DNA Methylation and Gene Expression Profiling

for Parkinson's Biomarker Discovery

by

Adrienne Rose Henderson

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2019 by the
Graduate Supervisory Committee:

Matthew Huentelman, Co-Chair
Jason Newbern, Co-Chair
Travis Dunckley
Kendall Jensen
Melissa Wilson

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder, diagnosed late in the disease by a series of motor deficits that manifest over years or decades. It is characterized by degeneration of mid-brain dopaminergic neurons with a high prevalence of dementia associated with the spread of pathology to cortical regions. Patients exhibiting symptoms have already undergone significant neuronal loss without chance for recovery. Analysis of disease specific changes in gene expression directly from human patients can uncover invaluable clues about a still unknown etiology, the potential of which grows exponentially as additional gene regulatory measures are questioned. Epigenetic mechanisms are emerging as important components of neurodegeneration, including PD; the extent to which methylation changes correlate with disease progression has not yet been reported. This collection of work aims to define multiple layers of PD that will work toward developing biomarkers that not only could improve diagnostic accuracy, but also push the boundaries of the disease detection timeline. I examined changes in gene expression, alternative splicing of those gene products, and the regulatory mechanism of DNA methylation in the Parkinson's disease system, as well as the pathologically related Alzheimer's disease (AD). I first used RNA sequencing (RNAseq) to evaluate differential gene expression and alternative splicing in the posterior cingulate cortex of patients with PD and PD with dementia (PDD). Next, I performed a longitudinal genome-wide methylation study surveying ~850K CpG methylation sites in whole blood from 189 PD patients and 191 control individuals obtained at both a baseline and at a follow-up visit after 2 years. I also considered how symptom management medications could affect the regulatory mechanism of DNA methylation. In the last chapter of this work, I intersected RNAseq and DNA methylation array datasets from whole blood patient samples for integrated differential analyses of both PD and AD. Changes in gene expression and DNA methylation reveal clear patterns of pathway dysregulation that can be seen across brain and blood, from one study to the next. I present a thorough survey of molecular changes occurring within the idiopathic Parkinson's disease patient and propose candidate targets for potential molecular biomarkers.

DEDICATION

To my mom, Nancy, my dear departed dad, Dan, and my brothers, Lenny and Danny.

To my friends and extended family.

Thank you for supporting me unconditionally and putting up with me all along the way.

ACKNOWLEDGMENTS

Funding support: The first chapter in this work was made possible by the Banner Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona for the provision of human biological materials (or specific description, e.g. brain tissue, cerebrospinal fluid). The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke, the National Institute on Aging, the Arizona Department of Health Services, the Arizona Biomedical Research Commission, and the Michael J. Fox Foundation for Parkinson's Research.

The second chapter was supported by the Michael J Fox Foundation (MJFF) and NIH grants. The Harvard Biomarkers Study is supported by the Harvard NeuroDiscovery Center (HNDC), MJFF, the Parkinson's Disease Biomarkers Program (PDBP) under grants of the NINDS, and the Massachusetts Alzheimer's Disease Research Center (ADRC) under grant of the National Institute on Aging. I want to thank all study participants, their families, and friends for their support and participation, and the study coordinators. I also want to thank the Harvard Biomarkers Study (HBS), the many investigators involved, and Co-Directors of Harvard NeuroDiscovery Center: Clemens R. Scherzer, Bradley T. Hyman, Adrian J. Ivinson.

The final chapter was also supported by the Michael J Fox Foundation and the Mayo Clinic of Arizona.

The author is eternally grateful for the consistent support offered by several investigators in the Neurogenomics division of TGen and Arizona State University, namely Matthew Huentelman, Travis Dunckley, and Kendall Jensen. I have worked for each of them, and am lucky to have done so. Dr. Jensen opened her door and generously allowed me to intern for her when I was fresh out of college. My next opportunity was given by Dr. Dunckley who has always encouraged, supported, and taught me. Finally, Dr. Huentelman brought me to his team enabling me to continue on with my doctoral studies, also with unwavering support. They have each extended invaluable mentorship and personal generosity for quite nearly a decade now and I will always be thankful to them for that, as well as for making me laugh always, and for making this possible for me.

In the lab, I have everyone to thank. If not for direct scientific support in being taught, or provided with services and expertise, it is for the experiences and new ways of thinking everyone has offered. Matt, Ignazio, Ryan, and Marcus have answered so many of my questions, showed me how to analyze something, or did something for me. There are a thousand little things that have added up and I appreciate them all. Ashley, Wayne and Candace are very special friends and mentors, in science and life. Chris and Josh, thank you for making me laugh and for supporting me in the lab, and otherwise. Bessie taught me almost everything in the beginning, Ben did too, and also the Liang lab along the way. Jason, we miss you— thank you for teaching me.

I am incredibly grateful to my committee members for fostering my growth and leading me in the right direction all along: Matthew Huentelman, Jason Newbern, Travis Dunckley, Kendall Jensen, and Melissa Wilson.

TABLE OF CONTENTS

Page

LIST OF TABLES vii

LIST OF FIGURES ix

LIST OF SYMBOLS / NOMENCLATURE x

PREFACE xii

CHAPTER

1 NEXT-GENERATION PROFILING TO IDENTIFY THE MOLECULAR ETIOLOGY OF PARKINSON'S DEMENTIA 1

Introduction 1

Methods 3

Results 5

Discussion 17

2 DNA METHYLATION CHANGES ASSOCIATED WITH PARKINSON'S DISEASE PROGRESSION: OUTCOMES FROM THE FIRST LONGITUDINAL GENOME-WIDE METHYLATION ANALYSIS IN BLOOD 23

Introduction 23

Materials and Methods 25

Results 28

Discussion 46

3 DNA METHYLATION AND EXPRESSION PROFILES OF WHOLE BLOOD IN PARKINSON'S AND ALZHEIMER'S DISEASES 51

Introduction 51

Methods 53

Results 58

Overlapping Results Across Studies 90

Discussion 92

Conclusions 94

CHAPTER Page

Future Directions 95

REFERENCES 98

LIST OF TABLES

Table		Page
1.	Differential Gene Expression in PD and PDD	8
2.	Gene Ontology Pathways in PD and PDD	9
3.	SpliceSeq Splicing Statistics for qRT-PCR Validated Events	12
4.	Splicing Events and qRT-PCR Primer Sequences	14
5.	Clinical and Demographic Characterization of the HBS Study Cohort	26
6.	Differentially Methylated Probes, Cross-Sectional Analysis	34
7.	Differentially Methylated Regions Associated with PD	37
8.	Longitudinal Changes in Methylation in PD Cases	39
9.	Longitudinal Changes in Methylation in PD Cases Receiving Medication	42
10.	Longitudinal Methylation Changes in PD Cases Not Receiving Medication	45
11.	Differentially Methylated Positions in AD	60
12.	Differentially Methylated Positions in PD	63
13.	Differentially Methylated Positions in AD and PD	64
14.	Differentially Methylated Regions in AD	66
15.	Differentially Methylated Regions in PD	69
16.	Differentially Methylated Regions in AD and PD	71
17.	Differentially Expressed Genes in AD	72
18.	Differentially Expressed Genes in PD	73
19.	Differentially Expressed Genes in AD and PD	74
20.	Integrated Analysis: Combined P-values in AD	76
21.	Integrated Analysis: Combined P-values in PD	78
22.	Integrated Analysis: Combined P-values in AD and PD	80
23.	Integrated Analysis: meQTL in AD	81
24.	Integrated Analysis: meQTL in PD	84
25.	Integrated Analysis: meQTL in AD and PD	86
26.	Functional Epigenetic Modules for AD and PD	88

Table Page

27. Across-Study Differential Expression hits 90

28. Across-Study Differential Methylation hits 91

LIST OF FIGURES

Figure		Page
1.	SpliceSeq “Splice Graphs” qRT-PCR Tested Genes	12
2.	dPSI Values for SpliceSeq-Predicted Alternatively Spliced Exons	15
3.	qRT-PCR Fold Change Comparison of Splice vs. No-Splice Primer Sites	16
4.	Comparison of Individual Cell Type	31
5.	Case vs. Control Differentially Methylated Position Plots	33
6.	Case vs. Control Differentially Methylated Region Plots	36
7.	Cross-talk Between One-Carbon Metabolism and Levodopa Catabolism	41
8.	Longitudinal Methylation Changes in PD and Medication Effect	43
9.	Sample Group PCA	55
10.	Cell-Type Composition PCA Plots	57
11.	AD Manhattan Plot of Global Differential Methylation Significance	59
12.	PD Manhattan Plot of Global Differential Methylation Significance	62
13.	Differentially Methylated Regions in AD	67,68
14.	Differentially Methylated Regions in PD	70
15.	Top DE and DMP by Combined P-value in AD	77
16.	Top DE and DMP by Combined P-value in PD	79
17.	Representative Plot of a *cis* meQTL in AD	83
18.	Representative Plot of a *cis* meQTL in PD	85
19.	Functional Epigenetic Modules for AD and PD	89

LIST OF SYMBOLS / NOMENCLATURE

Symbol	Name
AB	Amyloid beta
ACTH	Adrenocorticotropin
AD	Alzheimer's disease
ALS	Amyotrophic lateral sclerosis
APP	Amyloid precursor protein
ATP	Adenosine triphosphate
BACE	Beta-Secretase
cAMP	Cyclic adenosine monophosphate
CCL	Chemokine (C-C motif) ligand
CF	Cerebrospinal fluid
CNS	Central nervous system
COMT	Catechol-O-methyltransferase
CRH	Corticotropin-releasing hormone
CSF	Cerebrospinal fluid
CT	Control
CXCL	Chemokine (C-X-C motif) ligand
CYP	Cytochrome P450
DA	Dopamine
DE	Differential expression
DG	Dentate gyrus
DLB	Dementia with Lewy Bodies
DM	Differential methylation
DMP	Differentially methylated position
DMR	Differentially methylated region
DN	Dopaminergic neurons
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
EGFR	Epidermal growth factor receptor
EWAS	Epigenome-wide association study
FDR	False discovery rate
FEM	Functional epigenetic module
GABA	Gamma-aminobutyric acid
GO	Gene ontology
GTP	Guanosine-5'-triphosphate
HBS	Harvard Biomarker Study
HDAC	Histone Deacetylase
HERC	HECT And RLD Domain Containing E3 Ubiquitin Protein
HSP	Heat shock protein
JAK/STAT	Janus kinase signal transducer and activator of transcription
JNK	c-Jun N-terminal kinase
LB	Lewy body

Symbol	Name
LRRK	Leucine Rich Repeat Kinase
MAPK	mitogen-activated protein kinase
MAPT	Microtubule Associated Protein, Tau
meQTL	Methylation quantitative trait loci
MHC	Major Histocompatibility Complex
NFkB	Nuclear Factor Kappa B
OHDA	hydroxydopamine
PCA	Principal component analysis
PD	Parkinson's disease
PDD	Parkinson's disease with dementia
PI3K/AKT	Phosphatidylinositol 3-kinase / Protein Kinase B
PSEN	Presenilin
PSI	Percent spliced-in
qRT-PCR	Quantitative real-time polymerase chain reaction
RNA	Ribonucleic acid
RPKM	Reads Per Kilobase of transcript, per Million mapped reads
SAM	S-Adenosyl methionine
SNCA	Alpha-synuclein
SNP	Single nucleotide polymorphism
TSS	Transcription start site
UPDRS	Unified Parkinson's Disease Rating Scale
UTR	Untranslated region

PREFACE

This dissertation is the culminating collection of work initiated to study the molecular profiles of Parkinson's disease (PD) patients with the overarching goal of impacting future biomarker development. The author completed this dissertation while in the Neurogenomics division of the Translational Genomics Research Institute (TGen), in coordination with Arizona State University (ASU). The mission of TGen is to produce research that can be translated to clinical applications, which drove the conceptualization of each of the studies contained here.

The chapters in this dissertation where developed closely with Dr. Travis Dunckley. Leadership and study guidance where provided throughout by Drs. Dunckley and Matthew Huentelman. The author became involved in the first chapter's work before beginning her PhD studies at ASU in 2014, completing and publishing the study shortly thereafter as first author. Chapter 2 was a collaborative effort divided between Drs. Dunckley and Paula Desplats of the University of California, San Diego (UCSD). The study was published in March 2019 with co-first authorship. Chapter 3 is a completed manuscript by the author and data analysis was performed almost entirely by the author. This will be submitted for publication as a first author work as well. All studies where efforts completed with the input and support from teams of people in the laboratories of Drs. Huentelman and Dunckley, and in the laboratory of Dr. Desplats for the work in chapter 2.

CHAPTER 1

NEXT-GENERATION PROFILING TO IDENTIFY THE MOLECULAR ETIOLOGY OF PARKINSON'S DEMENTIA

INTRODUCTION

Parkinson's disease (PD) is the second most common neurologic disorder, characterized by degeneration of mid-brain dopaminergic neurons and a high prevalence of dementia. Neuroinflammation, oxidative stress, mitochondrial dysfunction, protein processing, and aberrant alternative splicing are among the commonly considered pathways of dysfunction in Parkinson's disease(Cooper-Knock et al. 2012).

Gene expression profiling via microarrays has been a widely implemented and useful method for disease pathway elucidation, though its application is limited by the requirement for prior sequence information for probe design. The more recent development and use of next-generation, high-throughput RNA sequencing (RNA-seq) has facilitated the expansion of gene expression knowledge by filling missing gene and isoform data gaps, in addition to reducing background noise and optimizing specificity by avoiding probe cross-hybridization(Mills and Janitz 2012).

It is important to consider that microarrays are biased as they are limited by probe design and available gene annotations. For example, Loraine et al. identified at least 500 annotated pollen genes in their RNA-seq data that were not included on the ATH1 (*Arabidopsis thaliana*) microarray(Loraine et al. 2013). Even for extensively annotated genomes, microarrays lack complete sequence inclusion of 3'/5' UTRs, introns, and alternative splice variants(Loraine et al. 2013). More specific exon arrays have been successfully applied to alternative splicing studies(Shoemaker et al. 2001; Gardina et al. 2006; Moore and Silver 2008), but they remain insufficient for comprehensive novel splicing and isoform discovery. RNA-seq produces accurate, high resolution coverage of the transcriptome versus a probe-limited subset(Trapnell et al. 2013) and is currently the superior method for gene and isoform expression profiling for discovery, particularly when investigating the functions and effects of alternative splicing(Mills and Janitz 2012; Nookaew et al. 2012; Sanchez-Pla et al. 2012; Giorgi et al. 2013; Loraine et al. 2013).

Gene expression studies of PD and other neurological disorders, via microarray or RNA-seq, have produced thematic results. Transcriptome microarray analysis of PD and controls showed that the PD-linked *LRRK2* mutation is associated with expression changes in peripheral blood mononuclear cells (PBMCs)(Mutez et al. 2011). The same study highlighted alterations in the *MAPK* and GTPase pathways, which are involved in axonal guidance, actin cytoskeleton, and vesicular dynamics, among other functions. Evidence of downregulated synaptic genes (*NSF*, *SYNGR3*) and cytoskeletal maintenance genes (*MARK1*, *MAP2*, and *DNAI1*) in the substantia nigra suggest malfunction of cytoskeletal dynamics(Miller et al. 2006). In our previous microarray study, we found that axonal transport, cell adhesion, and mRNA splicing are the most prevalent dysregulated processes in PD. We also reported decreased expression in neurite outgrowth and cell adhesion proteins, all occurring prior to dementia onset(Stamper et al. 2008).

Since expression profiling provided evidence of significant spliceosome alterations(Stamper et al. 2008), in the present study we leveraged mRNA-seq to incorporate splice variant analysis. Analyses are restricted to coding RNAs, though non-coding and micro-RNA (miRNA) studies offer good examples of how transcriptional changes are relevant on multiple levels(Minones-Moyano et al. 2011; Wood et al. 2013). Much attention has been given to overall expression changes between disease and controls, but we report here that differential alternative splicing is associated with PD and PD-dementia, possibly contributing to the etiology of these diseases. Alternative splicing of transcripts encoding components of immune response pathways and RNA processing correlate with disease states and the genes displaying the highest degree of alternative splicing in this dataset are somewhat disconnected from the greatest expression changes, though overlapping cellular processes are evident. Compared to gene expression profiling alone, alternative splice analysis of mRNA-seq data affords precision profiling in addition to an overview of global expression changes.

METHODS

Tissue Collection

Posterior cingulate cortex samples were obtained from the Banner Sun Health Research Institute Brain Bank. All cases signed informed consent at Banner Sun Health Research Institute and were prospectively followed until death and autopsied according to previously published protocols(Beach et al. 2015). Samples were selected with a postmortem interval less than 3 hours. Posterior cingulate cortex was sectioned at 8 μ thickness and placed into standard 1.5 mL microcentrifuge tubes and stored at -80°C until RNA extraction.

RNA Extraction

TRIzol reagent (Life Tech) was added to each sample tube containing a fresh frozen brain tissue slice, then pipetted to homogenize. The solution was incubated at room temperature for 3 minutes followed by adding 100µl chloroform (Sigma). Tubes were hand-shaken for 15 seconds and incubated at room temperature for 2 minutes then centrifuged at 12,000g for 15 minutes at 4°C. The top aqueous layer was transferred into a new, cold microcentrifuge tube and 1 equal volume cold 70% EtOH was added. The RNA solution was then processed through the Qiagen RNeasy Mini kit, with DNase treatment, using the manufacturer's protocol.

mRNAseq Library Preparation

1 µg of total RNA was used to generate mRNA-seq libraries for sequencing using Illumina's RNA Sample Prep Kit (catalog#FC-122-1001) using the manufacturer's protocol. In summary, poly-A selection was performed using oligo dT magnetic beads to capture mRNA transcripts, and double stranded cDNA was generated and fragmented to a target size of 400bp using sonication on the Covaris. Fragmented samples were then end repaired and adenylated at the 3' end, and TruSeq Indexed adapters were ligated on. Libraries were enriched using the TruSeq PCR Master Mix and primer cocktail. Final libraries were cleaned and quantified using the Agilent Bioanalyzer and Invitrogen Qubit. Libraries were equimolarly pooled for sequencing.

Paired-End Sequencing

Denatured and diluted libraries with a 1% phiX spike-in were used to generate clusters on Illumina's HiSeq Paired End v3 flowcell on the Illumina cBot using Illumina's TruSeq PE Cluster

Kit v3 (cat#PE-401-3001). The clustered flowcell was sequenced by synthesis on the Illumina HiSeq 2000 for paired 100bp reads using Illumina's TruSeq SBS Kit V3 (cat#FC-401-3001).

Sequencing and Differential Expression Analysis

An average of 109 million paired-end, passing-filter reads (90mers) were generated for each sample. Reads were trimmed of adapter sequences and aligned to H. sapiens GRCh37.62 with *TopHat* (version 2.0.8, bowtie version 0.12.7) using default parameters. A table of read counts was assembled in R with the *easyRNASeq* package. Pairwise differential expression analyses between the groups was conducted with the *DESeq* package.

Gene Ontology

GeneMANIA is a functional association tool used for data visualization and statistical analysis. It builds gene ontologies from gene input lists and is freely available as a web application (also available as a Cytoscape plugin); figures and tables are available for output organization.

Alternative Splicing Analysis

SpliceSeq is an RNA-seq data visualization and statistical analysis tool. Comprehensive documentation is available from MD Anderson Bioinformatics, on the developer's website. We used this program to investigate significant differential splicing events in disease groups (PD or PD-D) relative to controls. SpliceSeq returns gene RPKM and exon OPKM (Observations Per Kilobases of exon/splice per Million aligned reads) normalized read values, defined respectively as "Reads" and "Observations" per kilobase of transcript length per million aligned reads. OPKM provides a measure of exon expression; a read that contains at least 4 bases of an exon is "observed".

Reads "spliced in", i.e. reads that map to an exon, are calculated against the total of reads that span the flanking exons, but skip it, plus the spliced in reads. This is the "percent spliced in" (PSI) value assigned to each exon. The difference between the average exon PSI of "group 2" and "group 1" (e.g., PD minus CON) is termed delta PSI (dPSI).

dPSI sample calculation:

SRRM1 deltaPSI = PSI PD-D group – PSI CON group

$$= \left(\frac{11}{(11+11)} \right) - \left(\frac{27}{(10+27)} \right)$$

To obtain a $\log_2$ ratio of exon expression, OPKM values are measured against RPKMs (corrected for overall gene expression). The $\log_2$ ratio (mentioned in Figure 1) dictates whether an exon is highlighted as over-represented (green), under-represented (red), or normally represented (no highlight).

qRT-PCR Validation

RNA extracted (described above) from PD, PD-D, and CON tissues was processed with the SuperScript III First-Strand Synthesis kit (Life Technologies) for cDNA synthesis, per the manufacturer's protocol. qRT-PCR was completed using the Roche LightCycler 480 manufacturer protocol with their corresponding detection reagent, LightCycler 480 SYBR Green I Master.

Primer pairs for these regions were designed so that at least one primer site was situated within the spliced exon. Each primer set designed to cover a splice event was paired in the same gene with a primer set located in adjacent or nearby exons without predicted splicing activity. Placing primer sets in these two distinct areas served as a means to compare baseline expression levels without alternative splicing, to the alternative splice sites at which expression was predicted to vary.

RESULTS

We performed mRNA-seq on RNA samples isolated from the posterior cingulate cortex (Brodmann Area 23) of individuals with either Parkinson's disease (PD; N=13; mean age 79.3 years +/- 6.8, 23% female), Parkinson's disease with Dementia (PD-D; N=10; mean age 75.5 years +/- 7.0, 30% female), or healthy controls (N=11; mean age 77.9 years +/- 7.9, 9% female)). Data analysis for this study focused on differential gene expression (DE) and alternative splicing (AS) events. DE and AS changes in PD and PD-D were quantified relative to unaffected controls. We utilized SpliceSeq(Ryan et al. 2012) for AS analysis, developed with RNA-seq data and alternative splicing detection in mind, and DESeq(Anders and Huber 2010) for DE analysis.

Genes from DE and AS analyses were then compared for overlapping results and the data were run through GeneMANIA(Ashburner et al. 2000; Montojo et al. 2010), a web-based gene ontology (GO) program, to view common pathways or functions.

Gene Ontology & Differential Expression.

We first compared gene-level expression profiles across disease groups (Table 1). DE lists were restricted to the top 20 genes differentially expressed by fold change (fc) ≥ |0.2| and $p \leq 0.001$ (multiple test-corrected with the Benjamini-Hochberg procedure), as determined by DESeq analysis. The top 20 DE genes from each group comparison were compared to gene networks for GO analysis(Montojo et al. 2010). Significant functions (Table 2) associated with the top 20 overexpressed PD genes included cell chemotaxis, cytokine receptor binding, cAMP-mediated signaling, and ion homeostasis, while overexpressed PD-D genes were involved in unfolded or incorrectly folded protein responses. Underexpressed PD and PD-D genes shared a common involvement in hormone signaling.

Comparison of PD to control revealed changes in immune function and neuropeptide hormone function, including upregulation of chemokine (C-X-C motif) ligand 10 (*CXCL10*; $p = 1.78 \times 10^{-16}$, fc = 15.845) and downregulation of proenkephalin (*PENK*; $p = 5.73 \times 10^{-8}$, fc = -3.355). In the comparison of PD-D to control, genes involved in the unfolded protein response were significantly upregulated, reflected by increased expression of dnaJ (Hsp40) homolog (*DNAJB1*; $p = 3.16 \times 10^{-8}$,fc = 3.876) and several heat-shock protein genes, and an overall downregulation of hormone activity and ion transport, indicated by decreased expression of corticotropin-releasing hormone (*CRH*; $p = 6.4 \times 10^{-7}$, fc = -4.724), *PENK* ($p = 1.57 \times 10^{-4}$, fc = -3.049), somatostatin (*SST*; $p = 4.44 \times 10^{-9}$, fc = -5.144), chemokine (C-C motif) ligand 3 (*CCL3*; $p = 4.5 \times 10^{-5}$, fc = -4.492) and chemokine (C-C motif) ligand 4 (*CCL4*; $p = 9.47 \times 10^{-7}$, fc = -8.737).

Predominant pathways shared by both PD and PD-D included upregulated inflammatory responses, with increased expression of granulocyte colony-stimulating factor (*CSF3*; $p_{PD} = 8.25 \times 10^{-19}$, $fc_{PD} = 9.920$; $p_{PD\text{-}D} = 2.00 \times 10^{-35}$, $fc_{PD\text{-}D} = 49.470$), selectin-E (*SELE*; $p_{PD} = 4.24 \times 10^{-12}$, $fc_{PD} = 4.808$; $p_{PD\text{-}D} = 7.10 \times 10^{-29}$, $fc_{PD\text{-}D} = 25.334$), and additional cytokines. Shared downregulated genes for signaling and cytoskeletal structure were shown by decreased expression of *PENK*

(p_{PD} = 5.73x10^{-8}, fc_{PD} = -3.355; p_{PD-D} = 1.57x10^{-4}, fc_{PD-D} = -3.049) and keratin 5 (*KRT5*; p_{PD} = 1.47x10^{-7}, fc_{PD} = -4.718; p_{PD-D} = 6.75x10^{-4}, fc_{PD-D} = -4.373); Vgf nerve growth factor (*VGF*; p_{PD} = 1.06x10^{-10}, fc_{PD} = -3.821; p_{PD-D} = 1.61x10^{-8}, fc_{PD-D} = -5.621) was also downregulated across disease states.

A. PD vs. CON Overexpressed			B. PD vs. CON Underexpressed		
Gene	**P-value**	**Fold Change**	**Gene**	**P-value**	**Fold Change**
LMX1A	1.87×10^{-9}	18.08	FABP1	7.03×10^{-9}	-9.14
CXCL10	1.78×10^{-16}	15.85	HBG2	5.06×10^{-8}	-6.97
CSF3	8.25×10^{-19}	9.92	KRT5	1.47×10^{-7}	-4.72
S100A14	8.48×10^{-8}	8.60	(uncharacterized)	3.61×10^{-11}	-4.50
GRHL3	4.01×10^{-15}	8.39	FREM3	2.82×10^{-8}	-4.04
C6	1.14×10^{-13}	6.44	VGF	1.06×10^{-10}	-3.82
CHI3L2	2.59×10^{-8}	5.86	CNGB1	2.28×10^{-8}	-3.71
SELE	4.24×10^{-12}	4.81	FAM163A	6.90×10^{-8}	-3.69
CXCL11	1.29×10^{-8}	4.50	GLP2R	2.93×10^{-8}	-3.41
CD163	1.43×10^{-7}	4.34	CUX2	2.68×10^{-8}	-3.38
CCL2	6.90×10^{-8}	4.33	PENK	5.73×10^{-8}	-3.35
NTS	6.40×10^{-9}	4.31	SLC32A1	5.29×10^{-7}	-3.29
PZP	1.64×10^{-10}	4.11	PCDH8	7.57×10^{-9}	-3.29
SERPINA3	5.73×10^{-8}	4.10	SLC17A6	3.71×10^{-7}	-3.27
FCGR2B	2.82×10^{-8}	4.08	MEPE	3.73×10^{-7}	-3.01
RASSF9	7.70×10^{-9}	3.68	CALB2	1.57×10^{-8}	-2.94
CD44	6.97×10^{-8}	3.59	CBLN4	8.53×10^{-7}	-2.86
EMP1	4.59×10^{-8}	3.52	LAMP5	9.39×10^{-7}	-2.84
LTF	3.75×10^{-8}	3.26	CTXN3	2.25×10^{-7}	-2.62
TNC	1.69×10^{-7}	3.16	CARTPT	1.70×10^{-7}	-2.16
C. PD-D vs. CON Overexpressed			**D. PD-D vs. CON Underexpressed**		
Gene	**P-value**	**Fold Change**	**Gene**	**P-value**	**Fold Change**
CSF3	2.00×10^{-35}	49.47	C5orf17	6.95×10^{-4}	-11.12
SELE	7.10×10^{-29}	25.33	CCL4	9.47×10^{-7}	-8.74
SFN	1.00×10^{-10}	14.27	ACTG2	1.36×10^{-5}	-6.52
IGHG4	1.67×10^{-7}	13.13	DES	8.37×10^{-7}	-6.32
VHLL	4.76×10^{-6}	10.28	MYO1A	3.66×10^{-9}	-5.66
HSPA6	2.42×10^{-18}	9.40	VGF	1.61×10^{-8}	-5.62
IL8	9.70×10^{-17}	9.00	SST	4.44×10^{-9}	-5.14
IGHG1	4.59×10^{-9}	8.31	CRH	6.40×10^{-7}	-4.72
IL1R2	8.37×10^{-7}	7.41	CCL3	4.50×10^{-5}	-4.49
IGHG2	5.66×10^{-6}	7.39	KRT5	6.75×10^{-4}	-4.37
CXCR1	3.16×10^{-8}	6.94	HRNR	2.07×10^{-5}	-4.12
IGKC	1.90×10^{-6}	6.33	(uncharacterized)	1.13×10^{-3}	-4.08
IL6	2.10×10^{-10}	5.94	FAM216B	6.95×10^{-4}	-3.60
CXCL1	1.57×10^{-10}	5.72	ST8SIA2	1.13×10^{-3}	-3.52
NPC1L1	1.71×10^{-10}	4.85	DNAH3	1.38×10^{-3}	-3.46

SERPINH1	1.83×10^{-8}	4.24	GPD1	3.47×10^{-4}	-3.14
DNAJB1	3.16×10^{-8}	3.88	PENK	1.57×10^{-4}	-3.05
HSPA1L	3.66×10^{-8}	3.83	PPEF1	1.77×10^{-3}	-3.01
HSPA1A	1.51×10^{-6}	3.14	HBA2	1.29×10^{-3}	-2.54
HSPA1B	2.30×10^{-6}	3.08	HBA1	1.60×10^{-3}	-2.51

Table 1. PD over- (A) and underexpressed (B) genes and PD-D over- (C) and underexpressed (D) genes make up the lists used for GO analysis ($p \leq 0.001$, $fc \geq |0.2|$).

A. PD Overexpressed			
GO Function	**FDR**	**Network Genes**	**Genome Genes**
cell chemotaxis	5.22×10^{-7}	9	157
cytokine receptor binding	5.60×10^{-6}	8	147
positive regulation of cAMP-mediated signaling	2.74×10^{-5}	4	11
chemokine receptor binding	2.74×10^{-5}	5	32
positive regulation of homeostatic process	1.04×10^{-4}	5	46
regulation of cAMP-mediated signaling	1.04×10^{-4}	4	17
positive regulation of release of sequestered calcium ion into cytosol	1.04×10^{-4}	4	17
metal ion homeostasis	2.72×10^{-4}	8	288
positive regulation of intracellular transport	3.87×10^{-4}	6	122
positive regulation of calcium ion transport into cytosol	4.40×10^{-4}	4	26
B. PD Underexpressed			
GO Function	**FDR**	**Network Genes**	**Genome Genes**
neuropeptide hormone activity	6.57×10^{-2}	3	15
hormone activity	9.19×10^{-1}	3	44
C. PD-D Overexpressed			
GO Function	**FDR**	**Network Genes**	**Genome Genes**
response to unfolded protein	3.66×10^{-18}	14	106
response to topologically incorrect protein	4.71×10^{-18}	14	113
protein folding	3.81×10^{-5}	7	124
cell chemotaxis	2.57×10^{-3}	6	157
COP9 signalosome	2.57×10^{-3}	4	34
cytokine receptor binding	2.57×10^{-3}	6	147
leukocyte chemotaxis	6.76×10^{-3}	5	106
blood microparticle	6.76×10^{-3}	5	108
unfolded protein binding	8.52×10^{-3}	4	52
defense response to Gram-negative bacterium	8.73×10^{-3}	3	16
D. PD-D Underexpressed			
GO Function	**FDR**	**Network Genes**	**Genome Genes**

neuropeptide hormone activity	3.04×10^{-2}	3	15
hormone activity	3.04×10^{-2}	4	44
calcium-mediated signaling	7.66×10^{-2}	4	72
bicarbonate transport	1.20×10^{-1}	3	29
regulation of system process	1.78×10^{-1}	5	203
blood microparticle	1.90×10^{-1}	4	108
second-messenger-mediated signaling	2.78×10^{-1}	4	124
gas transport	4.05×10^{-1}	2	10
hydrogen peroxide catabolic process	4.08×10^{-1}	2	14
cellular ion homeostasis	4.08×10^{-1}	5	292

Table 2. GeneMANIA measures significance by false discovery rate (FDR). Numbers of genes involved in a particular cellular function are shown, both within the resulting network and from the entire genome. The first 10 functions (of 87) are listed from overexpressed PD genes, and the first 10 (of 56) from overexpressed PD-D genes (A,C). All functions from underexpressed genes are shown (B,D).

Alternative Splicing.

SpliceSeq identified significant disease-associated differences in transcript variants. To make predictions of alternative splicing events, SpliceSeq builds a “splice graph” based on current Ensembl transcript isoform data to which it aligns experimental transcript data for each sample, followed by normalization and statistical calculations (see Methods). User-defined groups can then be compared for expression differences. In this case, “groups” are defined by patient category, either PD, PD-D, or control (CON). SpliceSeq-predicted events were initially filtered for significance ($p < 0.01$) and dPSI > |0.2|, where dPSI (or “differential percent spliced in”) is the change in exon expression of disease vs. controls. Additional filtering parameters considered the occurrence of the splicing event across all transcripts in the group (magnitude) and the number of samples within the group experiencing the event (percent observed)– both set to > |0.8| (1.0 = 100%). Finally, genes with the highest reads per kilobase of transcript length per million aligned reads (RPKM) within the filtered lists (15-100 RPKM) were preferentially picked for validation (Table 3).

Transcription factors, in concert with immune responses and indicators of cellular stress and RNA processing, were among the most highly alternatively spliced mRNAs. Five genes (total = 40) from the PD-D/CON comparison (*ATXN2*, *DST*, *HSPH1*, *RELA*, *SRRM1*) and two genes

(total = 10) from the PD/CON comparison (*LRRFIP1*, *TRIM9*) were chosen for qRT-PCR validation of the specific AS events (Figure 1). Heat shock protein 105 kDa (*HSPH1*), rel-like domain-containing (*RELA*), leucine-rich repeat flightless-interacting protein 1 (*LRRFIP1*), and tripartite motif 9 (*TRIM9*) may be grouped by processes of cellular maintenance or stress response (proliferation, apoptosis, etc.), and immune responses. *RELA*, *LRRFIP1*, and *TRIM9* are also transcription factors. Ataxin-2 (*ATXN2*) and serine/arginine repetitive matrix 1 (*SRRM1*) perform RNA processing functions. Dystonin (*DST*) is a cytoskeletal linking protein.

We performed qRT-PCR on exons predicted by SpliceSeq to experience an event, either exclusion from, or inclusion in the final transcript (see Table 4 for exon sites and primer sequences). SpliceSeq-predicted exon level expression changes (Figure 2) were validated by comparison to qRT-PCR fold changes ("fc"; $2^{-e\ pe}$). Significant fold change values less than one for primers covering exon skip regions, and greater than one for the exon inclusion event, were considered validated.

In PD-D, skips were predicted by SpliceSeq at *ATXN2* exon 21 (dPSI= -0.255, fc= 0.674), DST exon 104 (dPSI= -0.255, fc= 1.160), *HSPH1* exon 13 (dPSI= -0.231, fc= 0.952), and *SRRM1* exon 16, which disrupts a region reported as necessary for speckles and matrix localization (dPSI= -0.250, fc= 0.598). Exon 10.2 of *RELA* was predicted to be spliced in (dPSI= 0.286, fc= 3.467). This is a 687bp inclusion containing a premature stop site that would result in a 377 amino acid (AA) truncated isoform, from the wild-type 551 AA isoform, and consequently eliminate an activation domain.

In PD, skips were predicted in *LRRFIP1* exons 18-19 within a coiled-coil structural region (dPSI= -0.257, fc= 0.821) and *TRIM9* exons 8-10 within a fibronectin type III and B30.2/SPRY domain (dPSI= -0.264, fc= 0.930).

PD-D						
Gene (exon #)	**P-Value**	**dPSI**	**Magnitude**	**% Observed**	**$RPKM_{CON}$**	**$RPKM_{PD-D}$**
ATXN2 (21)	7.81×10^{-3}	-0.231	1.000	100%	15.9	14.0
DST (104)	3.21×10^{-3}	-0.255	1.000	100%	81.4	85.6
HSPH1 (13)	9.50×10^{-3}	-0.231	0.962	100%	48.8	99.6
RELA (10.2)	5.09×10^{-4}	0.286	0.978	100%	16.1	24.9
SRRM1 (16)	4.95×10^{-3}	-0.250	0.994	100%	34.2	25.6
PD						
Gene (exon #)	**P-Value**	**dPSI**	**Magnitude**	**% Observed**	**$RPKM_{CON}$**	**$RPKM_{PD}$**
LRRFIP1 (18-19)	8.76×10^{-3}	-0.257	1.000	100%	12.5	8.7
TRIM9 (8-10)	1.83×10^{-3}	-0.264	0.932	100%	22.9	22.6

Table 3. SpliceSeq-predicted Event Statistics for Genes to be Validated by qRT-PCR. Top alternatively spliced genes obtained by filtering for $p < 0.01$, dPSI > |0.2|, magnitude and % observed > |0.8| (1.0 = 100%). The resulting lists contained 40 PD-D genes and 10 PD genes. Finally, genes with the highest RPKM values were chosen for validation by qRT-PCR.

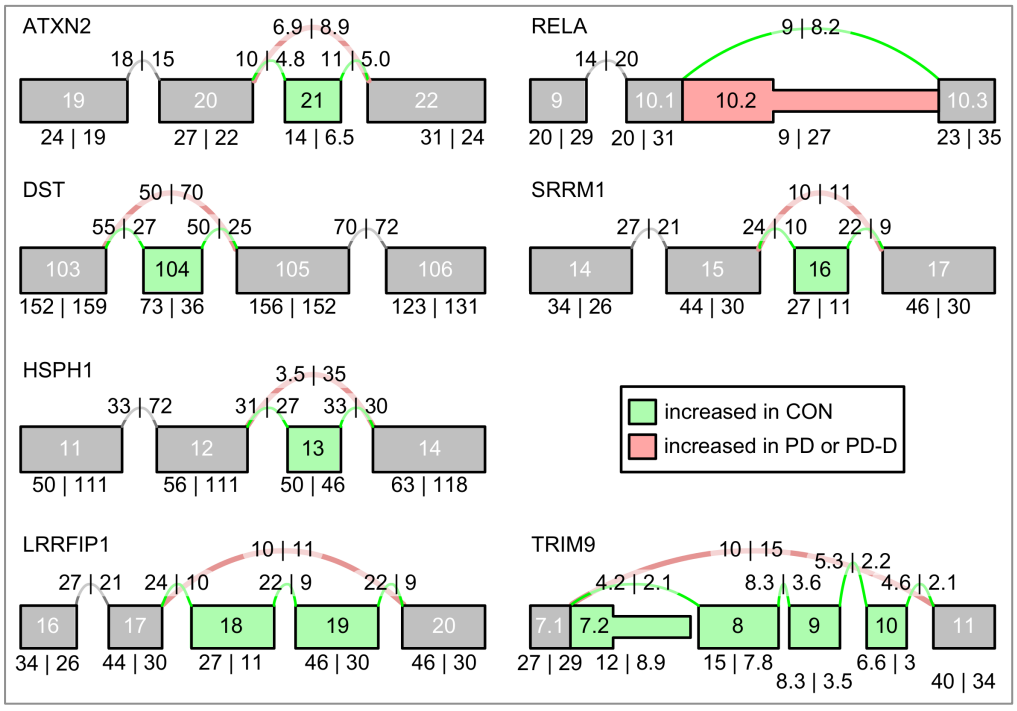


Fig 1. SpliceSeq "splice graphs" for the seven qRT-PCR tested genes. RNA-seq transcript reads are aligned to composite gene graphs built from all possible isoforms existing in the Ensembl genome database. The graphs in this figure are partial splice graphs, replicated from the SpliceSeq software, to show the alternative splicing regions discussed in this paper. Values on graphs are splice OPKMs, but can be changed in SpliceSeq to display read counts or log2 ratios (see Methods for explanation of terms). Exon highlighting indicates up- or down-regulated exon expression or splicing as follows: green exons and splice arcs represent increased activity in the control group and red exons and splice arcs represent either the PD or PD-D disease group. Gene splice graphs shown are specific for PD-D vs. CON (ATXN2, DST, HSPH1, RELA, SRRM1) or PD vs. CON (LRRFIP1, TRIM9).

Gene	Predicted Splice Site Exon	Locus	Group & Event Type	Primer Site(s)	Primer Sequence	"No Splice" Primer Site(s)	Primer Sequence
ATXN2	21	chr12: 111902519-111902466	PDD skip	21-22	F: CAACAAGGAGACAAGCCCTTC	19-20	F: TGTTGCCTACAGTCCTCAGC
					R: GGCTGAGGGTGTGGAGTATG		R: CTCATGAGCCCCGTACTGAG
DST	104	chr6: 56329554-56329483	PDD skip	104-105	F: GCCAAAGGAAGGACAAACATGG	105-106	F: CCAGGGTATGGCTGCTTTCC
					R: TGGATCTGTTGGGTGAAGCG		R: TGGAGAATCTTGGGTGTGGTG
HSPH1	13	chr13: 31718060-31717929	PDD skip	13	F: ACAGCCCCAGGTACAAACTG	11-12	F: ACCATGCTGCTCCTTTCTCC
					R: GGCAGCTCAACATTCACCAC		R: GGGTGTTGACTCGCACTTTG
RELA	10.2	chr11: 65423158-65422472	PDD inclusion	10.2	F: CAAGATGTGTGTGCCTGTGC	8.1-10.1	F: TCAGTGAGCCCATGGAATTCC
					R: AGCCCACGAAACTCTTCCAG		R: CTTGGGGACAGAAGCTGAGC
SRRM1	16	chr1: 24989674-24989715	PDD skip	16-17	F: GATGGGAAAGCGATGGCAATC	14-15	F: TCTGACTCTGGCTCCTCCTC
					R: TGGTGGTGGTGTGGGAGTC		R: TCGTGGTGATGGAGAAGCAC
LRRFIP1	18-19	chr2: 238666099-238666191; 238667372-238667464	PD skip	18-19	F: CAGCAGAAACAGGCGAGTTC	16-17	F: CTAGGCGGCAGTACGAAGAG
					R: TCATCCCGTTCACTCCTTACG		R: CAGGGCCTCCTTGACTTCAG
TRIM9	8-10	chr14: 51452862-51452674; 51450139-51450098; 51449683-51449660	PD skip	8-10	F: GTGCCTTCGGAAAGATTGCC	4-5	F: CAGTGAAATTGCGCCAGACC
					R: AATTGCATCCCGGTGCATTG		R: AGTGTGCCTTTACCCCACTG

Table 4. Splicing events and qRT-PCR primer sequences.

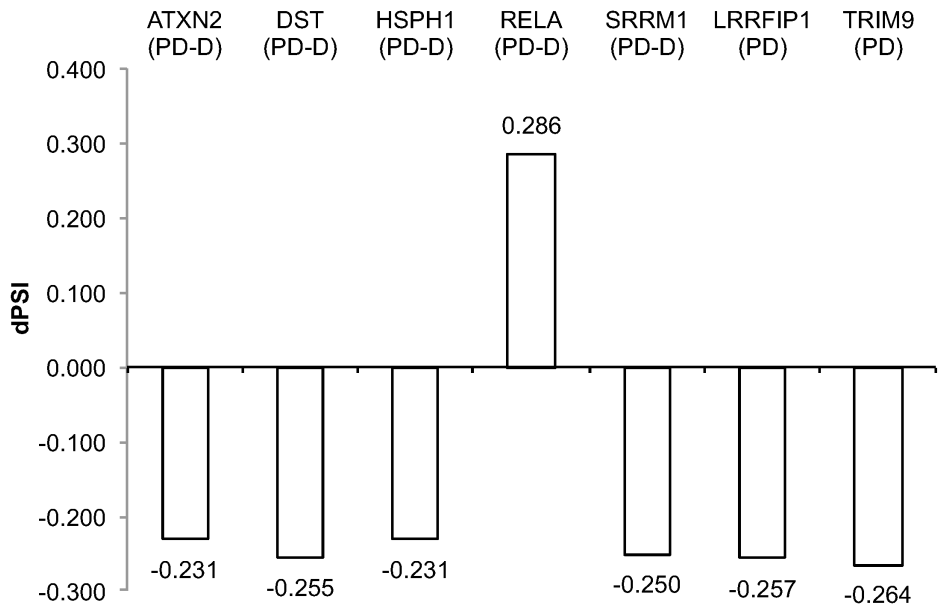


Fig 2. dPSI values for SpliceSeq-predicted alternatively spliced exons within the labeled genes. Exon dPSI represents the difference in percent spliced in, disease vs. controls (see Table 3 for exon details).

qRT-PCR.

Fold changes of the predicted alternative splice sites (fc_{as}) were compared to fold changes of adjacent sites lacking any predicted splicing activity, referred to as "no-splice" (fc_{ns}). Six of the seven genes were predicted to have exon skips, and one, an exon inclusion (*RELA*).

PD-D specific alternative splicing. At splice-predicted sites of *ATXN2* (fc_{as}= 0.674, fc_{ns}= 2.201), *HSPH1*(fc_{as}= 0.952, fc_{ns}= 1.924), and *SRRM1* (fc_{as}= 0.598, fc_{ns}= 0.780), fold changes were less than 1 relative to controls, in addition to having reduced expression compared to sites without predicted splice events. This is consistent with exon exclusion events in these transcripts. *RELA* (fc_{as}= 3.467, fc_{ns}= 1.323), at the predicted inclusion site, was greater than 1 and

overexpressed relative to the no-splice site. This indicates specific inclusion of exon 10.2. A change in *DST* was indiscernible from controls and its no-splice site (fc_{as}= 1.160, fc_{ns}= 1.171).

PD specific alternative splicing. *LRRFIP1* (fc_{as}= 0.794, fc_{ns}= 1.075) and *TRIM9* (fc_{as}= 0.667, fc_{ns}= 0.936) fc values (<1) likewise indicate reduced exon expression of the predicted exon for each gene. Figure 3 summarizes these results by comparing the fold change of alternatively spliced transcripts (fc_{as}) to the main transcript isoforms (no-splice (fc_{ns})) as expression ratios, represented by percent exon inclusion of the alternative splice site. Exon skips are less than 100%, while the inclusion is greater than 100%. Results demonstrate that the vast majority of predicted AS events selected for validation were confirmed by secondary qRT-PCR measures.

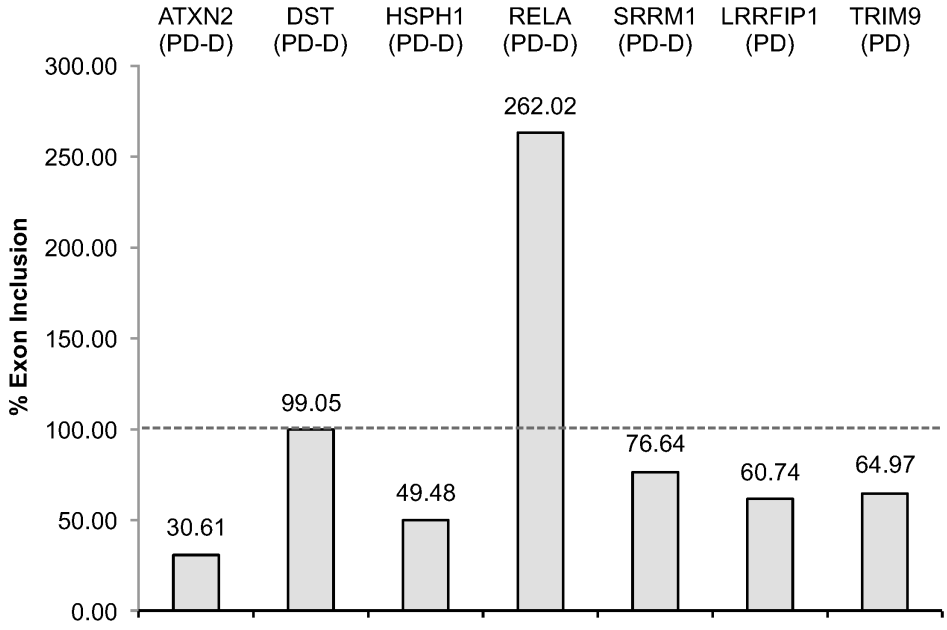


Fig 3. qRT-PCR disease vs. controls fold change comparison of splice vs. no-splice primer sites.
Exon regions with expected alternative splicing activity are compared to flanking exon regions without predicted alternative splicing, "no-splice". The dotted line indicates 100% inclusion (of a queried exon).

DISCUSSION

We present an mRNA-seq reference dataset to further characterize the molecular etiology of PD and PD-D. In addition to significant gene level differential mRNA expression, RNA-seq clearly reveals underlying differential alternative splicing in the posterior cingulate cortex during the course of PD and PD-D. Alternative splicing of *ATXN2*, *HSPH1*, *SRRM1*, *RELA*, *LRRFIP1*, and *TRIM9* suggest dysregulation of genes within immune and inflammation responses, as well as transcription and RNA processing. Compared to this alternative splicing profile, genes among those that were the most significantly differentially expressed (*CSF3*, *SELE*, *PENK*, *VGF*, *KRT5*, *CRH*, and *SST*) illustrate the breadth of PD gene dysregulation, beyond one level of expression analysis. In overall expression changes, immune activity is prominent, but there is also evidence of disrupted neuronal signaling pathways.

Differential Gene Expression.

Overexpressed in both PD and PD-D relative to controls: *CSF3* and *SELE*. One proposed contributing mechanism to PD etiology has been dysregulation of the immune system. Upregulation of both Granulocyte colony-stimulating factor (*CSF3*; $p_{PD} = 8.25x10^{-19}$, $fc_{PD} = 9.920$; $p_{PD-D} = 2.00x10^{-35}$, $fc_{PD-D} = 49.470$) and Selectin-E (*SELE*; $p_{PD} = 4.24x10^{-12}$, $fc_{PD} = 4.808$; $p_{PD-D} = 7.10x10^{-29}$, $fc_{PD-D} = 25.334$) in PD patients, prior to significant pathology in the posterior cingulate cortex, is consistent with an early role of immune induction in the progression of PD.

CSF3 (or *G-CSF*) is one of three major *CSFs*, distinguished for its granulocyte involvement in differentiation and apoptosis inhibition(Schneider et al. 2005). It is a major cytokine regulator of neutrophils, and promotes their increased production in bone marrow(Panopoulos and Watowich 2008; Fife et al. 2009). Due to its immune functions, normal *CSF3* levels in healthy individuals are relatively low and increase greatly under disease state. Pharmacological administration of *G-CSF* has been used as an immunotherapy to expedite and maintain the natural immune response and its receptors have been found in hematopoietic and non-hematopoietic cells, including in central nervous system (CNS) neurons and adult neural stem cells(Panopoulos and Watowich 2008). Comparable to their immune system role of promoting

neutrophil production by stem cell differentiation, *CSF3* can induce neuronal differentiation from adult progenitor cells while also protecting neurons from apoptosis(Schneider et al. 2005).

SELE is a cell adhesion molecule (CAM) that aids leukocytes in attaching to the endothelium; monocyte CAMs are activated in neurodegeneration(Hochstrasser et al. 2010). Elevation of serum *SELE* has been detected in Guillain-Barre syndrome and MS, but when investigated in Alzheimer's disease patients, levels did not differ(Rentzos et al. 2005). To our knowledge, this is the first report showing induction in cortex of PD patients.

Underexpressed in both PD and PD-D relative to controls: *PENK*, *VGF*, and *KRT5*. Underexpressed genes in both PD and PD-D suggest potential deficits in neurotransmitter signaling and altered cytoskeletal function.

Proenkephalin (*PENK*; $p_{PD} = 5.73x10^{-8}$, $fc_{PD} = -3.355$; $p_{PD-D} = 1.57x10^{-4}$, $fc_{PD-D} = -3.049$) is a precursor protein of signal peptides that function as neurotransmitters, autocrine and paracrine factors, and hormones(Denning et al. 2008). Consistent with our expression results, Backman *et al* report evidence of significant down-regulation, relative to young controls (mean age 31), in both aging controls (mean age 77) and PD patients(Backman et al. 2007). *PENK* is also linked to bipolar disorder(Altar et al. 2009).

Vgf nerve growth factor (*VGF*; $p_{PD} = 1.06x10^{-10}$, $fc_{PD} = -3.821$; $p_{PD-D} = 1.61x10^{-8}$, $fc_{PD-D} = -5.621$) is a neurosecretory protein most often expressed in neuroendocrine cells and neurons(Ferri et al. 2011). It has been studied and targeted as a promising disease biomarker due to reduced expression in cerebrospinal fluid (CF) for both AD and ALS patients (Zhao et al. 2008; Jahn et al. 2011); expression in parietal cortex is reportedly reduced in PD(Cocco et al. 2010). Ferri *et al.* support *VGF* as a biomarker of neurodegenerative and other diseases for its differential CF expression and cell type specific profiles within neuroendocrine organs(Ferri et al. 2011).

Keratin 5 (*KRT5*; $p_{PD} = 1.47x10^{-7}$, $fc_{PD} = -4.718$; $p_{PD-D} = 6.75x10^{-4}$, $fc_{PD-D} = -4.373$) encodes a major keratin of the basal cell layer as part of the intermediate filament cytoskeleton in basal keratinocytes(Betz et al. 2006; Chamcheu et al. 2011). Epidermolysis bullosa simplex causes intraepidermal blistering and is caused in part by a *KRT5* mutation(Chamcheu et al.

2011). Keratins are likely necessary for cell adhesion organization, organelle transport and nuclear anchorage(Betz et al. 2006).

Underexpression unique to PD-D: *CRH* and *SST*. Downregulation of certain hormone proteins suggest a potential hormone imbalance associated with dementia onset. Corticotropin-releasing hormone (*CRH*; $p = 6.4x10^{-7}$, fc = -4.724) and somatostatin (*SST*; $p = 4.44x10^{-9}$, fc = -5.144) are underexpressed, as well as *PENK* ($1.57x10^{-4}$, fc = -3.049), all of which are indicated by gene ontology to be involved with an altered hormone signaling state.

CRH is released along the stress response pathway by the hypothalamus and acts on the pituitary to release adrenocorticotropin (ACTH) to subsequently release a corticosteroid like cortisol(Bao et al. 2008).

SST (or *SRIF*, somatotrophin-release-inhibitory factor) is a neurotransmitter and a neuromodulator(Lu and Stoessl 2002). Exogenous expression may increase neuronal dopamine release and modulate dopamine-dependent behavior(Ionov and Pushinskaya 2013).

Alternatively Spliced Genes.

The genes that emerged from alternative splicing analyses did not exhibit a high degree of differential expression when assessed at the gene level. Nor did these genes clearly overlap with the top differentially expressed genes, although they often function in overlapping pathways. However, within the AS group, there are a variety of overlapping functions or instances of close networking.

Tripartite motif (*TRIM*) proteins belong to a subgroup of RING finger proteins, a structural feature that functions as a ubiquitin ligase (E3) to conjugate ubiquitin to target proteins; some *TRIMs* are important within innate and adaptive immune pathways(Tanji et al. 2010; Shi et al. 2014). *TRIM9* is thought to be a targeting signal for proteasomal degradation, may be involved in neuron development and synaptic vesicle exocytosis, and is a negative regulator of transcription factor NF-κB activation(Short and Cox 2006; Tanji et al. 2010; Shi et al. 2014). It has been found in Lewy Bodies (LB) of both dementia with LBs (DLB) and PD; expression is decreased in DLB and rabies virus infected brains, another case in which degeneration occurs(Tanji et al. 2010). Skipping of exons 8-10 approaching the C-terminus (protein position 550-636; dPSI = -0.264),

shown in our data, affects a fibronectin type III and B30.2/SPRY domain. This domain is associated with microtubule binding among the RBCC (N-terminal RING finger/B-box/coiled coil) protein family, to which TRIM9 and many other *TRIM* proteins belong. Mutations in the C-terminal B30.2-like domain of another RBCC protein, MID1, disrupts localization and ubiquitin targeting(Short and Cox 2006).

p65 protein is encoded by *RELA* (rel-like domain-containing protein) and is a subunit of the NF-κB transcription factor complex. They are part of a family of proteins involved in immune and inflammatory responses, apoptosis, and cell proliferation and differentiation(Huang et al. 2009). NF-κB nuclear localization, from the cytosol, is inhibited by IκB association with the p50-p65 heterodimer, which is liberated upon phosphorylation of IκB(Sharif-Askari et al. 2010). Though both *TRIM9* and p65 interact with the NF-κB pathway, it has been shown that *TRIM9* affects activation steps subsequent to the activation steps directed on p65(Shi et al. 2014). Of the splicing activity predicted and validated here, the exon inclusion (exon 10.2, protein position 347-381; dPSI = 0.286) within *RELA* leading to a possible protein truncation, with an activation domain deletion, is presumably the most detrimental alteration, predicted to cause overproduction of a truncated, inactive form of the protein.

Heat shock proteins (HSP) are molecular chaperones that assist in protein folding or degradation, ER stress response regulation and steps in the apoptosis cascade(Meares et al. 2008). *HSPH1*, a 105 kDa HSP (*HSP105*), directs caspase-3 mediated apoptosis following ER stress, and interacts with α-tubulin to suppress disorganization during heat shock(Saito et al. 2003; Meares et al. 2008). *HSP105β* is a truncated isoform (92 kDa) induced specifically at 42°C, not by other stressors(Ishihara et al. 1999). Our data shows a skip (exon 13, protein position 529-572; dPSI = -0.231), elevated in PD-D compared to controls, consistent with this β isoform. Normal physiological temperature ranges from peripheral 33°C to fever-induced 39°C and does not reach the extreme temperature (42°C) associated with a heat shock response(Katkere et al. 2010). If the high-heat responding β isoform is overproduced instead of the constitutively active form, signaling for apoptosis could be dampened as a result of reduced functional activity within the normal physiological temperature range.

Serine/arginine repetitive matrix 1, *SRm160* (or *SRRM1*, SR-related matrix protein of 160 kD), is a pre-mRNA splicing co-activator that associates with the nuclear matrix along with its binding partner SRm300, aiding in recruitment of splicing factors to speckled regions and assembly of complexes; *SRm160* may also be involved in mature mRNA transport within the space of the nucleus(Blencowe et al. 1998; Wagner et al. 2003). Disruption to a sequence portion that is necessary for speckle and matrix localization, both structures at which pre-mRNA splicing is shown to occur within the nucleus, could hinder recruitment of pre-mRNA to these areas and therefore disrupt proper splicing and processing to mature mRNA. In the PD-D patients, exon 16 (protein position 558-570; dPSI = -0.250), which is within the region encoding speckles and matrix localization specificity, is preferentially included in the mature mRNA.

Ataxin-2 is an RNA binding protein involved in RNA processing and endocytosis, impacting membrane receptor presence. In vivo knockout of the mouse homologue *SCA2* leads to increased insulin production with a concurrent reduction in insulin receptors, resulting in obesity. *ATXN2* also associates with EGFR (epidermal growth factor) in endocytic trafficking(Nonis et al. 2008), a process linked with huntingtin and parkin genes implicated in Huntington's and Parkinson's diseases, respectively. Exclusion of exon 21 (protein position 1127-1145; dPSI = -0.231), as reported here, was also previously identified by a broad sequencing and characterization study, though the effect of this particular event is unknown. While a poly-Q expansion within *ATXN2* is a notorious culprit of genetic malfunction, the exon 21 skip discussed here is far downstream (>2Kbp) from the well-studied region in exon 1. Similar to the aforementioned mouse study, alternatively spliced *ATXN2* might interfere with cellular responses by reducing receptor production or membrane incorporation. However, the specific functional effect of this alternative splicing event is unclear.

Leucine-rich repeat flightless-interacting protein 1 is a promoter-binding transcriptional repressor that imposes regulatory control on EGFR, and also associates with the platelet cytoskeleton during clot formation(Rikiyama et al. 2003; Goodall et al. 2010; Alkelai et al. 2011). *LRRFIP1* also may regulate the proinflammatory cytokine TNF and is a component of the TNFR superfamily(Alkelai et al. 2011; Li et al. 2014). These functions suggest *LRRFIP1* is closely tied to

cell damage and injury response. Exons 18 and 19 (protein position 398-459; dPSI = -0.257) skipping were increased in our PD group. Together, they partially encode coiled-coil domains and a prefoldin superfamily domain, protein features that promote protein-protein interactions and molecular chaperoning, respectively(Siegert et al. 2000; Rikiyama et al. 2003). Exon skipping in this region has the potential to decrease binding affinity with *LRRFIP1* targets and to interrupt any chaperone activity this protein may provide. Ontological query shows that other proteins with the prefoldin domain tend to be involved with protein folding and cytoskeletal binding. Therefore, this particular domain might influence the interaction of *LFFRIP1* with the cytoskeleton.

The Potential Roles of Alternative Splicing in PD.

Though alternative splicing is a normal mechanism of expression alteration and control, the pathways in which it is relatively increased suggest the most active responses in the disease state. Analysis of alternative splicing can reveal aberrant splicing of key disease genes. A number of scenarios are possible. An aberrantly spliced transcript may drive a pathway to become overactive, contributing to disease onset or progression. Another possibility is that cells in distress during the progression of PD and PD-D may experience altered splicing as a consequence of widespread dysfunction. Distinguishing between these possibilities is an important goal of future work.

CHAPTER 2

DNA METHYLATION CHANGES ASSOCIATED WITH PARKINSON'S DISEASE PROGRESSION: OUTCOMES FROM THE FIRST LONGITUDINAL GENOME-WIDE METHY

INTRODUCTION

Parkinson disease (PD) is the second most common neurodegenerative disorder of the elderly, currently affecting~2% of the population over 60 years of age(Dauer and Przedborski 2003). PD presents clinically as a progressive movement disorder with resting tremor and postural instability; and it is characterized neuropathologically by intracytoplasmic α-synuclein (α-syn) aggregates in Lewy bodies (Spillantini et al. 1997). Neurodegeneration occurs primarily in dopaminergic neurons of the substantia nigra, but Lewy body pathology occurs in limbic and cortical areas as PD progresses (Mattila et al. 2000).

PD is a multifactorial disease where environmental and genetic factors are intricately associated. In idiopathic PD, 60-70% of dopaminergic neurons have already been lost by the time someone presents clinical symptoms sufficient for a diagnosis (Ferrer et al. 2011b). Multiple pre-motor biomarkers are actively being investigated for their potential to identify early-stage PD or patients at risk for developing PD (Haas et al. 2012), including clinical measures (rapid eye movement behavior disorder (RBD), olfactory deficits, mood disorders); molecular measures (α-syn in cerebrospinal fluid and blood); and brain imaging.

Epigenetic mechanisms are emerging as important factors in the molecular etiology of neurodegenerative diseases, including PD (Pavlou and Outeiro 2017). Hypomethylation of the α-synuclein gene (SNCA) promoter region has been reported in substantia nigra of PD patients(Jowaed et al. 2010; Matsumoto et al. 2010). Moreover, this *SNCA* promoter hypomethylation has been shown to increase *SNCA* protein expression in cell culture, possibly contributing to the pathology of PD. Interestingly, L-dopa therapy has been associated with hypermethylation of the *SNCA* promoter, suggesting that current PD therapy may alter methylation (Schmitt et al. 2015). While results on altered *SNCA* methylation in PD have not been replicated by other studies using smaller cohorts (Richter et al. 2012; Guhathakurta et al. 2017); epigenomic changes associated with other genes including hypomethylation of *NPAS2*(Lin et al.

2012) and *CYP2E1*(Kaut et al. 2012); and hypermethylation of *PGC1-d*(Su et al. 2015) and the H1 haplotype of Tau (*MAPT*)(Coupland et al. 2014), have also been implicated in PD.

We previously demonstrated alterations in the intracellular localization of DNA methyltransferase 1 (DNMT1), which catalyzes the addition of methyl groups to the DNA, and that appeared retained in the cytoplasm in neurons from PD patients (Desplats et al. 2011), due to interaction with misfolded α-syn. These observations provided a potential mechanistic explanation for the hypomethylation of multiple genes in PD and suggested global alterations in the methylome of PD patients. Indeed, we subsequently identified methylation changes in brain and blood samples from PD cases in comparison to control subjects in a pilot study using a small cohort. Notably, we identified concordant methylation changes in matching brain and blood samples, thus supporting the idea that blood may serve as a surrogate tissue for brain methylation analyses in PD(Masliah et al. 2013). Subsequently, Moore et al.(Moore et al. 2014) reported a subset of CpG sites with altered methylation in PD and confirmed two CpG sites via bisulfite sequencing in a second cohort of 219 PD patients versus 223 control individuals. Most recently, an epigenome-wide association study (EWAS) using 335 PD and 237 control blood DNA samples identified 82 CpG sites of altered DNA methylation in PD patients, also using the previous generation Infinium 450K HumanMethylation beadchip(Chuang et al. 2017).

These studies provide indications that there are significant alterations to the methylome of PD patients in both brain and blood tissues. Noteworthy, these previous studies were based on single-time point samples, hindering the observation of potential epigenomic changes associated with disease progression. Here we report results of the first longitudinal analysis of genome-wide methylation surveying over 850,000 CpG sites in 189 PD patients and 191 controls enrolled in the Harvard Biomarkers Study at both baseline and a follow-up visit that was ~2 years later. We present evidence of specific methylation changes associated with PD status, disease progression, and PD medication.

MATERIALS AND METHODS

Study cohort

This longitudinal case-control study is nested within the Harvard Biomarkers Study (HBS) (Ding et al. 2013; Locascio et al. 2015; Liu et al. 2016; Liu et al. 2017). The HBS is an ongoing case-control study including individuals with PD, Alzheimer's disease, and controls without neurologic disease, and collecting high-quality biosamples and high-resolution clinical phenotypes longitudinally over a five-year period (under funding from the Harvard NeuroDiscovery Center). Clinical characteristics of patients with PD enrolled in HBS that were selected for this study are shown in Table 5. Individuals with early-stage PD and controls (CT) were enrolled into HBS from Massachusetts General Hospital and Brigham & Women's Hospital. Inclusion criteria for cases with PD were age ≥ 21; diagnosis of PD according to UK brain bank criteria; MMSE score > 21. Main exclusion criteria for cases with PD were diagnosis of a blood or bleeding disorder, known hematocrit < 30. Cases for the nested longitudinal case-control methylation study were selected from the HBS population based on additional criteria that included the availability of follow-up visit(s), age ≥ 55, baseline Hoehn and Yahr stage ≤ 3, baseline MMSE > 27. Sex-matched and age-similar controls were selected who had a baseline MMSE > 27 and available follow-up visit(s). Cases and controls with a past medical history of cancer (that might affect methylation status) were excluded. To increase diagnostic certainty, subjects with a diagnosis change on longitudinal follow-up were excluded from the current study. Familial PD cases carrying a G2019S LRRK2 mutation were also excluded.

Phenotype	PD	Control
Female/Male	59/130	64/127
Age at baseline visit [1]	67.31 (6.95)/68.08 (7.05)	66.83 (2.63)/ 70.89 (6.16) ***
Age at follow-up visit	69.58 (7.34)/70.33 (7.05)	69.05 (2.54)/ 72.98 (6.03) **
Education at baseline [2]	15.25 (1.57)/15.22 (1.93)	15.63 (1.18)/14.98 (2.07)
Ever Smoked	27/53	20/37
Current Smoker (baseline)	1/0	2/5
Current Smoker (follow-up)	1/0	2/5
Disease Duration at baseline [3]	3.73 (3.99)/3.72 (3.97)	
HY [4] (baseline)	2.00 (0.48) / 2.01 (0.42)	
HY (follow-up)	2.33 (0.60) / 2.20 (0.42) ***	
t-test HY [5]	0.001717/0.0005481	
MMSE [6] (baseline)	29.02 (1.03)/ 28.98 (1.08)	29.26 (0.92) / 28.84 (1.17)
MMSE (follow-up)	28.71 (1.95)/ 28.73 (1.61)	29.22 (1.24) /28.96 (1.27)
WBC [7] (baseline)	6.28 (1.33) / 6.70 (1.59)	6.11 (1.53) / 6.41 (1.75)
WBC (follow-up)	6.48 (1.54) / 6.54 (1.55)	6.15 (1.71) / 6.40 (1.65)
RBC [8] (baseline)	4.56 (1.23) / 4.70 (0.42)	4.45 (0.34) / 4.75 (0.42)
RBC (follow-up)	4.37 (0.39) /4.58 (0.52)	4.44 (0.36) / 4.69 (0.42)
De novo [9] (baseline)	5/14	
De novo (follow-up)	1/2	
on medication[10] (baseline)	38/82	
on medication (follow-up)	50/108	

Table 5. Clinical and demographic characterization of the HBS study cohort. Values are expressed means with (S.D.) *** indicates p value <0.001 and ** indicates p value <0.01 for inter-group comparisons as per Student's *t* test. (1) Age expressed in years. (2) Education expressed in years. College or above =16; High school= 12; elementary school = 5. (3) Disease duration is calculated in years since diagnosis. A value of 0 is assigned is at baseline if the patient has received a diagnosis of PD during the same year of enrolling in the study. (4) Modified Hoehn and Yahr scale for clinical staging of Parkinson's disease (Goetz et al. 2004). (5) Indicates pvalue of Welch two-sided two-sample t-test comparing the indicated category between enrollment and follow-up visits. Female and Male groups were analyzed separately. Only provided for significant differences. (6) Mini-Mental State Examination. (7) White blood cells count. (8) Red blood cells count. (9) De novo patients that yet did not receive any type of anti-parkinsonian medication. (10) Based on Parkinson's disease treatment that may affect one-carbon metabolism as defined in our study, including Sinemet; Comtan and Stalevo. Data was not available for: HY enrollment 2 cases; HY follow-up 13 PD cases; MMSE enrollment 61 CT cases; MMSE follow-up 67 CT and 2 PD; WBC/RBC enrollment 10 CT and 6 PD; WBC/RBC follow-up 48 CT and 50 PD cases.

Array Processing

Genomic DNA (1µg) samples received from HBS were coded and randomized with respect to disease status. DNA was bisulfite converted (EZ DNA Methylation kits, Zymo Research, D5003) per Illumina's recommendation. The samples were processed and hybridized to Infinium MethylationEPIC BeadChip (Illumina, WG-317-1002) and signal was scanned with

Illumina's iScan. Longitudinal sample pairs corresponding to the same subjects were run on the same chip to avoid batch effects. Raw IDAT files were exported for processing in R.

Data Normalization and Quality Control

The study was conducted at the laboratories of Dr. Dunckley at Arizona State University (ASU) and Dr. Desplats at University of California San Diego (UCSD). Both laboratories applied unified standard operating procedures (SOPs) according to Illumina's recommendations. We performed careful quality control and pre-processing steps using the R Bioconductor package Minfi v. 1.22.1 (Aryee et al. 2014). Detection P values were calculated. No samples had mean detection P value > 0.05. Sex prediction was performed and 8 samples with discordant calls were removed from the analysis. Subject identity in paired-samples was determined using SNPs-matching probes contained in the EPIC array (Heiss and Just 2018). Samples with ratios of non-methylated/methylated sites (uMeth/mMeth) < 10.5 were also removed. The call rate was calculated as the proportion of probes in each sample with a detP of < 0.01. Samples with detP > .99 were removed. We ran technical replicates across sites and batches for control. Replicates were removed by taking the sample with the highest call rate. As a result, a total of 36 samples were removed from downstream analyses and the remaining 792 samples (two time-point longitudinal samples from 197 PD cases and 199 controls) were normalized using ssNoob, a method recommended for EPIC array data processing (Fortin et al. 2017). After normalization, probes were removed that failed in one or more samples (detP > 0.01), were located on sex chromosomes, had SNPs at the CpG site, or documented to be cross-reactive from Pidsley et al.(Pidsley et al. 2016), leaving 755,625 probes for analysis. In addition, we evaluated the correlation of intensities between sample replicates and longitudinal pairs repeated across the arrays to verify association.

Methylation Data Analysis: PD vs Control with repeated measures

All probes were used to build multi-dimensional scale plots to visualize the variation in the data. Probe-wise differential methylation analysis was performed with the Bioconductor package limma. Beta values were converted to M-values for statistical analysis. Consensus correlation was calculated for longitudinal samples using the limma duplicateCorrelation function using

10,000 randomly selected probes and patient ID as the block parameter. The model design for the longitudinal analysis was ~0+condition_visit + batch_site + age + sex + smoke + duration_at_baseline + CD8T+CD4T+NK+Bcell+Mono+Gran, and was adjusted as needed to test PD vs CT . The limma function lmFit including patient ID as the block variable and the consensus correlation from duplicate-Correlation() function was run with the specified design on M-values, followed by fitting the desired contrasts and running the limma function eBayes to calculate differentially methylated probes. The corresponding Beta value is also included in tables and was used to calculate differential methylation as Delta Beta between the indicated comparisons. Differentially methylated regions (DMRs) were analyzed using DMRcate(Peters et al. 2015a).

Longitudinal analysis

Longitudinal analysis was performed on different subsets of the data (Control only, PD only, Medication true only and Medication false only). On each of these datasets, we fit a linear model by robust regression using an M estimator with the R function rlm in the R package MASS. The model is outlined below and includes the covariates baseline age (blineAge), sex, cell type composition, site, smoking status (smoked) and duration at baseline plus time between visits (duration_plus_time). The term “time” in the model is composed by the sum of disease duration at baseline (which = 0 in control subjects or PD cases diagnosed less than one year before baseline sample) and the time interval between baseline and follow-up samples.

Rlm (x~blineAge+Sex+CD8T+CD4T+NK+Bcell+Mono+Gran+Site+Smoked+duration plus time -1, maxit=100).

RESULTS

Our study included 380 participants from the Harvard Biomarker Study, 189 patients with PD (31% females; average age of 68 at enrollment visit) and 191 control subjects (34% females; average age of 69 at baseline, Table 5). From the 380 participants, 313 self-reported as white non-Hispanic; 40 as white Hispanic/Latino; 1 as Asian non-Hispanic; 2 as African-American Hispanic/Latino. Data was not available for the remaining 24 subjects. Levels of education did not significantly vary between cases and controls and overall 82% of participants attended

college. There were no significant differences in red or white blood cell count (as reported in blood tests) or in cognitive status (as per Mini-Mental State Examination, MMSE) between cases and controls at baseline or follow-up. We profiled samples obtained at the enrollment visit (baseline point) and a longitudinal sample corresponding to the second visit, which was collected on average 2.2 years later (S.D.=0.85) and used as the follow-up point. The time elapsed between visits ranged between 0.8 and 11.98 years and the mean of distributions was similar between controls and PD cases ($p=0.2206$ as per unpaired t-test). We selected PD cases with confirmed clinical diagnoses and with Hoehn and Yahr scale scores ≤ 3 at baseline, representing early or mild PD. Disease duration was similar between female and male PD patients, with an average duration of 3.7 years (S.D.=3.97) at baseline. During the period elapsed between baseline and follow-up visits no changes in cognition were observed in controls, while PD patients showed decay in MMSE performance, which was only significant for males (Table 5). Smoking has been largely reported as a protective factor and inversely correlated with the occurrence of PD(Ascherio and Schwarzschild 2016) . In our study 41.8 % of PD cases and 29.8 % of controls reported to have previously smoked. The number of active smokers was low with only 1 PD case being an active smoker in comparison to 7 control subjects still smoking (Table 5). Smoking status did not change between baseline and follow up visits and was included as a covariate in our analysis. Overall, PD patients showed a worsening in clinical manifestations as per HY scores, which changed significantly between baseline and follow up visits ($p<0.0001$, as per paired t-test), indicating disease progression. Similarly, the Unified Parkinson's Disease Rating Scale (UPDRS) total scores increased between baseline and follow-up visits at an average rate of 3.05 points/year ($p<0.0001$, as per paired t-test), consistent with disease worsening.

Estimation of blood cell composition using methylation data.

We used whole blood DNA to profile methylation; therefore, different lymphocyte cell type distributions between cases and controls may confound the analysis. We used distinctive cell-specific methylation profiles to estimate the proportional abundance of blood cell types and to evaluate whether alterations in white blood cell composition may be associated with PD pathology and have the potential to drive differential methylation between cases and controls. We

applied the “estimate-CellCounts” function in minfi (Jaffe and Irizarry 2014) to estimate the proportional abundance of blood cell types in our study samples by integrating the intensity of specific probes present in the EPIC array. We observed that granulocytes (as a group, including neutrophils) were the most abundant cells in blood, as expected (**Fig. 4**). Overall blood cell composition varied between control and PD groups. At baseline, PD patients showed higher estimated levels of granulocytes (p=4.0E-6, as per t-test) and lower estimated B-cells (p=0.0019) and NKs (p=0.00055) in comparison to controls. These differences only persisted for granulocytes, which were higher in PD cases (p=0.0066) and natural killers, which were lower in PD (p=0.00065) in the follow-up visit. Intra-group analysis showed that only granulocytes (p=0.00063) changed longitudinally in control subjects, while no changes were observed in PD cases between the time points analyzed (**Fig. 4**).

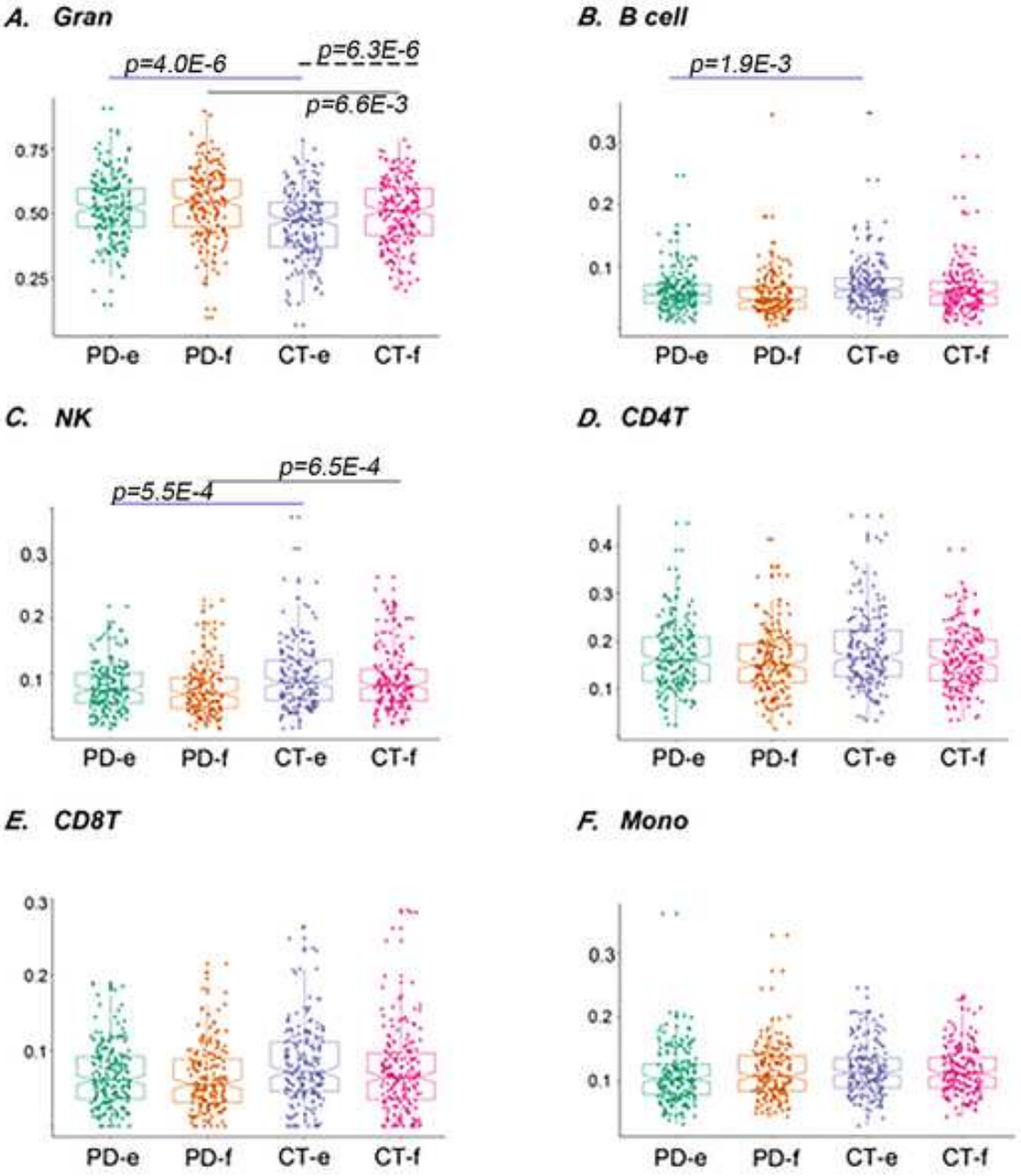


Figure 4. Comparison of individual cell type across control (CT) and PD groups at enrollment (e) and follow-up **(f).** Abundance of specific blood cell types was estimated based on unique methylation markers for cell identity. Shown in (**A**) granulocytes, in (**B**) B cells, in (**C**) natural killer cells, in (**D**) CD4T cells, in (**E**) CD8T cells, and in (**F**) monocytes. Blue solid line indicates comparison between PD cases vs. CT subjects at enrollment; black solid line indicates comparison between PD cases vs. CT subjects at follow-up; dash blue line indicates comparison between PD cases at follow-up vs. enrollment time points; dash black line indicates comparison between CT subjects at follow-up vs. enrollment time points. P-value for the differences in cell composition estimates across groups as per Wilcoxon test after correction for multiple observations are indicated.

Methylation changes associated with Parkinson's disease: cross-sectional analysis.

We first conducted a cross-sectional analysis by comparing methylation profiles between cases and controls regardless of time of visit to define disease-associated changes in methylation, while accounting for repeated measures, using limma. We identified 7 probes showing differential methylation (DMPs) in PD with genome-wide scale significance at a p-value < 5.0E-7 (adjusted pvalue<0.05) and another 46 DMPs with marginal significance of adj. p < 0.2 (Fig. 5; Table 6). Overall changes in methylation were modest, with $|\log_2 FC| < 0.6$ (in M-values). Among the top changes in this group, we identified multiple CpGs clustering around 200 bp of the TSS of the Lamin Tail Domain Containing 1 (IFLTD1/LMNTD1), an intermediate filament protein, and Delta Like Non-Canonical Notch Ligand 1 (DLK1), a transmembrane protein involved in differentiation of multiple cell types. Additional functional categories represented in this list (Table 6) include microtubule associated proteins Doublecortin Like Kinase 1 (DCLK1: p=3.19E-6) and Dynein Cytoplasmic 1 Heavy Chain 1 (DYNC1H1: p=3.24E-6); the transcription factor LIM Domain Only 3 (LMO3: p= 3.68E-6), and a neurotransmission regulator synaptotagmin 12 (SYT12: p=2.79E-7). These represent diverse cellular processes, some of which would not be expected to be important in blood cells (LMO3, which is highly expressed in brain(Sang et al. 2014); SYT12, essential in synaptic transmission(Kaeser-Woo et al. 2013)). Thus, these findings may be reflective of similar alterations in the brain. This is supported by the finding of CYP2E1 in this study (Table 6) and previous reports of altered CYP2E1 methylation in the PD brain(Kaut et al. 2012).

Noteworthy, many DMPs associated with PD present intermediate levels of methylation (ecific s>0.2 and <0.8), which represent the more dynamic portion of the methylome and more likely to respond to environmental/physiological cues, and which may reflect intrinsic alterations due to disease progression.

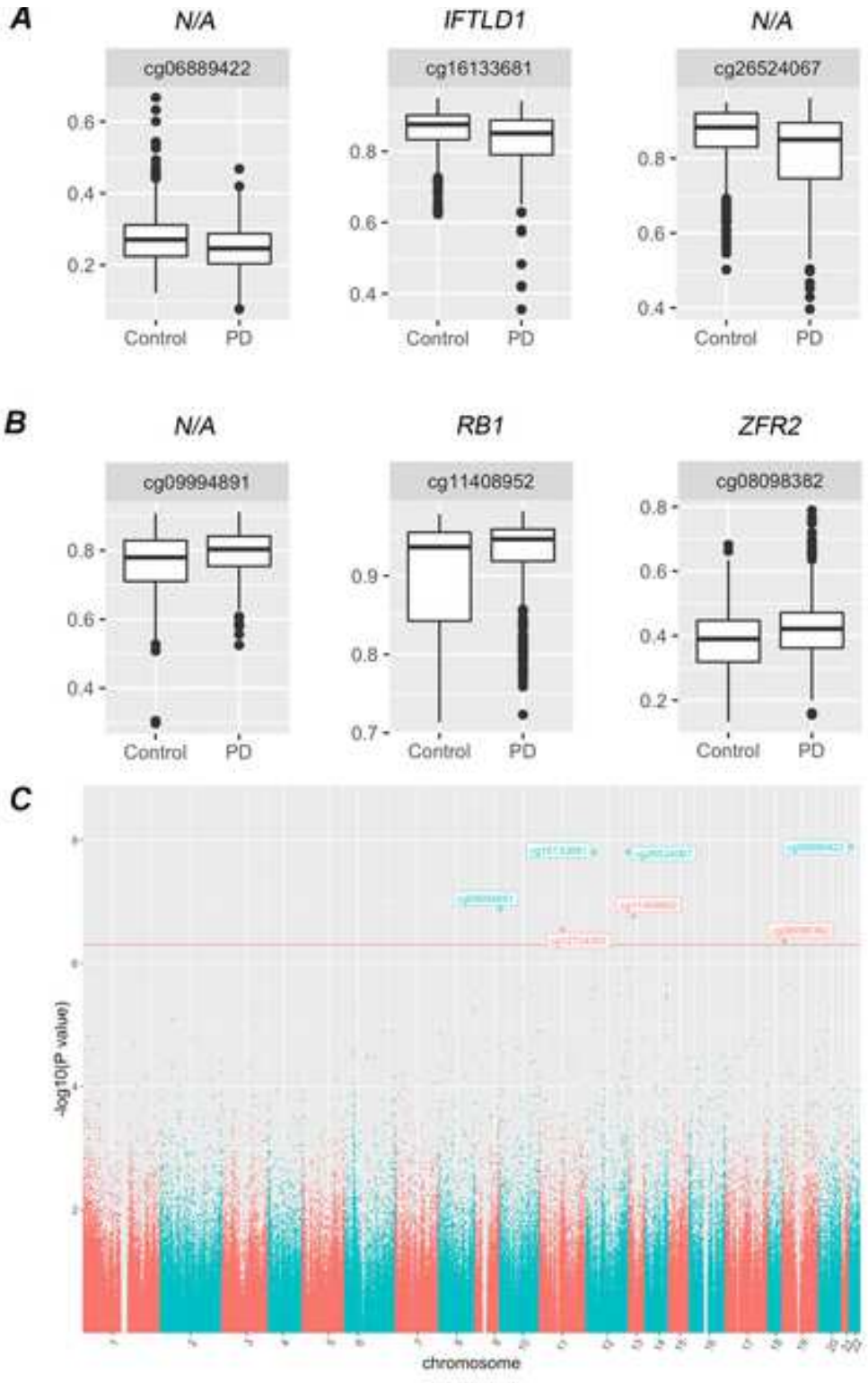

A
N/A
cg06889422
0.6
0.4
0.2
Control
PD
IFTLD1
cg16133681
0.8
0.6
0.4
Control
PD
N/A
cg26524067
0.8
0.6
0.4
Control
PD
B
N/A
cg09994891
0.8
0.6
0.4
Control
PD
RB1
cg11408952
0.9
0.8
0.7
Control
PD
ZFR2
cg08098382
0.8
0.6
0.4
0.2
Control
PD
C
-log10(p value)
chromosome

Figure 5. Case versus control comparisons in the HBS study. Representative box plots of top significant DMPs showing decreased (**A**) or increased (**B**) methylation in PD cases in comparison to control subjects. (**C**) Manhattan plot compiling genome-wide methylation sites and highlighting significant DMPs for the cross-sectional comparison PD vs. CT. Red line

Name	Chr	Position	Relation to TSS	Gene ID	Fold Change (log_2)	P-Value	FDR
cg06889422	22	24627294	Body	N/A	-0.284	1.28×10^{-8}	**0.0039**
cg16133681	12	25801621	TSS200	IFLTD1	-0.386	1.54×10^{-8}	**0.0039**
cg26524067	12	133003928	Open Sea	N/A	-0.497	1.55×10^{-8}	**0.0039**
cg09994891	10	2173024	Open Sea	N/A	0.319	1.31×10^{-7}	**0.0248**
cg11408952	13	48892244	Body	RB1	0.486	1.72×10^{-7}	**0.0260**
cg12724357	11	66790285	TSS1500	SYT12	0.241	2.79×10^{-7}	**0.0351**
cg08098382	19	3869345	TSS1500	ZFR2	0.307	4.35×10^{-7}	**0.0470**
cg23979954	12	25801601	TSS200	IFLTD1	-0.306	6.65×10^{-7}	0.0623
cg18279536	14	101194748	Body	DLK1	0.183	8.04×10^{-7}	0.0623
cg10405605	10	6188149	TSS200	PFKFB3	0.152	8.24×10^{-7}	0.0623
cg04741728	12	133003907	Open Sea	N/A	-0.594	1.17×10^{-6}	0.0768
cg20787649	1	17636898	Body	N/A	0.530	1.22×10^{-6}	0.0768
cg03681383	12	25801522	TSS200	IFLTD1	-0.304	1.56×10^{-6}	0.0907
cg12342048	19	11465311	Body	AC024575.1	-0.247	2.16×10^{-6}	0.1081
cg19628497	14	101194267	Body	DLK1	0.205	2.22×10^{-6}	0.1081
cg13211181	12	25801455	1stExon	IFLTD1	-0.268	2.52×10^{-6}	0.1081
cg00515755	19	1005248	Body	N/A	0.204	2.57×10^{-6}	0.1081
cg11469325	10	75012359	1stExon	MRPS16	-0.299	2.57×10^{-6}	0.1081
cg04224786	8	144222401	Open Sea	N/A	0.270	3.03×10^{-6}	0.1096
cg08766508	13	36430582	TSS1500	DCLK1	0.315	3.19×10^{-6}	0.1096
cg25588820	12	108070383	Open Sea	N/A	-0.278	3.19×10^{-6}	0.1096
cg12127149	14	102487020	Body	DYNC1H1	0.094	3.24×10^{-6}	0.1096
cg05763097	14	103569340	Body	EXOC3L4	0.297	3.33×10^{-6}	0.1096
cg01181415	12	16757954	5'UTR	LMO3	-0.140	3.68×10^{-6}	0.1126
cg18121862	14	101195312	Body	N/A	0.153	3.72×10^{-6}	0.1126
cg13315147	10	135341528	Body	CYP2E1	0.356	4.02×10^{-6}	0.1170
cg18397450	14	105830631	Body	PACS2	-0.380	4.56×10^{-6}	0.1220
cg21435367	3	133574742	Body	RAB6B	-0.252	4.80×10^{-6}	0.1220
cg15756507	17	65471461	Body	N/A	0.139	5.05×10^{-6}	0.1220

Table 6. Top differentially methylated probes (DMPs) in PD in the cross-sectional analysis. Name indicates probe designation at Illumina EPIC human methylation array; chr is chromosome location of the CG; position refers to Genome Reference Consortium Human Build 37 (GRCh37) hg19; N/A indicates no annotated gene associated with the probe.

Identification of Differentially Methylated Regions.

DNA methylation is influenced by CpG topology and methylation in one site is dependent on the methylation status of nearby CpGs by potential cooperation in recruiting methylating/demethylating factors, extending their activity in wider local DNA domains (Lovkvist et al. 2016). CpG clusters showing concerted changes in methylation are deemed highly relevant in the modulation of transcription. We searched for differentially methylated regions (DMRs), or groups of at least 4 CpGs within proximal genomic locations using DMRCate(Peters et al. 2015a), a stringent model that links proximal sites after testing their significance as individual DMPs (Fig. 6, Table 7). Among the top DMRs in PD, we identified *CYP2E1*, with 13 CGs altered (p=1.22E-22). This gene was one of the first reported as differentially methylated in PD brains (Kaut et al. 2012). Therefore, our results in whole blood are consistent with this previous finding and support a role for methylation in regulating this gene, which contributes to cholesterol and lipid metabolism, pathways that are altered in PD, also highlighting the utility of whole blood profiling. Another notable DMR is *LY6G5C*, with 5 CGs altered (p=3.92E-9), which has recently been identified in brain tissue as an epigenetic marker of schizophrenia (Wockner et al. 2015).

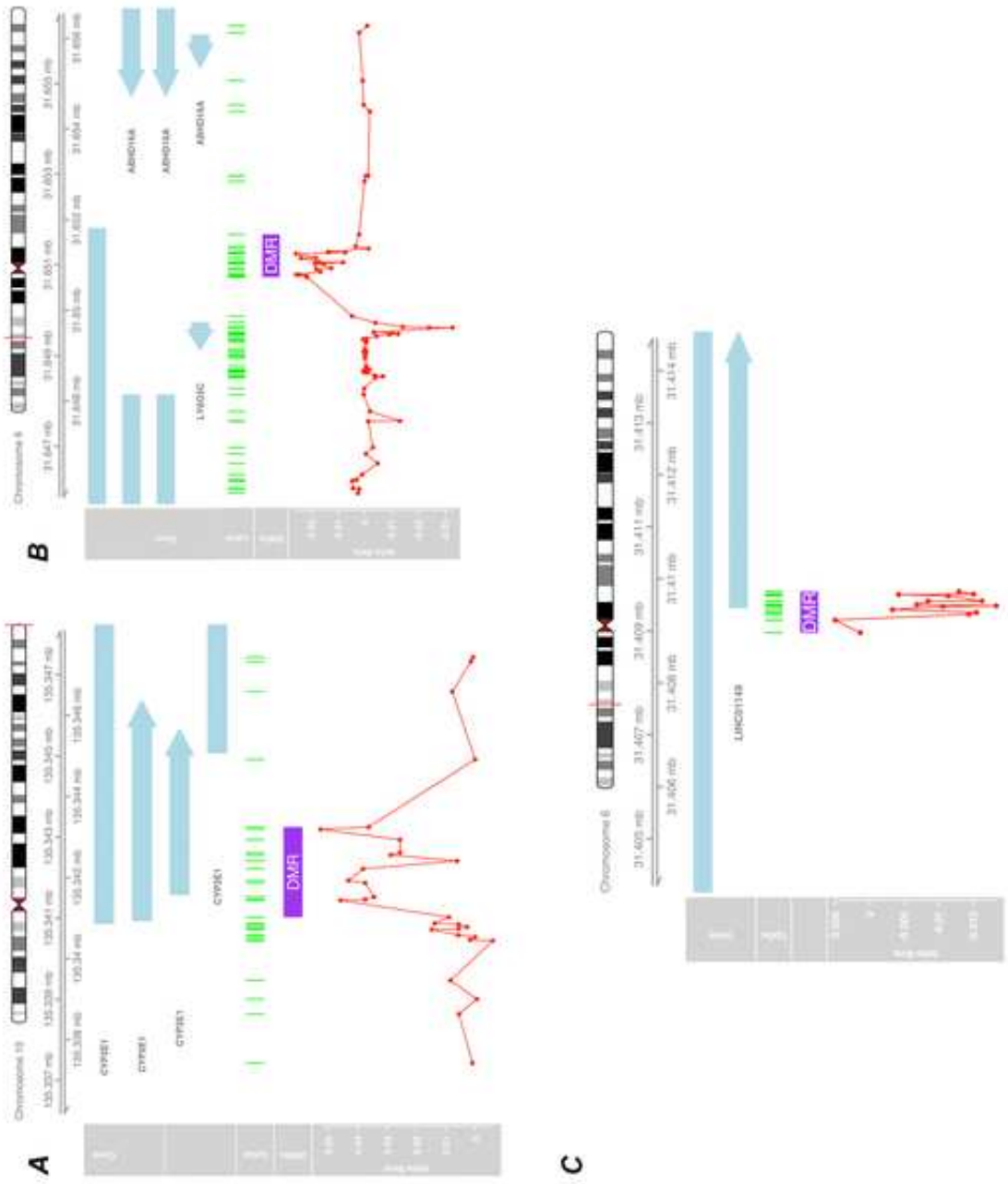


Figure 6. Graphical representation of differentially methylated regions (DMRs) in PD cases in the cross-sectional study. Representative schematics for top DMRs (Table 3) associated with CYP2E1 (A); LY6G5C (B) and CCDC89 **(C)**. Genomic location is indicated by chromosome position (based on Genome Reference Consortium Human Build 37 (GRCh37) hg19). Transcripts are indicated by light blue arrows. CGs appear as green lines. Red line represents Delta Beta of PD vs CT comparison for all the CGs constituting the significant region.

Chr	start	end	probes in region	minFDR	Stouffer	max beta fold change	overlapping promoters
Regions with increased methylation in PD cases							
10	135341025	135343248	13	1.22×10^{-22}	0.00004	0.06768	**CYP2E1**
19	54945959	54946993	6	4.14×10^{-15}	0.00019	0.05752	**AC008746.3**
18	72916776	72917390	6	1.14×10^{-10}	0.03901	0.04425	**ZADH2**
8	144222015	144222455	5	2.54×10^{-18}	0.00001	0.04395	**NA**
1	153599479	153600156	15	1.05×10^{-12}	0.10454	0.04178	**S100A1**
11	2889840	2891360	35	1.04×10^{-20}	0.81658	0.03802	KCNQ1DN
14	101194145	101195312	4	9.83×10^{-17}	0.00000	0.03520	DLK1
6	31650786	31650930	5	3.92×10^{-9}	0.38577	0.03252	LY6G5C
6	28983835	28984341	5	6.2×10^{-12}	0.19530	0.03166	NA
17	65471303	65471507	5	2.45×10^{-9}	0.02194	0.02850	NA
6	29599012	29599390	9	1.19×10^{-11}	0.23989	0.02814	GABBR1
10	6187854	6188415	5	2.87×10^{-14}	0.00086	0.02319	PFKFB3
15	91473059	91473569	10	2.38×10^{-11}	0.02671	0.02146	UNC45A
6	32144667	32146779	35	8.45×10^{-25}	0.05366	0.01962	AGPAT1
16	3114847	3115809	12	7.41×10^{-17}	0.00671	0.01755	IL32
Regions with decreased methylation in PD cases							
6	28945189	28945507	7	2.77×10^{-13}	0.00338	-0.04644	RN7SL471P
8	57350735	57351067	5	1.06×10^{-10}	0.06139	-0.04393	PENK
6	164506692	164507305	9	3.77×10^{-12}	0.03187	-0.04304	RP1
12	25801455	25801945	6	9.6×10^{-29}	0.00000	-0.04099	IFLTD1
13	111521981	111522651	5	2.12×10^{-12}	0.00725	-0.04075	LINC00346
10	77542302	77542585	9	1.28×10^{-12}	0.01495	-0.03571	LRMNDA
2	48844369	48845068	10	8.32×10^{-17}	0.04500	-0.03301	GTF2A1L
6	31409319	31409757	12	1.48×10^{-10}	0.22543	-0.02132	LINC01149
5	138210632	138211184	12	4.46×10^{-12}	0.05387	-0.01719	LRRTM2

Table 7. Differentially methylated regions (DMRs) associated with PD. Start and end indicate with the genomic coordinates for location of the defined DMR based on Genome Reference Consortium Human Build 37 (GRCh37) hg19. # CG indicates how many CGs were included in the DMR; minFDR is the minimum adjusted p-value from the CGs constituting the significant region; Stouffer is the Stouffer transformation of the group of FDRs for individual CGs at the DMR; max beta fold change is the maximum absolute fold change within the region.

Methylation changes associated with PD progression: longitudinal analysis.

The main goal of our study was to investigate whether blood DNA methylation changes as PD pathology progresses. We fit linear models by robust regression using an M estimator using the rlm() function in the R MASS package to identify genomic sites where methylation

varies over time in PD cases. We identified 138 DMPs that significantly changed over time in PD cases only at p<1.0E-7 with a rate of change ranging from a 0.8 % increase and a 0.6% decrease in methylation per year and supporting dynamic methylation changes in the blood methylome associated with disease progression (Table 8). Aging is an important determinant of DNA methylation (Fraga and Esteller 2007; Horvath et al. 2012). Although we controlled for age in our models; we also calculated the association of methylation changes with age to rule out that aging was the main driver of the longitudinal epigenetic changes we observed in PD. We observed a significant association with age for only 2/138 probes changing in PD over time. Furthermore, none of these longitudinal DMPs overlap with probes composing the epigenetic clock from Horvath (Horvath 2013), further supporting that the observed changes are due to disease progression. We used Ingenuity Pathway analysis to interrogate pathway enrichment in the differentially methylated probes. For this analysis we applied a less stringent cut-off criteria at p<1.0E-6 (886 DMPs and 534 mapped genes entered into the analysis). Notably, the top category enriched in the Disease and Function annotation was neurological disorders; including 13 genes associated with Alzheimer's disease (p=1.08E-2). Among enriched canonical pathways, biotin-carboxyl carrier protein showed a significant p-value of 2.25E-3, including DMPs mapping to Acetyl-CoA carboxylase alpha (*ACACA*), the rate limiting enzyme in the synthesis of long chain fatty acids. Interestingly, decreased long-chain acylcarnitines have been recently proposed as potential early diagnostic markers for PD (Saiki et al. 2017).

Probe ID	Chr	Position	Gene ID	Change rate PD	Pvalue PD	Change rate Controls	Pvalue Controls
Methylation decreasing over time							
cg26275301	17	81041828	METRNL	-0.604	7.56×10^{-8}	-0.00026	6.93×10^{-1}
cg07856430	18	42339206		-0.525	8.67×10^{-9}	-0.00058	2.85×10^{-1}
cg13153353	7	4175865	SDK1	-0.524	5.43×10^{-8}	0.00244	2.30×10^{-5}
cg03565777	12	125028339	NCOR2	-0.508	8.40×10^{-8}	-0.00276	1.38×10^{-6}
cg24136431	19	47016623		-0.392	6.40×10^{-8}	-0.00045	2.93×10^{-1}
cg16655626	19	45886078		-0.334	2.46×10^{-9}	0.00152	4.38×10^{-1}
cg04843111	1	156617074	BCAN	-0.328	1.00×10^{-9}	-0.00088	6.64×10^{-1}
cg08988821	7	74075293	GTF2I	-0.326	4.22×10^{-8}	-0.00107	6.71×10^{-1}
cg01791421	1	19996240		-0.309	8.38×10^{-8}	-0.00105	2.19×10^{-3}
cg13994376	13	112554122	LINC00354	-0.292	1.25×10^{-8}	0.00005	8.71×10^{-1}
cg15987655	5	139196457		-0.278	5.86×10^{-8}	-0.00285	1.42×10^{-1}
cg03389720	16	8780048	ABAT	-0.277	2.00×10^{-8}	-0.00091	1.88×10^{-3}
cg15765398	21	46409994	LINC00163	-0.272	5.57×10^{-9}	-0.00114	4.16×10^{-1}
cg09750643	11	1718086	KRTAP5	-0.257	8.04×10^{-9}	-0.00022	4.07×10^{-1}
Methylation increasing over time							
cg17046825	13	21081216		0.834	5.35×10^{-8}	-0.00311	6.60×10^{-4}
cg26126295	4	119095240		0.775	1.78×10^{-8}	0.00030	9.44×10^{-1}
cg06688960	20	1504932		0.713	1.04×10^{-9}	-0.00053	8.49×10^{-1}
cg25979148	4	871629		0.210	2.55×10^{-8}	-0.00001	9.95×10^{-1}
cg15540764	7	36919658	ELMO1	0.140	7.33×10^{-8}	-0.00040	9.61×10^{-3}
cg05933219	2	240099221	HDAC4	0.134	5.61×10^{-8}	-0.00058	4.87×10^{-1}
cg14946911	20	56772325		0.128	2.24×10^{-9}	0.00001	9.90×10^{-1}
cg09593391	8	144737908	ZNF623	0.126	9.53×10^{-8}	-0.00017	8.51×10^{-1}
cg18496624	17	38094692		0.123	5.44×10^{-9}	-0.00005	7.14×10^{-1}
cg10401356	8	140712424		0.123	1.11×10^{-8}	0.00037	6.62×10^{-1}
cg08561469	16	81944150	PLCG2	0.116	3.38×10^{-9}	0.00002	8.52×10^{-1}
cg15475168	7	101860421		0.115	1.16×10^{-8}	-0.00023	7.69×10^{-1}
cg20397902	16	88624807		0.109	1.06×10^{-8}	-0.00023	7.43×10^{-1}
cg01202950	15	74943647	EDC3	0.106	6.20×10^{-9}	-0.00027	6.90×10^{-1}

Table 8. Top longitudinal changes in methylation in PD cases. Name indicates probe designation at Illumina EPIC human methylation array; chr is chromosome location of the CG; position refers to Genome Reference Consortium Human Build 37 (GRCh37) hg19; N/A indicates no annotated gene associated with the probe. Rate of change indicates percentage methylation change/year.

Longitudinal methylation changes associated with PD medication.

Dopamine replacement is the standard of clinical care for Parkinson's disease and the vast majority of patients receive dopamine precursors, like levodopa/carbidopa (commercialized as Sinemet or Stalevo) and/or inhibitors of Catechol-O-methyltransferase (COMT; commercial name Comtan). Importantly, the metabolism of these compounds directly impact the one-carbon pathway, potentially affecting the supply of methyl-group donor molecules and the activity of DNA methyltransferases (**Fig.7**). Although current levopoda products are formulated to prevent break-down in the gastrointestinal tract and systemic circulation, the potential impact of these drugs on blood methylation has not been explored at genome-wide scale before. Therefore, we analyzed the effect of anti-parkinsonian therapy on DNA methylation by defining the category "PD Medication". PD cases receiving Sinemet; Stalevo and/or Comtan, either as single drugs or in combination at any dose and at any time point were categorized as the PD medication group. PD cases that did not take any of these drugs by the time of blood sampling were coded as PD NOT medicated. According to this classification, 69 patients were grouped as "PD NOT Medicated", from which 27 remained unexposed to these drugs at the follow-up visit.

We first analyzed longitudinal changes in methylation in the group of PD cases receiving PD medication. We identified 237 probes showing significant changes of methylation over time at a $p<1.0E-7$ and showing modest changes in methylation ranging in the order of 0.63 % reduction and 0.86 % increase in methylation/year (Table 9; Fig. 8). Longitudinal methylation in these sites did not change significantly in control subjects. Only 23/237 of the probes changing in the PD medicated group were significantly associated with aging, The remaining 214 CpG sites showed longitudinal changes in patients taking medication, which are likely involved in both the effect of medication and PD progression over time. Genes tagged by these CpGs (Table 9) function as transcription factors (ZNF544, ZNF623, GTF2I), extracellular matrix proteins (BCAN), non-coding RNAs (Y_RNA, LINC00163), neural cell adhesion (PCDH1), and synaptic transmission (RIMBP2), once again suggesting that signals related to central nervous system dysfunction can be detected in peripheral blood DNA.

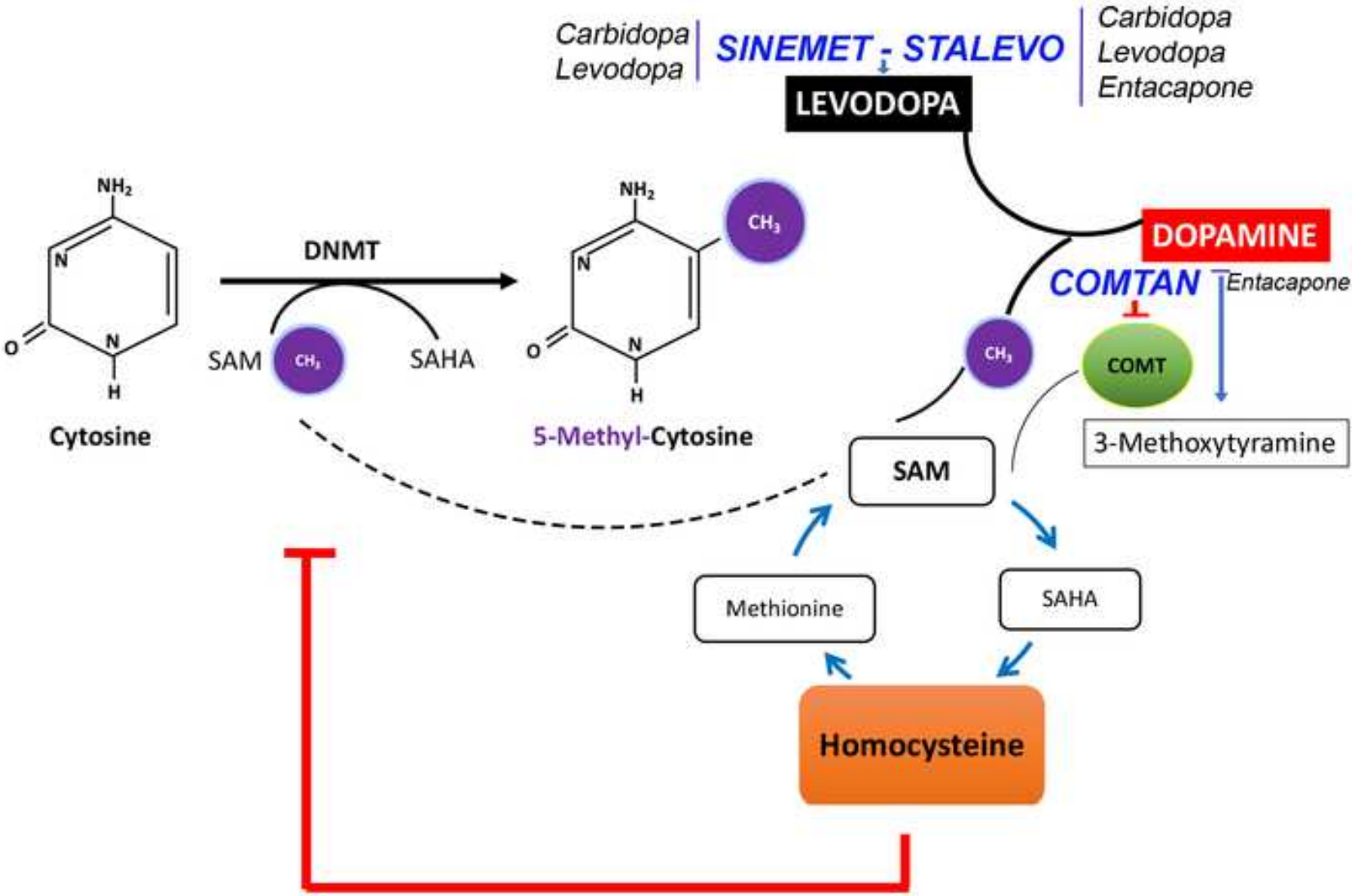
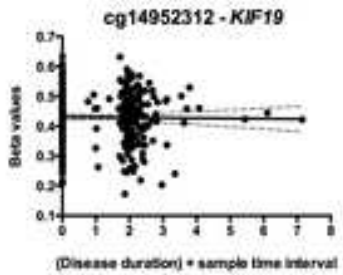
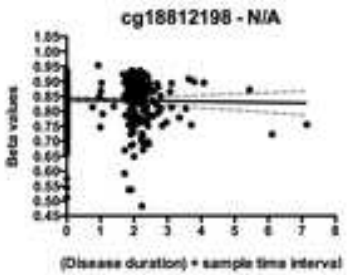

Carbidopa
Levodopa
SINEMET - STALEVO
Carbidopa
Levodopa
Entacapone
LEVODOPA
DOPAMINE
COMTAN
Entacapone
COMT
3-Methoxytyramine
DNMT
SAM
SAHA
Cytosine
5-Methyl-Cytosine
SAM
Methionine
SAHA
Homocysteine

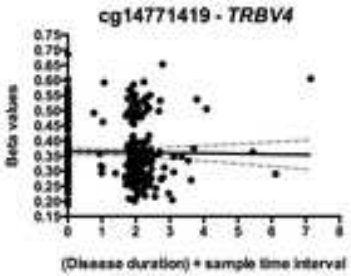
Figure 7. Cross-talk between one-carbon metabolism and Levodopa catabolism. Diagram depicting the effects of dopamine metabolism on the one-carbon metabolic pathway. Conversion of levodopa to dopamine requires the breakage of S-adenosyl methionine as methyl group donor which is the same source of methyl groups used by DNA methyl transferases (DNMTs) to methylation cytosine residues. In addition, homocysteine produced by the conversion levodopa-dopamine is an inhibitor of DNMTs activity and may alter DNA methylation. COMT further mediates the conversion of dopamine into O-methyl derivatives (3-methoxytyramine) by consuming methyl groups from SAM. The potential interaction with PD medication are indicated, with Sinemet and Stalevo increasing levodopa levels and COMTAN inhibiting COMT activity.

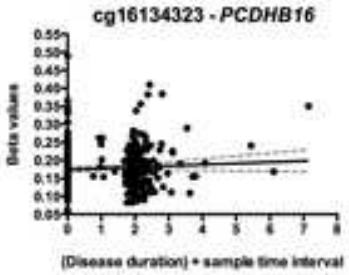
Probe ID	Chr	Position	Gene ID	Change rate PD Med	Pvalue PD Med	Change rate Controls	Pvalue Control
Methylation decreasing over time							
cg13643943	5	138629019	MATR3	-0.0632	7.25×10^{-8}	-0.00417	2.26×10^{-1}
cg08988821	7	74075293	GTF2I	-0.0396	7.24×10^{-9}	-0.00107	6.71×10^{-1}
cg16655626	19	45886078		-0.0373	2.20×10^{-8}	0.00152	4.38×10^{-1}
cg14386312	19	58740861	ZNF544	-0.0346	2.92×10^{-8}	-0.00204	2.90×10^{-1}
cg15765398	21	46409994	LINC00163	-0.0344	1.11×10^{-11}	-0.00114	4.16×10^{-1}
cg15987655	5	139196457		-0.0329	1.98×10^{-8}	-0.00285	1.42×10^{-1}
cg04843111	1	156617074	BCAN	-0.0327	8.26×10^{-8}	-0.00088	6.64×10^{-1}
cg14240188	1	11621054		-0.0281	1.77×10^{-8}	0.00038	8.03×10^{-1}
cg04551581	18	44226771		-0.0273	5.48×10^{-8}	-0.00150	3.08×10^{-1}
cg15298173	15	29903900	RP11-300A12	-0.0262	7.18×10^{-8}	-0.00297	5.48×10^{-2}
cg15202607	5	149520120	PDGFRB	-0.0234	4.44×10^{-8}	-0.00151	1.89×10^{-1}
cg10429957	5	141245719	PCDH1	-0.0197	8.39×10^{-8}	-0.00139	2.38×10^{-1}
Methylation increasing over time							
cg26126295	4	119095240		0.0862	8.67×10^{-8}	0.00030	9.44×10^{-1}
cg25979148	4	871629		0.0252	4.47×10^{-9}	-0.00001	9.95×10^{-1}
cg10224806	12	131188144	RIMBP2	0.0221	9.04×10^{-8}	0.00156	1.16×10^{-1}
cg18629514	7	5388895		0.0178	8.06×10^{-8}	0.00025	7.94×10^{-1}
cg03501539	11	114760598		0.0166	6.57×10^{-8}	0.00014	8.87×10^{-1}
cg15008072	17	7440546	Y_RNA	0.0166	7.51×10^{-8}	0.00113	2.20×10^{-1}
cg05933219	2	240099221	HDAC4	0.0162	1.37×10^{-9}	-0.00058	4.87×10^{-1}
cg18763089	1	1683738	NADK	0.0155	1.10×10^{-8}	0.00067	3.78×10^{-1}
cg05120150	14	105912598	MTA1	0.0143	3.80×10^{-8}	0.00006	9.39×10^{-1}
cg09593391	8	144737908	ZNF623	0.0139	8.19×10^{-8}	-0.00017	8.51×10^{-1}
cg10401356	8	140712424		0.0138	3.24×10^{-9}	0.00037	6.62×10^{-1}
cg22511774	13	114746915		0.0137	3.27×10^{-8}	0.00036	6.52×10^{-1}

A. Control subjects

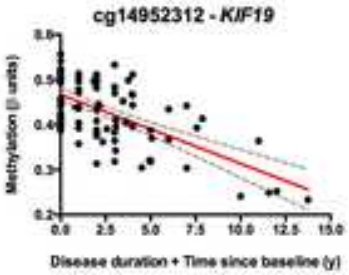


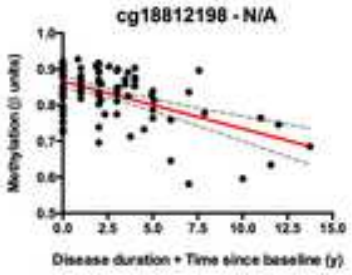


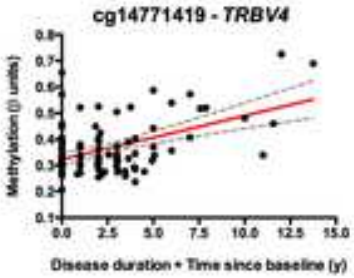


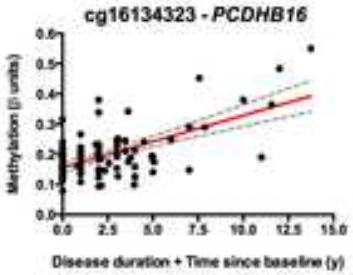


B. PD cases NO medication

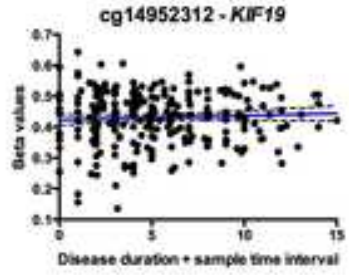


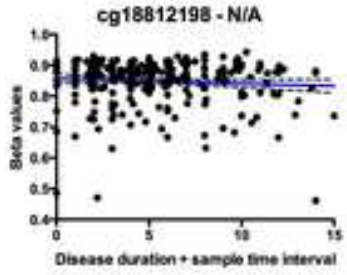


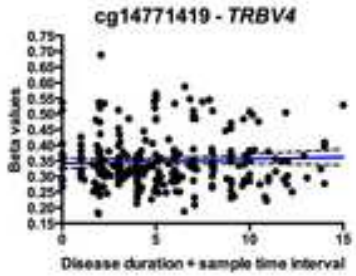




C. PD cases on medication







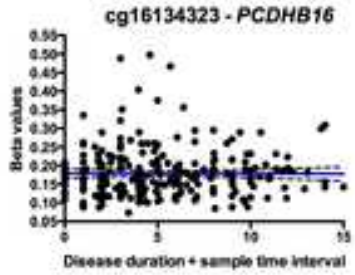


Table 9. Top Longitudinal changes in methylation in PD cases receiving medication. Name indicates probe designation at Illumina EPIC human methylation array; Chr is chromosome location of the CG; position refers to Genome Reference Consortium Human Build 37 (GRCh37) hg19. Rate of change indicates percentage methylation change/year.

Figure 8. Longitudinal changes in the blood methylome associated with PD progression are compensated by PD medication. Representative plots showing regression of methylation (individual β-values) as a function of disease duration and time between baseline and follow-up samples. Comparison of top changing probes between controls (A); PD cases NOT receiving medication (B) and PD cases on medication. Coefficient of change as determined by mixed linear models (Tables 5 and 6) and corresponding p-values are indicated.

Longitudinal methylation changes in the absence of medication associated solely with PD progression.

One of the most valuable analyses in our study is the possibility of exploring the blood methylome in PD patients that did not yet receive PD medication. Despite the limitation of a reduced cohort size (N=27), this comparison may unveil epigenomic alterations only due to PD systemic pathophysiology. We thus analyzed changes in methylation over time in PD NOT medicated cases (according to the criteria for PD medication described in the previous section). We identified 24 probes showing significant changes of methylation over time at a $p<1.0E-7$. Notably these sites showed the largest size effects compared to the other analysis groups, with methylation changes ranging from a 1.5 % reduction to a 1.7 % increase in methylation/year (Table 10 and Fig. 8). Longitudinal methylation in these sites did not change significantly in control subjects and none of the sites showed association with age. Genes showing differential methylation in this group (Table 10) include those involved in cytoskeletal functioning, like microtubules (KIF19, KIF22, TAOK2, MAPK13, CEP70) and actin (SWAP70). In addition, genes involved in immune response are also noted (RAIT1E and TRBV4). Interestingly, changes observed in this group are distinct from those identified in the PD medicated group above (Table 9). This supports the conclusion that dopamine replacement therapy indeed alters the methylome of PD patients (see Discussion).

Probe ID	Chr	Position	Gene ID	Change rate PD No Med	Pvalue PD Med	Change rate Controls	Pvalue Control
Methylation decreasing over time							
cg14952312	17	72350710	KIF19	-1.537	2.41×10^{-9}	-0.00156	6.41×10^{-1}
cg18812198	11	68715658		-1.324	7.66×10^{-8}	0.00114	6.95×10^{-1}
cg15969149	10	77352287		-1.054	1.61×10^{-8}	0.00087	7.34×10^{-1}
cg15531997	10	34999169		-0.981	5.39×10^{-12}	0.00023	8.71×10^{-1}
cg19174044	20	18446362	DZANK1	-0.914	3.71×10^{-8}	0.00091	6.32×10^{-1}
cg13937758	6	150217448	RAET1E	-0.711	3.55×10^{-8}	-0.00227	1.83×10^{-1}
Methylation increasing over time							
cg14771419	7	142012988	TRBV4	1.713	1.03×10^{-9}	-0.00114	7.31×10^{-1}
cg16134323	5	140562034	PCDHB16	1.507	7.50×10^{-10}	-0.00042	8.29×10^{-1}
cg13252209	16	51603716		1.279	1.78×10^{-8}	-0.00192	5.70×10^{-1}
cg03700944	6	46703468	PLA2G7	1.093	8.90×10^{-9}	-0.00262	2.94×10^{-1}
cg06688960	20	1504932		0.988	3.41×10^{-8}	-0.00053	8.49×10^{-1}
cg26421310	1	25257058	RUNX3	0.093	2.43×10^{-8}	-0.00055	3.31×10^{-3}
cg27380788	16	4526765	NMRAL1	0.073	9.24×10^{-8}	-0.00046	2.78×10^{-3}
cg11107196	16	71918265	ZNF821	0.065	5.36×10^{-9}	-0.00019	1.37×10^{-1}
cg09585751	6	24646282	KIAA0319	0.056	9.21×10^{-8}	-0.00021	4.51×10^{-2}
cg14934866	16	29985156	TAOK2	0.053	7.30×10^{-8}	-0.00023	7.87×10^{-2}
cg21645604	6	36098567	MAPK13	0.043	3.31×10^{-8}	-0.00014	5.70×10^{-2}
cg07038191	16	29801882	KIF22	0.041	5.83×10^{-8}	-0.00028	3.27×10^{-4}
cg23993697	11	9685562	SWAP70	0.040	7.05×10^{-8}	-0.00012	2.22×10^{-1}
cg17129217	9	91933717	SECISBP2	0.039	6.52×10^{-8}	-0.00022	5.45×10^{-3}
cg15193793	14	90422250	EFCAB11	0.038	8.37×10^{-8}	-0.00012	2.83×10^{-1}
cg19116545	3	138313166	CEP70	0.037	3.02×10^{-8}	-0.00022	8.33×10^{-3}
cg18104674	18	60190218	ZCCHC2	0.036	5.58×10^{-8}	-0.00011	9.65×10^{-2}
cg20390702	12	8850385	RIMKLB	0.036	6.54×10^{-8}	-0.00019	1.85×10^{-2}

Table 10. Longitudinal changes in methylation in PD cases NOT receiving medication. Name indicates probe designation at Illumina EPIC human methylation array; chr is chromosome location of the CG; position refers to Genome Reference Consortium Human Build 37 (GRCh37) hg19; N/A indicates no annotated gene associated with the probe. Rate of change

DISCUSSION

Specific blood cell types are altered in Parkinson's disease.

Blood methylation profiles have been used to estimate the relative abundances of specific blood cell types in Parkinson's disease(Horvath 2013). In the HBS cohort, we observed similar significant increases in granulocytes of PD patients (p=4.0E-6, as per t-test) and lower estimated B-cells (p=0.0019) and NKs (p=0.00055) in comparison to controls. These differences remained in the longitudinal samples (~ 2 years later) for granulocytes, which were higher in PD cases (p=0.0066) and natural killers, which were lower in PD (p=0.00065) in the follow-up visit. The finding of increased granulocytes in PD confirms prior reports. However, in contrast to these reports, we also observe significant and persistent reductions to NK cells in PD, whereas no changes were previously observed for this blood cell type. While the mechanistic explanations for these changes will require additional study, it is tempting to speculate that changes in these immune related cells could reflect ongoing inflammatory responses that occur in the brain during PD.

Specific whole blood methylation changes correlate with Parkinson's disease.

Epigenetic changes are emerging as contributing factors to PD and other neurodegenerative diseases. To identify a comprehensive set of differentially methylated sites associated with PD, we performed a case versus control comparison on extremely well characterized and phenotyped patient samples acquired by the Harvard Biomarkers Study (HBS). Leveraging the full cohort size of 792 samples, a comparison of PD versus control groups identified 7 DMPs significantly associated with PD, many of them tagging genetic loci with interesting functional consequences with respect to pathways and cellular functions previously implicated in PD.

Analysis of neighbor sites with consistent alterations in methylation or differentially methylated regions (DMRs), more likely to affect gene expression, identified altered methylation in 13 CpG sites (p=1.22E-22) at the cytochrome P450 2E1 (*CYP2E1*) locus in whole blood of PD patients. Interestingly, *CYP2E1*, which has been reported to be hypomethylated in PD brain samples (Kaut et al. 2012), encodes a member of the cytochrome P450 mixed-function oxidase

system responsible for metabolizing environmental toxins. Since environmental factors and external toxins contribute to PD vulnerability(Helley et al. 2017), altered regulation of *CYP2E1* may signal the response to environmental cues that associate or contribute to PD onset. We note, however, that previous reports describe hypomethylation of the *CYP2E1* locus in PD brain. Here we find consistent increased methylation in several CpG sites across the 5' region of the gene in blood (Fig. 6A and Table 7). Reasons for this tissue specific difference remain to be clarified, but identification of this locus across independent studies and separate tissues suggests an important role in PD. Another notable DMR includes 1-acylglycerol-3-phosphate O-acyltransferase 1 (AGPAT1: 35 CGs hypermethylated; p=8.45E-25). AGPAT1 catalyzes the conversion of lysophosphatidic acid (LPA) into phosphatidic acid (PA). LPA is required for dopamine neuron development and, in the 6-OHDA model of PD, reduced LPA has implicated in dopamine neuron degeneration through activity at the LPA1 receptor(Yang et al. 2015). In addition, among the DMRs identified was one in the lymphocyte antigen 6 family member G5C (*LY6G5C*: 5 CGs hypermethylated; p=3.92E-9), recently identified as an epigenetic marker of schizophrenia(Wockner et al. 2015); and clustering in the major histocompatibility complex (MHC) class III region of chromosome 6, encoding genes with critical functions in immunity. Genetic variation of MHC associates with sporadic PD (Hamza et al. 2010), suggesting that altered immunity in PD may be trackable via epigenetic changes in blood.

Longitudinal analyses reveal methylome changes over time in PD patients.

Analysis of samples from patients at baseline and follow-up, which averaged 2.2 years later, provided the opportunity to interrogate the extent to which methylation profiles of PD patients change over time. During the sampling interval, PD cases showed modest but highly significant disease progression as assessed by both HY ($p<0.0001$, as per paired t-test) and UPDRS scores ($p<0.0001$; see also Table 5).

Single locus hits showing longitudinal methylation changes in PD with the highest statistical significance (Table 8) highlight pathways of potential importance to nervous system function. These include 4-aminobutyrate aminotransferase (ABAT: p=2.0E-8), which is responsible for breaking down the neurotransmitter γ-aminobutyric acid (GABA, and Enhancer Of

mRNA Decapping 3 (EDC3: p=6.2E-9); EDC3 promotes removal of the 5' cap structure of mRNAs during their degradation. Recessive mutations in EDC3 cause intellectual disability(Ahmed et al. 2015), suggesting an impairment of neurological functions linking mRNA decapping to normal cognition. In addition, multiple cytoskeletal and extracellular matrix associated proteins showed differential methylation over time in PD, including Keratin Associated Protein 5-5 (KRTAP5: p=8.04E-9); Engulfment And Cell Motility 1 (ELMO1: p=7.33E-8); Brevican (BCAN: p=1.0E-9); and Sidekick Cell Adhesion Molecule 1 (SDK1: p=5.43E-8). Lastly, Nuclear Receptor Corepressor 2 (NCOR2: p=8.4E-8) and Histone Deacetylase 4 (HDAC4: p=5.61E-8) were hypo- and hyper-methylated, respectively, in PD patients at the follow-up visit relative to baseline, suggesting potential alterations to the epigenetic machinery itself. This could be due either to PD progression or, potentially, to the administration of PD medications that alter one-carbon metabolism pathways (see below).

Dopamine replacement therapies alter the methylome in PD patients.

Dopamine replacement therapies directly impact one carbon metabolism, consuming methyl groups that are required for DNA methylation and increasing homocysteine levels, which directly inhibits the activity of DNA-methyl transferases (**Figure 7**). Determining the extent to which common PD medications impact methylation profiles is, therefore, critical for studying differential methylation in the context of PD. The HBS contains a number of PD patients that were not taking any medication associated with PD treatment at time of enrollment. Samples from 25 of these patients were included in our study. In addition, we were particularly interested in exploring a potential interaction between levodopa and COMT-inhibitors on DNA methylation profiles, therefore we grouped all PD cases that ever received Sinemet, Stalevo and Comtan as "PD Medication".

Sinemet contains two active ingredients: levodopa and carbidopa. Levodopa is a dopamine agonist. Carbidopa is a peripheral inhibitor of DOPA-decarboxylase, preventing conversion of Levodopa to dopamine outside the brain. *Stalevo* is a combination of carbidopa, levodopa, and entacapone- an inhibitor of catechol-O-methyltransferase (COMT). *Comtan* is the commercial denomination for entacapone.

The conversion of levodopa to dopamine consumes methyl-groups from the donor (SAM) used by DNMTs to methylate DNA. COMT further metabolizes dopamine into 3-methoxytyramine at expense of SAM. Therefore the 3 drugs may alter the availability of SAM and the levels of homocysteine, which in turn inhibits DNMTs.

The majority of patients were taking Sinemet during the study (n=104 at baseline and n=140 at follow-up); from which 13 cases at baseline and 15 at follow-up were also on Comtan. Comtan was never administered alone or in combination with Stalevo. Use of Stalevo was reported for 21 cases at baseline and 26 at follow-up. The cases using Sinemet + Comtan may have equivalent effects to those on Stalevo. All patients taking any combination or formulation of these drugs are receiving levodopa/carbidopa.

While we recognize that the 25 non-medicated PD cases is a relatively small sample size, preliminary analysis of longitudinal changes in methylation in patients naïve to these drugs identified 217 DMPs. These DMPs may denote true PD associations separate from any modifications that may be imposed by dopamine replacement medications. To the best of our knowledge, this is also the first genome-wide analysis of blood methylation in medication naïve PD cases. Importantly, the comparison between medicated versus not medicated patients showed larger changes in methylation longitudinally, suggesting that medication modifies the epigenome. Therefore, as the methylome of naïve cases better resembles the epigenetic profiles of the disease, they provide additional value for early diagnosis. Study of prodromal cases before phenoconversion and of a properly powered naïve cohort will be fundamental for the appraisal of methylation as an early PD classifier tool.

One important observation from the longitudinal analysis is the overall damping effect that PD medication has on the blood methylome, illustrated in the group comparison of top DMPs changing longitudinally in **Figure 8.** On one hand, as our cohort is dominated by PD cases receiving medication, the changes in methylation are smaller when using the entire cohort for the analysis. On the other hand, the larger changes in methylation in PD NOT medicated cases supports the utility of blood methylation as an early disease classifier and a potential indicator of disease progression and, eventually, drug efficacy/history. Additional studies including more

subjects are needed to corroborate these findings, despite the scarcity of samples from non-medicated/naïve patients.

One of the limitations of our analysis is the number of time points investigated. Not only having only two data sets per subject restricts the analysis, but also the average longitudinal time lapse of ~2 years may not be sufficient to detect significant changes in methylation beyond the high variability of the population. UPDRS total scores are estimated to increase by 4.7 points per year in PD patients not taking medication (Holden et al. 2018). In contrast, we observed only 3.05 points/year increase in UPRDS in cases, suggesting slow disease progression in our study cohort, an effect likely due to 79.9 % of PD cases under anti-parkinsonian treatment at the time of the study. While the correlates between clinical motor scales and molecular mechanisms like DNA methylation are not determined, the slow disease progression may result into smaller changes in methylation. Future studies including additional longitudinal points and spanning longer periods of time may detect additional epigenetic changes relevant to pathology.

In summary, we present evidence demonstrating that changes to the methylome in PD are detectable in blood; change over time; and in many cases reflect cellular processes implicated in ongoing neurodegeneration in the brains of PD patients. In particular, the longitudinal sampling of our study emphasizes that DNA methylation is dynamic in PD and that common PD medication, including levodopa formulations and COMT-inhibitors impact methylation. Taken in all, these studies support the potential of blood DNA methylation as an epigenetic biomarker of disease, although additional profiling of large longitudinal cohorts is needed to complete the characterization of DNA methylation changes during the onset and progression of Parkinson's disease.

CHAPTER 3

DNA METHYLATION AND EXPRESSION PROFILES OF WHOLE BLOOD IN PARKINSON'S AND ALZHEIMER'S DISEASES

INTRODUCTION

Alzheimer's disease (AD) and Parkinson's disease (PD) are the two most common age-related neurodegenerative diseases(Mayeux and Stern 2012; Tysnes and Storstein 2017), collectively afflicted over 6.5 million people in the U.S. alone.

PD characteristically manifests as overt motor defects following the destruction of dopaminergic neurons in the substantia nigra and is pathologically associated with α-synuclein protein aggregation into intracellular cytoplasmic inclusions, termed Lewy bodies. This brain pathology is proposed to result from initial insults via olfaction or in the gut, with retrograde trafficking into affected brain regions (Hawkes et al. 2009; Planken et al. 2017). Neurodegeneration occurs primarily in dopaminergic neurons of the substantia nigra, but Lewy body pathology occurs in limbic and cortical areas as PD progresses(Mattila et al. 2000).

Currently, PD diagnosis is predominantly based on clinical manifestations of the disease, namely by the findings of tremor, rigidity, and bradykinesia. Extrastriatal, non-motor symptoms of PD, including cognitive problems, apathy, depression, anxiety, hallucinations, and psychosis as well as sleep disorders, fatigue, autonomic dysfunction, sensory problems, and pain can begin years before diagnosis, accompany the course of disease progression, and are major factors in reduced quality of life(Barone et al. 2009). Clinical diagnostic accuracy of 53% in these early stages of PD, is unacceptably low(Mehta and Adler 2016). Ongoing prevention therapy research currently underway(Olanow and Schapira 2013), would be greatly facilitated by increased diagnostic accuracy of early stage PD. Indeed, currently on average, over half of all dopaminergic neurons in the substantia nigra are already lost by the time of accurate clinical diagnosis(Delenclos et al. 2016), making prevention approaches problematic. A combination of multiple biomarker approaches as a diagnostic panel could accelerate improvements in early diagnostic accuracy. This will be important in pushing the point at which diagnosis, or high risk prediction, can be made to an even earlier time point in pre-motor prodromal stages. To this end,

multiple pre-motor biomarkers are actively being investigated for their potential to identify early-stage PD or patients at risk for developing PD (Haas et al. 2012), including clinical measures (rapid eye movement behavior disorder (RBD), olfactory deficits, mood disorders); molecular measures (α-syn in cerebrospinal fluid and blood); and brain imaging.

Epigenetic factors, which can be modified by both environment and ongoing disease processes, are emerging as important components of neurodegenerative diseases, including PD(Pavlou and Outeiro 2017). The dynamic process of DNA methylation is one of the most commonly studied epigenetic regulators in human disease. Addition of a methyl group primarily occurs on cytosine bases that are next to guanines, referred to as CpG sites, and it occurs to a much lesser extent on cytosines next to other bases(Messerschmidt et al. 2014). Initial hypotheses regarding methylation function were mainly centered around the polarizing effects that it can confer on gene expression (e.g., imprinting). However, the effects of DNA methylation on gene expression are far more nuanced and can be heavily context-dependent.

Hypomethylation of the α-synuclein gene (*SNCA*) promoter region has been reported in substantia nigra of PD patients in some studies(Jowaed et al. 2010; Matsumoto et al. 2010), yet not replicated in others (Richter et al. 2012; Guhathakurta et al. 2017). Epigenomic changes associated with other genes, including hypomethylation of *NPAS2*(Lin et al. 2012) and *CYP2E1*(Kaut et al. 2012); and hypermethylation of *PGC1-α*(Su et al. 2015) and the H1 haplotype of Tau (*MAPT*)(Coupland et al. 2014), have also been implicated in PD. More recently, epigenome wide association studies (EWAS) have identified concordant dysregulation of the methylome in brain and blood of PD patients(Masliah et al. 2013). Two additional EWAS have identified sites of altered DNA methylation throughout the genome in blood samples of PD patients versus controls(Moore et al. 2014; Chuang et al. 2017), suggesting that PD is associated with altered methylation that can be detected in peripheral blood samples and raising the possibility that a peripheral epigenetic biomarker of PD could be possible.

The primary functional consequence of DNA methylation is a resulting effect on regulation of gene expression. Thus altered DNA methylation would be predicted to alter mRNA expression levels as well. In the current work, we profile the methylome of whole blood from PD,

AD, and healthy age-matched controls using the Illumina Infinium 450K Human Methylation beadchip and also perform mRNA-seq on the same blood samples to identify DNA methylation loci that are associated with differential gene expression in each disease. Findings provide information about coordinate epigenome-wide regulation of gene expression across PD and AD.

METHODS

Patient demographics

All studies were approved by Institutional Review Boards in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants. Study participants completed a comprehensive neurologic examination and have received a clinical diagnosis of AD (n=10), PD (n=15), or Control (CON; n=15). PD, AD, and CON patients were 27%, 50%, and 53% female, respectively. Average ages per group at sample collection were 68 (CON) and 72 years old (PD and AD). PD individuals were cognitively normal based on the Kokmen Short Test of Mental Status with a score of 37 or greater.

Blood collection, DNA & RNA extraction

Peripheral blood was collected from patients using standard venipuncture techniques into PreAnalytiX PAXgene blood DNA and RNA tubes. Vacutainers were inverted several times and stored at -80°C. Samples were isolated from peripheral leukocytes according to manufacturer instructions with the PAXgene DNA or RNA isolation kits (Qiagen). Isolated samples were stored at -20°C.

Methylation array procedure

We examined methylation using Illumina Infinium HumanMethylation450K Beadchips per the manufacturer protocol. Genomic DNA samples were first bisulfite converted using the Zymo EZ DNA methylation kit, per manufacturer instructions with the alternative incubation conditions specified in the protocol for compatibility with the Illumina Infinium Methylation Assays. Bisulfite converted DNA was amplified, fragmented, precipitated, resuspended, and hybridized to beadchips following the manual protocol. Fluorescent staining was automated with Illumina's Tecan system. The chips were coated and vacuum-dried for preservation before scanning to

retrieve fluorescence intensity data, representing methylated or unmethylated positions, with Illumina's iScan.

Methylation analysis

Raw intensity files were processed with the R package, minfi. The data were normalized using the "functional normalization" method, sex chromosomes and positions at known SNPs were excluded from further analysis. Figure 9 shows sample clustering following removal of these positions. Biologically typical differential methylation between males and females is significant enough to skew any other analyses that should ideally be performed without that bias as much as possible. Single position differential methylation and region-wise differential methylation were measured with the R packages, limma(Ritchie et al. 2015) and DMRcate(Peters et al. 2015b), respectively. The model design input to limma was ~0+Group+Slide, using AD vs. controls and PD vs. controls as the desired contrasts, and beadchips ("Slide") as an additional variable. The linear model was fit to the data and the empirical Bayes function was applied to calculate differentially methylated probes. Peripheral blood cell-type counts were estimated using the function available in the minfi package. Only one patient sample was excluded from analysis as a PCA outlier for a disproportionately high b-cell balance. Cell-type composition PCAs are shown in figure 10.

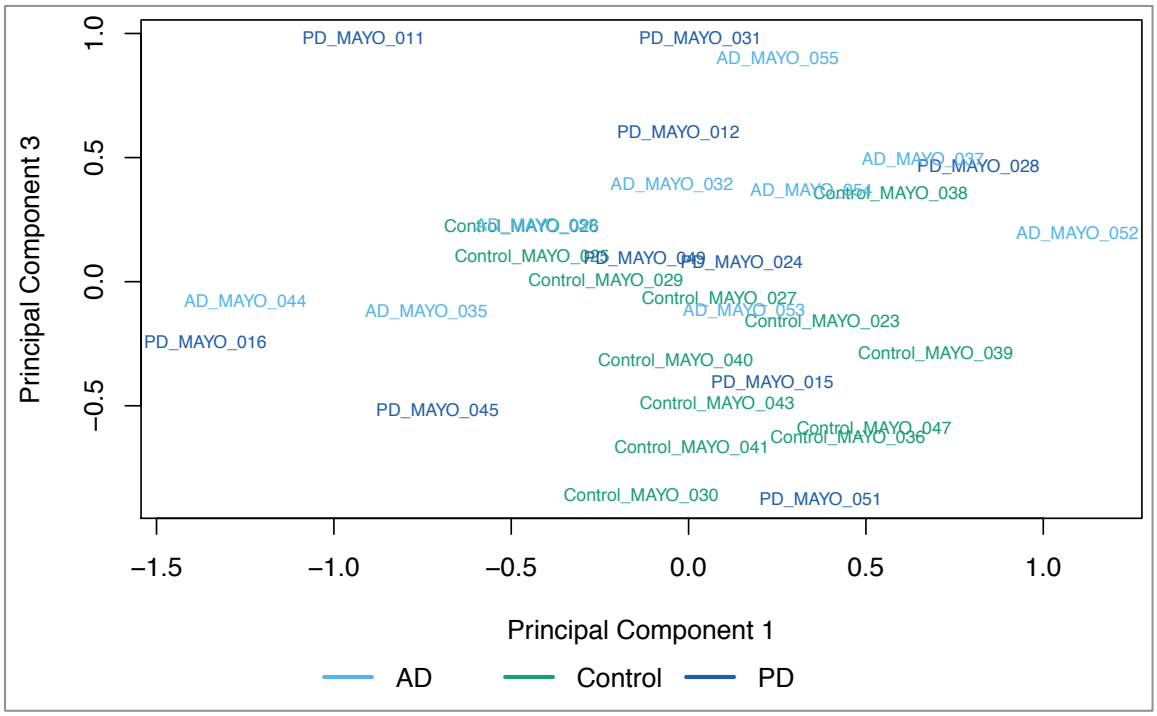
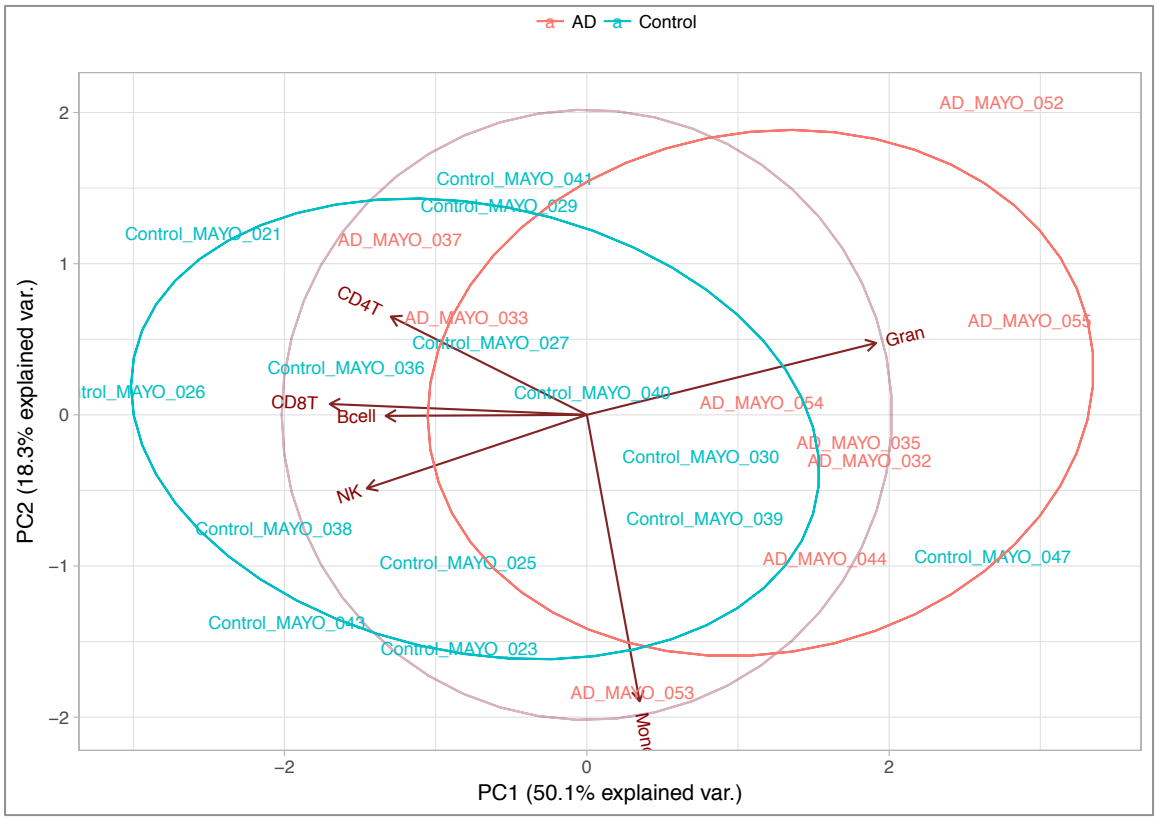


Figure 9. Group clustering following removal of sex chromosomes and positions at known SNPs.


AD
Control
AD_MAYO_052
Control_MAYO_041
Control_MAYO_029
Control_MAYO_021
AD_MAYO_037
CD4T
AD_MAYO_033
Control_MAYO_027
Gran
AD_MAYO_055
Control_MAYO_036
trol_MAYO_026
CD8T
Bcell
Control_MAYO_040
AD_MAYO_054
AD_MAYO_035
AD_MAYO_032
Control_MAYO_030
NK
Control_MAYO_039
Control_MAYO_038
Control_MAYO_025
AD_MAYO_044
Control_MAYO_047
Control_MAYO_043
Control_MAYO_023
AD_MAYO_053
PC2 (18.3% explained var.)
PC1 (50.1% explained var.)

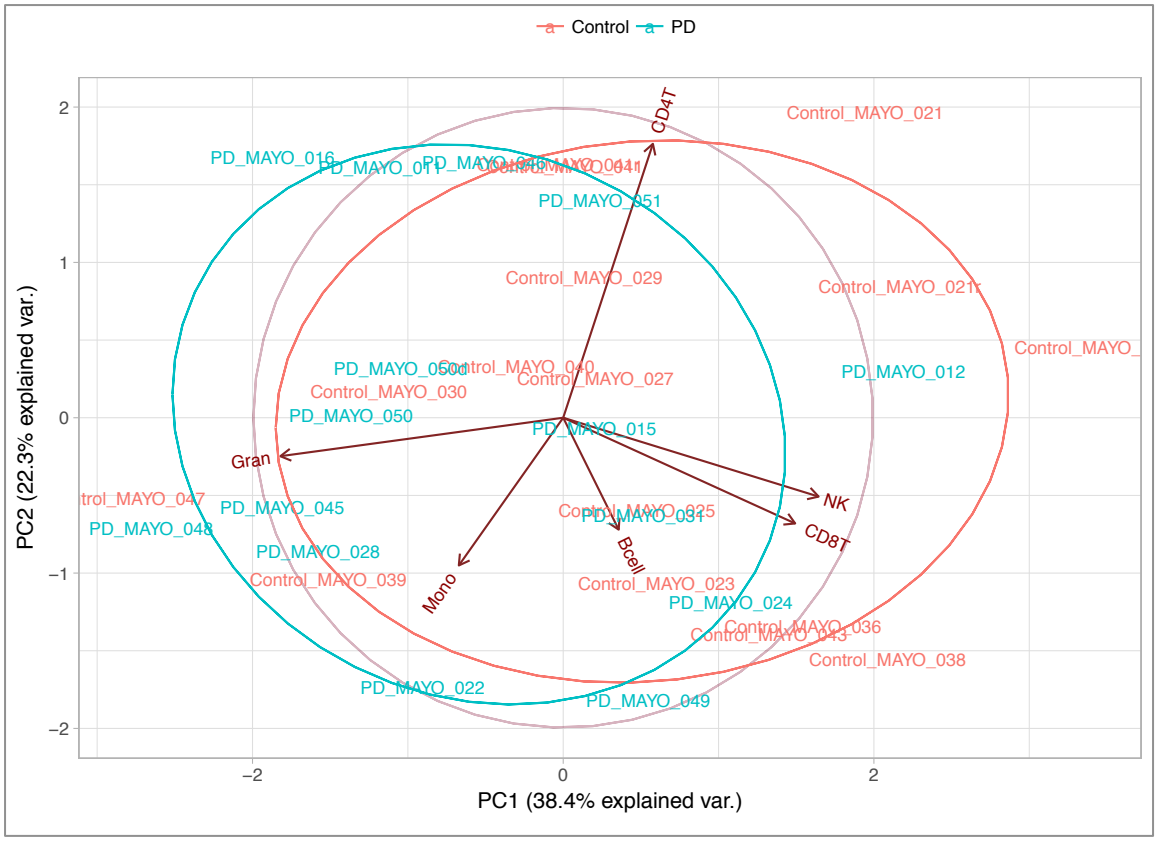

Control
PD
CD4T
Control_MAYO_021
PD_MAYO_016
PD_MAYO_051
Control_MAYO_029
Control_MAYO_021
PD_MAYO_012
Control_MAYO_040
Control_MAYO_027
Control_MAYO_030
PD_MAYO_050
PD_MAYO_015
Gran
NK
CD8T
trol_MAYO_047
PD_MAYO_045
PD_MAYO_048
PD_MAYO_028
Bcell
Mono
Control_MAYO_039
Control_MAYO_023
PD_MAYO_024
Control_MAYO_036
Control_MAYO_038
PD_MAYO_022
PD_MAYO_049
PC2 (22.3% explained var.)
PC1 (38.4% explained var.)

Figure 10. Cell-type composition principle component analysis (PCA) plots for AD and PD.

RNA sequencing (RNAseq) & Expression analysis

mRNA was sequenced using the Illumina TruSeq RNA Library Prep kit on the HiSeq 2000 platform following the manufacturer's protocol. Paired end RNA read transcripts were quantified with Kallisto(Bray et al. 2016), a program based on pseudoalignment. Kallisto's differential expression analysis companion is Sleuth, which outputs beta values as the measure for differential expression. This is similar to fold change, and different than the beta values reported for differential methylation output.

Integrated analysis: Combined p-values

Significance values for genes nearest to differentially methylated positions (DMPs) and differentially expressed (DE) genes were combined to reveal genes that were most significantly different in case vs. control(Ozer and Sezerman 2015). This method is widely used and was first described in *Statistical Methods for Research Workers* by R.A. Fisher(Fisher 1930) as a method for obtaining a single test of significance for a set of aggregated data based on the product of the individually observed probabilities. The basis is that the probability of rejecting the global null hypothesis is related to the intersection of each individual test probability(Winkler 2012).

Integrated analysis: meQTL

We identified associations between differentially methylated loci and changes in gene expression at those sites using the MatrixEQTL(Shabalin 2012) package in R. Methylation quantitative trait loci (meQTL) were measured in both *cis* (<10,000 bp from a gene) and *trans* (>10,000 bp from a gene).

GO enrichment

Enrichment analysis was performed with the R package topGO(Alexa and Rahnenfuhrer 2018), by checking significant DMPs and meQTLs against our unfiltered gene expression list as the gene universe. Simple gene ontology analysis for prominent pathways in gene lists was done with GeneMANIA(Montojo et al. 2010), a web-based software.

Functional network analysis

A supervised functional network analysis was performed using an algorithm called Functional Epigenetic Modules (FEM), developed by Jiao et al., which identifies gene modules from protein-protein interaction networks that exhibit dysregulation in both DNA methylation and gene expression(Jiao et al. 2014).

RESULTS

Differentially Methylated Positions in Alzheimer's Disease patients.

There were a variety of biological pathways related to differentially methylated positions in AD, including cyclase and lyase activity, many metabolic and ion homeostasis processes. Aberrant adenylate cyclase activity was previously reported in AD(Kelly 2018). Table 11 shows the top differentially methylated positions that are unique to the AD dataset.

Syntrophin Gamma 1 (*SNTG1*; ch.8.1157478F, logFC = 2.56, p = 3.55 x 10^{-3}) and Activating Transcription Factor 6 (*ATF6*; cg22522357, logFC = 2.03, p = 5.07 x 10^{-3}) are genes of interest near DMPs within the top hypermethylated positions. SNTG1, primarily expressed in the brain, is an adapter protein involved in subcellular organization of proteins and neurotrophic signaling. A variant of *SNTG1* has been associated with a delay of AD onset in patients who also carry the *PSEN1* E280A mutation(Vélez et al. 2018). ATF6 is a transcription factor that activates genes in the unfolded protein response pathway when there is endoplasm reticulum (ER) stress. Xu et al. found that overexpression of mesencephalic astrocyte-derived neurotrophic factor (MANF) reduced ATF6 levels, along with other ER stress molecules, in an $A\beta_{42}$ toxicity reduction study(Xu et al. 2019).

Hypomethylated positions are near genes related to several cell transport processes including ATP synthesis proton transport, transmembrane activity, and purine-containing biosynthesis. Calcitonin Related Polypeptide Alpha (*CALCA*; cg24136292, logFC = -3.02, p = 5.44 x 10^{-4}) and Ataxin 1 (*ATXN1*; cg19300401, logFC = -2.78, p =1.27 x 10^{-3}) are genes of interest near hypomethylated positions. CALCA is a regulatory neuropeptide involved in activation of PKA which has been shown to upregulate α-subunit of the acetylcholine receptor (AChR) in the

brain, which is a target of AD therapy research(Vallés et al. 2014; Singh et al. 2017). *ATXN1* affects amyloid precursor protein (APP) protein levels and may be a risk factor for late onset AD; additionally, mutations in *ATXN1* lead to spinocerebellar ataxia type 1 (SCA1)(Zhang et al. 2010).

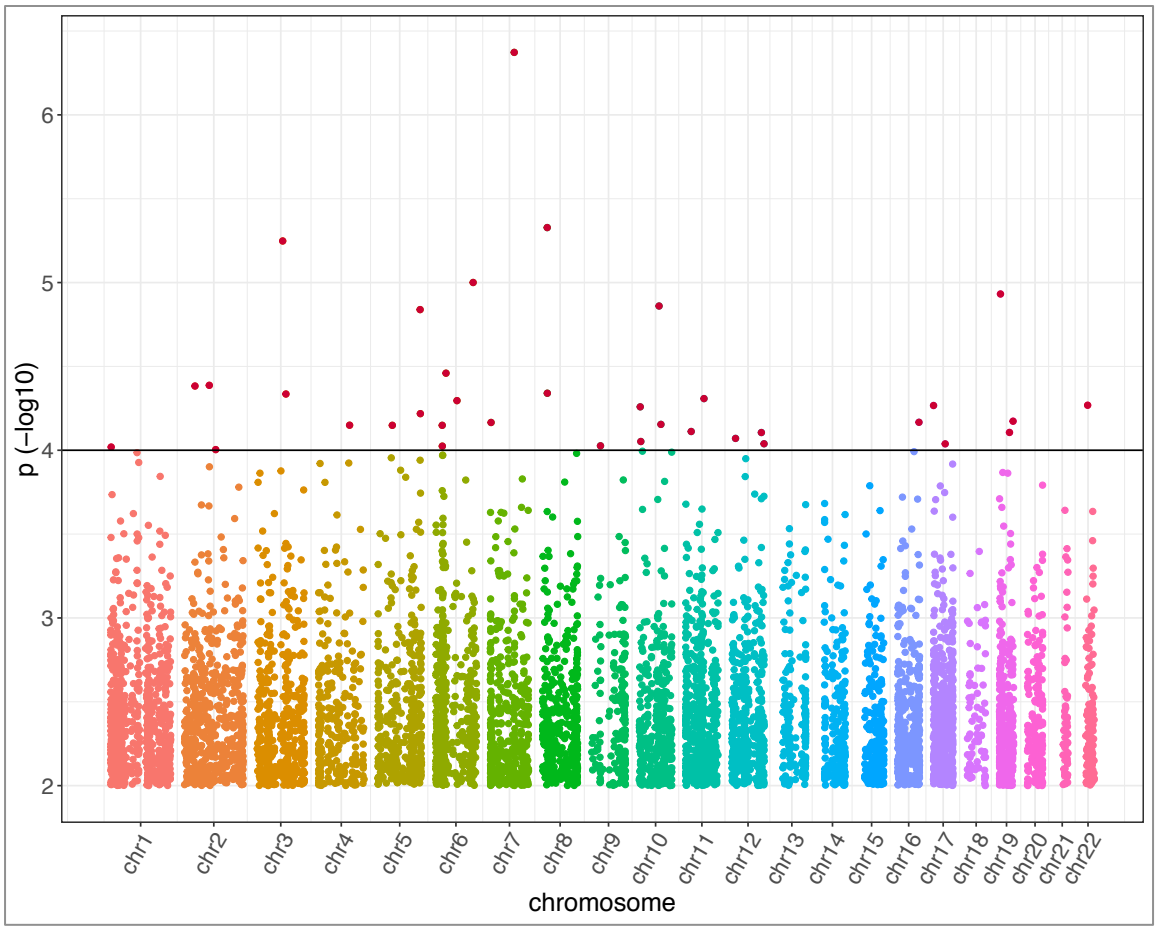


Figure 11. AD Manhattan plot of global differential methylation significance.

Probe	Position	Fold Change (log2)	P-Value	Target Gene Name	Position in Target Gene	Nearest Gene
			Hypermethylated			
cg07056794	chr9:139318309	3.18	6.23×10^{-3}			INPP5E
cg18339359	chr8:23423757	2.97	7.55×10^{-3}	SLC25A37	Body	ENTPD4
cg14655569	chr9:95473718	2.86	1.61×10^{-3}	BICD2	3'UTR	BICD2
cg22627029	chr6:32520615	2.81	6.77×10^{-3}	HLA-DRB6	Body	HLA-DQA1
cg14926231	chr2:906044	2.61	4.16×10^{-3}	C2orf90	TSS200	LINC01115
cg21223191	chr6:32583741	2.56	2.59×10^{-3}			HLA-DQA1
ch.8.1157478F	chr8:52618305	2.56	3.55×10^{-3}	PXDNL	Body	SNTG1
cg19353052	chr12:113516445	2.51	1.05×10^{-3}	DTX1	Body	DTX1
cg24307368	chr11:107582884	2.43	3.39×10^{-3}	SLN	TSS200	SLN
cg05161773	chr17:75378036	2.15	9.80×10^{-3}	SEPT9	5'UTR; Body	SEPT9
cg22522357	chr1:161872770	2.03	5.07×10^{-3}	ATF6	Body	OLFML2B
			Hypomethylated			
cg22777560	chr5:177612982	-3.44	8.97×10^{-3}	GMCL1L	Body	HNRNPAB
cg01778345	chr1:118427435	-3.39	6.13×10^{-3}	GDAP2	Body	GDAP2
cg02978201	chr16:11374865	-3.35	8.88×10^{-3}	PRM1	Body	PRM1
cg27187580	chr3:112927486	-3.28	6.18×10^{-4}			CFAP44
cg24136292	chr11:15180922	-3.02	5.44×10^{-4}	INSC	Body	CALCA
cg19300401	chr6:16962712	-2.78	1.27×10^{-3}			ATXN1
cg00256329	chr17:724374	-2.76	1.37×10^{-3}	NXN	Body	MRM3
cg22304519	chr2:227560785	-2.61	5.81×10^{-3}			IRS1
cg26813483	chr13:111980537	-2.54	3.02×10^{-3}	C13orf16	Body	TEX29
cg19475903	chr6:39271655	-2.53	1.62×10^{-3}	KCNK17	Body	KCNK17
cg10771931	chr19:34972145	-2.50	1.58×10^{-3}	WTIP	TSS1500	SCGB1B2P
cg18203203	chr12:113939171	-2.39	4.24×10^{-3}			SDSL
cg18096251	chr5:2205553	-2.36	2.52×10^{-3}			IRX4
cg15876198	chr14:95106927	-2.25	2.48×10^{-3}	SERPINA13	TSS200	SERPINA13P

Table 11. Top differentially methylated positions in AD.

Differentially Methylated Positions in Parkinson's Disease patients.

Pathways among differentially methylated positions in the PD cohort are primarily involved in antigen presentation and T-cell activation, and processes related to the endoplasmic reticulum. Differentially methylated positions are presented in table 12.

HCN2 (cg10662395, logFC = 2.11, p = 9.34 x 10^{-3}) and Trafficking Kinesin Protein 1 (*TRAK1*; cg24284460, logFC = 2.00, p = 5.16 x 10^{-4}) are genes of interest near hypermethylated positions. *HCN2* is one isoform of a hyperpolarization-activated cyclic nucleotide-gated (HCN) channel with a role in the transmission of neuropathic and inflammatory pain. HCN channels control oscillatory activity of dopaminergic neurons in the midbrain and their dysfunction has been proposed to play a role in PD pathogenesis(DiFrancesco and DiFrancesco 2015). TRAK1 is involved in endosome to lysosome trafficking and regulation of mitochondrial fusion. A mutation leading to truncated TRAK1 causes hypertonia in mice, which initiated further study of this gene in the context of neurological disorders, including PD; it is also implicated in several cancers and childhood absence epilepsy(Lee et al. 2018).

Cytotoxic T-Lymphocyte Associated Protein 4 (*CTLA4*; cg05092371, logFC = -2.96, p = 1.42 x 10^{-3}) and Epidermal Growth Factor Receptor Pathway Substrate 15 (*EPS15*; cg06751175, logFC = -2.05, p = 5.17 x 10^{-3}) are near two of the top hypomethylated positions. CTLA4 is an immunosuppressive modulator of T-cell activation. Mkhikian et al. studied the interplay between dysregulated loss of N-glycosylation and the subsequent loss of CTLA4 cell surface expression in concert with reduced T-cell activation thresholds in the pathology of multiple sclerosis(Mkhikian et al. 2011). CTLA4 was also shown to contribute to increased survival of neural cell porcine xenografts delivered to nonhuman primates in a study aimed at improving PD cell transplantation therapies(Aron Badin et al. 2016). In 2006, EPS15 was functionally linked to the PD susceptibility gene, *Parkin*, through demonstration that EPS15 is phosphorylated and ubiquitinated in response to exposure to epidermal growth factor (EGF), and that EPS15 ubiquitination is mediated by parkin, which also showed parkin's role in trafficking and PI3K-Akt signaling(Fallon et al. 2006).

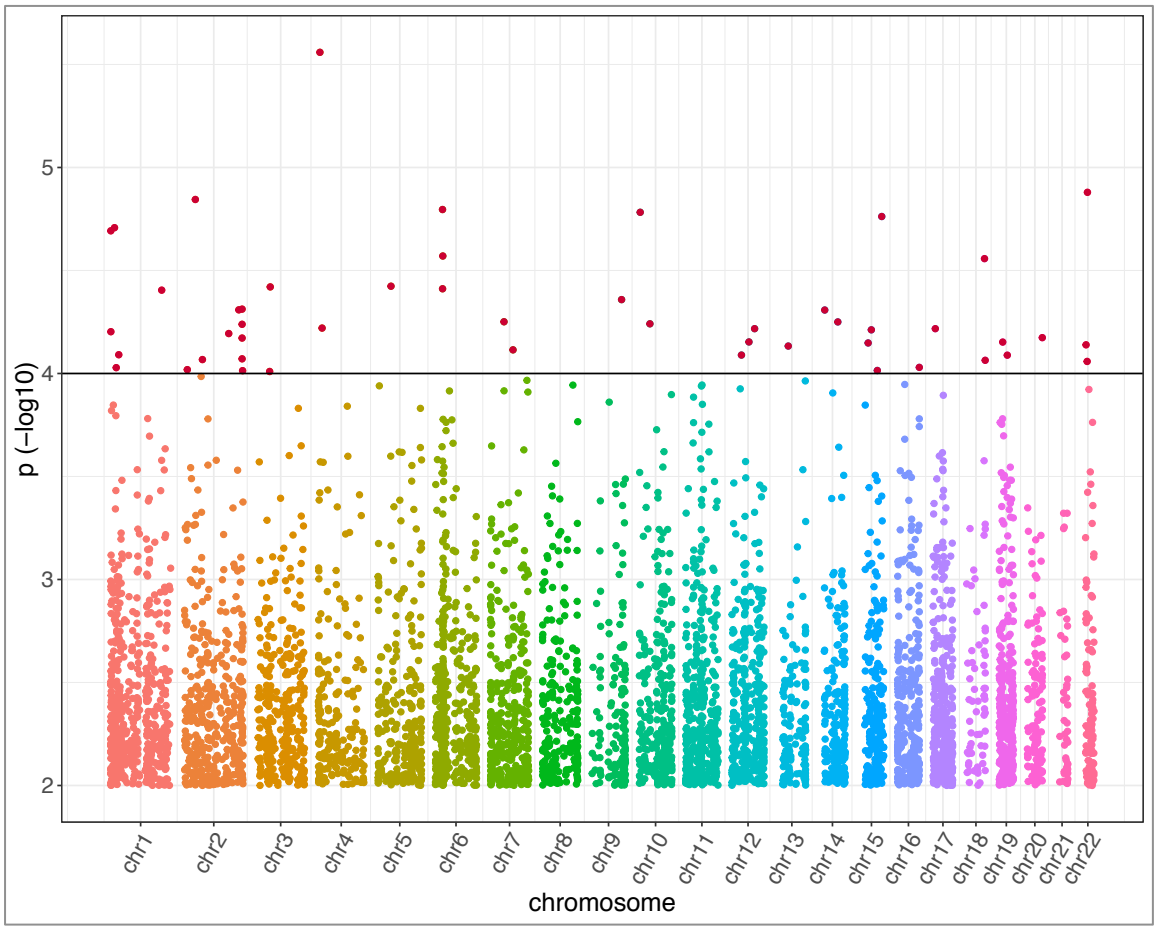


Figure 12. PD Manhattan plot of global differential methylation significance

Probe	Position	Fold Change (log2)	P-Value	Target Gene Name	Position in Target Gene	Nearest Gene
			Hypermethylated			
cg21211688	chr9:136403935	2.79	7.27×10^{-3}	ADAMTSL2	Body	ADAMTSL2
cg09727210	chr7:2962240	2.76	1.62×10^{-3}	CARD11	Body	CARD11
cg13353337	chr22:21336550	2.61	1.03×10^{-3}	LZTR1	TSS200	THAP7
cg00546757	chr5:170845058	2.60	8.29×10^{-3}			MIR3912
cg03329597	chr3:108125523	2.58	2.61×10^{-3}	MYH15	Body	HHLA2
cg26264314	chr19:56510965	2.57	3.59×10^{-3}	NLRP5	TSS200	NLRP13
cg10662395	chr19:597635	2.11	9.34×10^{-3}	HCN2	Body	POLRMT
cg02249577	chr10:52434778	2.09	6.31×10^{-3}			SGMS1
cg24284460	chr3:42139507	2.00	5.16×10^{-4}	TRAK1	Body	ULK4
cg24533989	chr4:42659536	1.95	7.27×10^{-3}	ATP8A1	TSS1500	ATP8A1
cg13452812	chr9:2234921	1.92	8.75×10^{-3}			VLDLR-AS1
cg24166916	chr5:78282669	1.88	9.59×10^{-3}	ARSB	TSS1500	ARSB
cg00616572	chr16:10772249	1.87	4.09×10^{-3}	TEKT5	Body	NUBP1
			Hypomethylated			
cg21046080	chr12:131506092	-4.85	3.63×10^{-4}	GPR133	Body	STX2
cg05092371	chr2:204731519	-2.96	1.42×10^{-3}	CTLA4	TSS1500	CTLA4
cg13604933	chr6:40145993	-2.93	1.52×10^{-3}			LINC00951
cg00035449	chr6:169539646	-2.87	6.51×10^{-3}			THBS2
cg14114910	chr9:124924045	-2.75	7.91×10^{-3}	MORN5	Body	MORN5
cg26296371	chr6:5442953	-2.72	8.88×10^{-3}	FARS2	Body	LYRM4
cg07304760	chr7:127514192	-2.72	2.29×10^{-3}	SND1	Body	LRRC4
cg13251750	chr14:94943234	-2.70	2.35×10^{-3}	SERPINA9	TSS1500	SERPINA9
cg13248811	chr4:24796803	-2.63	9.16×10^{-3}	SOD3	TSS1500	CCDC149
cg01324343	chr3:183735012	-2.55	8.36×10^{-4}	ABCC5	5'UTR	ABCC5
cg02375258	chr19:52074470	-2.42	3.83×10^{-3}	ZNF175	TSS200	ZNF175
cg00874817	chr16:46824055	-2.35	4.13×10^{-4}			C16orf87
cg24536782	chr8:216659	-2.30	1.91×10^{-3}			RPL23AP53
cg08762424	chr6:31275881	-2.29	1.52×10^{-3}			HLA-B
cg24167747	chr19:54614832	-2.23	3.16×10^{-3}	TFPT	Body	TFPT
cg24849373	chr8:143376371	-2.21	2.48×10^{-3}	TSNARE1	Body	LINC00051
cg07490070	chr2:97505464	-2.11	2.78×10^{-4}	ANKRD23	Body	CNNM3
cg03654560	chr5:178266175	-2.10	5.54×10^{-3}			AACSP1
cg15652532	chr2:30669759	-2.09	8.80×10^{-3}	LCLAT1	TSS1500	CAPN13
cg08270148	chr8:914818	-2.06	7.75×10^{-3}			ERICH1
cg06751175	chr1:52083188	-2.05	5.17×10^{-3}	OSBPL9	Body; 5'UTR	EPS15

Table 12. Top differentially methylated positions in PD.

Differentially Methylated Positions in Alzheimer's & Parkinson's Disease patients.

Among the overlapping DMPs, present in both PD and AD, two positions near potential genes of interest were hypermethylated. Arylsulfatase B (*ARSB*; cg19547330, AD logFC = 1.49, PD logFC = 1.54) is a lysosomal enzyme associated with a lysosomal storage disorder called MPS VI. Jansen *et al*. identified a loss of function variant in *ARSB* of four PD patients. In the same study, RNAi knockdown of the *ARSB Drosophila* homolog enhanced human a-synuclein transgene mediated neurodegeneration in the fly retina(Jansen et al. 2017). The second shared gene of interest, Zinc Finger DHHC-Type Containing 17 (*ZDHHC17;* cg26864661, AD logFC = 3.10, PD logFC = 2.69), codes for Huntingtin interacting protein 14 (HIP14) and is a highly conserved palmitoyl acyltransferase that acts upon many synaptic proteins(Sanders et al. 2016). It affects nicotinamide mononucleotide adenylyltransferase 2 (NMNAT2)-dependent axon survival in mouse primary neurons and Drosophila axon extension *in vivo*(Milde and Coleman 2014).

Probe	Nearest Gene	AD Fold Change	AD P-Value	PD Fold Change	PD P-Value
		Hypermethylated			
cg04663285	SIGIRR	1.39	8.67×10^{-3}	1.64	1.92×10^{-3}
cg14955916	ZMYND11	1.34	4.51×10^{-3}	1.38	2.93×10^{-3}
cg18493115	MOB2	1.75	1.65×10^{-3}	1.40	8.43×10^{-3}
cg19547330	ARSB	1.49	3.03×10^{-3}	1.54	1.82×10^{-3}
cg24906015	FANCL	2.25	3.35×10^{-3}	2.20	3.43×10^{-3}
cg26864661	ZDHHC17	3.10	1.51×10^{-3}	2.69	4.20×10^{-3}
cg27468880	GET4	2.49	6.47×10^{-3}	3.19	5.81×10^{-4}
		Hypomethylated			
cg00587941	RASA3	-2.61	3.73×10^{-3}	-2.59	3.34×10^{-3}
cg05270750	PRDM7	-1.06	1.70×10^{-3}	-1.29	1.81×10^{-4}
cg05394663	WNT3A	-1.13	2.88×10^{-3}	-1.31	6.06×10^{-4}
cg06295223	GAS8	-1.19	7.78×10^{-4}	-1.40	9.34×10^{-5}
cg10765922	HLA-DQA1	-1.09	8.64×10^{-3}	-1.11	6.82×10^{-3}
cg15729439	HPS5	-1.37	8.91×10^{-3}	-1.33	9.76×10^{-3}

cg16580391	HLA-DPB2	-1.04	2.56×10^{-3}	-1.12	1.08×10^{-3}
cg16611967	GAS8	-1.98	4.08×10^{-3}	-2.68	1.66×10^{-4}
cg17553353	GCSAM	-1.14	8.36×10^{-4}	-1.13	7.05×10^{-4}
cg22730047	CFAP126	-1.26	7.28×10^{-3}	-1.80	2.01×10^{-4}
cg27226147	ZNF184	-1.47	6.00×10^{-3}	-1.98	3.04×10^{-4}
cg27601198	ORC6	-1.16	9.82×10^{-3}	-1.68	3.05×10^{-4}

Table 13. Top differentially methylated positions in AD and PD.

Differentially Methylated Regions in Alzheimer's Disease patients.

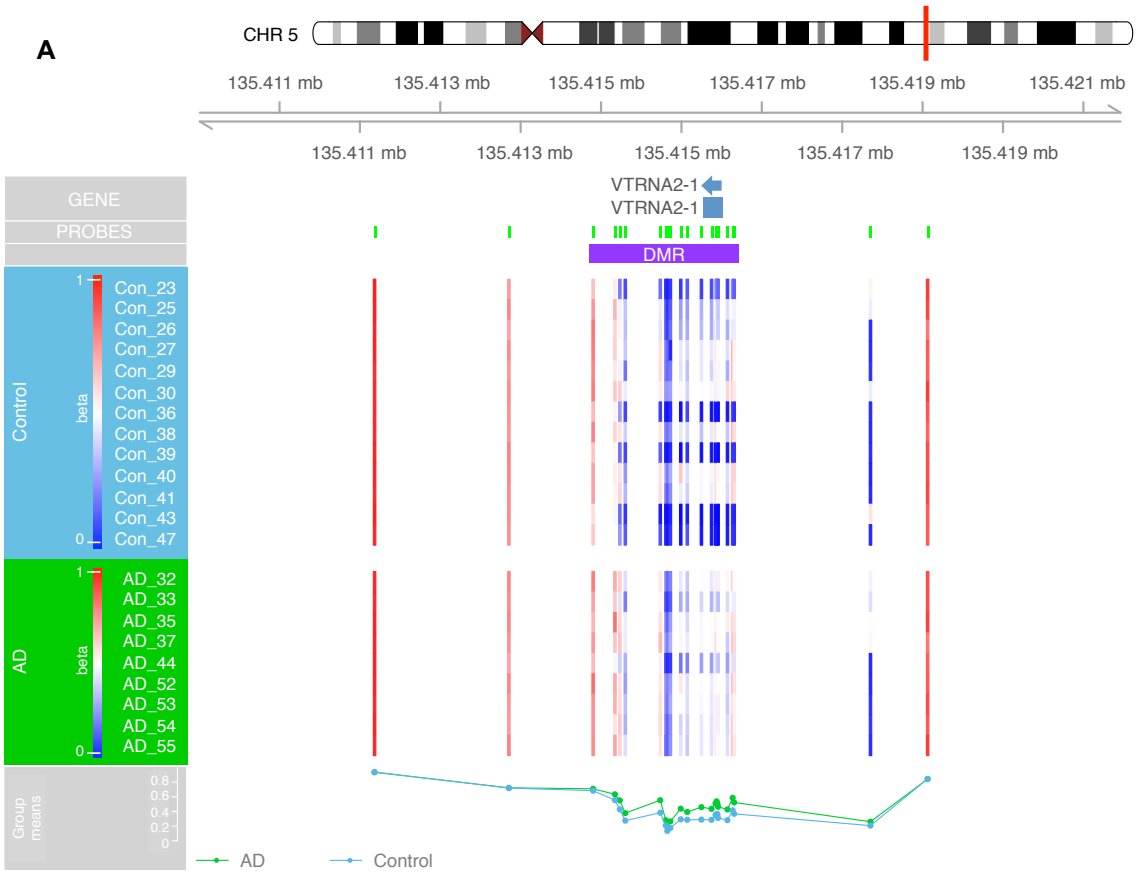
Pathways significantly involved with differentially methylated regions (DMRs), where at least two CpG positions are within 1000 nucleotides of each other, are associated with neurotrophin and tyrosine kinase (TRK) signaling, retinoic acid receptor signaling, vitamin metabolism, and transcription repression.

Five regions were hypermethylated and 14 were hypomethylated at FDR ≤ 0.001 (Table 14). Vault RNA 2-1 (*VTRNA2-1;* region start = chr5:135414858, probes in DMR = 20, fold change = 0.128, FDR = 6.88×10^{-6}) is near a hypermethylated region of interest. Its hypermethylation is associated with several cancers(Romanelli et al. 2014) and upregulation has been proposed as an early event in Parkinson's disease that contributes to neuronal dysfunction(Minones-Moyano et al. 2013).

Dual Specificity Phosphatase 22 (*DUSP22*; region start = chr6:291687, probes in DMR =11, fold change = -0.136, FDR = 2.86×10^{-4}) is near a hypomethylated region. The promoter of *DUSP22* was previously found to be hypermethylated in patients with AD, conversely to our finding, and the gene is involved in PKA-mediated tau phosphorylation(Sanchez-Mut et al. 2014).

Region Start	Width	Probes in DMR	FDR	Fold Change	Overlapping Promoters	Nearest Gene
Hypermethylated						
chr5:135414858	1756	20	6.88×10^{-6}	0.128		VTRNA2-1
chr6:28945182	326	8	2.33×10^{-4}	0.130	LINC01556-001	ZNF311
chr6:31148332	417	16	2.91×10^{-4}	0.095		PSORS1C3
chr12:132880790	2083	5	3.41×10^{-4}	0.087	CHFR-010	GALNT9
chr6:29648161	932	25	5.19×10^{-4}	0.140		ZFP57
Hypomethylated						
chr6:28874479	892	7	2.04×10^{-6}	-0.084		TRIM27
chr4:184908254	765	9	3.51×10^{-6}	-0.085		STOX2
chr10:1975000	632	4	1.04×10^{-4}	-0.083		LINC00700
chr4:76996414	277	2	1.64×10^{-4}	-0.081		ART3
chr6:291687	1645	11	2.86×10^{-4}	-0.136	DUSP22	DUSP22
chr19:5799340	128	3	5.10×10^{-4}	-0.089		DUS3L
chr1:202171635	1278	7	5.61×10^{-4}	-0.096		LGR6
chr7:1003645	1932	8	5.67×10^{-4}	-0.074		COX19;ADAP1
chr11:3175007	630	4	5.84×10^{-4}	-0.071		OSBPL5
chr1:23886730	1087	4	5.89×10^{-4}	-0.077		ID3
chr19:2546598	470	5	6.54×10^{-4}	-0.073		GNG7;GADD45B
chr14:102676728	567	4	7.90×10^{-4}	-0.071		WDR20;MOK
chr2:32490766	70	2	8.11×10^{-4}	-0.073		NLRC4
chr5:54275155	44	2	9.42×10^{-4}	-0.088		ESM1

Table 14. Top differentially methylated regions in AD.


A
CHR 5
135.411 mb
135.413 mb
135.415 mb
135.417 mb
135.419 mb
135.421 mb
135.411 mb
135.413 mb
135.415 mb
135.417 mb
135.419 mb
GENE
PROBES
VTRNA2-1
VTRNA2-1
DMR
Control
beta
1
0
Con_23
Con_25
Con_26
Con_27
Con_29
Con_30
Con_36
Con_38
Con_39
Con_40
Con_41
Con_43
Con_47
AD
beta
1
0
AD_32
AD_33
AD_35
AD_37
AD_44
AD_52
AD_53
AD_54
AD_55
Group means
0.8
0.6
0.4
0.2
0
AD
Control

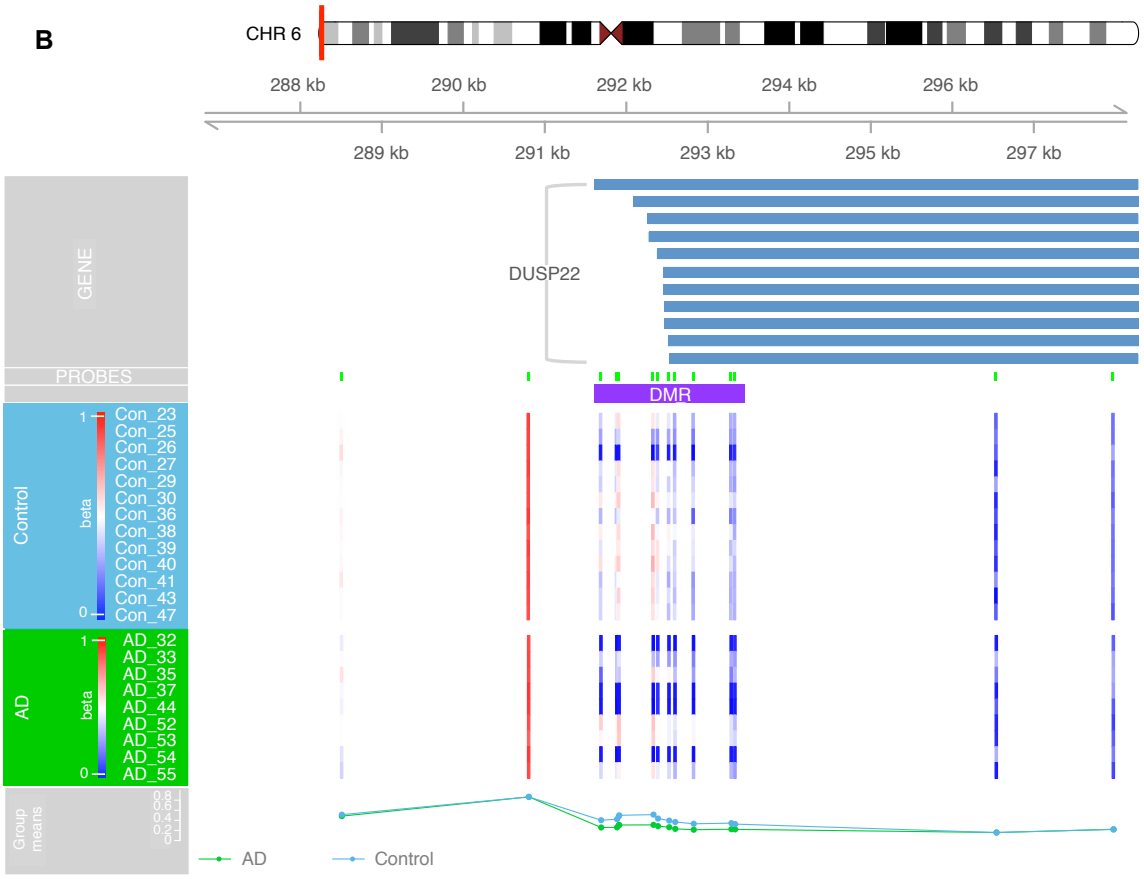


Figure 13. DMRs in AD. (A) *VTRNA2-1* and (B) *DUSP2*.

Differentially Methylated Regions in Parkinson's Disease patients.

Pathways related to genes near DMRs in PD are related to digestive tract development and several pathways are associated with transcription regulation, including JAK-STAT signaling. Implication of the enteric nervous system in the course, or even the etiology of PD makes this finding particularly interesting, suggesting that there are digestive tract genes being differentially regulated in the context of disease.

One region, near Growth Regulating Estrogen Receptor Binding 1 (*GREB1*; region start = chr2:11672761, probes in DMR = 2, fold change = 0.057, FDR = 9.77×10^{-4}), was hypermethylated at FDR ≥ 0.001. *GREB1* is a steroid hormone induced gene that is active in breast and prostate cancer cell proliferation, and plays a role in renal disease, but has not previously shown associations with neurological disorders(Hodgkinson and Vanderhyden 2014).

Six DMRs were hypomethylated, including one near SMAD Family Member 6 (*SMAD6*; region start = chr15:66947066, probes in DMR = 6, fold change = -0.117, FDR = 3.57×10^{-5}). SMAD6 is involved in the recruitment of A20 de-ubiquitinating enzyme to E3 ubiquitin ligases allowing for negative feedback following toll-like receptor (TLR) activation during innate immune responses. TLRs and the propagation of chronic pro-inflammatory responses are heavily and broadly implicated in neuropathogenesis(Kinsella et al. 2018).

Region Start	Width	Probes in DMR	FDR	Fold Change	Nearest Gene
Hypermethylated					
chr2:11672761	600	2	9.77×10^{-4}	0.057	GREB1
Hypomethylated					
chr12:116043958	75	3	8.80×10^{-6}	-0.126	MED13L;MIR620
chr15:66947066	552	6	3.57×10^{-5}	-0.117	SMAD6
chr6:166232637	505	5	3.65×10^{-4}	-0.114	LINC00473
chr11:65359521	989	6	3.65×10^{-4}	-0.107	EHBP1L1
chr11:18477153	983	10	5.99×10^{-4}	-0.090	LDHAL6A
chr8:215923	866	6	9.27×10^{-4}	-0.165	ZNF596

Table 15. Top differentially methylated regions in PD.

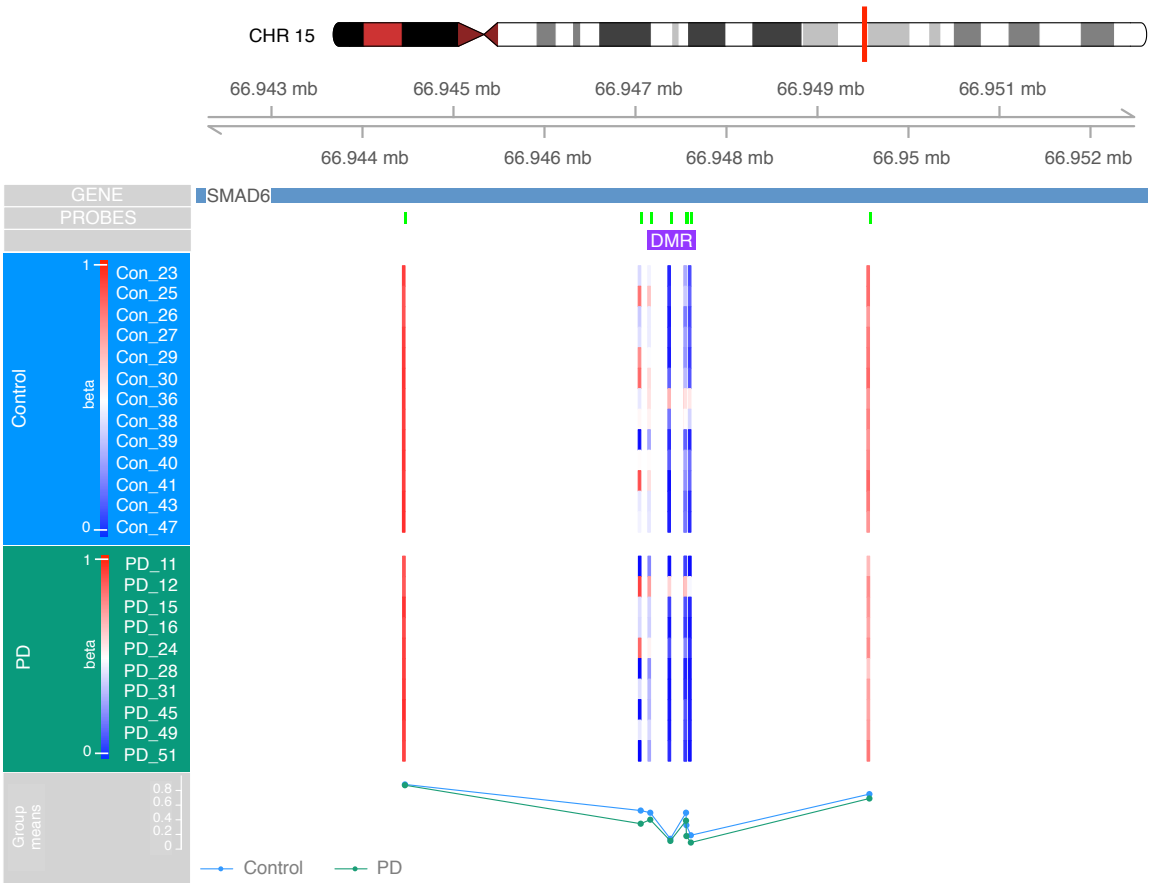


Figure 14. *SMAD6* DMR in PD.

Differentially Methylated Regions in Alzheimer's & Parkinson's Disease patients.

Alkaline phosphatase (*ALPL*; region start = chr1:21900987, probes in DMR = 3, fold change = -0.06, AD FDR = 6.95 x 10^{-5}, PD FDR = 1.43 x 10^{-5}) is near a hypomethylated DMR in AD and PD. Increased ALPL activity has been associated with central nervous system injury and the neurotoxic effect of extracellular tau protein in AD(Vardy et al. 2012). Serpin Family B Member 9 (*SERPIN9B*; region start = chr6:2891973, probes in DMR = 5, AD fold change = 0.07, PD fold change = 0.05) is hypermethylated and may contribute to autoinflammatory disease through incomplete caspase-1 inhibition resulting in uncontrolled interleukin-1b inhibition(Burgh et al. 2016). One study found an association between increased expression of SERPIN9B in females with Alzheimer's disease to be associated with greater amyloidosis(Deming et al. 2018).

Region Start	Width	Probes in DMR	AD FDR	AD Fold Change	PD FDR	PD Fold Change	Overlapping Promoters	Nearest Gene
Hypomethylated								
chr1:21900987	274	3	6.95×10^{-5}	-0.06	1.43×10^{-5}	-0.06		ALPL
chr1:26880207	1122	7	2.93×10^{-4}	-0.03	7.18×10^{-3}	-0.03		RPS6KA1
chr1:38412519	300	6	4.07×10^{-3}	-0.02	8.10×10^{-3}	-0.02		INPP5B
chr10:4867773	1316	13	3.08×10^{-8}	-0.06	5.92×10^{-5}	-0.05		AKR1E2
chr11:64980819	778	7	2.92×10^{-10}	-0.05	3.01×10^{-3}	-0.03		SLC22A20P
chr11:67069814	2072	15	5.52×10^{-11}	-0.03	2.60×10^{-12}	-0.02		SSH3
chr11:89867385	592	10	3.64×10^{-4}	-0.04	4.16×10^{-3}	-0.04	RP11-81M19.3-001	NAALAD2
chr12:122018177	2837	21	1.09×10^{-10}	-0.03	2.49×10^{-3}	-0.02	BCL7A, RP11-87C12.5	KDM2B
chr16:90143751	1038	6	6.51×10^{-7}	-0.08	2.83×10^{-12}	-0.10	RNU6-355P-201	PRDM7
chr16:90172237	783	5	7.53×10^{-4}	-0.07	2.51×10^{-3}	-0.06	FAM157C-201	PRDM7
chr17:41923945	633	12	4.29×10^{-3}	-0.02	7.12×10^{-4}	-0.02		CD300LG
chr19:14049018	1076	7	1.14×10^{-4}	-0.03	1.83×10^{-4}	-0.03		PODNL1
chr2:242808427	481	5	3.85×10^{-3}	-0.07	1.95×10^{-4}	-0.08		RTP5
chr20:1927546	366	4	3.62×10^{-3}	-0.02	1.88×10^{-3}	-0.02		SIRPA
chr20:20036653	498	6	1.65×10^{-3}	-0.05	2.24×10^{-4}	-0.05	CRNKL1	CRNKL1
chr22:23522307	1830	12	2.10×10^{-4}	-0.02	8.55×10^{-3}	-0.01		BCR
chr22:45071340	1826	11	9.30×10^{-6}	-0.03	1.09×10^{-3}	-0.02		PRR5
chr3:111717534	1032	12	1.40×10^{-5}	-0.05	6.14×10^{-3}	-0.04		TAGLN3
chr4:74864165	1338	9	1.62×10^{-5}	-0.03	8.09×10^{-6}	-0.03		CXCL5
chr6:30623775	995	15	8.47×10^{-8}	-0.03	2.91×10^{-3}	-0.02		DHX16
chr6:44186914	761	12	9.02×10^{-9}	-0.01	2.08×10^{-9}	-0.01		SLC29A1
chr7:27260102	484	6	1.49×10^{-4}	-0.03	3.88×10^{-7}	-0.03		HOTTIP
Hypermethylated								
chr10:459697	272	3	3.94×10^{-3}	0.08	5.38×10^{-3}	0.08		DIP2C
chr11:44626708	43	2	2.92×10^{-4}	0.03	3.20×10^{-3}	0.03		CD82
chr16:87734816	909	7	2.46×10^{-8}	0.03	1.12×10^{-3}	0.02		KLHDC4
chr6:2891973	180	5	1.07×10^{-6}	0.07	8.95×10^{-3}	0.05		SERPINB9

Table 16. Top differentially methylated regions in AD and PD.

Differential Expression in Alzheimer's Disease.

Genes differentially expressed in AD are largely involved with protein translation, driven by underexpressed genes. Cytokine production is upregulated.

CCS (beta = 2.00, p = 1.32×10^{-3}) and Integrin Subunit Beta 2 (*ITGB2*; beta = 0.94, p = 1.69×10^{-3}) were overexpressed in our results. Copper chaperone for super oxide dismutase (CCS) binds to the intracellular domain of Beta-Secretase 1 (BACE1) and to X11a, both of which

interact with APP. CCS homozygous knockout in mouse neurons resulted in greater levels of AB production along with increased processing of APP at the BACE1 site(Gray et al. 2010). ITGB2 has been previously identified to be highly associated with AD onset(Zhang et al. 2013) and another RNAseq study assigned this gene as part of a module of therapeutic candidates for cognitive resilience in early stages of AD using a pre-symptomatic AD mouse model(Neuner et al. 2019).

Gene	Beta	P-Value	Transcript
		Overexpressed	
PRAM1	1.02	1.18×10^{-3}	ENST00000600262.1
CD2BP2	1.79	1.32×10^{-3}	ENST00000564525.1
CCS	2.00	1.32×10^{-3}	ENST00000310190.8
AMPD2	1.42	1.61×10^{-3}	ENST00000528270.5
ITGB2	0.94	1.69×10^{-3}	ENST00000302347.9
KLHL21	1.31	1.74×10^{-3}	ENST00000496707.5
IQGAP1	0.99	1.80×10^{-3}	ENST00000268182.9
PQLC3	2.28	1.82×10^{-3}	ENST00000441908.6
SF3A1	1.99	1.99×10^{-3}	ENST00000444440.1
ACTG1	2.66	2.07×10^{-3}	ENST00000573283.5
		Underexpressed	
C16orf74	-1.07	1.03×10^{-3}	ENST00000602675.5
ODF3B	-2.01	1.07×10^{-3}	ENST00000403326.5
ZNF592	-2.48	1.27×10^{-3}	ENST00000299927.4
PTRHD1	-1.23	1.30×10^{-3}	ENST00000467797.1
SNHG8	-0.83	1.42×10^{-3}	ENST00000602414.5
QARS	-1.63	1.64×10^{-3}	ENST00000637281.1
RPL23A	-2.01	1.69×10^{-3}	ENST00000496182.5
MED25	-2.99	1.80×10^{-3}	ENST00000618715.4
PSME2	-1.41	1.80×10^{-3}	ENST00000560410.5
PIEZO1	-1.64	2.05×10^{-3}	ENST00000472168.1

Table 17. Differentially expressed genes in AD.

Differential Expression in Parkinson's Disease.

Genes that are differentially expressed are involved in innate immune responses, primarily indicated by genes involved in type I interferon signaling, which is driven by underexpressed genes. Overexpressed genes are involved in RAS signaling, implicating cell growth and division.

Serine/Arginine Repetitive Matrix 2 (*SRRM2;* beta = -1.97, p = 2.24 x 10^{-2}) was underexpressed in our results and was previously reported to be alternatively spliced in PD resulting in lowered expression of the long isoform and higher expression of the short(Fu et al. 2013). SRRM1 was alternatively spliced in our previous RNA-seq PD brain study(Henderson-Smith et al. 2016)

Gene	Beta	P-Value	Transcript
Overexpressed			
MORC3	2.27	1.01 x 10^{-2}	ENST00000400485.5
GOLM1	2.20	1.01 x 10^{-2}	ENST00000388712.7
NCOA1	2.14	1.12 x 10^{-2}	ENST00000348332.7
SLC43A3	1.93	1.14 x 10^{-2}	ENST00000395124.5
CANX	2.61	1.24 x 10^{-2}	ENST00000638706.1
PPP1R18	2.03	1.49 x 10^{-2}	ENST00000615527.1
VCPKMT	1.81	1.66 x 10^{-2}	ENST00000491402.5
PPP2R5D	2.03	1.77 x 10^{-2}	ENST00000485511.5
MBNL1	1.90	1.89 x 10^{-2}	ENST00000497971.1
DYSF	2.22	2.19 x 10^{-2}	ENST00000258104.7
Underexpressed			
GYS1	-1.98	1.02 x 10^{-2}	ENST00000263276.6
HLA-B	-3.04	1.08 x 10^{-2}	ENST00000639001.1, ENST00000639564.1
FHL3	-2.04	1.19 x 10^{-2}	ENST00000485803.5
SPNS1	-2.06	1.48 x 10^{-2}	ENST00000323081.12
CCL5	-2.18	1.60 x 10^{-2}	ENST00000605140.5, ENST00000618639.4
ELF4	-2.21	1.77 x 10^{-2}	ENST00000615377.4
HLA-B	-2.41	1.83 x 10^{-2}	ENST00000462100.1
HLA-B	-2.02	1.91 x 10^{-2}	ENST00000494335.5
AP2M1	-1.96	2.07 x 10^{-2}	ENST00000461733.5
SRRM2	-1.97	2.24 x 10^{-2}	ENST00000573692.1

Table 18. Differentially expressed genes in PD.

Differential Expression in Alzheimer's & Parkinson's Disease patients.

Cathepsin D (CTSD) is a lysosomal protease that is overexpressed in PD and AD. It has previously been studied in the context of both diseases, as well as for its involvement in childhood-onset neurodegenerative disorders(Cermak et al. 2016; Kettwig et al. 2018). It was determined to be a candidate PD susceptibility gene by Robak et al. whose study highlighted the

connection between PD and lysosomal storage disorders, e.g., Gaucher Disease(Robak et al. 2017).

Sequestosome 1 (*SQSTM1*) encodes the p62 scaffold protein that binds autophagy specific markers, including ubiquitinated proteins, in addition to possessing functional domains involved in inflammation and oxidative stress(Ma et al. 2019). It is also involved in NF-κB activation(Sivandzade et al. 2019).

NUMB, Endocytic Adaptor Protein (NUMB) is known and studied for a variety of functions, including as an indirect suppressor of NOTCH signaling, an endocytic adapter in intracellular trafficking, lysosomal regulation, and autophagy modulation(Sun et al. 2017). It was also studied for its role in APP processing and γ-secretase activity related to Vitamin D receptors and other neuronal membrane proteins(Dursun and Gezen-Ak 2017).

Using a cell line with enhanced γ-secretase activity, Magold et al. found that uridine phosphorylase (*UPP1*) was the top overexpressed gene in their profiling experiment. UPP1 is an enzyme involved in processing uridine, which is crucial for membrane biosynthesis as well as DNA and RNA. Oral administration of uridine has been used to improve AD phenotypes(Magold et al. 2009).

Gene	AD Beta	AD P-Value	PD Beta	PD P-Value	Transcript
			Overexpressed		
YTHDF3	2.76	3.71×10^{-3}	3.36	1.27×10^{-4}	ENST00000539294.5
PPP1R18	3.13	2.47×10^{-3}	2.56	7.06×10^{-3}	ENST00000399199.7, ENST00000449705.6, ENST00000451544.6
PPP1R18	2.69	3.02×10^{-3}	2.28	6.44×10^{-3}	ENST00000274853.7, ENST00000437121.2, ENST00000443517.2
CTSD	2.55	4.24×10^{-3}	2.19	7.65×10^{-3}	ENST00000636571.1
NUMB	2.40	3.33×10^{-3}	2.05	6.44×10^{-3}	ENST00000359560.7
LMBRD1	2.12	3.35×10^{-3}	2.05	2.05×10^{-3}	ENST00000472827.1
APP	2.77	1.33×10^{-4}	1.90	4.46×10^{-3}	ENST00000346798.7
CHD4	2.19	2.19×10^{-3}	1.78	7.05×10^{-3}	ENST00000544040.5
SQSTM1	1.63	7.70×10^{-3}	1.65	3.50×10^{-3}	ENST00000626660.1
CD14	2.10	1.24×10^{-4}	1.60	1.48×10^{-3}	ENST00000498971.6

FBXO11	1.73	5.35×10^{-3}	1.57	6.18×10^{-3}	ENST00000403359.7
PISD	1.64	3.25×10^{-3}	1.36	7.90×10^{-3}	ENST00000491342.1
BAHD1	1.47	7.36×10^{-3}	1.35	7.76×10^{-3}	ENST00000416165.5
HERC1	1.05	9.26×10^{-3}	1.00	7.14×10^{-3}	ENST00000561348.1
CMTM1	1.27	1.77×10^{-3}	1.00	7.99×10^{-3}	ENST00000533953.5
SERPINA1	1.26	1.55×10^{-4}	0.99	1.20×10^{-3}	ENST00000557492.5, ENST00000629092.2
Underexpressed					
QARS	-3.29	5.85×10^{-5}	-2.53	7.96×10^{-4}	ENST00000414533.5
UPP1	-1.88	6.98×10^{-3}	-1.84	4.19×10^{-3}	ENST00000417464.6

Table 19. Differentially expressed genes in AD and PD.

Integrated Analysis

Differential methylation and differential expression datasets were queried with several integrated methods: the combined p-value approach, methylation quantitative loci analysis, and functional epigenetic module searching. We also looked for genes that were alternatively spliced from our integrated lists.

Since the effect of DNA methylation is highly context dependent, we did not limit our filtering to inverse relationships (e.g., high methylation and low expression). The combined p-value tables (20-22) show that many genes are positively associated and contextually varied, as expected. The DMP tables (10-12) show the CpG probe positions relative to the CpG island associated with that probe, the location of which can be near or far from the probe itself.

Integrated Analysis: Combined p-values for Alzheimer's Disease.

Immune and inflammatory responses are the top pathways in the combined p-value AD dataset. Genes near DMPs that were also differentially expressed were also enriched for transmembrane transport and transport to the ER. Several of the top genes that were differentially expressed with nearby differentially methylated CpG sites have been previously reported for their relationships to AD-associated processes.

Beta tubulin (TUBB) was overexpressed in our results and two DMPs were nearest to this gene, one hypermethylated and one hypomethylated.

Hypericin imparts anti-inflammatory effects and can inhibit transcription of pro-inflammatory cytokines induced by oligomeric Aβ42 in vitro and in vivo(Zhang et al. 2016). One of hypericin's targets is Myocardin Related Transcription Factor A (MKL1), which is a nuclear cofactor recruited by p65 of the NF-κB complex and it was hypermethylated and underexpressed in our results (methylation fold change = 0.83, expression beta = -1.90, $p_{combined} = 4.21 \times 10^{-4}$).

SQSTM1 is involved in autophagy and was introduced in the shared AD & PD DE hits section of our results. Two different hypomethylated DMPs were near an overexpressed transcript (methylation fold change = -0.49 & -0.44, expression beta = 1.63, $p_{combined} = 4.49 \times 10^{-5}$ & 6.8×10^{-5}). APP was also hypomethylated and overexpressed in our results (methylation fold change = -0.59, expression beta = 2.77, $p_{combined} = 6.85 \times 10^{-6}$).

Gene	Transcript	Probe	P-Val Combined	Methylation Fold Change	Expression Beta	Probe Position Relative to Island
TMEM176B	ENST00000429904.6	cg14909495	3.02×10^{-7}	-0.30	-3.06	OpenSea
DDX24	ENST00000553400.1 ,ENST00000629321.1	cg06992688	4.32×10^{-6}	-0.58	-1.70	N_Shore
NAPA	ENST00000594001.5	cg23314866	6.13×10^{-6}	-0.49	-1.95	N_Shore
AOAH	ENST00000491444.5	cg25983531	2.24×10^{-5}	-0.60	-1.49	OpenSea
MKRN1	ENST00000480552.5	cg19494591	3.71×10^{-5}	-0.33	-1.05	N_Shelf
TUBB	ENST00000383564.8	cg05978527	3.74×10^{-5}	0.25	4.33	S_Shore
NELFA	ENST00000467661.5	cg17584633	3.96×10^{-5}	-0.21	-1.85	N_Shelf
Hypermethylated & Underexpressed						
ATP5O	ENST00000290299.6	cg27037013	2.03×10^{-5}	0.60	-0.52	OpenSea
CD3D	ENST00000300692.8	cg24612198	4.22×10^{-5}	0.54	-0.61	OpenSea
CD8B	ENST00000393759.6	cg22999502	1.16×10^{-5}	0.44	-1.72	OpenSea
DTX1	ENST00000547974.5	cg19353052	1.08×10^{-4}	2.51	-0.62	S_Shore
ADPGK	ENST00000569693.5	cg03950246	4.16×10^{-4}	0.67	-1.48	N_Shore
MKL1	ENST00000620651.4	cg25276892	4.21×10^{-4}	0.83	-1.90	OpenSea
Hypomethylated & Overexpressed						
APP	ENST00000346798.7	cg14414154	6.85×10^{-6}	-0.59	2.77	N_Shelf
IER3	ENST00000383560.4	cg06126421	1.49×10^{-5}	-0.69	2.97	OpenSea
CD2BP2	ENST00000564525.1	cg05099952	2.14×10^{-5}	-0.49	1.79	S_Shore
IER3	ENST00000383560.4	cg14753356	2.49×10^{-5}	-0.58	2.97	OpenSea
TUBB	ENST00000383564.8	cg16137928	2.67×10^{-5}	-0.41	4.33	OpenSea
CDC42EP3	ENST00000457889.1	cg26458748	2.80×10^{-5}	-0.70	1.94	OpenSea
TUBB	ENST00000383564.8	cg00985729	3.12×10^{-5}	-0.44	4.33	S_Shore

SQSTM1	ENST00000626660.1	cg08299521	4.49×10^{-5}	-0.49	1.63	OpenSea
SQSTM1	ENST00000626660.1	cg01152073	6.80×10^{-5}	-0.44	1.63	N_Shore

Table 20. Differentially methylated and differentially expressed genes in AD.

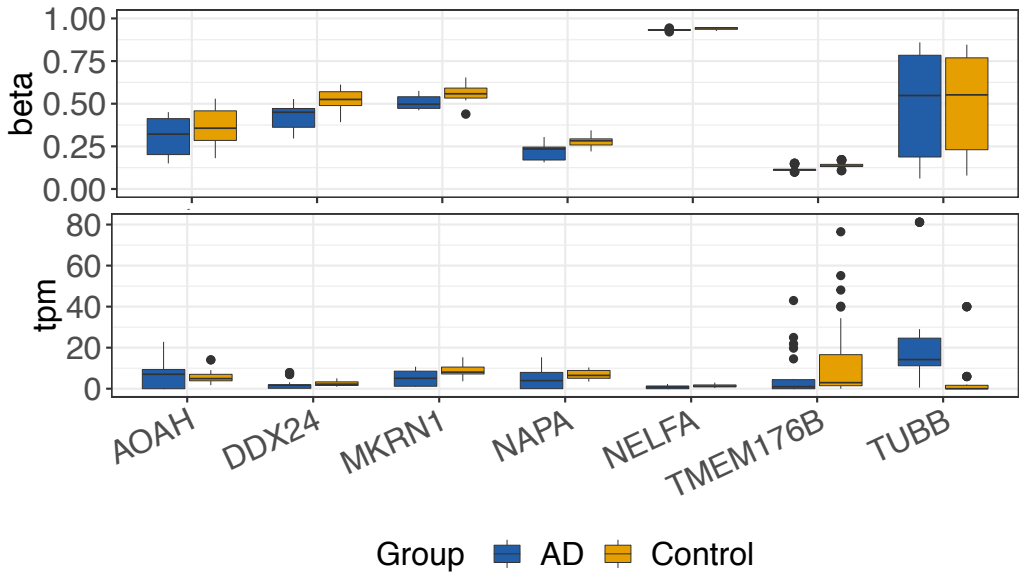


Figure 15. Top DE and DMP by combined p-value in AD.

Integrated Analysis: Combined p-values for Parkinson's Disease.

Combined p-value analysis of genes nearest DMPs that are also differentially expressed show that apoptosis, autophagy, and cell division are the main pathways active in both datasets. Genes near differentially methylated positions that were also differentially expressed were enriched for stress granule assembly, response to cold, and bone morphogenesis. When reducing the results to genes that are inversely related, the prominent pathways in the hypomethylated and overexpressed set are involved in RNA processing and splicing, which was made up of 110 genes.

ISG15 Ubiquitin-Like Modifier (*ISG15*) was hypermethylated and overexpressed in our results and it functions closely with parkin. It behaves similarly to ubiquitin in that it covalently

binds to target proteins (ISGylation) for biochemical modification and this mechanism is associated with JNK and NF-κB signal transduction and anti-viral responses. Parkin is ISGylated at two lysine residues by ISG15, mediated by HERC5 E3-ligase, promoting parkin's E3 ligase activity and increasing its cytoprotective effect(Im et al.). ISG15 is activated by interferons alpha and beta and is a ubiquitin pathway antagonist that inhibits polyubiquitylation(Desai et al. 2018).

Peroxisomal Biogenesis Factor 10 (PEX10) is part of a collection of peroxisomal biogenesis proteins that import peroxisomal matrix proteins and are associated with autosomal recessive neurodegenerative disorders called peroxisome biogenesis disorders (PBD) on the Zellweger syndrome spectrum (PBD-ZSS). *PEX10* mutations associated with this disease can result in a truncated isoform that lacks a c-terminal ring finger domain(Yik et al. 2009). PEX5 also arose as a gene of interest in our study, to be discussed in a later section. Seventeen HLA-B transcripts were among the top integrated hits, all hypomethylated and underexpressed. These genes were excluded from Table 21 to show additional genes of interest, though this highlights the important presence of T-cell activation and the immune response involved in this disease.

Gene	Transcript	Probe	P-Val Combined	Methylation Fold Change	Expression Beta	Probe Position Relative to Island
ISG15	ENST00000379389.4	cg12257384	4.34×10^{-6}	-0.58	-0.79	Island
ITM2C	ENST00000335005.10	cg09972569	3.45×10^{-5}	-0.73	-2.22	Island
YIF1B	ENST00000592694.5	cg11272232	6.40×10^{-5}	-0.75	-2.15	OpenSea
TREM1	ENST00000334475.10	cg26223899	2.21×10^{-4}	-0.52	-0.80	N_Shelf
PEX10	ENST00000447513.6	cg01947769	4.45×10^{-4}	-0.53	-0.89	S_Shelf
FKBP8	ENST00000597547.1	cg10011232	4.81×10^{-4}	-0.53	-1.02	S_Shore
Hypermethylated and Underexpressed						
SUMF2	ENST00000529457.1	cg10135877	2.44×10^{-5}	1.45	-1.03	S_Shelf
BAD	ENST00000309032.7	cg21986781	3.35×10^{-5}	0.54	-0.30	S_Shelf
IGFBP4	ENST00000269593.4	cg05484949	2.08×10^{-4}	0.53	-0.40	OpenSea
LGALS9	ENST00000467111.5	cg26227119	4.36×10^{-4}	0.29	-0.41	OpenSea
C22orf34	ENST00000405854.5	cg19015962	6.32×10^{-4}	0.78	-0.79	N_Shelf
SIGIRR	ENST00000528116.5	cg04663285	7.84×10^{-4}	1.64	-0.71	Island
MOB2	ENST00000531976.1	cg27373015	8.42×10^{-4}	0.60	-0.20	OpenSea
LGALS9	ENST00000486774.1	cg26227119	8.72×10^{-4}	0.29	-0.26	OpenSea
Hypomethylated and Overexpressed						
CD14	ENST00000498971.6	cg19008097	5.19×10^{-5}	-0.54	1.60	S_Shore

SUPT4H1	ENST00000581204.1	cg22481606	9.70×10^{-5}	-0.52	1.56	Island
ORC4	ENST00000392857.9	cg05172887	1.20×10^{-4}	-0.57	1.72	OpenSea
HSPH1	ENST00000602786.5	cg18351683	1.33×10^{-4}	-0.99	1.19	OpenSea
FCGR2A	ENST00000461298.1	cg19240319	1.47×10^{-4}	-0.71	0.59	Island
FCGR2A	ENST00000461298.1	cg00975746	1.61×10^{-4}	-1.12	0.59	N_Shore
PI4KB	ENST00000368875.6	cg00107782	1.87×10^{-4}	-0.51	1.06	Island
ACTN1	ENST00000394419.8	cg27036347	4.34×10^{-4}	-1.33	1.99	Island
PDIA3	ENST00000455250.1	cg02953927	5.40×10^{-4}	-0.85	1.35	Island
CPSF2	ENST00000555244.1	cg09996633	8.35×10^{-4}	-0.78	0.86	OpenSea
SUMO1	ENST00000392244.7	cg08267047	8.42×10^{-4}	-0.68	1.25	S_Shore

Table 21. Differentially methylated and differentially expressed genes in PD.

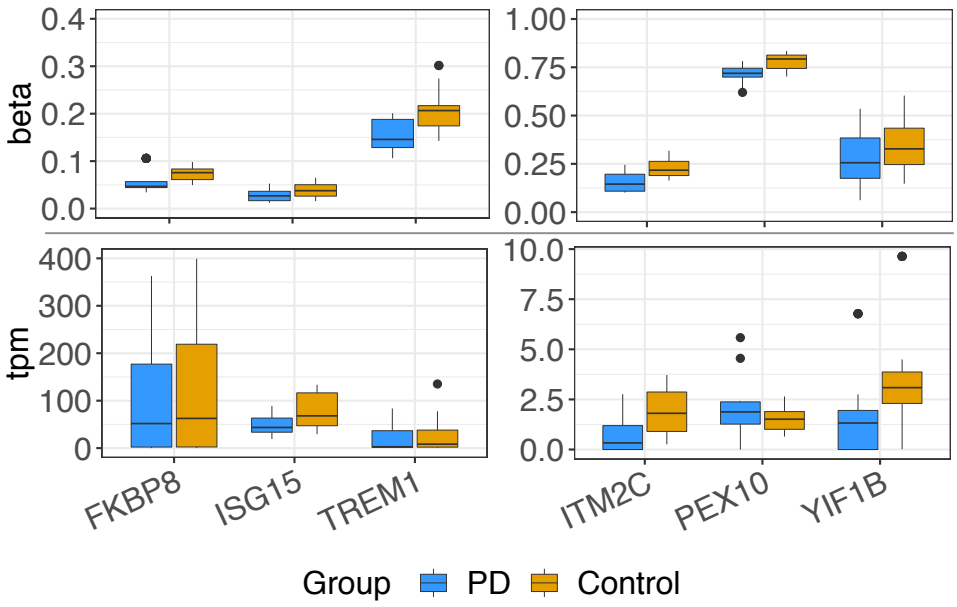


Figure 16. Top DE and DMP by combined p-value in PD.

Integrated Analysis: Combined p-values for *Alzheimer's & Parkinson's Disease*

SLC12A6 codes for a potassium/chloride cotransporter (KCC3). A mutation in this gene is known to cause hereditary motor and sensory neuropathy associated with agenesis of the corpus callosum (HMSN/ACC) which has developmental and neurodegenerative features(Dupré et al. 2006; Auer et al. 2016).

CUGBP Elav-Like Family Member 2 (CELF2) regulates alternative splicing and has been shown to repress inclusion of exon 10 in *MAPT*, tau protein, an aberration present in myotonic dystrophy type 1 (DM1) neuromuscular disease(Dhaenens et al. 2011).

POLD4 is the delta 4 subunit of DNA polymerase, involved in replication and repair. A study that used a chromosomal replication array found this gene to be underexpressed in schizophrenia versus controls, but found no difference when they looked at DNA methylation between the groups(Okazaki et al. 2016).

Gene	Transcript & Probe	AD P-Val Combined	AD Methyl Fold Change	AD Expr Beta	PD P-Val Combined	PD Methyl Fold Change	PD Expr Beta	Probe Position Relative to Island
SLC12A6	ENST00000290209.9 cg27566947	2.38×10^{-4}	0.50	0.25	2.05×10^{-3}	0.40	0.20	N_Shore
POLD4	ENST00000312419.7 cg08662757	2.05×10^{-4}	-0.72	-0.26	7.85×10^{-5}	-0.70	-0.28	Island
POLD4	ENST00000312419.7 cg25753024	1.05×10^{-4}	-0.81	-0.26	1.82×10^{-5}	-0.87	-0.28	Island
MCTP2	ENST00000357742.8 cg21467108	1.10×10^{-4}	-0.42	0.34	9.30×10^{-4}	-0.40	0.24	OpenSea
SNX18	ENST00000381410.4 cg23001370	4.82×10^{-4}	-0.29	0.39	2.87×10^{-5}	-0.40	0.37	S_Shore
CELF2	ENST00000416382.6 cg11832281	3.72×10^{-4}	-0.43	0.49	1.29×10^{-3}	-0.36	0.44	S_Shelf
PRR5	ENST00000475850.1 cg23240895	2.85×10^{-6}	-0.42	-0.50	5.21×10^{-5}	-0.42	-0.34	OpenSea
POLRMT	ENST00000587057.5 cg02707162	4.21×10^{-4}	-0.61	-0.39	1.02×10^{-3}	-0.63	-0.31	N_Shore

Table 22. Differentially methylated and differentially expressed genes in AD and PD.

Integrated Analysis: meQTL for Alzheimer's Disease Patients.

Trans meQTLs showed a number of HLA, immunoglobulin and other immune response involved transcripts associated with distant probes. NF-κB signaling, regulation and RNA processing are among the top pathways after immune related functions.

Neuron Navigator 1 (*NAV1*; beta (effect = -0.04, FDR = 2.24 x 10^{-3}) and RB Binding Protein 4, Chromatin Remodeling Factor (*RBBP4*; beta (effect = 1.01, FDR = 9.64 x 10^{-6}) are both *cis* meQTLs. NAV1 is primarily expressed in the nervous system and is thought to play a role in growth cone dynamics of axons, dendrites, and migrating neurons(Martínez-López et al. 2005).

Tubulin Folding Cofactor D (TBCD) is a *cis* meQTL involved with post-translational protein folding. One CpG site is positively associated with three transcripts in our results (Table 23). Upon upregulation of RBBP4 in DG of aged mice, novel object recognition and spatial reference memories improved comparably, independent of neurogenesis, to that of WT young mice(Pavlopoulos et al. 2013).

CNOT2 is a marginal *cis* meQTL that is related to exonuclease activity pathway and mRNA processing. It induces ubiquitination of p62 (*SQSTM1*), a protein previously discussed in this work(Jeong et al. 2017).

Lactotransferrin (LTF; beta (effect = -105.89, FDR = 7.36 x 10^{-6}) is an iron-binding glycoprotein that confers antioxidant effects and is a *trans* meQTL for AD. In addition to blood, it is expressed in the brain and is associated with aging and neurological disorders, including AD, as it has been found in neurofibrillary tangles and it may positively modulate the AKT pathway resulting in reduced AD pathology(Mohamed et al. 2019).

Transcript ID	Probe	Gene	FDR	Beta (effect)
CIS				
ENST00000587905	cg01299279	LDLRAD4	2.22 x 10^{-5}	-0.60
ENST00000588042	cg02278683	AC009303.2	8.95 x 10^{-3}	-0.28
ENST00000430015	cg12468255	NAV1	2.24 x 10^{-3}	-0.04
ENST00000571796	cg00960700	TBCD	6.18 x 10^{-3}	40.00

ENST00000574886	cg00960700	TBCD	6.77×10^{-4}	36.38
ENST00000300784	cg00960700	FN3K	7.92×10^{-3}	19.57
ENST00000592411	cg23784912	ZNF519	5.87×10^{-3}	4.26
ENST00000570734	cg00960700	FN3K	4.74×10^{-2}	4.19
ENST00000531983	cg17339202	RBBP4	9.64×10^{-6}	1.01
ENST00000382149	cg15401952	KCNIP4	2.38×10^{-2}	0.33
ENST00000525171	cg14304817	NAALAD2	9.64×10^{-6}	0.11
TRANS				
ENST00000421349	cg02553872	HLA-B	1.04×10^{-5}	-4095.98
ENST00000415187	cg21120951	TMSB4XP6	9.19×10^{-6}	-1745.62
ENST00000466304	cg02553872	HLA-B	3.16×10^{-6}	-314.34
ENST00000466340	cg02553872	SELL	1.04×10^{-5}	-234.83
ENST00000476677	cg03184194	HLA-DPB1	1.04×10^{-5}	-229.25
ENST00000638375	cg18626958	HLA-A	3.35×10^{-6}	-124.12
ENST00000399500	cg18626958	HLA-DPB1	2.96×10^{-6}	-111.09
ENST00000492796	cg02553872	CDC42SE1	9.19×10^{-6}	-107.60
ENST00000426532	cg10814415	LTF	7.36×10^{-6}	-105.89
ENST00000384161	cg22620689	RNU6-48P	1.04×10^{-5}	-75.51
ENST00000487685	cg03184194	HLA-DPB1	1.04×10^{-5}	-70.63
ENST00000395388	cg00960700	HLA-DRA	9.19×10^{-6}	5426.51
ENST00000384161	cg23784912	RNU6-48P	1.04×10^{-5}	321.31
ENST00000482769	cg14304817	IGKV2-28	1.04×10^{-5}	228.80
ENST00000633682	cg14304817	IGKV2-28	1.04×10^{-5}	228.80
ENST00000613056, ENST00000614745	cg23784912	MBOAT7	1.05×10^{-5}	212.19
ENST00000530802	cg00960700	TCIRG1	9.19×10^{-6}	115.48
ENST00000466602	cg00960700	CYP27A1	9.19×10^{-6}	100.45
ENST00000636081	cg23784912	EHMT1	1.04×10^{-5}	89.95
ENST00000516601	cg00960700	RNU5B-3P	9.19×10^{-6}	73.67

Transcript ID	Probe	Gene	P-value	FDR	Beta (effect)
		CIS			
ENST00000595791	cg22266001	MAG	5.18×10^{-5}	2.49×10^{-1}	4.16
ENST00000551661	cg19390271	CNOT2	6.75×10^{-5}	2.97×10^{-1}	-3.34
ENST00000550641	cg19390271	CNOT2	7.72×10^{-5}	3.06×10^{-1}	-2.12
ENST00000574975	cg00960700	TBCD	1.30×10^{-5}	8.41×10^{-2}	20.19

Table 23. Methylation quantitative trait loci (meQTLs) in AD.

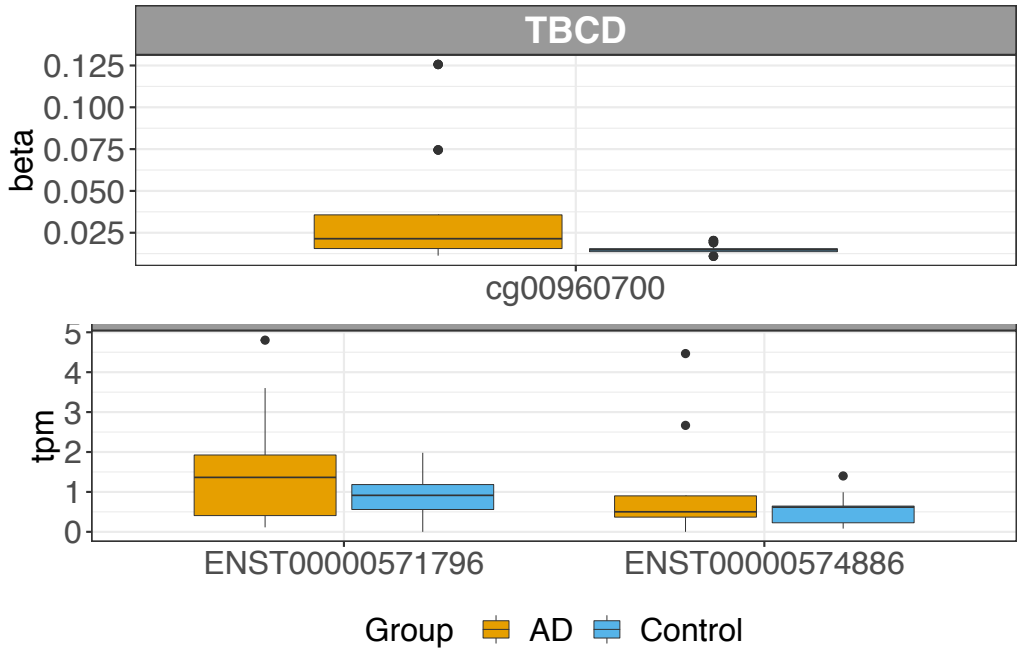


Figure 17. Representative plot of a *cis* meQTL in AD.

Integrated Analysis: meQTL for Parkinson's Disease Patients.

Gene set enrichment analysis showed associations between DMPs and expression changes in pathways related to telomere regulation and chromatin organization for genes in *cis* and protein-DNA complexing and transcription in *trans*. As with the AD cases, trans meQTLs for PD reflect an immune response with a relatively high number of HLA and other immune involved transcripts associated with distant probes. Aside from the prominent presence HLA and immunoglobulin meQTLs, genes involved in neurotrophin signaling and glycogen metabolic processes made up the next most prominent pathways for this set.

Colony Stimulating Factor 3 Receptor (*CSF3R*; beta (effect = -105.89, FDR = 7.36 x 10^{-6}) is a trans meQTL for PD. CSF1R is necessary for microglia survival(Yang et al. 2018). Additional cis meQTLs of interest are a block involved in similar functions. Histone Deacetylase 10 (*HDAC10*), Nudix Hydrolase 21 (*NUDT21*), Protein Kinase N1 (*PKN1*), and Nuclear Receptor Corepressor 2 (*NCOR2*) are involved in histone deacetylase binding and ERCC Excision Repair

1, Endonuclease Non-Catalytic Subunit (*ERCC1*) is involved with chromosome organization.

NCOR2 was also a gene of interest from the second chapter presented in this dissertation.

Transcript ID	Probe	Gene	FDR	Beta (effect)
CIS				
ENST00000483148	cg14288209	APTX	2.95×10^{-2}	-77.24
ENST00000586305	cg18480790	RASGRP4	4.40×10^{-2}	-47.56
ENST00000441983	cg02494999	TRAPPC12	2.61×10^{-2}	-29.64
ENST00000454664	cg13777287	CNKSR3	3.94×10^{-3}	-0.21
ENST00000525171	cg14304817	NAALAD2	2.08×10^{-5}	0.11
ENST00000575679	cg02214414	FLYWCH1	8.23×10^{-3}	7.00×10^{-8}
TRANS				
ENST00000638576	cg12376805	HLA-A	1.35×10^{-6}	-132.23
ENST00000391737	cg04422985	LILRA2	1.03×10^{-7}	-121.07
ENST00000412634	cg04422985	HLA-DRB1	6.24×10^{-7}	-76.77
ENST00000373106	cg24430871	CSF3R	2.24×10^{-7}	-71.42
ENST00000587728	cg26796043	ADGRE5	6.59×10^{-6}	-59.11
ENST00000373106	cg26796043	CSF3R	6.39×10^{-8}	-56.27
ENST00000463148	cg26868306	HLA-A	1.69×10^{-5}	-51.25
ENST00000490676	cg15471402	HLA-B	4.69×10^{-6}	1311.16
ENST00000390283	cg11382477	IGLV8-61	7.64×10^{-6}	838.34
ENST00000399088	cg15471402	HLA-DQB1	2.48×10^{-5}	297.73
ENST00000412634	cg23502778	HLA-DRB1	2.43×10^{-5}	155.28
ENST00000565375	cg15471402	HLA-V	2.15×10^{-11}	97.30
ENST00000430299	cg15471402	HLA-B	1.31×10^{-10}	84.96
ENST00000391738	cg23502778	LILRA2	2.54×10^{-5}	83.04
ENST00000639848	cg15471402	HLA-B	5.79×10^{-8}	81.13
ENST00000633660	cg15471402	TRBJ2-7	1.77×10^{-6}	78.22
ENST00000632846	cg11382477	IGHV3-66	1.26×10^{-6}	66.85
ENST00000625402	cg04865110	SCARNA4	2.76×10^{-6}	59.85
ENST00000384986	cg04865110	MIR623	2.76×10^{-6}	57.15
ENST00000547013	cg15471402	TMBIM6	2.92×10^{-6}	54.91

Transcript ID	Probe	Gene	P-value	FDR	Beta (effect)
		CIS			
ENST00000436302	cg05351601	AGBL3	1.31×10^{-5}	5.71×10^{-2}	7.81
ENST00000592410	cg26322788	ERCC1	1.71×10^{-5}	5.73×10^{-2}	4.56
ENST00000497952	cg23762037	HDAC10	2.11×10^{-5}	6.06×10^{-2}	-810.96
ENST00000405201	cg16713625	NCOR2	1.26×10^{-4}	2.27×10^{-1}	-22.82
ENST00000563362	cg11323230	NUDT21	2.30×10^{-4}	2.75×10^{-1}	-171.57
ENST00000591461	cg20468939	PKN1	8.66×10^{-4}	5.92×10^{-1}	-189.66

Table 24. Methylation quantitative trait loci (meQTLs) in PD.

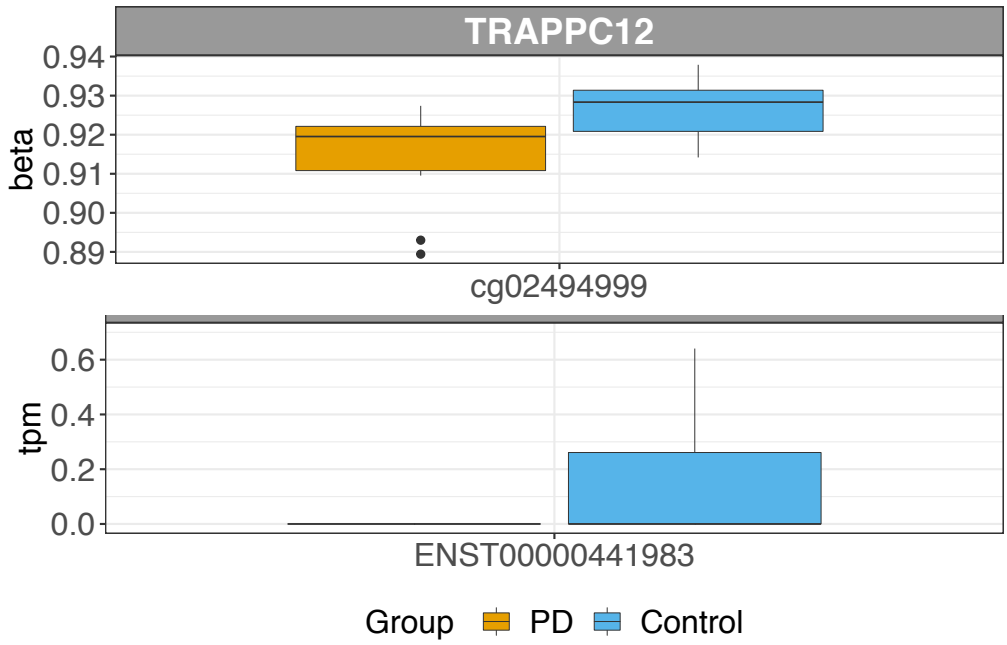


Figure 18. Representative plot of a *cis* meQTL in PD

Integrated Analysis: meQTLs for Alzheimer's & Parkinson's Disease.

NAALAD2 is part of the transferrin receptor family involved in iron transfer. It cleaves NAAG, a neuropeptide expressed in the peripheral and central nervous systems, and dysregulation of this system has been suggested to contribute to several neurological disorders through altered glutamate regulation(Pangalos et al. 1999). It is the only *cis* acting meQTL present for both AD and PD in our results.

Two CpG sites shared by AD and PD have *trans* acting associations with a variety of gene transcripts. Both diseases produced *trans* meQTLs for a relatively large number of noncoding RNA species, including pseudogenes and miRNAs.

Transcript	Probe	Gene	AD FDR	AD Beta (effect)	PD FDR	PD Beta (effect)
			CIS			
ENST00000525171	cg14304817	NAALAD2	9.64×10^{-6}	0.11	2.08×10^{-5}	0.11
			TRANS			
ENST00000632846	cg14304817	IGHV3-66	2.94×10^{-5}	35.60	1.98×10^{-6}	34.99
ENST00000390283	cg14304817	IGLV8-61	8.81×10^{-6}	455.36	4.54×10^{-6}	441.86
ENST00000383411	cg17339202	MPIG6B	1.22×10^{-5}	3.88	1.70×10^{-5}	3.79
ENST00000525902	cg17339202	CTNND1	1.49×10^{-5}	7.28	2.36×10^{-5}	7.09
ENST00000432428	cg14304817	FLOT1	1.40×10^{-5}	18.71	2.46×10^{-5}	18.19
ENST00000362918	cg14304817	RNY3P2	1.72×10^{-5}	9.58	2.55×10^{-5}	9.31
ENST00000368330	cg14304817	YY1AP1	1.72×10^{-5}	5.38	2.55×10^{-5}	5.23
ENST00000385017	cg14304817	MIR553	1.72×10^{-5}	6.36	2.55×10^{-5}	6.18
ENST00000385302	cg14304817	MIR320A	1.72×10^{-5}	5.44	2.55×10^{-5}	5.29
ENST00000408135	cg14304817	MIR1292	1.72×10^{-5}	6.54	2.55×10^{-5}	6.35
ENST00000408346	cg14304817	MIR1237	1.72×10^{-5}	9.58	2.55×10^{-5}	9.31
ENST00000408874	cg14304817	MIR548H2	1.72×10^{-5}	5.75	2.55×10^{-5}	5.59
ENST00000411464	cg14304817	MRPL55	1.72×10^{-5}	3.37	2.55×10^{-5}	3.28
ENST00000458862	cg14304817	SNORA47	1.72×10^{-5}	8.97	2.55×10^{-5}	8.72
ENST00000459332	cg14304817	SNORD108	1.72×10^{-5}	5.97	2.55×10^{-5}	5.80
ENST00000466367	cg14304817	ARRDC1	1.72×10^{-5}	3.69	2.55×10^{-5}	3.59
ENST00000521949	cg14304817	NIPAL2	1.72×10^{-5}	6.75	2.55×10^{-5}	6.56
ENST00000533417	cg14304817	CD81	1.72×10^{-5}	3.72	2.55×10^{-5}	3.61
ENST00000579005	cg14304817	SAP30BP	1.72×10^{-5}	5.12	2.55×10^{-5}	4.98
ENST00000582739	cg14304817	MIR548Y	1.72×10^{-5}	11.69	2.55×10^{-5}	11.36
ENST00000584654	cg14304817	MIR3116-1	1.72×10^{-5}	5.82	2.55×10^{-5}	5.66
ENST00000620884	cg14304817	MIR6837	1.72×10^{-5}	3.36	2.55×10^{-5}	3.27
ENST00000637097	cg14304817	SMARCA2	1.72×10^{-5}	8.80	2.55×10^{-5}	8.56
ENST00000625868	cg14304817	SERF1A	1.75×10^{-5}	3.61	2.55×10^{-5}	3.51
ENST00000461736, ENST00000488182	cg17339202	ABHD16A	2.59×10^{-5}	4.79	3.15×10^{-5}	4.67
ENST00000479622, ENST00000481723	cg17339202	ABHD16A	2.53×10^{-5}	6.83	3.15×10^{-5}	6.66
ENST00000415983	cg17339202	HLA-U	2.84×10^{-5}	5.07	3.15×10^{-5}	4.95
ENST00000616820	cg17339202	LILRB4	2.84×10^{-5}	5.90	3.15×10^{-5}	5.76

Table 25. Methylation quantitative trait loci (meQTLs) in AD and PD.

Integrated Analysis: FEM for Alzheimer's Disease Patients

Functional epigenetic modules identifies gene modules in protein-protein interaction networks associated with differentially methylated and differentially expressed loci. The FEM networks in AD are driven by five genes: *CSGALNACT1* ($p = 0.00$), *PNPLA2* ($p = 5.00 \times 10^{-3}$), *PILRA* ($p = 1.40 \times 10^{-2}$), *AGTRAP* ($p = 3.90 \times 10^{-2}$), and *LPPR2* ($p = 4.10 \times 10^{-2}$).

Paired Immunoglobin Like Type 2 Receptor Alpha (PILRA) is a cell surface receptor expressed on innate immune cells, including microglia, that inhibits signal transduction through recruitment of phosphatases for control by dephosphorylation. Rathore at al. describe a missense variant in PILRA that may confer protection from a known AD risk locus (7q21 locus (rs1476679)) through reduced ligand interaction and subsequent increased microglia activation(Rathore et al. 2018).

Integrated Analysis: FEM for Parkinson's Disease Patients

FEM networks for PD produced the following gene modules: *PEX5* ($p = 2.00 \times 10^{-3}$), *RAB3IP* ($p = 5.00 \times 10^{-3}$), *PPP4R2* ($p = 7.00 \times 10^{-3}$), *MOB4* ($p = 1.30 \times 10^{-2}$), *CSGALNACT1* ($p = 1.70 \times 10^{-2}$), *SLC4A7* ($p = 2.60 \times 10^{-2}$), and *KCNJ11* ($p = 2.60 \times 10^{-2}$).

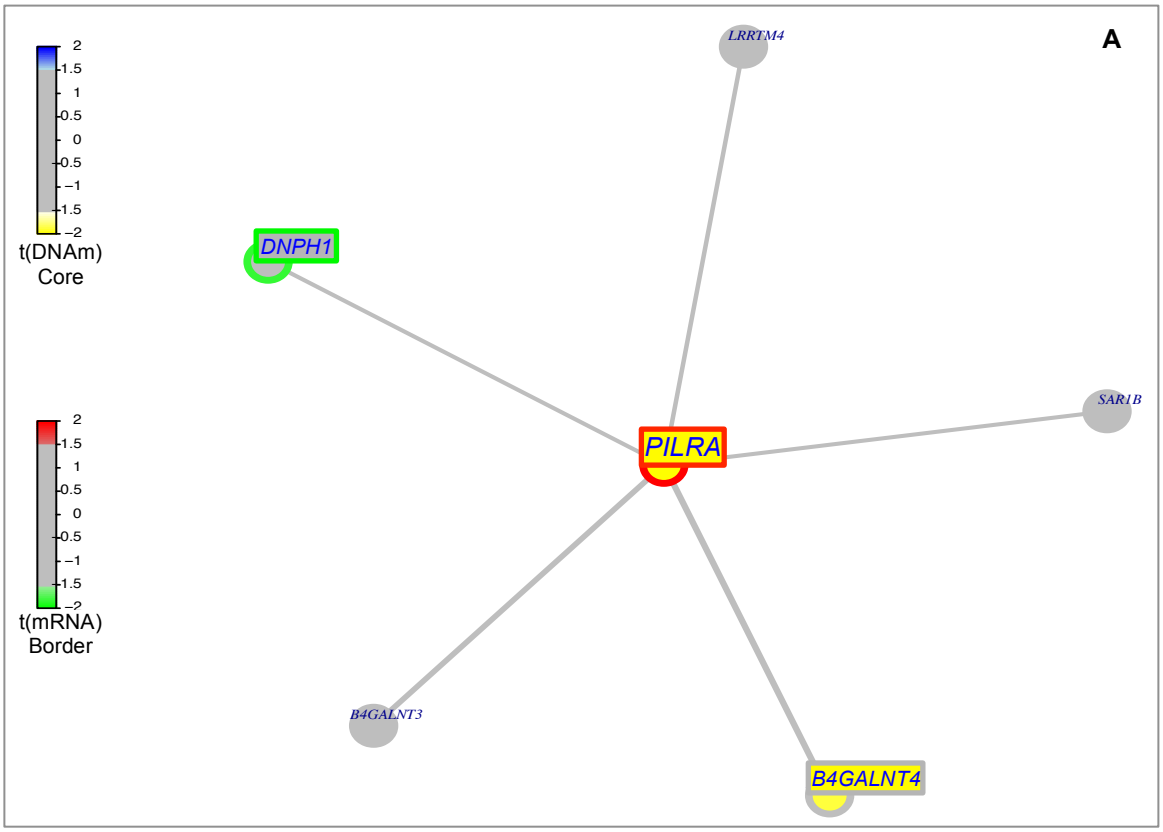
Peroxisomal Biogenesis Factor 5 (PEX5) binds to PTS1-type peroxisomal targeting signal and is essential for peroxisomal protein import. *PEX5* has been studied for its potential role in PD from the context of peroxisome biogenesis disorders and the effect of oxidative stress on alpha-synuclein(Yakunin et al. 2010). *PEX10* also revealed itself in the present study as a gene that was both differentially methylated and expressed, as well as alternatively spliced (results not shown).

Integrated Analysis: FEM for Alzheimer's & Parkinson's Disease

Chondroitin Sulfate N-Acetylgalactosaminyltransferase 1 (CSGALNACT1) is an enzyme that functions to elongate chondroitin sulfate chains and is associated with the progression of multiple sclerosis. This gene is the driver of a module in both AD and PD.

Gene (Seed)	Module Size	P-value	Genes in Module
AD modules			
CSGALNACT1	11	0.000	CSGALNACT1, TMEM214, SLC30A1, TMUB1, CSGALNACT2, TMX4, KIAA1467, GPRC5B, B3GALNT1, SLC8A3, SLC8A2
PNPLA2	6	0.005	PNPLA2, SERPINF1, PHYHIP, ABHD5, FAM131A, PLIN1
PILRA	6	0.014	PILRA, SAR1B, LRRTM4, DNPH1, B4GALNT3, B4GALNT4
AGTRAP	15	0.039	AGTRAP, TMEM14B, NDRG4, C12orf10, GAD2, ZNF391, PCTP, SYT16, NFU1, PITPNC1, ACSF2, TBRG4, MIEF1, GAD1, SYT14
LPPR2	6	0.041	LPPR2, C9orf40, PCDH9, DNAJC12, UQCC2, METAP1
PD modules			
PEX5	24	0.002	PEX5, PEX7, CAPRIN2, RANBP6, TOMM7, ACOT2, LONP2, GDPD5, HMGCL, EPHX2, PEX10, ZADH2, ECI2, S100A6, PEX2, TYSND1, PEX12, ZNF772, ACOX3, MLYCD, ACOX2, TM6SF1, MS4A7, PEX5L
RAB3IP	5	0.005	RAB3IP, CCDC146, FAM124A, RAB3IL1, RAB3D
PPP4R2	6	0.007	PPP4R2, WDR55, LCMT1, FAM133B, AMPD2, CUTA
MOB4	18	0.013	MOB4, STRIP1, STRN, STK25, ZRANB1, PTPRS, EGLN2, KIDINS220, SLMAP, CTTNBP2NL, SIKE1, STRIP2, PLEKHM1, CTTNBP2, RASSF3, RUFY3, KIAA1456, ASH1L
CSGALNACT1	7	0.017	CSGALNACT1, ARFGEF1, SLC30A1, CSGALNACT2, TMX4, KIAA1467, GPRC5B
SLC4A7	11	0.026	SLC4A7, SLC4A4, GDE1, TEX29, STEAP3, CA4, SLC6A15, CA2, SLC4A8, PTS, FNDC5
KCNJ11	8	0.026	KCNJ11, KCNJ8, FAM63B, ABCC8, CYP2J2, SLC2A8, XXYLT1, NCR3LG1

Table 26. Functional epigenetic modules for AD and PD. Protein-protein interaction sub-networks identified as differentially methylated (promoter region) and differentially expressed, with an inverse relationship.


A
t(DNAm)
Core
t(mRNA)
Border
LRRTM4
DNPH1
SAR1B
PILRA
B4GALNT3
B4GALNT4

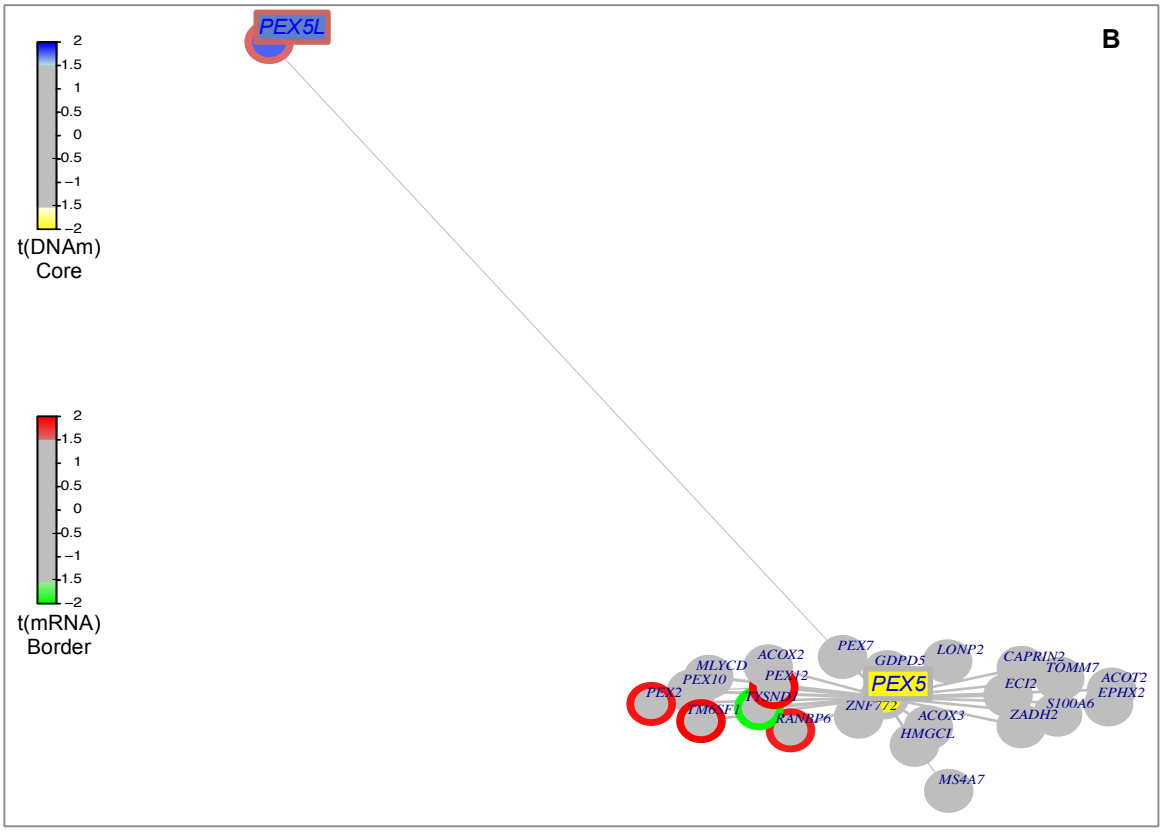

B
t(DNAm)
Core
t(mRNA)
Border
PEX5L
PEX7
GDPD5
LONP2
CAPRIN2
TOMM7
ACOT2
EPHX2
MLYCD
ACOX2
PEX10
PEX12
PEX5
ECI2
S100A6
PEX2
ZADH2
ACOX3
HMGCL
MS4A7

Figure 19. Functional epigenetic modules. (A) *PILRA*, in AD, is a hypermethylated and overexpressed gene driving this module and (B) *PEX5*, in PD, is hypomethylated connected with several overexpressed genes, and one closely related gene that is hypermethylated and overexpressed (PEX5L).

OVERLAPPING RESULTS ACROSS STUDIES

HSPB1 was overexpressed in PDD brain tissue (Chapter 1) and underexpressed in PD blood (Chapter 3). Five of the 7 significant overlapping expression hits across these two studies were hemoglobin (Hb) subunit (α or β) genes (*HBA2* or *HBB*), all underexpressed in both tissue types. Overlapping expression results are presented in Table 27.

There were about 30 differentially methylated positions in common between our chapter 2 and chapter 3 studies (Table 28). Histone deacetylase activity and Notch pathway signaling are the most prominent pathways in this list.

Transcript ID	Gene Name	Fold Change (log2) PD Brain	adjP-val PD Brain	beta_DE PD Blood	pval PD Blood
PD brain and PD blood overlapping results					
ENST00000347063.8	RNF175	-0.77	8.08×10^{-3}	0.46	3.23×10^{-4}
ENST00000377103.2	THBD	1.11	1.47×10^{-3}	0.69	3.79×10^{-4}
ENST00000482949.5	LCK	1.57	3.84×10^{-3}	2.02	3.09×10^{-3}
ENST00000246006.4	CD93	1.56	3.24×10^{-5}	0.39	6.97×10^{-3}
ENST00000334475.10	TREM1	1.33	3.68×10^{-3}	-0.80	8.12×10^{-3}
ENST00000552456.1	P2RX5	-1.02	3.34×10^{-3}	-0.56	9.22×10^{-3}
ENST00000497360.5	SLC4A1	-1.74	1.12×10^{-4}	-0.82	9.80×10^{-3}
PDD brain and PD blood overlapping results					
ENST00000429938.1	HSPB1	1.62	1.06×10^{-4}	-0.43	4.69×10^{-3}
ENST00000475226.1	HBB	-1.33	3.29×10^{-3}	-0.88	1.17×10^{-2}
ENST00000397806.1	HBA2	-1.34	1.29×10^{-3}	-0.84	1.42×10^{-2}
ENST00000482565.1	HBA2	-1.34	1.29×10^{-3}	-0.74	1.99×10^{-2}
ENST00000485743.1	HBB	-1.33	3.29×10^{-3}	-0.79	2.00×10^{-2}
ENST00000387372.1	MT-TQ	2.20	2.17×10^{-5}	-0.66	2.80×10^{-2}
ENST00000633227.1	HBB	-1.33	3.29×10^{-3}	-0.78	2.94×10^{-2}

Table 27. Differentially expressed genes shared across PD or PDD *brain* (chapter 1), and PD *blood* (chapter 3).

Probe	Gene ID	Study 3 Fold Change (log_2)	Study 3 P.Value
Cross-sectional (S2 DMPs – Table 6)			
cg02938407	EXOC3L4	-0.27	1.68×10^{-3}
cg10800369	LMO3	-0.84	2.16×10^{-3}
cg26354439	DYNC1H1	-0.28	2.23×10^{-3}
cg04347164	ZFR2	-0.27	3.05×10^{-3}
cg04154697	ZFR2	-0.30	3.09×10^{-3}
cg18253473	DYNC1H1	-0.23	3.73×10^{-3}
cg27109672	DLK1	-0.26	4.82×10^{-3}
cg03578918	EXOC3L4	-0.29	4.97×10^{-3}
cg25285743	LMO3	0.80	5.31×10^{-3}
cg03163184	LMO3	-0.62	6.35×10^{-3}
cg22382058	ZFR2	-0.28	7.72×10^{-3}
Longitudinal (S2 DMPs – Table 8)			
cg03508817	HDAC4	-0.36	2.52×10^{-3}
cg14120784	GTF2I	-0.39	2.57×10^{-3}
cg16713625	NCOR2	-0.74	3.83×10^{-3}
cg13956095	METRNL	-0.55	5.13×10^{-3}
cg20105848	KRTAP5	-0.43	5.31×10^{-3}
cg07055330	HDAC4	0.31	5.49×10^{-3}
cg01390479	HDAC4	-0.24	5.57×10^{-3}
cg01764079	ZNF623	-0.22	6.10×10^{-3}
cg26451769	METRNL	-0.24	6.99×10^{-3}
cg21248987	SDK1	-0.93	7.49×10^{-3}
cg22166325	EDC3	-0.24	7.91×10^{-3}
cg02001099	HDAC4	0.72	8.00×10^{-3}
cg17647904	NCOR2	-0.29	8.29×10^{-3}
cg24978805	SDK1	-0.24	8.39×10^{-3}
cg16114651	GTF2I	-0.35	8.41×10^{-3}
cg18493115	KRTAP5	1.40	8.43×10^{-3}
cg06228737	ABAT	0.37	8.70×10^{-3}
cg05865340	HDAC4	-0.51	8.81×10^{-3}
cg21831587	GTF2I	-1.05	9.07×10^{-3}
cg02329576	SDK1	-0.44	9.72×10^{-3}

Probe	Gene ID	Study 3 Fold Change (log_2)	Study 3 P.Value
Longitudinal, with medication (S2 DMPs – Table 9)			
cg26247309	PDGFRB	-0.42	2.11×10^{-3}
cg03508817	HDAC4	-0.36	2.52×10^{-3}
cg14120784	GTF2I	-0.39	2.57×10^{-3}
cg10543797	NADK	0.31	3.12×10^{-3}
cg07055330	HDAC7	0.31	5.49×10^{-3}
cg01390479	HDAC8	-0.24	5.57×10^{-3}
cg01764079	ZNF623	-0.22	6.10×10^{-3}
cg02001099	HDAC5	0.72	8.00×10^{-3}
cg16114651	GTF2I	-0.35	8.41×10^{-3}
cg05865340	HDAC6	-0.51	8.81×10^{-3}
cg21831587	GTF2I	-1.05	9.07×10^{-3}
Longitudinal, No medication (S2 DMPs – Table 10)			
cg17087669	MAPK13	-0.36	1.81×10^{-3}
cg02711976	KIAA0319	-0.35	3.56×10^{-3}
cg17211404	PLA2G7	-0.45	4.18×10^{-3}
cg18865832	PCDHB16	0.76	7.68×10^{-3}
cg14983698	CEP70	-0.24	9.25×10^{-3}
cg08908843	SECISBP2	-0.20	9.93×10^{-3}

Table 28. Differentially methylated positions shared across chapter 2 and chapter 3 studies in PD blood.

DISCUSSION

Many known AD and PD associated pathways continually arose in our datasets. This is important to reinforce certain known, or suspected mechanisms associated with disease, but we also presented additional targets to be exploited for potential therapeutic applications which would follow the necessary functional validations. The other important aspect of thematic gene and pathway hits is the prospect for biomarker development, especially for hits that are present in the three separate studies, including across tissues, presented here.

Our results showed links to two different HERC E3 ligase proteins, HERC5 and HERC2, associated with known PD susceptibility genes, parkin and LRRK2, respectively. These results highlighted the ubiquitin pathway and Notch signaling, both good targets for PD research extending into earlier disease stages. PKA, PI3K/AKT, and JAK/STAT are additional pathways of particular continued interest in both AD and PD. Mechanisms that affect transcription regulation and chromatin organization are relevant to our interests in gene expression and regulation via DNA methylation, but there is also a strong immune response presence is these diseases, particularly evident in the HLA and immunoglobulin hits in both brain and blood tissues. Active immune-related pathways is not a new finding, but it supports that this facet of neurological disorders must remain a focus of study and cannot be dismissed as a side-effect of inevitable neuroinflammation owing to some other causal early life insult to the nervous system.

Differential Expression in AD & PD.

The presence of NUMB, APP and UPP1 together implicates the involvement of γ-secretase activity in both diseases. Another common thread among the genes differentially expressed in PD and AD is protein degradation, whether that be a contributor to pathology or a consequence thereof.

Integrated Analysis: Combined p-values for AD & PD.

The unifying theme in the integrated PD analysis, combining differentially methylated positions and differentially expressed genes, is ubiquitination and the mechanistically related process of mitophagy. These pathways are widely known for their involvement in aberrant functioning within neurodegenerative disease systems, like ALS and AD.

Overlapping Results Across Studies.

Overexpressed HSPBs. Small heat shock proteins (HSPBs) are a common focus within PD research for their involvement in protein folding. HSPB1, along with several other HSPBs, was found to be a potent aggregation suppressor of mutated PARK2 isoform (C289G), a prevalent cause of juvenile onset PD(Minoia et al. 2014). The authors of that study suggested that HSPB1 aggregation interference was involved with an autophagy or protein degradation related function.

Underexpressed hemoglobin (Hb) subunit (α or β) genes (*HBA2* or *HBB*). *HBA2* and *HBB* are preferentially expressed in A9 dopamine (DA) neurons of the substantia nigra (the primary site associated with PD pathology) over A10 DA neurons of the adjacent ventral tegmental area (not associated with PD pathology). Codrich et al. suggest Hb related impairments could be related to several functions already tied to PD, including mitochondrial function and autophagy impairment. Among their findings, they reported reduced DA with Hb overexpression in vivo; additionally, they cited down-regulation of both Hb subunits in DA neurons of rotenone treated rats, but up-regulation in PD brains(Codrich et al. 2017). In line with our underexpressed results, Ferrer et al. reported underexpression of *HBB* and *HBA2* in the vast majority of neurons containing hyper-phosphorylated tau deposits in the frontal cortex and hippocampus of AD brains, which is also relevant to the late stages of PD pathology, and PDD, where tau pathology is prevalent. They also found that Hb was barely present in about 80% of DA substantia nigra neurons that contained α-synuclein inclusions(Ferrer et al. 2011a). Also of interest in this overlapping results list were Mitochondrially Encoded tRNA Glutamine (*MT-TQ*) and AD-associated Triggering Receptor Expressed On Myeloid Cells (*TREM1*), overexpressed in PDD or PD brain, respectively, and underexpressed in PD blood.

Differential methylation related to HDAC activity and Notch pathway signaling. Previously discussed, NUMB was overexpressed in our AD and PD blood differential expression results, and *NCOR2* was a hit from the longitudinal PD study in chapter 2, which is also a DMP in the present study. Notch signaling affects cell differentiation and development, including neurite outgrowth, synaptic plasticity and olfactory functions in the adult brain. Imai et al. reported that

LRRK2, a key PD susceptibility gene, modulates Notch signaling along its endocytic pathway via the LRRK2-associated proteins, NEURL4 and HERC2(Imai et al. 2015).

CONCLUSIONS

These studies were initiated with RNAseq to study gene expression in PD, then the query schema was elevated by incorporating epigenetic, and other expression modifiers, to ultimately join these data types into “multi-omic” analyses, allowing for a more nuanced and powerful characterization of the disease state.

In my first chapter, I compared gene expression levels across PD, PDD, and control brain tissues. I again applied this approach in chapter 3 to blood samples from PD, AD, and controls, with the added addition of DNA methylation measurements in the same patients.

Chapter 1 results revealed new, and some expected, molecular markers of progression from PD to PDD and we highlighted the largest differences when compared to healthy controls. Importantly, we found genes that may be more indicative of the advanced pathological state when focusing on the PD vs. PDD comparison. In particular, *PENK* may be of interest as a marker of disease progression, especially since this gene was also in a differentially methylated region, shown in chapter 2. In our alternative splicing analysis we validated a retained intron event for PDD patients in a gene called *RELA*, which suggested perturbed signaling within the immune response NFκB pathway by influencing nuclear localization capabilities. We also identified an exon skip event in *SRRM1*, which could impact proper splicing in PD.

In chapter 3 I presented integrated analyses of differential gene expression and differential methylation datasets to leverage the full potential of collecting data that represents these highly interconnected biological processes. The individual datasets showing methylation sites and genes, unique to PD or AD, yielded some interesting results alone. However, intersecting these data types pulled out possibly the most functionally relevant genes, since they were not only differentially expressed but also more likely to be under the influence of disease-specific DNA methylation changes. In addition to finding unique, disease-specific changes, I showed genes from each chapter 3 analysis that were shared between AD and PD. Combining

the top hits from those lists offers additional targets for studies of dysregulation in neurodegeneration and subsequent dementias.

Chapter 2 provided a solid preliminary step toward more extensive longitudinal tracking of DNA methylation changes in the progression of PD. Even with only two time points, we identified several genes of interest that are potential biomarker candidates. Three genes in particular were bolstered by results in chapter 3. *NCOR2*, *DLK1*, and *HDAC4* were top hits in both studies; five total *HDACs* came up as shared hits across both chapters which may be indicative of dynamic regulation by DNA methylation of these sites late in the disease stage. Four of them arose specifically from the comparison of medicated PD patients vs. controls in study 2, which implicates this class of molecules as influential in the beneficial medication effects conferred upon PD patients. It is also feasible to consider then, that perhaps these *HDAC*s target some early sites of molecular pathology making them a class to be tracked in a future longitudinal study over many years. NCOR2 and DLK1 hits across studies suggests that these genes are also good candidates for PD biomarkers, and their involvement in NOTCH signaling provides a target pathway for functional validation studies as well.

Taken together, this collection of work has added more information to the field about the importance of splicing in Parkinson's disease brains, and I provided additional evidence that some changes in the brain are reflected in the blood. We also uncovered DNA methylation hits from two different blood studies that might be good PD biomarkers, for diagnosis, for progression of the disease, and for monitoring of medication efficacy. Additional work is needed to develop these candidates into a true methylation biomarker panel that can be clinically tested, but I have narrowed these datasets down to at least, but not limited to, five potential genes as a starting point: *PENK*, *CYP2E1*, *NCOR2*, *DLK1*, and *HDAC4*.

FUTURE DIRECTIONS

The foundational work presented here marks an important starting point for further PD biomarker development using signature DNA methylation changes as the basis for disease detection. In order to move closer to disease prediction in the pre-motor stage and strengthen our

candidate markers, we should expand our profiling to include additional PD subsets to obtain a more complete characterization along the PD spectrum.

First, the chapter 2 longitudinal study can essentially be lengthened to cover more years of progression, and also include in an early-onset cohort (<60 years old), aiming for ten blood collections from each patient over ten years, for example. Alternatively, track patients for 5 years and collect blood samples every six months. The latter timeline and collection points would be more beneficial to clinicians interested in tracking medication efficacy, and would also give a slightly better idea of the dynamic variability in methylation status across 6-month periods. In either case, track the defined set of proposed markers across groups, including as many medication naïve patients as possible for a separate comparison. Changes in the biomarker panel would be monitored with yearly or semiannual progression to look for associations in methylation changes, near *NCOR2*, *HDAC4*, and others, with rate of progression. We would also include comparisons of early-onset cases with late-onset ones, as there may be distinct differences between these PD subsets that lead to divergent biomarker panels. Conversely, it would be just as important to look for a set of target genes that consistently arise across PD subsets, which would amount to the most general PD biomarker that is still distinct from healthy controls. Since I showed preliminary evidence of PD medication effects on DNA methylation status in chapter 2, this follow-up study should again compare patients taking medication to patients not taking medication, and we would look for normalizing effects on methylation status over time in the medicated set.

In addition to tracking of the aforementioned gene regions, I would consider tracking a class or family of molecules (e.g., ten additional genes) that represented a relatively high proportion of our top gene hits, like HDACs or include additional NOTCH pathway genes in the panel. If it happens that we hit on a couple pathways that are highly active late in the disease, it may be more promising to also detect those changes much earlier in the disease state.

Second, we could target a cohort of high risk patients for yearly tracking as well. For example, REM sleep behavior disorder (RBD) is a primary prodromal PD risk factor that represents a good early detection study cohort. PD patients are diagnosed on average in their

early 60s, and RBD patients in the range of 50-60 years old. Therefore, it would be useful to track an RBD cohort beginning at least several years before the average conversion age to PD with continued monitoring several years after any PD diagnoses. To find the best disease markers spanning pre-motor to late-stage PD, the resulting hits would be compared to RBD patients who did not convert to PD, and to hits that arise in studies of already diagnosed PD patients. This design would do well with a five year split before and after conversion, expanding the total duration to an earlier starting point as patient availability and funding permit.

Finally, the target genes previously listed may also be good candidates for functional validations. I would design experiments in an effort to uncover their specific molecular roles using in vitro and in vivo models of Parkinson's disease. Validation studies could lend to improved biomarker specificity, in terms of gained knowledge of pathway functionality and DNA methylation effects, and provide solid leads for clinical trial work.

REFERENCES

Ahmed I, Buchert R, Zhou M, Jiao X, Mittal K, Sheikh TI, Scheller U, Vasli N, Rafiq MA, Brohi MQ et al. 2015. Mutations in DCPS and EDC3 in autosomal recessive intellectual disability indicate a crucial role for mRNA decapping in neurodevelopment. *Human molecular genetics* **24**: 3172-3180.

Alexa A, Rahnenfuhrer J. 2018. topGO: Enrichment Analysis for Gene Ontology. p. R package

Alkelai A, Lupoli S, Greenbaum L, Giegling I, Kohn Y, Sarner-Kanyas K, Ben-Asher E, Lancet D, Rujescu D, Macciardi F et al. 2011. Identification of new schizophrenia susceptibility loci in an ethnically homogeneous, family-based, Arab-Israeli sample. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **25**: 4011-4023.

Altar CA, Vawter MP, Ginsberg SD. 2009. Target identification for CNS diseases by transcriptional profiling. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **34**: 18-54.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Aron Badin R, Vadori M, Vanhove B, Nerriere-Daguin V, Naveilhan P, Neveu I, Jan C, Leveque X, Venturi E, Mermillod P et al. 2016. Cell Therapy for Parkinson's Disease: A Translational Approach to Assess the Role of Local and Systemic Immunosuppression. *Am J Transplant* **16**: 2016-2029.

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**: 1363-1369.

Ascherio A, Schwarzschild MA. 2016. The epidemiology of Parkinson's disease: risk factors and prevention. *Lancet Neurol* **15**: 1257-1272.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**: 25-29.

Auer RN, Laganiere JL, Robitaille YO, Richardson J, Dion PA, Rouleau GA, Shekarabi M. 2016. KCC3 axonopathy: neuropathological features in the central and peripheral nervous system. *Mod Pathol* **29**: 962-976.

Backman CM, Shan L, Zhang Y, Hoffer BJ, Tomac AC. 2007. Alterations in prodynorphin, proenkephalin, and GAD67 mRNA levels in the aged human putamen: correlation with Parkinson's disease. *Journal of neuroscience research* **85**: 798-804.

Bao AM, Meynen G, Swaab DF. 2008. The stress system in depression and neurodegeneration: focus on the human hypothalamus. *Brain research reviews* **57**: 531-553.

Barone P, Antonini A, Colosimo C, Marconi R, Morgante L, Avarello TP, Bottacchi E, Cannas A, Ceravolo G, Ceravolo R et al. 2009. The PRIAMO study: A multicenter assessment of nonmotor symptoms and their impact on quality of life in Parkinson's disease. *Mov Disord* **24**: 1641-1649.

Beach TG, Adler CH, Sue LI, Serrano G, Shill HA, Walker DG, Lue L, Roher AE, Dugger BN, Maarouf C et al. 2015. Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. *Neuropathology : official journal of the Japanese Society of Neuropathology* **35**: 354-389.

Betz RC, Planko L, Eigelshoven S, Hanneken S, Pasternack SM, Bussow H, Van Den Bogaert K, Wenzel J, Braun-Falco M, Rutten A et al. 2006. Loss-of-Function Mutations in the Keratin 5 Gene Lead to Dowling-Degos Disease. *The American Journal of Human Genetics* **78**.

Blencowe BJ, Issner R, Nickerson JA, Sharp PA. 1998. A coactivator of pre-mRNA splicing. *GENES & DEVELOPMENT* **12**.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* **34**: 525.

Burgh R, Meeldijk J, Jongeneel L, Frenkel J, Bovenschen N, Gijn M, Boes M. 2016. Reduced serpinB9-mediated caspase-1 inhibition can contribute to autoinflammatory disease. *Oncotarget* **7**.

Cermak S, Kosicek M, Mladenovic-Djordjevic A, Smiljanic K, Kanazir S, Hecimovic S. 2016. Loss of Cathepsin B and L Leads to Lysosomal Dysfunction, NPC-Like Cholesterol Sequestration and Accumulation of the Key Alzheimer's Proteins. *PLoS One* **11**: e0167428.

Chamcheu JC, Navsaria H, Pihl-Lundin I, Liovic M, Vahlquist A, Torma H. 2011. Chemical chaperones protect epidermolysis bullosa simplex keratinocytes from heat stress-induced keratin aggregation: involvement of heat shock proteins and MAP kinases. *The Journal of investigative dermatology* **131**: 1684-1691.

Chuang YH, Paul KC, Bronstein JM, Bordelon Y, Horvath S, Ritz B. 2017. Parkinson's disease is associated with DNA methylation levels in human blood and saliva. *Genome Med* **9**: 76.

Cocco C, D'Amato F, Noli B, Ledda A, Brancia C, Bongioanni P, Ferri GL. 2010. Distribution of VGF peptides in the human cortex and their selective changes in Parkinson's and Alzheimer's diseases. *Journal of anatomy* **217**: 683-693.

Codrich M, Bertuzzi M, Russo R, Francescatto M, Espinoza S, Zentilin L, Giacca M, Cesselli D, Beltrami AP, Ascenzi P et al. 2017. Neuronal hemoglobin affects dopaminergic cells' response to stress. *Cell Death &Amp; Disease* **8**: e2538.

Cooper-Knock J, Kirby J, Ferraiuolo L, Heath PR, Rattray M, Shaw PJ. 2012. Gene expression profiling in human neurodegenerative disease. *Nature reviews Neurology* **8**: 518-530.

Coupland KG, Mellick GD, Silburn PA, Mather K, Armstrong NJ, Sachdev PS, Brodaty H, Huang Y, Halliday GM, Hallupp M et al. 2014. DNA methylation of the MAPT gene in Parkinson's disease cohorts and modulation by vitamin E in vitro. *Mov Disord* **29**: 1606-1614.

Dauer W, Przedborski S. 2003. Parkinson's disease: mechanisms and models. *Neuron* **39**: 889-909.

Delenclos M, Jones DR, McLean PJ, Uitti RJ. 2016. Biomarkers in Parkinson's disease: Advances and strategies. *Parkinsonism & related disorders* **22 Suppl 1**: S106-110.

Deming Y, Dumitrescu L, Barnes LL, Thambisetty M, Kunkle B, Gifford KA, Bush WS, Chibnik LB, Mukherjee S, De Jager PL et al. 2018. Sex-specific genetic predictors of Alzheimer's disease biomarkers. *Acta Neuropathol* **136**: 857-872.

Denning GM, Ackermann LW, Barna TJ, Armstrong JG, Stoll LL, Weintraub NL, Dickson EW. 2008. Proenkephalin expression and enkephalin release are widely observed in non-neuronal tissues. *Peptides* **29**: 83-92.

Desai S, Juncker M, Kim C. 2018. Regulation of mitophagy by the ubiquitin pathway in neurodegenerative diseases. *Exp Biol Med (Maywood)* **243**: 554-562.

Desplats P, Spencer B, Coffee E, Patel P, Michael S, Patrick C, Adame A, Rockenstein E, Masliah E. 2011. Alpha-synuclein sequesters Dnmt1 from the nucleus: a novel mechanism for epigenetic alterations in Lewy body diseases. *The Journal of biological chemistry* **286**: 9031-9037.

Dhaenens CM, Tran H, Frandemiche ML, Carpentier C, Schraen-Maschke S, Sistiaga A, Goicoechea M, Eddarkaoui S, Van Brussels E, Obriot H et al. 2011. Mis-splicing of Tau exon 10 in myotonic dystrophy type 1 is reproduced by overexpression of CELF2 but not by MBNL1 silencing. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1812**: 732-742.

DiFrancesco JC, DiFrancesco D. 2015. Dysfunctional HCN ion channels in neurological diseases. *Front Cell Neurosci* **6**: 174.

Ding H, Dhima K, Lockhart KC, Locascio JJ, Hoesing AN, Duong K, Trisini-Lipsanopoulos A, Hayes MT, Sohur US, Wills AM et al. 2013. Unrecognized vitamin D3 deficiency is common in Parkinson disease: Harvard Biomarker Study. *Neurology* **81**: 1531-1537.

Dupré N, Howard H, Rouleau G. 2006. *Hereditary Motor and Sensory Neuropathy with Agenesis of the Corpus Callosum.*, Seattle (WA): University of Washington, Seattle.

Dursun E, Gezen-Ak D. 2017. Vitamin D receptor is present on the neuronal plasma membrane and is co-localized with amyloid precursor protein, ADAM10 or Nicastrin. *PLoS One* **12**: e0188605.

Fallon L, Belanger CM, Corera AT, Kontogiannea M, Regan-Klapisz E, Moreau F, Voortman J, Haber M, Rouleau G, Thorarinsdottir T et al. 2006. A regulated interaction with the UIM protein Eps15 implicates parkin in EGF receptor trafficking and PI(3)K-Akt signalling. *Nat Cell Biol* **8**: 834-842.

Ferrer I, Gómez A, Carmona M, Huesa G, Porta S, Riera-Codina M, Biagioli M, Gustincich S, Aso E. 2011a. Neuronal Hemoglobin is Reduced in Alzheimer's Disease, Argyrophilic Grain Disease, Parkinson's Disease, and Dementia with Lewy Bodies. *Journal of Alzheimer's Disease* **23**: 537-550.

Ferrer I, Martinez A, Blanco R, Dalfo E, Carmona M. 2011b. Neuropathology of sporadic Parkinson disease before the appearance of parkinsonism: preclinical Parkinson disease. *J Neural Transm (Vienna)* **118**: 821-839.

Ferri GL, Noli B, Brancia C, D'Amato F, Cocco C. 2011. VGF: an inducible gene product, precursor of a diverse array of neuro-endocrine peptides and tissue-specific disease biomarkers. *Journal of chemical neuroanatomy* **42**: 249-261.

Fife M, Gibson MS, Kaiser P. 2009. Identification of chicken granulocyte colony-stimulating factor (G-CSF/CSF3): the previously described myelomonocytic growth factor is actually CSF3. *Journal of Interferon & Cytokine Research* **29**: 339.

Fisher R. 1930. *Statistical methods for research workers*. Oliver & Boyd, Oxford, England.

Fortin JP, Triche TJ, Jr., Hansen KD. 2017. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics* **33**: 558-560.

Fraga MF, Esteller M. 2007. Epigenetics and aging: the targets and the marks. *Trends in genetics : TIG* **23**: 413-418.

Fu R-H, Liu S-P, Huang S-J, Chen H-J, Chen P-R, Lin Y-H, Ho Y-C, Chang W-L, Tsai C-H, Shyu W-C et al. 2013. Aberrant Alternative Splicing Events in Parkinson's Disease. *Cell Transplantation* **22**: 653-661.

Gardina PJ, Clark TA, Shimada B, Staples MK, Yang Q, Veitch J, Schweitzer A, Awad T, Sugnet C, Dee S et al. 2006. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC genomics* **7**: 325.

Giorgi FM, Del Fabbro C, Licausi F. 2013. Comparative study of RNA-seq- and microarray-derived coexpression networks in Arabidopsis thaliana. *Bioinformatics* **29**: 717-724.

Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, Counsell C, Giladi N, Holloway RG, Moore CG, Wenning GK et al. 2004. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov Disord* **19**: 1020-1028.

Goodall AH, Burns P, Salles I, Macaulay IC, Jones CI, Ardissino D, Bono Bd, Bray SL, Deckmyn H, Dudbridge F et al. 2010. Transcription profiling in human platelets reveals LRRFIP1 as a novel protein regulating platelet function. *PLATELETS AND THROMBOPOIESIS* **116**.

Gray EH, De Vos KJ, Dingwall C, Perkinton MS, Miller CC. 2010. Deficiency of the copper chaperone for superoxide dismutase increases amyloid-beta production. *J Alzheimers Dis* **21**: 1101-1105.

Guhathakurta S, Evangelista BA, Ghosh S, Basu S, Kim YS. 2017. Hypomethylation of intron1 of alpha-synuclein gene does not correlate with Parkinson's disease. *Mol Brain* **10**: 6.

Haas BR, Stewart TH, Zhang J. 2012. Premotor biomarkers for Parkinson's disease - a promising direction of research. *Transl Neurodegener* **1**: 11.

Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Yearout D, Kay DM, Doheny KF, Paschall J, Pugh E et al. 2010. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nature genetics* **42**: 781-785.

Hawkes CH, Del Tredici K, Braak H. 2009. Parkinson's disease: the dual hit theory revisited. *Annals of the New York Academy of Sciences* **1170**: 615-622.

Heiss JA, Just AC. 2018. Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO. *Clin Epigenetics* **10**: 73.

Helley MP, Pinnell J, Sportelli C, Tieu K. 2017. Mitochondria: A Common Target for Genetic Mutations and Environmental Toxicants in Parkinson's Disease. *Front Genet* **8**: 177.

Henderson-Smith A, Corneveaux JJ, De Both M, Cuyugan L, Liang WS, Huentelman M, Adler C, Driver-Dunckley E, Beach TG, Dunckley TL. 2016. Next-generation profiling to identify the molecular etiology of Parkinson dementia. *Neurol Genet* **2**: e75.

Hochstrasser T, Weiss E, Marksteiner J, Humpel C. 2010. Soluble cell adhesion molecules in monocytes of Alzheimer's disease and mild cognitive impairment. *Experimental gerontology* **45**: 70-74.

Hodgkinson KM, Vanderhyden BC. 2014. Consideration of GREB1 as a potential therapeutic target for hormone-responsive or endocrine-resistant cancers. *Expert Opinion on Therapeutic Targets* **18**: 1065-1076.

Holden SK, Finseth T, Sillau SH, Berman BD. 2018. Progression of MDS-UPDRS Scores Over Five Years in De Novo Parkinson Disease from the Parkinson's Progression Markers Initiative Cohort. *Mov Disord Clin Pract* **5**: 47-53.

Horvath S. 2013. DNA methylation age of human tissues and cell types. *Genome Biol* **14**: R115.

Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA. 2012. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome biology* **13**: R97.

Huang B, Yang XD, Zhou MM, Ozato K, Chen LF. 2009. Brd4 coactivates transcriptional activation of NF-kappaB via specific binding to acetylated RelA. *Molecular and cellular biology* **29**: 1375-1387.

Im E, Yoo L, Hyun M, Shin Woo H, Chung Kwang C. Covalent ISG15 conjugation positively regulates the ubiquitin E3 ligase activity of parkin. *Open Biology* **6**: 160193.

Imai Y, Kobayashi Y, Inoshita T, Meng H, Arano T, Uemura K, Asano T, Yoshimi K, Zhang C-L, Matsumoto G et al. 2015. The Parkinson's Disease-Associated Protein Kinase LRRK2 Modulates Notch Signaling through the Endosomal Pathway. *PLOS Genetics* **11**: e1005503.

Ionov ID, Pushinskaya, II. 2013. Somatostatin antagonist induces catalepsy in the aged rat. *Psychopharmacology* **227**: 273-276.

Ishihara K, Yasuda K, Hatayama T. 1999. Molecular cloning, expression and localization of human 105 kDa heat shock protein, hsp105. *Biochimica et biophysica acta* **1444**: 138-142.

Jaffe AE, Irizarry RA. 2014. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* **15**: R31.

Jahn H, Wittke S, Zurbig P, Raedler TJ, Arlt S, Kellmann M, Mullen W, Eichenlaub M, Mischak H, Wiedemann K. 2011. Peptide Fingerprinting of Alzheimer's Disease in Cerebrospinal Fluid: Identification and Prospective Evaluation of New Synaptic Biomarkers. *PLoS One* **6**.

Jansen IE, Ye H, Heetveld S, Lechler MC, Michels H, Seinstra RI, Lubbe SJ, Drouet V, Lesage S, Majounie E et al. 2017. Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing. *Genome Biol* **18**: 22.

Jeong K, Kwon HY, Jeong MS, Sohn EJ, Kim S-H. 2017. CNOT2 promotes degradation of p62/SQSTM1 as a negative regulator in ATG5 dependent autophagy. *Oncotarget* **8**: 46034-46046.

Jiao Y, Widschwendter M, Teschendorff AE. 2014. A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. *Bioinformatics* **30**: 2360-2366.

Jowaed A, Schmitt I, Kaut O, Wullner U. 2010. Methylation regulates alpha-synuclein expression and is decreased in Parkinson's disease patients' brains. *J Neurosci* **30**: 6355-6359.

Kaeser-Woo YJ, Younts TJ, Yang X, Zhou P, Wu D, Castillo PE, Südhof TC. 2013. Synaptotagmin-12 Phosphorylation by cAMP-Dependent Protein Kinase Is Essential for Hippocampal Mossy Fiber LTP. *The Journal of Neuroscience* **33**: 9769.

Katkere B, Rosa S, Caballero A, Repasky EA, Drake JR. 2010. Physiological-range temperature changes modulate cognate antigen processing and presentation mediated by lipid raft-restricted ubiquitinated B cell receptor molecules. *Journal of immunology* **185**: 5032-5039.

Kaut O, Schmitt I, Wullner U. 2012. Genome-scale methylation analysis of Parkinson's disease patients' brains reveals DNA hypomethylation and increased mRNA expression of cytochrome P450 2E1. *Neurogenetics* **13**: 87-91.

Kelly MP. 2018. Cyclic nucleotide signaling changes associated with normal aging and age-related diseases of the brain. *Cellular signalling* **42**: 281-291.

Kettwig M, Ohlenbusch A, Jung K, Steinfeld R, Gartner J. 2018. Cathepsin D Polymorphism C224T in Childhood-Onset Neurodegenerative Disorders: No Impact for Childhood Dementia. *J Pediatr Genet* **7**: 14-18.

Kinsella S, Fichtner M, Watters O, König H-G, Prehn JHM. 2018. Increased A20-E3 ubiquitin ligase interactions in bid-deficient glia attenuate TLR3- and TLR4-induced inflammation. *Journal of Neuroinflammation* **15**: 130.

Lee CA, Chin L-S, Li L. 2018. Hypertonia-linked protein Trak1 functions with mitofusins to promote mitochondrial tethering and fusion. *Protein & cell* **9**: 693-716.

Li WQ, Yu HY, Li YM, Wang X, He J, Yan HZ, Yang DH, Wu XJ, Hou LJ, Liu HM et al. 2014. Higher LRRFIP1 expression in glioblastoma multiforme is associated with better response to teniposide, a type II topoisomerase inhibitor. *Biochem Biophys Res Commun* **446**: 1261-1267.

Lin Q, Ding H, Zheng Z, Gu Z, Ma J, Chen L, Chan P, Cai Y. 2012. Promoter methylation analysis of seven clock genes in Parkinson's disease. *Neurosci Lett* **507**: 147-150.

Liu G, Boot B, Locascio JJ, Jansen IE, Winder-Rhodes S, Eberly S, Elbaz A, Brice A, Ravina B, van Hilten JJ et al. 2016. Specifically neuropathic Gaucher's mutations accelerate cognitive decline in Parkinson's. *Annals of neurology* **80**: 674-685.

Liu G, Locascio JJ, Corvol JC, Boot B, Liao Z, Page K, Franco D, Burke K, Jansen IE, Trisini-Lipsanopoulos A et al. 2017. Prediction of cognition in Parkinson's disease with a clinical-genetic score: a longitudinal analysis of nine cohorts. *Lancet Neurol* **16**: 620-629.

Locascio JJ, Eberly S, Liao Z, Liu G, Hoesing AN, Duong K, Trisini-Lipsanopoulos A, Dhima K, Hung AY, Flaherty AW et al. 2015. Association between alpha-synuclein blood transcripts and early, neuroimaging-supported Parkinson's disease. *Brain* **138**: 2659-2671.

Loraine AE, McCormick S, Estrada A, Patel K, Qin P. 2013. RNA-Seq of Arabidopsis Pollen Uncovers Novel Transcription and Alternative Splicing. *Plant physiology* **162**: 1092-1109.

Lovkvist C, Dodd IB, Sneppen K, Haerter JO. 2016. DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic acids research* **44**: 5123-5132.

Lu JQ, Stoessl AJ. 2002. Somatostatin modulates the behavioral effects of dopamine receptor activation in parkinsonian rats. *Neuroscience* **112**: 261-266.

Ma S, Attarwala I, Xie X-Q. 2019. SQSTM1/p62: a potential target for neurodegenerative disease. *ACS Chemical Neuroscience* doi:10.1021/acschemneuro.8b00516.

Magold AI, Cacquevel M, Fraering PC. 2009. Gene Expression Profiling in Cells with Enhanced γ-Secretase Activity. *PLoS One* **4**: e6952.

Martínez-López MJ, Alcántara S, Mascaró C, Pérez-Brangulí F, Ruiz-Lozano P, Maes T, Soriano E, Buesa C. 2005. Mouse Neuron navigator 1, a novel microtubule-associated protein involved in neuronal migration. *Molecular and Cellular Neuroscience* **28**: 599-612.

Masliah E, Dumaop W, Galasko D, Desplats P. 2013. Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes. *Epigenetics* **8**: 1030-1038.

Matsumoto L, Takuma H, Tamaoka A, Kurisaki H, Date H, Tsuji S, Iwata A. 2010. CpG demethylation enhances alpha-synuclein expression and affects the pathogenesis of Parkinson's disease. *PLoS One* **5**: e15522.

Mattila PM, Rinne JO, Helenius H, Dickson DW, Roytta M. 2000. Alpha-synuclein-immunoreactive cortical Lewy bodies are associated with cognitive impairment in Parkinson's disease. *Acta Neuropathol* **100**: 285-290.

Mayeux R, Stern Y. 2012. Epidemiology of Alzheimer disease. *Cold Spring Harb Perspect Med* **2**.

Meares GP, Zmijewska AA, Jope RS. 2008. HSP105 interacts with GRP78 and GSK3 and promotes ER stress-induced caspase-3 activation. *Cellular signalling* **20**: 347-358.

Mehta SH, Adler CH. 2016. Advances in Biomarker Research in Parkinson's Disease. *Current neurology and neuroscience reports* **16**: 7.

Messerschmidt DM, Knowles BB, Solter D. 2014. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev* **28**: 812-828.

Milde S, Coleman MP. 2014. Identification of palmitoyltransferase and thioesterase enzymes that control the subcellular localization of axon survival factor nicotinamide mononucleotide adenylyltransferase 2 (NMNAT2). *J Biol Chem* **289**: 32858-32870.

Miller RM, Kiser GL, Kaysser-Kranich TM, Lockner RJ, Palaniappan C, Federoff HJ. 2006. Robust dysregulation of gene expression in substantia nigra and striatum in Parkinson's disease. *Neurobiology of disease* **21**: 305-313.

Mills JD, Janitz M. 2012. Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases. *Neurobiology of aging* **33**: 1012 e1011-1024.

Minoia M, Grit C, Kampinga HH. 2014. HSPA1A-Independent Suppression of PARK2 C289G Protein Aggregation by Human Small Heat Shock Proteins. *Molecular and cellular biology* **34**: 3570.

Minones-Moyano E, Friedlander MR, Pallares J, Kagerbauer B, Porta S, Escaramis G, Ferrer I, Estivill X, Marti E. 2013. Upregulation of a small vault RNA (svtRNA2-1a) is an early event in Parkinson disease and induces neuronal dysfunction. *RNA Biol* **10**: 1093-1106.

Minones-Moyano E, Porta S, Escaramis G, Rabionet R, Iraola S, Kagerbauer B, Espinosa-Parrilla Y, Ferrer I, Estivill X, Marti E. 2011. MicroRNA profiling of Parkinson's disease brains identifies early downregulation of miR-34b/c which modulate mitochondrial function. *Human molecular genetics* **20**: 3067-3078.

Mkhikian H, Grigorian A, Li CF, Chen HL, Newton B, Zhou RW, Beeton C, Torossian S, Tatarian GG, Lee SU et al. 2011. Genetics and the environment converge to dysregulate N-glycosylation in multiple sclerosis. *Nature communications* **2**: 334.

Mohamed WA, Salama RM, Schaalan MF. 2019. A pilot study on the effect of lactoferrin on Alzheimer's disease pathological sequelae: Impact of the p-Akt/PTEN pathway. *Biomed Pharmacother* **111**: 714-723.

Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, Morris Q, Bader GD. 2010. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* **26**: 2927-2928.

Moore K, McKnight AJ, Craig D, O'Neill F. 2014. Epigenome-wide association study for Parkinson's disease. *Neuromolecular Med* **16**: 845-855.

Moore MJ, Silver PA. 2008. Global analysis of mRNA splicing. *Rna* **14**: 197-203.

Mutez E, Larvor L, Lepretre F, Mouroux V, Hamalek D, Kerckaert JP, Perez-Tur J, Waucquier N, Vanbesien-Mailliot C, Duflot A et al. 2011. Transcriptional profile of Parkinson blood mononuclear cells with LRRK2 mutation. *Neurobiology of aging* **32**: 1839-1848.

Neuner SM, Heuer SE, Zhang J-G, Philip VM, Kaczorowski CC. 2019. Identification of Pre-symptomatic Gene Signatures That Predict Resilience to Cognitive Decline in the Genetically Diverse AD-BXD Model. *Front Genet* **10**: 35.

Nonis D, Schmidt MH, van de Loo S, Eich F, Dikic I, Nowock J, Auburger G. 2008. Ataxin-2 associates with the endocytosis complex and affects EGF receptor trafficking. *Cellular signalling* **20**: 1725-1739.

Nookaew I, Papini M, Pornputtapong N, Scalcinati G, Fagerberg L, Uhlen M, Nielsen J. 2012. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae. *Nucleic acids research* **40**: 10084-10097.

Okazaki S, Boku S, Otsuka I, Mouri K, Aoyama S, Shiroiwa K, Sora I, Fujita A, Shirai Y, Shirakawa O et al. 2016. The cell cycle-related genes as biomarkers for schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* **70**: 85-91.

Olanow CW, Schapira AH. 2013. Therapeutic prospects for Parkinson disease. *Annals of neurology* **74**: 337-347.

Ozer B, Sezerman OU. 2015. A novel analysis strategy for integrating methylation and expression data reveals core pathways for thyroid cancer aetiology. *BMC genomics* **16 Suppl 12**: S7.

Pangalos MN, Neefs JM, Somers M, Verhasselt P, Bekkers M, van der Helm L, Fraiponts E, Ashton D, Gordon RD. 1999. Isolation and expression of novel human glutamate carboxypeptidases with N-acetylated alpha-linked acidic dipeptidase and dipeptidyl peptidase IV activity. *J Biol Chem* **274**: 8470-8483.

Panopoulos AD, Watowich SS. 2008. Granulocyte colony-stimulating factor: molecular mechanisms of action during steady state and 'emergency' hematopoiesis. *Cytokine* **42**: 277-288.

Pavlopoulos E, Jones S, Kosmidis S, Close M, Kim C, Kovalerchik O, Small SA, Kandel ER. 2013. Molecular Mechanism for Age-Related Memory Loss: The Histone-Binding Protein RbAp48. *Science Translational Medicine* **5**: 200ra115.

Pavlou MAS, Outeiro TF. 2017. Epigenetics in Parkinson's Disease. *Adv Exp Med Biol* **978**: 363-390.

Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, R VL, Clark SJ, Molloy PL. 2015a. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* **8**: 6.

Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, V Lord R, Clark SJ, Molloy PL. 2015b. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* **8**: 6.

Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Djik S, Muhlhausler B, Stirzaker C, Clark SJ. 2016. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome biology* **17**: 208.

Planken A, Kurvits L, Reimann E, Kadastik-Eerme L, Kingo K, Koks S, Taba P. 2017. Looking beyond the brain to improve the pathogenic understanding of Parkinson's disease: implications of whole transcriptome profiling of Patients' skin. *BMC Neurol* **17**: 6.

Rathore N, Ramani SR, Pantua H, Payandeh J, Bhangale T, Wuster A, Kapoor M, Sun Y, Kapadia SB, Gonzalez L et al. 2018. Paired Immunoglobulin-like Type 2 Receptor Alpha G78R variant alters ligand binding and confers protection to Alzheimer's disease. *PLOS Genetics* **14**: e1007427.

Rentzos M, Michalopoulou M, Nikolaou C, Cambouri C, Rombos A, Dimitrakopoulos A, Vassilopoulos D. 2005. The role of soluble intercellular adhesion molecules in neurodegenerative disorders. *Journal of the neurological sciences* **228**: 129-135.

Richter J, Appenzeller S, Ammerpohl O, Deuschl G, Paschen S, Bruggemann N, Klein C, Kuhlenbaumer G. 2012. No evidence for differential methylation of alpha-synuclein in leukocyte DNA of Parkinson's disease patients. *Mov Disord* **27**: 590-591.

Rikiyama T, Curtis J, Oikawa M, Zimonjic DB, Popescu N, Murphy BA, Wilson MA, Johnson AC. 2003. GCF2: expression and molecular analysis of repression. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1629**: 15-25.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**: e47-e47.

Robak LA, Jansen IE, van Rooij J, Uitterlinden AG, Kraaij R, Jankovic J, International Parkinson's Disease Genomics C, Heutink P, Shulman JM. 2017. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain* **140**: 3191-3203.

Romanelli V, Nakabayashi K, Vizoso M, Moran S, Iglesias-Platas I, Sugahara N, Simon C, Hata K, Esteller M, Court F et al. 2014. Variable maternal methylation overlapping the nc886/vtRNA2-1 locus is locked between hypermethylated repeats and is frequently altered in cancer. *Epigenetics* **9**: 783-790.

Ryan MC, Cleland J, Kim R, Wong WC, Weinstein JN. 2012. SpliceSeq: a resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics* **28**: 2385-2387.

Saiki S, Hatano T, Fujimaki M, Ishikawa KI, Mori A, Oji Y, Okuzumi A, Fukuhara T, Koinuma T, Imamichi Y et al. 2017. Decreased long-chain acylcarnitines from insufficient beta-oxidation as potential early diagnostic markers for Parkinson's disease. *Sci Rep* **7**: 7328.

Saito Y, Yamagishi N, Ishihara K, Hatayama T. 2003. Identification of α-tubulin as an hsp105α-binding protein by the yeast two-hybrid system. *Experimental Cell Research* **286**: 233-240.

Sanchez-Mut JV, Aso E, Heyn H, Matsuda T, Bock C, Ferrer I, Esteller M. 2014. Promoter hypermethylation of the phosphatase DUSP22 mediates PKA-dependent TAU phosphorylation and CREB activation in Alzheimer's disease. *Hippocampus* **24**: 363-368.

Sanchez-Pla A, Reverter F, Ruiz de Villa MC, Comabella M. 2012. Transcriptomics: mRNA and alternative splicing. *Journal of neuroimmunology* **248**: 23-31.

Sanders SS, Parsons MP, Mui KKN, Southwell AL, Franciosi S, Cheung D, Waltl S, Raymond LA, Hayden MR. 2016. Sudden death due to paralysis and synaptic and behavioral deficits when Hip14/Zdhhc17 is deleted in adult mice. *BMC Biology* **14**: 108.

Sang M, Ma L, Sang M, Zhou X, Gao W, Geng C. 2014. LIM-domain-only proteins: multifunctional nuclear transcription coregulators that interacts with diverse proteins. *Molecular Biology Reports* **41**: 1067-1073.

Schmitt I, Kaut O, Khazneh H, deBoni L, Ahmad A, Berg D, Klein C, Frohlich H, Wullner U. 2015. L-dopa increases alpha-synuclein DNA methylation in Parkinson's disease patients in vivo and in vitro. *Mov Disord* **30**: 1794-1801.

Schneider A, Kuhn HG, Schäbitz WR. 2005. A Role for G-CSF (Granulocyte-Colony Stimulating Factor) in the Central Nervous System. *Cell Cycle* **4**: 1753-1757.

Shabalin AA. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics (Oxford, England)* **28**: 1353-1358.

Sharif-Askari E, Vassen L, Kosan C, Khandanpour C, Gaudreau MC, Heyd F, Okayama T, Jin J, Rojas ME, Grimes HL et al. 2010. Zinc finger protein Gfi1 controls the endotoxin-mediated Toll-like receptor inflammatory response by antagonizing NF-kappaB p65. *Molecular and cellular biology* **30**: 3929-3942.

Shi M, Cho H, Inn KS, Yang A, Zhao Z, Liang Q, Versteeg GA, Amini-Bavil-Olyaee S, Wong LY, Zlokovic BV et al. 2014. Negative regulation of NF-kappaB activity by brain-specific TRIpartite Motif protein 9. *Nature communications* **5**: 4820.

Shoemaker DD, E. E. Schadt*, C. D. Armour, Y. D. He, P. Garrett-Engele, P. D. McDonagh, P. M. Loerch, A. Leonardson, P. Y. Lum, G. Cavet et al. 2001. Experimental annotation of the human genome using microarray technology. *Nature* **409**.

Short KM, Cox TC. 2006. Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. *J Biol Chem* **281**: 8970-8980.

Siegert R, Leroux MR, Scheufler C, Hartl FU, Moarefi I. 2000. Structure of the molecular chaperone prefoldin: unique interaction of multiple coiled coil tentacles with unfolded proteins. *Cell* **103**: 621-632.

Singh Y, Gupta G, Shrivastava B, Dahiya R, Tiwari J, Ashwathanarayana M, Sharma RK, Agrawal M, Mishra A, Dua K. 2017. Calcitonin gene-related peptide (CGRP): A novel target for Alzheimer's disease. *CNS Neurosci Ther* **23**: 457-461.

Sivandzade F, Prasad S, Bhalerao A, Cucullo L. 2019. NRF2 and NF-B interplay in cerebrovascular and neurodegenerative disorders: Molecular mechanisms and possible therapeutic approaches. *Redox Biol* **21**: 101059.

Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M. 1997. Alpha-synuclein in Lewy bodies. *Nature* **388**: 839-840.

Stamper C, Siegel A, Liang WS, Pearson JV, Stephan DA, Shill H, Connor D, Caviness JN, Sabbagh M, Beach TG et al. 2008. Neuronal gene expression correlates of Parkinson's disease with dementia. *Mov Disord* **23**: 1588-1595.

Su X, Chu Y, Kordower JH, Li B, Cao H, Huang L, Nishida M, Song L, Wang D, Federoff HJ. 2015. PGC-1alpha Promoter Methylation in Parkinson's Disease. *PLoS One* **10**: e0134087.

Sun H, Liu Y, Zhang L, Shao X, Liu K, Ding Z, Liu X, Jiang C, Li H, Li H. 2017. Numb positively regulates autophagic flux via regulating lysosomal function. *Biochem Biophys Res Commun* **491**: 780-786.

Tanji K, Kamitani T, Mori F, Kakita A, Takahashi H, Wakabayashi K. 2010. TRIM9, a novel brain-specific E3 ubiquitin ligase, is repressed in the brain of Parkinson's disease and dementia with Lewy bodies. *Neurobiology of disease* **38**: 210-218.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* **31**: 46-53.

Tysnes OB, Storstein A. 2017. Epidemiology of Parkinson's disease. *J Neural Transm (Vienna)* **124**: 901-905.

Vallés AS, Borroni MV, Barrantes FJ. 2014. Targeting Brain α7 Nicotinic Acetylcholine Receptors in Alzheimer's Disease: Rationale and Current Status. *CNS Drugs* **28**: 975-987.

Vardy ER, Kellett KA, Cocklin SL, Hooper NM. 2012. Alkaline phosphatase is increased in both brain and plasma in Alzheimer's disease. *Neurodegener Dis* **9**: 31-37.

Vélez JI, Lopera F, Creagh PK, Piñeros LB, Das D, Cervantes-Henríquez ML, Acosta-López JE, Isaza-Ruget MA, Espinosa LG, Easteal S et al. 2018. Targeting Neuroplasticity, Cardiovascular, and Cognitive-Associated Genomic Variants in Familial Alzheimer's Disease. *Molecular Neurobiology* doi:10.1007/s12035-018-1298-z.

Wagner S, Chiosea S, Nickerson JA. 2003. The spatial targeting and nuclear matrix binding domains of SRm160. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 3269-3274.

Winkler A. 2012. The logic of the Fisher method to combine P-values.

Wockner LF, Morris CP, Noble EP, Lawford BR, Whitehall VL, Young RM, Voisey J. 2015. Brain-specific epigenetic markers of schizophrenia. *Transl Psychiatry* **5**: e680.

Wood SH, Craig T, Li Y, Merry B, de Magalhaes JP. 2013. Whole transcriptome sequencing of the aging rat brain reveals dynamic RNA changes in the dark matter of the genome. *Age* **35**: 763-776.

Xu S, Di Z, He Y, Wang R, Ma Y, Sun R, Li J, Wang T, Shen Y, Fang S et al. 2019. Mesencephalic astrocyte-derived neurotrophic factor (MANF) protects against Abeta toxicity via attenuating Abeta-induced endoplasmic reticulum stress. *J Neuroinflammation* **16**: 35.

Yakunin E, Moser A, Loeb V, Saada A, Faust P, Crane DI, Baes M, Sharon R. 2010. alpha-Synuclein abnormalities in mouse models of peroxisome biogenesis disorders. *Journal of neuroscience research* **88**: 866-876.

Yang X, Ren H, Wood K, Li M, Qiu S, Shi F-D, Ma C, Liu Q. 2018. Depletion of microglia augments the dopaminergic neurotoxicity of MPTP. *The FASEB Journal* **32**: 3336-3345.

Yang XY, Zhao EY, Zhuang WX, Sun FX, Han HL, Han HR, Lin ZJ, Pan ZF, Qu MH, Zeng XW et al. 2015. LPA signaling is required for dopaminergic neuron development and is reduced through low expression of the LPA1 receptor in a 6-OHDA lesion model of Parkinson's disease. *Neurol Sci* **36**: 2027-2033.

Yik WY, Steinberg SJ, Moser AB, Moser HW, Hacia JG. 2009. Identification of novel mutations and sequence variation in the Zellweger syndrome spectrum of peroxisome biogenesis disorders. *Human Mutation* **30**: E467-E480.

Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA, Zhang C, Xie T, Tran L, Dobrin R et al. 2013. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**: 707-720.

Zhang C, Browne A, Child D, Divito JR, Stevenson JA, Tanzi RE. 2010. Loss of function of ATXN1 increases amyloid beta-protein levels by potentiating beta-secretase processing of beta-amyloid precursor protein. *J Biol Chem* **285**: 8515-8526.

Zhang M, Wang Y, Qian F, Li P, Xu X. 2016. Hypericin inhibits oligomeric amyloid β42-induced inflammation response in microglia and ameliorates cognitive deficits in an amyloid β injection mouse model of Alzheimer's disease by suppressing MKL1. *Biochemical and Biophysical Research Communications* **481**: 71-76.

Zhao Z, Lange DJ, Ho L, Bonini S, Shao B, Salton SR, Thomas S, Pasinetti GM. 2008. Vgf is a novel biomarker associated with muscle weakness in amyotrophic lateral sclerosis (ALS), with a potential role in disease pathogenesis. *International Journal of Medical Sciences* **5**: 92-99.