Title

Stochastic Epigenetic Mutations Influence Parkinson's Disease Risk, Progression, and Mortality

Journal

Journal of Parkinson's Disease, 12(2)

ISSN

1877-7171

Authors

Chen, Gary K
Yan, Qi
Paul, Kimberly C
et al.

Publication Date

2022

DOI

10.3233/jpd-212834

Peer reviewed

HHS Public Access

Author manuscript

J Parkinsons Dis. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

J Parkinsons Dis. 2022 ; 12(2): 545–556. doi:10.3233/JPD-212834.

Stochastic Epigenetic Mutations Influence Parkinson's Disease Risk, Progression, and Mortality

Gary K. Chen[a], **Qi Yan**[b], **Kimberly C. Paul**[c], **Cynthia D.J. Kusters**[d], **Aline Duarte Folle**[b], **Melissa Furlong**[e], **Adrienne Keener**[c], **Jeff Bronstein**[c], **Steve Horvath**[d,f], **Beate Ritz**[b,*]

[a]Independent Researcher

[b]Department of Epidemiology, UCLA Fielding School of Public Health, Los Angeles, CA, USA

[c]Department of Neurology, David Geffen School of Medicine, Los Angeles, CA, USA

[d]Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[e]Department of Community, Environment and Policy, University of Arizona Mel and Enid Zuckerman College of Public Health, Tucson, AZ, USA

[f]Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, USA

Abstract

Background: Stochastic epigenetic mutations (SEM) reflect a deviation from normal site-specific methylation patterns. Epigenetic mutation load (EML) captures the accumulation of SEMs across an individual's genome and may reflect dysfunction of the epigenetic maintenance system in response to epigenetic challenges.

Objective: We investigate whether EML is associated with PD risk and time to events (i.e., death and motor symptom decline).

Methods: We employed logistic regression and Cox proportional hazards regression to assess the association between EML and several outcomes. Our analyses are based on 568 PD patients and 238 controls from the Parkinson's disease, Environment and Genes (PEG) study, for whom blood-based methylation data was available.

Results: We found an association for PD onset and EML in all genes (OR = 1.90; 95% CI 1.52–2.37) and PD-related genes (OR = 1.87; 95% CI 1.50–2.32). EML was also associated with time to a minimum score of 35 points on the motor UPDRS exam (OR = 1.28; 95% CI 1.06–1.56) and time to death (OR = 1.29, 95% CI 1.11–1.49). An analysis of PD related genes only revealed five intragenic hotspots of high SEM density associated with PD risk.

*Correspondence to: Beate Ritz, UCLA, Epidemiology, Box 951772, Los Angeles, CA 90095, USA. britz@ucla.edu.

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JPD212834.

Conclusion: Our findings suggest an enrichment of methylation dysregulation in PD patients in general and specifically in five PD related genes. EML may also be associated with time to death and motor symptom progression in PD patients.

Keywords

Parkinson's disease; epigenetics

INTRODUCTION

Parkinson's disease (PD), the second most common neurodegenerative disorder after Alzheimer's disease, affects approximately 2% of individuals over 60 years of age [1]. PD is clinically characterized by motor symptoms such as rigidity, resting tremor, bradykinesia, and postural instability, but little is known about factors that influence symptom worsening [2]. Neuropathology in PD is marked by the loss of dopamine producing neurons in the substantia nigra pars compacta. While evidence for genetic contributions to PD risk has been accumulating over decades [3–5], research on epigenetic processes and their role in PD etiology is still incomplete [6–8]. There is also mounting evidence that the environment and its interplay with genetic variation and susceptibility plays a role in increasing PD risk and some of this may be through epigenetic mechanisms that affect protein expression levels. For example, epigenetic processes involving methylation of cytosine nucleotides can suppress transcription of genes to proteins. Epigenetic mutations have previously been shown to be useful biomarkers of biological aging which in turn potentially contributes to both risk and progression of neurodegenerative disease [9]. Based on the hypothesis that methylation patterns across the genome play a role in biologic mechanisms that contribute to aging and disease risk, investigators have conducted epigenome-wide association studies (EWAS) to map regions to traits of interest [10]. The EWAS approach is predicated on the notion that expression levels for a limited number of genes are primary drivers for disease risk, possibly mediated by environmental factors. Our group as well as others who have investigated links between PD and methylation patterns have employed EWAS to map loci that confer risk for PD [7, 8, 11]. For example, Chuang et al. identified 82 CpGs and 5 CpGs in blood and saliva respectively that were either hyper or hypomethylated in PD cases when compared to controls [7].

Complementary work has suggested that overall dysregulation of methylation levels, rather than changes in methylation levels at a specific set of loci, are important drivers for outcomes such as aging [12], pre-term birth [13], or cancer [14]. Recently, a metric known as a SEM (stochastic epigenetic mutation) has been suggested as one approach for quantifying dysregulation at a particular genomic site [15, 16]. SEM reflects the accumulation of epigenetic maintenance machinery failures (stochastic mutations) and the number of SEMs increases exponentially with chronological age [12]. As an example of the applicability of this metric in earlier work, we aggregated SEM status across all methylation sites into a genome-wide score we called EML (epigenetic mutation load) to investigate the relationship between methylation levels and biological clocks [17]. Here, we employ this measure (denoted in this paper as EML_{genome}) to improve our understanding of PD etiology and progression. Additionally, we devise a new measure called EML_{PD} that is purely

composed of methylation sites that map to a set of curated genes previously linked to PD risk. Our primary objective is to assess whether (log-transformed) EML_{genome} and EML_{PD} scores are associated with PD risk, motor symptom progression, and/or time-to-death.

MATERIALS AND METHODS

Study population

We draw on blood samples from participants of the Parkinson's disease, Environment and Genes (PEG) study, including 511 individuals of European ancestry (292 PD patients and 219 controls) and 61 individuals of Hispanic ancestry (42 PD patients and 19 controls). PEG was designed to investigate links between environmental factors and PD in agricultural regions of central California [18, 19]. Eligible patients recruited via community outreach, local neurologists, and clinics were enrolled into the PEG study in two waves, denoted in this paper as PEG1 and PEG2. Analyses related to PD progression were restricted to either patients (PEG1 and PEG2 combined) or PEG1 controls (for comparison purposes). Standardized interviews were used for collecting information such as demographic and lifestyle factors, comorbidities, and medication use. PD patients enrolled in PEG1 were recently diagnosed (within 3 years) and in PEG2 on average 4 years before they were examined by UCLA movement disorder specialists (JB, AK) at least once (i.e., at baseline). Patients and controls were ascertained from Kern, Tulare, and Fresno counties (2001–2007 for PEG1 and 2011–2015 for PEG2). PEG1 patients were identified by neurologists, large medical groups, or public service announcements. PEG2 patients were identified through the California PD Registry operating in the same counties.

Of 1,167 PD patients initially invited in PEG1, 563 patients were eligible based on the threshold of diagnosis of 3 years or less. 363 PEG1 patients were confirmed as having probable idiopathic PD through examination by movement disorder specialists. At first re-contact for follow-up examination, 83 (22.9%) of the 363 patients were deceased, 25 (6.9%) withdrew, 9 (2.5%) could not be found, and 4 (1.1%) were too ill to participate. For follow-up, 242 (66.6%) were re-examined with 6 patients participating by mail/phone only (no in-person follow-up exam and therefore not contributing to this analysis) and 3 patients were re-classified by us during follow-up as not having idiopathic PD. Of 233 patients re-examined and confirmed to have idiopathic PD, 178 (76.4%) were seen twice and 55 were seen only once during follow-up (21 were deceased prior to and 14 pending a second exams, 8 were too ill to participate, 4 could not be located, 3 withdrew, and 5 participated by mail/phone only).

From 2011 to 2015, 734 PD patients were eligible at first screening for entry into the PEG2 cohort. Of these, 85 potential PEG2 patients refused to participate, 47 could not be examined by UCLA movement disorder specialists at a local clinic, and 119 were excluded after an exam as they did not have idiopathic PD, leaving a baseline of 483 PD patients who met our enrollment criteria Of the 227 patients with methylation data available, 134 (59%) provided some follow-up data (124 of whom have motor symptom data), while 59 (26%) were deceased, 10 (4%) refused to participate, 17 (8%) could not be found, 5 (2%) are pending to have a second exam, and 2 (1%) were too ill to participate.

Population controls from PEG1 from the same communities were randomly sampled from Medicare lists (before HIPAA enactment) and through residential tax assessor's records as well as through community outreach. The controls for PEG2 only contributed saliva samples for DNA extraction and thus are not part of this investigation. PD risk analysis was restricted to $n = 806$ participants ($n = 238$ controls, $n = 341$ PD patients from PEG1, $n = 227$ PD patients from PEG2) for whom we currently have quality control validated blood-based DNA methylation data available.

Methylation data

DNA methylation data were obtained from peripheral blood samples. Methylation status was measured in bisulfite-treated genomic DNA using the Illumina Infinium HumanMethylation-450 assay. For the main analyses assessing the relationship between EML and all PD outcomes, we included all participants from PEG1 and PEG2, controlling for study wave status. To assess robustness of our inferences, we also conducted a sensitivity analysis, limiting to only PEG1 cases for the following reasons. We computed the distribution of our predictor of interest (i.e., the SEM score as defined in the following sections) based on our reference group, i.e., the community controls sampled along cases in PEG1 only, because only PEG1 required control subjects to provide a blood sample while controls for PEG2 only provided a saliva sample. Furthermore, we observed differences in the overall methylation levels at certain sites between PEG1 and PEG2 patients, possibly indicating either technical artifacts (e.g., a batch effect) or biological differences (e.g., differences in age, environmental factors between the two birth cohorts), which could potentially skew distributions of methylation events between PEG1 and PEG2 cases. Therefore, we present results for the whole study population and also for PEG1 only.

Quality control

Raw signal intensities were retrieved from Illumina iDat format files using the function read.metharray.exp of the R package *minfi* from the open-source library *Bioconductor* [20], followed by linear dye bias correction, *noob* background correction, and BMIQ normalization using the same R package [21–23]. The β-values, determined by calculating the ratio of intensities between methylated and unmethylated sites, were used for all analyses. Probes on the array were excluded based on the following criteria: 845 probes that were statistically likely to be detected above the background signal level regardless of methylation status (based on alpha-level $p < 0.05$), 645 probes with bead counts less that 3 in at least 5% of the samples, 11,334 probes mapping to sex chromosomes, 7,306 probes that included a known SNP within the CpG interrogation site, and 27,332 probes that were cross-reactive. In total, 438,050 remaining probes were included for downstream analyses.

Outcomes

In this paper, we examined the relationship between methylation pattern and four outcomes: first, overall risk of PD, then disease progression and time to death in PD patients. PD diagnosis was determined by UCLA movement disorder specialists as described above and we compared cases to our community controls. Patients were examined clinically while not receiving PD medications, i.e., being functionally in an 'off' state (82% at the baseline examination and 80% at follow-up). For patients whom we could only examine

on medication, we estimated an off-score by adding the difference of the whole study population's mean off- and mean on- scores at the time of exam to the patient's on-score, as described in previous work [24]. Second, we modelled time to death as a continuous variable indicating the number of years from the baseline exam to death. The last two outcomes were based on two complementary measures, capturing PD symptom progression that have been widely used in PD research: the Unified Parkinson's Disease Rating Scale part 3 (motor – mUPDRS) and Hoehn & Yahr staging [25], rated by UCLA movement disorder specialists who conducted physical examinations at each examination to assess PD progression. We used time to a score of 35 or higher on the mUPDRS to indicate progression to more severe PD motor symptoms, as we did in previous analyses [26]. A mUPDRS score ≥ 35 (higher score represents worse motor function) was chosen as a meaningful threshold for motor progression because it represents, on average, motor progression to a stage where patients start presenting some dependency for functional activities, equivalent to a HY stage 3 and to 60% on the Schwab and England scale [27]. In our analysis, we considered time to conversion to mUPDRS>=35 and time to conversion to Hoehn & Yahr stage 3 or greater (HY3).

PD-specific gene set

The Parkinson's UK Annotation Initiative has provided the research community a free resource in the form of a list of 811 curated genes involved in over 4,500 GO annotations that relate to processes relevant to PD [28]. Based on this data, we hypothesized that methylation patterns at these genes may be informative and would be associated with PD disease risk. We intersected this list of 811 curated genes to probes annotated with the same gene symbols based on a methylation probe descriptor file provided by Illumina, resulting in 747 candidate regions.

Metrics for assessing methylation dysregulation: SEM and EML

After exclusion criteria were applied on methylation probes that failed quality control, at each subject-probe datapoint across 438,050 probes and 806 study subjects, we computed its SEM status as a binary score. Specifically, at each locus, we first determined the interquartile range (IQR) among all 238 controls. Relying on a previously published definition, a SEM is observed for a given person at a specific CpG site if an individual's methylation level is more than three times the interquartile range (IQR) lower than the 25th percentile (Q1 – 3×IQR), or more than three times the IQR higher than the 75th percentile (Q3 + 3×IQR) [12].

We assessed whether downstream statistical inferences would remain stable when using different threshold criteria via sensitivity analyses. Namely, we also considered two generalized scores, SEM_{α} and SEM_{β}, defined as:

$$\mathrm{SEM}_{\alpha} = \mathrm{I}(x < \mathrm{Q1} - \alpha * \mathrm{IQR} \cup\ x > \mathrm{Q3} + \alpha * \mathrm{IQR}) \tag{1}$$

$$\mathrm{SEM}_{\beta} = \mathrm{I}((1-\beta)x < \mathrm{Q1} - 3 * \mathrm{IQR} \cup (\beta)x > \mathrm{Q3} + 3 * \mathrm{IQR}) \tag{2}$$

where we re-fit all our logistic regression models for predicting PD risk after varying values of α and β.

Based on the SEM scores defined at each locus, for each subject, we constructed three EML scores to capture the overall burden of SEM counts for selected sets of valid methylation probes for each subject [17]. We denote these EML measures as the EML_{genome}, EML_{PD}, and $EML_{<window>}$ scores. The EML_{genome} score is equivalent to the definition introduced in Yan et al. (2020), i.e., the EML taken over all available probes spanning the genome. A subset of probes that map to the 747 genes from the PD geneset described earlier are used to construct the EML_{PD} score. We hypothesized that the EML_{PD} score may also predict PD disease risk and compared results of analyses using both EML scores. Finally, we applied a sliding window approach in order to consider the possibility that there are certain hotspots, defined as regions mapped to a subgroup of probes within the PD geneset that are most strongly associated with PD risk. We define notation for this measure as $EML_{<window>}$, where the subscript <window> (e.g., $EML_{Chr5:100502}$-120000) represents the physical base pair coordinates of the start and end boundaries of the sliding window.

Figure 1 suggests a strong right skewing in the distribution of the EML_{genome} score. This was confirmed based on a Kolmogorov-Smirnov (KS) test, comparing to a normal distribution, for both EML_{genome} and EML_{PD} (deviance of 1.0 ($p < 1e\text{-}16$)). Hence, for regression models, a log transformation was applied to all EML definitions to ameliorate violations to normality in regression model residuals. To aid in interpretation of the unit size of regression coefficients, we further applied a second transformation, mapping all log-transformed EML scores to a Z-score.

Statistical analyses

To study the relationship between our EML measures and PD status or disease severity (i.e., time-to-event as described above), we constructed logistic regression models to estimate odds ratios in cases and controls, and Cox proportional hazards models for estimating hazard ratios among PD cases only. We also conducted additional analyses, stratified by gender. All regression models were adjusted for potential confounding variables: age (continuous variable with 31–99), sex (binary), ethnicity (white/non-white), and five quantities that estimated blood cell composition since methylation is tissue cell type composition dependent. For the time to event (Cox) models, to minimize confounding on methylation due to disease duration, we also included a variable measuring the number of years between diagnosis and the baseline interview date. The cell composition variables distinguish between naïve CD8 cells, CD8 + CD28-CD45RA-T cells, plasma blasts, CD4 T cells, and granulocytes. These variables were estimated from an online tool (https://horvath.genetics.ucla.edu/html/dnamage/), based on methods described by Horvath [29].

To assess robustness of the results, as a sensitivity analysis we added three additional variables based on knowledge gleaned from previous work investigating potential factors for methylation changes: educational attainment, smoking status, and a biomarker measuring epigenetic age acceleration. An EWAS of 4152 participants demonstrated an education-related gradient of methylome changes that reflect cigarette smoke exposure signatures [30]. For this study, we encoded years of educational attainment and smoking status as categorical

variables: (educational attainment: less than 12, 12, or greater than 12 years of education and smoking status: ever smoked, never smoked, or quit smoking). Previous studies have shown that a biomarker known as extrinsic epigenetic age acceleration (EEAA) was associated with increased mortality [31], PD risk [32], and EML [17]. Thus, given its association with EML, we considered EEAA as a potential confounder. We compared regression results from the standard adjusted model (i.e., with all control variables described above) to a model with the added continuous variable measuring EEAA, and two categorical measures of educational attainment and smoking status in the sensitivity analysis.

In models of disease severity (i.e., time to death and PD symptom progression), we also included total levodopa equivalent dosage (LED) reported at the baseline exam as levodopa medication may potentially affect methylation levels, and patients with more severe symptoms of PD may be prescribed a more aggressive levodopa regiment. Previous analyses suggested that methylation levels confer a risk of developing PD, and it has been shown that methylation is affected by genes, i.e., that the average heritability of DNA methylation sites is 18.7% [33]. Thus, we also included as a covariate in our models a polygenic risk score (PRS) to control for any heritable variation. The PRS was constructed as a weighted sum of allele counts, where the weights are effect estimates from known PD risk SNPs [34].

For the purpose of testing associations between PD and $\text{EML}_{<\text{window}>}$ (i.e., window-specific EML scores), we applied the R package *wgscan* to test a series of candidate sliding windows, comprising contiguous methylation probes, for association with either PD disease status or disease severity. The *wgscan* program, which was originally applied for mapping a contiguous set of rare SNPs to a given outcome [35], can be applied to other measures (e.g., CNV, methylation counts) under the assumption that a cluster of variation at neighboring sites is relevant to the trait. Methods for rare variant analysis in general apply either a burden test (i.e., scores are summed) or a test of overdispersion in the variance of the effect sizes on a selected set of contiguous variants, where the latter test can potentially be more powerful than the former if variants within a window do not share the same directionality of effects [36]. In the context of our analysis, we treat SEM calls at each methylation probe as the variant of interest. The program applies a sliding window approach, systematically considering various window sizes. Candidate window sizes can be configured by the user, and significance levels are adjusted for the increased type 1 error due to the large number of windows (hypotheses) being tested. For each window tested against the trait of interest (PD), we included in the model the standard set of control variables described above. At each window in the provided region, the program reports *p*-values for the burden and overdispersion tests. A global region-specific significance level, adjusted for the number of windows tested is also reported to the user. Finally, we computed a genome-wide significance level α_{global} to correct for the number of regions tested. For k regions (genes), this threshold is defined as $\alpha_{\text{global}} = 1/(1/\alpha_{\text{gene-1}} + 1/\alpha_{\text{gene-2}} + \ldots 1/\alpha_{\text{gene-k}})$ where the region-specific significance levels (e.g., α_{gene}-1) is computed to account for the multiple comparisons inherent in testing multiple sliding windows.

RESULTS

Table 1 summarizes the characteristics of the 806 subjects included in this study whose quality checked methylation data were available for analyses.

Distribution of SEM frequency across loci

Inspection of the distribution of SEM frequency across all methylation probes showed that the probability of a SEM event at any particular locus was rare. A histogram of SEM counts at each methylation probe is depicted in Supplementary Figure 1. The median frequency at a given site was 0.74% (6/806) with a range of 0–211.

Estimated effects of EML_{genome} on PD risk

A visual comparison of the distributions of the EML_{genome} score between cases and controls, as depicted in the violin plot of Fig. 1, suggests that EML levels are higher in cases than controls. Furthermore, the large range in scores for the full dataset appears to be driven more strongly by female cases. To test our hypothesis that overall methylation dysregulation is enriched in PD cases, we fitted regression models with PD status as the outcome of interest.

Table 2 summarizes results from logistic regression models for all PEG1/PEG2 subjects and men and women separately. For comparison, we also present results after excluding PEG2 subjects as part of a sensitivity analysis. As expected, for effects estimated with different case inclusion or exclusion criteria (i.e., all PEG waves versus PEG1 subjects only), the directionality of effects was consistent, with somewhat diminished effect sizes for the latter comparison. To quantify differences in the distribution of EML scores between the two datasets, we regressed the EML measures on PEG wave status, controlling for the same confounding variables in the logistic regression models. We did not observe any formally statistically significant heterogeneity of effects as indicated in Supplementary Table 1.

Risk of PD was statistically significantly associated with both EML scoring criteria when comparing all cases to controls (EML_{genome}: OR = 1.90; 95% CI (1.52–2.37); EML_{PD}: OR = 1.87; 95% CI (1.50–2.32). The EML_{genome} variable exhibited the strongest estimated effect size on PD risk for female cases only (all subjects: OR = 2.28; 95% CI 1.59–3.25; PEG1 only: OR = 2.05; 95% CI 1.45–2.92). The association between EML_{genome} and PD risk was weaker in males (all subjects: OR = 1.66; 95% CI 1.22–2.24; PEG1 only: OR = 1.46; 95% CI 1.08–1.97). The magnitude of the estimated effect size of PD risk was smaller with EML_{PD} than with EML_{genome}, but the direction of the effect was consistent, with females always showing a stronger association than males. To assess the possibility of sex specific heterogeneity, we fitted models after adding an interaction term between sex and either EML variable. Based on this interaction term, we observed moderate evidence of heterogeneity by sex for EML_{genome} (OR = 1.53; 95% CI 1.03–2.26; $p = 0.0338$) and weaker evidence for EML_{PD} (OR = 1.41; 95% CI 0.96–2.06; $p = 0.0787$), suggesting a stronger joint effect of female sex and EML_{genome} than expected.

We conducted additional sensitivity analyses on the full dataset to assess the robustness of our findings. First, adding the potential confounders EEAA age acceleration, educational

attainment, and smoking status did not change our results (Supplementary Table 2). Second, when we defined EML according to SEM_{α} and SEM_{α} odds ratio estimates remained very similar (Supplementary Table 3).

Mapping of $EML_{<window>}$ to PD risk

We split the genome-wide methylation dataset into 747 regions, each of which corresponds to a gene that has previously been associated with PD and received a GO annotation as listed on the Parkinson's UK Annotation Initiative website. A total of 9400 windows ranging from size 5 kB to 50 kB were tested across this subset of the genome.

We did not find any formally statistically significant associations between the $EML_{<window>}$ scores and the continuous outcomes measuring disease severity. However, burden and dispersion tests indicated that five of these $EML_{<window>}$ scores were associated with PD risk, based on the genome-wide significance level α_{global} of $3e^{-7}$, as shown in Table 3. Notably, these regions map to the genes encoding for proteins that are related to cellular stress response (e.g., STK39), endocytosis (e.g., NEDD4), ER protein trafficking (e.g., PI4KA), and neuronal/microglia function (e.g., P2Y12). Although we configured the program to construct candidate window sizes ranging from 5kB to 50 kB, none of the scores corresponding to windows greater than 5 kB in size were significantly associated with PD risk. The Manhattan scatter plot in Supplementary Figure 2 depicts the landscape of $-\log_{10}$ *p*-values of window-specific burden (shown in red) and dispersion (shown in blue) tests.

Effect of EML_{genome}/EML_{PD} on PD progression and mortality

Next, we assessed whether either EML_{genome} or EML_{PD} were associated with measures of disease progression in PD cases only, applying a Cox proportional hazards ratio time to event model for each of the three outcome variables. We did not observe statistically significant associations between the EML measures and the HY3 time to event variable. For the time to a mUPDRS score of 35 or greater, we estimated increased hazard ratios and their confidence intervals were formally statistically significant for all cases (EML_{genome} HR = 1.28, 95% CI 1.06–1.56 and EML_{PD} HR = 1.20, 95% CI 1.08–1.56) and the female subset of cases (EML_{genome} HR = 1.43, 95% CI 1.10–1.86 and EML_{PD} HR = 1.41, 95% CI 1.10–1.86), as indicated in Table 4. Neither EML_{genome} nor EML_{PD} were associated with time to a HY stage>=3 (EML_{genome} HR = 1.13, 95% CI 0.94–1.37 and EML_{PD} HR = 1.13, 95% CI 0.94–1.36).

Likewise, hazard ratios were increased for time to death for all cases (EML_{genome} HR = 1.29, 95% CI 1.11–1.49 and EML_{PD} HR = 1.29, 95% CI 1.12–1.48) and the female subset (EML_{genome} HR = 1.33, 95% CI 1.07–1.64 and EML_{PD} HR = 1.31, 95% CI 1.07–1.60), also presented in Table 4. Effect sizes did not change substantively in models including different sets of control variables (Supplementary Table 4). To assess whether this result was specific to PD cases, we also assessed time to death between EML_{genome} and EML_{PD} in controls. While we did not observe associations in controls that were formally statistically significant, the effect estimates were above the null value of one (EML_{genome} HR = 1.21, 95% CI 0.85–1.74 and EML_{PD} HR = 1.22, 95% CI 0.86–1.73).

DISCUSSION

We investigated whether dysregulation of blood cell methylation plays a role in PD etiology and contributes to mortality in patients with PD in a population-based study of patients and their community controls from California's Central Valley. For these participants, we have been able to gather a rich set of outcome data with patients followed for progression, and all participants for mortality over decades. Using a high-density assay from Illumina, we queried over 450 k probes after extensive QC filtering to assess SEM, a surrogate for measuring methylation dysregulation. Our findings suggest an increase ranging from 66–128% in risk of PD for each standard deviation of EML, with a stronger estimated effect size in women than in men. Although our analyses were based on pooled cases from two different birth cohorts (PEG1 and PEG2) compared against only one reference group (PEG1 controls), our sensitivity analyses (as indicated in the Table 2 and Supplementary Table 1) suggest that our findings are robust against confounding due to a cohort effect in that effect sizes were diminished in the smaller case control set but generally findings and trends remained the same (i.e., EML_{genome} had stronger associations than EML_{PD}, and females showed stronger associations than males).SEM events in blood cells are stochastic in nature, affect the genome overall, and are driven by some external process but seem to nevertheless not only affect PD risk but also motor decline especially among female patients as well as time to death in all patients. Similar to PD risk, for time to death and motor symptom progression, we observed positive associations with both EML measures (i.e., EML_{PD} and EML_{genome}).

Using an agnostic sliding window analysis, where only contiguous probes within a window were used to construct a score called $EML_{<window>}$ from all PD risk genes previously identified, several hotspots of dysregulation emerged as genome-wide statistically significant according to both a burden and a dispersion test for the region. Stochastic epimutations in and around five PD risk genes were identified as associated with PD risk. The serine/threonine kinase (STK39) phosphorylates alpha-synuclein, and a hyperphosphorylated state has been implicated in PD [37]. NEDD4 protein, which attaches ubiquitin chains to alpha synuclein, is critical to marking potential synuclein aggregates for degradation by the lysosomal machinery [38]. Deficiency in expression of this gene may be a driving factor in the alpha-synuclein inclusions that are a hallmark of PD. Knockout experiments in mice of the phosphoinositide kinase PI4KA resulted in diminished myelination in Schwann cells [39]; notably the authors noted that these mice displayed a neurodegenerative-like phenotype. The PI4KA enzyme can also play an important role in the immune system, where it critically binds to the CD7 glycoprotein to participate in proper intracellular signaling that modulates T cell functionality [40]. Injured neurons, which release nucleotides (purines), normally trigger an inflammatory response where microglia migrate toward the source of these nucleotides, mediated via the purinergic receptor P2Y12 [41]. Since neurodegenerative diseases such as PD are characterized by pro-inflammatory responses [42], over-expression of P2Y12 may be a driver of disease risk and/or progression. The Dickkopf-related protein 3 (DKK3), which is involved in the Wnt receptor signaling pathway, has been suggested as a therapeutic target for PD [43]. Both the 'classic' (through β-catenin) and 'non-classical' pathways play roles in dopaminergic cell development

and synaptic function. PD-associated proteins encoded by PARK2 (protein: parkin) and LRRK2 have been shown to modify classic Wnt signaling [44]. While our study did not find evidence for epigenetic modifications in these PD genes, we identified other genes (APC and AXIN1) involved in classic Wnt signaling in enrichment analysis. Interestingly, decreased Wnt signaling has also been reported in AD [45]. CALM2 is one of the AD-related genes we associated in blood based methylation analysis with PD status [7] and it has a function in calcium ion binding which is important in the non canonical Wnt signaling pathway [46].

We note some caveats to our results and interpretations. Reported *p*-values for the various regression models were not corrected for the number of outcomes tested as these outcomes are highly correlated as measures of disease progression/risk and applying a Bonferroni correction may be too stringent. Although we estimated relatively strong effect sizes for PD risk, the heterogeneity in effect sizes between sexes and their wide confidence intervals need to be addressed in further studies with a larger sample size. For example, a lack of a statistically significant association between EML and time to death in controls may have been due to the small sample size relative to cases. To consider the possibility of heterogeneity between cases and controls, we fitted a time to event model for cases and controls that included an interaction term between case status and EML and the interaction term was not statistically significant ($p > 0.5$). We should note that not all PEG patients with methylation data had follow-up data for mUPDRS scores (time to death follow-up data was complete however). For this reason, we considered the possibility that hazard ratios estimated for time to an increase of 35+ in the UPDRS score may have been biased, as our outcome of interest (EML_{genome}) was associated with probability of follow-up based on a logistic regression. After computing case weights based on the inverse probability censoring weights method, we refit our Cox proportional hazards model that incorporated these weights. The revised hazard ratios for this model, however, remained stable at 1.28 (95% CI: 1.0451–1.564) at a precision of 3 significant figures. Independent replication will be crucial in determining whether and how blood methylation dysregulation contributes to PD risk, motor symptom progression, and/or time to death. Notably, as these signatures of methylation dysregulation were measured in blood and not brain tissue, they may suggest the involvement of peripheral organs such as the immune system in progression and time to death in PD. Also, as methylation patterns were assessed at only one time point in our study after PD onset, we cannot determine directions of action, i.e., whether disease onset caused these blood cell methylation changes or whether the methylation dysregulation contributed to PD risk representing a potential biomarker of processes leading up to PD. It is plausible that SEMs could accumulate in higher density in certain regions (e.g., the five regions identified in the sliding window analysis) in blood and other tissues of patients if certain repair mechanisms fail. Indeed, it has been shown that epigenetic fingerprints in blood can reflect upstream epigenetic changes in other tissues [47, 48]. We have previously observed a significant increase in DNA methylation age in PD patients compared to controls, as assessed by measures of both intrinsic and extrinsic epigenetic age acceleration of blood [32]. In our recent study using the same methylation dataset, we found evidence for a correlation between EML and dysfunction of the epigenetic maintenance system as captured by various measures of methylation age [17].

Given the mounting evidence that methylation plays a major role in aging and aging-related diseases, findings from this study provide interesting implications for potential mechanisms contributing to PD onset, which represent an interplay between genes and the environment. Our results suggest that certain dysfunctions in epigenome repair mechanisms affecting PD gene methylation and pathways may not only be critical in keeping the blood methylome stable but may also influence motor symptom progression in women and time to death in all PD patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by the National Institute of Environmental Health Science (grant numbers R21ES024356, 2R01ES010544, U54ES012078).

REFERENCES

[1]. Willis AW (2013) Parkinson disease in the elderly adult. Mo Med 110, 406–410. [PubMed: 24279192]

[2]. Poewe W, Mahlknecht P (2009) The clinical progression of Parkinson's disease. Parkinsonism Relat Disord 15, S28–S32. [PubMed: 20123553]

[3]. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, Tan M, Kia DA, Noyce AJ, Xue A, Bras J, Young E, von Coelln R, Simón-Snáchez J, Schulte C, Sharma M, Krohn L, Pihlstrøm L, Siitonen A, Iwaki H, Leonard H, Faghri F, Gibbs JR, Hernandez DG, Scholz SW, Botia JA, Martinez M, Corvol JC, Lesage S, Jankovic J, Shulman LM, Sutherland M, Tienari P, Majamaa K, Toft M, Andreassen OA, Bangale T, Brice A, Yang J, Gan-Or Z, Gasser T, Heutink P, Shulman JM, Wood NW, Hinds DA, Hardy JA, Morris HR, Gratten J, Visscher PM, Graham RR, Singleton AB, Team 23andMe Research, Consortium SG of PD, Consortium IPDG (2019) Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol 18, 1091–1102. [PubMed: 31701892]

[4]. Simón-Sánchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, Paisan-Ruiz C, Lichtner P, Scholz SW, Hernandez DG, Krüger R, Federoff M, Klein C, Goate A, Perlmutter J, Bonin M, Nalls MA, Illig T, Gieger C, Houlden H, Steffens M, Okun MS, Racette BA, Cookson MR, Foote KD, Fernandez HH, Traynor BJ, Schreiber S, Arepalli S, Zonozi R, Gwinn K, van der Brug M, Lopez G, Chanock SJ, Schatzkin A, Park Y, Hollenbeck A, Gao J, Huang X, Wood NW, Lorenz D, Deuschl G, Chen H, Riess O, Hardy JA, Singleton AB, Gasser T (2009) Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet 41, 1308–1312. [PubMed: 19915575]

[5]. Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, Kawaguchi T, Tsunoda T, Watanabe M, Takeda A, Tomiyama H, Nakashima K, Hasegawa K, Obata F, Yoshikawa T, Kawakami H, Sakoda S, Yamamoto M, Hattori N, Murata M, Nakamura Y, Toda T (2009) Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat Genet 41, 1303–1307. [PubMed: 19915576]

[6]. Ammal Kaidery N, Tarannum S, Thomas B (2013) Epigenetic landscape of Parkinson's disease: emerging role in disease mechanisms and therapeutic modalities. Neurotherapeutics 10, 698–708. [PubMed: 24030213]

[7]. 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. [PubMed: 28851441]

[8]. Chuang YH, Lu AT, Paul KC, Folle AD, Bronstein JM, Bordelon Y, Horvath S, Ritz B (2019) Longitudinal epigenome-wide methylation study of cognitive decline and motor progression in Parkinson's disease. J Parkinsons Dis 9, 389–400. [PubMed: 30958317]

[9]. Lardenoije R, Iatrou A, Kenis G, Kompotis K, Steinbusch HW, Mastroeni D, Coleman P, Lemere CA, Hof PR, van den Hove DL, Rutten BP (2015) The epigenetics of aging and neurodegeneration. Prog Neurobiol 131, 21–64. [PubMed: 26072273]

[10]. Rakyan VK, Down TA, Balding DJ, Beck S (2011) Epigenome-wide association studies for common human diseases. Nat Rev Genet 12, 529–541. [PubMed: 21747404]

[11]. Kaut O, Schmitt I, Tost J, Busato F, Liu Y, Hofmann P, Witt SH, Rietschel M, Fröhlich H, Wüllner U (2017) Epigenome-wide DNA methylation analysis in siblings and monozygotic twins discordant for sporadic Parkinson's disease revealed different epigenetic patterns in peripheral blood mononuclear cells. Neurogenetics 18, 7–22. [PubMed: 27709425]

[12]. Gentilini D, Garagnani P, Pisoni S, Bacalini MG, Calzari L, Mari D, Vitale G, Franceschi C, Di Blasio AM (2015) Stochastic epigenetic mutations (DNA methylation) increase exponentially in human aging and correlate with X chromosome inactivation skewing in females. Aging 7, 568–578. [PubMed: 26342808]

[13]. Spada E, Calzari L, Corsaro L, Fazia T, Mencarelli M, Di Blasio AM, Bernardinelli L, Zangheri G, Vignali M, Gentilini D (2020) Epigenome wide association and stochastic epigenetic mutation analysis on cord blood of preterm birth. Int J Mol Sci 21, 5044.

[14]. Gagliardi A, Dugué PA, Nøst TH, Southey MC, Buchanan DD, Schmidt DF, Makalic E, Hodge AM, English DR, Doo NW, Hopper JL, Severi G, Baglietto L, Naccarati A, Tarallo S, Pace L, Krogh V, Palli D, Panico S, Sacerdote C, Tumino R, Lund E, Giles GG, Pardini B, Sandanger TM, Milne RL, Vineis P, Polidoro S, Fiorito G (2020) Stochastic epigenetic mutations are associated with risk of breast cancer, lung cancer, and mature B-cell neoplasms. Cancer Epidemiol Biomark Prev 29, 2026–2037.

[15]. Feinberg AP, Irizarry RA (2010) Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc Natl Acad Sci U S A 107, 1757–1764. [PubMed: 20080672]

[16]. Gentilini D, Scala S, Gaudenzi G, Garagnani P, Capri M, Cescon M, Grazi GL, Bacalini MG, Pisoni S, Dicitore A, Circelli L, Santagata S, Izzo F, Di Blasio AM, Persani L, Franceschi C, Vitale G (2017) Epigenome-wide association study in hepatocellular carcinoma: Identification of stochastic epigenetic mutations through an innovative statistical approach. Oncotarget 8, 41890–41902. [PubMed: 28514750]

[17]. Yan Q, Paul KC, Lu AT, Kusters C, Binder AM, Horvath S, Ritz B (2020) Epigenetic mutation load is weakly correlated with epigenetic age acceleration. Aging 12, 17863–17894. [PubMed: 32991324]

[18]. Ritz BR, Paul KC, Bronstein JM (2016) Of pesticides and men: a California story of genes and environment in Parkinson's disease. Curr Env Health Rep 3, 40–52. [PubMed: 26857251]

[19]. Costello S, Cockburn M, Bronstein J, Zhang X, Ritz B (2009) Parkinson's disease and residential exposure to maneb and paraquat from agricultural applications in the central valley of California. Am J Epidemiol 169, 919–926. [PubMed: 19270050]

[20]. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5, R80. [PubMed: 15461798]

[21]. Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD (2013) Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res 41, e90. [PubMed: 23476028]

[22]. 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. [PubMed: 24478339]

[23]. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, Greenwood CM, Hansen KD (2014) Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol 15, 503. [PubMed: 25599564]

[24]. Ritz B, Rhodes SL, Bordelon Y, Bronstein J (2012) α-Synuclein genetic variants predict faster motor symptom progression in idiopathic Parkinson disease. PLoS One 7, e36199. [PubMed: 22615757]

[25]. Hoehn MM, Yahr MD (1998) Parkinsonism: Onset, progression, and mortality. Neurology 50, 318. [PubMed: 9484345]

[26]. Folle AD, Paul KC, Kusters CD, Bronstein JM, Kenner AM, Ritz B (2020) Parkinson's disease motor and non-motor features accompanying insomnia and excessive daytime sleepiness symptoms, a large population-based study. OBM Geriatr 4, 1–17.

[27]. Shulman LM, Gruber-Baldini AL, Anderson KE, Fishman PS, Reich SG, Weiner WJ (2010) The clinically important difference on the unified Parkinson's disease rating scale. Arch Neurol 67, 64–70. [PubMed: 20065131]

[28]. Foulger RE, Denny P, Hardy J, Martin MJ, Sawford T, Lovering RC (2016) Using the gene ontology to annotate key players in Parkinson's disease. Neuroinformatics 14, 297–304. [PubMed: 26825309]

[29]. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14, R115. [PubMed: 24138928]

[30]. van Dongen J, Bonder MJ, Dekkers KF, Nivard MG, van Iterson M, Willemsen G, Beekman M, van der Spek A, van Meurs JBJ, Franke L, Heijmans BT, van Duijn CM, Slagboom PE, Boomsma DI, BIOS consortium (2018) DNA methylation signatures of educational attainment. NPJ Sci Learn 3, 7. [PubMed: 30631468]

[31]. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai PC, Roetker NS, Just AC, Demerath EW, Guan W, Bressler J, Fornage M, Studenski S, Vandiver AR, Moore AZ, Tanaka T, Kiel DP, Liang L, Vokonas P, Schwartz J, Lunetta KL, Murabito JM, Bandinelli S, Hernandez DG, Melzer D, Nalls M, Pilling LC, Price TR, Singleton AB, Gieger C, Holle R, Kretschmer A, Kronenberg F, Kunze S, Linseisen J, Meisinger C, Rathmann W, Waldenberger M, Visscher PM, Shah S, Wray NR, McRae AF, Franco OH, Hofman A, Uitterlinden AG, Absher D, Assimes T, Levine ME, Lu AT, Tsao PS, Hou L, Manson JE, Carty CL, LaCroix AZ, Reiner AP, Spector TD, Feinberg AP, Levy D, Baccarelli A, van Meurs J, Bell JT, Peters A, Deary IJ, Pankow JS, Ferrucci L, Horvath S (2016) DNA methylation-based measures of biological age: meta-analysis predicting time to death. Aging 8, 1844–1865. [PubMed: 27690265]

[32]. Horvath S, Ritz BR (2015) Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. Aging 7, 1130–1142. [PubMed: 26655927]

[33]. McRae AF, Powell JE, Henders AK, Bowdler L, Hemani G, Shah S, Painter JN, Martin NG, Visscher PM, Montgomery GW (2014) Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome Biol 15, R73. [PubMed: 24887635]

[34]. Paul KC, Schulz J, Bronstein JM, Lill CM, Ritz BR (2018) Association of polygenic risk score with cognitive decline and motor progression in Parkinson disease. JAMA Neurol 75, 360–366. [PubMed: 29340614]

[35]. He Z, Xu B, Buxbaum J, Ionita-Laza I (2019) A genome-wide scan statistic framework for whole-genome sequence data analysis. Nat Commun 10, 3018. [PubMed: 31289270]

[36]. Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X (2013) Sequence kernel association tests for the combined effect of rare and common variants. Am J Hum Genet 92, 841–853. [PubMed: 23684009]

[37]. Dzamko N, Zhou J, Huang Y, Halliday GM (2014) Parkinson's disease-implicated kinases in the brain; insights into disease pathogenesis. Front Mol Neurosci 7, 57. [PubMed: 25009465]

[38]. Tofaris GK, Kim HT, Hourez R, Jung JW, Kim KP, Goldberg AL (2011) Ubiquitin ligase Nedd4 promotes alpha-synuclein degradation by the endosomal-lysosomal pathway. Proc Natl Acad SciUSA 108, 17004–17009.

[39]. Alvarez-Prats A, Bjelobaba I, Aldworth Z, Baba T, Abebe D, Kim YJ, Stojilkovic SS, Stopfer M, Balla T (2018) Schwann-cell-specific deletion of phosphatidylinositol 4-kinase alpha causes aberrant myelination. Cell Rep 23, 2881–2890. [PubMed: 29874576]

[40]. Subrahmanyam G, Rudd CE, Schneider H (2003) Association of T cell antigen CD7 with type II phosphatidylinositol-4 kinase, a key component in pathways of inositol phosphate turnover. Eur J Immunol 33, 46–52. [PubMed: 12594831]

[41]. Webster CM, Hokari M, McManus A, Tang XN, Ma H, Kacimi R, Yenari MA (2013) Microglial P2Y12 deficiency/inhibition protects against brain ischemia. PLoS One 8, e70927. [PubMed: 23940669]

[42]. Stojkovska I, Wagner BM, Morrison BE (2015) Parkinson's disease and enhanced inflammatory response. Exp Biol Med Maywood 240, 1387–1395. [PubMed: 25769314]

[43]. Berwick DC, Harvey K (2012) The importance of Wnt signalling for neurodegeneration in Parkinson's disease. Biochem Soc Trans 40, 1123–1128. [PubMed: 22988876]

[44]. Rawal N, Corti O, Sacchetti P, Ardilla-Osorio H, Sehat B, Brice A, Arenas E (2009) Parkin protects dopaminergic neurons from excessive Wnt/β-catenin signaling. Biochem Biophys Res Commun 388, 473–478. [PubMed: 19591802]

[45]. Inestrosa NC, Montecinos-Oliva C, Fuenzalida M (2012) Wnt signaling: role in Alzheimer disease and schizophrenia. J Neuroimmune Pharmacol 7, 788–807. [PubMed: 23160851]

[46]. Kohn AD, Moon RT (2005) Wnt and calcium signaling: β-Catenin-independent pathways. Cell Calcium 38, 439–446. [PubMed: 16099039]

[47]. Lowe R, Rakyan VK (2014) Correcting for cell-type composition bias in epigenome-wide association studies. Genome Med 6, 23. [PubMed: 25031617]

[48]. Breeze CE, Paul DS, van Dongen J, Butcher LM, Ambrose JC, Barrett JE, Lowe R, Rakyan VK, Iotchkova V, Frontini M, Downes K, Ouwehand WH, Laperle J, Jacques P-É, Bourque G, Bergmann AK, Siebert R, Vellenga E, Saeed S, Matarese F, Martens JHA, Stunnenberg HG, Teschendorff AE, Herrero J, Birney E, Dunham I, Beck S (2016) eFORGE: a tool for identifying cell type-specific signal in epigenomic data. Cell Rep 17, 2137–2150. [PubMed: 27851974]

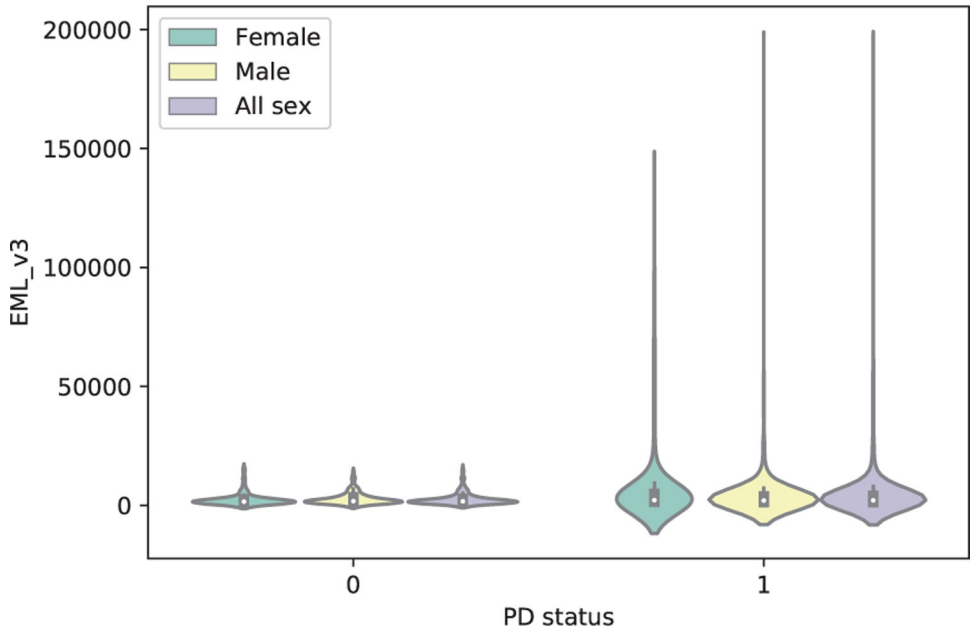

Fig. 1.

Distribution of EML_{genome} comparing cases (PD status = 1) and controls (PD status = 0) for PEG study wave 1 and 2.

J Parkinsons Dis. Author manuscript; available in PMC 2023 January 01.

Table 1

Characteristics of PEG participants

		Controls 238 (30%)	PEG1 Cases 341 (42%)	PEG2 Cases 227 (28%)
Sex	Female	111	144	69
	Male	127	197	158
Race	White	219	292	168
	African American	0	3	0
	Asian	0	4	7
	Hispanics	19	42	51
	Other	0	0	1
Years education	<12	27	62	38
	12	50	91	38
	>12	161	188	151
Years follow-up [1]	Mean (range)	NA	6.5 (1–15.7)	3.8 (2.3–6.8)
Baseline mUPDRS	Mean (std dev)	NA	21.6 (10.9)	24.0 (11.9)
Baseline HY	Mean (std dev)	NA	2.2 (0.8)	2.3 (0.8)
Smoker Status	Never	96	180	122
	Quit	125	144	97
	Current	17	17	7
	NA	0	0	1
Age at diagnosis of cases or interview of controls	Mean (range)	67.4 (35.1–91.9)	70.2 (36.5–90.5)	70.9 (46.4–92)
Age at death	Mean (range)	80.5 (50–96)	80.9 (52.4–97)	79.9 (54–93.9)
Years PD duration[2]	Mean (range)	NA	2.1 (0–9)	3.8 (0–15)

[1] Time span in years between first and last exams for motor progression.

[2] Time span in years from diagnosis until baseline exam.

Table 2

Odds ratios of PD onset for the variables EML_{genome} and EML_{PD}

		PEG1 and PEG2 (*n* = 806)		PEG1 only (sensitivity analysis, *n* = 579)	
Stratum	**EML scoring criterion**	**Odds Ratio (95% CI)**	***p***	**Odds Ratio (95% CI)**	***p***
All	genome	1.90 (1.52–2.37)	2.1e-8	1.72 (1.38–2.13)	1e-6
	PD-genes	1.87 (1.50–2.32)	2.1e-8	1.66 (1.36–2.04)	2.1e-6
Female	genome	2.28 (1.59–3.25)	6.5e-6	2.05 (1.45–2.92)	5.6e-5
	PD-genes	2.13 (1.52–2.97)	9.0e-6	1.88 (1.36–2.60)	1.2e-4
Male	genome	1.66 (1.22–2.24)	1.1e-3	1.46 (1.08–1.97)	0.013
	PD-genes	1.67 (1.23–2.27)	1.1e-3	1.44 (1.07–1.93)	0.017

Models control for sex, age, ethnicity, and cell counts (Gran, NK, CD8pCD28nCD45RAn, CD4T).

Table 3

Mapping of EML<windows> to PD risk

Region*	Gene description	Suggested function	Start position	Dispersion p-value	Burden p-value
2.STK39	serine/threonine kinase 39	this kinase may serve as an intermediate in the response to cellular stress	169079247	2.38E-07	2.38E-07
3.P2RY12	Purinergic Receptor P2Y12	chemoreceptor for adenosine diphosphate (ADP)[5] [6] that belongs to the Gi class of a group of G protein-coupled (GPCR) purinergic receptors. In the central nervous system, this receptor has been found expressed exclusively on microglia, where it is necessary for physiological and pathological microglial actions, such as monitoring neuronal functions and microglial neuroprotection	151055545	1.86E-07	1.74E-08
11.DKK3	Dickkopf-related protein 3	involved in embryonic development through its interactions with the Wnt signaling pathway	12026850	7.22E-09	6.11E-08
15.NEDD4	E3 ubiquitin-protein ligase	regulates a large number of membrane proteins, such as ion channels and membrane receptors, via ubiquitination and endocytosis;	56282277	4.61E-08	2.78E-08
22.PI4KA	1-phosphatidylinositol 4-kinase	Responsible for biosynthetic trafficking of two different classes of proteins from the ER to the Golgi complex.	21138859	4.58E-08	5.90E-08

*Chromosome.GeneSymbol

Table 4

Hazard ratios for EML_{genome}/EML_{PD} measures and time to event - (having a score of 35+ on mUPDRS) and time to death in PD patients

Outcome	Stratum	EML scoring criterion	Hazard Ratio (95% CI)	*p*
mUPDRS 35	All (*n* = 455)	genome	1.28 (1.06–1.56)	0.011
		PD-genes	1.30 (1.08–1.56)	0.062
	Female (*n* = 165)	genome	1.43 (1.10–1.86)	7.7e-3
		PD-genes	1.41 (1.09–1.82)	8.0e-3
	Male (*n* = 290)	genome	1.06 (0.79–1.43)	0.69
		PD-genes	1.11 (0.83–1.48)	0.50
Time to death	All (*n* = 357)	genome	1.29 (1.11–1.49)	6.9e-4
		PD-genes	1.29 (1.12–1.48)	3.6e-4
	Female (*n* = 138)	genome	1.33 (1.07–1.64)	9.1e-3
		PD-genes	1.31 (1.07–1.60)	0.088
	Male (*n* = 165)	genome	1.16 (0.94–1.45)	0.17
		PD-genes	1.21 (0.97–1.49)	0.088

Models control for sex, age, ethnicity, cell counts (Gran, NK, CD8pCD28nCD45RAn, CD4T), total levodopa equivalent (total_led), PRS (polygenic risk score), and PD duration in years.