

**Title: Characterization of baseline and longitudinal DNA Methylation in patients with sporadic Parkinson's disease**

**Running head: DNA methylation in Parkinson's Disease**

**Authors:** Paulina Gonzalez-Latapi<sup>1</sup> MD, MS; Bernabe Bustos<sup>1</sup>, PhD; Siyuan Dong<sup>2</sup> MS; Steven Lubbe<sup>1</sup> PhD; Tanya Simuni<sup>1</sup> MD; Dimitri Krainc<sup>1</sup> MD, PhD

1. Department of Neurology, Northwestern University Feinberg School of Medicine
2. Biostatistics Collaboration Center, Northwestern University Feinberg School of Medicine

*Corresponding author:*

Paulina Gonzalez-Latapi MD, MSc

Email: [paulina.latapi@northwestern.edu](mailto:paulina.latapi@northwestern.edu)

Address: 710 N Lake Shore Drive, Ste 1120, Chicago, IL, 60611

**Word count:**

**Abstract:** 248 words

**Introduction:** 357 words

**Discussion:** 1080 words

**Body:** 2594 words

**References:** 40

**Tables:** 4

**Figures:** 4

**Keywords:** Parkinson's Disease, DNA methylation, PPMI

## Abstract

**Objective:** To characterize DNA methylation differences between sporadic Parkinson's Disease and healthy control individuals enrolled in the Parkinson's Progression Markers Initiative.

**Methods:** We characterized cross-sectional and longitudinal DNA methylation differences between individuals with sporadic (i.e., non-genetic) PD and healthy controls. We included 282 individuals (196 Parkinson's Disease individuals and 86 healthy control individuals). DNA methylation data was collected at the time of enrollment and longitudinally over three years.

**Results:** This analysis revealed 81,604 differentially methylated positions and 5,281 differentially methylated regions between sporadic PD and healthy controls. Gene ontology analysis revealed that these differentially methylated positions and regions were associated with genes involved in diverse cellular processes, including several with specific functions in the brain ("Focal adhesion", "Cholinergic synapse", "Glutamatergic synapse", "Dopaminergic synapse"). Integration of both differentially methylated sites and expressed genes showed 20 genes that were hypomethylated and overexpressed and one gene, CTSB that was hypermethylated and associated with reduced expression.

**Interpretation of Results:** Our study provides evidence that alterations in the methylome in Parkinson's Disease are discernible in blood, evolve over time, and reflect cellular processes linked to ongoing neurodegeneration. These findings lend support to the potential of blood DNA methylation as an epigenetic biomarker for Parkinson's Disease. To fully comprehend DNA methylation changes throughout the progression of Parkinson's Disease, additional profiling at longer intervals and during the prodromal stage will be necessary.

## Introduction

Parkinson's Disease (PD) is the second most common neurodegenerative disorder, and its prevalence is expected to double by 2040<sup>1,2</sup>. Although various genetic causes of PD have been identified, genetic factors account for only a minority of disease heritability<sup>3</sup>. Accumulating epidemiological evidence has shown that environmental factors can increase the risk of PD<sup>4</sup>, but not everyone responds in the same way to environment. Presently, it appears that the complex interplay between genetic and environmental factors underlies the pathogenesis of PD<sup>5</sup>. Some of the features of PD risk may be reflected in the epigenome, which serves as an interface between genetic and environmental risk factors.

Epigenetic modifications do not affect the DNA sequence and can result in alterations in gene expression. DNA methylation (DNAm) is the most widely studied epigenetic phenomenon<sup>6-8</sup>. As such, there is growing interest in exploring the role of DNAm in PD. Some studies suggest that methylation patterns in the blood of PD patients differ between cases and controls and may be related to the onset of PD<sup>8-10</sup>. However, most of these studies have been conducted in heterogeneous or small cohorts. Therefore, there is a need for large-scale studies involving DNA samples from deeply clinically and biologically characterized PD cases and suitable controls, with rigorous collection and storage protocols.

The Parkinson's Progression Markers Initiative (PPMI) is an observational, international, multicenter study designed to establish biomarker-defined cohorts and identify clinical, imaging, genetic, and biospecimen (e.g., blood sample, cerebrospinal fluid) markers of PD progression. PPMI was established in 2010, and participants have been followed longitudinally, with clinical and biological samples, including DNAm, collected on a yearly basis<sup>11</sup>. Longitudinal studies with multiple time points where methylation data was assessed, could provide novel insights into methylation trajectories in disease<sup>12</sup>.

The aim of this study was to characterize DNAm differences between individuals with sporadic PD (i.e. non-genetic) and healthy controls (HC) enrolled in the PPMI study. Importantly, we report on cross-sectional analyses as well as longitudinal DNAm changes in these individuals. Additionally, we provide initial data on the use of DNAm to develop an epigenetic signature to differentiate sPD and HC individuals.

## Methods

### *Study cohort*

Data were downloaded from PPMI LONI database (December 13, 2021). The aims and methodology of PPMI have been published<sup>11</sup>. For this study, we used the analytical dataset for sporadic Parkinson Disease (sPD) and HC participants. All study participants were enrolled at the initial stage of PPMI. Importantly, sPD individuals were treatment naïve and within two years of diagnosis at enrollment. Individuals with a variant in a known PD gene were excluded from this analysis. Individuals with a known condition that could impact DNAm (i.e. diagnosis of a blood or bleeding disorders, past medical history of cancer, autoimmune conditions and inflammatory bowel disease) were also excluded from our analysis.

### *DNA methylation analysis*

For each study participant, DNAm data of ~ 850,000 CpG-sites were obtained from the Infinium MethylationEPIC array (Illumina) generated using bisulfite converted blood DNA. Raw IDAT files were available from PPMI. The *ChAMP* pipeline<sup>13, 14</sup> for Infinium MethylationEPIC BeadChip (Illumina) was used for all analyses, following the default workflow. Raw IDAT files were exported for processing in R (version 4.1). While  $\beta$  values were used for interpretation of the results,  $M$  values [ $M$  value =  $\log_2(\text{methylated intensity}/\text{unmethylated intensity})$ ] were used for

the differential methylation analysis<sup>15</sup>. Quality control and filtering steps were applied to exclude probes with detection  $p$ -value  $> 0.01$ , probes with  $< 3$  beads in at least 5% of samples, non-CpG probes, probes that fall in single-nucleotide polymorphism (SNP), multi-hit probes and probes located in sex chromosomes<sup>16</sup>. Data were normalized with BMIQ Method<sup>17</sup> and batch effects were corrected by *ComBat*<sup>18</sup>. Cell type influence on the whole blood data was corrected with *RefbaseEWAS*<sup>19</sup>. Differentially methylated positions (DMPs) were detected using *limma*<sup>20</sup>. For the longitudinal analysis, we fit linear regression models with cluster-robust standard errors using the `robust_se()` function in the R *sandwich* package. Age, disease duration, sex, blood cell composition and levodopa equivalent daily dose (LEDD) were included as covariates in these analyses.

Differentially methylated regions (DMRs) were detected using the *DMRcate*<sup>21</sup>. A bandwidth of 1000 nucleotides ( $\lambda = 1000$ ) and a scaling factor of 2 ( $C = 2$ ) were used as recommended by the *DMRcate* authors. An adjusted (FDR)  $p$  value  $< 0.05$  was considered statistically significant for all methylation analyses.

We will have over 80% power at 70% of methylation sites to detect a difference in methylation at significance level  $9e-8$ , as recommended for the EPIC array<sup>22</sup>.

#### *RNASeq analyses*

Differential gene expression analyses were performed only for the genes that were mapped to or near ( $\pm 1000$  base pairs) to the DMRs. Blood gene expression was obtained from the PPMI RNASeq raw read counts. Differential expression analysis was done using the R package *DESeq2*<sup>23</sup>, and an adjusted FDR  $p$  value  $< 0.05$  was used to select significantly differentially expressed genes.

## **Results**

Our study included a total of 282 participants from the PPMI cohort. Among these, 196 individuals belonged to the sPD group, with 34% being female, mean age of 63.31 years, and a mean disease duration from diagnosis of 0.5 years at enrollment. Additionally, 86 healthy control participants were included, with 32% being female and a mean age of 63.95 at enrollment. There were no significant differences in age at DNA collection, sex ratio, or racial composition between the sPD and healthy control groups (**Table 1**).

Baseline and longitudinal clinical characteristics are presented in **Table 2**. Individuals with sPD exhibited significantly lower Montreal Cognitive Assessment (MoCA) scores compared to HC, although their scores still fell within the range for normal cognition. As expected, both motor and non-motor scores were significantly worse in sPD individuals, and these scores declined further during follow-up. None of the sPD participants were taking any dopaminergic medication at baseline.

### **Blood cell composition differs between sPD and healthy controls**

Given that we used whole blood DNA for methylation profiling, we considered the potential impact of differential lymphocyte cell type distributions on our analysis. To address this, we utilized distinctive cell-specific methylation profiles to estimate the proportional abundance of blood cell types in our samples based on the EPIC array's specific probes. At baseline, the sPD group exhibited a significantly higher proportion of granulocytes compared to HC ( $t$  test: 0.64 vs. 0.62,  $p < 0.0001$ ), as shown in **Fig1**. This difference persisted during follow-up. Additionally, at 3-year follow up the proportion of CD8+ and CD4+ cells decreased in the sPD group and was significantly lower compared to the HC group (0.051 vs 0.060,  $p=0.041$  for CD8+; 0.127 vs 0.148,  $p = 0.0001$  for CD4+) (**Fig1**).

## DNA Methylation differs between sPD and healthy controls

We conducted a cross-sectional analysis by comparing methylation profiles between sPD cases and HC at enrollment. This analysis revealed 81,604 differentially methylated sites between PD and HC. Applying a cutoff of a 2% absolute change in average methylation ( $\Delta\beta$ ), resulted in 5,183 retained positions. The top differentially methylated positions are presented in **Table 3**. PD individuals exhibited hypermethylation in 2,683 DMPs and hypomethylation in 2,495 DMPs (**Fig2A**). Approximately 30% of DMPs were enriched in promoter regions (**Fig2B**). Gene ontology analysis revealed that these DMPs were associated with genes involved in diverse cellular processes, including several with specific functions in the brain (“Focal adhesion”, “Cholinergic synapse”, “Glutamatergic synapse”, “Dopaminergic synapse”) (**Fig2C**).

Since methylation at one site can be dependent on the methylation status of nearby CpGs, we decided to interrogate DMRs. Consequently, 13,071 differentially methylated regions (DMRs), between sPD and HC at baseline, were identified. Using a logFC cutoff of  $\geq \pm 0.05$  to focus on positions with substantial differential methylation, we identified 5,281 DMRs. The top 20 DMRs are presented in **Table 4**, and gene ontology analysis similarly indicated a connection with brain-specific functions (“Dopaminergic synapse”, “Tyrosine metabolism”, “Notch signaling pathway”) (**Fig3**). A complete list of DMPs and DMRs results at each time point are presented in **Supplementary Tables 1-6**.

## DNAm influences gene expression in sPD at baseline.

To delve deeper into the relationship between DNAm changes and gene expression, we analyzed RNASeq data from PPMI. 75 significantly differentially expressed genes between sPD and HC were identified, with 71 being upregulated and 4 downregulated (**Supplementary Table 7**). Integration of both differentially methylated sites and expressed genes showed 20 genes that were hypomethylated and overexpressed and one gene, *CTSH* (beta = 0.005; adjusted P value = 0.022), that was hypermethylated and associated with reduced expression (**Fig4**).

## Longitudinal Analysis

We extended our analysis to longitudinally profile samples collected over a span of 3 years. Demographic characteristics at each time point are detailed in **Table 2**. Linear regression models with cluster-robust standard errors identified 579 positions that significantly changed over time in sPD. To ascertain that age and LEDD were not the primary drivers of the longitudinal epigenetic alterations observed in sPD, we additionally assessed the relationship between methylation changes and these variables. We also found that several hypomethylated CpGs (7) were near the *CYP2E1* gene, while several hypermethylated CpGs (4) associated with the *NDRG4* gene (**Table 5**).

## Discussion

This study delves into the emerging role of epigenetic changes in PD and other neurodegenerative diseases. Specifically, we conducted a cross-sectional and longitudinal analysis of DNA methylation in whole blood samples from individuals diagnosed with sPD within the well-characterized and phenotyped PPMI cohort. This data also allowed us to examine DNAm patterns in a group of medication-free individuals at baseline, allowing us to uncover DNAm alterations more closely tied to disease mechanisms rather than medication-induced changes. Moreover, the longitudinal tracking of these individuals, coupled with comprehensive clinical and demographic data, offers a comprehensive view of PD's DNAm dynamics. Significantly, despite our use of blood-derived data, many of these DMRs were linked to genes with established roles in brain function (e.g. *CTSH*, *NDRG4*, *CYP2E1*, *BRSK2*).

By comparing DNAm across a span of three years, we identified 579 DMPs, even after controlling for age and LEDD. These findings suggest a potential role of DNAm in disease progression. Of interest, several hypomethylated DMPs were associated with the *CYP2E1* gene that was also a highly significant DMR at baseline. While other groups have reported on differential methylation in this gene on cross-sectional data<sup>10, 24, 25</sup>, our analysis demonstrates that this continues to be the case on follow up, suggesting a role of aberrant DNAm in *CYP2E1* in disease progression. Cytochrome P450 2E1, encoded by *CYP2E1*, plays a pivotal role in generating potentially toxic metabolites linked to dopaminergic degeneration. Furthermore, a SNP in this gene was identified as a genetic risk factor for PD in a Swedish cohort<sup>26</sup>. *CYP2E1* polymorphisms have also been suggested as indicators of susceptibility to DNA damage in individuals exposed to pesticides, a finding of significance given the well-established association between pesticide exposure and PD<sup>27, 28</sup>. Additionally, we also found several hypermethylated genes associated with the *NDRG4* gene; the protein encoded by this gene contributes to the maintenance of intracerebral BDNF levels and is required for cell cycle progression and survival in primary astrocytes. This gene has been found to have reduced expression in the substantia nigra and cingulate gyrus of individuals with PD<sup>29, 30</sup>. Future studies analyzing DNAm at longer intervals will be essential to gain a more comprehensive understanding of the link between DNAm and clinical progression,

Most of the differentially methylated positions at each time point were situated within gene bodies. Gene body methylation refers to the methylation of CpG sites in the transcribed regions of genes. Growing evidence suggests that methylation in CpG-sparse regions, such as enhancers and gene bodies, can profoundly influence gene expression<sup>31</sup>. While methylation in promoter regions typically silences genes, intragenic methylation may have a more varied effect, such as promoting gene expression, gene silencing, or alternative splicing [35]. Therefore, further exploration through functional studies of these intragenic methylation sites may provide crucial insights into their role in PD.

To explore the link between DNAm and gene expression, we analyzed RNA Sequencing data in individuals with sPD. This analysis unveiled differences in gene expression profiles between sPD and HC individuals. Additionally, a comparison between differentially methylated genes and differentially expressed genes revealed an overlap of 21 genes. Here, we identified that the *CTSH* gene was hypermethylated and with reduced expression. *CTSH* encodes the Cathepsin H protein, a lysosomal cysteine protease involved in lysosomal protein metabolism<sup>32, 33</sup>. Lysosomal dysfunction has been associated with PD<sup>34-36</sup>, and lysosomal related genes are associated with increased PD risk. As such, abnormal DNAm of this protein emerges as a promising novel avenue for investigating PD pathogenesis. Additionally, while we were only able to perform a cross-sectional analysis of RNASeq data, origination and analysis of longitudinal gene expression data will aid in contextualizing these findings and their role in disease progression.

Our analysis of methylation profiles from whole blood also allowed us to investigate the abundance of specific blood cell types in sPD. At baseline, individuals with sPD exhibited an increase in granulocytes that persisted upon follow-up, in agreement with previous reports. Our study also revealed a decrease in CD4T and CD8T cells upon follow-up (3 years after enrollment in PPMI). These changes may indicate immune system dysregulation in sPD, a finding of interest given the potential role of inflammation in the disease<sup>37, 38</sup>.

These findings also suggest that DNAm could be explored as a marker for PD diagnosis and disease monitoring, likely as part of a broader biomarker panel. This will likely require the use of machine learning algorithms<sup>39</sup>. Furthermore, as we move towards an era where the inclusion of



pre-symptomatic individuals in clinical trials for PD becomes feasible, characterizing these changes in the prodromal stage of PD will be valuable for understanding the mechanisms underlying the disease onset and assessing the potential of DNAm as a biomarker for identifying individuals at high risk of progressing to motor symptoms.

Future research will also leverage the complete PPMI dataset, including genetic-PD individuals and validate our findings in other cohorts. This will be of importance given that most individuals included in PPMI are of European descent. As such, it is imperative to perform further analyses in diverse populations to ascertain if there are population specific methylation signatures that could explain PD risk. Additionally, there is a lack of easily accessible DNAm data in large PD cohorts. This study underscores the importance of collecting these biological data in future cohort studies, which will aid in replication and validation of this and subsequent epigenetic studies.

A limitation of our analysis is the absence of environmental data for these subjects. As mentioned earlier, DNAm can be influenced by environmental exposures. Future directions will involve analyses on environmental factors such as pesticide exposure and lifestyle factors like smoking and alcohol consumption collected as part of PPMI. Further functional characterization of some of these findings will be vital for uncovering new disease mechanisms and potential intervention avenues. Also, integration of DNAm and other biomarkers will be informative. As an example, now that PPMI has data on alpha-synuclein seed amplification assay ( $\alpha$ Syn-SAA)<sup>40</sup>, this analysis will be rerun comparing individuals who are  $\alpha$ Syn-SAA positive vs those who are  $\alpha$ Syn-SAA negative.

In summary, our study provides evidence that alterations in the methylome in PD are discernible in blood, evolve over time, and reflect cellular processes linked to ongoing neurodegeneration. These findings lend support to the potential of blood DNA methylation as an epigenetic biomarker for PD. To fully comprehend DNA methylation changes throughout the progression of PD, additional profiling at longer intervals and during the prodromal stage will be necessary.

### **Acknowledgments**

PPMI, a public-private partnership, is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including 4D Pharma, Abbvie, AcureX, Allergan, Amathus Therapeutics, Aligning Science Across Parkinson's, AskBio, Avid Radiopharmaceuticals, BIAL, Biogen, Biohaven, BioLegend, BlueRock Therapeutics, Bristol-Myers Squibb, Calico Labs, Celgene, Cerevel Therapeutics, Coave Therapeutics, DaCapo Brainscience, Denali, Edmond J. Safra Foundation, Eil Lilly, GE HealthCare, Genentech, GSK, Golub Capital, Gain Therapeutics, Handl Therapeutics, Insitro, Janssen Neuroscience, Lundbeck, Merck, Meso Scale Discovery, Mission Therapeutics, Neurocrine Biosciences, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi, Servier, Sun Pharma Advanced Research Company, Takeda, Teva, UCB, Vanqua Bio, Verily, Voyager Therapeutics, the Weston Family Foundation and Yumanity Therapeutics

### **Author Contributions**

PGL, DK, TS contributed to the conception and design of the study;  
PGL, BB, SD contributed to the acquisition and analysis of data;  
PGL, BB, SL, TS, DK contributed to drafting a significant portion of the text.

### **Potential Conflicts of Interest**

Nothing relevant to this article to report

### **Data Availability**

The data that support the findings of this study are available from the corresponding author on reasonable request.

## References

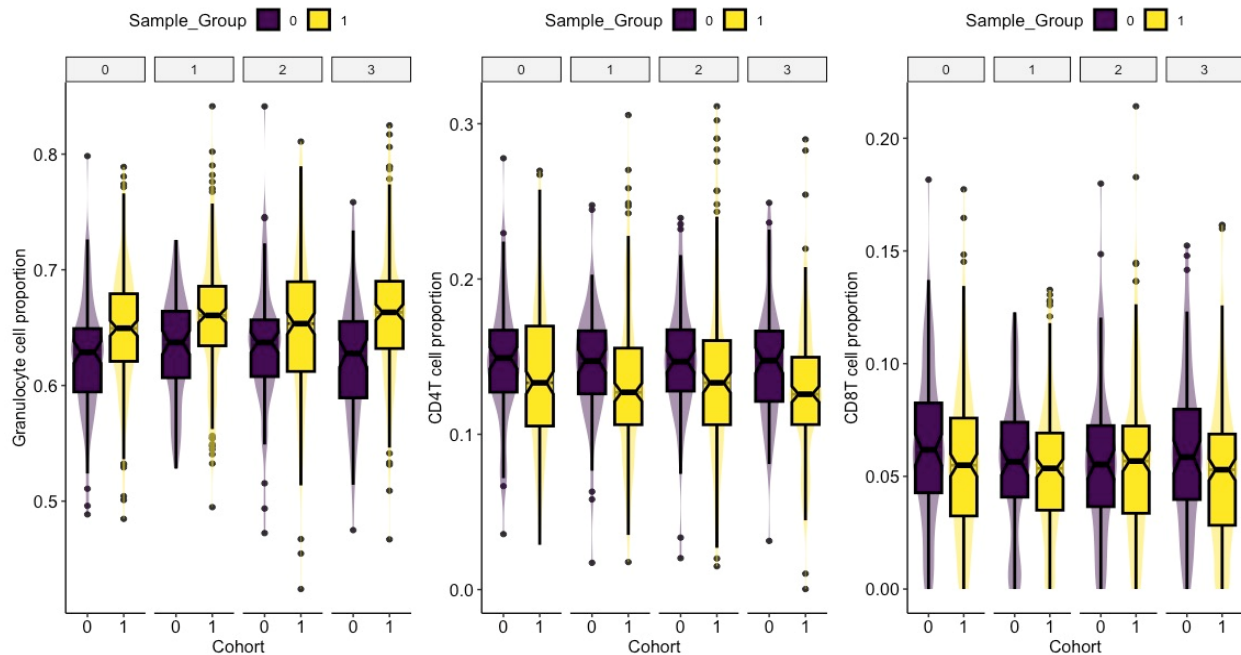
1. Dorsey ER, Sherer T, Okun MS, Bloem BR. The Emerging Evidence of the Parkinson Pandemic. *J Parkinsons Dis.* 2018;8(s1):S3-S8.
2. Marino BLB, de Souza LR, Sousa KPA, et al. Parkinson's Disease: A Review from Pathophysiology to Treatment. *Mini Rev Med Chem.* 2020;20(9):754-67.
3. Tran J, Anastacio H, Bardy C. Genetic predispositions of Parkinson's disease revealed in patient-derived brain cells. *NPJ Parkinsons Dis.* 2020;6:8.
4. Pang SY, Ho PW, Liu HF, et al. The interplay of aging, genetics and environmental factors in the pathogenesis of Parkinson's disease. *Transl Neurodegener.* 2019;8:23.
5. Bellou V, Belbasis L, Tzoulaki I, Evangelou E, Ioannidis JP. Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses. *Parkinsonism Relat Disord.* 2016 Feb;23:1-9.
6. Pavlou MAS, Outeiro TF. Epigenetics in Parkinson's Disease. *Adv Exp Med Biol.* 2017;978:363-90.
7. Halder R, Hennion M, Vidal RO, et al. DNA methylation changes in plasticity genes accompany the formation and maintenance of memory. *Nat Neurosci.* 2016 Jan;19(1):102-10.
8. Chuang YH, Paul KC, Bronstein JM, Bordelon Y, Horvath S, Ritz B. Parkinson's disease is associated with DNA methylation levels in human blood and saliva. *Genome Med.* 2017 Aug 30;9(1):76.
9. Chuang YH, Lu AT, Paul KC, et al. Longitudinal Epigenome-Wide Methylation Study of Cognitive Decline and Motor Progression in Parkinson's Disease. *J Parkinsons Dis.* 2019;9(2):389-400.
10. Henderson-Smith A, Fisch KM, Hua J, et al. DNA methylation changes associated with Parkinson's disease progression: outcomes from the first longitudinal genome-wide methylation analysis in blood. *Epigenetics.* 2019 Apr;14(4):365-82.
11. Marek K, Chowdhury S, Siderowf A, et al. The Parkinson's progression markers initiative (PPMI) - establishing a PD biomarker cohort. *Ann Clin Transl Neurol.* 2018 Dec;5(12):1460-77.
12. Campagna MP, Xavier A, Lechner-Scott J, et al. Epigenome-wide association studies: current knowledge, strategies and recommendations. *Clin Epigenetics.* 2021 Dec 4;13(1):214.
13. Morris TJ, Butcher LM, Feber A, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics.* 2014 Feb 1;30(3):428-30.
14. Fortin JP, Triche TJ, Jr., Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics.* 2017 Feb 15;33(4):558-60.
15. Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics.* 2010 Nov 30;11:587.
16. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* 2017 Feb 28;45(4):e22.
17. Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics.* 2013 Jan 15;29(2):189-96.
18. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007 Jan;8(1):118-27.
19. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012 May 8;13:86.
20. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 2004;3:Article3.



21. Peters TJ, Buckley MJ, Chen Y, Smyth GK, Goodnow CC, Clark SJ. Calling differentially methylated regions from whole genome bisulphite sequencing with DMRcate. *Nucleic Acids Res.* 2021 Nov 8;49(19):e109.
22. Mansell G, Gorrie-Stone TJ, Bao Y, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics.* 2019 May 14;20(1):366.
23. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
24. Kaut O, Schmitt I, Wullner U. Genome-scale methylation analysis of Parkinson's disease patients' brains reveals DNA hypomethylation and increased mRNA expression of cytochrome P450 2E1. *Neurogenetics.* 2012 Feb;13(1):87-91.
25. Kaut O, Schmitt I, Stahl F, et al. Epigenome-Wide Analysis of DNA Methylation in Parkinson's Disease Cortex. *Life (Basel).* 2022 Mar 29;12(4).
26. Brolin K, Bandres-Ciga S, Blauwendraat C, et al. Insights on Genetic and Environmental Factors in Parkinson's Disease from a Regional Swedish Case-Control Cohort. *J Parkinsons Dis.* 2022;12(1):153-71.
27. Tanner CM. Advances in environmental epidemiology. *Mov Disord.* 2010;25 Suppl 1:S58-62.
28. Ross GW, Abbott RD, Petrovitch H, et al. Association of brain heptachlor epoxide and other organochlorine compounds with lewy pathology. *Mov Disord.* 2019 Feb;34(2):228-35.
29. Phung DM, Lee J, Hong S, Kim YE, Yoon J, Kim YJ. Meta-Analysis of Differentially Expressed Genes in the Substantia Nigra in Parkinson's Disease Supports Phenotype-Specific Transcriptome Changes. *Front Neurosci.* 2020;14:596105.
30. Young JI, Sivasankaran SK, Wang L, et al. Genome-wide brain DNA methylation analysis suggests epigenetic reprogramming in Parkinson disease. *Neurol Genet.* 2019 Aug;5(4):e342.
31. Gutierrez-Arcelus M, Ongen H, Lappalainen T, et al. Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLoS Genet.* 2015 Jan;11(1):e1004958.
32. Fuchs R, Gassen HG. Nucleotide sequence of human preprocathepsin H, a lysosomal cysteine proteinase. *Nucleic Acids Res.* 1989 Nov 25;17(22):9471.
33. Wang Y, Zhao J, Gu Y, et al. Cathepsin H: Molecular characteristics and clues to function and mechanism. *Biochem Pharmacol.* 2023 Jun;212:115585.
34. Mc Donald JM, Krainc D. Lysosomal Proteins as a Therapeutic Target in Neurodegeneration. *Annu Rev Med.* 2017 Jan 14;68:445-58.
35. Navarro-Romero A, Montpeyo M, Martinez-Vicente M. The Emerging Role of the Lysosome in Parkinson's Disease. *Cells.* 2020 Nov 2;9(11).
36. Peng W, Schroder LF, Song P, Wong YC, Krainc D. Parkin regulates amino acid homeostasis at mitochondria-lysosome (M/L) contact sites in Parkinson's disease. *Sci Adv.* 2023 Jul 21;9(29):eadh3347.
37. Gonzalez-Latapi P, Marras C. Epidemiological Evidence for an Immune Component of Parkinson's Disease. *J Parkinsons Dis.* 2022;12(s1):S29-S43.
38. Pajares M, A IR, Manda G, Bosca L, Cuadrado A. Inflammation in Parkinson's Disease: Mechanisms and Therapeutic Implications. *Cells.* 2020 Jul 14;9(7).
39. Harvey J, Reijnders RA, Cavill R, et al. Machine learning-based prediction of cognitive outcomes in de novo Parkinson's disease. *NPJ Parkinsons Dis.* 2022 Nov 7;8(1):150.
40. Siderowf A, Concha-Marambio L, Lafontant DE, et al. Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using alpha-synuclein seed amplification: a cross-sectional study. *Lancet Neurol.* 2023 May;22(5):407-17.

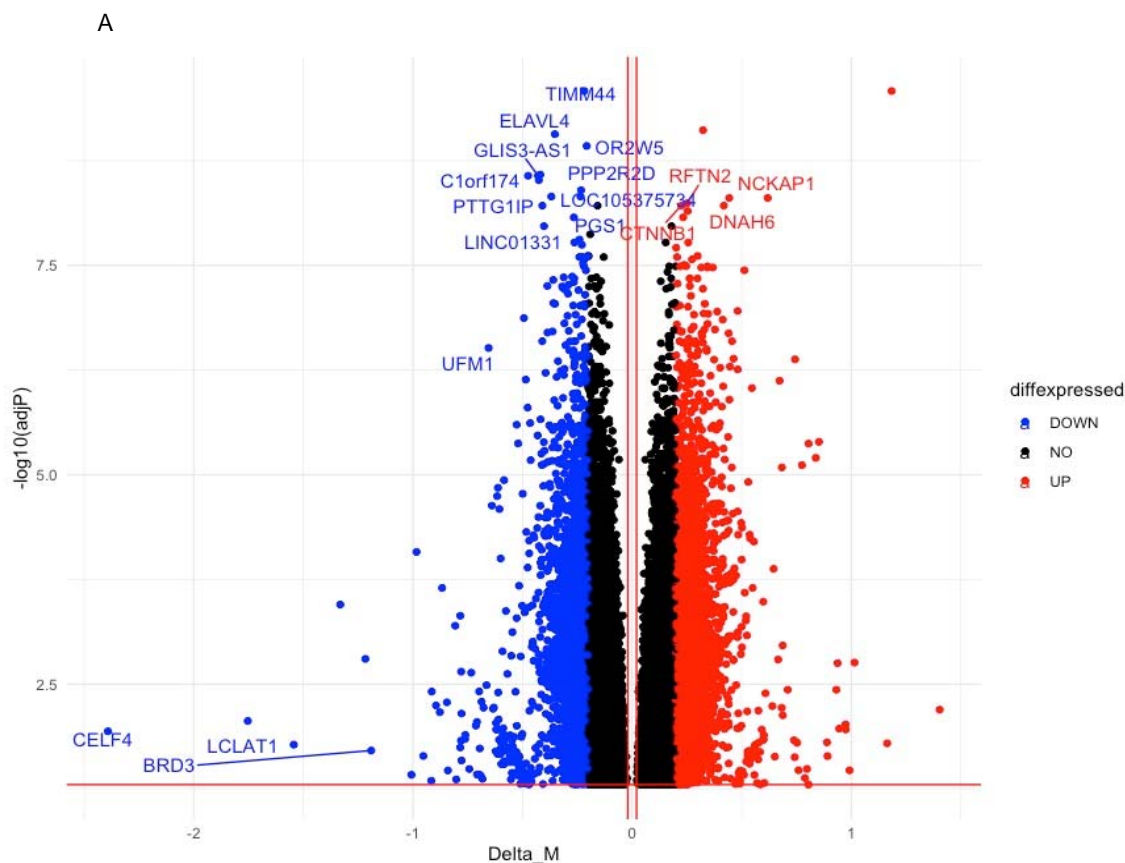
## Figure Legends

**Figure 1. Blood cell composition in sporadic Parkinson Disease individuals**



## Figure 2. Differentially Methylated Probes (DMPs)

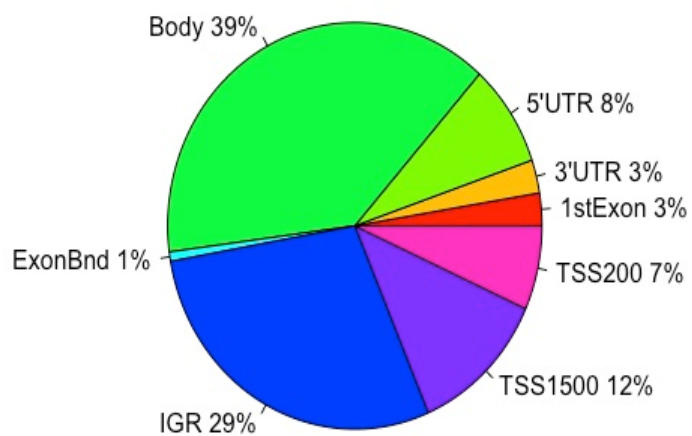
### A. Genes associated with DMPs



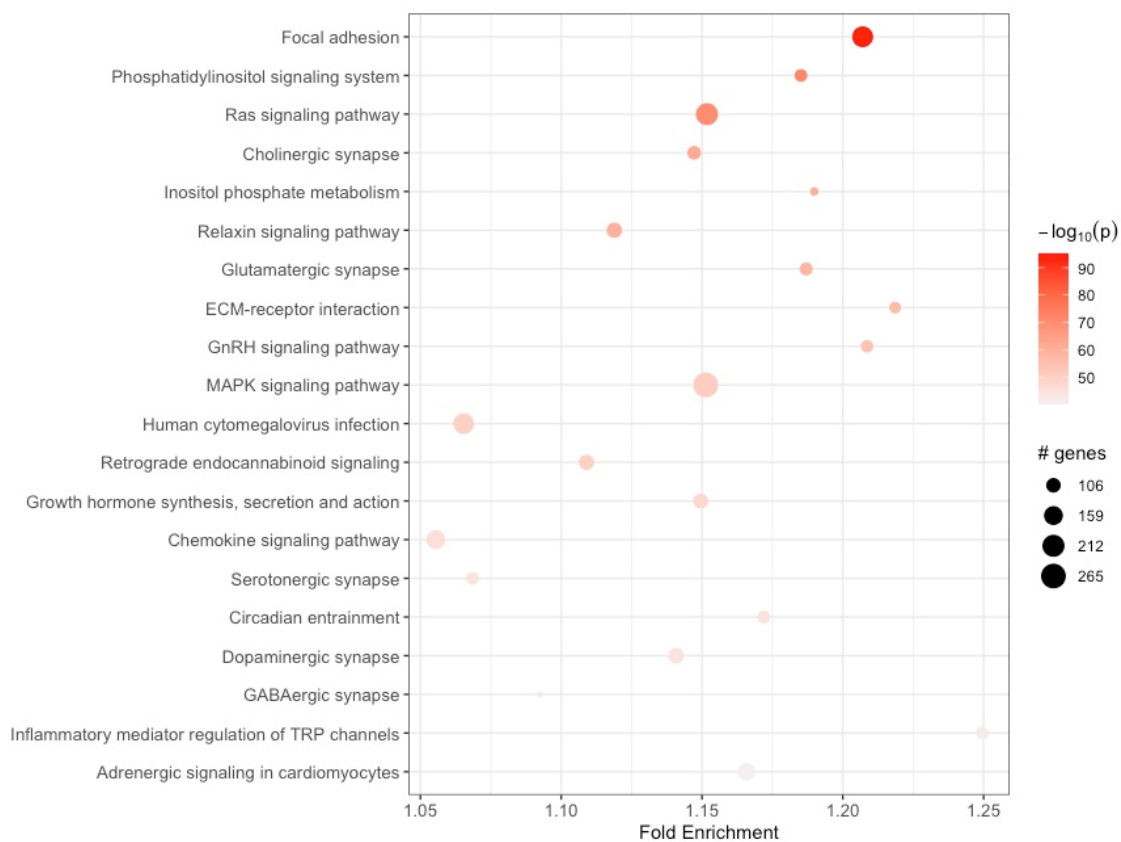
Volcano plot of differentially methylated cytosine-guanine sites (CpGs). Each point represents an individual probe. Nearest associated gene is presented. UP=Hypermethylation; Down=Hypomethylation. Of note, differential methylation is presented as M values (log of B values).

## B. Distribution of CpGs

### CpG Distribution

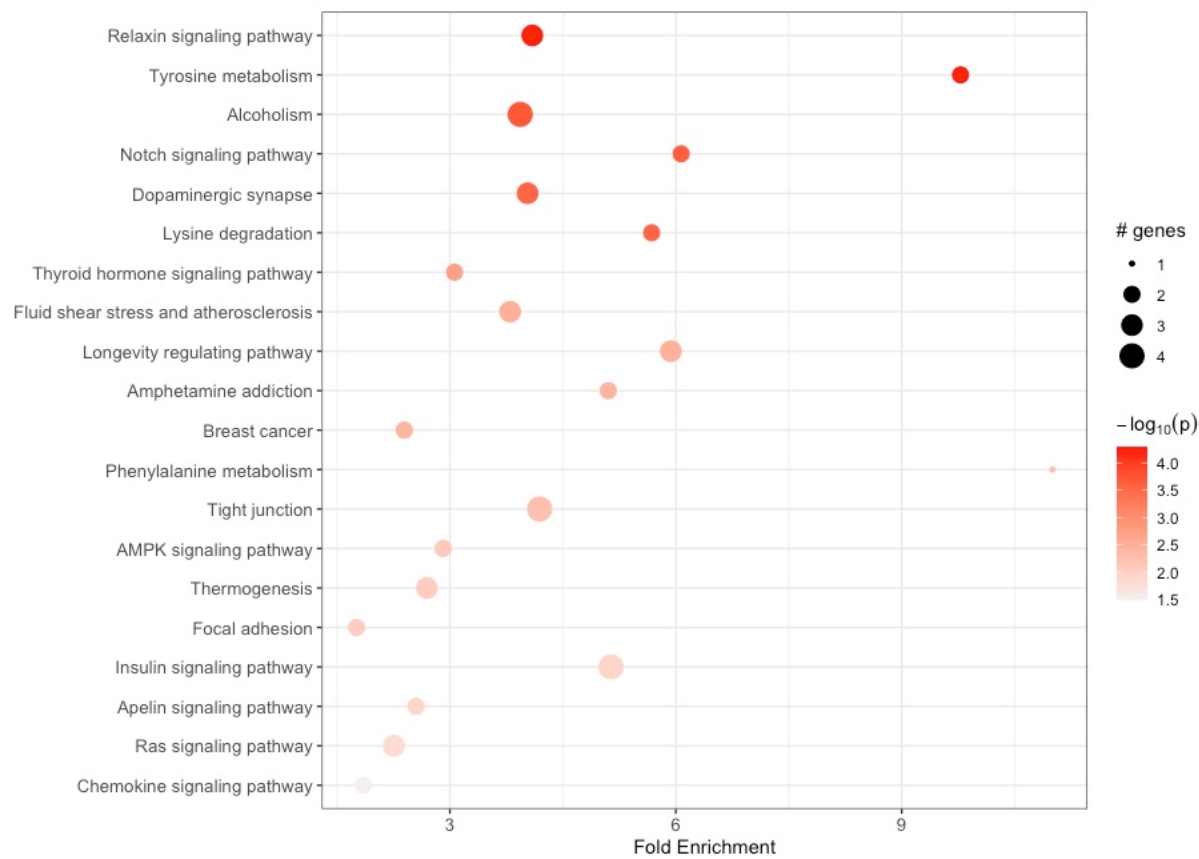


### C. Pathway enrichment for DMPs at baseline

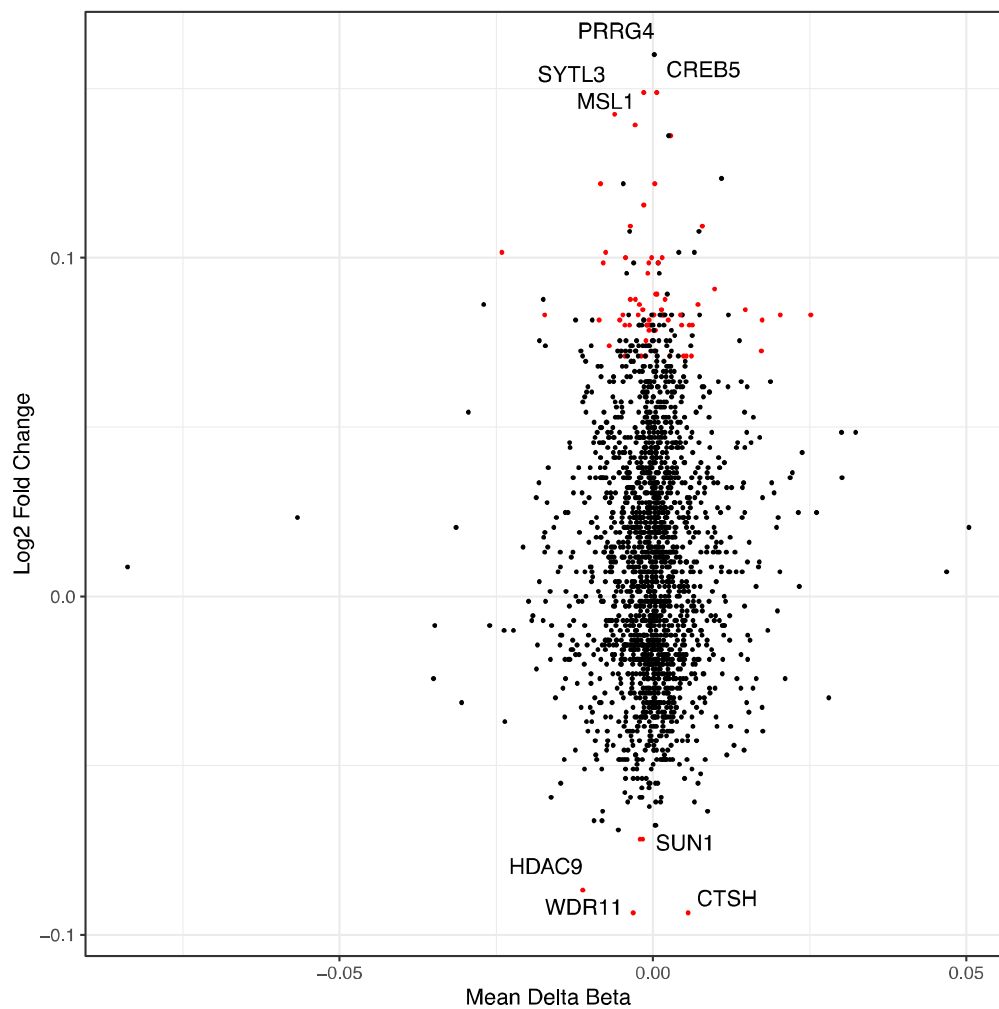




**Figure 3. Pathway enrichment for Differentially Methylated Regions (DMRs) at baseline**



**Figure 4. Overlap of differentially methylated and differentially expressed genes.**



Volcano plot of differentially methylated genes and differentially expressed genes. Each point represents an individual probe. Nearest associated gene is presented. UP=Hypermethylation; Down=Hypomethylation.

## Tables

<b>Table 1. Demographic characteristics at baseline</b>			
	<b>sPD(n=196)</b>	<b>HC (n=86)</b>	<b>p (95% CI)</b>
Mean age at baseline visit (SD)	61.80 (9.36)	62.15 (10.78)	0.795 (-2.30,2.99)
Female/Male (n)	58/129	28/58	0.79
White (n)	185	77	0.38
Median disease duration at baseline (range)	0.4 (0 -2)	NA	NA
Median symptom age (range)	62 (35-81)	NA	NA

**Table 3. Top 20 differentially methylated probes (DMPs) at baseline**

**Table 2. Clinical characteristics at baseline and on longitudinal follow up**

	Baseline			Year 1			Year 2			Year 3		
	sPD (196)	HC (86)	p (95% CI)	sPD (178)	HC (83)	p (95% CI)	sPD (179)	HC (80)	p (95% CI)	sPD (181)	HC (80)	p (95% CI)
Mean age (SD)	61.80 (9.36)	62.15 (10.78)	0.795 (-2.30, 2.99)	62.59 (9.43)	63.58 (10.43)	0.46 (-1.66, 3.65)	64.05 (10.78)	64.45 (9.50)	0.772 (-2.36, 3.17)	64.94 (9.62)	65.75 (10.66)	0.55 (-1.93, 3.56)
MDS-UPDRS Part I (mean)	1.40	0.63	< .001 (-1.10, -0.42)	1.76	0.65	< .001 (-1.50, -0.73)	1.82	0.57	< .001 (-1.62, -0.86)	1.96	0.62	< .001 (-1.73, -0.93)
MDS-UPDRS Part II (mean)	6.11	0.52	< .001 (-6.24, -4.93)	7.62	0.55	< .001 (-7.90, -6.23)	8.07	0.86	< .001 (-8.11, -6.31)	8.97	0.63	< .001 (-9.21, -7.45)
MDS-UPDRS-III (mean)	21.55	1.54	< .001 (-21.41, -18.59)	25.72	1.90	< .001 (-25.67, -21.96)	28.47	1.95	< .001 (-28.71, -24.37)	30.67	1.47	< .001 (-31.33, -27.07)
MOCA (mean)	27.21	28.15	< .001 (0.54, 1.33)	26.38	27.08	0.029 (0.06, 1.32)	26.28	26.98	0.060 (-0.03, 1.43)	26.69	27.4	0.023 (0.09, 1.30)
Mean LEDD (SD)	NA	NA	NA	311.81 (220.27)			381.54 (280.66)			381.54 (280.66)		

CpG	Chr	Position	Relation to TSS	Nearest associated gene	logFC	FDR
cg18213661	11	93681423	IGR		-0.2946842	0.00868945
cg25403368	5	130554320	IGR		-0.1673612	0.00035482
cg06378142	19	50119633	Body	<i>PRR12</i>	-0.1543515	0.03815787
cg13388618	2	238814771	Body	<i>RAMP1</i>	-0.1355725	0.00063473
cg17509989	5	176798049	Body	<i>RGS14</i>	-0.1284415	0.00682996
cg15355235	18	35001518	Body	<i>CELF4</i>	-0.1198882	0.01153713
cg10326673	2	30669757	TSS1500	<i>LCLAT1</i>	-0.1107039	0.04497683
cg03646189	5	2756101	IGR		-0.109624	0.00989133
cg06087988	2	46567260	Body	<i>EPAS1</i>	-0.109153	2.34E-05
cg14121685	15	56299462	IGR		-0.1054255	1.80E-05
cg27586797	5	13664584	IGR		0.19857441	0.0160684
cg06612594	20	43682078	Body	<i>STK4</i>	0.133571	0.01565868
cg01938825	7	1563708	IGR		0.13325065	0.04997444
cg12858166	6	33033176	3'UTR	<i>HLA-DPA1</i>	0.11387422	0.03364694
cg18661872	5	131436611	IGR		0.11101836	4.20E-07
cg05804568	5	159592696	IGR		0.1099409	2.63E-10
cg07796016	1	152779584	TSS1500	<i>LCE1C</i>	0.10961198	0.0417662
cg26251192	14	74003199	TSS1500	<i>ACOT1</i>	0.10695994	0.00160629
cg24853868	1	146555624	IGR		0.10477105	0.00406238
cg05194426	10	135343193	Body	<i>CYP2E1</i>	0.10295235	7.68E-06

**Table 4. Top 20 significant differentially methylated regions (DMRs) at baseline**



Chromosome	Start	End	CpG	FDR	Max $\Delta\beta$	Gene
<i>Hypomethylated</i>						
chr5	159592081	159592696	3	7.43E-10	-0.1063226	N/A
chr5	131436611	131436890	2	8.35E-07	-0.1030687	N/A
chr10	135340445	135343248	22	2.48E-05	-0.0981492	CYP2E1
chr20	43681878	43682078	2	0.01865295	-0.097719	STK4
chr11	1474588	1475899	7	2.87E-06	-0.0899853	BRSK2
chr1	232007841	232008510	2	8.71E-05	-0.0888528	DISC1
chr16	19777410	19778059	4	0.00648286	-0.0842839	IQCK
chr13	30053833	30055092	6	0.00095081	-0.0756862	MTUS2,
chr8	143751325	143751796	7	0.00430513	-0.0740626	PSCA, JRK
chr5	179740743	179741120	4	0.04429012	-0.0697626	GFPT2
<i>Hypermethylated</i>						
chr2	46567260	46567451	2	4.66E-05	0.09948858	EPAS1
chr2	30669385	30670304	12	0.01696551	0.09649313	LCLAT1
chr15	56299462	56299636	2	1.40E-05	0.09613655	N/A
chr12	7900072	7900651	3	0.00298332	0.09112004	CLEC4C
chr5	176796807	176798049	7	0.00116872	0.0875691	RGS14
chr1	115825602	115825641	2	0.02418216	0.08476224	N/A
chr11	1412270	1414627	9	0.00147298	0.08060391	BRSK2
chr1	101003634	101003924	3	0.02727681	0.07447863	GPR88
chr7	948246	949834	6	0.00280961	0.07007024	ADAP1
chr6	32305106	32305145	2	0.00142024	0.06978072	C6orf10

**Table 5. Top Longitudinal changes in sPD cases**

CpG_Site	Coefficient	SE	FDR	$\Delta\beta$ over three years	Nearest Gene
----------	-------------	----	-----	--------------------------------	--------------

cg10596483	0.31088382	0.01116813	1.75E-115	-0.0247078	JRK
cg17416644	0.53757281	0.00913787	1.50E-273	-0.0189683	BRSK2
cg21941354	0.2000633	0.00637213	7.49E-136	-0.0170572	PSCA
cg11445109	0.15685194	0.00880022	6.66E-59	-0.0150986	CYP2E1
cg04008252	0.2226635	0.00607315	3.09E-165	-0.0148353	NPFFR2
cg05473257	0.16573402	0.00645931	6.16E-103	-0.0127294	CYP2E1
cg24984195	0.34316818	0.00607679	2.96E-263	-0.0114182	UBE2I
cg21299458	0.17011373	0.0043338	3.15E-179	-0.0102815	SCARF2
cg19469447	0.1371077	0.00581347	5.06E-91	-0.0100304	CYP2E1
cg03134882	0.19896345	0.00548764	5.23E-163	-0.0100248	CYP2E1
cg17273911	0.21393103	0.00433243	1.45E-230	-0.0092127	N/A
cg09977703	0.19796984	0.00326851	5.65E-281	-0.0089486	PPP5C
cg05194426	0.25591425	0.00867698	5.36E-125	-0.0089095	CYP2E1
cg01873886	0.25430321	0.00524984	4.87E-226	-0.0089066	BMP4
cg23400446	0.14237323	0.00617015	4.17E-88	-0.0087313	CYP2E1
cg10862468	0.19063746	0.0061232	2.34E-134	-0.0083889	CYP2E1
cg12158483	0.18460978	0.00499163	5.51E-167	-0.0083389	BAIAP2L2
cg12754571	0.51125879	0.00774553	2.83E-303	-0.0076333	LOC148824
cg06378142	0.5259074	0.01546718	1.59E-150	0.03692832	PRR12
cg07504457	0.47968392	0.01020797	5.86E-219	0.02384914	VPS28
cg05725404	0.35295282	0.0064493	1.86E-255	0.02259907	NDRG4
cg00758881	0.24403566	0.00682108	2.27E-160	0.02230257	NDRG4
cg17457090	0.2751549	0.00549316	5.79E-234	0.02229852	NDRG4
cg27113419	0.27983815	0.00544638	4.67E-240	0.02203448	NDRG4
cg06211550	0.31122216	0.00835066	1.54E-168	0.02119219	N/A
cg22772380	0.43071726	0.0069473	5.76E-287	0.01959139	ZNF12