

**TITLE:** Plasma acellular transcriptome contains Parkinson's disease signatures that can inform clinical diagnosis

**AUTHORS**

Aleksandra Beric, Department of Psychiatry, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA. NeuroGenomics and Informatics Center, Washington University in Saint Louis School of Medicine, St. Louis, MO 63108, USA. ORCID ID: 0000-0002-9351-585X

Alejandro Cisterna-García, Departamento de Ingeniería de la Información y las Comunicaciones; Universidad de Murcia, Murcia, Spain.

Charissa Martin, Department of Psychiatry, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA.

Ravindra Kumar, Department of Psychiatry, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA. NeuroGenomics and Informatics Center, Washington University in Saint Louis School of Medicine, St. Louis, MO 63108, USA.

Isabel Alfradique-Dunham, Department of Neurology, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA.

Kevin Boyer, Department of Genetics, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA.

Ibrahim Olabayode Saliu, Department of Genetics, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA.

Shinnosuke Yamada, Department of Genetics, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA.

Jessie Sanford, Department of Psychiatry, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA. NeuroGenomics and Informatics Center, Washington University in Saint Louis School of Medicine, St. Louis, MO 63108, USA.

Daniel Western, Department of Psychiatry, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA. NeuroGenomics and Informatics Center, Washington University in Saint Louis School of Medicine, St. Louis, MO 63108, USA. ORCID: 0000-0003-3725-9553

Menghan Liu, Department of Psychiatry, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA. NeuroGenomics and Informatics Center, Washington University in Saint Louis School of Medicine, St. Louis, MO 63108, USA.

Ignacio Alvarez, Department of Pathology, Germans Trias I Pujol Research Institute, Badalona, Spain.

Joel S. Perlmutter, Neurology, Radiology, Neuroscience, Physical Therapy, Occupational Therapy, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA.

Scott A. Norris, Departments of Neurology and Radiology, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA.

Pau Pastor, Department of Neurology, Hospital Germans Trias i Pujol, and Genomics and Transcriptomics of Synucleinopathies (GTS), The Germans Trias i Pujol Research Institute (IGTP), Badalona, Barcelona, Spain.

Guoyan Zhao, Department of Genetics, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA. Department of Neurology, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA. ORCID ID: 0000-0001-5615-6774

Juan Botia, Departamento de Ingeniería de la Información y las Comunicaciones; Universidad de Murcia, Murcia, Spain. Department of Neurodegenerative Diseases, Institute of Neurology, University College London, London UK.

Laura Ibanez, Department of Psychiatry, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA. NeuroGenomics and Informatics Center, Washington University in Saint Louis School of Medicine, St. Louis, MO 63108, USA. Department of Neurology, Washington University in Saint Louis School of Medicine, St. Louis, MO 63110, USA. ORCID ID: 0000-0003-2381-7059

**Corresponding Author:**

Laura Ibanez, PhD  
Washington University in Saint Louis School of Medicine  
4444 Forest Park Ave  
Campus Box 8134  
Saint Louis, MO 63110  
Tel: 314-273-3641  
ibanezl@wustl.edu

## ABSTRACT

We aimed to identify plasma cell-free transcripts (cfRNA) associated with Parkinson's disease (PD) that also have a high predictive value to differentiate PD from healthy controls. Leveraging two independent populations from two different movement disorder centers we identified 2,188 differentially expressed cfRNAs after meta-analysis. The identified transcripts were enriched in PD relevant pathways, such as PD ( $p=9.26 \times 10^{-4}$ ), ubiquitin-mediated proteolysis ( $p=7.41 \times 10^{-5}$ ) and endocytosis ( $p=4.21 \times 10^{-6}$ ). Utilizing in-house and publicly available brain, whole blood, and acellular plasma transcriptomic and proteomic PD datasets, we found significant overlap across dysregulated biological species in the different tissues and the different biological layers. We developed three predictive models containing increasing number of transcripts that can distinguish PD from healthy control with an area under the ROC Curve (AUC)  $\geq 0.85$ . Finally, we showed that several of the predictive transcripts significantly correlate with symptom severity measured by UPDRS-III. Overall, we have demonstrated that cfRNA contains pathological signatures and has the potential to be utilized as biomarker to aid in PD diagnostics and monitoring.

## INTRODUCTION

Parkinson's disease (PD) is a slowly progressing, complex neurodegenerative disorder, with higher prevalence in males.<sup>1,2</sup> It is one of the most common neurodegenerative diseases (NDDs), second only to Alzheimer's disease (AD).<sup>3,4</sup> As with other NDDs, the greatest risk factor for PD development is age, with incidence peaking after 80 years of age, with contributions from environmental and genetic factors.<sup>2</sup> PD is characterized pathologically by formation of Lewy bodies (LBs) and early death of dopaminergic neurons, resulting in a typical clinical presentation including bradykinesia, rest tremor and rigidity. Other clinical hallmarks of PD include a number of non-motor symptoms, such as sleep, gastrointestinal and olfactory disorders, which may precede motor disorders by over a decade.<sup>2,4,5</sup> At the molecular level, LBs are primarily comprised of misfolded  $\alpha$ -synuclein, which can spread between the cells, serving as a template for further  $\alpha$ -synuclein misfolding.<sup>5</sup>

While PD diagnoses largely depend on patient history and physical examination, no currently available tests enable definitive diagnosis of PD in the early stages.<sup>2,4</sup> Instead, definitive diagnostics presently depends on neuropathological analyses upon death, typically occurring many years after disease onset.<sup>1,2,4</sup> Several imaging methods can aid in confirm nigrostriatal deficits that occur in PD, but are not diagnostic.<sup>6-9</sup> Dopamine transporter single-photon emission computed tomography (DaT SPECT) can detect cell loss in PD patients<sup>4,6</sup>, while positron emission tomography (PET) scan can point to early signs of dopaminergic neuron damage.<sup>4,7,10,11</sup> Conversely, magnetic resonance imaging (MRI) methods provide modest benefit for diagnosis of PD.<sup>4,6,7,10</sup> In addition to imaging, the field has strived to develop PD-specific cerebrospinal fluid (CSF) biomarkers, independent of clinical representation of the disease.<sup>12</sup> CSF levels of  $\alpha$ -synuclein have been the focal point of a number of studies.  $\alpha$ -synuclein seed

amplification assays (SAA) have shown a lot of promise, with the ability to differentiate between PD and healthy controls.<sup>13,14</sup> However, results have been variable possibly due to clinical heterogeneity, cross-contamination with blood, or experimental differences, requiring further validation prior to clinical implementation. Lysosomal enzymes and neurofilament light chain emerged as candidates for biomarker panels, though they also require further investigation.<sup>12,15,16</sup> Another unmet need is the differential diagnosis of PD from other NDDs. Though it is a very common neurodegenerative disorder, PD is misdiagnosed in clinical practice with error rates reported to range from 15% to 24%.<sup>4</sup> The prevailing reason for disagreement between clinical and neuropathological diagnoses is the heterogeneity of parkinsonism with non-PD pathologies, including multiple system atrophy (MSA) and progressive supranuclear palsy (PSP).<sup>17</sup> Indeed, even established PD cases are greatly heterogeneous in the age of onset, rate of progression as well as clinical presentation, which led to the establishment of several PD subtypes.<sup>4,6</sup>

While most blood-based biomarker studies focus on measuring levels of various proteins, circulating nucleic acids have found their place in clinical practice. Analyses of cell-free DNA (cfDNA) have revolutionized the field of obstetrics and antenatal testing by allowing the identification of fetal aneuploidies in a sample of mother's blood, thus reducing test-related risk of miscarriage.<sup>18-20</sup> cfDNA has been utilized as a biomarker for cancer<sup>20-22</sup>, metabolic disorders<sup>20</sup> and a way to assess the health of donor organs in a recipient's body upon organ transplantation.<sup>20,23</sup> In addition to cfDNA, cell-free RNA (cfRNA) can also be captured from plasma and provides a temporal snapshot of cellular processes throughout the body, as it is released from cells as part of normal cell death.<sup>24</sup> Numerous studies are investigating the potential of using cfRNAs as biomarkers for prenatal testing<sup>19</sup>, cancer<sup>21,25-28</sup> and AD.<sup>29-33</sup> Furthermore, a recent study provided evidence of circulating micro RNAs being involved in the regulation of PD-associated genes<sup>34</sup>, adding support to the utility of non-protein biomarkers in the PD field.

In this study, we used plasma cfRNA from two independent populations of PD participants to capture transcriptional changes caused by PD pathology. We biologically contextualized our findings via pathway analyses and multiomic data integration by accessing whole blood and brain transcriptomic datasets and plasma and CSF proteomic datasets. Then we leveraged those to build a predictive model that could accurately predict PD using a limited number of transcripts, with the potential for translation into clinical practice. We further evaluated the capabilities of the best performing models to discriminate between PD and AD, dementia with Lewy bodies (DLB) and frontotemporal dementia (FTD) to ensure that captured changes were specific to PD pathobiology.

## METHODS

### *Study design*

We analyzed acellular RNAseq data from two independent movement disorder clinical cohorts, Hospital Universitari Mutua Terrassa (HUMT) in Barcelona, Spain, and Washington University in

Saint Louis School of Medicine (WUSM) in Saint Louis, US. We performed differential expression (DE) analyses in HUMT (206 participants) and WUSM (175 participants) cohorts separately, followed by meta-analysis. Transcripts with Benjamini-Hochberg corrected p-values below 0.05 in the meta-analysis were considered DE. To understand the biological significance of the DE transcripts, we performed pathway analyses and leveraged in-house and publicly available datasets to contextualize the expression of the identified transcripts in whole blood and brain, as well as the accumulation of corresponding proteins in plasma and CSF. To assess the diagnostic capabilities of cfRNAs, we developed several predictive models, focusing on the DE transcripts. We then employed a third independent dataset, consisting of participants with dementia with DLB, AD, and FTD, to test the specificity of the predictive models for PD. AD, DLB, and FTD participants were diagnosed according to the clinical criteria contained in the Uniform Data Set (UDS), the standard set of clinical data collected in all participants enrolled in any of the 37-federally funded ADRCs. Finally, utilizing the information available about participants' motor symptom severity, measured by Unified Parkinson's Disease Rating Scale Part III (UPDRS-III), and cognitive status measured by The Montreal Cognitive Assessment (MoCA), we assessed the clinical relevance of transcripts included in the predictive models. Research in this study was conducted in accordance with recommended protocols. Written informed consent was obtained from all participants or their family members. The Washington University in Saint Louis Institutional Review Board approved the study (IRB ID 201701124 and 202004010).

### *Study Participants*

We obtained plasma samples from two independent cohorts, HUMT and WUSM. HUMT cohort included a total of 206 plasma samples (87 PD participants and 119 healthy controls), while the WUSM cohort included 175 samples (94 PD participants and 81 healthy controls). Given the different geographical location and standards of care, the two cohorts show some differences (Table 1). They are comparable in proportion of female participants (HUMT 47.57%, WUSM 48.57%;  $p=0.93$ ). The HUMT population shows a lower mean age ( $68.26\pm 8.39$ ) compared to the WUSM population ( $72.97\pm 6.78$ ;  $p=3.30\times 10^{-9}$ ). Differences are also observed in motor symptom severity as measured by the UPDRS-III scale. Participants in the WUSM population displayed greater UPDRS-III ( $26.03\pm 9.10$ ), compared to HUMT participants ( $18.69\pm 8.99$ ;  $p=6.39\times 10^{-4}$ ). Finally, dementia was assessed only in the WUSM participants, who presented with mild or no signs of dementia ( $25.32\pm 4.22$ ), as measured by the MoCA scale. Similarly, no therapeutic information was available for the HUMT dataset.

### *RNA Extraction and Sequencing*

Whole blood samples were collected from all participants. Within 20 min of collection, blood samples were centrifuged for 10 min at 1500rpm to obtain plasma and subsequently stored at  $-80^{\circ}\text{C}$ , as previously described.<sup>35</sup> Plasma samples were thawed on ice and centrifuged for 5 min at 2000 rpm prior to RNA extraction to remove any cells present and avoid cellular RNA contamination. Total plasma cfRNA was extracted from 0.5 mL of plasma using the Maxwell RSC

miRNA from plasma or serum kit (Ambion) and ribodepleted (NEBNext rRNA Depletion Kit). After total RNA quantification, libraries were generated using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs) using 1ng of RNA as input. Libraries were cleaned for adapter dimers prior to sequencing. We targeted 40 million 100 base pair single-end reads for each sample using an Illumina NovaSeq 6000.

#### *Data Processing and Quality Control*

FastQC (v0.11.7)<sup>36</sup> was used to evaluate the sequencing quality of each sample. Reads were aligned to the human reference genome GRCh38 using STAR (v2.7.1a).<sup>37</sup> The quality of sequences and alignments was assessed with PICARD (v2.26)<sup>38</sup> and SAMtools<sup>39</sup>, and transcripts quantified using Salmon (v0.11.3).<sup>40</sup> Quality control measures were gathered via MultiQC (v1.9)<sup>41</sup> followed by stringent quality control (QC). Briefly, all genes with less than ten reads in 90% or more of the individuals were removed. Subsequently, transcriptome Principal Component Analysis (PCA) were performed and screened for correlation with technical and methodological variables to detect potential biases. A strong correlation was observed with total reads and coding bases; thus, all samples with less than 10% of coding bases and less than 1,000,000 total reads were removed. Outlier samples identified via transcriptome PCA, defined as samples whose first two principal component values deviated more than three standard deviations from mean values of either of the respective principal components, were also removed.

Despite following the same protocol for data generation, processing, and QC, the two datasets included in the present study (HUMT and WUSM) were sequenced at different timepoints. Each dataset underwent QC separately. ComBat\_seq function from the sva package<sup>42</sup> was used to adjust for technical variation within each dataset. Batch effect correction was followed by PCA and removal of any additional outliers. As previously described, there is RNA degradation associated with plasma long-term storage (up to 20 years).<sup>33</sup> Consequently, we addressed degradation using DESeq2 (v1.22.2)<sup>43</sup> to find transcripts associated with storage time in control participants as previously reported.<sup>33</sup> All transcripts nominally ( $p < 0.05$ ) associated with storage time were removed from the analyses from both HUMT (n=486 transcripts) and WUSM (n=221 transcripts) datasets. Further, due to prevalence of PD therapies and lack of medication data for HUMT population, transcripts associated with PD-medication usage were identified in the WUSM dataset using DESeq2<sup>43</sup> and any nominally significant transcripts ( $p < 0.05$ ) were removed from further analyses from both datasets (n=630 transcripts). Overall, 27832 transcripts passed QC and were included used in subsequent analyses.

#### *Differential Expression Analyses and Pathway Analyses*

Differential expression (DE) analyses were performed using DESeq2<sup>43</sup> in HUMT and WUSM data separately, followed by meta-analysis using metaRNASeq.<sup>44</sup> All analyses were adjusted by sex and age at blood draw. Benjamini-Hochberg correction (FDR) was used to correct for multiple testing, considering meta-analysis FDR p-values lower than 0.05 as significant. No effect size ( $\log_2$  fold change) value cut-off was applied. Pathway enrichment analyses were carried out

using clusterProfiler<sup>45</sup> to functionally characterize the identified transcripts and FDR p-values below 0.05 were regarded as significant.

### *Multi-omic data integration*

To biologically contextualize our findings, we cross-checked the results from the cfRNA DE meta-analysis with: (i) in-house brain bulk RNAseq, (ii) publicly available, whole blood bulk RNAseq<sup>46</sup>, (iii) publicly available plasma and (iv) CSF proteomic data<sup>46</sup>, and (v) in house and (vi) publicly available brain single-cell RNAseq data<sup>47,48</sup> including substantia nigra scRNAseq (Figure 1). Following differential expression or accumulation analyses in each of the RNAseq or proteomic datasets, we identified the overlap between nominally significant findings and the DE plasma acellular transcripts and each accessed dataset. The significance of each overlap was tested using hypergeometric test (phyper function in R), p-values lower than 0.05 considered significant.

Additionally, to investigate whether any of the DE transcripts mapped to PD associated loci identified by genome wide association analyses (GWAS), we leveraged the information available in the PD GWAS locus browser and in the latest multi-ancestry genome-wide association study.<sup>49,50</sup> We browsed the PD GWAS locus browser and recorded the scores that rank the genes based on the amount of supporting evidence compiled in the PD GWAS locus browser.<sup>50</sup> Next, to compare our results to the multi-ancestry GWAS, we converted locations of all lead SNPs from hg19 version of the human genome to the respective locations in hg38 version of the human genome and then found coordinates 500kb upstream and downstream from each SNP.<sup>49</sup> All regions that overlapped after the coordinate conversion were collapsed into 69 non-overlapping genomic regions. Subsequently, collapsed PD-associated genomic regions were overlapped with genomic start and end coordinates of the identified DE genes using bedtools intersect from BEDTools tool suite.<sup>51</sup>

### *Predictive Models Construction and Evaluation*

To build and evaluate predictive models we used the previously published in-house pipeline.<sup>33</sup> In brief, glmnet (v4.1.6)<sup>52</sup> was leveraged to produce a suitable classifier to identify PD cases based on plasma acellular gene expression. HUMT was used as the training and WUSM as the testing dataset. After ComBat\_seq<sup>42</sup> regression, transcript counts were further scaled between the two datasets by computing the z-score using the mean and standard deviation. With the transcripts that were significantly DE in the meta-analysis after multiple test correction (FDR<0.05), we calculated Kullback-Leibler divergence (KLD) between the training (HUMT) and the testing (WUSM) dataset for each transcript using R package entropy v1.3.1.<sup>53</sup> A hundred L2 regularization linear models were trained with increasing number of transcripts, ranging in KLD value from 0.01 to 1 in increments of 0.01. The Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) value was computed for all models in the training dataset.

### *Specificity Analyses*

Performance of the best predictive models was evaluated in 44 AD, 16 FTD and 17 DLB cases from a publicly available dataset from the Knight-ADRC<sup>33,54</sup>, as well as twelve DLB participants available in the HUMT dataset. Knight-ADRC data has been generated and processed as described above for the HUMT and WUSM datasets. Transcript counts were scaled by computing the z-score using the mean and standard deviation. Then, risk score for each individual was calculated using the previously defined predictive models. Scores higher than 0.50 were considered cases. ROC curve was computed by comparing the predicted to the actual disease status for each compared group. We assessed the ability of the predictive models to differentiate between AD, DLB or FTD and healthy controls, as well as between AD, DLB or FTD and PD. Additionally, we evaluated if the addition of the *APOE* genotype to the cfRNA predictor improved the model performance in differentiating between PD and AD, as *APOE* is a crucial genetic risk factor for AD. *APOE* genotype data was available for AD and WUSM PD participants. To discern whether the effect of *APOE* was captured by the predictor, we included the *APOE* genotype in the model coded by two variables representing the number of  $\epsilon 2$  alleles and  $\epsilon 4$  alleles.

#### *Evaluation of predictive models' clinical relevance*

To explore the relationship of selected transcripts with PD clinical manifestations, we calculated Spearman correlations between normalized and age and sex adjusted transcript counts with UPDRS-III or MoCA scores. We did it for those PD participants with available data. UPDRS-III information was available for 76 participants from HUMT and 27 participants in the WUSM population. To maximize our sample size and statistical power for this exploratory analysis, the two populations were combined for correlation with UPDRS-III. MoCA scores were not available for the HUMT population, thus correlation to MoCA analysis was carried out only in the WUSM population (n=28). Correlations were considered significant if p-value was lower than 0.05.

## **RESULTS**

### Acellular transcriptomic patterns translate to changes of plasma protein abundances

We leveraged two independent movement disorder clinics (Hospital Universitari MutuaTerrassa (HUMT), and Washington University School of Medicine (WUSM)) with a total of 181 PD participants ( $n_{\text{HUMT}}=87$ ;  $n_{\text{WUSM}}=94$ ) and 200 healthy control participants ( $n_{\text{HUMT}}=119$ ;  $n_{\text{WUSM}}=81$ ) (Figure 1, Table 1). All PD participants had a clinical diagnosis of PD at the time of sample collection.<sup>55</sup> After stringent quality control (QC), we performed differential expression (DE) analyses comparing PD to healthy control participants using DESeq2<sup>43</sup>, followed by integration of HUMT and WUSM results through meta-analysis. We performed meta-analyses on 6,496 transcripts which had same direction of effect in both HUMT and WUSM populations, and identified 2,188 DE transcripts, 1,101 of which were upregulated and 1,087 downregulated (Supplementary Figure 1, Supplementary Table 1). To evaluate the biological relevance of the DE transcripts, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. We found significant enrichment in PD ( $p=9.26 \times 10^{-4}$ ), endocytosis



( $p=4.21 \times 10^{-6}$ ) and ubiquitin mediated proteolysis ( $p=7.41 \times 10^{-5}$ ), along with other nervous system diseases such as Huntington’s disease (HD;  $p=6.35 \times 10^{-4}$ ) and amyotrophic lateral sclerosis (ALS;  $p=6.10 \times 10^{-4}$ ; Figure 3A, Supplementary Table 2). Additionally, we performed Gene Ontology (GO) enrichment analyses and found enrichment in the cellular components primary lysosome ( $p=4.19 \times 10^{-3}$ ) and ubiquitin ligase complex ( $p=2.12 \times 10^{-3}$ ), and biological processes such as late endosome to lysosome transport ( $p=8.18 \times 10^{-4}$ ) and exocytosis ( $p=9.48 \times 10^{-6}$ ; Supplementary Table 3).

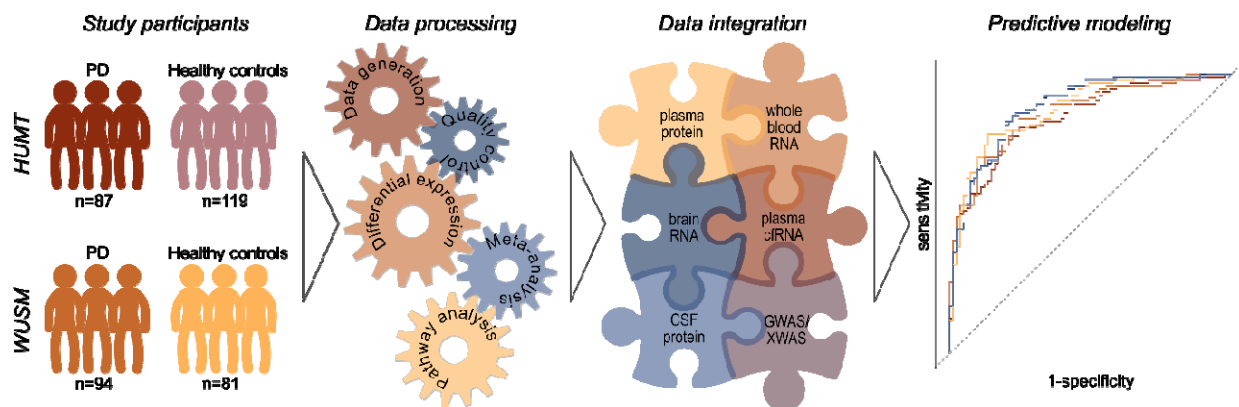
To further replicate our results, we used publicly available whole blood RNAseq data<sup>56</sup> from the Parkinson’s Progression Markers Initiative (PPMI)<sup>46</sup> and Parkinson’s Disease Biomarkers Program (PDBP)<sup>57</sup> and found 424 of the 2,188 DE transcripts identified in plasma were also DE in whole blood<sup>56</sup> (Figure 2A), which represents a significant overlap ( $p=4.65 \times 10^{-60}$ ). These 424 transcripts were enriched in ubiquitin mediated proteolysis ( $p=8.96 \times 10^{-5}$ ; Supplementary Table 2). Additionally, we investigated if the abundances of the proteins encoded by the identified RNAs were also significantly different. We leveraged plasma proteomic data from PPMI generated with Olink. We found a significant overlap ( $p=5.00 \times 10^{-3}$ ) of twelve differentially accumulated proteins out of the 2188 transcripts DE in plasma (Figure 2A). Further, six of the twelve proteins have the same direction of effect as the respective mRNAs from plasma (Supplementary Table 1). Looking back to the pathway analysis, we find that the twelve proteins were present in the immune response pathways. Two of the twelve acellular transcripts that are differentially accumulated in plasma, at both transcript and protein level, are also DE in whole blood (Figure 2A, Supplementary Table 1).

**Table 1.** Summary demographics of the two populations included in the main analyses

	Hospital Universitari Mutua Terrassa (Barcelona, Spain)		Washington University School of Medicine (Saint Louis, US)	
	Healthy Controls	Parkinson’s Disease	Healthy Controls	Parkinson’s Disease
<b>Participants (N)</b>	119	87	81	94
<b>Mean age (IQR)</b>	67.78 (61.00- 74.00)	68.91 (65.00- 74.00)	74.07 (69.00- 78.00)	72.02 (67.00- 77.00)
<b>Female (N, %)</b>	63 (52.94%)	35 (43.21%)	49 (60.49%)	36 (38.29%)
<b>Mean UPDRS-III (IQR)</b>	-	18.70 (14.00- 22.00)	-	26.03 (20.69- 30.00)
<b>Mean MoCA (IQR)</b>	-	-	-	25.32 (24.00- 28.00)

N=Sample size; IQR=Interquartile Range; UPDRS=Unified Parkinson’s Disease Rating Scale; MoCA=Montreal Cognitive Assessment; US=United States

**Figure 1. Study Design Infographic** depicting the study participants from two independent cohorts, summary of the experimental approach including data processing and analyses, biological contextualization of the results via multiomic data integration with available independent datasets, and leverage of results to build predictive models.



### Expression patterns in plasma follow those observed in brain

We compared our findings to an in-house brain RNAseq dataset<sup>54</sup> to test whether the changes we capture in plasma might have their origin in pathological changes taking place in the brain. Out of the 2,188 transcripts, we found 537 transcripts that were DE in both plasma and brain ( $p=4.22 \times 10^{-105}$ ; Figure 2B, Supplementary Table 1), 278 of which were upregulated and 259 downregulated. Of the 537 transcripts, 286 (142 upregulated and 144 downregulated) have the same direction of effect in plasma and brain. Transcripts that were DE in both plasma and brain were enriched in neurodegenerative diseases including PD ( $p=7.86 \times 10^{-7}$ ), HD ( $p=6.27 \times 10^{-8}$ ), and AD ( $p=2.16 \times 10^{-6}$ ), as well as endocytosis ( $p=8.91 \times 10^{-4}$ ) and ubiquitin mediated proteolysis ( $p=2.47 \times 10^{-3}$ ; Supplementary Table 2).

Next, we compared our DE transcripts with differentially expressed genes derived from three independent single-nucleus RNA-seq datasets from individuals with PD and controls: i) subcortical putamen tissue<sup>48</sup>, ii) anterior cingulate cortex<sup>47</sup>, and iii) *in-house* and publicly available<sup>58</sup> substantia nigra. Out of the 2,188 transcripts DE in plasma, 222 were also DE in at least one cell type in the subcortical putamen (Figure 2C, Supplementary Table 4). The greatest overlap between plasma and subcortical putamen was in transcripts expressed by ependymal cells (a type of glial cells), with 162 shared DE transcripts, followed by 41 transcripts in GRIK3-enriched neuronal cells and eleven transcripts in astrocytes (Supplementary Table 4). Transcripts DE in both plasma and subcortical putamen were nominally enriched in HD ( $p=3.51 \times 10^{-3}$ ) and general pathways of neurodegeneration ( $p=0.03$ ; Supplementary Table 2). Next, we found 77 DE transcripts shared between plasma and the anterior cingulate cortex. The greatest overlap between plasma and anterior cingulate cortex was in neuronal cells (47 transcripts), followed by myeloid cells (27 transcripts), while only three transcripts were shared between plasma cfRNA and cortical astrocytes (Supplementary Table 5). These transcripts were enriched in PD ( $p=7.59 \times 10^{-4}$ ), as well as pathways of neurodegeneration ( $p=1.35 \times 10^{-3}$ , Supplementary Table 2). Of the 222 transcripts shared between plasma and subcortical putamen and 77 transcripts shared between plasma and anterior cingulate cortex, 35 were DE in both subcortical putamen and anterior cingulate cortex (Figure 2C, Supplementary Tables 4 and 5).

*In-house* substantia nigra had 411 DE transcripts in common with plasma (Figure 2C, Supplementary Table 6). These transcripts were enriched in PD ( $p=3.35\times 10^{-4}$ ), endocytosis ( $p=4.26\times 10^{-5}$ ) and ubiquitin mediated proteolysis ( $p=6.92\times 10^{-4}$ ; Supplementary Table 2). Similar to subcortical putamen, ependymal cells from the *in-house* substantia nigra dataset showed the greatest overlap with plasma cfRNA transcripts (245 transcripts; Supplementary Table 6). In parallel, 370 of the 2,188 DE cfRNA were DE in the public substantia nigra dataset<sup>58</sup> (Figure 2C, Supplementary Table 7). They were enriched in ubiquitin mediated proteolysis ( $p=6.06\times 10^{-5}$ ), dopaminergic synapse ( $p=8.19\times 10^{-3}$ ) and general pathways of neurodegeneration ( $p=9.10\times 10^{-3}$ ; Supplementary Table 2). Furthermore, 229 plasma DE transcripts were shared with both the *in-house* and the public substantia nigra. Additionally, 72 of those 229 were in the same direction of effect in the two substantia nigra datasets.

Finally, using publicly available CSF proteomic data generated with Somalogic (PPMI)<sup>46</sup>, we tested whether any of the DE transcripts correspond to differentially abundant CSF proteins. We uncovered a significant overlap ( $p=4.90\times 10^{-6}$ ) of 89 acellular plasma transcripts and their respective protein products in CSF (Figure 2B), which were enriched in endocytosis ( $p=8.10\times 10^{-7}$ ) and nominally enriched in HD ( $p=1.83\times 10^{-2}$ ), and general pathways of neurodegeneration ( $p=4.40\times 10^{-2}$ ; Supplementary Table 2). Of the 89 transcripts whose protein products were differentially accumulated in CSF, 24 are also DE in bulk brain RNAseq (Supplementary Table 1).

**Figure 2. Multiomic integration summary.** Integration between differential expression analysis of cfRNA (maroon) with **A.** whole blood RNAseq (orange) and plasma proteomics (light yellow); **B.** brain RNAseq (dark blue) and CSF proteomics (light blue); **C.** single nuclei datasets; and **D.** GWAS data (rosy-brown). The y-axes in panels A, B and D represent  $-\log_{10}$  of differential expression p-value, while x-axes represent genomic coordinates of the respective transcript/gene. Dotted lines symbolize the same transcript/gene or its protein product that are significantly differentially expressed across the multiple *omic* layers displayed in each respective panel, A, B or D.



nominated genes in the PD GWAS locus browser<sup>50</sup> with the DE transcripts identified in our analyses to assess if there is an agreement with the nominated gene. We found varying levels of support for 121 of the 190 autosomal genes, with an average score of  $4.42 \pm 1.88$  (Supplementary Table 8). Genes located on the X chromosome cannot be evaluated due to the unavailability of sex chromosome data in the GWAS browser. We checked if any of the 121 genes were DE in whole blood or brain RNAseq, or if their protein products were differentially accumulated in plasma or CSF. We found 26 of the 121 genes DE in whole blood and 24 of the 121 DE in brain (Supplementary Tables 1 and 8). Similarly, we observed that corresponding protein products of two of the 121 genes (ITGAM and PARP1) were differentially accumulated in plasma, and four (GCH1, HEXIM2, LGALS3 and UFC1) differentially accumulated in CSF (Supplementary Tables 1 and 8).

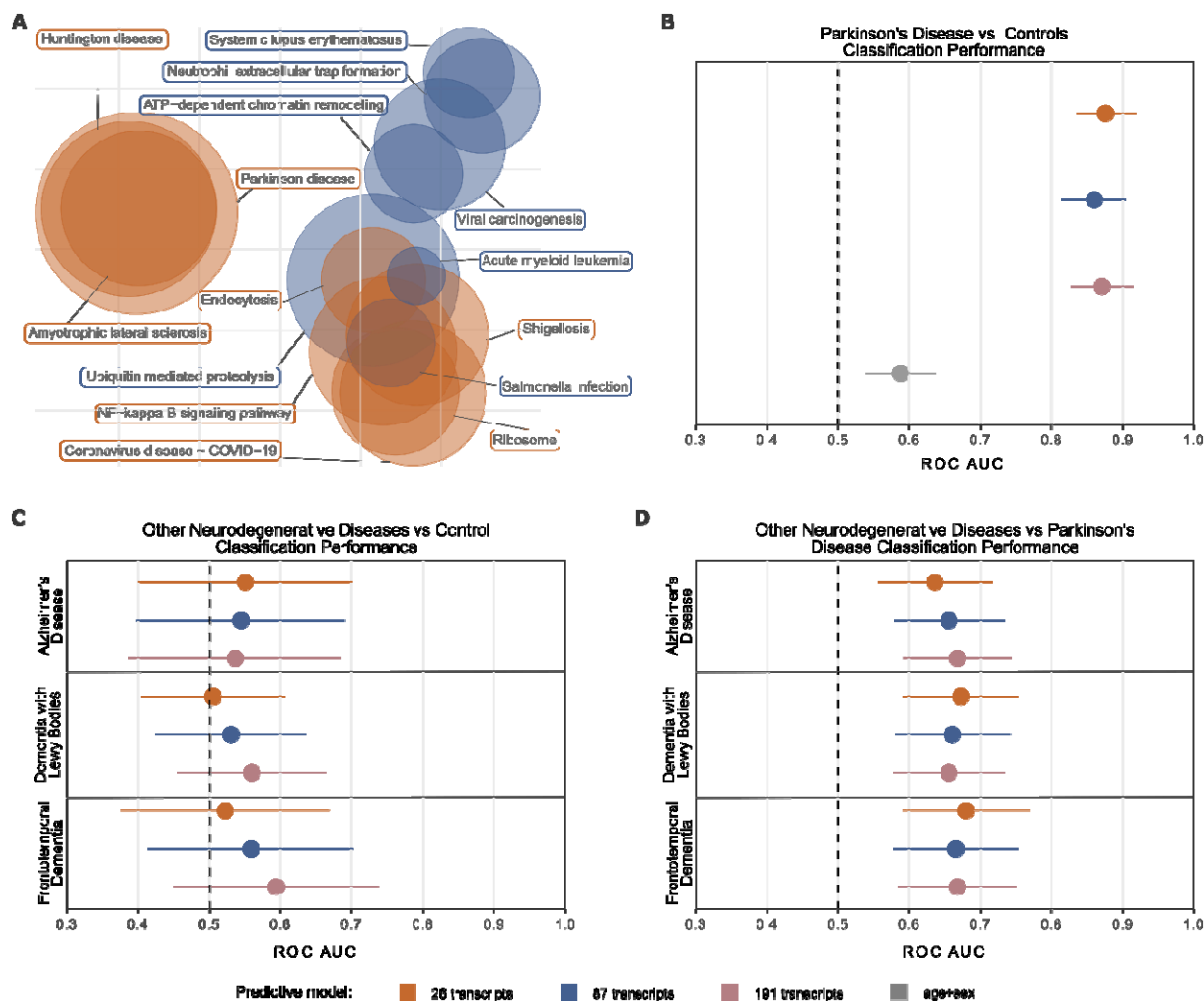
### cfRNA captures transcriptomic signatures corresponding to Parkinson's disease

We wanted to assess if cfRNA changes can be leveraged to build predictive models. To utilize the two independent RNAseq data sets for the development of predictive models, we employed an approach similar to that used previously focusing only on the 2,188 DE transcripts.<sup>33</sup> HUMT was used as training and WUSM as testing. We generated a total of 100 predictive models corresponding to 100 KLD threshold values (increments of 0.01). Based on the balance between ROC-AUC and number of transcripts included in the model we selected three transcript subsets for further follow up (Supplementary Figure 2). Of note, each larger subset is inclusive of the smaller ones. The three selected subsets contained 26, 87 and 191 transcripts, with ROC-AUC values of 0.86, 0.87 and 0.88 in the testing data set (Figure 3B, Supplementary Table 9), respectively.

To contextualize the role of transcripts included in the predictive models, we checked whether any of the transcripts were enriched in the pathways identified in KEGG pathway enrichment analyses described including all the DE transcripts. We found that our smallest subset, consisting of 26 transcripts, captured transcripts that pertain to PD, ubiquitin mediated proteolysis, HD, and ALS (Figure 3A). The next subset, with 87 transcripts, further captures transcripts involved in endocytosis (Figure 3A). Next, we checked whether proteins translated from any of the transcripts included in the predictive models were differentially accumulated in plasma and found two proteins, BMP6 and PARP1, to be differentially accumulated.

Finally, we explored whether the selected transcripts reflected motor symptom severity, measured by Unified Parkinson's Disease Rating Scale (UPDRS) Part III (UPDRS-III), and cognitive status measured by The Montreal Cognitive Assessment (MoCA). UPDRS-III scores were available for both datasets, but not for all participants. Due to sample size ( $n_{\text{HUMT}}=78$ ,  $n_{\text{WUSM}}=28$ ), we combined HUMT and WUSM data for UPDRS-III analysis and observed that three (*PLAC8*,  $p=1.82 \times 10^{-4}$ ; *PTK2B*,  $p=0.02$ ; *ATP5F1B*,  $p=0.04$ ) of the 191 selected transcripts correlated negatively, and one (*RPS27P8*,  $p=0.02$ ) correlated positively, with UPDRS-III (Supplementary Table 10). MoCA information was available for a subset of the WUSM population ( $n=28$ ). We found two cfRNA transcripts (*FCGR3A*,  $p=0.01$ ; *RERE*,  $p=0.04$ ) with significant positive correlations to MoCA scores (Supplementary Table 10).

**Figure 3. Plasma cfRNA capture signatures associated with PD pathology.** **A.** Scatter plot summarizing the significant results of KEGG pathway analysis, in gold are pathways first captured in all the predictive models and in blue pathways captured in 87 and 191 transcript models; Whisker plot showing the performance of the evaluated predictive models to differentiate between **B.** PD and HC; **C.** AD, DLB, or FTD, and HC; and **D.** AD, DLB, or FTD, and PD.



### The predictive models are specific to PD

To assess whether the predictive models were specific to PD, we tested the models in samples from Dementia with Lewy bodies (DLB, n=29), Alzheimer's disease (AD, n=44), and Frontotemporal dementia (FTD, n=16), using two approaches. Firstly, we evaluated if the models could differentiate between each neurodegenerative disease and healthy controls and secondly, if they could discern between PD and DLB, AD, or FTD. The models exhibited low predictive power to differentiate between healthy controls and AD ( $0.53 < AUC < 0.55$ ), DLB

(0.51<AUC<0.57), or FTD (0.52<AUC<0.59; Figure 3C, Supplementary Table 11), confirming that the models are specific to PD. Models performed slightly better in classifying AD (0.64<AUC<0.67), DLB (0.65<AUC<0.67), or FTD (0.67<AUC<0.68; Figure 3D, Supplementary Table 11) when compared to PD. Given the known association between *APOE* genotype and AD risk<sup>60</sup>, we tested whether the addition of *APOE* genotype affects the ability of the predictive models to differentiate between AD and PD. With the inclusion of *APOE* information, we observed an increase in the power of our predictive models to discern between AD and PD (0.85<AUC<0.88).

To gain more insight about whether any of the transcripts included in the predictive models were commonly dysregulated across different NDDs, we investigated plasma expression patterns across PD, DLB, AD and FTD. We observe the most remarkable difference in FTD, where the same transcripts seem to be dysregulated in opposite direction to PD (Supplementary Figure 3).

## DISCUSSION

In the present study, we leveraged two independent datasets to identify plasma cfRNAs that were dysregulated in PD participants and, for the first time in PD, employ plasma cfRNAs to develop predictive models that can distinguish between PD and healthy controls. Overall, we identified 2,188 transcripts that were DE in plasma of PD participants. Through pathway analyses we showed that the identified transcripts are part of PD-associated pathways, such as endocytosis<sup>61</sup>, ubiquitin mediated proteolysis<sup>62-64</sup>, PD, and PD-related cellular components and biological processes such as ubiquitin ligase complex<sup>62-64</sup>, primary lysosome, lysosome transport<sup>65</sup> and exocytosis.<sup>66,67</sup> Furthermore, we show that our findings in plasma are consistent with those in blood by utilizing publicly available PD blood bulk RNAseq data to replicate 424 of our findings. Additionally, we find that protein products of twelve of the identified acellular transcripts are also dysregulated in plasma. One of the twelve transcripts, *COL6A3*, has previously been indicated in other neurologic disorders, namely muscular dystrophy<sup>68</sup> and dystonia<sup>69</sup>. Interestingly, three of the twelve transcripts, *ITGAM*<sup>70</sup>, *SERPINB8*<sup>71</sup>, and *SLC27A4*<sup>72,73</sup>, are associated with dry, flaky skin, which is a common symptom of PD.

Leveraging an in-house brain bulk RNAseq dataset, we showed that plasma cfRNA captures changes occurring in the brains of PD participants, most likely due to blood-brain barrier (BBB) leakage.<sup>74,75</sup> Moreover, we found a significant overlap between transcripts DE in plasma and the corresponding proteins in CSF. Specifically, 24 differentially accumulated CSF proteins are produced from transcripts DE in both plasma and brain. Several of these proteins/transcripts are associated with other neurodegenerative or movement disorders, such as AD (*KLC1*<sup>76</sup>) and dystonia (*SYNE2*<sup>77</sup>), or PD-related pathologies and pathways (*ARF3*<sup>78,79</sup>, *DCTN2*<sup>80</sup>, *HERC1*). Specifically, *ARF3* contributes to the disruption of Golgi apparatus<sup>78,79</sup> and there is evidence of Golgi fragmentation in PD.<sup>81</sup> *DCTN2* colocalizes with phosphorylated *SNCA* in Lewy bodies in participants with PD and DLB.<sup>80</sup> *HERC1* is an E3 ubiquitin protein ligase whose dysregulation

leads to alterations in the endosomal system<sup>82</sup>, and both atypical ubiquitination and dysfunction of endosomal system are hallmarks of PD pathobiology.<sup>63</sup>

Lastly, we showed that 196 of the 2188 identified transcripts originate from genomic regions associated with PD. Of those, 121 have some level of support as the driving gene for a given locus.<sup>49,50</sup> Remarkably, six of these transcripts are encoded by genes nominated by the PD GWAS studies (*ADORA2B*, *DYRK1A*, *GCH1*, *PNA1*, *MCCC1*, *TMEM163*), adding additional evidence to the already prioritized genes.<sup>49,83</sup> All but *ADORA2B* had high supporting scores, ranging from seven (*MCCC1*) to ten (*DYRK1A*), in the GWAS browser. Interestingly, *DYRK1A* has the same level of support as *SNCA*, a well-known PD-associated gene, in the locus browser<sup>50</sup>, adding evidence of the involvement of *DYRK1A* in the pathobiology of PD.

To date, there are no established biomarkers for accurate PD diagnosis, yet with  $\alpha$ -synuclein seed amplification assays (SAA) having promising performance in CSF.<sup>14</sup> We utilized the identified DE transcripts to develop minimally-invasive predictive models capable of differentiating between PD and healthy controls. We developed three scalable models with high predictive power to classify PD, with AUC values between 0.86 and 0.88. Two of the predictive transcripts, *BMP6* and *PARP1*, are also dysregulated on a protein level in plasma, which opens additional possibilities to leverage the proteins as biomarkers. Interestingly, increased levels of *BMP6*<sup>84</sup> have been associated with AD, while *PARP1* is connected with  $\alpha$ -synuclein pathology and PD<sup>85</sup>, but they have not been explored as biomarkers. Regardless, we showed that the transcriptomic models are specific to PD, as they are unable to differentiate between DLB, FTD or AD and healthy controls and capture pathways relevant to the known pathobiology of PD. Notably, we observed that the predictive transcripts are dysregulated in FTD in the opposite direction to PD, suggesting that underlying molecular pathways could be shared between PD and FTD, though regulated differently. Due to the complexity of PD, and the biological interplay between DNA, RNA, and protein functions, we believe that future PD biomarkers might benefit from integrating proteomic and transcriptomic data. Mapping the predictive model transcripts to the pathway analysis results, we show that model transcripts are involved in PD relevant pathways, such as PD, ubiquitin mediated proteolysis, as well as other nervous system disorders, like HD and ALS, adding further evidence of the potential translatability of these models, and their biological relevance.

Four transcripts, *PLAC8*, *PTK2B*, *RPS27P8*, *ATP5F1B*, included in the models correlate significantly with motor symptoms measured by UPDRS-III scale, several of which have known links to PD. *PTK2B* is highly expressed in the nervous system and has been indicated in AD for its role in synaptic homeostasis.<sup>87-89</sup> In PD, *PTK2B* is associated with variant rs11060180 ( $p=1.12 \times 10^{-4}$ ), considered to be a PD-risk allele.<sup>90</sup> Similarly, *ATP5F1B* has no known correlation to PD, but is associated with dystonia, which, like PD, is a movement disorder, further supporting its potential relation with UPDRS-III.<sup>92</sup> Finally, two transcripts, *RERE* and *FCGR3A*, correlate with dementia symptoms measured by MoCA scores. The latter of the two have already been associated with memory disorders, highlighting the biological relevance of our predictive models.<sup>93</sup> Together, this further shows that acellular transcripts are truly capturing



PD pathology and the complexity of the movement disorder diseases and potentially points to different mechanisms causing the motor versus cognitive symptoms in PD participants.

This study has several limitations. While we accessed two independent populations from different repositories, we are limited by sample size and in the amount of clinical information we had available for HUMT population. This has made it challenging to account for all relevant biological, medical, and technical variables. Further, samples have been stored in the freezer for varying amounts of time, which is known to affect the quality of RNA, and thus can potentially impact our findings. To minimize the effects of missing clinical information and storage time, we removed from the analyses any transcripts that showed selective degradation or were associated with medication in WUSTL population as previously described.<sup>33</sup> Finally, larger sample sizes of both PD samples, as well as other NDDs, would allow for further validation and sensitivity testing of the developed predictive models.

Nonetheless, this study is the first of its kind and shows that cfRNAs have a potential to aid in diagnosis of PD as cost effective, minimally invasive biomarkers. We identified several plasma transcripts that have already been associated to PD or relevant pathways, correlate with symptom severity, and could potentially be leveraged for disease monitoring. On top of that not only are some of those transcripts dysregulated in different tissues, but they are also encoded in known PD loci. Additionally, some of them result in alterations in protein abundance when compared to healthy controls in relevant tissues. Overall, we believe that we have demonstrated that cfRNA has the power to capture changes relevant to PD pathology, has the potential to be translated to the clinic, and if replicated and validated in larger samples sizes, would benefit the whole PD community by providing non-invasive biomarkers, that are cost-effective and can be implemented in remote areas, providing access to care for all.

## **DATA AVAILABILITY**

Data is available from the NIAGADS Data Sharing Service (NIAGADS dss), accession number NG00142. All original code has been deposited at GitHub (<https://github.com/Ibanez-Lab/PlasmaCellFreeRNA-ParkinsonDisease>) and is publicly available as of the date of publication. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## **ACKNOWLEDGEMENTS**

We thank all the participants and their families along with the institutions and all the staff who provided plasma tissue, without whom this study would not have been possible. This work was supported by access to equipment made possible by the Hope Center for Neurological Disorders, the Neurogenomics and Informatics Center (NGI: <https://neurogenomics.wustl.edu/>) and the Departments of Neurology and Psychiatry at Washington University School of Medicine. This work was supported by grants from the National Institutes of Health K99/R00-AG062723, P01AG003991, RO1 NS075321, P30 AG066444, P01 AG03991, and P01 AG026276, the Michael J Fox Foundation (MJFF-021599 and MJFF-025292), the Alzheimer's Drug Discovery

Foundation (GDAPB-201807-2015632), Bright Focus Foundation (A2021033S), and the Department of Defense (W81XWH2010849).

#### **AUTHOR CONTRIBUTIONS**

LI, and AB conceived and wrote this article. LI conceptualized and designed the research plan. LI and AB designed the analysis plan. AB, ACG, CM, RK, IAD, KB, and IS processed all the data and performed the analyses. IA and SY obtained and processed the samples. IA and ML obtained and curated clinical information. AB, ACG, CM, RK, IAD, KB, IS, SY, JS, DW, ML, IA, JP, SN, PP, GZ, JB, and LI discussed the project, revised the manuscript, and provided critical feedback.

#### **COMPETING INTERESTS**

The funders of the study had no role in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. SY is an employee of Daiichi Sankyo Co., Ltd. LI and AB are named as inventors on the invention disclosure filed in relation to this work. The rest of the authors report no conflict of interest.

## REFERENCES

1. Optimizing Parkinson's disease diagnosis: the role of a dual nuclear imaging algorithm | npj Parkinson's Disease. <https://www.nature.com/articles/s41531-018-0041-9>.
2. Kalia, L. V. & Lang, A. E. Parkinson's disease. *The Lancet* **386**, 896–912 (2015).
3. Memou, A. *et al.* Defining (and blocking) neuronal death in Parkinson's disease: Does it matter what we call it? *Brain Research* **1771**, 147639 (2021).
4. Tolosa, E., Garrido, A., Scholz, S. W. & Poewe, W. Challenges in the diagnosis of Parkinson's disease. *The Lancet Neurology* **20**, 385–397 (2021).
5. Henderson, M. X., Trojanowski, J. Q. & Lee, V. M.-Y.  $\alpha$ -Synuclein pathology in Parkinson's disease and related  $\alpha$ -synucleinopathies. *Neuroscience Letters* **709**, 134316 (2019).
6. Armstrong, M. J. & Okun, M. S. Diagnosis and Treatment of Parkinson Disease: A Review. *JAMA* **323**, 548–560 (2020).
7. Surguchov, A. Biomarkers in Parkinson's Disease. in *Neurodegenerative Diseases Biomarkers: Towards Translating Research to Clinical Practice* (eds. Peplow, P. V., Martinez, B. & Gennarelli, T. A.) 155–180 (Springer US, New York, NY, 2022). doi:10.1007/978-1-0716-1712-0\_7.
8. Hastings, A. *et al.* Neuropathologic Validation and Diagnostic Accuracy of Presynaptic Dopaminergic Imaging in the Diagnosis of Parkinsonism. *Neurology* **102**, e209453 (2024).
9. Di Luca, D. G. & Perlmutter, J. S. Time for Clinical Dopamine Transporter Scans in Parkinsonism?: Not DAT Yet. *Neurology* **102**, e209558 (2024).
10. Miller, D. B. & O'Callaghan, J. P. Biomarkers of Parkinson's disease: Present and future. *Metabolism* **64**, S40–S46 (2015).
11. Karimi, M. *et al.* Validation of nigrostriatal positron emission tomography measures: Critical limits. *Annals of Neurology* **73**, 390–396 (2013).

12. Parnetti, L. *et al.* CSF and blood biomarkers for Parkinson's disease. *The Lancet Neurology* **18**, 573–586 (2019).
13. Orrù, C. D. *et al.* A rapid  $\alpha$ -synuclein seed assay of Parkinson's disease CSF panel shows high diagnostic accuracy. *Ann Clin Transl Neurol* **8**, 374–384 (2020).
14. Siderowf, A. *et al.* Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using  $\alpha$ -synuclein seed amplification: a cross-sectional study. *The Lancet Neurology* **22**, 407–417 (2023).
15. Parnetti, L. *et al.* Cerebrospinal fluid biomarkers in Parkinson disease. *Nat Rev Neurol* **9**, 131–140 (2013).
16. Kwon, E. H. *et al.* Update on CSF Biomarkers in Parkinson's Disease. *Biomolecules* **12**, 329 (2022).
17. Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis Author Response | Neurology. <https://n.neurology.org/content/87/2/237.long>.
18. Everett, T. R. & Chitty, L. S. Cell-free fetal DNA: the new tool in fetal medicine. *Ultrasound Obstet Gynecol* **45**, 499–507 (2015).
19. Wright, C. F. & Chitty, L. S. Cell-free fetal DNA and RNA in maternal blood: implications for safer antenatal testing. *BMJ* **339**, b2451 (2009).
20. Drag, M. H. & Kilpeläinen, T. O. Cell-free DNA and RNA—measurement and applications in clinical diagnostics with focus on metabolic disorders. *Physiological Genomics* **53**, 33–46 (2021).
21. Papadopoulou, E. *et al.* Cell-free DNA and RNA in Plasma as a New Molecular Marker for Prostate and Breast Cancer. *Annals of the New York Academy of Sciences* **1075**, 235–243 (2006).
22. Bronkhorst, A. J., Ungerer, V. & Holdenrieder, S. The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomolecular Detection and Quantification* **17**, 100087 (2019).

23. Gielis, E. M. *et al.* Cell-Free DNA: An Upcoming Biomarker in Transplantation. *American Journal of Transplantation* **15**, 2541–2551 (2015).
24. Tzimagiorgis, G., Michailidou, E. Z., Kritis, A., Markopoulos, A. K. & Kouidou, S. Recovering circulating extracellular or cell-free RNA from bodily fluids. *Cancer Epidemiol* **35**, 580–589 (2011).
25. Redova, M., Sana, J. & Slaby, O. Circulating miRNAs as new blood-based biomarkers for solid cancers. *Future Oncology* **9**, 387–402 (2013).
26. Zaporozhchenko, I. A., Ponomaryova, A. A., Rykova, E. Y. & Laktionov, P. P. The potential of circulating cell-free RNA as a cancer biomarker: challenges and opportunities. *Expert Review of Molecular Diagnostics* **18**, 133–145 (2018).
27. Larson, M. H. *et al.* A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. *Nat Commun* **12**, 2357 (2021).
28. Roskams-Hieter, B. *et al.* Plasma cell-free RNA profiling distinguishes cancers from pre-malignant conditions in solid and hematologic malignancies. *npj Precis. Onc.* **6**, 1–11 (2022).
29. Müller, M., Kuiperij, H. B., Claassen, J. A., Küsters, B. & Verbeek, M. M. MicroRNAs in Alzheimer's disease: differential expression in hippocampus and cell-free cerebrospinal fluid. *Neurobiology of Aging* **35**, 152–158 (2014).
30. Sheinerman, K. S. *et al.* Circulating brain-enriched microRNAs as novel biomarkers for detection and differentiation of neurodegenerative diseases. *Alzheimers Res Ther* **9**, 89 (2017).
31. Zendjabil, M. Circulating microRNAs as novel biomarkers of Alzheimer's disease. *Clinica Chimica Acta* **484**, 99–104 (2018).
32. Toden, S. *et al.* Noninvasive characterization of Alzheimer's disease by circulating, cell-free messenger RNA next-generation sequencing. *Science Advances* **6**, eabb1654 (2020).
33. Cisterna-García, A. *et al.* Cell-free RNA signatures predict Alzheimer's disease. *iScience* **26**, (2023).

34. Cai, M. *et al.* Aberrant Expression of Circulating MicroRNA Leads to the Dysregulation of Alpha-Synuclein and Other Pathogenic Genes in Parkinson's Disease. *Frontiers in Cell and Developmental Biology* **9**, (2021).
35. Inverse relation between in vivo amyloid imaging load and cerebrospinal fluid Abeta42 in humans - PubMed. <https://pubmed.ncbi.nlm.nih.gov/16372280/>.
36. Andrews, Simon. FastQC A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
37. STAR: ultrafast universal RNA-seq aligner | Bioinformatics | Oxford Academic. <https://academic.oup.com/bioinformatics/article/29/1/15/272537?login=true>.
38. Picard Tools - By Broad Institute. <http://broadinstitute.github.io/picard/>.
39. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
40. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat Methods* **14**, 417–419 (2017).
41. Ewels, P., Magnusson, M., Lundin, S. & Källér, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
42. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
43. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
44. Differential meta-analysis of RNA-seq data from multiple studies - PubMed. <https://pubmed.ncbi.nlm.nih.gov/24678608/>.

45. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).
46. Parkinson Progression Marker, I. The Parkinson Progression Marker Initiative (PPMI). *Prog Neurobiol* **95**, 629–35 (2011).
47. Feleke, R. *et al.* Cross-platform transcriptional profiling identifies common and distinct molecular pathologies in Lewy body diseases. *Acta Neuropathol* **142**, 449–474 (2021).
48. Xu, J. *et al.* Human striatal glia differentially contribute to AD- and PD-specific neurodegeneration. *Nat Aging* **3**, 346–365 (2023).
49. Kim, J. J. *et al.* Multi-ancestry genome-wide association meta-analysis of Parkinson’s disease. *Nat Genet* **56**, 27–36 (2024).
50. Grenn, F. P. *et al.* The Parkinson’s Disease Genome-Wide Association Study Locus Browser. *Mov Disord* **35**, 2056–2067 (2020).
51. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
52. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1–22 (2010).
53. Hausser, J. & Strimmer, K. *Entropy: Estimation of Entropy, Mutual Information and Related Quantities*. (2021).
54. Fernandez, M. V. *et al.* Genetic and multi-omic resources for Alzheimer disease and related dementia from the Knight Alzheimer Disease Research Center. *Sci Data* **11**, 768 (2024).
55. Hughes, A. J., Daniel, S. E. & Lees, A. J. Improved accuracy of clinical diagnosis of Lewy body Parkinson’s disease. *Neurology* **57**, 1497–1499 (2001).
56. Beric, A. *et al.* Circulating blood circular RNA in Parkinson’s Disease; a systematic study. 2024.01.22.24301623 Preprint at <https://doi.org/10.1101/2024.01.22.24301623> (2024).

57. Ofori, E., Du, G., Babcock, D., Huang, X. & Vaillancourt, D. E. Parkinson's disease biomarkers program brain imaging repository. *Neuroimage* **124**, 1120–1124 (2016).
58. Wang, Q. *et al.* Molecular profiling of human substantia nigra identifies diverse neuron types associated with vulnerability in Parkinson's disease. *Science Advances* **10**, eadi8287 (2024).
59. Le Guen, Y. *et al.* Common X-Chromosome Variants Are Associated with Parkinson Disease Risk. *Ann Neurol* **90**, 22–34 (2021).
60. Narasimhan, S. *et al.* Apolipoprotein E in Alzheimer's disease trajectories and the next-generation clinical care pathway. *Nat Neurosci* 1–17 (2024) doi:10.1038/s41593-024-01669-5.
61. Vidyadhara, D. J., Lee, J. E. & Chandra, S. S. Role of the endolysosomal system in Parkinson's disease. *J Neurochem* **150**, 487–506 (2019).
62. McNaught, K. S. & Jenner, P. Proteasomal function is impaired in substantia nigra in Parkinson's disease. *Neurosci Lett* **297**, 191–194 (2001).
63. Buneeva, O. & Medvedev, A. Atypical Ubiquitination and Parkinson's Disease. *Int J Mol Sci* **23**, 3705 (2022).
64. Bi, M., Du, X., Jiao, Q., Chen, X. & Jiang, H. Expanding the role of proteasome homeostasis in Parkinson's disease: beyond protein breakdown. *Cell Death Dis* **12**, 1–16 (2021).
65. Navarro-Romero, A., Montpeyó, M. & Martínez-Vicente, M. The Emerging Role of the Lysosome in Parkinson's Disease. *Cells* **9**, 2399 (2020).
66. Ramezani, M. *et al.* Regulation of exocytosis and mitochondrial relocalization by Alpha-synuclein in a mammalian cell model. *npj Parkinsons Dis.* **5**, 1–17 (2019).
67. Xie, Y. X. *et al.* Lysosomal exocytosis releases pathogenic  $\alpha$ -synuclein species from neurons in synucleinopathy models. *Nat Commun* **13**, 4918 (2022).
68. Demir, E. *et al.* Mutations in COL6A3 cause severe and mild phenotypes of Ullrich congenital muscular dystrophy. *Am J Hum Genet* **70**, 1446–1458 (2002).



69. Zech, M. *et al.* Recessive mutations in the  $\alpha 3$  (VI) collagen gene COL6A3 cause early-onset isolated dystonia. *Am J Hum Genet* **96**, 883–893 (2015).
70. Fagerholm, S. C., MacPherson, M., James, M. J., Sevier-Guy, C. & Lau, C. S. The CD11b-integrin (ITGAM) and systemic lupus erythematosus. *Lupus* **22**, 657–663 (2013).
71. Pigors, M. *et al.* Loss-of-Function Mutations in SERPINB8 Linked to Exfoliative Ichthyosis with Impaired Mechanical Stability of Intercellular Adhesions. *Am J Hum Genet* **99**, 430–436 (2016).
72. Klar, J. *et al.* Mutations in the fatty acid transport protein 4 gene cause the ichthyosis prematurity syndrome. *Am J Hum Genet* **85**, 248–253 (2009).
73. Li, S., Green, J. F. & Jin, M. Impacts of deletion and ichthyosis prematurity syndrome-associated mutations in fatty acid transport protein 4 on the function of RPE65. *FEBS Lett* **594**, 540–552 (2020).
74. Al-Bachari, S., Naish, J. H., Parker, G. J. M., Emsley, H. C. A. & Parkes, L. M. Blood-Brain Barrier Leakage Is Increased in Parkinson's Disease. *Front Physiol* **11**, 593026 (2020).
75. Sweeney, M. D., Sagare, A. P. & Zlokovic, B. V. Blood-brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nat Rev Neurol* **14**, 133–150 (2018).
76. Chen, X.-Q., Das, U., Park, G. & Mobley, W. C. Normal levels of KIF5 but reduced KLC1 levels in both Alzheimer disease and Alzheimer disease in Down syndrome: evidence suggesting defects in anterograde transport. *Alzheimers Res Ther* **13**, 59 (2021).
77. Lee, F.-C., Wu, S., Hsu, C.-S., Huang, S.-M. & Hong, J.-S. Association of SYNE2 variants in accelerating the progress of DYT1 early-onset isolated dystonia.
78. Sakamoto, M. *et al.* De novo ARF3 variants cause neurodevelopmental disorder with brain abnormality. *Human Molecular Genetics* **31**, 69–81 (2022).
79. El Ghouzzi, V. & Boncompain, G. Golgipathies reveal the critical role of the sorting machinery in brain and skeletal development. *Nat Commun* **13**, 7397 (2022).
80. Shen, C. *et al.* Dynactin is involved in Lewy body pathology. *Neuropathology* **38**, 583–590 (2018).

81. Rendón, W. O., Martínez-Alonso, E., Tomás, M., Martínez-Martínez, N. & Martínez-Menárguez, J. A. Golgi fragmentation is Rab and SNARE dependent in cellular models of Parkinson's disease. *Histochem Cell Biol* **139**, 671–684 (2013).
82. Limanaqi, F. *et al.* Promiscuous Roles of Autophagy and Proteasome in Neurodegenerative Proteinopathies. *Int J Mol Sci* **21**, 3028 (2020).
83. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology* **18**, 1091–1102 (2019).
84. Crews, L. *et al.* Increased BMP6 Levels in the Brains of Alzheimer's Disease Patients and APP Transgenic Mice Are Accompanied by Impaired Neurogenesis. *J. Neurosci.* **30**, 12252–12262 (2010).
85. Salemi, M. *et al.* Poly (ADP-ribose) polymerase 1 and Parkinson's disease: A study in *post-mortem* human brain. *Neurochemistry International* **144**, 104978 (2021).
86. Gisbert, S. *et al.* Complexin-1 and Foxp1 Expression Changes Are Novel Brain Effects of Alpha-Synuclein Pathology. *Mol Neurobiol* **52**, 57–63 (2015).
87. Beck, T. N., Nicolas, E., Kopp, M. C. & Golemis, E. A. Adaptors for disorders of the brain? The cancer signaling proteins NEDD9, CASS4, and PTK2B in Alzheimer's disease. *Oncoscience* **1**, 486–503 (2014).
88. Astillero-Lopez, V. *et al.* Proteomic analysis identifies HSP90AA1, PTK2B, and ANXA2 in the human entorhinal cortex in Alzheimer's disease: Potential role in synaptic homeostasis and A $\beta$  pathology through microglial and astroglial cells. *Brain Pathology* **34**, e13235 (2024).
89. Niso-Santano, M., Fuentes, J. M. & Galluzzi, L. Immunological aspects of central neurodegeneration. *Cell Discov* **10**, 1–14 (2024).
90. Chan, G. *et al.* Trans-pQTL study identifies immune crosstalk between Parkinson and Alzheimer loci. *Neurol Genet* **2**, e90 (2016).

91. Zubenko, G. S., Hughes, H. B. & Zubenko, W. N. D10S1423 Identifies a Susceptibility Locus for Alzheimer's Disease (AD7) in a Prospective, Longitudinal, Double-Blind Study of Asymptomatic Individuals: Results at 14 Years. *Am J Med Genet B Neuropsychiatr Genet* **153B**, 359–364 (2010).
92. Nasca, A. *et al.* Variants in ATP5F1B are associated with dominantly inherited dystonia. *Brain* **146**, 2730–2738 (2023).
93. Li, D. *et al.* Bioinformatic identification of potential biomarkers and therapeutic targets in carotid atherosclerosis and vascular dementia. *Front Neurol* **13**, 1091453 (2023).