ross, O.A., et al.; SNCA Cognition Study Group (2016). α-synuclein genetic variability: A biomarker for dementia in Parkinson disease. Ann. Neurol. *79*, 991–999.
54. Hughes, A.J., Daniel, S.E., Kilford, L., and Lees, A.J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J. Neurol. Neurosurg. Psychiatry *55*, 181–184.
55. Schaid, D.J., Rowland, C.M., Tines, D.E., Jacobson, R.M., and Poland, G.A. (2002). Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am. J. Hum. Genet. *70*, 425–434.
56. Machiela, M.J., and Chanock, S.J. (2015). LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics *31*, 3555–3557.
57. Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., et al. (2018). The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol. *19*, 151.
58. Jeong, J.S., Jiang, L., Albino, E., Marrero, J., Rho, H.S., Hu, J., Hu, S., Vera, C., Bayron-Poueymiroy, D., Rivera-Pacheco, Z.A., et al. (2012). Rapid identification of monospecific monoclonal antibodies using a human proteome microarray. Mol. Cell. Proteomics *11*.
59. Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C., et al. (2013). DNA methylation presents distinct binding sites for human transcription factors. eLife *2*, e00726.
60. Heintz, N. (2004). Gene expression nervous system atlas (GENSAT). Nat. Neurosci. *7*, 483.
61. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.
62. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.
63. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.
64. Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser—A database of tissue-specific human enhancers. Nucleic Acids Res. *35*, D88–D92.
65. Stott, S.R.W., Metzakopian, E., Lin, W., Kaestner, K.H., Hen, R., and Ang, S.-L. (2013). Foxa1 and foxa2 are required for the maintenance of dopaminergic properties in ventral midbrain neurons at late embryonic stages. J. Neurosci. *33*, 8022–8034.
66. Arenas, E. (2008). Foxa2: The rise and fall of dopamine neurons. Cell Stem Cell *2*, 110–112.
67. Prakash, N., and Wurst, W. (2006). Development of dopaminergic neurons in the mammalian brain. Cell. Mol. Life Sci. *63*, 187–206.
68. Smits, S.M., Ponnio, T., Conneely, O.M., Burbach, J.P.H., and Smidt, M.P. (2003). Involvement of Nurr1 in specifying the neurotransmitter identity of ventral midbrain dopaminergic neurons. Eur. J. Neurosci. *18*, 1731–1738.
69. Hook, P.W., McClymont, S.A., Cannon, G.H., Law, W.D., Morton, A.J., Goff, L.A., and McCallion, A.S. (2018). Single-cell RNA-Seq of mouse dopaminergic neurons informs candidate gene selection for sporadic Parkinson disease. Am. J. Hum. Genet. *102*, 427–446.
70. Sugiaman-Trapman, D., Vitezic, M., Jouhilahti, E.-M., Mathelier, A., Lauter, G., Misra, S., Daub, C.O., Kere, J., and Swoboda, P. (2018). Characterization of the human RFX transcription factor family by regulatory and target gene analysis. BMC Genomics *19*, 181.
71. Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. Nature *403*, 916–921.

72. Kittappa, R., Chang, W.W., Awatramani, R.B., and McKay, R.D.G. (2007). The foxa2 gene controls the birth and spontaneous degeneration of dopamine neurons in old age. PLoS Biol. *5*, e325.

73. Caiazzo, M., Dell'Anno, M.T., Dvoretskova, E., Lazarevic, D., Taverna, S., Leo, D., Sotnikova, T.D., Menegon, A., Roncaglia, P., Colciago, G., et al. (2011). Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. Nature *476*, 224–227.

74. Sung, M.-H., Guertin, M.J., Baek, S., and Hager, G.L. (2014). DNase footprint signatures are dictated by factor dynamics and DNA sequence. Mol. Cell *56*, 275–285.

75. Zarow, C., Lyness, S.A., Mortimer, J.A., and Chui, H.C. (2003). Neuronal loss is greater in the locus coeruleus than nucleus basalis and substantia nigra in Alzheimer and Parkinson diseases. Arch. Neurol. *60*, 337–341.

76. Kastenhuber, E., Kratochwil, C.F., Ryu, S., Schweitzer, J., and Driever, W. (2010). Genetic dissection of dopaminergic and noradrenergic contributions to catecholaminergic tracts in early larval zebrafish. J. Comp. Neurol. *518*, 439–458.

77. Rink, E., and Wullimann, M.F. (2001). The teleostean (zebrafish) dopaminergic system ascending to the subpallium (striatum) is located in the basal diencephalon (posterior tuberculum). Brain Res. *889*, 316–330.

78. Seidel, K., Mahlke, J., Siswanto, S., Krüger, R., Heinsen, H., Auburger, G., Bouzrou, M., Grinberg, L.T., Wicht, H., Korf, H.-W., et al. (2015). The brainstem pathologies of Parkinson's disease and dementia with Lewy bodies. Brain Pathol. *25*, 121–135.

79. Wakabayashi, K., Mori, F., Tanji, K., Orimo, S., and Takahashi, H. (2010). Involvement of the peripheral nervous system in synucleinopathies, tauopathies and other neurodegenerative proteinopathies of the brain. Acta Neuropathol. *120*, 1–12.

80. Wakabayashi, K., and Takahashi, H. (1997). Neuropathology of autonomic nervous system in Parkinson's disease. Eur. Neurol. *38* (*Suppl 2*), 2–7.

81. Braak, H., Braak, E., Yilmazer, D., de Vos, R.A., Jansen, E.N., Bohl, J., and Jellinger, K. (1994). Amygdala pathology in Parkinson's disease. Acta Neuropathol. *88*, 493–500.

82. Langston, J.W., and Forno, L.S. (1978). The hypothalamus in Parkinson disease. Ann. Neurol. *3*, 129–133.

83. Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D.U., Zhang, Y., Sos, B.C., Afzal, V., Dickel, D.E., et al. (2018). Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat. Neurosci. *21*, 432–439.

84. Greffard, S., Verny, M., Bonnet, A.-M., Beinis, J.-Y., Gallinari, C., Meaume, S., Piette, F., Hauw, J.-J., and Duyckaerts, C. (2006). Motor score of the Unified Parkinson Disease Rating Scale as a good predictor of Lewy body-associated neuronal loss in the substantia nigra. Arch. Neurol. *63*, 584–588.

85. Soldner, F., Stelzer, Y., Shivalila, C.S., Abraham, B.J., Latourelle, J.C., Barrasa, M.I., Goldmann, J., Myers, R.H., Young, R.A., and Jaenisch, R. (2016). Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. Nature *533*, 95–99.

86. Gupta, R.M., Hadaya, J., Trehan, A., Zekavat, S.M., Roselli, C., Klarin, D., Emdin, C.A., Hilvering, C.R.E., Bianchi, V., Mueller, C., et al. (2017). A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. Cell *170*, 522–533.e15.

The American Journal of Human Genetics, Volume *103*

Supplemental Data

Parkinson-Associated *SNCA* Enhancer Variants Revealed by Open Chromatin in Mouse Dopamine Neurons

Sarah A. McClymont, Paul W. Hook, Alexandra I. Soto, Xylena Reed, William D. Law, Samuel J. Kerans, Eric L. Waite, Nicole J. Briceno, Joey F. Thole, Michael G. Heckman, Nancy N. Diehl, Zbigniew K. Wszolek, Cedric D. Moore, Heng Zhu, Jennifer A. Akiyama, Diane E. Dickel, Axel Visel, Len A. Pennacchio, Owen A. Ross, Michael A. Beer, and Andrew S. McCallion

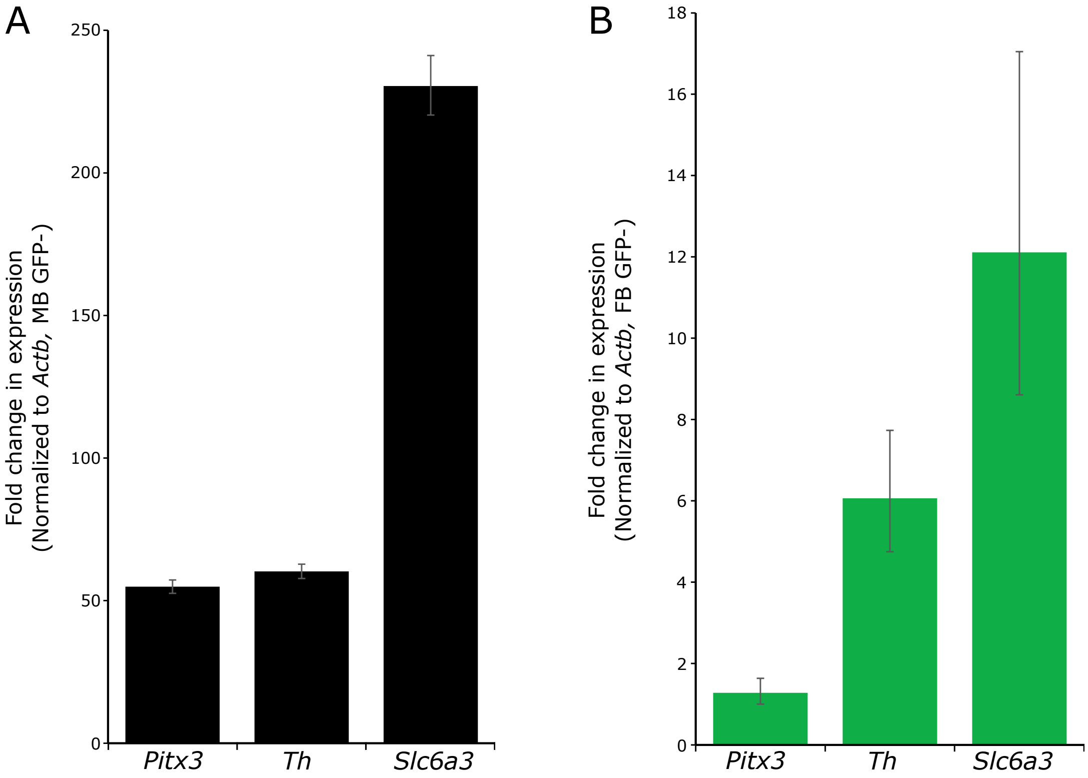


Figure S1 RT-qPCR of key DA neuron markers

Expression of key DA neuron markers (*Pitx3, Th, Slc6a3*) in MB FACS-isolated (**A**) and FB FACS-isolated (**B**) cells confirms isolation of purified MB and FB DA neurons. Error bars represent the fold change range after incorporation of the standard deviation values (n = 3 technical replicates).

A

	MB1	MB2	MB3	FB1	FB2	FB3
Reads sequenced	27,464,563	28,786,058	42,760,942	16,849,958	50,580,772	21,497,236
% aligned	97.39%	97.28%	90.89%	97.97%	96.06%	96.68%
% duplicate	25.04%	9.79%	39.34%	5.46%	16.84%	8.46%
% mitochondrial	7.48%	5.80%	1.79%	3.32%	4.93%	3.51%
Peaks called	62,646	74,018	22,858	48,525	45,222	52,493
% blacklisted	1.48%	1.16%	4.77%	1.84%	1.91%	1.66%
Fraction of reads in peaks	30.65%	37.62%	10.49%	25.04%	22.16%	25.06%

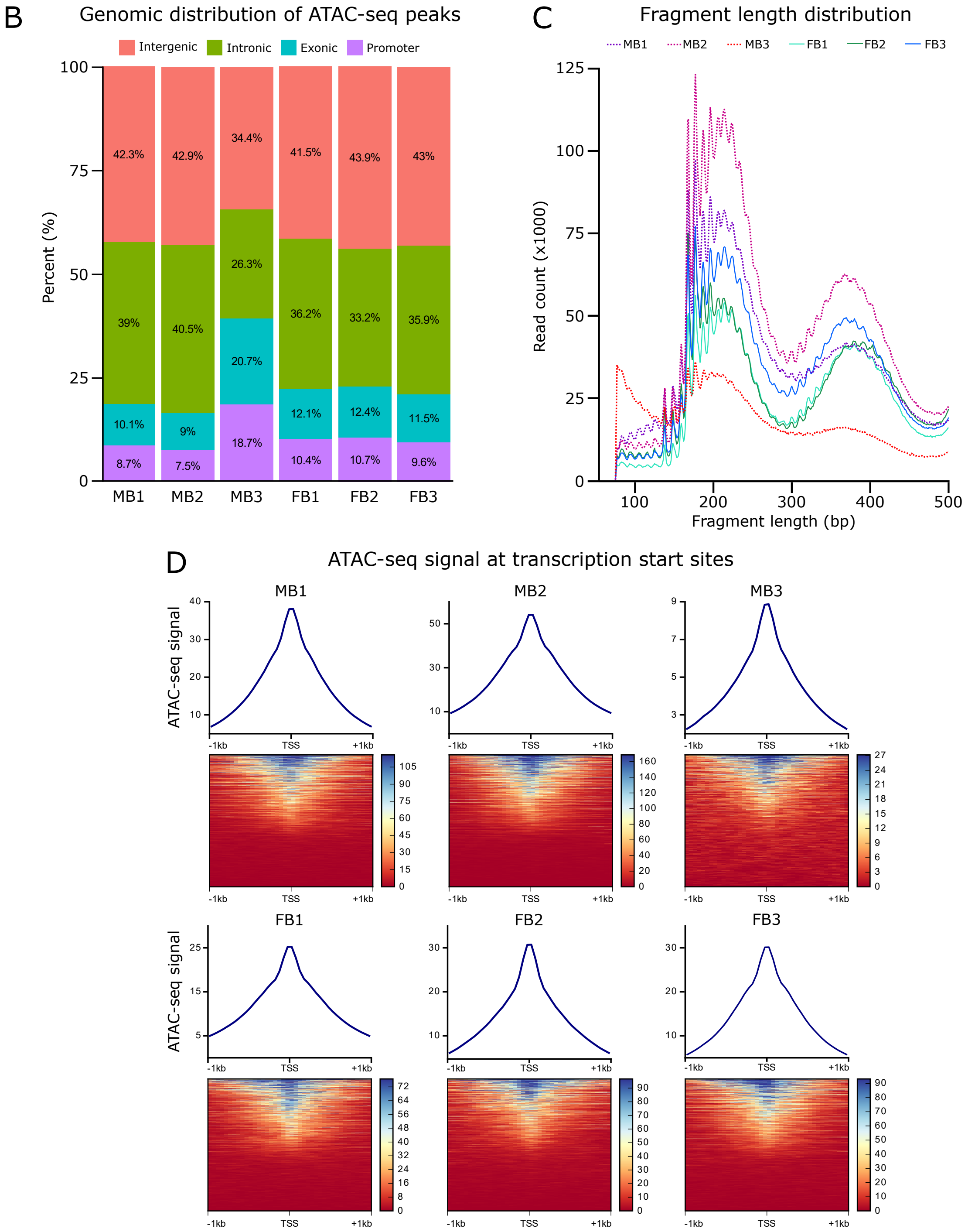


Figure S2 *in silico* quality control metrics for the ATAC-seq libraries

(**A**) Sequencing statistics for the ATAC-seq libraries indicate all six libraries are of sufficient quality. (**B**) The genomic distribution of ATAC-seq peaks indicate a preference for promoters and intergenic regions. (**C**) The fragment length distribution of the ATAC-seq libraries indicate the presence of a nucleosome ladder (with one nucleosome fragments, perhaps, being selected against in the bead clean-up). (**D**) All ATAC-seq libraries display an abundance of reads overlapping gene promoters, genome-wide.

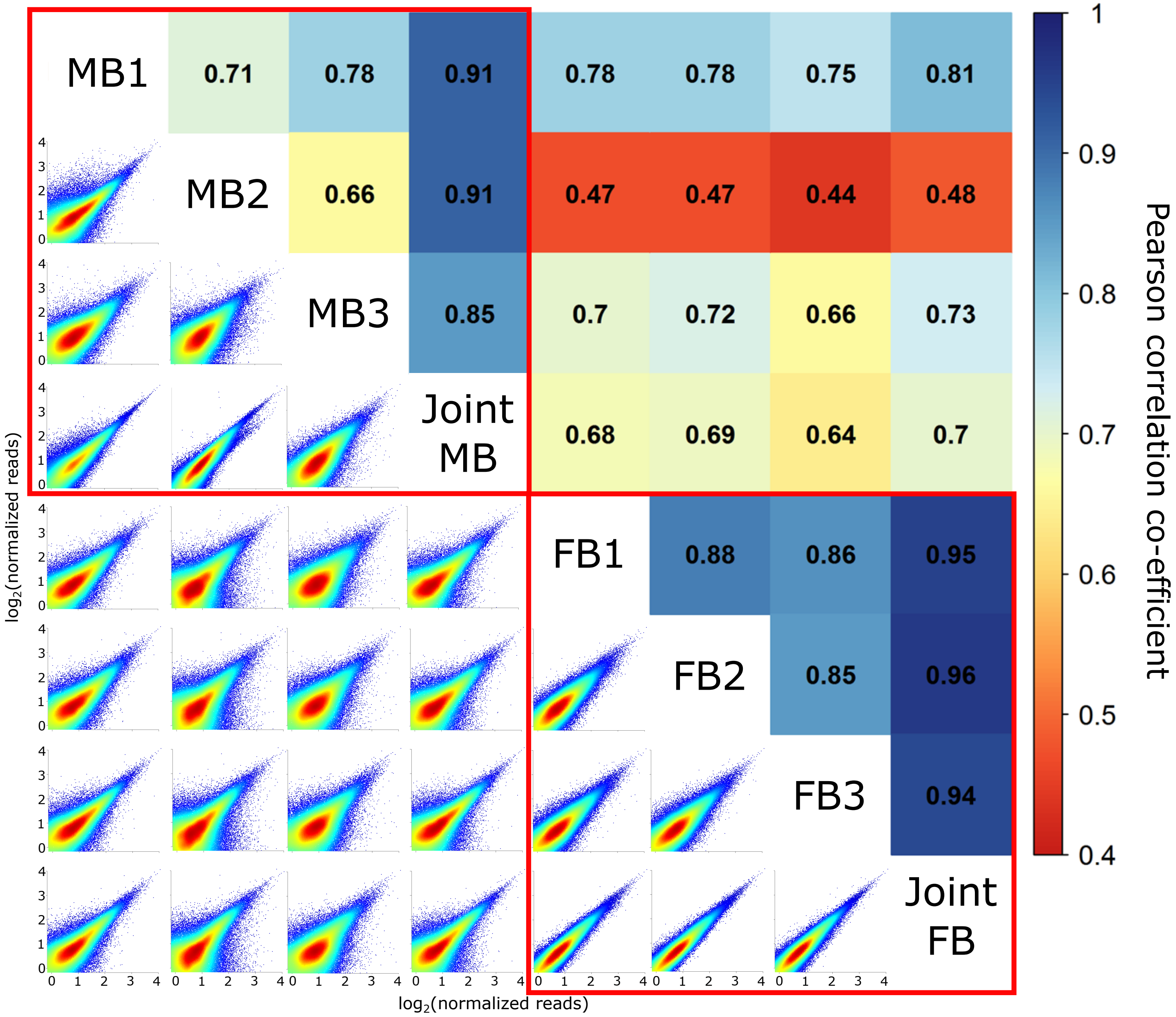


Figure S3 Correlation analysis of all ATAC-seq libraries

Genome-wide correlation within replicates (red boxed areas) and between brain regions indicate there is strong correlation within a brain region across replicates, with correlation to a lesser extent between brain regions.

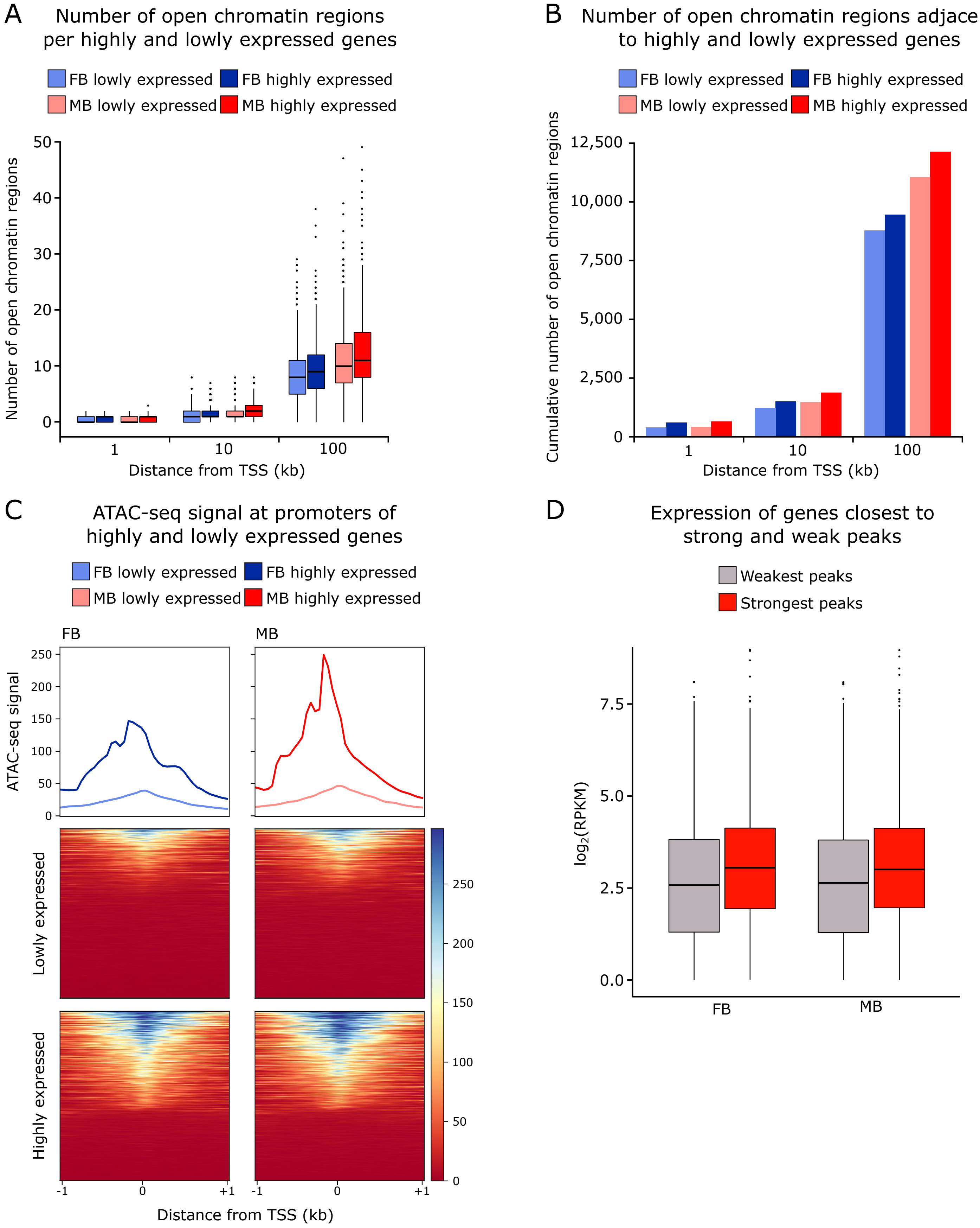


Figure S4 Relating RNA-seq and ATAC-seq data

Broad analyses indicate that highly expressed genes are under greater regulatory control, in that there are more proximal regulatory elements (**A, B**) and their promoters are more open (**C**) compared to lowly expressed genes. (**D**) Additionally, the genes closest to the strongest ATAC-seq peaks are more highly expressed than those adjacent the weakest peaks.

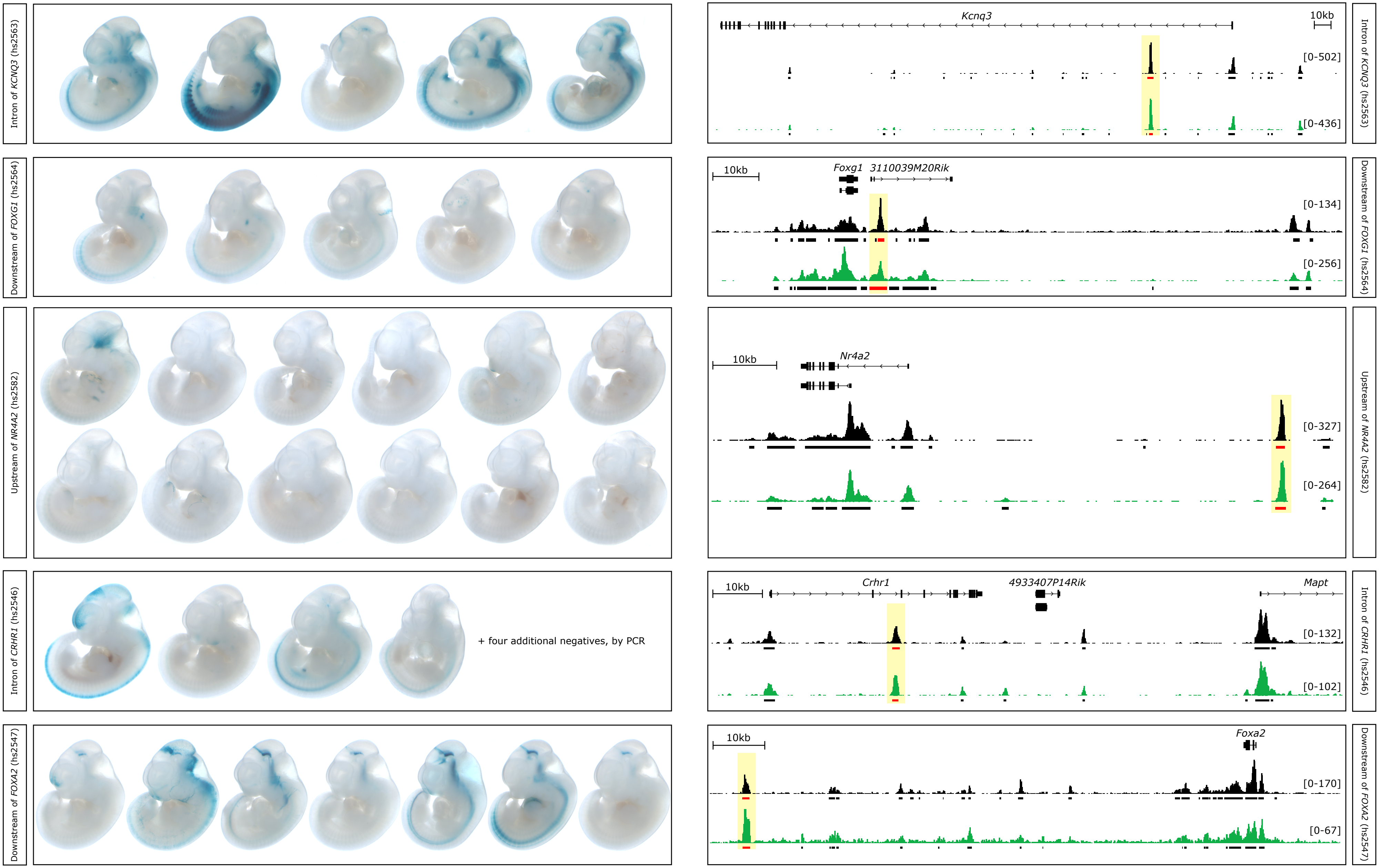


Figure S5 All *lacZ* reporter mice and the mouse genomic locations of the putative CREs

All transgenic mouse embryos assayed for *lacZ* reporter activity for each of the five putative CREs tested *in vivo* (left) and the genomic location and context of those putative CREs (right). MB: Black track, FB: Green track. Red peaks in yellow boxes: The putative CREs that were lifted over to hg19 and tested *in vivo*.

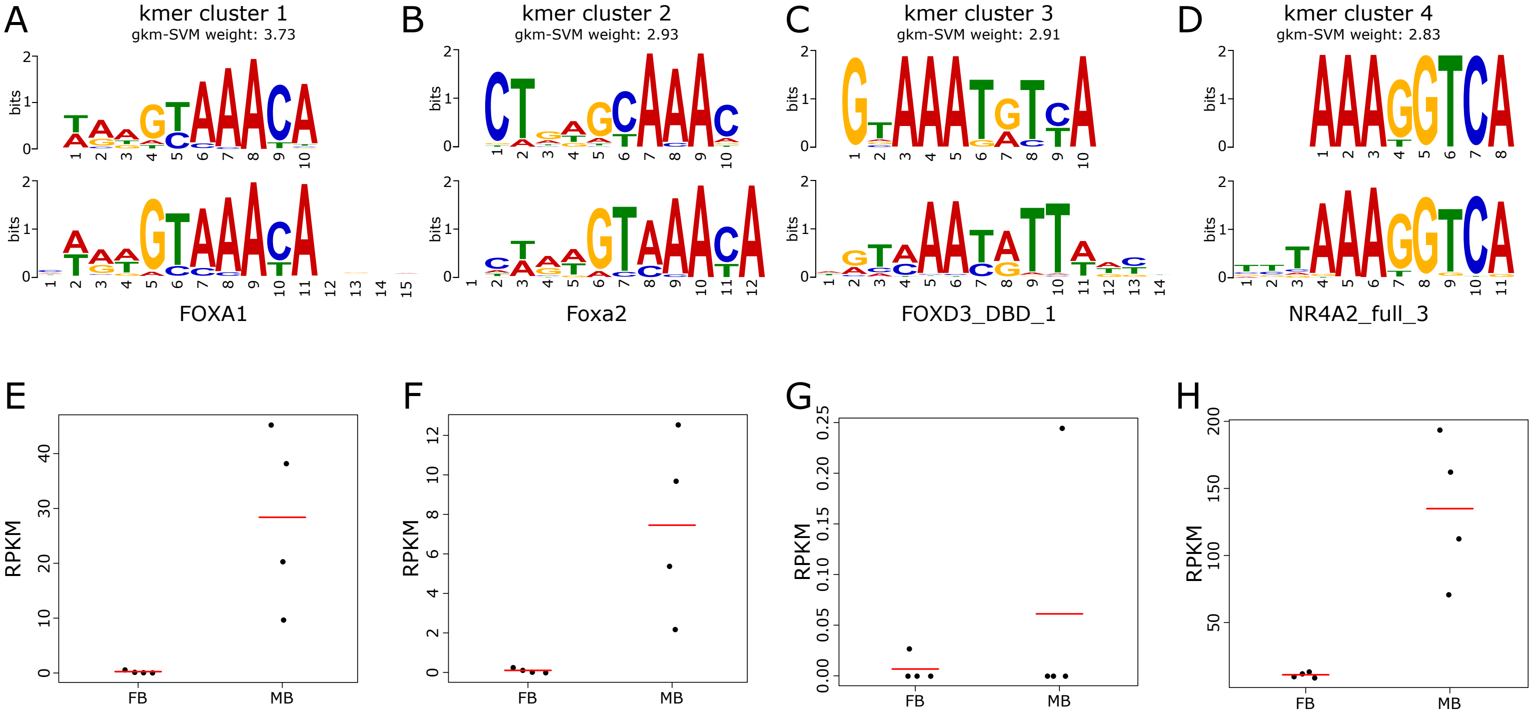
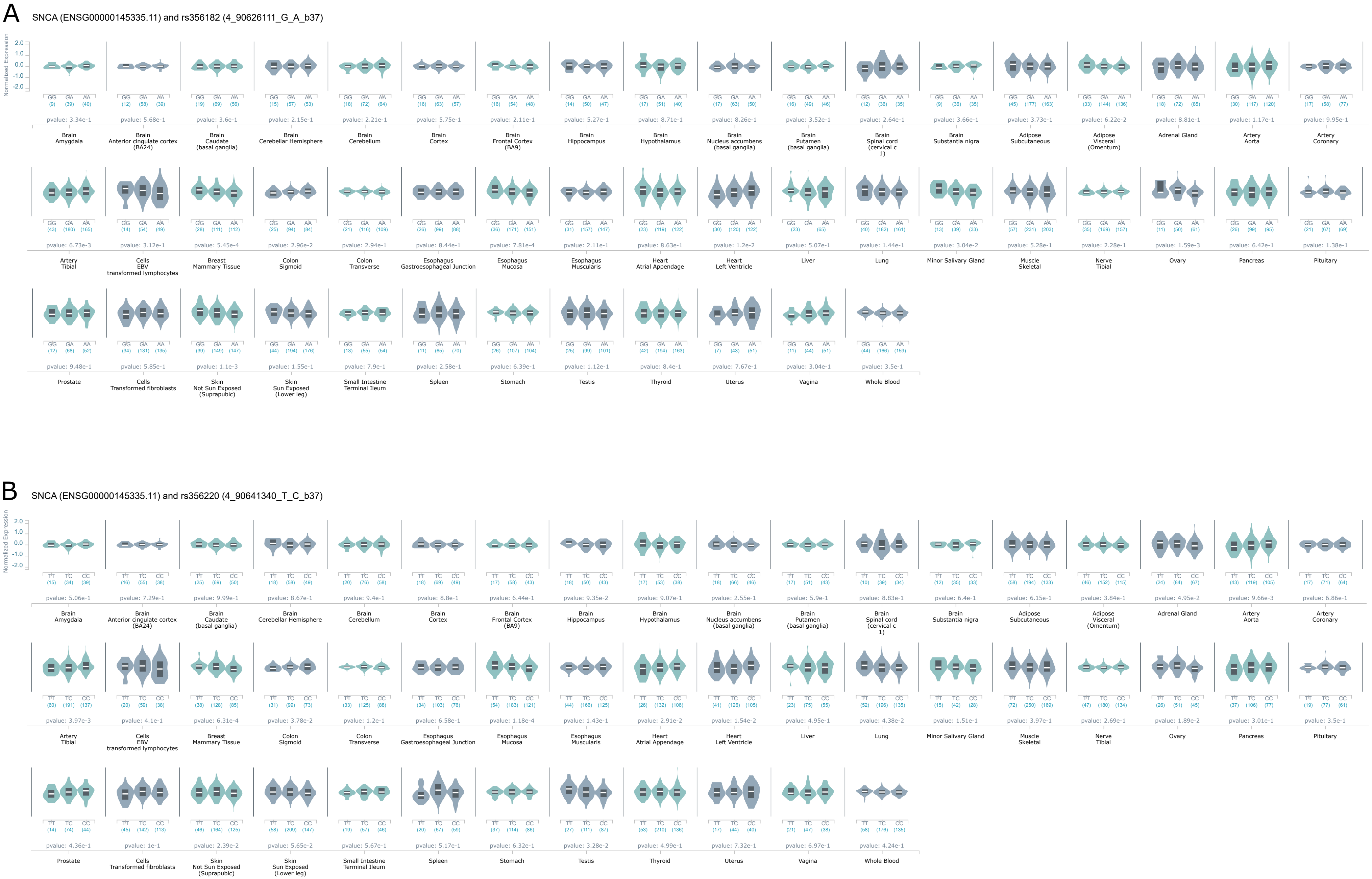
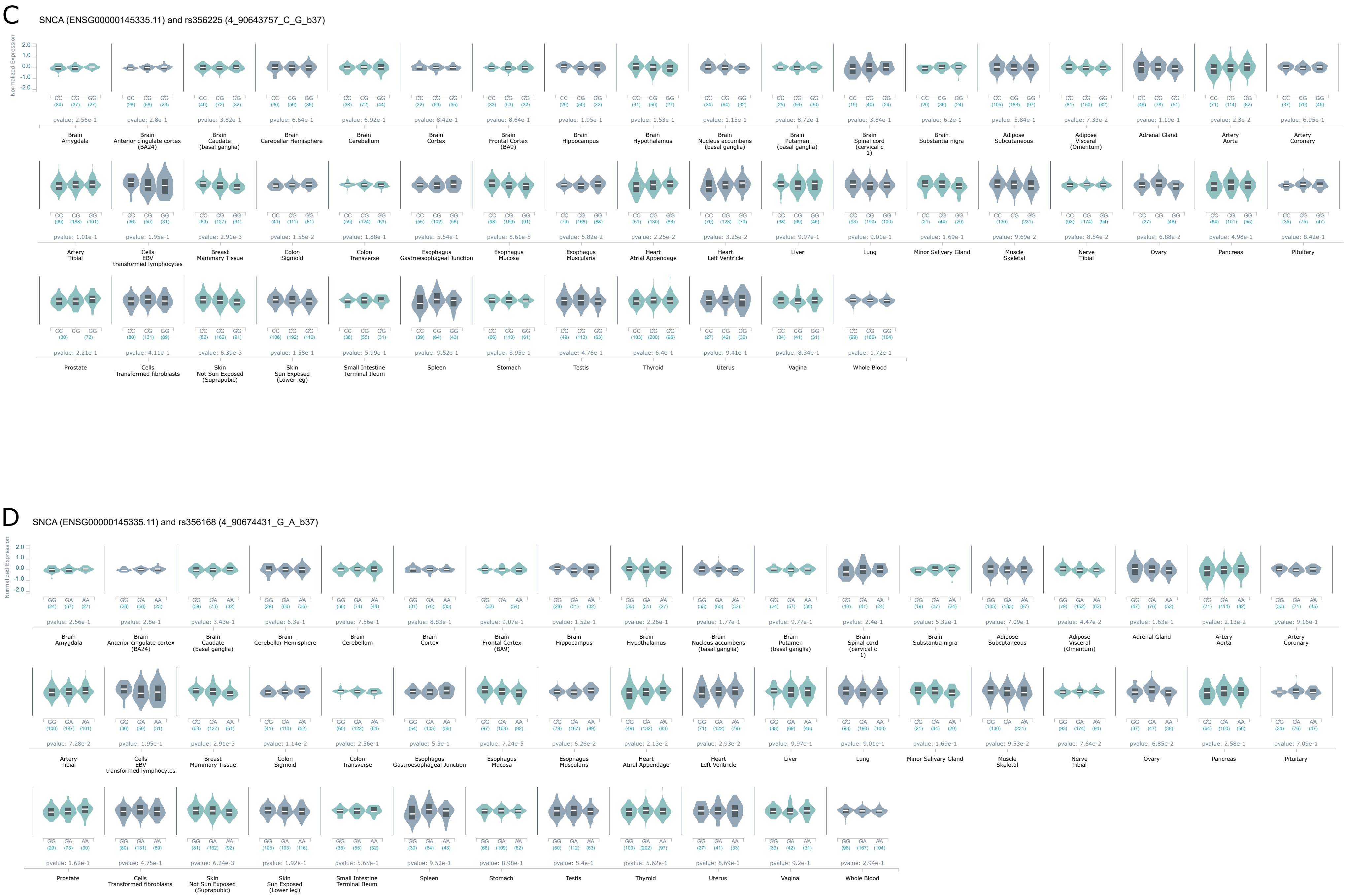
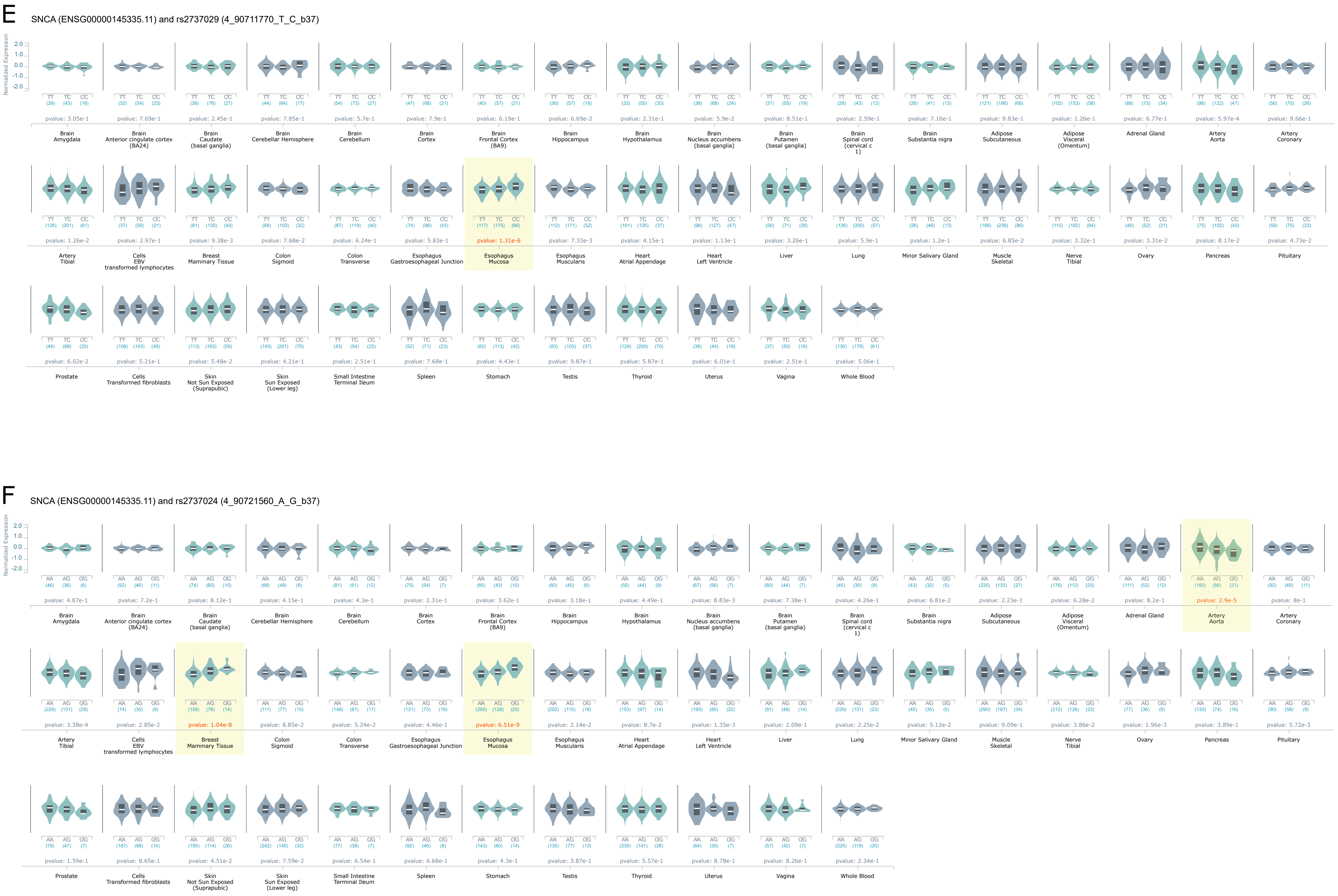


Figure S6 Motif analysis identifies transcription factors (TFs) important specifically for MB regulatory potential (**A-D**) The motifs with the greatest regulatory potential specific to the MB and the potential TF matching that motif were identified. (**E-H**) Expression analysis of these identified TFs confirm the sequence based analysis for Foxa1 (**E**), Foxa2 (**F**), and Nr4a2 (**H**). *Foxd3* (**G**), while prioritized on the basis of sequence composition, is not expressed in MB or FB DA neurons (≤1 RPKM) and was likely identified as a consequence of the sequence degeneracy within TF families.


A
SNCA (ENSG00000145335.11) and rs356182 (4_90626111_G_A_b37)
Normalized Expression
2.0
1.0
0.0
-1.0
-2.0
GG GA AA (9) (39) (40) pvalue: 3.34e-1 Brain Amygdala
GG GA AA (12) (58) (39) pvalue: 5.68e-1 Brain Anterior cingulate cortex (BA24)
GG GA AA (19) (69) (56) pvalue: 3.6e-1 Brain Caudate (basal ganglia)
GG GA AA (15) (57) (53) pvalue: 2.15e-1 Brain Cerebellar Hemisphere
GG GA AA (18) (72) (64) pvalue: 2.21e-1 Brain Cerebellum
GG GA AA (16) (63) (57) pvalue: 5.75e-1 Brain Cortex
GG GA AA (16) (54) (48) pvalue: 2.11e-1 Brain Frontal Cortex (BA9)
GG GA AA (14) (50) (47) pvalue: 5.27e-1 Brain Hippocampus
GG GA AA (17) (51) (40) pvalue: 8.71e-1 Brain Hypothalamus
GG GA AA (17) (63) (50) pvalue: 8.26e-1 Brain Nucleus accumbens (basal ganglia)
GG GA AA (16) (49) (46) pvalue: 3.52e-1 Brain Putamen (basal ganglia)
GG GA AA (12) (36) (35) pvalue: 2.64e-1 Brain Spinal cord (cervical c 1)
GG GA AA (9) (36) (35) pvalue: 3.66e-1 Brain Substantia nigra
GG GA AA (45) (177) (163) pvalue: 3.73e-1 Adipose Subcutaneous
GG GA AA (33) (144) (136) pvalue: 6.22e-2 Adipose Visceral (Omentum)
GG GA AA (18) (72) (85) pvalue: 8.81e-1 Adrenal Gland
GG GA AA (30) (117) (120) pvalue: 1.17e-1 Artery Aorta
GG GA AA (17) (58) (77) pvalue: 9.95e-1 Artery Coronary
GG GA AA (43) (180) (165) pvalue: 6.73e-3 Artery Tibial
GG GA AA (14) (54) (49) pvalue: 3.12e-1 Cells EBV transformed lymphocytes
GG GA AA (28) (111) (112) pvalue: 5.45e-4 Breast Mammary Tissue
GG GA AA (25) (94) (84) pvalue: 2.96e-2 Colon Sigmoid
GG GA AA (21) (116) (109) pvalue: 2.94e-1 Colon Transverse
GG GA AA (26) (99) (88) pvalue: 8.44e-1 Esophagus Gastroesophageal Junction
GG GA AA (36) (171) (151) pvalue: 7.81e-4 Esophagus Mucosa
GG GA AA (31) (157) (147) pvalue: 2.11e-1 Esophagus Muscularis
GG GA AA (23) (119) (122) pvalue: 8.63e-1 Heart Atrial Appendage
GG GA AA (30) (120) (122) pvalue: 1.2e-2 Heart Left Ventricle
GG GA AA (23) (65) pvalue: 5.07e-1 Liver
GG GA AA (40) (182) (161) pvalue: 1.44e-1 Lung
GG GA AA (13) (39) (33) pvalue: 3.04e-2 Minor Salivary Gland
GG GA AA (57) (231) (203) pvalue: 5.28e-1 Muscle Skeletal
GG GA AA (35) (169) (157) pvalue: 2.28e-1 Nerve Tibial
GG GA AA (11) (50) (61) pvalue: 1.59e-3 Ovary
GG GA AA (26) (99) (95) pvalue: 6.42e-1 Pancreas
GG GA AA (21) (67) (69) pvalue: 1.38e-1 Pituitary
GG GA AA (12) (68) (52) pvalue: 9.48e-1 Prostate
GG GA AA (34) (131) (135) pvalue: 5.85e-1 Cells Transformed fibroblasts
GG GA AA (39) (149) (147) pvalue: 1.1e-3 Skin Not Sun Exposed (Suprapubic)
GG GA AA (44) (194) (176) pvalue: 1.55e-1 Skin Sun Exposed (Lower leg)
GG GA AA (13) (55) (54) pvalue: 7.9e-1 Small Intestine Terminal Ileum
GG GA AA (11) (65) (70) pvalue: 2.58e-1 Spleen
GG GA AA (26) (107) (104) pvalue: 6.39e-1 Stomach
GG GA AA (25) (99) (101) pvalue: 1.12e-1 Testis
GG GA AA (42) (194) (163) pvalue: 8.4e-1 Thyroid
GG GA AA (7) (43) (51) pvalue: 7.67e-1 Uterus
GG GA AA (11) (44) (51) pvalue: 3.04e-1 Vagina
GG GA AA (44) (166) (159) pvalue: 3.5e-1 Whole Blood
B
SNCA (ENSG00000145335.11) and rs356220 (4_90641340_T_C_b37)
Normalized Expression
2.0
1.0
0.0
-1.0
-2.0
TT TC CC (15) (34) (39) pvalue: 5.06e-1 Brain Amygdala
TT TC CC (16) (55) (38) pvalue: 7.29e-1 Brain Anterior cingulate cortex (BA24)
TT TC CC (25) (69) (50) pvalue: 9.99e-1 Brain Caudate (basal ganglia)
TT TC CC (18) (58) (49) pvalue: 8.67e-1 Brain Cerebellar Hemisphere
TT TC CC (20) (76) (58) pvalue: 9.4e-1 Brain Cerebellum
TT TC CC (18) (69) (49) pvalue: 8.8e-1 Brain Cortex
TT TC CC (17) (58) (43) pvalue: 6.44e-1 Brain Frontal Cortex (BA9)
TT TC CC (18) (50) (43) pvalue: 9.35e-2 Brain Hippocampus
TT TC CC (17) (53) (38) pvalue: 9.07e-1 Brain Hypothalamus
TT TC CC (18) (66) (46) pvalue: 2.55e-1 Brain Nucleus accumbens (basal ganglia)
TT TC CC (17) (51) (43) pvalue: 5.9e-1 Brain Putamen (basal ganglia)
TT TC CC (10) (39) (34) pvalue: 8.83e-1 Brain Spinal cord (cervical c 1)
TT TC CC (12) (35) (33) pvalue: 6.4e-1 Brain Substantia nigra
TT TC CC (58) (194) (133) pvalue: 6.15e-1 Adipose Subcutaneous
TT TC CC (46) (152) (115) pvalue: 3.84e-1 Adipose Visceral (Omentum)
TT TC CC (24) (84) (67) pvalue: 4.95e-2 Adrenal Gland
TT TC CC (43) (119) (105) pvalue: 9.66e-3 Artery Aorta
TT TC CC (17) (71) (64) pvalue: 6.86e-1 Artery Coronary
TT TC CC (60) (191) (137) pvalue: 3.97e-3 Artery Tibial
TT TC CC (20) (59) (38) pvalue: 4.1e-1 Cells EBV transformed lymphocytes
TT TC CC (38) (128) (85) pvalue: 6.31e-4 Breast Mammary Tissue
TT TC CC (31) (99) (73) pvalue: 3.78e-2 Colon Sigmoid
TT TC CC (33) (125) (88) pvalue: 1.2e-1 Colon Transverse
TT TC CC (34) (103) (76) pvalue: 6.58e-1 Esophagus Gastroesophageal Junction
TT TC CC (54) (183) (121) pvalue: 1.18e-4 Esophagus Mucosa
TT TC CC (44) (166) (125) pvalue: 1.43e-1 Esophagus Muscularis
TT TC CC (26) (132) (106) pvalue: 2.91e-2 Heart Atrial Appendage
TT TC CC (41) (126) (105) pvalue: 1.54e-2 Heart Left Ventricle
TT TC CC (23) (75) (55) pvalue: 4.95e-1 Liver
TT TC CC (52) (196) (135) pvalue: 4.38e-2 Lung
TT TC CC (15) (42) (28) pvalue: 1.51e-1 Minor Salivary Gland
TT TC CC (72) (250) (169) pvalue: 3.97e-1 Muscle Skeletal
TT TC CC (47) (180) (134) pvalue: 2.69e-1 Nerve Tibial
TT TC CC (26) (51) (45) pvalue: 1.89e-2 Ovary
TT TC CC (37) (106) (77) pvalue: 3.01e-1 Pancreas
TT TC CC (19) (77) (61) pvalue: 3.5e-1 Pituitary
TT TC CC (14) (74) (44) pvalue: 4.36e-1 Prostate
TT TC CC (45) (142) (113) pvalue: 1e-1 Cells Transformed fibroblasts
TT TC CC (46) (164) (125) pvalue: 2.39e-2 Skin Not Sun Exposed (Suprapubic)
TT TC CC (58) (209) (147) pvalue: 5.65e-2 Skin Sun Exposed (Lower leg)
TT TC CC (19) (57) (46) pvalue: 5.67e-1 Small Intestine Terminal Ileum
TT TC CC (20) (67) (59) pvalue: 5.17e-1 Spleen
TT TC CC (37) (114) (86) pvalue: 6.32e-1 Stomach
TT TC CC (27) (111) (87) pvalue: 3.28e-2 Testis
TT TC CC (53) (210) (136) pvalue: 4.99e-1 Thyroid
TT TC CC (17) (44) (40) pvalue: 7.32e-1 Uterus
TT TC CC (21) (47) (38) pvalue: 6.97e-1 Vagina
TT TC CC (58) (176) (135) pvalue: 4.24e-1 Whole Blood


C
SNCA (ENSG00000145335.11) and rs356225 (4_90643757_C_G_b37)
Normalized Expression
2.0
1.0
0.0
-1.0
-2.0
CC CG GG (24) (37) (27) pvalue: 2.56e-1 Brain Amygdala
CC CG GG (28) (58) (23) pvalue: 2.8e-1 Brain Anterior cingulate cortex (BA24)
CC CG GG (40) (72) (32) pvalue: 3.82e-1 Brain Caudate (basal ganglia)
CC CG GG (30) (59) (36) pvalue: 6.64e-1 Brain Cerebellar Hemisphere
CC CG GG (38) (72) (44) pvalue: 6.92e-1 Brain Cerebellum
CC CG GG (32) (69) (35) pvalue: 8.42e-1 Brain Cortex
CC CG GG (33) (53) (32) pvalue: 8.64e-1 Brain Frontal Cortex (BA9)
CC CG GG (29) (50) (32) pvalue: 1.95e-1 Brain Hippocampus
CC CG GG (31) (50) (27) pvalue: 1.53e-1 Brain Hypothalamus
CC CG GG (34) (64) (32) pvalue: 1.15e-1 Brain Nucleus accumbens (basal ganglia)
CC CG GG (25) (56) (30) pvalue: 8.72e-1 Brain Putamen (basal ganglia)
CC CG GG (19) (40) (24) pvalue: 3.84e-1 Brain Spinal cord (cervical c 1)
CC CG GG (20) (36) (24) pvalue: 6.2e-1 Brain Substantia nigra
CC CG GG (105) (183) (97) pvalue: 5.84e-1 Adipose Subcutaneous
CC CG GG (81) (150) (82) pvalue: 7.33e-2 Adipose Visceral (Omentum)
CC CG GG (46) (78) (51) pvalue: 1.19e-1 Adrenal Gland
CC CG GG (71) (114) (82) pvalue: 2.3e-2 Artery Aorta
CC CG GG (37) (70) (45) pvalue: 6.95e-1 Artery Coronary
CC CG GG (99) (188) (101) pvalue: 1.01e-1 Artery Tibial
CC CG GG (36) (50) (31) pvalue: 1.95e-1 Cells EBV transformed lymphocytes
CC CG GG (63) (127) (61) pvalue: 2.91e-3 Breast Mammary Tissue
CC CG GG (41) (111) (51) pvalue: 1.55e-2 Colon Sigmoid
CC CG GG (59) (124) (63) pvalue: 1.88e-1 Colon Transverse
CC CG GG (55) (102) (56) pvalue: 5.54e-1 Esophagus Gastroesophageal Junction
CC CG GG (98) (169) (91) pvalue: 8.61e-5 Esophagus Mucosa
CC CG GG (79) (168) (88) pvalue: 5.82e-2 Esophagus Muscularis
CC CG GG (51) (130) (83) pvalue: 2.25e-2 Heart Atrial Appendage
CC CG GG (70) (123) (79) pvalue: 3.25e-2 Heart Left Ventricle
CC CG GG (38) (69) (46) pvalue: 9.97e-1 Liver
CC CG GG (93) (190) (100) pvalue: 9.01e-1 Lung
CC CG GG (21) (44) (20) pvalue: 1.69e-1 Minor Salivary Gland
CC CG GG (130) (231) pvalue: 9.69e-2 Muscle Skeletal
CC CG GG (93) (174) (94) pvalue: 8.54e-2 Nerve Tibial
CC CG GG (37) (48) pvalue: 6.88e-2 Ovary
CC CG GG (64) (101) (55) pvalue: 4.98e-1 Pancreas
CC CG GG (35) (75) (47) pvalue: 8.42e-1 Pituitary
CC CG GG (30) (72) pvalue: 2.21e-1 Prostate
CC CG GG (80) (131) (89) pvalue: 4.11e-1 Cells Transformed fibroblasts
CC CG GG (82) (162) (91) pvalue: 6.39e-3 Skin Not Sun Exposed (Suprapubic)
CC CG GG (106) (192) (116) pvalue: 1.58e-1 Skin Sun Exposed (Lower leg)
CC CG GG (36) (55) (31) pvalue: 5.99e-1 Small Intestine Terminal Ileum
CC CG GG (39) (64) (43) pvalue: 9.52e-1 Spleen
CC CG GG (66) (110) (61) pvalue: 8.95e-1 Stomach
CC CG GG (49) (113) (63) pvalue: 4.76e-1 Testis
CC CG GG (103) (200) (96) pvalue: 6.4e-1 Thyroid
CC CG GG (27) (42) (32) pvalue: 9.41e-1 Uterus
CC CG GG (34) (41) (31) pvalue: 8.34e-1 Vagina
CC CG GG (99) (166) (104) pvalue: 1.72e-1 Whole Blood
D
SNCA (ENSG00000145335.11) and rs356168 (4_90674431_G_A_b37)
Normalized Expression
2.0
1.0
0.0
-1.0
-2.0
GG GA AA (24) (37) (27) pvalue: 2.56e-1 Brain Amygdala
GG GA AA (28) (58) (23) pvalue: 2.8e-1 Brain Anterior cingulate cortex (BA24)
GG GA AA (39) (73) (32) pvalue: 3.43e-1 Brain Caudate (basal ganglia)
GG GA AA (29) (60) (36) pvalue: 6.3e-1 Brain Cerebellar Hemisphere
GG GA AA (36) (74) (44) pvalue: 7.56e-1 Brain Cerebellum
GG GA AA (31) (70) (35) pvalue: 8.83e-1 Brain Cortex
GG GA AA (32) (54) pvalue: 9.07e-1 Brain Frontal Cortex (BA9)
GG GA AA (28) (51) (32) pvalue: 1.52e-1 Brain Hippocampus
GG GA AA (30) (51) (27) pvalue: 2.26e-1 Brain Hypothalamus
GG GA AA (33) (65) (32) pvalue: 1.77e-1 Brain Nucleus accumbens (basal ganglia)
GG GA AA (24) (57) (30) pvalue: 9.77e-1 Brain Putamen (basal ganglia)
GG GA AA (18) (41) (24) pvalue: 2.4e-1 Brain Spinal cord (cervical c 1)
GG GA AA (19) (37) (24) pvalue: 5.32e-1 Brain Substantia nigra
GG GA AA (105) (183) (97) pvalue: 7.09e-1 Adipose Subcutaneous
GG GA AA (79) (152) (82) pvalue: 4.47e-2 Adipose Visceral (Omentum)
GG GA AA (47) (76) (52) pvalue: 1.63e-1 Adrenal Gland
GG GA AA (71) (114) (82) pvalue: 2.13e-2 Artery Aorta
GG GA AA (36) (71) (45) pvalue: 9.16e-1 Artery Coronary
GG GA AA (100) (187) (101) pvalue: 7.28e-2 Artery Tibial
GG GA AA (36) (50) (31) pvalue: 1.95e-1 Cells EBV transformed lymphocytes
GG GA AA (63) (127) (61) pvalue: 2.91e-3 Breast Mammary Tissue
GG GA AA (41) (110) (52) pvalue: 1.14e-2 Colon Sigmoid
GG GA AA (60) (122) (64) pvalue: 2.56e-1 Colon Transverse
GG GA AA (54) (103) (56) pvalue: 5.3e-1 Esophagus Gastroesophageal Junction
GG GA AA (97) (169) (92) pvalue: 7.24e-5 Esophagus Mucosa
GG GA AA (79) (167) (89) pvalue: 6.26e-2 Esophagus Muscularis
GG GA AA (49) (132) (83) pvalue: 2.13e-2 Heart Atrial Appendage
GG GA AA (71) (122) (79) pvalue: 2.93e-2 Heart Left Ventricle
GG GA AA (38) (69) (46) pvalue: 9.97e-1 Liver
GG GA AA (93) (190) (100) pvalue: 9.01e-1 Lung
GG GA AA (21) (44) (20) pvalue: 1.69e-1 Minor Salivary Gland
GG GA AA (130) (231) pvalue: 9.53e-2 Muscle Skeletal
GG GA AA (93) (174) (94) pvalue: 7.64e-2 Nerve Tibial
GG GA AA (37) (47) (38) pvalue: 6.85e-2 Ovary
GG GA AA (64) (100) (56) pvalue: 2.58e-1 Pancreas
GG GA AA (34) (76) (47) pvalue: 7.09e-1 Pituitary
GG GA AA (29) (73) (30) pvalue: 1.62e-1 Prostate
GG GA AA (80) (131) (89) pvalue: 4.75e-1 Cells Transformed fibroblasts
GG GA AA (81) (162) (92) pvalue: 6.24e-3 Skin Not Sun Exposed (Suprapubic)
GG GA AA (105) (193) (116) pvalue: 1.92e-1 Skin Sun Exposed (Lower leg)
GG GA AA (35) (55) (32) pvalue: 5.65e-1 Small Intestine Terminal Ileum
GG GA AA (39) (64) (43) pvalue: 9.52e-1 Spleen
GG GA AA (66) (109) (62) pvalue: 8.98e-1 Stomach
GG GA AA (50) (112) (63) pvalue: 5.4e-1 Testis
GG GA AA (100) (202) (97) pvalue: 5.62e-1 Thyroid
GG GA AA (27) (41) (33) pvalue: 8.69e-1 Uterus
GG GA AA (33) (42) (31) pvalue: 9.2e-1 Vagina
GG GA AA (98) (167) (104) pvalue: 2.94e-1 Whole Blood


E
SNCA (ENSG00000145335.11) and rs2737029 (4_90711770_T_C_b37)
Normalized Expression
2.0
1.0
0.0
-1.0
-2.0
TT (29) TC (43) CC (16) pvalue: 3.05e-1 Brain Amygdala
TT (32) TC (54) CC (23) pvalue: 7.69e-1 Brain Anterior cingulate cortex (BA24)
TT (39) TC (78) CC (27) pvalue: 2.45e-1 Brain Caudate (basal ganglia)
TT (44) TC (64) CC (17) pvalue: 7.85e-1 Brain Cerebellar Hemisphere
TT (54) TC (73) CC (27) pvalue: 5.7e-1 Brain Cerebellum
TT (47) TC (68) CC (21) pvalue: 7.9e-1 Brain Cortex
TT (40) TC (57) CC (21) pvalue: 6.19e-1 Brain Frontal Cortex (BA9)
TT (36) TC (57) CC (18) pvalue: 6.69e-2 Brain Hippocampus
TT (33) TC (55) CC (20) pvalue: 2.31e-1 Brain Hypothalamus
TT (38) TC (68) CC (24) pvalue: 5.9e-2 Brain Nucleus accumbens (basal ganglia)
TT (37) TC (55) CC (19) pvalue: 8.51e-1 Brain Putamen (basal ganglia)
TT (28) TC (43) CC (12) pvalue: 2.59e-1 Brain Spinal cord (cervical c 1)
TT (26) TC (41) CC (13) pvalue: 7.16e-1 Brain Substantia nigra
TT (121) TC (196) CC (68) pvalue: 9.83e-1 Adipose Subcutaneous
TT (102) TC (153) CC (58) pvalue: 1.26e-1 Adipose Visceral (Omentum)
TT (68) TC (73) CC (34) pvalue: 6.77e-1 Adrenal Gland
TT (98) TC (122) CC (47) pvalue: 5.97e-4 Artery Aorta
TT (56) TC (70) CC (26) pvalue: 9.66e-1 Artery Coronary
TT (126) TC (201) CC (61) pvalue: 1.26e-2 Artery Tibial
TT (37) TC (59) CC (21) pvalue: 2.97e-1 Cells EBV transformed lymphocytes
TT (81) TC (130) CC (40) pvalue: 9.38e-3 Breast Mammary Tissue
TT (68) TC (103) CC (32) pvalue: 7.68e-2 Colon Sigmoid
TT (87) TC (119) CC (40) pvalue: 6.24e-1 Colon Transverse
TT (74) TC (96) CC (43) pvalue: 5.83e-1 Esophagus Gastroesophageal Junction
TT (117) TC (175) CC (66) pvalue: 1.31e-6 Esophagus Mucosa
TT (112) TC (171) CC (52) pvalue: 7.33e-3 Esophagus Muscularis
TT (101) TC (126) CC (37) pvalue: 4.15e-1 Heart Atrial Appendage
TT (98) TC (127) CC (47) pvalue: 1.13e-1 Heart Left Ventricle
TT (56) TC (71) CC (26) pvalue: 3.26e-1 Liver
TT (126) TC (200) CC (57) pvalue: 5.9e-1 Lung
TT (26) TC (46) CC (13) pvalue: 1.2e-1 Minor Salivary Gland
TT (166) TC (239) CC (86) pvalue: 6.85e-2 Muscle Skeletal
TT (115) TC (182) CC (64) pvalue: 3.32e-1 Nerve Tibial
TT (49) TC (52) CC (21) pvalue: 3.31e-2 Ovary
TT (75) TC (102) CC (43) pvalue: 8.17e-2 Pancreas
TT (59) TC (75) CC (23) pvalue: 4.73e-2 Pituitary
TT (44) TC (68) CC (20) pvalue: 6.02e-2 Prostate
TT (108) TC (143) CC (49) pvalue: 5.21e-1 Cells Transformed fibroblasts
TT (113) TC (163) CC (59) pvalue: 5.48e-2 Skin Not Sun Exposed (Suprapubic)
TT (143) TC (201) CC (70) pvalue: 4.21e-1 Skin Sun Exposed (Lower leg)
TT (43) TC (54) CC (25) pvalue: 2.51e-1 Small Intestine Terminal Ileum
TT (52) TC (71) CC (23) pvalue: 7.68e-1 Spleen
TT (82) TC (113) CC (42) pvalue: 4.43e-1 Stomach
TT (83) TC (105) CC (37) pvalue: 9.87e-1 Testis
TT (129) TC (200) CC (70) pvalue: 5.87e-1 Thyroid
TT (38) TC (44) CC (19) pvalue: 6.01e-1 Uterus
TT (37) TC (50) CC (19) pvalue: 2.51e-1 Vagina
TT (130) TC (178) CC (61) pvalue: 5.06e-1 Whole Blood
F
SNCA (ENSG00000145335.11) and rs2737024 (4_90721560_A_G_b37)
Normalized Expression
2.0
1.0
0.0
-1.0
-2.0
AA (46) AG (36) GG (6) pvalue: 4.87e-1 Brain Amygdala
AA (52) AG (46) GG (11) pvalue: 7.2e-1 Brain Anterior cingulate cortex (BA24)
AA (74) AG (60) GG (10) pvalue: 8.12e-1 Brain Caudate (basal ganglia)
AA (68) AG (49) GG (8) pvalue: 4.15e-1 Brain Cerebellar Hemisphere
AA (81) AG (61) GG (12) pvalue: 4.3e-1 Brain Cerebellum
AA (75) AG (54) GG (7) pvalue: 2.31e-1 Brain Cortex
AA (65) AG (43) GG (10) pvalue: 3.62e-1 Brain Frontal Cortex (BA9)
AA (60) AG (45) GG (6) pvalue: 3.18e-1 Brain Hippocampus
AA (55) AG (44) GG (9) pvalue: 4.49e-1 Brain Hypothalamus
AA (67) AG (56) GG (7) pvalue: 8.83e-3 Brain Nucleus accumbens (basal ganglia)
AA (60) AG (44) GG (7) pvalue: 7.38e-1 Brain Putamen (basal ganglia)
AA (45) AG (30) GG (8) pvalue: 4.26e-1 Brain Spinal cord (cervical c 1)
AA (43) AG (32) GG (5) pvalue: 6.81e-2 Brain Substantia nigra
AA (225) AG (133) GG (27) pvalue: 2.23e-1 Adipose Subcutaneous
AA (176) AG (112) GG (25) pvalue: 6.28e-2 Adipose Visceral (Omentum)
AA (111) AG (52) GG (12) pvalue: 8.2e-1 Adrenal Gland
AA (160) AG (86) GG (21) pvalue: 2.9e-5 Artery Aorta
AA (92) AG (49) GG (11) pvalue: 8e-1 Artery Coronary
AA (229) AG (131) GG (28) pvalue: 3.38e-4 Artery Tibial
AA (74) AG (35) GG (8) pvalue: 2.85e-2 Cells EBV transformed lymphocytes
AA (159) AG (78) GG (14) pvalue: 1.04e-8 Breast Mammary Tissue
AA (111) AG (77) GG (15) pvalue: 8.85e-2 Colon Sigmoid
AA (148) AG (81) GG (17) pvalue: 5.24e-2 Colon Transverse
AA (121) AG (73) GG (19) pvalue: 4.46e-1 Esophagus Gastroesophageal Junction
AA (205) AG (128) GG (25) pvalue: 6.51e-9 Esophagus Mucosa
AA (202) AG (115) GG (18) pvalue: 2.14e-2 Esophagus Muscularis
AA (153) AG (97) GG (14) pvalue: 8.7e-2 Heart Atrial Appendage
AA (165) AG (85) GG (22) pvalue: 1.35e-3 Heart Left Ventricle
AA (91) AG (48) GG (14) pvalue: 2.09e-1 Liver
AA (229) AG (131) GG (23) pvalue: 2.25e-2 Lung
AA (45) AG (35) GG (5) pvalue: 5.12e-2 Minor Salivary Gland
AA (290) AG (167) GG (34) pvalue: 9.09e-1 Muscle Skeletal
AA (212) AG (126) GG (23) pvalue: 3.86e-2 Nerve Tibial
AA (77) AG (36) GG (9) pvalue: 1.96e-3 Ovary
AA (130) AG (74) GG (16) pvalue: 3.89e-1 Pancreas
AA (90) AG (58) GG (9) pvalue: 5.72e-3 Pituitary
AA (78) AG (47) GG (7) pvalue: 1.59e-1 Prostate
AA (187) AG (99) GG (14) pvalue: 8.65e-1 Cells Transformed fibroblasts
AA (195) AG (114) GG (26) pvalue: 4.51e-2 Skin Not Sun Exposed (Suprapubic)
AA (242) AG (140) GG (32) pvalue: 7.59e-2 Skin Sun Exposed (Lower leg)
AA (77) AG (38) GG (7) pvalue: 6.54e-1 Small Intestine Terminal Ileum
AA (92) AG (46) GG (8) pvalue: 6.68e-1 Spleen
AA (143) AG (80) GG (14) pvalue: 4.3e-1 Stomach
AA (135) AG (77) GG (13) pvalue: 3.87e-1 Testis
AA (230) AG (141) GG (28) pvalue: 5.57e-1 Thyroid
AA (64) AG (30) GG (7) pvalue: 8.78e-1 Uterus
AA (57) AG (42) GG (7) pvalue: 8.26e-1 Vagina
AA (225) AG (119) GG (25) pvalue: 2.34e-1 Whole Blood

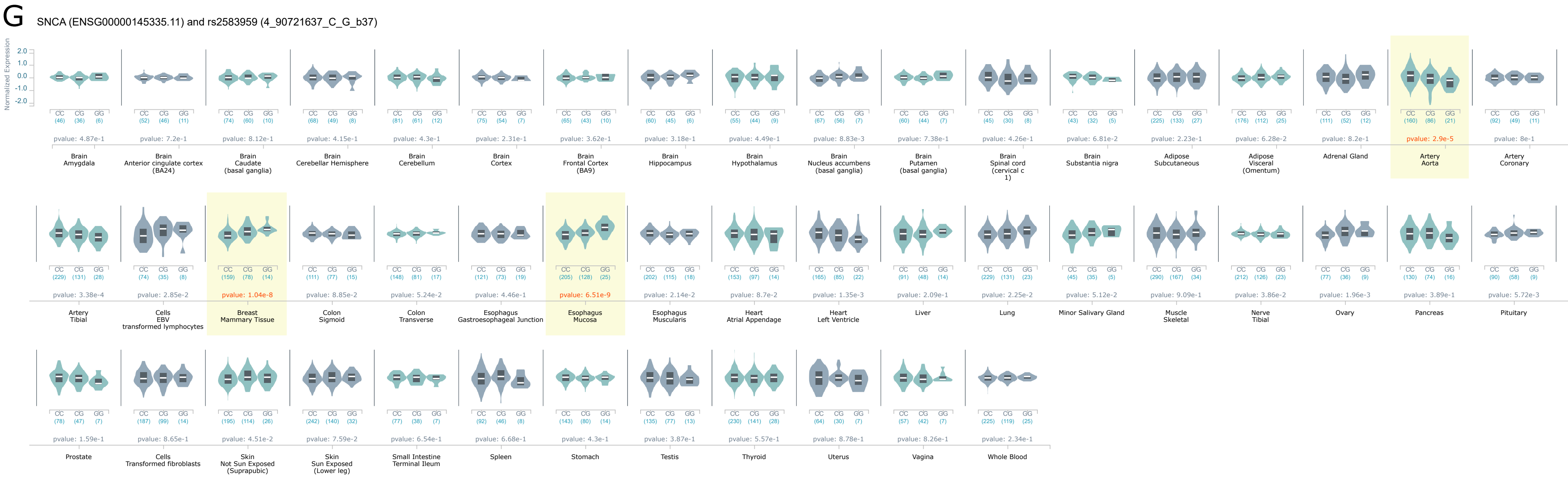


Figure S7 Violin plots of *SNCA* expression by genotype at the SNPs whose minor allele defines the PD-associated haplotype Only SNPs rs2737029 (**E**), rs2737024 (**F**), and rs2583959 (**G**) are eQTLs of *SNCA* in any tissues.

Characterizing mouse DA neurons by qPCR

	Forward	Reverse	Expected amplicon	mm9 co-ordinates
Pitx3	ACGCACTAGACCTCCCTCCAT	GCTTCTTCTTCAGAGAGCCGT	203	Pitx3 exons 1, 2, 3
Th	CTGTCCACGTCCCCAAGGTTCA	CAATGGGTTCCCAGGTTCCG	147	Th exons 1, 2
Slc6a3	GAGGCCCGATAAGAGCTCAAG	CCTTCTTCTTCGACTGCCTCC	111	Slc6a3 exons 1, 2
Actb	TGGCTCCTAGCACCATGAAG	AGCTCAGTAACAGTCCGCCTA	188	Actb exons 5. 6

Testing five putative CREs in *in vivo* reporter assays

	Forward	Reverse	Expected amplicon	hg19 co-ordinates
KCNQ3	ATAAAGCAAGTGACCGGGGA	GGCTGCTCTTGAGACATTCG	2744	chr8:133425146-133427889
FOXG1	CGGCAAAGGAACATGGAGAG	TCACATCCAGGGCCAAGAAT	2188	chr14:29242870-29245057
NR4A2	ATCAGCCTGTGTCCTGTTCT	AAGGAAGGGGCAGCTTAGAG	2447	chr2:157255824-157258270
CRHR1	CAGGACTATGACGGCTGACT	GGAACACACCCTCTCCATCA	1691	chr17:43889821-43891511
FOXA2	GTCTGATGTTCGTTCACCCAG	GCCGTTTTAAGCATTGGGAA	3288	chr20:22382513-22385800

Testing the *SNCA* enhancer in *in vivo* reporter assays

	Forward	Reverse	Expected amplicon	hg19 co-ordinates
SNCA	GGACTCCTTGCTTGAAGGAAAAAT	AGACAAAAGGAGTGCATTGATGT	1060	chr4:90,721,063-90,722,122

Testing protein binding at the two implicated SNPs

rs2737024-maj	acatcacattgtcctAttacattcttgcccaACCCTATAGTGAGTGCTATTA
rs2737024-min	acatcacattgtcctGttacattcttgcccaACCCTATAGTGAGTGCTATTA
rs2583959-maj	ctttgttaataaatcCttgtataaaccccacACCCTATAGTGAGTGCTATT
rs2583959-min	ctttgttaataaatcGttgtataaaccccacACCCTATAGTGAGTGCTATT

Table S1: Primer sequences used for qPCR, cloning, and protein binding assays

	Counts			Percentage	
	VISTA	**MB**	**FB**	**MB**	**FB**
Positive	**1264**	**786**	**719**	**62**	**57**
Neuronal	*652*	*498*	*465*	*76*	*71*
Forebrain	191	137	147	72	77
Midbrain	104	80	65	77	63
Hindbrain	94	66	58	70	62
Multiple regions	156	126	112	81	72
Whole brain	107	89	83	83	78
Non-neuronal	*612*	*288*	*254*	*47*	*42*
Negative	**1123**	**538**	**529**	**48**	**47**
TOTAL	2387	1324	1248	55	52

Table S2: Summary of counts and percent overlap with the VISTA enhancer browser, related to figure 2A-C

			Allele counts (frequency)		Genotype counts (frequency)		
Variant	MA	Population	Minor allele	Major allele	Homozygous Minor	Heterozygous	Homozygous Major
rs537518252	A	Control	0 (0%)	1910 (100%)	0 (0%)	0 (0%)	955 (100%)
		PD	1 (0.1%)	1909 (99.9%)	0 (0%)	1 (0.1%)	954 (99.9%)
rs78789649	A	Control	0 (0%)	1910 (100%)	0 (0%)	0 (0%)	955 (100%)
		PD	1 (0.1%)	1909 (99.9%)	0 (0%)	1 (0.1%)	954 (99.9%)
rs112174335	C	Control	2 (0.1%)	1908 (99.9%)	0 (0%)	2 (0.2%)	953 (99.8%)
		PD	0 (0%)	1910 (100%)	0 (0%)	0 (0%)	955 (100%)
rs28720123	T	Control	4 (0.2%)	1906 (99.8%)	0 (0%)	4 (0.4%)	951 (99.6%)
		PD	1 (0.1%)	1909 (99.9%)	0 (0%)	1 (0.1%)	954 (99.9%)
rs2737024	G	Control	515 (27%)	1395 (73%)	76 (8%)	363 (38%)	516 (54%)
		PD	609 (31.9%)	1301 (68.1%)	105 (11%)	399 (41.8%)	451 (47.2%)
chr4:90721581 T>C	C	Control	1 (0.1%)	1909 (99.9%)	0 (0%)	1 (0.1%)	954 (99.9%)
		PD	0 (0%)	1910 (100%)	0 (0%)	0 (0%)	955 (100%)
rs2583959	G	Control	518 (27.1%)	1390 (72.9%)	89 (9.3%)	340 (35.6%)	525 (55%)
		PD	606 (31.7%)	1304 (68.3%)	105 (11%)	396 (41.5%)	454 (47.5%)
chr4:90721702 G>A	T	Control	0 (0%)	1910 (100%)	0 (0%)	0 (0%)	955 (100%)
		PD	1 (0.1%)	1909 (99.9%)	0 (0%)	1 (0.1%)	954 (99.9%)
chr4:90721760 T>-	-	Control	0 (0%)	1910 (100%)	0 (0%)	0 (0%)	955 (100%)
		PD	1 (0.1%)	1909 (99.9%)	0 (0%)	1 (0.1%)	954 (99.9%)
rs189903574	A	Control	0 (0%)	1910 (100%)	0 (0%)	0 (0%)	955 (100%)
		PD	1 (0.1%)	1909 (99.9%)	0 (0%)	1 (0.1%)	954 (99.9%)
rs17016188	C	Control	116 (6.1%)	1794 (93.9%)	4 (0.4%)	108 (11.3%)	843 (88.3%)
		PD	156 (8.2%)	1754 (91.8%)	5 (0.5%)	146 (15.3%)	804 (84.2%)
rs28536191	G	Control	2 (0.1%)	1908 (99.9%)	0 (0%)	2 (0.2%)	953 (99.8%)
		PD	0 (0%)	1910 (100%)	0 (0%)	0 (0%)	955 (100%)
chr4:90721974 T>A	A	Control	0 (0%)	1910 (100%)	0 (0%)	0 (0%)	955 (100%)
		PD	1 (0.1%)	1909 (99.9%)	0 (0%)	1 (0.1%)	954 (99.9%)
rs7684892	A	Control	131 (6.9%)	1777 (93.1%)	9 (0.9%)	113 (11.8%)	832 (87.2%)
		PD	121 (6.3%)	1789 (93.7%)	3 (0.3%)	115 (12%)	837 (87.6%)

Table S4: Allele and genotype counts and frequencies in PD cases and controls of all variants identified by sequencing within the intronic *SNCA* enhancer

	rs2737029	rs356168	rs356220	rs356225	rs3857057	rs62306323	rs7689942	rs7684892	rs28536191	rs17016188	rs2583959	chr4:90721581 T>C	rs2737024	rs28720123	rs112174335
rs10018362	0.185	0.14	0.02	0.137	0.417	0.018	0.547	0.534	0.008	0.242	0.039	0.004	0.046	0.017	<0.001
rs2737029	----	0.663	0.523	0.672	0.101	0.07	0.103	0.096	0.002	0.096	0.498	0.001	0.542	0.003	0.001
rs356168	----	----	0.675	0.985	0.085	0.063	0.078	0.078	0.001	0.072	0.307	0.001	0.341	0.002	0.001
rs356220	----	----	----	0.686	0.12	0.03	0.002	0.002	0.002	0.101	0.373	<0.001	0.411	0.003	0.002
rs356225	----	----	----	----	0.082	0.062	0.078	0.078	0.001	0.073	0.31	0.001	0.344	0.002	0.001
rs3857057	----	----	----	----	----	0.007	0.023	0.023	0.014	0.583	0.025	<0.001	0.028	0.027	<0.001
rs62306323	----	----	----	----	----	----	0.01	0.006	<0.001	0.008	0.047	<0.001	0.05	<0.001	0.001
rs7689942	----	----	----	----	----	----	----	0.974	<0.001	0.005	0.022	0.007	0.026	<0.001	<0.001
rs7684892	----	----	----	----	----	----	----	----	<0.001	0.005	0.021	0.007	0.023	<0.001	<0.001
rs28536191	----	----	----	----	----	----	----	----	----	0.016	<0.001	<0.001	<0.001	0.5	<0.001
rs17016188	----	----	----	----	----	----	----	----	----	----	0.019	<0.001	0.024	0.033	<0.001
rs2583959	----	----	----	----	----	----	----	----	----	----	----	<0.001	0.934	0.001	<0.001
chr4:90721581 T>C	----	----	----	----	----	----	----	----	----	----	----	----	<0.001	<0.001	<0.001
rs2737024	----	----	----	----	----	----	----	----	----	----	----	----	----	0.001	<0.001
rs28720123	----	----	----	----	----	----	----	----	----	----	----	----	----	----	<0.001
rs112174335	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Table S5: r^2 values measuring linkage disequilibrium between *SNCA* variants in controls

			Allele counts (frequency)		Genotype counts (frequency)		
Variant	MA	Population	Minor allele	Major allele	Homozygous Minor	Heterozygous	Homozygous Major
rs10018362	C	Control	213 (11.1%)	1709 (88.9%)	9 (0.9%)	195 (20.3%)	757 (78.8%)
		PD	228 (12.1%)	1656 (87.9%)	7 (0.7%)	214 (22.7%)	721 (76.5%)
rs2737029	C	Control	864 (41.1%)	1240 (58.9%)	161 (15.3%)	542 (51.5%)	349 (33.2%)
		PD	869 (46.2%)	1011 (53.8%)	208 (22.1%)	453 (48.2%)	279 (29.7%)
rs356168	G	Control	910 (47.2%)	1016 (52.8%)	227 (23.6%)	456 (47.4%)	280 (29.1%)
		PD	965 (51.2%)	919 (48.8%)	246 (26.1%)	473 (50.2%)	223 (23.7%)
rs356220	T	Control	734 (38.1%)	1190 (61.9%)	159 (16.5%)	416 (43.2%)	387 (40.2%)
		PD	827 (43.9%)	1055 (56.1%)	190 (20.2%)	447 (47.5%)	304 (32.3%)
rs356225	C	Control	904 (47%)	1018 (53%)	223 (23.2%)	458 (47.7%)	280 (29.1%)
		PD	966 (51.3%)	918 (48.7%)	246 (26.1%)	474 (50.3%)	222 (23.6%)
rs3857057	G	Control	137 (7.1%)	1791 (92.9%)	7 (0.7%)	123 (12.8%)	834 (86.5%)
		PD	179 (9.5%)	1705 (90.5%)	4 (0.4%)	171 (18.2%)	767 (81.4%)
rs62306323	T	Control	241 (12.6%)	1667 (87.4%)	19 (2%)	203 (21.3%)	732 (76.7%)
		PD	205 (10.9%)	1679 (89.1%)	10 (1.1%)	185 (19.6%)	747 (79.3%)
rs7689942	T	Control	125 (6.5%)	1801 (93.5%)	5 (0.5%)	115 (11.9%)	843 (87.5%)
		PD	117 (6.2%)	1767 (93.8%)	2 (0.2%)	113 (12%)	827 (87.8%)

Table S6: Allele and genotype counts and frequencies in PD cases and controls of all variants genotyped from the Guella *et al.* panel