

<https://doi.org/10.1038/s41531-025-00877-5>

# Epigenetic insights into neuropsychiatric and cognitive symptoms in Parkinson's disease: A DNA co-methylation network analysis

Check for updates

Joshua Harvey<sup>1</sup>, Adam R. Smith<sup>1</sup>, Luke S. Weymouth<sup>1</sup>, Rebecca G. Smith<sup>1</sup>, Isabel Castanho<sup>2</sup>, Leon Hubbard<sup>3</sup>, Byron Creese<sup>1,4</sup>, Catherine Bresner<sup>3</sup>, Nigel Williams<sup>3</sup>, Ehsan Pishva<sup>1,5</sup> & Katie Lunnon<sup>1</sup> ✉

Parkinson's disease is a highly heterogeneous disorder, encompassing a complex spectrum of clinical presentation including motor, sleep, cognitive and neuropsychiatric symptoms. We aimed to investigate genome-wide DNA methylation networks in post-mortem Parkinson's disease brain samples and test for region-specific association with common neuropsychiatric and cognitive symptoms. Of traits tested, we identify a co-methylation module in the substantia nigra with significant correlation to depressive symptoms. Notably, expression of the genes annotated to the methylation loci present within this module are found to be significantly enriched in neuronal subtypes within the substantia nigra. These findings highlight the potential involvement of neuronal-specific changes within the substantia nigra with regards to depressive symptoms in Parkinson's disease.

Parkinson's disease (PD) is the second most common neurodegenerative disease and is the fastest growing in prevalence of all neurological disorders, estimated to affect 6.1 million individuals worldwide based on a 2016 census<sup>1</sup>. Clinically, PD is defined by its cardinal motor symptoms (resting tremor, bradykinesia, rigidity and postural instability)<sup>2</sup>, but highly prevalent features of the disease encompass a range of cognitive and neuropsychiatric symptoms<sup>3</sup>. Common symptoms, reported in a high proportion of patients, include depression<sup>4</sup>, anxiety<sup>5</sup>, psychosis (most prominently hallucinations and delusions)<sup>6</sup>, apathy<sup>7</sup>, cognitive impairment and dementia<sup>8</sup>. The cumulative effect of these secondary symptoms greatly increases disease burden for patients and complicates treatment<sup>8,9</sup>. As examples, psychosis is an associated factor to increased nursing home placement<sup>10</sup>, mortality and caregiver burden in PD<sup>11</sup>. Dopaminergic therapies, highly prescribed for motor symptom treatment, reportedly increase individual risk for the emergence of psychosis symptoms<sup>12</sup>. The development of these secondary symptoms is not always timed after the diagnosis of the primary motor disorder, for example, depression is a common manifestation in premorbid PD and has been associated as a risk factor for motor symptom development<sup>13–15</sup>. Furthermore, the therapies that exist for these non-motor

symptoms are currently minimally effective, despite the considerable disease burden they represent.

Although the occurrence of neuropsychiatric and cognitive symptoms in PD is much more common than in age-matched populations<sup>8,9</sup>, individual to individual level susceptibility to these secondary features is highly variable<sup>16,17</sup>. Genetic liability has been implicated, for example a recent genome-wide association study (GWAS) of cognitive progression in PD highlighted the contribution of risk genes such as *GBA* with worsening cognitive decline over time<sup>18</sup> and meta-analyses of the gene have shown an association to the emergence of psychosis and depression symptoms<sup>19</sup>. However, given the high levels of heterogeneity within the condition, PD secondary symptoms likely share a complex underlying etiology, owing to additional factors aside from genetics. One potential contributing factor is epigenetic changes, which play an intermediary role between genetic and environmental risk, and regulate gene expression<sup>20</sup>. DNA methylation, which refers to the reversible addition of methyl groups to cytosines typically in a CpG dinucleotide, is the most studied epigenetic mechanism in neurological disorders<sup>21</sup>. Indeed, several studies have shown robust alterations in DNA methylation in a number of genes in different neurodegenerative

<sup>1</sup>University of Exeter Medical School, Department of Clinical and Biomedical Sciences, Faculty of Health and Life Sciences, University of Exeter, Exeter, UK.

<sup>2</sup>Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. <sup>3</sup>Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, Wales, UK. <sup>4</sup>Department of Life Sciences, College of Health, Medicine and Life Sciences, Brunel University of London, London, UK. <sup>5</sup>Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNS), Maastricht University, Maastricht, The Netherlands. ✉e-mail: [K.Lunnon@exeter.ac.uk](mailto:K.Lunnon@exeter.ac.uk)



diseases, in both the brain and blood, including Alzheimer's disease (AD)<sup>22–24</sup>, PD<sup>25–27</sup>, and Dementia with Lewy bodies (DLB)<sup>28</sup>. Interestingly, associations have also been reported for secondary symptoms of these neurodegenerative disease, for example with psychosis symptoms in AD<sup>29,30</sup> or cognition in PD<sup>26</sup>. However the analysis of DNA methylation signatures in relation to PD secondary symptoms is understudied and has predominantly been undertaken in peripheral tissues such as blood<sup>26</sup>.

In the current study we investigated the relationship between DNA methylation patterns and the occurrence of key secondary symptoms in PD (dementia, hallucinations, depression, anxiety, aggression, sleep disorder), using weighted gene correlation network analysis (WGCNA) in multiple disease-relevant brain regions. Subsequently, gene ontology and cell type enrichment analysis were performed on the genes comprising the significant modules to identify dysfunctional pathways and the cell types likely driving this. We highlight a core finding of a co-methylation module specific to the substantia nigra, significantly correlated to depressive symptom presentation. Assessing the expression of genes annotated to this module found enriched expression in specific neuronal sub-populations within the substantia nigra, indicative of neuronal changes within this region that may play a role in the development of depressive symptoms within PD. Utilizing genotyping and longitudinal clinical data from the Parkinson's Progression Marker's Initiative (PPMI) we find that refined polygenic risk scores (PRS) from loci annotated the depression epigenetic signature show evidence of association with depression risk within PD.

## Results

### A cohort to assess DNA methylation signatures of PD neuropsychiatric and cognitive symptoms

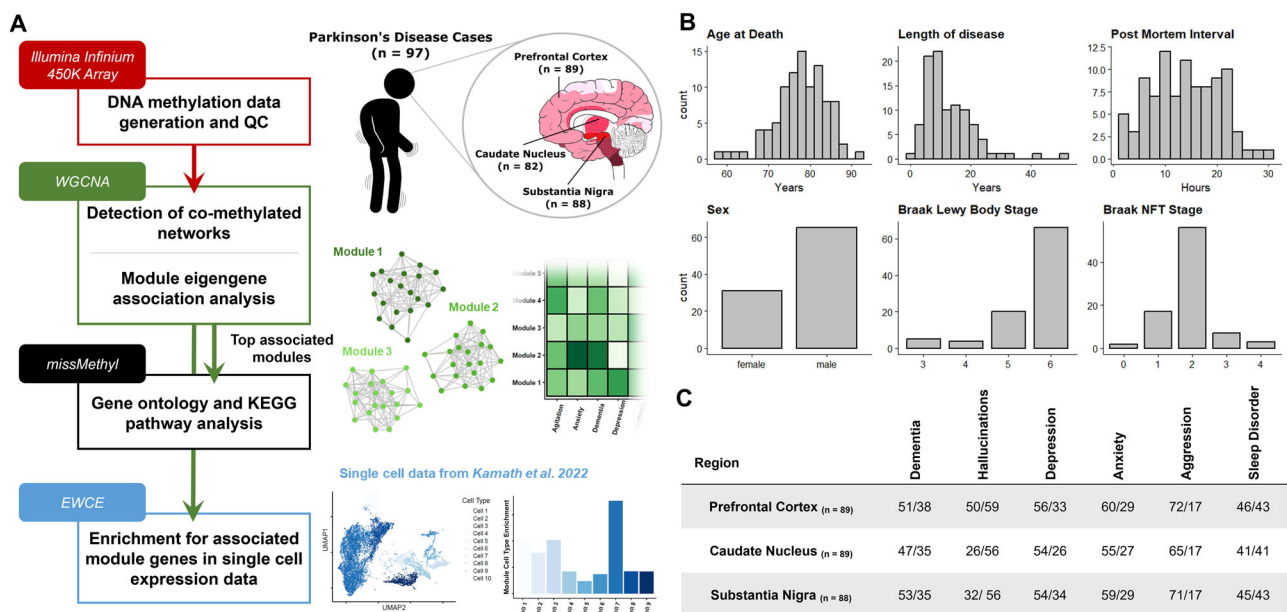
Our study comprised a cohort of 97 idiopathic Parkinson's Disease (PD) patients with post-mortem DNA methylomic profiling conducted on the Illumina Infinium 450 K array (Fig. 1A). Three brain regions were assessed: the substantia nigra (SN,  $n = 88$ ), caudate nucleus (CN,  $n = 82$ ) and prefrontal cortex (FC,  $n = 88$ ), with the majority of cases having all brain regions represented in this dataset (Fig. 1A, C). PD patients had a mean age of 78.25 years at death (SD = 6.17) with the average patient having had PD symptoms for over ten years (mean = 12.63, SD = 8.28). Pathologically, these patients predominantly showed late stage PD-associated Lewy body (LB)

pathology<sup>12</sup> (LB Braak stage mean = 5.54, SD = 0.81) with relatively mild AD-associated neurofibrillary tangle (NFT) pathology<sup>31,32</sup> (NFT Braak stage mean = 1.91, SD = 0.72) (Fig. 1B).

The primary hypothesis of this study posits that the prevailing neuropsychiatric and cognitive manifestations observed in PD exhibit a distinctive epigenetic profile in the brain, distinguishing individuals presenting with these symptoms from those who do not. To test this hypothesis, we annotated binary symptom prevalence from antemortem clinical records for six phenotypes: dementia, hallucinations, depression, anxiety, aggression, and sleep disorder (Methods, Fig. 1C). The majority of these sub-symptoms showed overlap in their presentation and demographic differences (Supplementary Fig. 1, Supplementary Table 1) as could be expected with cumulative disease burden<sup>5</sup>. To identify DNA methylation signatures associated with the phenotypes of interest, we investigated co-methylation changes by implementing WGCNA. This was also useful as a strategy to reduce the number of features and increase statistical power, given our modest sample size for conducting epigenome-wide association studies (EWAS), particularly when examining binary variables related to phenotypes of interest.

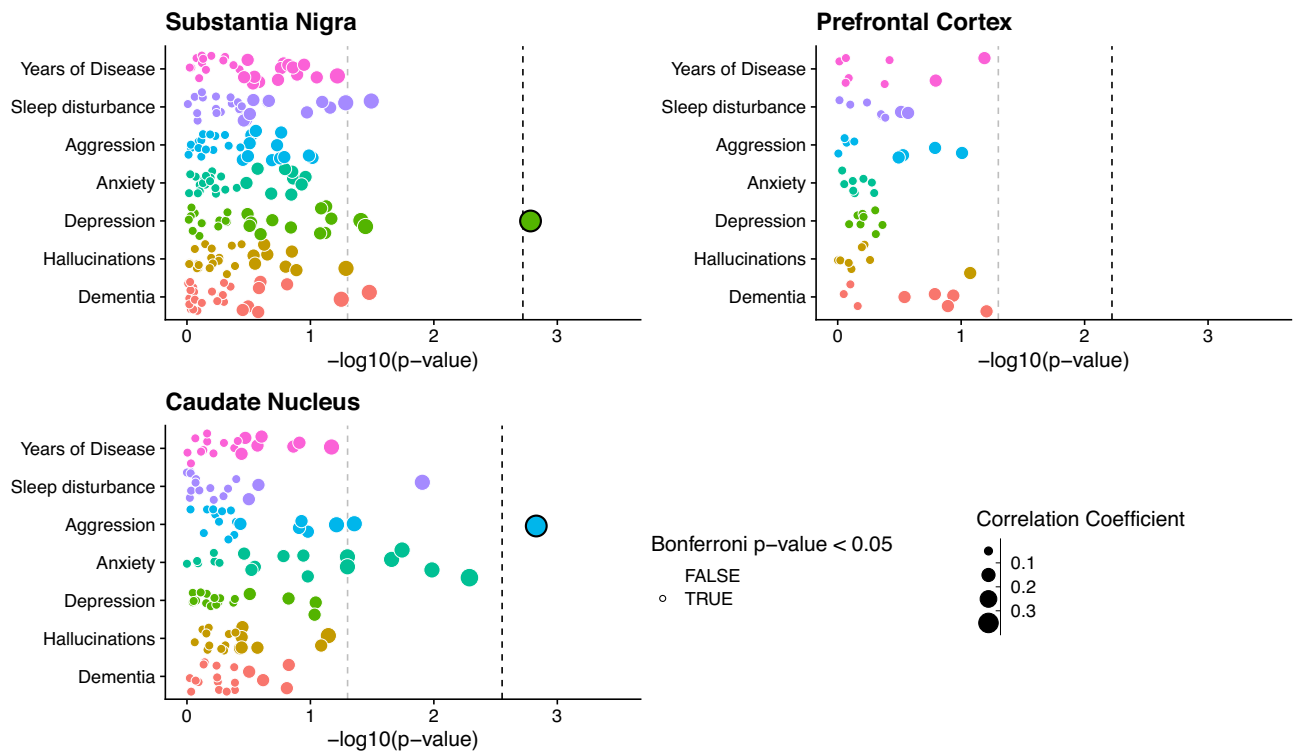
### DNA co-methylation networks show brain region specific correlation to depressive and aggression symptoms

To identify co-methylated modules within each brain region, we followed a standardized WGCNA protocol (Methods), and tested their association with sub-trait presentation, after regressing out key covariates (age, sex, technical batch, proportions of neurons, post-mortem interval (PMI)). The number of detected modules differed across each brain region, with 27 modules identified in the SN (Supplementary Fig. 2), eight in the FC (Supplementary Fig. 3) and 18 in the CN (Supplementary Fig. 4). The correlation of these modules to trait presentation also differed across brain regions. Stronger module-trait correlations were observed in the SN and CN, with two modules passing the Bonferroni significance threshold for the number of tests within each trait association (Fig. 2, SN:  $P < 0.0019$ , CN:  $P < 0.0028$ ), with no significant correlations in the FC (Fig. 2). The significant SN module correlated with depressive symptoms in PD (Spearman's Coefficient = 0.33,  $P = 0.0016$ ) and was comprised of 1375 distinct methylated loci, while the significant CN module that was significantly



**Fig. 1 | Overview of study design and samples used.** A Sample information and data analysis flowchart as detailed in Methods. Brain sagittal view showing Braak Lewy body staging representation<sup>32</sup>. Data used in the creation of UMAP plot from Kamath et al.<sup>34</sup> B Demographic summaries of PD patients profiled with histograms or bar-charts representing overall numbers for each variable. Length of disease corresponds

to time interval between recorded age of PD diagnosis and age at death. C Summary table of sub-symptoms tested in primary WGCNA association analysis. Symptoms were annotated for binary status in the ante-mortem clinical records. Symptom prevalence per each brain region tested is shown and is annotated as Absent/Present.



**Fig. 2 | Co-methylation network association to sub symptom occurrence in PD.** Points represent individual module eigengenes for the substantia nigra ( $n = 27$ ), frontal cortex ( $n = 8$ ) and caudate nucleus ( $n = 18$ ), repeatedly tested using correlation analysis in association with traits displayed along the y axis. Points are colored by the trait they are being tested for association with and sized by the absolute correlation coefficient value of the association. For clinical binary traits Spearman’s

correlation was used, while for years of disease Pearson’s correlation was used. The  $-\log_{10}(p\text{-value})$  of the association tests is displayed along the X-axis. The gray dashed line represents  $P\text{-value} = 0.05$ , while the Black dashed line represents the Bonferroni correction threshold for each brain region, controlling for the number of tests within each trait association, equivalent to 0.05 divided by the number of module eigengenes per region (SN:  $P < 0.0019$ , CN:  $P < 0.0028$ , FC  $P < 0.006$ ).

correlated to aggression presentation (Spearman’s Coefficient = 0.35,  $P = 0.0015$ ) was comprised of 475 distinct methylated loci.

When assessing module membership of these two significant modules, a weak but significant correlation was observed between  $P\text{-value}$  significance of depressive symptom association and module membership for methylated loci within the SN depression associated module (Pearson’s Coefficient = 0.12,  $p\text{-value} = 1.24 \times 10^{-5}$ , Supplementary Fig. 4A). By contrast the CN aggression associated module does not show any indication of correlation between module membership and probe significance from the aggression symptom association (Pearson’s Coefficient = 0.07,  $p\text{-value} = 0.13$ , Supplementary Fig. 4B).

Although no further modules passed our threshold for multiple testing correction, several other modules in these regions did show nominal significance in their correlation with trait presentation (Fig. 2, Supplementary Fig. 2, Supplementary Fig. 4). Of particular note, a set of four modules in the CN all showed correlation with anxiety symptoms. However, we have focused our downstream analyses on the Bonferroni-significant module identified in the SN (with respect to depression) and the CN (associated with aggression), henceforth referred to as the DepressionSN module and the AggressionCN module, respectively.

To test whether the association of the DepressionSN module was affected by onset of depression before motor symptoms we subset the depression group based on annotation of depressive symptoms before PD diagnosis (Premorbid depression,  $n = 9$ ) versus those without annotation preceding PD diagnosis (Depression,  $n = 23$ ) and compared both groups to the group without depression annotation (No Depression,  $n = 54$ ). Both premorbid depression and depression groups showed increased eigengene values compared to the non-depressed group (Supplementary Fig. 6A). A pairwise comparison of the three groups with a Wilcoxon rank sum test, with Benjamini-Hochberg (BH) correction found a significant difference

between the non-depressed and depressed group ( $q\text{-value} = 0.02$ ), while no significant difference was observed between any of the other groupwise comparisons with the premorbid or non-depressed group. To test if the DepressionSN association is attributable to variation in disease stage, we further explored the effect of regressing out Braak Lewy body stage, Braak neurofibrillary tangle stage and years of disease from the eigengene. We found this to have no significant effect on the depression association we had observed, with comparable directions of effect and Wilcoxon rank sum  $p\text{-values}$  in binary depression groupwise comparisons (Supplementary Fig. 6B).

### Genes annotated to the depression-associated DNA co-methylation module show significantly enriched expression in neuronal subtypes of the substantia nigra

Next, to gain insight into potential underlying molecular functions captured by these trait-associated modules, we performed Gene Ontology (GO) analysis in the missMethyl package, a method which tests for enrichment for gene symbols annotated to each methylated loci while correcting for coverage bias of the 450 K array<sup>33</sup>. Neither module tested showed FDR significant enrichment for any gene ontology terms (Supplementary Fig. 8, Supplementary Table 2). We noted a nominal enrichment (uncorrected  $P\text{-value} < 0.01$ ) in the SN depression module for a number terms relevant to neuronal processes (e.g. maintenance of synapse structure and presynaptic active zone) and next sought to elucidate the cell-specific expression of genes annotated to the DNA methylation loci in the DepressionSN module. For this we used a reference set of human single nucleus RNAseq data<sup>34</sup> generated in the SN and using Expression Weighted Cell Enrichment analysis (EWCE) to test for enrichment. Of the 617 genes overlapping between the DepressionSN module methylation dataset and the reference snRNAseq data, we observed significant enrichment of expression in neurons only (Fig.

3A, Supplementary Table 6). Of a total of 68 defined cell subtypes annotated in the original study by Kamath and colleagues, 15 showed a significant expression enrichment for DepressionSN annotated genes (BH corrected  $q$  value  $< 0.05$ ), corresponding to seven excitatory neuronal populations, four inhibitory neuronal populations and four dopaminergic neuronal populations. Of these populations, the two with the highest standard deviation shift from the mean expression were both excitatory: POSTN and OPRD1 (7.44 and 6.35 standard deviations from the mean, respectively).

A potential source of bias influencing this enrichment signal is the nature of the reference dataset used to define cell-type specific gene expression enrichment. Indeed, Kamath et al. had performed fluorescence activated nuclei sorting to enrich for dopaminergic neuron populations, which could bias our findings towards observing neuronal enrichment. To test that the neuronal enrichment we had observed in this analysis was not driven by reference data bias, we went on to perform the same EWCE analysis in two additional SN snRNAseq datasets that had not enriched for specific cell types prior to sequencing. In both of these datasets, we found that genes annotated to the depressionSN module were significantly enriched for expression within neuron subpopulations (Supplementary Tables 7–8). Specifically, these corresponded to GABAergic neurons in the Agarwal et al. dataset (Supplementary Table 7) and CADPS2 positive neurons in the Smajic et al. dataset (Supplementary Table 8). Although these three datasets do not provide a consensus on the exact neuronal subtype implicated in our DepressionSN module, they do provide evidence that the network of genes present within the DepressionSN module has functional relevance in disease as it is likely driven by SN neuronal cell types.

Because of the broad neuronal signature identified in the initial analysis, we next sought to refine the analysis by conditional testing of cell-type enrichment, controlling for other significantly enriched populations and testing which subtypes preserve significance. This was solely conducted in the Kamath dataset, as it had the highest representation of neuronal subtypes. As expected, controlling for each of the 15 significant neuronal subtypes showed a general reduction of significance in the others, with POSTN and OPRD1 excitatory neurons being the only populations retaining significance in 13/15 analyses conducted (Fig. 3B). Notably, all dopaminergic neuron populations did not retain significance in any of the conditional analyses, except for CALB1\_GEM which showed retained significance in 6/15 conditional analyses. From these, we can conclude that there is strongest evidence for involvement of POSTN and OPRD1 excitatory neurons, independent of the other significant cell types detected in the initial enrichment analysis.

### Refinement of depression polygenic risk using loci from the DepressionSN module shows evidence of predicting depression development risk in PD

A growing body of research has shown that refinement of polygenic risk scores (PRSs) based on prioritized genomic regions can explain additional disease heterogeneity and, in some cases, improve phenotypic association<sup>35</sup>. We next tested whether the regions of the genome prioritized by the methylated signal in the DepressionSN module refined heritable genetic risk scores for a recent GWAS of depression<sup>36</sup>, in association to depressive state in PD using data available from the Parkinson's Progression Markers Initiative (PPMI).

PRSs for all available variants (base) as well as PRSs refined to variants within one megabase of prioritized methylated loci in the DepressionSN module (refined) were tested at multiple  $p$ -value thresholds using PRSet (Fig. 4A, Supplementary Table 9). A pattern was observed where higher  $p$ -value threshold ( $P = 0.05 - 1$ ) PRSs showed greater significance in the refined subset compared to the base. This supported evidence that the refined subset improved the base PRS scores at less significantly associated GWAS variants, while at lower  $p$ -value thresholds, the base scores generally performed better than the refined, albeit non-significantly in most cases. When correcting for all  $p$ -value thresholds tested, only the refined PRS for depression at a  $p$ -value threshold of 0.05 surpassed FDR correction for depression association (R2 value = 0.024,  $p$ -value = 0.0014).

To further test both the predictive value of these PRS values with risk of depression development specifically within PD, we employed a cox-proportional hazards model specifically within the PD subset of PPMI, controlling for severity of parkinsonian motor symptoms (MDS-Unified PD Rating Scale, Part-III), age and sex for each individual. We observed that both the base (Hazard Ratio = 1.21, Standard Error = 0.06,  $p$ -value = 0.0016) and refined (Hazard Ratio = 1.20, Standard Error = 0.06,  $p$ -value = 0.0027) PRS scores showed very similar hazard ratios and significance of prediction of conversion to depression within PD. However, when stratifying individual samples into the extremes of their polygenic risk using the base and refined PRS, we observed a greater separation for those samples at the extreme end of risk (95–100% percentile) in the refined PRS (Hazard Ratio = 7.4, Standard Error = 0.53,  $p$ -value =  $1.67 \times 10^{-4}$ , Fig. 4B) compared to the base PRS (Hazard Ratio = 3.51, Standard Error = 0.47,  $p$ -value = 0.0077, Fig. 4C).

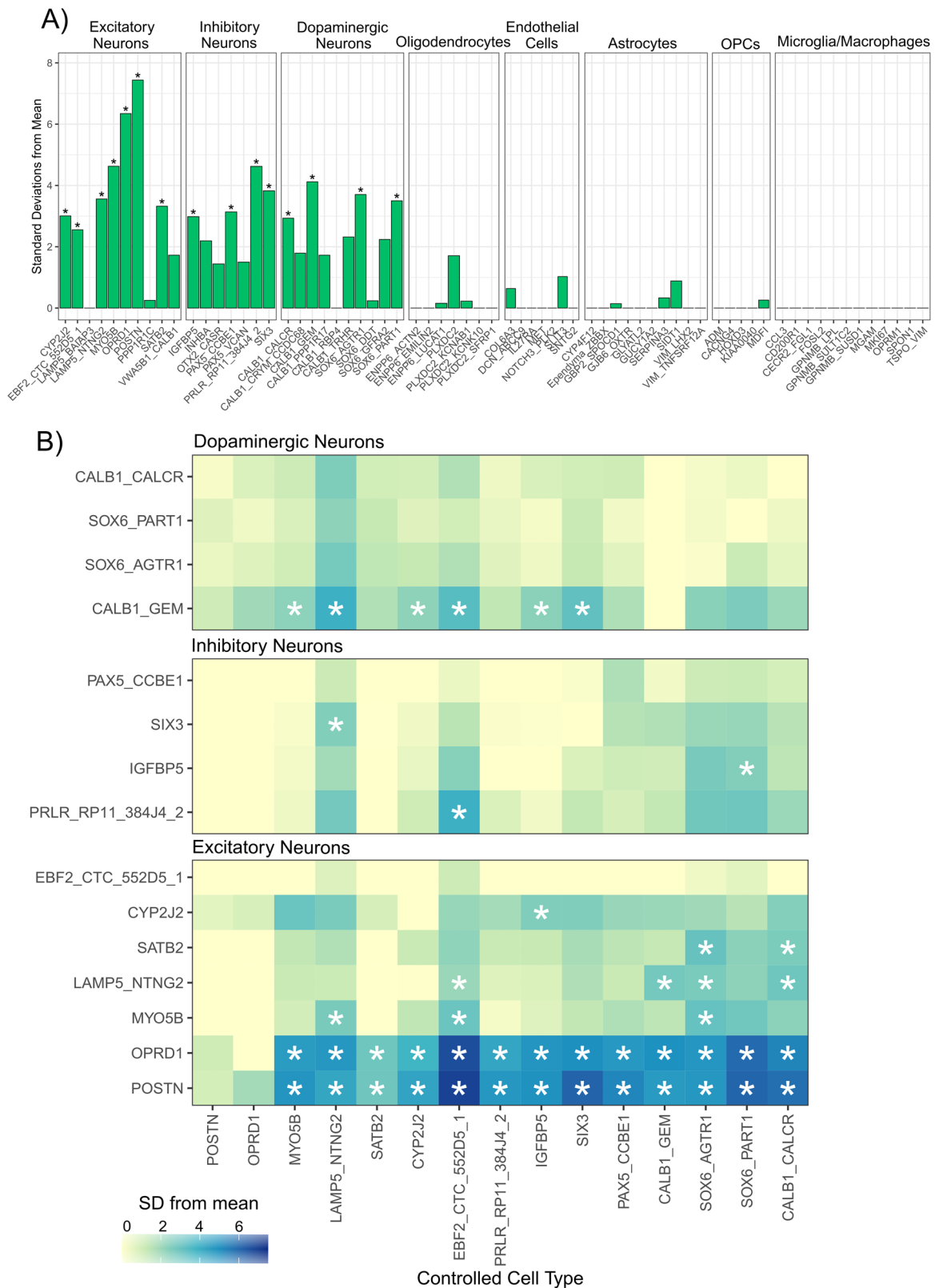
## Discussion

In this report we have investigated multiple common secondary symptom traits in PD across three disease-relevant brain regions and explored the contribution of DNA methylation (summarized into inter-correlated DNA methylation networks) with trait presentation. We report region-specific associations between DNA methylation networks and trait presentation, specifically in association with depression in the SN and aggression in the CN. Subsequent downstream analyses indicated that genes related to the depression-associated SN co-methylation network (DepressionSN module) show significant overrepresentation of genes that are expressed in neuronal cells in the SN, inferred from separate snRNAseq datasets. We finally provide evidence that the genomic regions prioritized by this co-methylation network refine associations of depression PRS within the context of PD.

Depression in PD has a prevalence of roughly 40–46%<sup>37</sup> and is a common premorbid symptom, being a risk factor for both PD development<sup>13</sup> and worse symptom progression over time<sup>38</sup>. The pathophysiology underlying depressive symptoms in PD however remains poorly understood, with multiple potential threads of evidence for its etiology and relation to PD pathological development. Our results relating to SN neuronal changes lend support to midbrain theories of PD depression onset. Previous studies have shown that depressed PD patients present with greater neuronal loss<sup>39</sup> and gliosis<sup>40</sup> in the SN than non-depressed patients. Furthermore, alpha-synuclein pathology in the SN has been reported to be significantly higher in the SN of depressed cases versus non-depressed<sup>39</sup>. This regional neurodegeneration and consequent disruption of neurotransmission may be a contributing factor to the epigenetic alterations we observe in our results. However, the epigenetic network identified appears to be most enriched for expression in non-dopaminergic excitatory neurons, contradicting the evidence that this effect is purely a result of dopaminergic neurodegeneration. Indeed, when we conducted a conditional analysis, the strongest association remained in the excitatory neuronal subpopulations, while the majority of inhibitory and dopaminergic neurons did not retain significance. The sole exception for the dopaminergic neuron populations was the CALB1\_GEM subpopulation, which did retain significant enrichments in certain conditional analyses. This is interesting in the context of the original publication by Kamath et al.<sup>34</sup> which reports increased proportions of this population in PD samples compared to controls. This may indicate that resilient neuronal populations to primary pathology may in fact contribute to the development of secondary symptoms within PD. Further research is needed to fully elucidate the contribution of these SN neuronal cell types in the context of PD depression.

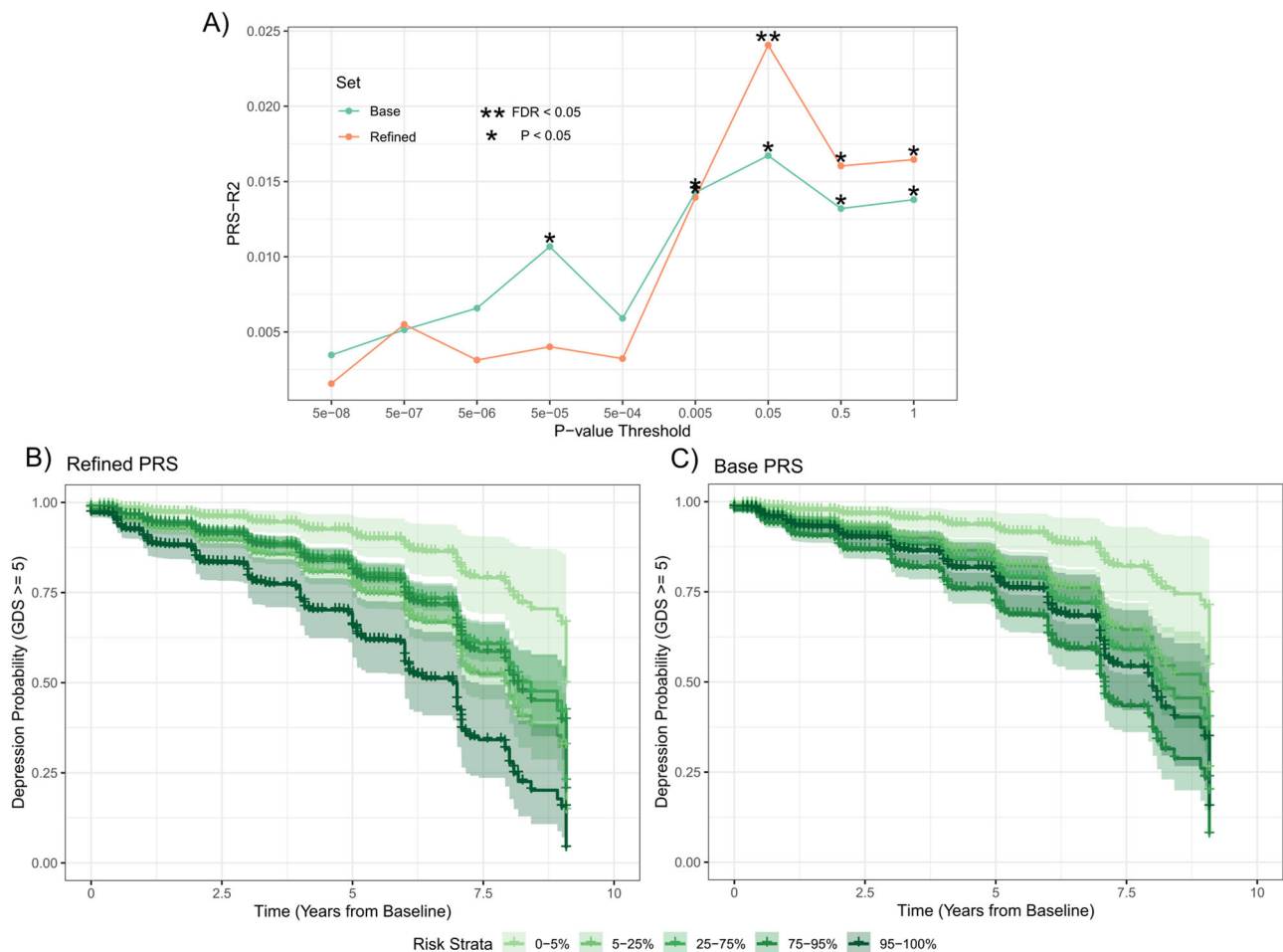
A potential avenue for further research could be in animal models of PD neurodegeneration, specifically in the context of PD depression. A number of rodent studies utilizing neurotoxic compounds such as 1-Methyl-4-Phenyl-1,2,3,6-Tetrahydropyridine (MPTP), which cause selective dopaminergic neuron degeneration, report depression-like behaviors, even manifesting before the onset of motor impairments<sup>41</sup>. Importantly the onset of this depression-like phenotype is variable<sup>42,43</sup>, potentially allowing





**Fig. 3 | Testing substantia nigra cell type enrichment of genes prioritised by the DNA methylation network associated with depression.** **A** Expression Weighted Cell Enrichment analysis of DepressionSN module genes within the snRNAseq dataset from Kamath et al.<sup>34</sup>. Sub-cell type annotations along the X-axis are grouped by their broader cell class. Benjamini-Hochberg (BH)-corrected significant enrichments ( $q < 0.05$ ) are annotated with asterisks. Standard deviations from the mean value is displayed along the Y-axis. Standard deviation from the mean indicates the number of standard deviations from the

mean level of expression of genes in the DepressionSN module, relative to the bootstrapped mean for that cell type. **B** Conditional expression weighted cell type enrichment analysis of the 15 neuronal subtypes identified. Heatmap displays the standard deviation from the mean for each of the tested cell types as fill, with stars indicating FDR significant  $q$ -value  $< 0.05$ . The cell type controlled for in each analysis is shown along the x-axis and the enrichment of the tested cell types is shown along the y-axis. Plots are subset by broad neuronal classification (i.e. dopaminergic, inhibitory and excitatory neurons).



**Fig. 4 | Polygenic Risk Score analysis for depression shows evidence of phenotypic association refinement of association with depression symptoms in PD.** A Line plot of  $p$ -value thresholding tests for PRS generation in the Base GWAS and refined scores based on variants proximal to the loci annotated to the DepressionSN module. R2 values calculated based on a linear model of binary depressive state presentation

over all available records. FDR correction applied to all 18 tests. B, C Kaplan Meier curve analysis for conversion to depression over time for PD samples for DepressionSN refined PRS (B) and Base GWAS (C). Both PRS methods calculated at a  $p$ -value threshold of 0.05. Individual lines show strata of polygenic risk based on percentile of PRS.

for a controlled model for assessment of specific cell type contributions to the variable onset of PD depression, as mediated by midbrain dopaminergic degeneration.

We find evidence of a refinement of depression genetic risk in the context of PD based on genomic regions prioritized by the DepressionSN module. Compared to a base PRS, the epigenetic refined PRS based on the DepressionSN module showed evidence of better association with lifetime depression occurrence for samples present within the PPMI cohort. Although both the base and refined PRSs showed evidence of significant prediction of longitudinal depression conversion within PD, a clear separation at the highest risk strata of the refined PRS was evident when compared to the base. This is to be expected, given the complex nature of how depression may develop within PD. The refinement of those at the highest genetic risk of depression using the refined PRS may be explained as an etiology of depression that is influenced directly by PD pathological change, while a number of cases that develop depression due to a separate process or as a response to the disease burden is being captured by the base PRS. One potential caveat to our polygenic findings is that the criteria for variant inclusion around our methylation signature does not assure direct genetic causality. We included a one megabase window around each site, based on the standard threshold set for local “cis” methylation quantitative trait loci<sup>44</sup>. Further work, looking at additional refined PRSs based on epigenetic association and refined by evidence of quantitative trait loci relationship

may show efficacy in explained symptom and disease heterogeneity within PD.

As is a common issue with DNA methylation studies, in particular in bulk tissue, the causality of any changes detected is unclear, in particular in a disease process where cell type proportion changes are implicit. Although we have controlled for inferred cell type proportions, we cannot exclude the fact that the perturbed DNA methylation network we observe in the SN may be a downstream consequence of broad neurodegeneration in this brain region. Furthermore, it is premature to conclude whether differential DNA methylation of genes present within this particular network lead to altered expression, an assumption relied-upon for the findings of the snRNAseq enrichment analysis. Further work, in appropriate powered cohorts to look at a depression trait within PD and testing gene expression changes in the SN is required to validate this. This study represents a foundational investigation into the underlying epigenomic changes associated with multiple symptom changes in PD and provides a basis for replication to confirm our findings.

A caveat to our findings is in the nature of the phenotyping data present and the clinical binary subsetting used in our trait annotations for discovery analysis. Although care has been taken to annotate these records ante-mortem, we are limited in our ability to resolve clinical traits and inaccuracy may be present within this labeling criteria. In particular, we do not have capacity to resolve timing of symptoms for certain individuals and are limited to binary presentation over lifetime. This may have had a

detrimental impact on identifying significant findings for specific outcomes tested in this study. However, we provide secondary validation evidence given our use of genetic instruments to test for association of the prioritized depression signature in PPMI, using the more robust Geriatric Depression Scale as an outcome.

To conclude, we find evidence of regional epigenetic changes in relation to the development of secondary symptoms in PD, investigating multiple common secondary symptom traits in PD across three relevant brain regions and exploring the contribution of DNA methylation (summarized into inter correlated networks) with trait presentation. We find brain region-specific correlations between these networks and trait presentation, specifically in association with depression in the SN. Furthermore, we find that expression of genes within this network are specifically enriched for expression within relevant neuronal subtypes, prioritizing neuronal changes in the SN and cell types with potential contribution to the onset of PD depression. Finally, we show that genetic variants localized around the regions annotated to the depression epigenetic signature, as summarized in PRS's, show evidence of association to longitudinal depression development in PD.

## Methods

### Parkinson's Sample Summary

PD samples profiled in this study have been summarized previously<sup>45</sup>. Tissue for 134 unique individuals was sourced from the Parkinson's Disease UK Brain Bank, covering the Substantia Nigra (SN), Caudate Nucleus (CN) and Frontal Cortex (FC). Ethical approval for these samples was provided under Research Ethics Committee (ID 14/WM/0129). All donors provided informed consent for the use of their tissue in research. Samples were excluded for having atypical parkinsonism noted on pathology reports, or early onset of disease (age of onset < 40). For our analysis, case notes were assessed for lifetime prevalence of secondary symptoms of depression, anxiety, aggression, dementia, hallucinations and sleep disturbance. As an example, for the annotation of hallucinations, evidence of the following psychiatric sections of clinical notes were used to evidence presence in three separate cases:

1. "deterioration, cognitive decline with visual hallucinations-animals"
2. "Hallucinations (visual, auditory, tactile?)"
3. "Hallucinations; Confusion"

Whereas three examples with evidence of absence of hallucinations had the following psychiatric sections of the clinical notes:

1. "Nightmares - vivid dreams; Anxiety; Poor memory (late); No dementia, no hallucinations"
2. "Somnolence & lethargy; impaired memory; dementia"
3. "Poor concentration (losing train of thought)"

Where relevant, acute symptoms were not included (e.g. situational short term depression). For sensitivity analysis of premorbid depression, premorbid depression was evidenced either by explicit annotation in the clinical notes (e.g. "Anxiety and depression prior to onset of motor symptoms") or from temporal staging of dated symptom entries. Years of disease, as defined by the number of years between diagnosis and death was also annotated as a separate outcome.

### DNA methylation profiling

Genome wide DNA methylation was profiled using the Illumina 450 K methylation array which interrogates ~450,000 methylation sites across the genome and has been described previously<sup>45</sup>. Data underwent stringent quality control and normalization as previously described using functions available in the *watermelon*<sup>46</sup> R package (version 1.26). Samples with low median methylated or unmethylated intensities ( $n=0$ ) and with low bisulfite conversion percentages as determined using the *bscon* () function ( $n=2$ ) were removed as part of quality control (QC). Using a principal component (PC) based method, samples were tested for overlap of reported and predicted biological sex and removed if discordant ( $n=2$ ). Using single

nucleotide polymorphism (SNP) probes included on the array, samples were checked for expected genetic relatedness for replicates across multiple brain regions ( $n=12$  removed for discordant expected relatedness). Using the *pfilter*() function samples were excluded if >1% of probes showed a detection value > 0.05 ( $n=0$ ) and probes were excluded if showing >1% of samples with detection value > 0.05 or beadcount < 3 in 5% of samples ( $n=2411$  probes). Samples were tested for outliers using the *outlyx*() function and visually assessed using PC analysis. As a subtle separation of data points on the PC analysis could be seen corresponding to the different brain regions, we normalized each brain region separately. Quantile normalization was conducted using the *dasen*() function with default settings. Normalization violence was assessed using the *qual*() function to determine samples with high degrees of difference between raw and normalized beta values, with no outlying samples apparent.

### Weighted Gene Correlation Network Analysis

**Data processing and module detection.** Due to the high number of CpG sites tested on the Illumina 450 K array and the low groupwise sample size available for this sample set, we aimed to reduce the multiple testing burden for association using Weighted Gene Correlation Network Analysis (WGCNA, version 1.69)<sup>47</sup>. WGCNA determines correlation networks present within a given dataset to identify distinct clusters of highly correlated data-points which may hold functional relevance based on its significant pairwise relationship to other datapoints. Datasets were first filtered for variable probes across each brain region separately as determined by median absolute deviation (MAD) for any individual probe > median MAD for the entire dataset. To ensure consistent probes were being fed into each analysis per brain region, all probes passing this threshold in every brain region were included for further analysis, resulting in a set of 243,783 probes. Following this, all brain regions were processed separately. To reduce the effect of unwanted technical and biological variance, multiple regression was used to regress out these effects from the dataset. Each CpG site was regressed against age, sex, technical batch, NeuN+ predicted cell proportion (estimated using the *estimateCellCounts*() function in *watermelon*<sup>46</sup>) and post mortem interval (PMI). The residuals from this regression were extracted and added to the intercept to give a methylation value controlling for the applied covariates and scaled similarly to the raw value. Residual corrected methylation values were then clustered by Euclidean distance and the first four PCs were visually assessed to test for outlying samples. From this analysis two samples were removed from the SN, one from the FC and one from the CN. Co-methylation network and module detection were determined in a block-wise method and set as unsigned, so to weight correlation between probes irrespective of the direction of correlation. As recommended in the WGCNA protocol, soft thresholding was applied, which raises the power of each correlation to a particular value with the aim to reduce noise within the dataset. A scale free topology graph was constructed for powers ranging from 1-20 in stepwise increments and assessed for balance between scale free topology and connectivity. From this, a value of eight was selected for the SN, 12 for the FC and nine for the CN. Finally, modules were identified using the *blockwiseModules*() function (unsigned network, min size = 100, max size = 10000).

**Module Trait Association Analysis.** Identified modules ranged in size and similarity and were labeled based on an arbitrary color value determined by the WGCNA process. Modules were additionally filtered at this point to remove any remaining modules retaining any significant ( $P < 0.05$ ) confounding trait association. For association testing, CpGs present within each module were aggregated into individual values representative of a weighted average of methylation within that module. These values, termed module eigengenes (MEs) are calculated using the eigenvalues from the first PC for all methylation values in that module with one module eigengene value determined for each individual case in the dataset per module. These module eigengenes were tested for association with phenotypic traits, using Pearson's correlation for continuous

traits and Spearman's correlation for binary traits. For multiple testing correction, Bonferroni correction was applied as 0.05 divided by the number of modules present in that particular brain region.

**Module membership analysis.** For modules that showed a significant association with any of the outcomes tested, individual relevant probes were assessed based on module membership (MM) and probe significance (PS). MM is calculated using Pearson's correlation between an individual probe and the ME of the module it is assigned and is thus representative of that individual probe's connectivity to the rest of the module. PS was determined using correlation analysis between individual methylation values and the trait found to be significantly associated with that module's eigengene in the same method as for overall ME association. MM was tested against  $-\log_{10}$  transformed PS using Pearson's correlation.

**Ontological enrichment analysis.** For ontological enrichment analysis, annotated gene-symbols from the Illumina manifest file were extracted for the corresponding modules. We used a background of all annotated gene-symbols for all 243,783 probes fed into the analysis. Modules were tested for ontological terms for biological pathways enriched in CpGs present within each module using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis using the *missMethyl* package<sup>33</sup> (version 1.30). As similar ontology terms were observed from this output due to overlapping gene sets, modules were merged based on semantic similarity using the web tool REVIGO (<http://revigo.irb.hr/>)<sup>48</sup>. Resnik's measure was used to compute the similarity of terms and a medium between terms similarity of 0.7 was allowed.

#### Single cell data processing and cell enrichment analysis

To determine the sub cellular localization of annotated genes determined from the WGCNA analysis, human SN single nuclei RNA sequencing (snRNAseq) data generated using the 10× Genomics Platform was sourced from three separate datasets. Kamath et al.<sup>34</sup> included data produced from NR4A2 positive and negative sorted nuclei populations from midbrain of 18 PD, Lewy body dementia and control donors ( $n = 387,483$  nuclei). Agarwal et al.<sup>49</sup> included data from the SN generated from five control samples ( $n = 5,943$  nuclei) and Smajic et al.<sup>50</sup> included data from six PD and five control samples ( $n = 41,435$  nuclei). Filtered human single nuclei barcodes, gene features and expression matrix along with processed UMAPs and metadata were downloaded and processed. Data was loaded using the *Read10X()* function in the *Seurat* R package (version 4.3.0). Loaded data was converted to a summarized experiment object using the *SummarizedExperiment()* function in the R package of the same name. *colData* for each profiled cell was assigned from the corresponding cell types based on annotations from the metadata files as determined from the original publications and annotated at two levels of granularity. Data was then processed for Expression Weighted Cell Type Enrichment analysis using functions within the *EWCE*<sup>51</sup> package (<https://github.com/NathanSkene/EWCE>) (version 1.4.0). First, genes with no overall expression or no significant differential expression between cell types at FDR adjusted q-value threshold  $< 1e-05$  were removed using the *drop\_uninformative\_genes()* function with the *Limma* setting. A normalized mean expression and specificity cell type dataset was calculated using the *generate\_celltype\_data()* function. Data quality was assessed at this point for potential artefacts by visual assessment of known marker gene expression in known cell types using the *plot\_ctd()* function. Genes annotated to methylated loci in each module determined from WGCNA were tested separately for cell type enrichment using the *bootstrap\_enrichment\_test()* function. Tests were conducted over 100,000 repetitions and tested for cell type and sub-cell type enrichment separately (Supplementary Fig. 7). Significant cell type enrichment was determined by Benjamini-Hochberg (BH) corrected q-values  $< 0.05$ . For plotting, similar modules were determined based on Euclidean distance of a binary significance module by cell type matrix.

To refine the cell-type signal observed in the Kamath et al. dataset, conditional enrichment analyses were conducted, controlling for significant cell types observed in the initial analysis. Each significant cell type was assigned as the controlled cell type iteratively using the *controlledCT* parameter in the *bootstrap\_enrichment\_test()* function. The corresponding controlled analyses were interpreted using standard deviations from the mean and BH-corrected q-values as in the initial analysis.

#### Parkinson's progression marker's initiative data processing and refined PRS analysis

Data from the PPMI cohort was sourced from the PPMI database (<https://ida.loni.usc.edu/>) as has been described previously<sup>52</sup>. Participating PPMI sites all received approval from an ethical standards committee before study initiation and written informed consent was obtained for all individuals participating in the study. The study was registered at [clinicaltrials.gov](http://clinicaltrials.gov) (NCT01141023). Neuro X SNP array genotyping data<sup>53</sup> for 619 whole-blood DNA samples, covering 292,313 variants was quality checked and imputed as previously described<sup>52</sup>. Following this, 2,288,759 variants were included from 582 samples covering 365 de-novo PD samples, 161 control samples and 56 samples with clinical PD but without evidence of dopaminergic deficit (SWEDD).

For refined PRS analysis, summary statistic data was sourced from the most recent and largest GWAS of depression to date by Als et al.<sup>36</sup> Summary statistics were pre-processed to remove variants with an imputation information score of  $< 0.9$ . The PRSet implementation within PRSice-2<sup>54</sup> was used to calculate depression PRS's within PPMI. DepressionSN module annotated regions were defined by adding a 1 mega-base extension around each CpG site coordinate within the module. Depression PRSs were calculated for the DepressionSN module signal and the base unrefined GWAS, testing *p*-value thresholds of 1, 0.5, 0.05, 0.005,  $5e^{-4}$ ,  $5e^{-5}$ ,  $5e^{-6}$ ,  $5e^{-7}$  and  $5e^{-8}$ , with default clumping parameters (1000 kb). PRS score R2 were tested for a binary trait of depression within all samples, based on a minimum of one geriatric depression score (GDS)  $\geq 5$  at a single time point within all available observations. Sex and the first three genetic derived principal components were included as covariates in this analysis.

For time to event analysis, longitudinal GDS records for de-novo PD samples were annotated from baseline to year 8. To control for severity of parkinsonian motor symptoms at each visit, Movement Disorder Society Unified PD Rating Scale Part-III (UPDRS-3), as measured in the off-state was merged onto the dataset. To confirm consistent, significant depression, samples required a minimum of three longitudinal records and GDS defined depression had to be present for a minimum of two annotated visits. Cox regression models were fit within the *survival* package (version 3.5). Base and DepressionSN module refined PRSs were defined as predictors of conversion, with sex defined as a fixed covariate and age and UPDRS-3 score as time dependent covariates. For Kaplan Meier curve analysis, PRS scores were subset into quantiles of risk.

#### Data availability

DNA methylation data used in this study was provided for the study by the authors of a previous publication<sup>45</sup>. It is available via figshare <https://doi.org/10.17035/cardiff.27195645.v1>. Data used in the preparation of this article were obtained on 1<sup>st</sup> June 2020 from the Parkinson's Progression Markers Initiative (PPMI) database (<https://www.ppmi-info.org/access-data-specimens/download-data>), RRID:SCR\_006431. This analysis used data openly available from PPMI. For up-to-date information on the study, visit <http://www.ppmi-info.org>.

#### Code availability

All codes are available at [https://github.com/JoshHarveyGit/PD\\_TraitNetworkAnalysis](https://github.com/JoshHarveyGit/PD_TraitNetworkAnalysis).

Received: 19 July 2023; Accepted: 24 January 2025;  
Published online: 02 March 2025



## References

- Ray Dorsey, E. et al. Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **17**, 939–953 (2018).
- Jankovic, J. Parkinson's disease: clinical features and diagnosis. *J. Neurol. Neurosurg. Psychiatry* **79**, 368–376 (2008).
- Jankovic, J. & Tan, E. K. Parkinson's disease: etiopathogenesis and treatment. *J. Neurol. Neurosurg. Psychiatry* **91**, 795–808 (2020).
- Institute, K. et al. Depression in Parkinson disease—epidemiology, mechanisms and management. *Nat. Rev. Neurol.* **8**, 35–47 (2011).
- Broen, M. P. G., Narayan, N. E., Kuijff, M. L., Dissanayaka, N. N. W. & Leentjens, A. F. G. Prevalence of anxiety in Parkinson's disease: A systematic review and meta-analysis. **31**, 1125–1133 (2016).
- Aarsland, D. et al. Range of neuropsychiatric disturbances in patients with Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **67**, 492 (1999).
- Den Brok, M. G. H. E. et al. Apathy in Parkinson's disease: A systematic review and meta-analysis. *Mov. Disord.* **30**, 759–769 (2015).
- Aarsland, D. et al. Parkinson disease-associated cognitive impairment. **7**, 1–21 (2021).
- Jones, S., Torsney, K. M., Scourfield, L., Berryman, K. & Henderson, E. J. Neuropsychiatric symptoms in Parkinson's disease: aetiology, diagnosis and treatment. *BJPsych Adv.* **26**, 333–342 (2020).
- Aarsland, D., Larsen, J. P., Tandberg, E. & Laake, K. Predictors of nursing home placement in Parkinson's disease: A population-based, prospective study. *J. Am. Geriatr. Soc.* **48**, 938–942 (2000).
- Fénelon, G. & Alves, G. Epidemiology of psychosis in Parkinson's disease. *J. Neurol. Sci.* **289**, 12–17 (2010).
- Munhoz, R. P. et al. Demographic and motor features associated with the occurrence of neuropsychiatric and sleep complications of Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **84**, 883–887 (2013).
- Bareeqa, S. B. et al. Prodromal depression and subsequent risk of developing Parkinson's disease: a systematic review with meta-analysis. **12**, 155–164 (2022).
- Gustafsson, H., Nordström, A. & Nordström, P. Depression and subsequent risk of Parkinson disease. *Neurology* **84**, 2422–2429 (2015).
- Schuurman, A. G. et al. Increased risk of Parkinson's disease after depression: a retrospective cohort study. *Neurology* **58**, 1501–1504 (2002).
- Sandor, C. et al. Universal clinical Parkinson's disease axes identify a major influence of neuroinflammation. *Genome Med* **14**, 1–15 (2022).
- Greenland, J. C., Williams-Gray, C. H. & Barker, R. A. The clinical heterogeneity of Parkinson's disease and its therapeutic implications. *Eur. J. Neurosci.* **49**, 328–338 (2019).
- Liu, G. et al. Genome-wide survival study identifies a novel synaptic locus and polygenic score for cognitive progression in Parkinson's disease. *Nat. Genet.* **53**, 787–793 (2021).
- Creese, B. et al. Glucocerebrosidase mutations and neuropsychiatric phenotypes in Parkinson's disease and Lewy body dementias: Review and meta-analyses. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **177**, 232–241 (2018).
- Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016).
- MacBean, L. F., Smith, A. R. & Lunnon, K. Exploring Beyond the DNA Sequence: A Review of Epigenomic Studies of DNA and Histone Modifications in Dementia. *Curr. Genet. Med. Rep.* **8**, 79–92 (2020).
- Roubroeks, J. A. Y. et al. An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. *Neurobiol. Aging* **95**, 26–45 (2020).
- Lunnon, K. et al. Methylopic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* **17**, 1164–1170 (2014).
- Smith, R. G. et al. A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. *Nat. Commun.* **12**, 1–13 (2021).
- Shao, X. et al. Dementia with Lewy bodies post-mortem brains reveal differentially methylated CpG sites with biomarker potential. *Commun. Biol.* **5**, 1–11 (2022).
- Chuang, Y.-H. H. et al. Longitudinal Epigenome-Wide Methylation Study of Cognitive Decline and Motor Progression in Parkinson's Disease. *J. Parkinsons. Dis.* **9**, 389–400 (2019).
- Young, J. I. et al. Genome-wide brain DNA methylation analysis suggests epigenetic reprogramming in Parkinson disease. *Neurol. Genet.* **5**, e342 (2019).
- Pihlström, L. et al. Epigenome-wide association study of human frontal cortex identifies differential methylation in Lewy body pathology. *Nat. Commun.* **13**, 1–10 (2022).
- Pishva, E. et al. Psychosis-associated DNA methylomic variation in Alzheimer's disease cortex. *Neurobiol. Aging* **89**, 83–88 (2020).
- Kouhsar, M. et al. A brain DNA co-methylation network analysis of psychosis in Alzheimer's disease. *Alzheimers Dement* **21**, e14501 (2025).
- Braak, H., Alafuzoff, I., Arzberger, T., Kretschmar, H. & Tredici, K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* **112**, 389–404 (2006).
- Braak, H., Ghebremedhin, E., Rüb, U., Bratzke, H. & Del Tredici, K. Stages in the development of Parkinson's disease-related pathology. *Cell Tissue Res* **318**, 121–134 (2004).
- Wang, Z., Wu, X. L. & Wang, Y. A framework for analyzing DNA methylation data from Illumina Infinium HumanMethylation450 BeadChip. *BMC Bioinforma.* **19**, 15–22 (2018).
- Kamath, T. et al. Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease. *Nat. Neurosci.* **25**, 588–595 (2022).
- Yang, H.-S. et al. Cell-type-specific Alzheimer's disease polygenic risk scores are associated with distinct disease processes in Alzheimer's disease. *Nat. Commun.* **14**, 7659 (2023).
- Als, T. D. et al. Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nat. Med.* **29**, 1832–1844 (2023).
- Jellinger, K. A. The pathobiological basis of depression in Parkinson disease: challenges and outlooks. *J. Neural Transm.* **129**, 1397–1418 (2022).
- Ng, A., Chander, R. J., Tan, L. C. S. & Kandiah, N. Influence of depression in mild Parkinson's disease on longitudinal motor and cognitive function. *Park. Relat. Disord.* **21**, 1056–1060 (2015).
- Patterson, L., Rushton, S. P., Attems, J., Thomas, A. J. & Morris, C. M. Degeneration of dopaminergic circuitry influences depressive symptoms in Lewy body disorders. *Brain Pathol.* **29**, 544–557 (2019).
- Frisina, P. G., Haroutunian, V. & Libow, L. S. The neuropathological basis for depression in Parkinson's disease. *Parkinsonism Relat. Disord.* **15**, 144 (2009).
- Mou, Y. K. et al. Application of neurotoxin-induced animal models in the study of parkinson's disease-related depression: profile and proposal. *Front. Aging Neurosci.* **14**, 504 (2022).
- Zhang, X. et al. Decrease of gene expression of astrocytic 5-HT2B receptors parallels development of depressive phenotype in a mouse model of Parkinson's disease. *Front. Cell. Neurosci.* **9**, (2015).
- Tang, J. et al. Crocin reverses depression-like behavior in Parkinson disease mice via VTA-mPFC pathway. *Mol. Neurobiol.* **57**, 3158–3170 (2020).
- McClay, J. L. et al. High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome Biol.* **16**, (2015).

45. Kia, D. A. et al. Identification of Candidate Parkinson Disease Genes by Integrating Genome-Wide Association Study, Expression, and Epigenetic Data Sets. *JAMA Neurol.* **78**, 1 (2021).
46. Pidsley, R. et al. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 1–10 (2013).
47. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinforma.* **9**, 1–13 (2008).
48. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One* **6**, e21800 (2011).
49. Agarwal, D. et al. A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. *Nat. Commun.* **11**, 1–11 (2020).
50. Smajic, S. et al. Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state. *Brain* **145**, 964–978 (2022).
51. Skene, N. G. & Grant, S. G. N. Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front. Neurosci.* **10**, 16 (2016).
52. Harvey, J. et al. Machine learning-based prediction of cognitive outcomes in de novo Parkinson's disease. *NPJ Park. Dis.* **8**, (2022).
53. Nalls, M. A. et al. Baseline genetic associations in the Parkinson's Progression Markers Initiative (PPMI). *Mov. Disord.* **31**, 79–85 (2016).
54. Choi, S. W. et al. PRSet: Pathway-based polygenic risk score analyses and software. *PLOS Genet* **19**, e1010624 (2023).

## Acknowledgements

This work was funded by research grants from the Medical Research Council (MRC (MR/S011625/1), the National Institute of Aging (NIA) of the National Institutes of Health (NIH) (R01AG067015), BRACE and the Charles Wolfson Charitable Trust to KL and from Parkinson's UK (Project Grants G-1309 and G-1502) to NW. This study was supported by the National Institute for Health and Care Research Exeter Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. We thank all participants and teams who contributed data to PPMI. 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 Therapeutics, Allergan, Amathus Therapeutics, Aligning Science Across Parkinson's (ASAP), Avid Radiopharmaceuticals, Bial Biotech, Biogen, BioLegend, Bristol Myers Squibb, Calico Life Sciences LLC, Celgene Corporation, DaCapo Brainscience, Denali Therapeutics, The Edmond J. Safra Foundation, Eli Lilly and Company, GE Healthcare, GlaxoSmithKline, Golub Capital, Handl Therapeutics, Insitro, Janssen Pharmaceuticals, Lundbeck, Merck & Co., Meso Scale Diagnostics LLC, Neurocrine Biosciences, Pfizer, Piramal

Imaging, Prevail Therapeutics, F. Hoffmann-La Roche and its affiliated company Genentech, Sanofi Genzyme, Servier, Takeda Pharmaceutical Company, Teva Neuroscience, UCB, Vanqua Bio, Verily Life Sciences, Voyager Therapeutics and Yumanity Therapeutics.

## Author contributions

L.H. and K.B. conducted laboratory experiments generating the DNA methylation data. J.H. undertook data analysis and bioinformatics, with support from RGS, IC and EP. A.R.S., L.S.W. and B.C. collated and interpreted the clinical data for the analysis. N.W. and K.L. conceived the project. E.P. and K.L. supervised the project. J.H., E.P., and K.L. drafted the manuscript. All authors read and approved the final submission.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at

<https://doi.org/10.1038/s41531-025-00877-5>.

**Correspondence** and requests for materials should be addressed to Katie Lunnon.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025