

ARTICLE OPEN



Early-stage idiopathic Parkinson's disease is associated with reduced circular RNA expression

Benjamin J. Whittle¹, Osagie G. Izuogu², Hannah Lowes³, Dasha Deen³, Angela Pyle³, Jon Coxhead⁴, Rachael A. Lawson⁵, Alison J. Yarnall⁵, Michael S. Jackson⁴, Mauro Santibanez-Koref⁴ and Gavin Hudson¹

Neurodegeneration in Parkinson's disease (PD) precedes diagnosis by years. Early neurodegeneration may be reflected in RNA levels and measurable as a biomarker. Here, we present the largest quantification of whole blood linear and circular RNAs (circRNA) in early-stage idiopathic PD, using RNA sequencing data from two cohorts (PPMI = 259 PD, 161 Controls; ICICLE-PD = 48 PD, 48 Controls). We identified a replicable increase in *TMEM252* and *LMNB1* gene expression in PD. We identified novel differences in the expression of circRNAs from *ESYT2*, *BMS1P1* and *CCDC9*, and replicated trends of previously reported circRNAs. Overall, using circRNA as a diagnostic biomarker in PD did not show any clear improvement over linear RNA, minimising its potential clinical utility. More interestingly, we observed a general reduction in circRNA expression in both PD cohorts, accompanied by an increase in *RNASEL* expression. This imbalance implicates the activation of an innate antiviral immune response and suggests a previously unknown aspect of circRNA regulation in PD.

npj Parkinson's Disease (2024)10:25; <https://doi.org/10.1038/s41531-024-00636-y>

INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disease characterised by a progressive loss of dopaminergic neurons in the substantia nigra¹. Diagnosis is predominantly based on the presentation of cardinal symptoms, including motor (such as bradykinesia and gait disturbances) and non-motor (sleep problems, cognitive impairment and dementia) symptoms¹, which result in substantial morbidity². Unfortunately, the clinical manifestation of PD typically occurs years after the onset of dopaminergic neuron loss, making early diagnosis challenging³. Measures to slow the rate of progression, therefore, remain a key goal in PD research⁴. The reliance on the observation of clinical symptoms hampers the detection in the earliest phases of neurodegeneration when disease-modifying intervention would be most beneficial⁵. Additionally, diagnostic accuracy is adversely affected by shared clinical presentations with other neurodegenerative conditions, resulting in an estimated misdiagnosis rate of ~25% in the early stages of disease³. The identification of biomarkers that can facilitate accurate early diagnosis and lead to the discovery of new therapeutic targets are therefore crucial for the future clinical management of PD, particularly in the early stages of disease⁶.

Despite significant effort, few PD biomarkers have been translated into clinical practice⁶, in part limited by a lack of replication⁷. Efforts to identify biomarkers of PD onset have been hampered by heterogeneity in clinical presentation and the rate of progression of patients⁸. Familial PD cases, both dominant and recessive, account for ~10–15% of cases⁹, whilst the majority of remaining cases are considered idiopathic. The current largest GWAS meta-analysis identified 90 variants explaining ~16–36% of PD heritability¹⁰, yet polygenic risk scoring is not currently at a stage to be used clinically to predict an individual's risk¹⁰. Neuroimaging show promising results,

with accuracies typically >80%¹¹. However, these techniques are expensive and usually most accurate in advanced disease¹² or when applied in combination¹³.

Biofluid-based biomarker candidates, including neurochemicals such as orexin¹⁴ and metabolites including lactate and acylcarnitines^{15,16} have been proposed. Recent work assessing α -synuclein seeding in CSF is promising in specific PD subgroups¹⁷, although the invasive nature of a lumbar puncture limits its utility. The accessibility of whole blood, as well as a high degree of overlap with neuronal expression (>80% shared gene expression)^{18,19}, has prompted research into assessing blood-based RNAs as biomarkers of PD. Early array-based and, more recently, RNA sequencing-based experiments have identified several messenger RNA (mRNA)^{20–28}, long non-coding RNAs (lncRNA)²⁹ and microRNAs (miRNAs)^{30–33} as potential biomarkers of PD. However, there is little concordance between studies^{24,26,34,35}, likely a result of differences in methodologies, sample cohorts and study design^{24,26,36}.

Circular RNA (circRNA) is an RNA species characterised by the formation of a back-splice junction (BSJ, Fig. 1)³⁷. CircRNA expression can vary across cell, tissue and development stages^{38–40}. Interestingly, circRNA expression can be uncoupled from that of its parental host gene^{40–42} and several modulators of global circRNA levels have been reported^{43–53}. CircRNAs are abundant in the brain⁵⁴, enriched in genes related to neuronal function⁵⁵, and readily identifiable in blood^{40,56}. Their reported resistance to RNA exonuclease activity and increased stability compared to their linear counterparts^{57,58}, have prompted investigations of their potential use as biomarkers in a range of human diseases, including PD^{59,60}. Differential expression of circRNAs in PD has been reported^{59–65}. However, these studies typically use small cohorts, identify specific circRNAs that are not replicated and show limited classification ability when compared

¹Wellcome Centre for Mitochondrial Research, Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK. ²European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, UK. ³Wellcome Centre for Mitochondrial Research, Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK. ⁴Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK. ⁵Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK. ✉email: mauro.santibanez-koref@newcastle.ac.uk; gavin.hudson@ncl.ac.uk

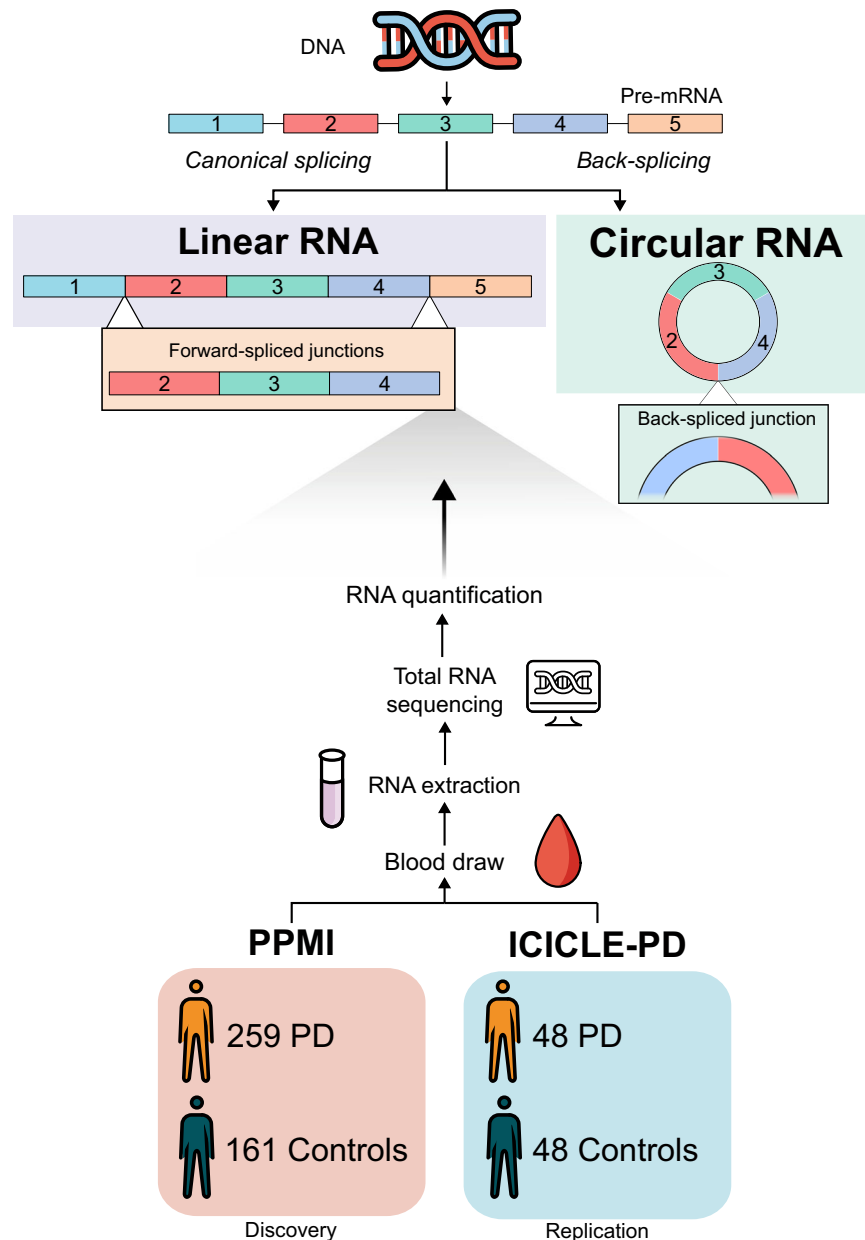


Fig. 1 Circular RNA formation and study overview. The canonical splicing of pre-mRNA produces a linear RNA molecule containing forward-spliced junctions (FSJ). Circular RNA molecules are formed by a back-splicing reaction and can be identified through the presence of a back-spliced junction (BSJ). Individuals from our discovery (PPMI, PD = 259, Controls = 161) and replication (ICICLE-PD, PD = 48, Controls = 48) cohorts underwent whole blood total RNA sequencing to detect and quantify linear and circular RNAs.

to existing clinical predictors. Thus, a large-scale, unbiased assessment of circRNAs in PD is warranted.

Here, we investigate the potential utility of circRNA as biomarkers for early-stage idiopathic PD in two large unrelated cohorts, The Michael J Fox Foundation Parkinson's Progression Markers Initiative⁶⁶ (PPMI) and The Incidence of Cognitive Impairment in Cohorts with Longitudinal Evaluation-PD (ICICLE-PD)⁶⁷, adopting a discovery-replication strategy. Our analysis indicates that, although circular RNAs and some genes are differently expressed, neither has sufficient discriminatory power to displace existing PD biomarkers. Interestingly, we observed a reproducible reduction in general circRNA expression levels in PD patients compared to controls, which was not recapitulated in

canonical splice junction expression, suggestive of a link between circRNA expression and PD.

RESULTS

We utilised high-depth total RNA sequencing of whole blood in two unrelated PD cohorts to investigate circular RNA expression (Fig. 1). CircRNA expression does not always reflect host gene expression^{40–42}, thus prior to analysing circRNAs, we compared gene expression (i.e., linear mRNA) between PD patients and controls in both PPMI and ICICLE-PD. We limited all our analysis to early-stage idiopathic Parkinson's disease (PD) patients (i.e., diagnosed <13 months and no known predisposing genetic

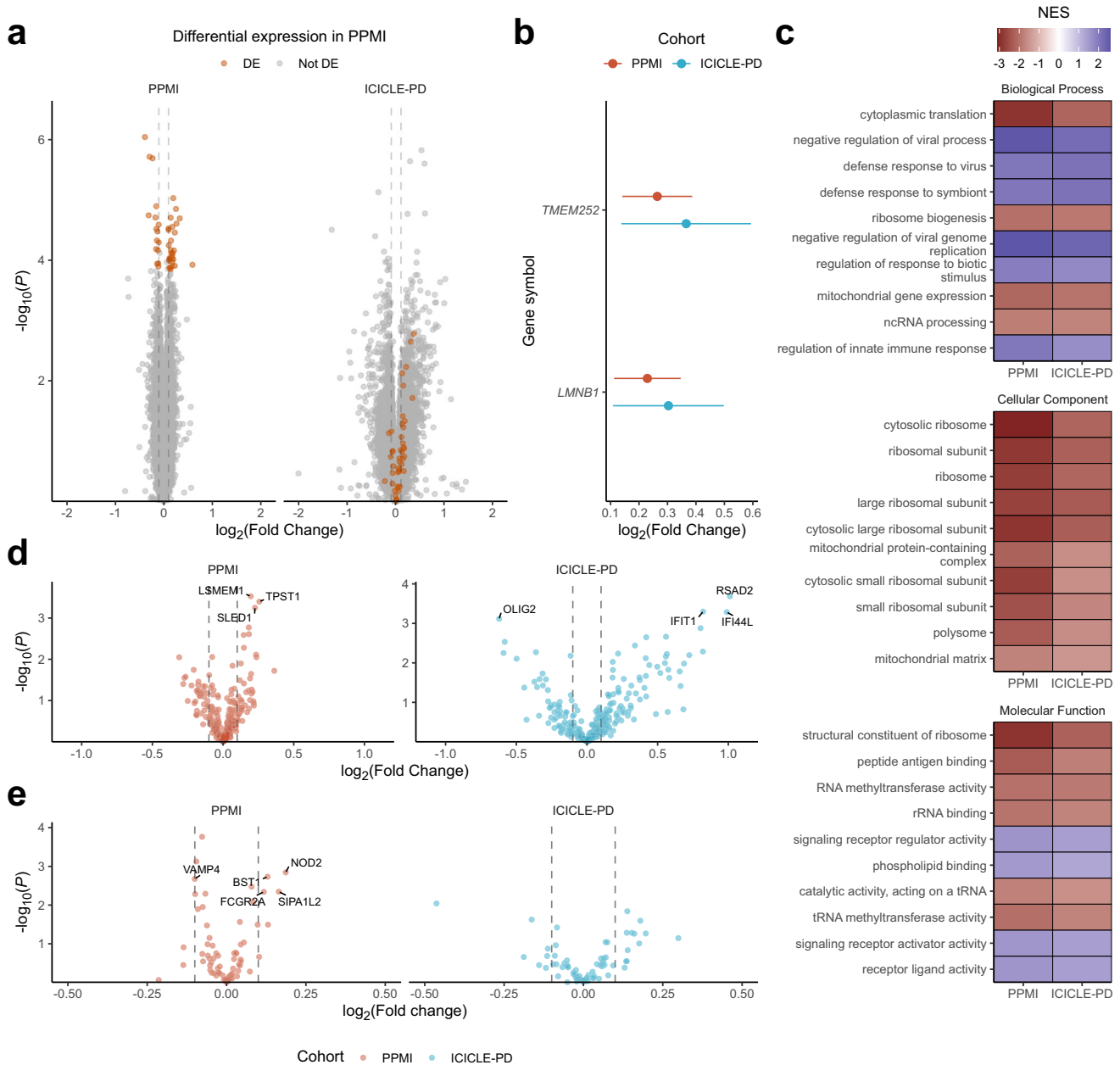


Fig. 2 Gene differential expression in early-stage idiopathic PD. Figures showing the comparative expression of genes between PD cases and controls in PPMI and ICICLE-PD. **a** Volcano plot of differential expression in PPMI and ICICLE-PD, highlighted (orange) are the 44 genes significantly differentially expressed in PPMI ($FDR < 0.05$ and \log_2 fold change $> 0.1 / < -0.1$) and the corresponding results in ICICLE-PD (Supplementary Data 1). **b** Graph showing the comparative \log_2 fold changes and 95% CI of genes (*TMEM252* and *LMNB1*) that were significant ($FDR < 0.05$) were replicated in ICICLE-PD. Displayed are the first 10 terms from each ontology as ranked by P -value in PPMI. NES = Normalised Enrichment Score (Supplementary Data 2). **c** Results from the gene set enrichment of Gene Ontology terms. Terms that were significantly enriched ($FDR < 0.05$) were replicated in ICICLE-PD. Displayed are the first 10 terms from each ontology as ranked by P -value in PPMI. NES = Normalised Enrichment Score (Supplementary Data 2). **d** Volcano plots showing the differential expression of genes reported as differentially expressed in PD from previous blood RNAseq studies^{70–73} (Supplementary Data 4). Genes significantly differentially expressed ($FDR < 0.05$ and \log_2 fold change $> 0.1 / < -0.1$) in each cohort are labelled. **e** Volcano plots showing the differential expression of risk genes highlighted in PD GWAS¹⁰ (Supplementary Data 5). Genes significantly differentially expressed ($FDR < 0.05$ and \log_2 fold change $> 0.1 / < -0.1$) in each cohort are labelled. Grey dashed lines on the volcano plots indicate the \log_2 fold changes 0.1 and -0.1 .

variation) to age and sex-matched controls, using PPMI as a discovery cohort and ICICLE-PD as a replication cohort (Fig. 1 and Supplementary Table 1).

Differential gene expression in early-stage idiopathic PD

Differential gene expression analysis was carried out on 16,191 (PPMI) and 15,852 (ICICLE-PD) genes respectively (overlap of

95.53%). In PPMI, we identified 44 significantly differentially expressed genes ($\log_2\text{FC} < -0.1 / > 0.1$, $FDR < 0.05$), with 28 upregulated and 16 downregulated in PD (Fig. 2a and Supplementary Data 1). Of these 44, *TMEM252* and *LMNB1* were similarly upregulated in ICICLE-PD ($\log_2\text{FC} > 0.1$, $FDR < 0.05$) (Fig. 2a, b and Supplementary Data 1).

To gain an overview of global alterations in gene expression patterns we performed gene set enrichment analysis (GSEA)⁶⁸. After

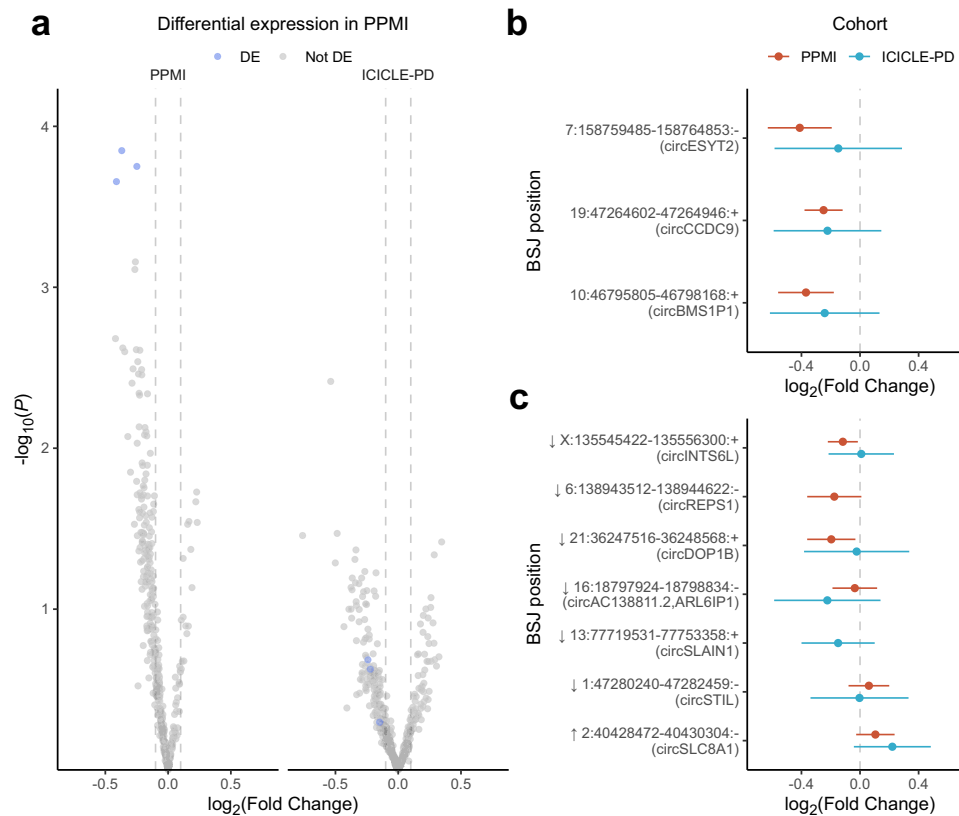


Fig. 3 Circular RNA differential expression in early-stage idiopathic PD. Figures showing the comparative expression of circular RNAs (shown as individual back-spliced junctions, BSJs) between PD cases and controls in PPMI and ICICLE-PD, highlighted (blue) are the three BSJs significantly differentially expressed in PPMI (FDR < 0.05 and \log_2 fold change > 0.1/ < -0.1) and the corresponding results in ICICLE-PD (Supplementary Data 8). Grey dashed lines indicate the \log_2 fold changes of 0.1 and -0.1. **b** Graph showing the comparative \log_2 fold changes of the three BSJs that were significantly differentially expressed in PPMI. None reached significance in ICICLE-PD (P -values > 0.05) (Supplementary Data 8). **c** Graph showing the comparative \log_2 fold changes of BSJs previously reported as being differentially expressed in PD^{61,62} (Supplementary Data 9). The arrows indicate the previously reported direction of change in PD relative to controls. Error bars in **b** and **c** show the 95% CI of the fold change. BSJ positions are reported as chromosome:start-end:strand (GRCh38).

classifying into GO terms, we identified 97 biological processes, 11 molecular functions and 29 cellular components that were significantly enriched in PPMI and subsequently replicated in ICICLE-PD (FDR < 0.05 with the same direction of enrichment) (Supplementary Data 2). Ontologies related to ribosomal function and translation were more likely to be decreased in PD. In contrast, ontologies related to the innate immune response, including antiviral and interferon signalling processes, tended to be increased in PD (the top 10 ranked by PPMI P -value in each ontology are shown in Fig. 2c with all results present in Supplementary Data 2). GSEA of KEGG pathways also showed the significant enrichment of the pathways *Ribosome* (PPMI FDR = 7.58×10^{-24} ; ICICLE-PD FDR = 5.57×10^{-10}) and *NOD-like receptor signalling* (PPMI FDR = 0.0016, ICICLE-PD FDR = 2.27×10^{-6}), part of the innate immune response⁶⁹ (Supplementary Data 3).

We were able to independently replicate (FDR < 0.05) genes previously reported as differentially expressed in the blood⁷⁰⁻⁷³ using PPMI (*LSMEM1*, *TPST1* and *SLED1*) and ICICLE-PD (*IFIT1*, *RSAD2*, *IFI44L* and *OLIG2*) datasets (Fig. 2d and Supplementary Data 4). Overlapping the expression of genes potentially underpinning PD GWAS risk loci¹⁰, identified five significantly differentially expressed genes (*BST1*, *FCGR2A*, *SIPA1L2*, *NOD2* and *VAMP4*, $\log_2\text{FC} > 0.1 / < -0.1$, FDR < 0.05) in PPMI, although none passed multiple testing correction in ICICLE-PD (Fig. 2e and Supplementary Data 5). Overlapping with highly penetrant pathogenic parkinsonism genes, we identified *PTRHD1* as significantly

decreased in PD (FDR = 0.015), with a similar trend reported in ICICLE-PD (FDR = 0.088) (Supplementary Data 6).

CircRNAs are differentially expressed in early-stage idiopathic PD

CircRNAs were detected based on the presence of a back-spliced junction (BSJ, Fig. 1), combining multiple detection algorithms to define high-confidence BSJs (see *Circular RNA detection and quantification*). Filtering of lower expressed circRNAs left a set of abundant circRNAs (PPMI = 403, ICICLE-PD = 457, 62.57% overlap). We observed no overrepresentation of GO terms after grouping circRNAs based on their host gene (Supplementary Data 7). We identified three BSJs significantly reduced in PD patients versus controls in PPMI ($\log_2\text{FC} < -0.1 / > 0.1$, FDR < 0.05) (Fig. 3a, b and Supplementary Data 8). These BSJs, derived from the genes *ESYT2*, *BMS1P1*, and *CCDC9* were similarly decreased in ICICLE-PD but did not reach statistical significance (FDR > 0.05) (Fig. 3a, b and Supplementary Data 8). Seven circRNAs previously reported as differentially expressed in PD^{61,62} were sufficiently expressed in the PPMI or ICICLE-PD cohorts (Supplementary Data 9). Two BSJs, derived from *DOP1B* and *INTS6L*, showed altered expression in PPMI (P -value < 0.05) (Supplementary Data 9). While no BSJ reached statistical significance after multiple testing correction (FDR > 0.05), we were able to replicate the direction of change of four BSJs in PPMI and three in ICICLE-PD (Fig. 3c and Supplementary Data 9).

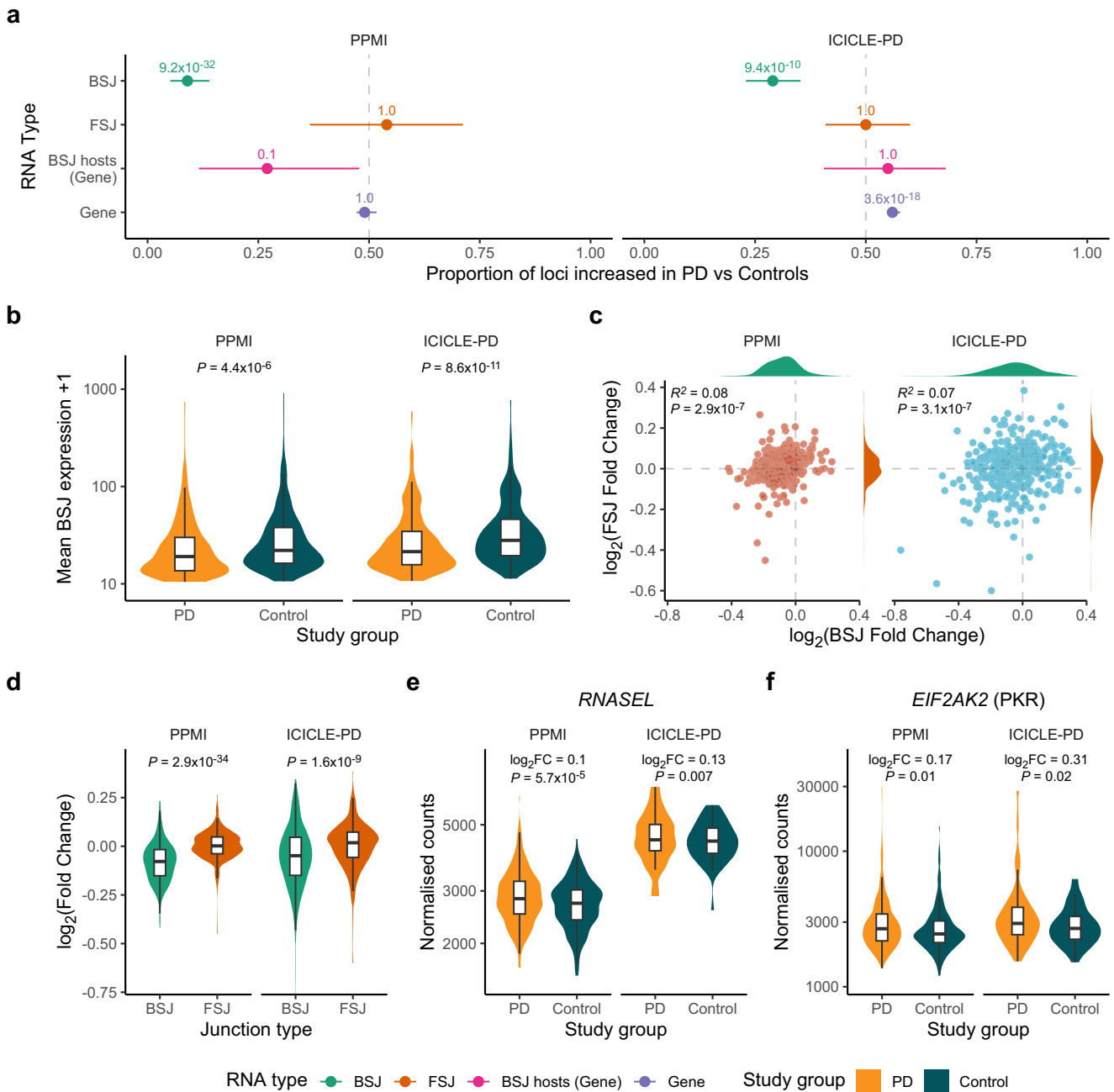


Fig. 4 Idiopathic PD patients exhibit a circular RNA-specific reduction in expression. **a** Proportion of loci increased in PD relative to controls based on differential expression testing of back-spliced junctions (BSJs), forward-spliced junctions (FSJs), genes which harbour a BSJ (BSJ hosts) and all genes (Gene). P -values obtained from a two-sided exact binomial test corrected for multiple testing (Bonferroni correction, four tests). Error bars show the 95% CI of the imbalance estimate. **b** Comparing the expression of all BSJs included in differential expression testing between PD and controls in PPMI and ICICLE-PD. P -values are derived from a Wilcoxon rank-sum test comparing BSJ expression between PD and controls. **c** Correlation between BSJ and FSJ expression fold changes in both PPMI and ICICLE-PD. The correlation coefficients are reported as Spearman's rho. Density plots show the distribution of the fold changes for each respective junction type. **d** Comparison of fold changes between BSJs and FSJs in PPMI and ICICLE-PD. P -values derived from a Wilcoxon rank-sum test comparing fold changes between each junction type. **e**, **f** Plots showing the normalised expression of the genes *RNASEL* and *EIF2AK2* (which encodes PKR) in PD and controls from PPMI and ICICLE-PD. \log_2 fold changes and P -values were obtained from the results of testing the differential expression of genes.

Early-stage idiopathic PD is characterised by a reduction in circRNA abundance

Whilst we observed differential expression of specific BSJs, perhaps the more striking observation was an imbalance of BSJ fold changes, i.e., where the expression of the majority of BSJs tended to be reduced in PD compared to controls (PPMI

imbalance = 0.09, Bonferroni P -value = 9.2×10^{-32} , ICICLE-PD imbalance = 0.29, Bonferroni P -value = 9.4×10^{-10} , Exact binomial test) (Fig. 4a, Supplementary Data 10). To explore whether reduced expression was specific to BSJs, we also examined the expression of other RNA types (Supplementary Fig. 7, Supplementary Data 10). Forward-spliced junction (FSJ) counts provide

the best approximation of the linear RNA as gene expression estimates will inevitably include circRNA-derived reads that do not span the BSJ and thus cannot be distinguished. Crucially, a reduction was not evident when examining the fold changes of FSJs (PPMI imbalance = 0.54, Bonferroni P -value = 1.0, ICICLE-PD imbalance = 0.50, Bonferroni P -value = 1.0, Exact binomial test) (Fig. 4a, Supplementary Data 10). A comparison of fold changes based only on genes that host circRNAs identified no reduction (PPMI imbalance = 0.27, Bonferroni P -value = 0.1, ICICLE-PD imbalance = 0.54, Bonferroni P -value = 1.0, Exact binomial test). Global gene expression also showed no reduction in PPMI (imbalance = 0.49, Bonferroni P -value = 1.0), yet was significantly increased in ICICLE-PD (imbalance = 0.56, Bonferroni P -value = 3.6×10^{-18}) (Fig. 4a, Supplementary Data 10).

As fold change imbalance estimates are dependent on the number of RNAs passing the fold change threshold, we also directly compared expression levels of all RNAs included in differential expression analyses. This also provided evidence of a significant reduction in circRNA expression in PD (PPMI P -value = 4.4×10^{-6} , ICICLE-PD P -value = 8.6×10^{-11} , Wilcoxon rank-sum test) (Fig. 4b), yet no reduction in the expression of flanking FSJs, expression of circRNA host genes or total gene expression (Supplementary Fig. 8).

To highlight differences in circRNA regulation, we compared fold changes between BSJ expression and their flanking FSJs. For a subset of junctions, this was not possible as no corresponding FSJs were detected (PPMI = 95/403 or 23.6%, ICICLE-PD = 92/457 or 20.1%). We observed a weak correlation in fold change differences between BSJs and corresponding FSJs (PPMI $R^2 = 0.08$, ICICLE-PD $R^2 = 0.07$, Fig. 4c). BSJ fold changes were consistently lower, albeit generally modest, across both cohorts (PPMI P -value = 2.9×10^{-34} , ICICLE-PD P -value = 1.6×10^{-9} , Wilcoxon rank-sum test) (Fig. 4d). Together, these results suggest that the observed imbalance in circRNA expression in early-stage idiopathic PD is not solely driven by a reduction in parental host gene expression.

***RNASEL* and *ADAR* expression is increased in early-stage idiopathic PD**

CircRNA biogenesis and degradation are linked to several *cis*- and *trans*-acting factors³⁷. We explored whether differential expression of previously reported circRNA regulators^{43–53} may be contributing to the reduced circRNA abundances in PD (regulators and evidence are given in Supplementary Data 11). In PPMI, *RNASEL* and *ADAR* were significantly increased in PD, with *RNASEL* also increased in ICICLE-PD and significant after multiple testing correction (Fig. 4e, Supplementary Data 11). *RNASEL* encodes Ribonuclease L, which is known to be involved in the degradation of circRNA, subsequently allowing the activation of PKR as part of the integrated stress response⁵⁰. We also observed significantly increased expression of *EIF2AK2*, which encodes PKR, in both cohorts (PPMI P -value = 0.010; ICICLE-PD P -value = 0.019) (Fig. 4f, Supplementary Data 1). Increased PKR activity has been reported in lymphocytes of PD patients⁷⁴. This, combined with the downregulation of protein synthesis genes and the upregulation of genes involved in the innate immune response (Fig. 2c, Supplementary Data 2), suggests that the antiviral integrated stress response may be active in early-stage idiopathic PD.

Evaluating RNA expression as a predictor of early-stage idiopathic PD

To evaluate the use of RNA expression as a predictor of early-stage idiopathic PD, we first examined the potential of BSJ and FSJ expression, as well as BSJ:FSJ expression ratios, to separate PD from controls. Area under the receiver operating characteristics curves (AUC) were generated using the expression of each abundant junction type as a predictor. The median AUC across all junction measures was higher in ICICLE-PD (0.57) compared to

PPMI (0.53). The strongest correlation between PPMI and ICICLE-PD AUCs was for BSJ:FSJ ratios (Spearman's $\rho = 0.33$), followed by FSJ (Spearman's $\rho = 0.22$) and BSJ counts (Spearman's $\rho = 0.18$) (Fig. 5a).

With no clear individual junction biomarker candidates, we included multiple features in multivariable regularised logistic regression models (see *Classification of Parkinson's status using RNA abundances*). Using cross-validation, we trained and evaluated the classification performance of models using the PPMI cohort. We then used the ICICLE-PD as an independent test cohort to assess the performance of classifiers trained using the PPMI cohort. The best performance in the PPMI cohort was achieved by combining gene and BSJ counts (AUC = 0.85 [0.81, 0.89], Fig. 5b) and appeared primarily driven by gene expression (AUC = 0.84 [0.81, 0.88], Fig. 5b) rather than BSJ counts (AUC = 0.61 [0.55, 0.66], Fig. 5b). In ICICLE-PD, combining gene and BSJ counts produced an AUC of 0.60 [0.48, 0.71] (Fig. 5b). Similarly, the performance of the gene expression strategy also decreased in the ICICLE-PD cohort (AUC = 0.59 [0.48, 0.71], Fig. 5b). All other classifiers had an AUC < 0.6 in the PPMI cohort (Fig. 5b, Supplementary Table 2). Using BSJ:FSJ ratio as a classifier achieved the best generalisability in the ICICLE-PD cohort (AUC = 0.63 [0.52, 0.74], Fig. 5b). All other unmentioned models achieved an AUC < 0.6 in the ICICLE-PD cohort (Fig. 5b, Supplementary Table 2).

DISCUSSION

We identified differences in gene and circRNA expression between early-stage idiopathic PD patients and controls, with elevated expression of *TMEM252* and *LMNB1* observed in both PPMI and ICICLE-PD PD cases compared to controls. We independently validated expression changes in genes shown to modulate PD risk¹⁰ and in genes that had been previously reported to be differentially expressed in PD^{70–73}. Additionally, we discovered three novel circRNAs that were decreased in PD, with similar trends observed in our replication cohort. After evaluating the performance of multivariable classification models, our results suggest that circRNAs are unlikely to be useful biomarkers for early-stage idiopathic PD, particularly when compared to classification using gene expression. Perhaps more interestingly, we observed a tendency towards generalised decreased circRNA expression in PD in both cohorts, which correlated alongside decreased expression of genes involved in protein translation and increased expression of genes involved in the immune response. These findings suggest that reduced circRNA levels in the blood may be a proxy measure of the underlying pathology and add to the growing body of evidence linking RNA metabolism and immune response to PD^{75,76}.

The global downregulation of circRNA expression we observed in blood is reminiscent of the reduction in microRNA expression seen in PPMI participants³³. CircRNA expression is influenced by *trans*-acting proteins, including Ribonuclease L (RNase L, encoded by *RNASEL*)⁵⁰ and adenosine deaminase RNA specific (*ADAR*, encoded by *ADAR*), an enzyme responsible for A-I RNA editing that can modulate the base-pairing of reverse complementary matches in the flanking introns of circRNAs, influencing their biogenesis^{53,54,77,78}. *ADAR* was significantly increased in PPMI PD cases and showed a similar trend in ICICLE-PD (Supplementary Data 11). Although *ADAR* expression levels do not necessarily correlate with editing activity⁷⁹, altered RNA editing has been observed in PD patients' blood and brain regions^{61,80}. RNase L degrades circRNAs as a response to viral infection⁵⁰. The degradation of circRNAs is required for the activation of Protein Kinase R (PKR, encoded by *EIF2AK2*), a regulator of the integrated stress response (ISR)^{50,81}. This is consistent with our results, where the reduction in circRNAs in PD we observed coincides with increased expression of *RNASEL* and *EIF2AK2* (Fig. 4e, f, Supplementary Data 1), as well as the upregulation of genes related to antiviral activity and the innate

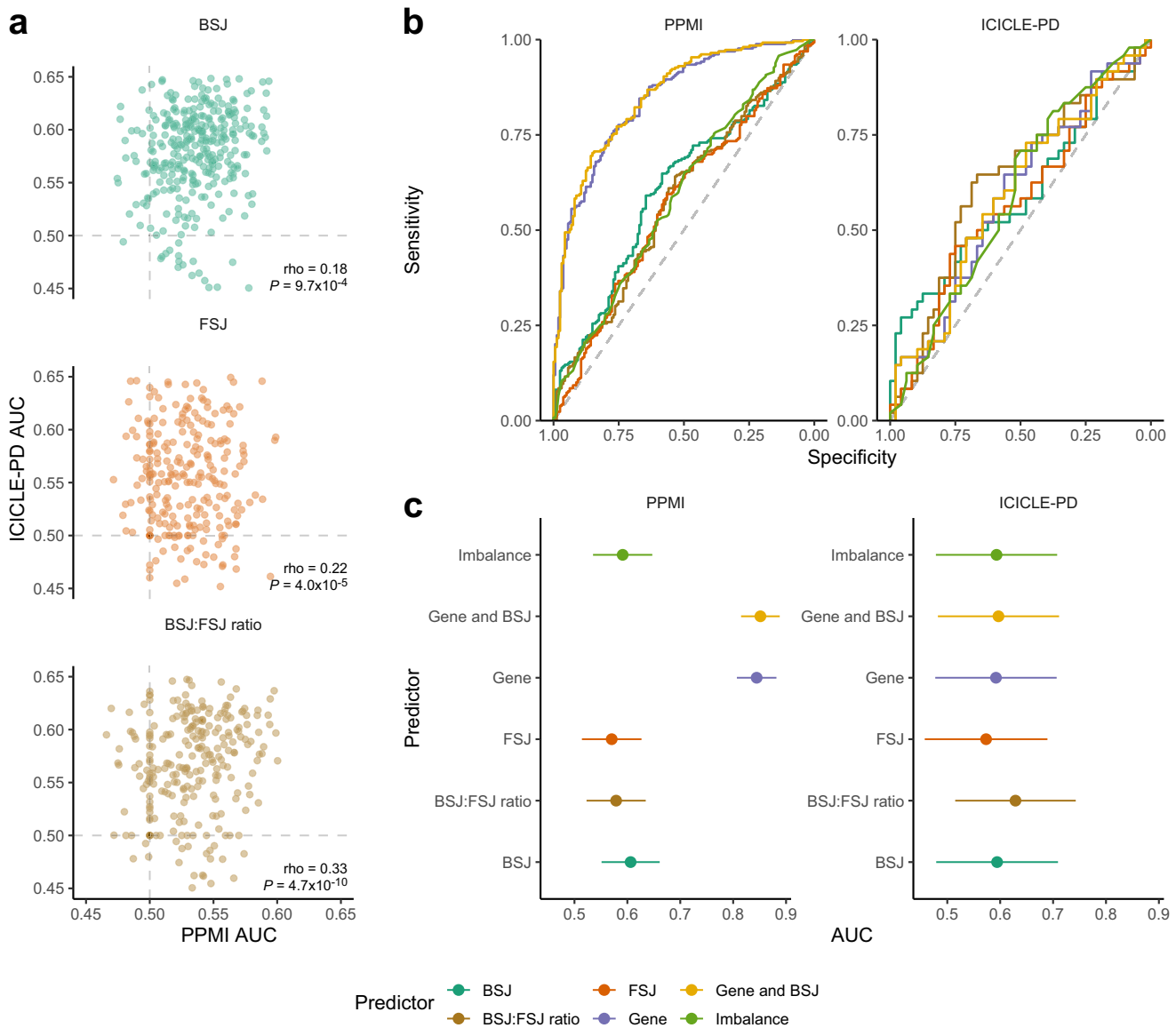


Fig. 5 Using RNA expression to classify early-stage idiopathic PD status. **a** Comparison of AUCs calculated in PPMI and ICICLE-PD cohorts using the expression of individual junctions (BSJ and FSJ) and BSJ:FSJ ratios to classify PD status. ROC curves were constructed using the expression of back-spliced junctions (BSJ, VST normalised), forward-spliced junctions (FSJs, VST normalised) and the circular to linear ratio (BSJ:FSJ ratio, calculated by CIRIquant) for each junction position. Spearman's rho demonstrates the correlation between the AUC in each cohort. Grey dashed lines indicate an AUC of 0.5. **b, c** Ability of each predictor to classify PD status. ROC curves (**b**) showing the sensitivity and specificity of each predictor (see methods) at various thresholds in PPMI and ICICLE-PD cohorts. The area under the ROC curve for each predictor is shown in **c**. Performance in the PPMI was evaluated on held-out data in the outer nested folds. Performance in ICICLE-PD was based on models fit on the PPMI data. Error bars give the 95% confidence intervals of the AUC.

immune response (Fig. 2c, Supplementary Data 2, 3). Although the role of RNase L in the antiviral immune response is well documented⁸², and despite the links between viral infection and PD⁸³, to our knowledge, there have been no reports showing *RNASEL* expression changes in PD. Conversely, PKR activation has been observed in PD^{74,84}. Endogenous dsRNA can lead to PKR activation^{85,86} and has been recently implicated in neurodegenerative diseases other than PD^{87–89}. In mice, exogenous dsRNA triggered α -synuclein aggregation and dopaminergic neuron loss, suggesting the existence of aberrant dsRNA may have a potential role in the pathogenesis of PD⁹⁰. Activated PKR leads to eIF2 α phosphorylation, and results in global protein synthesis attenuation⁸¹. Protein synthesis is reduced in PD patient cell lines^{91,92}, consistent with the reduced expression of genes related to protein

synthesis and ribosomal function we observed in PD (Fig. 2c, Supplementary Data 2, 3). Based on our findings, we propose that reduced circRNA levels in early-stage idiopathic PD is related to activation of an antiviral immune response, reflected through increased expression of *RNASEL*, *ADAR* and *EIF2AK2* (which encodes PKR) (Fig. 6). Reduced circRNA expression would allow activation of PKR, consistent with the increased activity of PKR in lymphocytes of PD patients⁷⁴, triggering the cellular ISR. Overall, these results highlight a potential role of RNase L and PKR in the pathogenesis and/or systemic response to PD (Fig. 6).

Using gene expression as a classifier of early-stage PD in the PPMI cohort was in line with previously published estimates (AUC = 0.84 compared to 0.80 in Makariou et al.⁹³, Fig. 5c). In our independent replication cohort (ICICLE-PD), the AUC of the gene

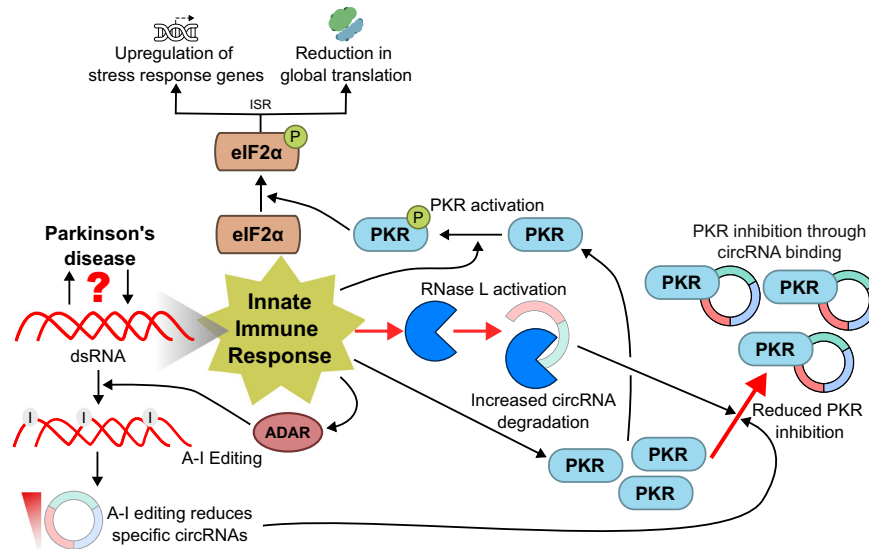


Fig. 6 A mechanism of circRNA reduction in Parkinson's disease. Innate immune responses can be induced by double-stranded RNA (dsRNA). A-I RNA editing by ADAR suppresses immune detection of endogenous dsRNA¹⁴⁹ and influences circRNA synthesis^{53,54,77,78}. Under normal conditions, circRNAs bound to PKR (encoded by *EIF2AK2*) act as inhibitors, undergoing degradation when activation is required⁵⁰. Activation of the endonuclease RNase L as part of an antiviral immune response leads to degradation of circRNAs⁵⁰. Decreased circRNAs in idiopathic PD could result in increased PKR, available for activation. Once activated, PKR phosphorylates eIF2 α , initiating the integrated stress response (ISR)⁸¹.

expression classifier was lower (ICICLE-PD AUC = 0.59, Fig. 5c). Despite promising results using circRNA as a biomarker in other diseases⁵⁹, circRNAs performed substantially worse than gene expression when classifying PD in PPMI (AUC = 0.61 vs 0.84, Fig. 5c), although similar performance was seen in ICICLE-PD (AUC = 0.59 vs 0.59, Fig. 5c). Similar to previous work assessing circRNAs as a lung cancer biomarker⁹⁴, combining circRNA expression with gene expression showed a small improvement in predictive ability over gene expression alone in both PPMI (AUC = 0.85 vs 0.84, Fig. 5c) and ICICLE-PD (AUC = 0.60 vs 0.59, Fig. 5c). Put into context, circRNAs perform worse than their linear counterparts when classifying PD as a group, leading us to the conclusion that circRNA expression in isolation has limited utility as a clinical PD biomarker. Although combining circRNAs and gene expression data does improve classification power, the performance of circRNAs either in isolation or combined with gene expression falls short of recent estimates using alpha-synuclein seeding as a biomarker¹⁷.

When comparing differential gene expression, we identified two genes (*TMEM252* and *LMNB1*) that were upregulated in PD cases in both cohorts, consistent with previous work²⁷. *TMEM252* encodes a transmembrane protein of unknown function that has been linked to cancer⁹⁵, but has not previously been implicated in PD. *LMNB1* encodes Lamin B1, a component of the nuclear lamina. Increased *LMNB1* expression has been reported in the dopaminergic neurons of PD patients⁹⁶ and a variant within *LMNB1* has been associated with cognitive outcomes in PD⁹⁷.

We were also able to independently replicate the differential expression of genes that have been previously reported in blood RNA-seq studies as differentially expressed in PD patients^{70–73} (*LSMEM1*, *TPST1* and *SLED1*, Fig. 2d). As previous studies were not limited to early-stage PD (diagnosed <13 months), these genes may reflect those with altered expression throughout PD development. Of the three genes replicated using PPMI, *SLED1* and *LSMEM1* encode proteins of unknown function. *TPST1* encodes tyrosylprotein sulfotransferase 1, an enzyme required for post-translational modification of proteins that also plays important roles in the inflammatory process, leukocyte movement and cytoskeleton, viral cell entrance, and other cell-cell and protein-

protein interactions⁹⁸. Interestingly, 3 out of the 4 previously reported genes replicated in ICICLE-PD (*IFIT1*, *RSAD2*, *IFI44I*, Fig. 2d) are induced by type-1 interferons⁹⁹. Their increase in expression is consistent with an increase in type-1 interferon signalling in PD¹⁰⁰. In addition, our data provide a potential mechanistic link between PD-associated genetic variation, identifying differential expression among genes associated with PD risk. These included *PTRHD1* (Supplementary Data 6), which encodes a peptidyl-tRNA hydrolase that has been linked to recessive parkinsonism^{101,102}, and genes identified by association studies (*BST1*, *FCGR2A*, *SIPA1L2*, *NOD2* and *VAMP4*, Fig. 2e)¹⁰.

Our analysis of circRNAs identified three that were over-expressed in PPMI PD cases compared to controls (within the genes *BMS1P1*, *CCDC9* and *ESYT2*) with similar trends observed in ICICLE-PD (Fig. 2a, b). These circRNAs, or their host genes, have not previously been associated with PD. The protein product of *CCDC9* is believed to be a member of the exon junction complex involved in RNA splicing¹⁰³. CircRNAs from *CCDC9* have been implicated in cancer, acting as a miRNA sponge to suppress tumorigenesis¹⁰⁴, and in stroke, suppressing NOTCH signalling in mouse models of ischaemia¹⁰⁵. Interestingly, alterations in NOTCH signalling has been linked to PD through the function of *LRRK2*¹⁰⁶. *ESYT2* encodes extended synaptotagmin 2, a member of the E-Syt family, which are endoplasmic reticulum (ER) localised proteins involved in tethering the ER to the cellular plasma membrane¹⁰⁷. There is suggestive evidence that circRNAs derived from *ESYT2* may be upregulated upon viral infection¹⁰⁸, in contrast to the decreased expression we observed in PD patients (Fig. 2a, b). *BMS1P1* encodes a pseudogene of ribosome biogenesis factor pseudogene 1 (BMS1), and there are no known disease associations for circRNAs produced by this gene.

We were also able to independently replicate trends in several previously reported differentially expressed circRNAs observed in PD blood and brain tissue^{61,62} although none reached statistical significance (Fig. 3c, Supplementary Data 9). This is likely due to methodological differences (e.g., circRNA detection and normalisation), tissue-specific expression (whole blood compared to the brain) or differences in blood cell composition (whole blood compared to peripheral blood mononuclear cells). Unfortunately,

23 of the 30 previously reported circRNAs (20 from the substantia nigra and three detected in peripheral blood mononuclear cells) were undetectable in either PPMI or ICICLE-PD. We identified altered expression of circRNAs (derived from *DOP1B* and *INTS6L*), whose host genes have not been functionally linked to PD (Supplementary Data 9). However, copy number variation in *DOP1B* has previously been linked to Alzheimer's disease¹⁰⁹, while deletion of *INTS6L* leads to a cardiomyopathic phenotype¹¹⁰. Nonetheless, our data extend the growing body of evidence linking circRNA dysregulation to PD⁶⁰.

As a key vehicle for immune cells, the differences we observe in blood RNA expression in PD patients add to the proposed role of inflammation and immune dysfunction in the development and response to PD⁷⁶. It is possible that changes in blood can be detected early on in disease duration, in line with the early presentation of non-motor symptoms and peripheral aggregations of α -Synuclein¹¹¹. Pre-diagnosis changes in lymphocyte levels and function have been reported^{112,113}. In addition, inflammation markers within blood have been associated extensively with both PD risk and symptom progression¹¹⁴.

In this work, we have drawn together and performed the largest whole-blood circular transcriptome analysis of PD, focusing on the potential utility of circRNAs as biomarkers. However, several limitations must be considered. Biomarker detection using whole-blood has numerous advantages, including accessibility and ease of processing, but we could not account for variation in the proportions of different blood cell types across samples^{115,116}. In addition, despite efforts to increase homogeneity (see *Cohorts*), both cohorts differ in some respects. Firstly, PPMI participants were not receiving treatment at sample collection, whereas most ICICLE-PD participants were. Dopaminergic treatment may influence biological measures, such as circulating cell-free mitochondrial DNA levels¹¹⁷. It must also be noted that some PPMI samples were recruited during a treatment interval (up to ~60 days before enrolment) and a large number of individuals were taking concomitant medications for a range of non-PD and PD-related symptoms⁶⁶. This may have influenced our results, however, the concordances we see between RNA expression as well as the reproducible reduction in circRNA expression in PD patients between cohorts suggests this is minimal. Secondly, it is inevitable that the difference between cohort sizes affected our ability to detect rarer RNAs, particularly circRNAs. We attempted to mitigate against this by using the largest cohort as discovery (PPMI) opting to replicate in the smaller cohort (ICICLE-PD). In addition, we limited our analysis to abundant genes and circular RNAs, with the majority (95.53% of genes and 62.57% of circRNAs) detectable in both cohorts. Finally, some circRNA quantification studies use circRNA enrichment steps such as RNase R treatment¹¹⁸ to enhance the detection of lesser expressed circRNAs. However, not depleting linear RNAs allowed us to quantify both linear and circular RNAs simultaneously.

In conclusion, we observed specific and consistent alterations in the linear and circular blood transcriptome in early-stage idiopathic PD patients. Changes in circRNA levels were not sufficient to facilitate reliable PD classification, particularly when compared to existing PD biomarkers. We did, however, identify a reproducible reduction in circRNA expression in PD. This imbalance, along with gene expression patterns, implicates the activation of an innate antiviral immune response, providing an opportunity for future investigations into this previously unknown aspect of circRNA regulation in PD.

METHODS

Cohorts

We utilised samples with corresponding demographic (e.g., age at sample collection, sex) and clinical data (e.g., dopaminergic

treatment status, disease duration) from two large cohorts of patients with Parkinson's disease (PD) and controls of similar ages and sex. The discovery cohort was obtained from The Michael J Fox Foundation Parkinson's Progression Markers Initiative⁶⁶ (PPMI, <https://www.ppmi-info.org/>), while The Incidence of Cognitive Impairment in Cohorts with Longitudinal Evaluation-PD⁶⁷ (ICICLE-PD, <https://www.bam-ncl.co.uk/iciclepd>) was used for replication. To study transcriptomic changes in the early stages of PD and to ensure parity between discovery and replication cohorts, only PD patients recently diagnosed with PD (<13 months) were included. PD patients harbouring causative variants in select genes (e.g., *LRRK2*, *GBA*, *SNCA*, *PINK1*, *PRKN*) were excluded and are thus all PD cases are idiopathic. Both studies were conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines after approval of local ethics committees of the participating sites (PPMI, <https://www.ppmi-info.org/> and ICICLE-PD, Newcastle and North Tyneside Research Ethics Committee)^{66,67}. All subjects provided written informed consent.

ICICLE-PD RNA isolation and sequencing

Between June 2009 and December 2011, newly diagnosed PD patients were recruited from the community and hospital outpatient clinics in Newcastle-upon-Tyne and Gateshead. Idiopathic PD was diagnosed by a movement disorder specialist and fulfilled Queen's Square Brain Bank criteria¹¹⁹. Full exclusion criteria have been published elsewhere⁶⁷. Briefly, participants were excluded if they had significant cognitive impairment at presentation (Mini Mental State Examination <24) or a pre-existing diagnosis of dementia, an atypical Parkinsonian syndrome, or insufficient English to complete assessments. We limited our analysis to 48 early-stage (diagnosed <13 months) idiopathic (i.e., without a known PD-related genetic diagnosis) PD patients and 48 control samples matched for age and sex. Total RNA was extracted from 5 ml of whole blood collected and stored in PAXgene tubes as per manufacturer's instructions (Blood RNA Kit, Qiagen), quality assessed using an Agilent 2100 Bioanalyzer system and stored at -80°C . Only samples with an RNA integrity number >8 were included. Library preparation, including ribosomal RNA and globin depletion (Globin-Zero Gold rRNA Removal Kit, Illumina), was carried out using the TruSeq Stranded Total RNA kit (Illumina). Sequencing was performed using an Illumina NovaSeq 6000 generating 150 bp paired-end reads. Median sequencing depth was estimated as 89.2 (IQR = 15.7) million paired-end reads (Supplementary Fig. 1).

PPMI RNA isolation and sequencing

Subject recruitment and eligibility criteria for the PPMI study have been previously published⁶⁶. To increase matching between ICICLE-PD, we limited our analysis to 287 early-stage (diagnosed <13 months) idiopathic (i.e., without a known PD-related genetic diagnosis) PD patients and 176 controls, additionally matching within PPMI for age and sex. Whole blood RNAseq data was downloaded as FASTQ files from the online PPMI repository (<https://www.ppmi-info.org/>). RNA collection, isolation and subsequent sequencing have been previously reported²⁷ and is described in the PPMI Biologicals Manual. Briefly, whole-blood RNA was extracted from PAXgene tubes (Blood RNA Kit, Qiagen), rRNA and globin depleted (Globin-Zero Gold rRNA Removal Kit, Illumina). Library preparation used the NEB/Kapa (NEBKAP) kit (see ref. ²⁷ for more information). Sequencing was performed using an Illumina NovaSeq 6000 generating 125–150 bp paired-end reads. Median sequencing depth was estimated as 107 (IQR = 31.2) million paired-end reads (Supplementary Figure 1).

RNA alignment and quantification

FASTQ files were aligned to the human genome (Ensembl GRCh38¹²⁰) using HISAT2 v2.1.0¹²¹. RNA transcripts were quantified using the Ensembl GRCh38 v101 cDNA reference and Salmon v1.3.0¹²² using the 'selective alignment mode' option and correcting for sequence-specific (*--seqBias*), GC-content (*--gcBias*) and positional biases (*--posBias*). Transcript counts were translated into gene-level counts using tximport v1.26.1¹²³ and annotated using the Ensembl v101 GRCh38 reference.

Circular RNA detection and quantification

Reads derived from circular RNA (circRNA) molecules were identified by the presence of a back-spliced junction (BSJ, Fig. 1) detected using three tools: CIRI2 v2.0.6¹²⁴, PTESfinder v2.0¹²⁵ (<https://github.com/osagie/pfv2>) and CIRCexplorer v2.3.8¹²⁶. First, FASTQ files were quality (Phred <15) and then adapter trimmed using Trim Galore v0.67 running cutadapt v4.2¹²⁷. Reads were aligned to the human genome (Ensembl GRCh38) according to each tool's specification: STAR v.2.7.10a¹²⁸ for CIRCexplorer2 and PTESfinder v2, bowtie v2.3.4¹²⁹ for PTESfinder v2, and BWA v0.7.17¹³⁰ for CIRI2. All BSJs were annotated using the Ensembl v101 GRCh38 reference.

Similar to others⁵⁶, we limited false positives by retaining BSJs with a read count >1 in at least two individuals. In addition, BSJs were only retained for downstream analysis if they were detected by at least two tools (Supplementary Figure 2). This reduced our initial set of BSJs (PPMI = 438,189 to 23,454, or 5.35%; ICICLE-PD = 222,133 to 15,345, or 6.91%). These BSJs, alongside the corresponding forward-spliced junctions (FSJs, Fig. 1), were quantified with CIRIquant v1.1.2¹³¹ using HISAT v2.2.0¹²¹ as the aligner.

Sample level quality control

Sample level quality control was assessed in both cohorts to identify sample failure, RNA contamination and abnormal global transcriptome issues. Fastq Screen v0.14.1¹³² and FastQC v0.11.7 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) were used to assess contamination and obtain general sequencing metrics. Alignment metrics were obtained using *CollectRnaSeqMetrics* from Picard v2.27.5 (<http://broadinstitute.github.io/picard/>) and *stats* from SAMtools v1.6¹³³. Additionally, PPMI samples that had been previously flagged due to QC issues were removed²⁷. Validation of each participant's clinically recorded sex was based on the normalised and variance-stabilising transformed¹³⁴ expression of the Y chromosomal genes *RPS4Y1*, *KDM5D*, *DDX3Y* and *USP9Y*. Principal component analysis was carried out and the first two principal components and was used to detect mismatches by visual inspection (Supplementary Fig. 3). This identified one incorrectly coded individual in ICICLE-PD, which was corrected for analysis. After QC, the final PPMI dataset comprised of 259 PD and 161 controls; the ICICLE-PD dataset comprised of 48 PD patients and 48 controls (summarised in Supplementary Table 1). We observed no significant differences between the age or sex profiles of cases and controls in either dataset (Supplementary Table 1).

Identification of sources of variation in RNA expression data

Biological and technical factors are known to impact the quantification of gene expression^{135,136}. Like previous large-scale transcriptomic studies, we quantified sources of expression variation at both the sample and RNA level^{27,137,138}. At the sample level, we used univariate linear regression to identify technical sequencing metrics (obtained from Picard and SAMtools) that explained a high proportion of the variance ($R^2 > 0.5$) associated with the first 10 principal components of gene and circRNA expression (Supplementary Figs. 4a, b and 5a, b). At the gene level,

based on recommendations¹³⁹, we excluded highly correlated factors (Spearman's $\rho > 0.9$, Supplementary Figs. 4c and 5c), subsequently quantifying the contribution of this reduced set of covariates to gene and circRNA expression variation using variancePartition v1.28.3 (Supplementary Figure 6). This final set of cohort-specific covariates were then included in regression modelling (see *Differential expression*).

Differential expression

Normalisation and differential expression analysis of genes (i.e., linear mRNA) and BSJs was carried out using DESeq2 v1.38.2¹⁴⁰. Gene and BSJ raw counts were filtered for low counts (>10 counts in the smallest sample group). Gene counts were library normalised using the default median-of-ratios method in DESeq2 (Anders and Huber, 2010). Junction counts (both BSJs and FSJs) were normalised using the sample size factors generated during the normalisation of gene counts.

Differential expression between PD and controls was assessed using a Wald test in DESeq2, adjusted for specific technical and biological covariates relevant to each cohort and RNA type (see *Identification of sources of variation in gene expression data*). Across all cohorts and RNA types, we adjusted for sex, age of collection and sequencing batch. For gene differential expression, we adjusted for the percentage of usable bases (PPMI), in addition to the percentage of coding and intronic bases (ICICLE-PD). For circRNA differential expression, we adjusted for the percentage of intronic bases (PPMI and ICICLE-PD), in addition to the median Coefficient of Variance of transcript coverage (ICICLE-PD). *P*-values were adjusted for multiple significance by the Benjamini-Hochberg procedure at a false discovery rate (FDR) of 5%¹⁴¹. Similar to previous work²⁷, significantly differentially expressed genes and BSJs were defined as those that had an FDR < 0.05 and a \log_2 fold change (\log_2FC) < -0.1 / > 0.1.

Ontology and pathway analysis

Enrichment of gene sets was performed using clusterProfiler v4.6.2¹⁴². Gene set enrichment analysis (GSEA)⁶⁸ was based on gene lists ranked by fold change. For each cohort, all genes included in differential expression testing were used as the background set. Gene sets were obtained as Gene Ontologies¹⁴³ (including biological processes, molecular functions, and cellular compartments) and KEGG pathways¹⁴⁴ using the *gseGO* and *gseKEGG* functions, respectively.

GSEA is not suitable for circRNA enrichment at the junction level as each gene can host multiple circRNAs. As such, we carried out circRNA enrichment using an over-representation test of the host genes using the *enrichGO* function from clusterProfiler. To detect groups enriched in common BSJs, those included in differential expression testing were compared against all abundant BSJs detected in that cohort (>10 reads in the smallest sample group). To detect categories enriched in differentially expressed BSJs, those that were significant were compared to all BSJs included in differential expression testing. For all enrichment analyses, where possible, categories that passed the multiple testing threshold in the PPMI (FDR < 0.05) were validated in the ICICLE-PD cohort.

Relating RNA expression to previous work

Genes reported to be differentially expressed in PD were taken from previously published blood RNA sequencing studies⁷⁰⁻⁷³, resulting in a total of 354 genes. GWAS loci associated with PD risk were obtained from the current largest GWAS meta-analysis¹⁰. Expression of the nearest genes to these loci was then compared with 70 detected in PPMI and 69 detected in ICICLE-PD. Genes that host established pathogenic parkinsonism variants were obtained from the Genomics England Parkinson Disease and Complex Parkinsonism panel v1.111 (accessed 10/22)¹⁴⁵. Out of 35 genes,

30 were detected at sufficient expression levels in the PPMI and ICICLE-PD cohorts.

Classification of Parkinson's disease status using RNA expression

The following features were used for classification: normalised counts (gene, BSJ and FSJ) counts after Variance Stabilising Transformation as implemented in DESeq2¹³⁴. Circular to linear (BSJ:FSJ) ratios were extracted from CIRIquant¹³¹. We also explored the use of circRNA differential expression imbalance as a classification feature. The principle was to compare the expression value of each BSJ from a sample we want to classify with the average expression value in controls, subsequently assessing whether the directional deviation from the mean (i.e., over- or under-expression) was consistent with the direction observed in the training group (PPMI). Each BSJ was scored as 1 when the directions agreed or as 0 when the directions disagreed. BSJ scores were summed per sample and the ability of this sum to discriminate between cases and controls was assessed.

For gene predictors, we selected those that were nominally significantly differentially expressed (P -value < 0.05). For BSJ, FSJ, BSJ:FSJ ratio and BSJ imbalance models, we used all highly expressed junctions. We investigated the following sets of features: Expression levels by gene (Gene), BSJ counts (BSJ), circular to linear ratios (BSJ:FSJ ratio), circular RNA imbalance (BSJ imbalance), expression levels by gene and BSJ counts (Gene and BSJ). With the exception of the BSJ imbalance (see above), the ability of the selected sets of features to classify cases and controls was assessed using regularised logistic regression as implemented in *glmnet* v4.1.6¹⁴⁶ and quantified using the area under the receiver operating characteristic curve (AUC). AUCs and their 95% confidence intervals (DeLong's method) were calculated using pROC v1.18.0¹⁴⁷. Classifiers were derived from the PPMI cohort, using the *nestcv.glmnet* function from *nestcdv* v0.4.4¹⁴⁸, and 10 fold cross-validation for both inner and outer cross-validation steps. Final model parameters were derived from a cross-validation fold fit on PPMI data. Model generalisability was assessed in the ICICLE-PD cohort.

Statistical analyses

All statistical analyses were carried out in R v4.2.1. Correlations were assessed by Spearman's rank correlation (*cor.test*), reported in the text as Spearman's rho (ρ). We use the term imbalance to describe whether there is an excess of loci (genes or junctions depending on the context) that are overexpressed in PD compared to controls. It describes the number of features showing \log_2 fold change >0.1 divided by the number of loci with \log_2 fold change >0.1 or <-0.1. Significance was assessed using a two-sided exact binomial test (*binom.test*). Group differences were assessed using a Wilcoxon rank-sum test (*wilcox.test*). Where appropriate, multiple testing correction was performed using Benjamini-Hochberg or Bonferroni corrections (*p.adjust*).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

PPMI raw RNA sequencing and corresponding clinical data are available from <https://www.ppmi-info.org/>. Due to Ethics, ICICLE-PD raw RNA sequence data is stored locally (<https://data.ncl.ac.uk/>), but is available for non-commercial, collaborative use upon reasonable request. Summary data used to generate summary statistics and figures are included as supplementary datasets. Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

CODE AVAILABILITY

The code used in this analysis is available at <https://github.com/bj-w/PPD-lin-circ-RNA-paper>.

Received: 25 July 2023; Accepted: 8 January 2024;

Published online: 20 January 2024

REFERENCES

- Poewe, W. et al. Parkinson disease. *Nat. Rev. Dis. Prim.* **3**, 1–21 (2017).
- Pont-Sunyer, C. et al. The Onset of Nonmotor Symptoms in Parkinson's disease (The ONSET PD Study). *Mov. Disord.* **30**, 229–237 (2015).
- Rizzo, G. et al. Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis. *Neurology* **86**, 566–576 (2016).
- Vijjaratnam, N., Simuni, T., Bandmann, O., Morris, H. R. & Foltynie, T. Progress towards therapies for disease modification in Parkinson's disease. *Lancet Neurol.* **20**, 559–572 (2021).
- Bloem, B. R., Okun, M. S. & Klein, C. Parkinson's disease. *Lancet* **397**, 2284–2303 (2021).
- Chen-Plotkin, A. S. et al. Finding useful biomarkers for Parkinson's disease. *Sci. Transl. Med.* **10**, eam6003 (2018).
- Li, S. & Le, W. Biomarker Discovery in Parkinson's Disease: Present Challenges and Future Opportunities. *Neurosci. Bull.* **33**, 481–482 (2017).
- 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).
- Karimi-Moghadam, A., Charsouei, S., Bell, B. & Jabalameli, M. R. Parkinson Disease from Mendelian Forms to Genetic Susceptibility: New Molecular Insights into the Neurodegeneration Process. *Cell. Mol. Neurobiol.* **38**, 1153–1178 (2018).
- 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. *Lancet Neurol.* **18**, 1091–1102 (2019).
- Pagano, G., Niccolini, F. & Politis, M. Imaging in Parkinson's disease. *Clin. Med.* **16**, 371–375 (2016).
- Bidesi, N. S. R., Vang Andersen, I., Windhorst, A. D., Shalgunov, V. & Herth, M. M. The role of neuroimaging in Parkinson's disease. *J. Neurochem.* **159**, 660–689 (2021).
- Depierreux, F. et al. Parkinson's disease multimodal imaging: F-DOPA PET, neuromelanin-sensitive and quantitative iron-sensitive MRI. *Npj Park. Dis.* **7**, 1–10 (2021).
- Froneczek, R. et al. Hypocretin (orexin) loss in Parkinson's disease. *Brain* **130**, 1577–1585 (2007).
- Burté, F. et al. Metabolic profiling of Parkinson's disease and mild cognitive impairment. *Mov. Disord. J. Mov. Disord. Soc.* **32**, 927–932 (2017).
- Havelund, J. F., Heegaard, N. H. H., Færgeman, N. J. K. & Gramsbergen, J. B. Biomarker Research in Parkinson's Disease Using Metabolite Profiling. *Metabolites* **7**, 42 (2017).
- Siderowf, A. et al. Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using α -synuclein seed amplification: a cross-sectional study. *Lancet Neurol.* **22**, 407–417 (2023).
- Liew, C.-C., Ma, J., Tang, H.-C., Zheng, R. & Dempsey, A. A. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J. Lab. Clin. Med.* **147**, 126–132 (2006).
- Qi, T. et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282 (2018).
- Scherzer, C. R. et al. Molecular markers of early Parkinson's disease based on gene expression in blood. *Proc. Natl Acad. Sci.* **104**, 955–960 (2007).
- Shehadeh, L. A. et al. SRRM2, a potential blood biomarker revealing high alternative splicing in Parkinson's disease. *PLoS One* **5**, e9104 (2010).
- Santiago, J. A., Littlefield, A. M. & Potashkin, J. A. Integrative transcriptomic meta-analysis of Parkinson's disease and depression identifies NAMPT as a potential blood biomarker for de novo Parkinson's disease. *Sci. Rep.* **6**, 34579 (2016).
- Shamir, R. et al. Analysis of blood-based gene expression in idiopathic Parkinson disease. *Neurology* **89**, 1676–1683 (2017).
- Santiago, J. A., Bottero, V. & Potashkin, J. A. Evaluation of RNA Blood Biomarkers in the Parkinson's Disease Biomarkers Program. *Front. Aging Neurosci.* **10**, 157 (2018).
- Jiang, F., Wu, Q., Sun, S., Bi, G. & Guo, L. Identification of potential diagnostic biomarkers for Parkinson's disease. *FEBS Open Bio* **9**, 1460–1468 (2019).
- Falchetti, M., Prediger, R. D. & Zanotto-Filho, A. Classification algorithms applied to blood-based transcriptome meta-analysis to predict idiopathic Parkinson's disease. *Comput. Biol. Med.* **124**, 103925 (2020).
- Craig, D. W. et al. RNA sequencing of whole blood reveals early alterations in immune cells and gene expression in Parkinson's disease. *Nat. Aging* **1**, 734–747 (2021).

28. 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**, 464–472 (2021).
29. Manna, I., Quattrone, A., De Benedittis, S., Iaccino, E. & Quattrone, A. Roles of Non-Coding RNAs as Novel Diagnostic Biomarkers in Parkinson's Disease. *J. Park. Dis.* **11**, 1475–1489 (2021).
30. Arshad, A. R. et al. MicroRNAs and Target Genes As Biomarkers for the Diagnosis of Early Onset of Parkinson Disease. *Front. Mol. Neurosci.* **10**, 352 (2017).
31. Dos Santos, M. C. T. et al. miRNA-based signatures in cerebrospinal fluid as potential diagnostic tools for early stage Parkinson's disease. *Oncotarget* **9**, 17455–17465 (2018).
32. Leggio, L. et al. microRNAs in Parkinson's Disease: From Pathogenesis to Novel Diagnostic and Therapeutic Approaches. *Int. J. Mol. Sci.* **18**, 2698 (2017).
33. Kern, F. et al. Deep sequencing of sncRNAs reveals hallmarks and regulatory modules of the transcriptome during Parkinson's disease progression. *Nat. Aging* **1**, 309–322 (2021).
34. Chikina, M. D. & Sealfon, S. C. Increasing Consistency of Disease Biomarker Prediction Across Datasets. *PLOS ONE* **9**, e91272 (2014).
35. Tönges, L. et al. Blood-based biomarker in Parkinson's disease: potential for future applications in clinical research and practice. *J. Neural Transm.* **129**, 1201–1217 (2022).
36. Chahine, L. M., Stern, M. B. & Chen-Plotkin, A. Blood-based biomarkers for Parkinson's disease. *Parkinsonism Relat. Disord.* **20**, S99–S103 (2014).
37. Kristensen, L. S. et al. The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* **20**, 675–691 (2019).
38. Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L. & Brown, P. O. Cell-Type Specific Features of Circular RNA Expression. *PLOS Genet.* **9**, e1003777 (2013).
39. Xia, S. et al. Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief. Bioinform.* **18**, 984–992 (2017).
40. Izuogu, O. G. et al. Analysis of human ES cell differentiation establishes that the dominant isoforms of the lncRNAs RMST and FIRRE are circular. *BMC Genomics* **19**, 276 (2018).
41. Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PloS One* **7**, e30733 (2012).
42. Vo, J. N. et al. The Landscape of Circular RNA in Cancer. *Cell* **176**, 869–881.e13 (2019).
43. Ashwal-Fluss, R. et al. circRNA Biogenesis Competes with Pre-mRNA Splicing. *Mol. Cell* **56**, 55–66 (2014).
44. Conn, S. J. et al. The RNA Binding Protein Quaking Regulates Formation of circRNAs. *Cell* **160**, 1125–1134 (2015).
45. Aktaş, T. et al. DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* **544**, 115–119 (2017).
46. Errichelli, L. et al. FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat. Commun.* **8**, 14741 (2017).
47. Li, X. et al. Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection. *Mol. Cell* **67**, 214–227.e7 (2017).
48. Di Liddo, A. et al. A combined computational pipeline to detect circular RNAs in human cancer cells under hypoxic stress. *J. Mol. Cell Biol.* **11**, 829–844 (2019).
49. Jia, R., Xiao, M.-S., Li, Z., Shan, G. & Huang, C. Defining an evolutionarily conserved role of GW182 in circular RNA degradation. *Cell Discov.* **5**, 1–4 (2019).
50. Liu, C.-X. et al. Structure and Degradation of Circular RNAs Regulate PKR Activation in Innate Immunity. *Cell* **177**, 865–880.e21 (2019).
51. Li, X., Ding, J., Wang, X., Cheng, Z. & Zhu, Q. NUDT21 regulates circRNA cyclization and ceRNA crosstalk in hepatocellular carcinoma. *Oncogene* **39**, 891–904 (2020).
52. Knupp, D., Cooper, D. A., Saito, Y., Darnell, R. B. & Miura, P. NOVA2 regulates neural circRNA biogenesis. *Nucleic Acids Res.* **49**, 6849–6862 (2021).
53. Shen, H. et al. ADARs act as potent regulators of circular transcriptome in cancer. *Nat. Commun.* **13**, 1508 (2022).
54. Rybak-Wolf, A. et al. Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol. Cell* **58**, 870–885 (2015).
55. You, X. et al. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat. Neurosci.* **18**, 603–610 (2015).
56. Memczak, S., Papavasileiou, P., Peters, O. & Rajewsky, N. Identification and Characterization of Circular RNAs As a New Class of Putative Biomarkers in Human Blood. *PLOS ONE* **10**, e0141214 (2015).
57. Jeck, W. R. et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA N. Y. N.* **19**, 141–157 (2013).
58. Eruka, Y. et al. Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. *Nucleic Acids Res.* **44**, 1370–1383 (2016).
59. Verduci, L., Tarcitano, E., Strano, S., Yarden, Y. & Blandino, G. CircRNAs: role in human diseases and potential use as biomarkers. *Cell Death Dis.* **12**, 1–12 (2021).
60. Doxakis, E. Insights into the multifaceted role of circular RNAs: implications for Parkinson's disease pathogenesis and diagnosis. *NPJ Park. Dis.* **8**, 7 (2022).
61. Hanan, M. et al. A Parkinson's disease CircRNAs Resource reveals a link between circSLC8A1 and oxidative stress. *EMBO Mol. Med.* **12**, e11942 (2020).
62. Ravanidis, S. et al. Differentially Expressed Circular RNAs in Peripheral Blood Mononuclear Cells of Patients with Parkinson's Disease. *Mov. Disord. Off. J. Mov. Disord. Soc.* <https://doi.org/10.1002/mds.28467> (2021).
63. Kong, F. et al. RNA-sequencing of peripheral blood circular RNAs in Parkinson disease. *Med. (Baltim.)* **100**, e25888 (2021).
64. Zhong, L., Ju, K., Chen, A. & Cao, H. Circulating CircRNAs Panel Acts as a Biomarker for the Early Diagnosis and Severity of Parkinson's Disease. *Front. Aging Neurosci.* **13**, 684289 (2021).
65. Xiao, Y. et al. The Potential Circular RNAs Biomarker Panel and Regulatory Networks of Parkinson's Disease. *Front. Neurosci.* **16**, 893713 (2022).
66. Marek, K. et al. The Parkinson Progression Marker Initiative (PPMI). *Prog. Neurobiol.* **95**, 629–635 (2011).
67. Yarnall, A. J. et al. Characterizing mild cognitive impairment in incident Parkinson disease: the ICICLE-PD study. *Neurology* **82**, 308–316 (2014).
68. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci.* **102**, 15545–15550 (2005).
69. Zhong, Y., Kinio, A. & Saleh, M. Functions of NOD-Like Receptors in Human Diseases. *Front. Immunol.* **4**, 333 (2013).
70. Infante, J. et al. Comparative blood transcriptome analysis in idiopathic and LRRK2 G2019S-associated Parkinson's disease. *Neurobiol. Aging* **38**, 214.e1–214.e5 (2016).
71. Garofalo, M. et al. Alzheimer's, Parkinson's Disease and Amyotrophic Lateral Sclerosis Gene Expression Patterns Divergence Reveals Different Grade of RNA Metabolism Involvement. *Int. J. Mol. Sci.* **21**, 9500 (2020).
72. Henderson, A. R. et al. DNA Methylation and Expression Profiles of Whole Blood in Parkinson's Disease. *Front. Genet.* **12**, 640266 (2021).
73. Kurvits, L. et al. Transcriptomic profiles in Parkinson's disease. *Exp. Biol. Med.* **246**, 584–595 (2021).
74. Pain, S. et al. Regulation of Protein Synthesis and Apoptosis in Lymphocytes of Parkinson Patients: The Effect of Dopaminergic Treatment. *Neurodegener. Dis.* **19**, 178–183 (2020).
75. Lu, B., Gehrke, S. & Wu, Z. RNA Metabolism in the Pathogenesis of Parkinson's Disease. *Brain Res.* **0**, 105–115 (2014).
76. Tansey, M. G. et al. Inflammation and immune dysfunction in Parkinson disease. *Nat. Rev. Immunol.* **22**, 657–673 (2022).
77. Ivanov, A. et al. Analysis of Intron Sequences Reveals Hallmarks of Circular RNA Biogenesis in Animals. *Cell Rep.* **10**, 170–177 (2015).
78. Kokot, K. E. et al. Reduction of A-to-I RNA editing in the failing human heart regulates formation of circular RNAs. *Basic Res. Cardiol.* **117**, 32 (2022).
79. Roth, S. H., Levanon, E. Y. & Eisenberg, E. Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nat. Methods* **16**, 1131–1138 (2019).
80. Wu, S. et al. The Potential Regulation of A-to-I RNA Editing on Genes in Parkinson's Disease. *Genes* **14**, 919 (2023).
81. Pakos-Zebrucka, K. et al. The integrated stress response. *EMBO Rep.* **17**, 1374–1395 (2016).
82. Bisbal, C. & Silverman, R. H. Diverse functions of RNase L and implications in pathology. *Biochimie* **89**, 789–798 (2007).
83. Olsen, L. K., Dowd, E. & McKernan, D. P. A role for viral infections in Parkinson's etiology. *Neuronal Signal.* **2**, NS20170166 (2018).
84. Bando, Y. et al. Double-strand RNA dependent protein kinase (PKR) is involved in the extrastriatal degeneration in Parkinson's disease and Huntington's disease. *Neurochem. Int.* **46**, 11–18 (2005).
85. Kim, Y. et al. PKR is activated by cellular dsRNAs during mitosis and acts as a mitotic regulator. *Genes Dev.* **28**, 1310–1322 (2014).
86. Kim, Y. et al. PKR Senses Nuclear and Mitochondrial Signals by Interacting with Endogenous Double-Stranded RNAs. *Mol. Cell* **71**, 1051–1063.e6 (2018).
87. Lee, H. et al. Cell Type-Specific Transcriptomics Reveals that Mutant Huntingtin Leads to Mitochondrial RNA Release and Neuronal Innate Immune Activation. *Neuron* **107**, 891–908.e8 (2020).
88. Rodriguez, S. et al. Genome-encoded cytoplasmic double-stranded RNAs, found in C9ORF72 ALS-FTD brain, propagate neuronal loss. *Sci. Transl. Med.* **13**, eaaz4699 (2021).
89. Ochoa, E. et al. Pathogenic tau-induced transposable element-derived dsRNA drives neuroinflammation. *Sci. Adv.* **9**, eabq5423 (2023).
90. Xu, W. et al. Double-stranded RNA-induced dopaminergic neuronal loss in the substantia nigra in the presence of Mac1 receptor. *Biochem. Biophys. Res. Commun.* **533**, 1148–1154 (2020).
91. Deshpande, P. et al. Protein synthesis is suppressed in sporadic and familial Parkinson's disease by LRRK2. *FASEB J.* **34**, 14217–14233 (2020).
92. Flinkman, D. et al. Regulators of proteostasis are translationally repressed in fibroblasts from patients with sporadic and LRRK2-G2019S Parkinson's disease. *Npj Park. Dis.* **9**, 1–13 (2023).

93. Makarious, M. B. et al. Multi-modality machine learning predicting Parkinson's disease. *NPJ Park. Dis.* **8**, 35 (2022).
94. D'Ambrosi, S. et al. Combinatorial Blood Platelets-Derived circRNA and mRNA Signature for Early-Stage Lung Cancer Detection. *Int. J. Mol. Sci.* **24**, 4881 (2023).
95. Lv, X. et al. Identification of potential key genes and pathways predicting pathogenesis and prognosis for triple-negative breast cancer. *Cancer Cell Int.* **19**, 172 (2019).
96. Simunovic, F. et al. Gene expression profiling of substantia nigra dopamine neurons: further insights into Parkinson's disease pathology. *Brain* **132**, 1795–1809 (2009).
97. Chung, S. J. et al. Genomic determinants of motor and cognitive outcomes in Parkinson's disease. *Parkinsonism Relat. Disord.* **18**, 881–886 (2012).
98. Stone, M. J., Chuang, S., Hou, X., Shoham, M. & Zhu, J. Z. Tyrosine sulfation: an increasingly recognised post-translational modification of secreted proteins. *N. Biotechnol.* **25**, 299–317 (2009).
99. Kyogoku, C. et al. Cell-Specific Type I IFN Signatures in Autoimmunity and Viral Infection: What Makes the Difference? *PLoS ONE* **8**, e83776 (2013).
100. Main, B. S. et al. Type-1 interferons contribute to the neuroinflammatory response and disease progression of the MPTP mouse model of Parkinson's disease. *Glia* **64**, 1590–1604 (2016).
101. Khodadadi, H. et al. PTRHD1 (C2orf79) mutations lead to autosomal-recessive intellectual disability and parkinsonism. *Mov. Disord.* **32**, 287–291 (2017).
102. Al-Kasbi, G. et al. Biallelic PTRHD1 Frameshift Variants Associated with Intellectual Disability, Spasticity, and Parkinsonism. *Mov. Disord. Clin. Pract.* **8**, 1253–1257 (2021).
103. Drew, K., Wallingford, J. B. & Marcotte, E. M. hu.MAP 2.0: integration of over 15,000 proteomic experiments builds a global compendium of human multi-protein assemblies. *Mol. Syst. Biol.* **17**, e10016 (2021).
104. Luo, Z. et al. Circular RNA circCCDC9 acts as a miR-6792-3p sponge to suppress the progression of gastric cancer through regulating CAV1 expression. *Mol. Cancer* **19**, 86 (2020).
105. Wu, L. et al. Circular RNA circCCDC9 alleviates ischaemic stroke ischaemia/reperfusion injury via the Notch. *Pathw. J. Cell. Mol. Med.* **24**, 14152–14159 (2020).
106. Imai, Y. et al. The Parkinson's Disease-Associated Protein Kinase LRRK2 Modulates Notch Signaling through the Endosomal Pathway. *PLoS Genet.* **11**, e1005503 (2015).
107. Giordano, F. et al. PI(4,5)P2-Dependent and Ca²⁺-Regulated ER-PM Interactions Mediated by the Extended Synaptotagmins. *Cell* **153**, 1494–1509 (2013).
108. Tagawa, T. et al. Characterizing Expression and Regulation of Gamma-Herpesviral Circular RNAs. *Front. Microbiol.* **12**, 670542 (2021).
109. Swaminathan, S. et al. Analysis of Copy Number Variation in Alzheimer's Disease in a Cohort of Clinically Characterized and Neuropathologically Verified Individuals. *PLoS ONE* **7**, e50640 (2012).
110. Spielmann, N. et al. Extensive identification of genes involved in congenital and structural heart disorders and cardiomyopathy. *Nat. Cardiovasc. Res.* **1**, 157–173 (2022).
111. Hawkes, C. H., Del Tredici, K. & Braak, H. Parkinson's disease: a dual-hit hypothesis. *Neuropathol. Appl. Neurobiol.* **33**, 599–614 (2007).
112. Lindestam Arlehamn, C. S. et al. α -Synuclein-specific T cell reactivity is associated with preclinical and early Parkinson's disease. *Nat. Commun.* **11**, 1875 (2020).
113. Jensen, M. P. et al. Lower Lymphocyte Count is Associated With Increased Risk of Parkinson's Disease. *Ann. Neurol.* **89**, 803–812 (2021).
114. Qu, Y. et al. A systematic review and meta-analysis of inflammatory biomarkers in Parkinson's disease. *Npj Park. Dis.* **9**, 1–14 (2023).
115. Alhasan, A. A. et al. Circular RNA enrichment in platelets is a signature of transcriptome degradation. *Blood* **127**, e1–e11 (2016).
116. Nicolet, B. P. et al. Circular RNA expression in human hematopoietic cells is widespread and cell-type specific. *Nucleic Acids Res.* **46**, 8168–8180 (2018).
117. Lowes, H., Pyle, A., Santibanez-Koref, M. & Hudson, G. Circulating cell-free mitochondrial DNA levels in Parkinson's disease are influenced by treatment. *Mol. Neurodegener.* **15**, 10 (2020).
118. Xiao, M.-S. & Wilusz, J. E. An improved method for circular RNA purification using RNase R that efficiently removes linear RNAs containing G-quadruplexes or structured 3' ends. *Nucleic Acids Res.* **47**, 8755–8769 (2019).
119. Hughes, A. J., Daniel, S. E., Kilford, L. & Lees, A. J. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J. Neurol. Neurosurg. Psychiatry* **55**, 181–184 (1992).
120. Cunningham, F. et al. Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995 (2022).
121. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
122. Patro, R., Duggal, G., Love, M. I., Lizarray, R. A. & Kingsford, C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat. Methods* **14**, 417–419 (2017).
123. Sonesson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521 (2015).
124. Gao, Y., Zhang, J. & Zhao, F. Circular RNA identification based on multiple seed matching. *Brief. Bioinform.* **19**, 803–810 (2018).
125. Izuogu, O. G. et al. PTESFinder: a computational method to identify post-transcriptional exon shuffling (PTES) events. *BMC Bioinforma.* **17**, 31 (2016).
126. Zhang, X.-O. et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* **26**, 1277–1287 (2016).
127. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
128. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
129. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
130. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
131. Zhang, J., Chen, S., Yang, J. & Zhao, F. Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat. Commun.* **11**, 90 (2020).
132. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **7**, 1338 (2018).
133. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
134. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
135. T Hoen, P. A. C. et al. Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nat. Biotechnol.* **31**, 1015–1022 (2013).
136. Su, Z. et al. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **32**, 903–914 (2014).
137. Wang, M. et al. Transformative Network Modeling of Multi-omics Data Reveals Detailed Circuits, Key Regulators, and Potential Therapeutics for Alzheimer's Disease. *Neuron* **109**, 257–272.e14 (2021).
138. Lopes, K. et al. Genetic analysis of the human microglial transcriptome across brain regions, aging and disease pathologies. *Nat. Genet.* **54**, 4–17 (2022).
139. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinforma.* **17**, 483 (2016).
140. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
141. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
142. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
143. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
144. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
145. Martin, A. R. et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat. Genet.* **51**, 1560–1565 (2019).
146. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).
147. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinforma.* **12**, 77 (2011).
148. Lewis, M. J. et al. nestedcv: an R package for fast implementation of nested cross-validation with embedded feature selection designed for transcriptomics and high-dimensional data. *Bioinforma. Adv.* **3**, vbad048 (2023).
149. Liddicoat, B. J. et al. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**, 1115–1120 (2015).

ACKNOWLEDGEMENTS

This work was funded by the Michael J. Fox Foundation (Grant ID MJFF-007574). G.H. receives funding from Wellcome (203105/Z/16/Z), the Michael J. Fox Foundation (MJFF-007574) and Parkinson's UK (G-2003 and G-2201). ICICLE-PD was funded by Parkinson's UK (J-0802, G-1301, G-1507) and supported by the Lockhart Parkinson's Disease Research Fund, National Institute for Health Research (NIHR) Newcastle Biomedical Research Unit and Centre based at Newcastle upon Tyne Hospitals NHS Foundation Trust and Newcastle University. R.A.L. is supported by a Janet Owens Parkinson's UK Senior Research Fellowship (F-1801). The authors would like to express their deepest gratitude to the Parkinson's Progression Markers Initiative and The Incidence of Cognitive Impairment in Cohorts with Longitudinal Evaluation-PD study team members as well as to the patients and staff at each study. The authors

wish to thank the Newcastle Genomics Core Facility for their assistance in experimental design and data generation. This research made use of the Rocket High-Performance Computing service at Newcastle University. In addition to the listed authors, we thank the following members of the ICICLE-PD Study Group who all made a significant contribution to the work reported in this paper: Patrick F Chinnery (Department of Clinical Neurosciences, University of Cambridge, UK, Principal Investigator); John T O'Brien (Department of Psychiatry, University of Cambridge, UK, Principal Investigator); Trevor W Robbins (Department of Psychology, University of Cambridge, UK, Principal Investigator); Caroline Williams Grey, Marta Camacho Gemma A Cummins, Jonathan Evans, Ruwani Wijeyekoon, Kirsten Scott, Tom Stoker, Julia Greenland, Natalie Valle Guzman, Lucy Collins, Simon Stott and Sarah Mason (John van Geest Centre for Brain Repair, University of Cambridge, UK, Site Investigators); David J Brooks (Translational and Clinical Research Institute, Newcastle University, UK, Principal Investigator); and James B Rowe (Behavioural and Clinical Neuroscience Institute, UK, Site Investigator). Figures 1 and 6 make use of icons from Biolcons and OpenMojii.

AUTHOR CONTRIBUTIONS

M.S.J., M.S.K. and G.H. designed the study. B.J.W., O.G.J., D.D. and A.P. performed data generation and analysis. A.P. and H.L. prepared blood samples for RNA sequencing. J.C. performed the RNA sequencing. R.A.L. and A.J.Y. provided clinical and patient samples (ICICLE-PD) and supported data interpretation. B.J.W., M.S.K. and G.H. conducted statistical analysis. B.J.W., M.S.K. and G.H. wrote the manuscript with feedback and input from all authors. All authors read and approved the final manuscript.

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/s41531-024-00636-y>.

Correspondence and requests for materials should be addressed to Mauro Santibanez-Koref or Gavin Hudson.

Reprints and permission 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024