

ARTICLE OPEN

Mitochondria function associated genes contribute to Parkinson's Disease risk and later age at onset

Kimberley J. Billingsley^{1,2}, Ines A. Barbosa³, Sara Bandrés-Ciga², John P. Quinn¹, Vivien J. Bubb¹, Charu Deshpande⁴, Juan A. Botia^{5,6}, Regina H. Reynolds⁶, David Zhang⁶, Michael A. Simpson³, Cornelis Blauwendraat², Ziv Gan-Or^{7,8,9}, J. Raphael Gibbs², Mike A. Nalls^{2,10}, Andrew Singleton², International Parkinson's Disease Genomics Consortium (IPDGC), Mina Ryten⁶ and Sulev Koks^{11,12}

Mitochondrial dysfunction has been implicated in the etiology of monogenic Parkinson's disease (PD). Yet the role that mitochondrial processes play in the most common form of the disease; sporadic PD, is yet to be fully established. Here, we comprehensively assessed the role of mitochondrial function-associated genes in sporadic PD by leveraging improvements in the scale and analysis of PD GWAS data with recent advances in our understanding of the genetics of mitochondrial disease. We calculated a mitochondrial-specific polygenic risk score (PRS) and showed that cumulative small effect variants within both our primary and secondary gene lists are significantly associated with increased PD risk. We further reported that the PRS of the secondary mitochondrial gene list was significantly associated with later age at onset. Finally, to identify possible functional genomic associations we implemented Mendelian randomization, which showed that 14 of these mitochondrial function-associated genes showed functional consequence associated with PD risk. Further analysis suggested that the 14 identified genes are not only involved in mitophagy, but implicate new mitochondrial processes. Our data suggests that therapeutics targeting mitochondrial bioenergetics and proteostasis pathways distinct from mitophagy could be beneficial to treating the early stage of PD.

npj Parkinson's Disease (2019)5:8; <https://doi.org/10.1038/s41531-019-0080-x>

INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative movement disorder characterized pathologically by the death of dopaminergic neurons in the substantia nigra (SN) and aggregation of α -synuclein protein (encoded by *SNCA*), within intraneuronal inclusions called Lewy bodies. The majority of PD cases are apparently sporadic in nature (~95%). Aging is a major risk factor for disease and due to population ageing the prevalence of PD is predicted to increase rapidly, making the identification of therapeutic targets a high priority.^{1–3}

Although there have been great advances in understanding both the genetic architecture and cellular processes involved in PD, the exact molecular mechanisms that underlie PD remain unknown.¹ However, it is suggested that PD has a complex etiology, involving several molecular pathways, and understanding these specific pathways will be key to establishing mechanistic targets for therapeutic intervention. While several key pathways are currently being investigated, including autophagy, endocytosis, immune response, and lysosomal function,^{4–7} mitochondrial function was the first biological process to be associated with PD.^{8,9}

An interest in mitochondrial function and PD began with the observation that exposure to the drug 1-methyl-4-phenyl-1,2,3,4-tetrahydropyridine (MPTP) can cause rapid parkinsonism and neuronal loss in the SN in humans, and that this is mediated through inhibition of complex I of the mitochondrial electron transport chain.^{7,10,11} Subsequent work suggested that individuals with sporadic PD have reduced complex I activity not only in the SN, but in other brain regions and peripheral tissues.¹² Genetic studies focusing on monogenic forms of PD provided further support for the involvement of mitochondrial dysfunction in the disease. Pathogenic mutations that lead to autosomal recessive forms of PD have been reported in *PINK1*, *PARK2*, *PARK7*, *CHCHD2*, and *VPS13C* and the proteins they encode are all now known to be involved in the mitochondrial quality control system and in particular mitophagy.^{13–16}

Therefore, in this paper, we aim to comprehensively assess the role of mitochondrial function in sporadic PD by leveraging improvements in the scale and analysis of PD genome-wide association study (GWAS) data with recent advances in our understanding of the genetics of mitochondrial disease. The availability of large-scale genome-wide association data in PD

¹Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, Liverpool, UK; ²Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA; ³Department of Medical and Molecular Genetics, King's College London School of Basic and Medical Biosciences, London SE1 9RT, UK; ⁴Clinical Genetics Unit, Guys and St. Thomas' NHS Foundation Trust, London SE1 9RT, UK; ⁵Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, 30100 Murcia, Spain; ⁶Department of Neurodegenerative Disease, UCL Institute of Neurology, 10-12 Russell Square House, London, UK; ⁷Montreal Neurological Institute, McGill University, Montréal, QC, Canada; ⁸Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada; ⁹Department of Human Genetics, McGill University, Montréal, QC, Canada; ¹⁰Data Tecnica International, Glen Echo, MD 20812, USA; ¹¹The Perron Institute for Neurological and Translational Science, 8 Verdun Street, Nedlands, WA 6009, Australia and ¹²Centre for Comparative Genomics, Murdoch University, Murdoch 6150, Australia

Correspondence: Sulev Koks (sulev.koks@perron.uwa.edu.au)

Members of the IPDGC are listed at the end of the paper.

These authors contributed equally: Kimberley J. Billingsley, Ines A. Barbosa, Mina Ryten and Sulev Koks

Received: 3 December 2018 Accepted: 15 April 2019

Published online: 22 May 2019

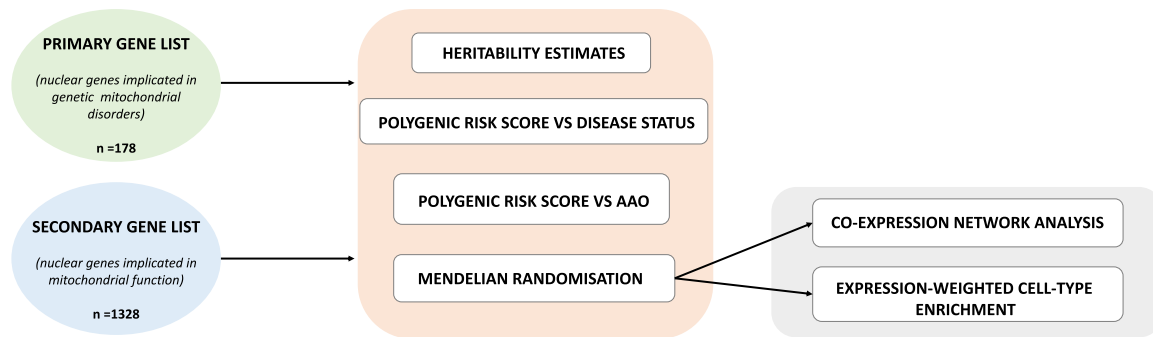


Fig. 1 Workflow of mitochondrial-function specific PD analysis

Table 1. Cohort-level heritability analysis for the primary and secondary mitochondrial gene lists

	Primary		Secondary	
	Heritability estimate	SE of heritability estimate	Heritability estimate	SE of heritability estimate
WTCCC PD GWAS (PMID:21044948)	0.00321	0.00277	0.00563	0.00688
Spanish Parkinson's (IPDGC) part2	0.00027	0.00314	0.00629	0.00932
NIA PD GWAS (PMID:19915575)	0.00945	0.00540	0.03616	0.01365
Dutch GWAS (PMID:21248740)	0.00000	0.00530	0.03562	0.01681

Reporting estimates for the WTCCC PD GWAS (PMID:21044948), Spanish Parkinson's (IPDGC) part2, NIA PD GWAS (PMID:19915575), Dutch GWAS (PMID:21248740) cohorts. Showing heritability estimates generated using GCTA and standard error of the estimates (SE)

cases and the rapid identification of genetic lesions that underlie mitochondrial disease provide an opportunity to systematically assess the role of genetic variability in mitochondrial linked genes in the context of risk for PD.¹⁷ In this study we combine these new resources with current statistical tools, such as polygenic risk scoring and Mendelian randomization, to explore the role of mitochondrial function in both PD risk and age at onset of disease to obtain novel insights.

RESULTS

Common variation within mitochondria function genes contributes to the heritable component of PD

The general workflow for the genetic analysis used in this study is shown in Fig. 1. First, to study the importance of mitochondrial function in sporadic PD, we investigated the heritability of PD specifically within genomic regions that contained genes annotated as important in mitochondrial function. The construction of this annotation was driven by the principle that genomic regions, which are known to be the sites of mutations in individuals with rare mitochondrial diseases or are candidate regions for such mutations provide the best evidence for involvement in mitochondrial function.

GCTA is a statistical method that estimates phenotypic variance of complex traits explained by genome-wide SNPs, including those not associated with the trait in a GWAS. Using GCTA, heritability estimates were first calculated for the four largest IPDGC GWAS datasets, including all variants (WTCCC PD GWAS (PMID:21044948), Spanish Parkinson's (IPDGC) part2, NIA PD GWAS (PMID:19915575), Dutch GWAS (PMID:21248740)). Owing to the low number of included cases, the heritability estimates in the other IPDGC datasets were deemed less reliable. Consistent with previous heritability estimates from both Keller et al.²⁴ (24%) and Chang et al.²¹ (21%), our random-effects meta-analysis for the four datasets identified 23% (95% CI 12–34, $p = 2.72 \times 10^{-5}$) phenotypic variance associated with all PD samples (Tables 1 and 2). There was a high degree of consistency across the cohorts.

After establishing the consistency of our heritability estimates, we next calculated heritability using only variants located within genetic regions specified as being of primary ($n = 176$) or secondary ($n = 1463$) importance in mitochondrial function. Initially, to assess the full contribution of the mitochondria pathway we ran the analysis including and excluding the PD risk genes.⁵ However, as shown in Supplementary Fig. 1 there was little difference overall in the heritability estimates. Therefore, to only to catalog mitochondria-specific genetic risk outside of known loci we focused on the analysis excluding these genes. The heritability estimate using a random-effects meta-analysis for the primary gene list excluding the PD genes was estimated to be a modest 0.26% (95% CI -0.11 – 0.66 , $p = 0.166$). However, the heritability estimate using a random-effects meta-analysis for the secondary list, namely genes implicated in mitochondrial function or morphology, as well as disease, was estimated to be 1.67% (95% CI -0.07 – 0.32 , $p = 0.041$).

Mitochondria function-specific polygenic risk score is significantly associated with disease status

Next, we calculated PRS to capture the additive effect of all common variants within genes implicated in mitochondria function on PD risk. PRS is a particularly powerful approach in this context because it is able to efficiently incorporate information from all hits, including sub-significant hits, which may nonetheless be etiologically relevant. Again, initially we ran the analysis including and excluding the PD risk genes and the comparison can be seen in Supplementary Fig. 2. However, to only report novel associations we focused on the lists excluding the PD risk genes.

Using this approach, the primary and secondary mitochondrial genomic annotations were found to be significantly associated with PD disease status. Remarkably, the primary gene list consisting of only 176 genes implicated in Mendelian mitochondrial disorders was associated with PD with an odds ratio of 1.12 per standard deviation increase in the PRS from the population mean (random-effects p -value = 6.00×10^{-4} , $\beta = 0.11$, SE = 0.03).

Table 2. Summary of random-effects meta-analysis for the primary and secondary mitochondrial gene lists

	Heritability estimate from random-effects	Lower 95% confidence interval	Upper 95% confidence interval	<i>p</i> -value from random effects	Heterogeneity of variance from random effects (%)	Heterogeneity <i>p</i> -value
All SNPs	0.2313	0.1233	0.3393	2.72E-05	0.0100	3.00E-03
Primary	0.0026	-0.0011	0.0062	1.66E-01	0.0000	4.85E-01
Secondary	0.0167	0.0007	0.0328	4.10E-02	0.0001	9.63E-02

Here, we show the random-effects meta-analysis of heritability estimates for all SNPs in the genome (All SNPs), estimate calculated with for the SNPs within the primary mitochondria list genes (Primary), and the SNPs within the secondary mitochondria list genes (Secondary)

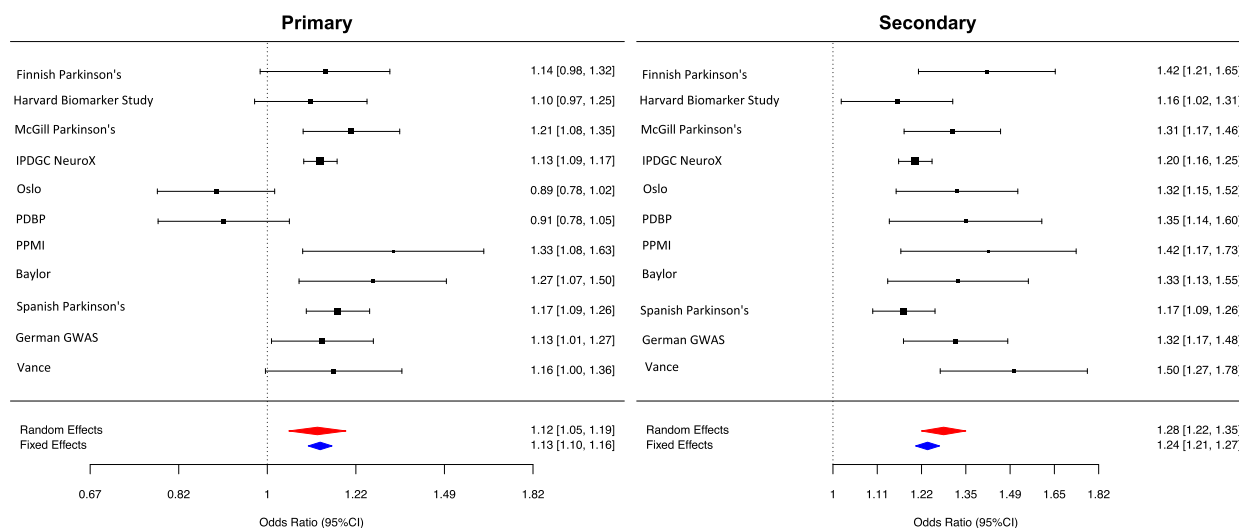


Fig. 2 Forest plots of PRS for Parkinson's Disease across cohorts. Random effect meta-analysis results are shown as red diamonds and fixed effects are shown as blue, with the centerline of each diamond representing the summary PRS for that dataset. IPDGC NeuroX = (Nalls et al. 2015, PMID:25444595), OSLO = Oslo Parkinson's Disease Study, PDBP = Parkinson's Disease Biomarker's Program, PPMI = Parkinson's Progression Markers Initiative, Baylor = Baylor College of Medicine/University of Maryland, German GWAS = (PMID:19915575), VANCE = Vance (dbGap phs000394)

The secondary gene list, which also included genes implicated in mitochondria function or morphology, was associated with PD with a higher odds ratio of 1.28 per standard deviation increase in the PRS from the population mean (random-effects *p*-value = 1.9E-22, beta = 0.25, SE = 0.03) (Fig. 2). Altogether, these analyses not only provide further support for importance of mitochondrial processes in PD, but potentially provide a tool for identifying PD patients most likely to benefit from treatments specifically targeting mitochondrial function.

Mitochondria function-specific polygenic risk score is significantly associated with later age at onset

Although multiple lines of evidence point to the importance of mitochondrial dysfunction as a primary cause of PD, impaired mitochondrial dynamics appears to be common to a wide range of neurodegenerative diseases, including Huntington's disease,^{18,19} amyotrophic lateral sclerosis,^{20,21} and Alzheimer's disease.^{22–25} The latter suggests that even when impaired mitochondrial function is not the primary event in disease pathogenesis, it is a common outcome and could contribute to disease progression. We sought to test this hypothesis by investigating the importance of common variation within our mitochondrial gene lists in determining the age at onset of PD (AAO). Given the significant lag period between PD pathophysiology and symptoms, AAO was used as an indirect measure of disease progression. This analysis was performed using PRS since it has been consistently found to be the main genetic predictor of

AAO^{6,26–28} with higher genetic risk scores being significantly associated with an overall trend for earlier AAO of disease. While the primary mitochondrial gene list was not significantly associated with AAO of disease, the secondary gene list was correlated with AAO. Contrary to expectation, the cumulative burden of common variants within the 1326 genes comprising the PRS for PD risk were positively correlated with AAO of PD. After meta-analyzing, we found that each 1SD increase in PRS, led to a 0.55 year increase in the AAO of disease (summary effect = 0.211, 95% CI (0.141–0.970), $I^2 = 68.49\%$, *p*-value = 9.00E-03, Fig. 3). As the forest plots demonstrate, although there was a relatively high heterogeneity across studies, the directionality and magnitude of effect on AAO were in concordance with the meta-analysis with the exception of the Oslo cohort. This finding could suggest that firstly, disease causation and progression are genetically separable processes in PD and that secondly the role of mitochondrial dysfunction in PD is likely to be highly complex with multiple distinct mitochondrial processes likely to be involved at different disease stages.

Mendelian randomization suggests potential causal association of 14 novel mitochondria function genes with PD risk

Given the robust evidence for the involvement of mitochondrial function in sporadic PD, next we used two-sample MR analysis to identify specific genes likely to be important in PD risk. MR uses genetic variants to identify if an observed association between a risk factor and an outcome is consistent with causal effect.²⁹ This

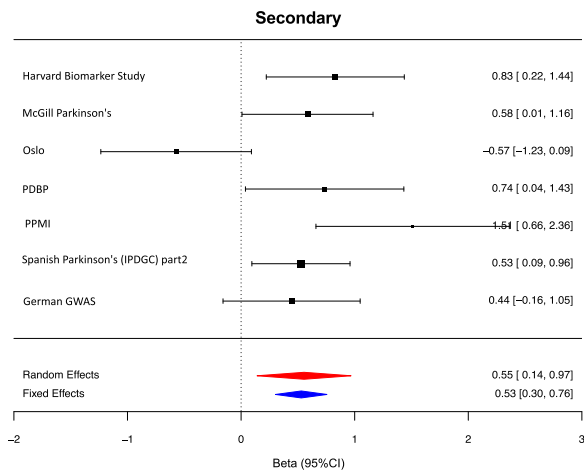


Fig. 3 Forest plots of PRS for the age at onset of Parkinson's Disease across cohorts. Random effect meta-analysis results are shown as red diamonds and fixed effects are shown as blue, with the centerline of each diamond representing the summary PRS for that dataset. OSLO = Oslo Parkinson's Disease Study, PDBP = Parkinson's Disease Biomarker's Program, PPMI = Parkinson's Progression Markers Initiative, German GWAS = (PMID:19915575), VANCE = Vance

method has been implemented in several recent genetic studies to identify association between expression quantitative trait loci (QTL), to more accurately nominate candidate genes within risk loci. Therefore, for this study, in the aim of identifying whether changes in expression of mitochondria function genes are potentially causally related to PD risk, two-sample MR was implemented.

Since we wanted to identify novel associations, we excluded genes already linked to PD through the most recent GWAS meta-analysis.⁶ This resulted in the exclusion of 31 genes linked to mitochondrial function and in linkage disequilibrium with the top PD risk variants. Analysis of the remaining 1432 genes (generated by combining the primary and secondary gene lists) resulted in the identification of 14 novel genes linked to mitochondrial function and causally associated with PD risk (Table 3). Of the 14 genes, the expression of five genes (*CLN8*, *MPI*, *LGALS3*, *CAPRIN2*, and *MUC1*) was positively associated with PD risk in blood. Similarly, in brain PD risk was associated with increased expression of *ATG14*, *E2F1*, and *EP300* in brain. However, negative associations in brain and blood expression were observed for *MRPS34* and *PTPN1* and *LMBRD1*, respectively. Finally, elevated methylation of *FASN* in the brain was found to be positively associated with PD risk but elevated methylation of *CRY2* was found to be inversely correlated.

Six of the fourteen novel PD risk genes we identified (*CLN8*, *EP300*, *LMBRD1*, *MPI*, *MRPS34*, and *MUC1*) are already associated with a monogenic disorder (Table 3). We noted that neurological abnormalities were a feature of the condition in five of the six cases with Combined Oxidative Phosphorylation Deficiency 32 due to biallelic mutations in *MRPS34* being perhaps of particular interest. In common with PD, this condition is associated with abnormalities of movement, including dystonia and choreoathetoid movements. Mutations causing this condition result in decreased levels of *MRPS34* protein causing destabilization of the small mitochondrial ribosome subunit and suggesting the involvement of mitochondrial processes distinct from mitophagy and mitochondrial homeostasis in PD. In this context, it is noteworthy that *MRPL43*, another nuclear gene encoding for a component of the large mitochondrial ribosome subunit is also highlighted by the MR analysis. Thus, this analysis not only enabled us to identify specific genes of interest, but also pointed

to the role of multiple mitochondrial processes in PD distinct from mitophagy.

Exploring the expression of the novel mitochondria risk genes provides additional support for their role in PD

We leveraged publicly available cell-specific and tissue-specific gene expression data to investigate the 14 mitochondria genes implicated in PD through MR. First, we used enrichment-weighted cell-type enrichment (EWCE) to determine whether the expression of mitochondrial PD-associated genes (as identified through MR and described above, $n = 14$) was enriched within a specific cell-type class or their subtypes. No significant enrichment of these genes was found in any of the tested neuronal and glial cell-type classes (Supplementary Table 6 and Supplementary Fig. 3). Next, we used co-expression network analysis to identify possible functional interactions between the 14 novel mitochondrial genes identified through MR and genes implicated in monogenic forms of PD. We found that five of the 14 genes assessed, *CLN8*, *FASN*, *MPI*, *MRPL43*, and *MRPS3*, were highly co-expressed with at least one gene already implicated in monogenic PD in multiple brain regions (> 3 brain regions, Supplementary Table 7). Interestingly, in the case of *CLN8*, *MRPL43*, and *MRPS4*, our novel genes were co-expressed with a monogenic PD gene already implicated in mitochondrial function such as *PARK7*. Furthermore, with the exception of *CLN8* (*FASN*, *MPI*, *MRPL43* and *MRPS3*), the novel mitochondrial gene was assigned to a co-expression module enriched for neuronal markers (Supplementary Table 8).

DISCUSSION

We first demonstrate that a proportion of the "missing heritability" of sporadic PD can be explained by additive common genetic variation within genes implicated in mitochondrial function, even after exclusion of genes previously linked to PD through linkage disequilibrium with the top risk variants.^{4-6,30-34} We identify that the overall heritability of PD is approximately 23% (Table 2) and that when heritability is calculated for the secondary gene list regions alone we can attribute around 7% of the overall heritability (23%) to common variation within the mitochondria function associated genes. Although this initially looks like a low estimate, to put this into context, when heritability was calculated in the most recent meta-analysis using the top PD risk hits, this estimate explained only 26–36% of the overall heritability.⁶ In addition, using PRS, which efficiently incorporates information from sub-significant hits, we showed that cumulative small effect variants within only 196 genes linked to monogenic mitochondrial disease significantly associated with increased PD risk (with odds ratios of 1.12 per standard deviation increase in PRS from the population mean). These findings are important for two main reasons. Firstly, given that risk profiling performed in the recent PD meta-analysis did not identify a significant association with mitochondrial function,^{4-6,17,30-34} Secondly, since the primary gene list consisted solely of the 196 genes mutated in monogenic mitochondrial disorders, this analysis highlights the increasingly close relationship between Mendelian and complex disease.⁷

In order to maximize the utility of this study, we used MR, which identified 14 specific mitochondrial genes of interest with putative functional consequences in PD risk. We found that although a number of the genes we identified are clearly linked to known PD-related pathways, such as lysosomal dysfunction in the case of *CLN8* and *LMBRD1* or autophagy in the case of *ATG14*, others appeared to point towards new processes. In particular, this analysis highlighted the mitochondrial ribosome through the identification of the genes, *MRPL43* and *MRPS34*, encoding components of the large and small mitochondrial ribosome subunits, respectively. Interestingly, biallelic mutations in *MRPS34* are known to cause a form of Leigh syndrome, characterized by

Table 3. Significant functional associations of mitochondrial function genes via two-sample Mendelian randomization

Genes	Beta	SE	P, FDR, adjusted	Probe	Data source	Analyte	Top QTL SNP	CHR, top QTL SNP	BP, top QTL SNP	Associated phenotype in OMIM	Neurological phenotypic features	Treatment response
<i>LMBRD1</i>	-0.173	0.050	4.71E-04	ENSG00000117245	Expression ⁵⁷	Brain	rs7763213	6	70416294	Methylmalonic aciduria and homocystinuria, CBLF type; MAHCF	Hypotonia, lethargy, developmental delay, impaired coordination	Responsive to vitamin B12 therapy
<i>ATG14</i>	0.113	0.028	4.21E-05	ENSG00000171612	Expression ⁵⁷	Brain	rs11621265	14	55822095	NA	NA	NA
<i>EZF1</i>	0.121	0.032	1.52E-04	ENSG00000159423	Expression ⁵⁷	Brain	rs57668191	20	32289763	NA	NA	NA
<i>EP300</i>	0.151	0.042	2.80E-04	ENSG00000090432	Expression ⁵⁷	Brain	rs139486	22	41627654	Rubinstein-taybi syndrome 2; RTS2	Behavioral difficulties, mental retardation (mild to moderate), low-normal intelligence, Autism spectrum disorder (in some patients), delayed psychomotor development, delayed gross motor development, speech delay	NA
<i>MRPS34</i>	-0.353	0.100	4.04E-04	ILMN_2210482	Expression ⁵⁸	Blood	rs2575369	16	1817431	Combined oxidative phosphorylation deficiency 32; COXPD32	Delayed psychomotor development, lack of speech, inability to walk, spasticity, hyperreflexia, dystonia, choreoathetoid movements, abnormal T2-weighted signals in the basal ganglia and brainstem	NA
<i>PTPN1</i>	-0.090	0.022	5.27E-05	ILMN_1681591	Expression ⁵⁸	Blood	rs17788127	20	49166548	NA	NA	NA
<i>MRPL43</i>	0.047	0.014	8.78E-04	ILMN_1652147	Expression ⁵⁸	Blood	rs2863095	10	102746503	NA	NA	NA
<i>CLN8</i>	0.148	0.047	1.54E-03	ILMN_1701094	Expression ⁵⁸	Blood	rs3812477	8	1734452	Ceroid lipofuscinosis, neuronal, 8; CLN8	Developmental regression, seizures, ataxia, speech and language difficulties, myoclonus, EEG abnormalities, cerebral atrophy, cerebellar atrophy, autofluorescent lipopigment in neurons	NA
<i>MPI</i>	0.214	0.068	1.52E-03	ILMN_1761262	Expression ⁵⁸	Blood	rs4886636	15	75196176	Congenital disorder of glycosylation, type Ib; CDG1B	Hypotonia	Responsive to oral mannose therapy
<i>LGALS3</i>	0.307	0.085	3.14E-04	ILMN_1803788	Expression ⁵⁸	Blood	rs7157678	14	55548739	NA	NA	NA
<i>CAPRN2</i>	0.337	0.101	8.49E-04	ILMN_2345739	Expression ⁵⁸	Blood	rs11051061	12	30914668	NA	NA	NA
<i>MUC1</i>	0.487	0.118	3.39E-05	ILMN_1756992	Expression ⁵⁸	Blood	rs6427184	1	155122783	Medullary cystic kidney disease 1; MCKD11	NA	NA
<i>CRY2</i>	-0.054	0.015	1.82E-04	ch.11.939596F	Methylation ⁵⁷	Brain	rs72902436	11	45881792	NA	NA	NA
<i>FASN</i>	0.068	0.019	4.47E-04	cg03407524	Methylation ⁵⁷	Brain	rs9905991	17	80052073	NA	NA	NA

Multi-SNP eQTL Mendelian randomization results focusing on the mitochondria-associated genes (combining the primary and secondary gene lists). Showing the 14 mitochondria function-associated genes that are significantly associated with PD risk after FDR adjustment.

neurodegeneration in infancy with dystonia and choreoathetoid movements due to basal ganglia dysfunction. Furthermore, we note that a recent study that utilized whole exome sequencing (WES) data from two PD cohorts to investigate^{35,36} rare variation in nuclear genes associated with distinct mitochondrial processes, not only provided support for the involvement of mitochondrial function in sporadic PD, but also identified the gene, *MRPL43*, which encodes a component of the large mitochondrial ribosomal subunit.³⁷ Interestingly, *MRPL43* and *MRPS34* were amongst five genes, which were also highly co-expressed in human brain with genes already known to cause monogenic forms of PD. Whereas *MRPL43* and *MRPS34* were highly co-expressed with *PARK7* in modules enriched for neuronal markers, *FASN* and *MPI* were co-expressed with *ATP13A2*, and *CLN8* was located in modules containing *FBXO7* and enriched for oligodendrocyte markers. While this form of analysis does not provide information at the single-cell level, it points to the possibility of pathway interactions between these gene sets. However, most importantly it implicates entirely distinct mitochondrial processes in PD risk.

Finally, and perhaps most remarkably using our mitochondrial gene lists we observe clear differences between disease causation and AAO in PD. Although PRS of the primary mitochondrial gene list was not significantly associated with AAO, the PRS of the secondary mitochondrial gene list was positively correlated (p -value = $3.6E-05$), indicating association with later age at onset. However, given these findings it seems plausible that some mitochondrial processes may contribute to PD risk. Thus, this analysis is consistent with the findings of the most recent and largest AAO PD GWAS, which reported that not all the well-established risk loci are associated with AAO and suggested a different mechanisms for PD causation and AAO.³⁸ This has also been shown in previous studies that have shown no association of risk loci with AAO in sporadic PD.^{35,39}

Although in this study we have comprehensively analyzed the largest PD datasets currently available with very specific and inclusive mitochondrial function gene lists, there are a number of limitations to our analyses. Firstly, there was a relative amount of heterogeneity in AAO within the AAO GWAS studies used. This was due to certain cohorts AAO being self-reported and other cohorts specifically recruiting younger onset cases. Nonetheless, the highly significant p -value we obtain for the association mitochondrial genes and AAO of PD (p -value = $3.56E-05$) and the recognized importance of mitochondrial function in aging would suggest that this finding is likely to be robust. Furthermore, it is important to recognize that our understanding of mitochondrial biology is far from complete and this is made evident by the fact that many individuals with probable genetic forms of mitochondrial disease remain undiagnosed. This in conjunction with the fact we also removed the mitochondria function-associated genes that are known risk hits (such as *SNCA*), suggests that our analysis likely represents an underestimate of the overall contribution of the mitochondria pathway to sporadic PD. This underestimate is illustrated in Supplementary Figs. 1 and 2, which shows that heritability estimates and PRS scores are higher when the PD risk genes are included. However, we note that our ultimate goal for this study was to catalog mitochondria-specific genetic risk outside of known risk loci, which we have subsequently reported.

Another possible limitation of this present study is that the statistical tools we have used in these analyses are currently limited. For example, MR ultimately relies on the availability of sufficient quantities of high-quality eQTL data. However, as there is a future focus to; increase dataset sample size, report and characterize phenotypes such as AAO more accurately and continue to increase the number of identified mitochondrial disease and function genes, we will be able to further explore the role of specific mitochondrial processes in more detail and identify their distinct contribution to disease causation and progression. In regard to the MR nominated genes in particular, further follow-up

functional studies will be crucial to validate how these genes contribute to disease risk. Finally, it is possible that our focus on one specific pathway could infer a selection bias to our analyses. However, the fact that our results are consistent across cohorts and the fact that this significant association is observed in multiple tests, adds validity to our data. In light of this, however, a large-scale unbiased approach should be the focus of future pathway related studies to avoid this potential bias, although this will be difficult given the scope of work and sample size needed.

In summary, in this study we provide robust evidence for the involvement of mitochondrial processes in sporadic PD, as opposed to its defined and well-established role in the monogenic forms of the disease. In relation to the 14 novel mitochondrial function genes that we have identified, our data suggests that it is not only mitochondrial quality control and homeostasis, which contributes to PD risk but other key mitochondrial processes, such as the function of mitochondrial ribosomes, mirroring the biological complexity of mitochondrial disorders. Thus, this study opens the way for the identification of novel drug targets in PD causation and progression.

METHODS

Samples and quality control of IPDGC datasets

All genotyping data was obtained from previously generated IPDGC datasets, consisting of 41,321 individuals (18,869 cases and 22,452 controls) of European ancestry. For the IPDGC datasets all participants donated DNA samples and provided informed consent for participation in genetics studies, which was approved by the local ethic committee for each of the datasets used (National Institutes of Health, Department of Health and Human Services; project Z01 AG000949). Detailed demographic and clinical characteristics are given in Supplementary Table 1 and are explained in further detail in refs.^{6,38} along with detailed quality control (QC) methods. For sample QC, in short, individuals with low call rate ($< 95\%$), discordance between genetic and reported sex, heterozygosity outliers (F -statistic cutoff of > -0.15 and < 0.15) and ancestry outliers ($+/- 6$ standard deviations from means of eigenvectors 1 and 2 of the 1000 Genomes phase 3 CEU and TSI populations from principal components⁴⁰) were excluded. Further, for genotype QC, variants with a missingness rate of $> 5\%$, minor allele frequency < 0.01 , exhibiting Hardy-Weinberg Equilibrium (HWE) $< 1E-5$ and palindromic SNPs were excluded. Remaining samples were imputed using the Haplotype Reference Consortium (HRC) on the University of Michigan imputation server under default settings with Eagle v2.3 phasing based on reference panel HRC r1.1 2016.^{41,42}

Curation of genes implicated in mitochondrial disorders and associated with mitochondrial function

Gene lists were built to encompass different levels of evidence for involvement of the respective protein products in disease phenotypes that relate to mitochondrial function. The list of genes implicated in genetic mitochondrial disorders ("primary" gene list, $n = 196$) has the most stringent criteria of evidence that the respective genes is related to mitochondrial dysfunction. It consists of 102 nuclear genes listed in MITOMAP (downloaded 2015) and 94 sourced from literature review as containing mutations that cause with mitochondrial disease.

The list of genes implicated in mitochondrial function ("secondary" gene list, $n = 1487$) was constructed using the OMIM API to identify all genes for which the word "mitochondria" (or derivatives) appeared in the free-text description, and by combining this information with MitoCarta v2.0 genes with no OMIM phenotype. This therefore gathered a list of plausible biological candidate genes, i.e., genes that are functionally implicated in mitochondrial function and morphology for which we may lack genetic evidence for mitochondrial disease association.

Next, to identify novel PD-associated genes, the 349 genes identified to be in LD with the PD risk variants of interest in the most recent PD meta-analysis⁶ were removed from both lists (removed genes listed in Supplementary Table 2). The final "primary" and "secondary" gene lists are given in Supplementary Table 3 and Supplementary Table 4 and following the removal the PD-associated genes were $n = 178$ and $n = 1328$, respectively.

Cohort-level heritability estimates and meta-analysis

Genome-wide complex trait analysis (GCTA) was used to calculate heritability estimates for the four largest IPDGC cohorts (WTCCC PD GWAS (PMID:21044948), Spanish Parkinson's (IPDGC) part2, NIA PD GWAS (PMID:19915575), Dutch GWAS (PMID:21248740) using non-imputed genotyping data for all variants within both mitochondria gene lists using the same workflow as in ref.⁴³ Genetic relationship matrices were calculated for each dataset to identify the genetic relationship between pairs of individuals. Genetic relationship matrices were then input into restricted maximum likelihood analyses to produce estimates of the proportion of phenotypic variance explained by the SNPs within each subset of data. Principal components (PCs) were generated for each dataset using PLINK (version 1.9). In order to adjust for factors accounting to possible population substructure, the top 20 generated eigenvectors from the PC analysis, age, sex, and prevalence were used as basic covariates. Disease prevalence standardized for age and gender based on epidemiological reports was specified at 0.002.^{43–47} Summary statistics from these estimates were produced for all four datasets and were included in the meta-analyses. Random-effects meta-analysis using the residual maximum likelihood method was performed using R (version 3.5.1) package metafor to calculate *p*-values and generate forest plots.⁴⁸

Risk profiles versus disease status and age at onset

Previous risk profiling methods have calculated polygenic risk scores (PRS) using only variants that exhibit genome-wide significant associated with disease risk. However, in the most recent PD meta-analysis, it is shown that using variants at thresholds below genome-wide significance improves genetic predictions of disease risk.^{6,43} Mirroring this workflow, but instead using only variants within gene regions outlined in both the primary and secondary gene lists, the R package PRSice2 was used to carry out PRS profiling in the standard weighted allele dose manner. In addition, PRSice2 performs permutation testing and *p*-value aware LD pruning to facilitate identifying the best *p*-value threshold for variant inclusion to construct the PRS. External summary statistics utilized in this phase of analysis included data from leave-one-out meta-analyses (LOOMAs) that exclude the study in which the PRS was being tested, avoiding overfitting/circularity to some degree. LD clumping was implemented under default settings (window size = 250 kb, $r^2 > 0.1$) and for each dataset 10,000 permutations of phenotype-swapping were used to generate empirical *p*-value estimates for each GWAS derived *p*-value threshold ranging from 5E-08 to 0.5, at a minimum increment of 5E-08. Each permutation test in each dataset provided a Nagelkerke's pseudo r^2 after adjustment for an estimated prevalence of 0.005 and study-specific PCs 1–5, age and sex as covariates. GWAS derived *p*-value threshold with the highest pseudo r^2 was selected for further analysis. Summary statistics were meta-analyzed using random effects (REML) per study-specific dataset using PRSice2.⁴⁹ For the age at onset risk profiling, the same workflow was followed, however instead, age at onset was used as a continuous variable, as previously reported.³⁸ To remove possible confounders that could possibly drive a false association with age we removed *APOE* and *FOXO3*, which are general markers for aging.⁵⁰

Mendelian randomization to explore possible causal effect of mitochondria function genes

Both mitochondria gene lists were combined, and all genes already associated with PD (i.e., that have been identified to be in LD with PD risk loci in the last meta-analysis) were removed, leaving 1432 unique mitochondria function gene regions.

We utilized four large-scale methylation and expression datasets through the summary-data-based Mendelian randomization (SMR) (<http://cns.genomics.com/software/smr>) resource. Summary statistics were compared to PD outcome summary statistics for the mitochondria variants of interest (extracted from refs.^{4–6,30–34}) to identify possible association using the R package TwoSampleMR.

Tissues were selected based on their relevance to the pathobiology of PD, which ultimately consisted of tissues from ten brain regions, whole blood, skeletal muscle, and nerve (a full list of the tissues utilized can be found in Supplementary Table 5). For the methylation QTLs “middle age” data was used, which was the oldest available time point. For each mitochondria function variant of interest (considered here the instrumental variable), Wald ratios were generated for each variable that tagged a cis-QTL (probes within each gene and meeting a QTL *p*-value of at least 5E-08 in the original QTL study) and for a methylation or expression probe

with a nearby gene. Using the default SMR protocols, linkage pruning and clumping were implemented. Finally, for each dataset *p*-values were adjusted by false discovery rate to account for multiple testing.

Co-expression network analysis

Co-expression network analysis was used to determine whether mitochondrial genes associated with PD using the SMR analysis are co-expressed with genes associated with monogenic forms of PD in human brain. This analysis was performed by using GTEx V6 gene expression data⁵¹ to generate co-expression networks for each of the 13 brain tissues included within the GTEx study. The raw FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values were corrected for known batch effects, age at death, sex and post-mortem interval, as well as unknown effects. The unknown effects were detected with the Surrogate Variable Analysis (SVA) R Package⁵² and correction was performed using ComBat⁵³. The resulting residuals were used to create a signed network using the blockwiseConsensusModules R function from the WGCNA R package⁵⁴ for each of the 13 tissues. Next, the modules obtained in each of the 13 networks were assigned to cell types using the user List Enrichment R function implemented in the WGCNA R package, which measures enrichment between module-assigned genes and defined brain-related lists using a hypergeometric test. The same approach was used to annotate modules with Gene Ontology, REACTOME⁵⁵ and KEGG⁵⁶ terms.

Expression-weighted cell-type enrichment (EWCE): evaluating enrichment of mitochondrial genes associated with PD risk

Expression-weighted cell-type enrichment (EWCE) (<https://github.com/NathanSkene/EWCE>) (EWCE) was used to determine whether mitochondrial genes associated with PD using the MR analysis have higher expression within a particular cell type than expected by chance. The input for the analysis was (1) neuronal and glial clusters of the central nervous system (CNS) identified in the Linnarsson single-cell RNA sequencing dataset (amounting to a subset of 114 of the original 265 clusters identified) (<http://mousebrain.org/>) and (2) our list of mitochondrial genes highlighted through the MR analysis (see Supplementary Table 6 for the full list of Linnarsson CNS neuronal clusters used).

For each gene in the Linnarsson dataset, cell-type specificity was estimated (the proportion of a gene's total expression in one cell-type compared to all cell types) using the “generate.celltype.data” function of the EWCE package. EWCE with the target list was run with 100,000 bootstrap replicates, which were sampled from a background list of genes that excluded all genes without a 1:1 mouse:human ortholog. In addition, transcript length and GC-content biases were controlled for by selecting bootstrap lists with comparable properties to the target list. The analysis was performed using major cell-type classes (e.g., “telencephalon inhibitory interneurons”, “telencephalon projecting excitatory neurons”, etc.) and subtypes of these classes (e.g., TEINH6 [“Interneuron-selective interneurons, cortex/hippocampus”], TEINH7 [“Interneuron-selective interneurons, hippocampus”] etc.). Data are displayed as standard deviations from the mean, and any values < 0, which reflect a depletion of expression, are displayed as 0. *p*-values were corrected for multiple testing using the Benjamini–Hochberg method over all cell types and gene lists displayed.

OMIM data

Phenotype relationships and clinical synopses of all OMIM genes were downloaded from <http://api.omim.org> on the 29th of May 2018. OMIM genes were filtered to exclude provisional, non-disease and susceptibility phenotypes retaining 2898 unique genes that were confidently associated to 4034 Mendelian diseases. The phenotypic information relating to all genes associated with mitochondrial disorders was collated.

DATA AVAILABILITY

For all datasets included in this study GWAS summary statistics are available at: <https://drive.google.com/file/d/1FZ9UL99LAqyWnyNBx6qOUIfAnubIN/view?usp=sharing> Underlying participant level IPDGC data is available to potential collaborators, please contact ipdgc.contact@gmail.com.

ACKNOWLEDGEMENTS

This work was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services; project ZO1 AG000949.

AUTHOR CONTRIBUTIONS

Data generation and analysis: K.J.B., I.A.B., S.B.-C., C.D., J.A.B., R.H.R., D.Z., M.A.S., C.B., Z. G.-O., J.R.B., M.A.N., A.S., M.R., S.K. Study concept and design: K.J.B., I.A.B., S.B.-C., J.Q., V. J.B., C.D., J.A.B., R.H.R., M.A.S., J.R.B., M.A.N., A.S., M.R., S.K. Critical revision of the manuscript for important intellectual content: all authors.

ADDITIONAL INFORMATION

Supplementary Information accompanies the paper on the *npj Parkinson's Disease* website (<https://doi.org/10.1038/s41531-019-0080-x>).

Competing interests: Mike A. Nalls' participation is supported by a consulting contract between Data Tecnica International and the National Institute on Aging, NIH, Bethesda, MD, USA, as a possible conflict of interest Dr. Nalls also consults for Neuro23 Inc, Lysosomal Therapeutics Inc, the Michael J. Fox Foundation, Illumina Inc and Vivid Genomics among others. Dr. Gan-or also consults for Lysosomal Therapeutics Inc., Denaly, Prevail Therapeutics, Idorsia, and Allergan. The other authors declare no competing interests.

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

REFERENCES

- Singleton, A. & Hardy, J. The evolution of genetics: Alzheimer's and Parkinson's Diseases. *Neuron* **90**, 1154–1163 (2016).
- Gasser, T. Genetics of parkinson's disease. *Ann. Neurol.* **44**, S53–S57 (1998).
- Reeve, A., Simcox, E. & Turnbull, D. Ageing and Parkinson's disease: why is advancing age the biggest risk factor? *Ageing Res. Rev.* **14**, 19–30 (2014).
- Chang, D. et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* **49**, 1511–1516 (2017).
- Nalls, M. A. et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).
- Nalls, M. A. et al. Parkinson's disease genetics: identifying novel risk loci, providing causal insights and improving estimates of heritable risk (2018). <https://doi.org/10.1101/388165>
- Robak, L. A. et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain* **140**, 3191–3203 (2017).
- Hardy, J. Genetic analysis of pathways to Parkinson disease. *Neuron* **68**, 201–206 (2010).
- Billingsley, K. J., Bandres-Ciga, S., Saez-Atienzar, S. & Singleton, A. B. Genetic risk factors in Parkinson's disease. *Cell Tissue Res.* **373**, 9–20 (2018).
- Langston, J. W., Ballard, P., Tetrud, J. W. & Irwin, I. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. *Science* **219**, 979–980 (1983).
- Langston, J. W. & Ballard, P. A. Parkinson's disease in a chemist working with 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine. *N. Engl. J. Med.* **309**, 310–310 (1983).
- Parker, W. D., Parks, J. K. & Swerdlow, R. H. Complex I deficiency in Parkinson's disease frontal cortex. *Brain Res.* **1189**, 215–218 (2008).
- Canet-Avilés, R. M. et al. The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfenic acid-driven mitochondrial localization. *Proc. Natl Acad. Sci. USA* **101**, 9103–9108 (2004).
- Funayama, M. et al. CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. *Lancet Neurol.* **14**, 274–282 (2015).
- Burchell, V. S. et al. The Parkinson's disease-linked proteins Fbxo7 and Parkin interact to mediate mitophagy. *Nat. Neurosci.* **16**, 1257–1265 (2013).
- Lesage, S. et al. Loss of VPS13C function in autosomal-recessive Parkinsonism causes mitochondrial dysfunction and increases PINK1/Parkin-dependent mitophagy. *Am. J. Hum. Genet.* **98**, 500–513 (2016).
- Gorman, G. S. et al. Mitochondrial diseases. *Nat. Rev. Dis. Prim.* **2**, 16080 (2016).
- Goebel, H. H., Heipertz, R., Scholz, W., Iqbal, K. & Tellez-Nagel, I. Juvenile Huntington chorea: Clinical, ultrastructural, and biochemical studies. *Neurology* **28**, 23–23 (1978).
- Carmo, C., Naia, L., Lopes, C. & Rego, A. C. Mitochondrial Dysfunction in Huntington's Disease. In *Advances in Experimental Medicine and Biology* 59–83 (2018).

- Atsumi, T. The ultrastructure of intramuscular nerves in amyotrophic lateral sclerosis. *Acta Neuropathol.* **55**, 193–198 (1981).
- Sasaki, S. & Iwata, M. Mitochondrial alterations in the spinal cord of patients with sporadic amyotrophic lateral sclerosis. *J. Neuropathol. Exp. Neurol.* **66**, 10–16 (2007).
- Moreira, P. I., Cardoso, S. M., Santos, M. S. & Oliveira, C. R. The key role of mitochondria in Alzheimer's disease. *J. Alzheimers Dis.* **9**, 101–110 (2006).
- Nunomura, A. et al. Oxidative damage is the earliest event in Alzheimer disease. *J. Neuropathol. Exp. Neurol.* **60**, 759–767 (2001).
- Moreira, P. I., Duarte, A. I., Santos, M. S., Cristina Rego, A. & Oliveira, C. R. An integrative view of the role of oxidative stress, mitochondria and insulin in Alzheimer's disease. *J. Alzheimers Dis.* **16**, 741–761 (2009).
- Fang, E. F. et al. Mitophagy inhibits amyloid- β and tau pathology and reverses cognitive deficits in models of Alzheimer's disease. *Nat. Neurosci.* **22**, 401–412 (2019).
- Nalls, M. A. et al. Genetic risk and age in Parkinson's disease: Continuum not stratum. *Mov. Disord.* **30**, 850–854 (2015).
- Escott-Price, V. et al. Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann. Neurol.* **77**, 582–591 (2015).
- Lill, C. M. et al. Impact of Parkinson's disease risk loci on age at onset. *Mov. Disord.* **30**, 847–850 (2015).
- Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
- International Parkinson Disease Genomics Consortium et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641–649 (2011).
- Satake, W. et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat. Genet.* **41**, 1303–1307 (2009).
- Simón-Sánchez, J. et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.* **41**, 1308–1312 (2009).
- Pihlstrøm, L. et al. Supportive evidence for 11 loci from genome-wide association studies in Parkinson's disease. *Neurobiol. Aging* **34**, 1708.e7–1708.e13 (2013).
- Lill, C. M. et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genet.* **8**, e1002548 (2012).
- Hill-Burns, E. M. et al. Identification of genetic modifiers of age-at-onset for familial Parkinson's disease. *Hum. Mol. Genet.* **25**, 3849–3862 (2016).
- Brockmann, K. et al. SNCA: major genetic modifier of age at onset of Parkinson's disease. -PubMed- NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/23674386>. (Accessed 1 Mar 2019)
- Gaare, J. J. et al. Rare genetic variation in mitochondrial pathways influences the risk for Parkinson's disease. *Mov. Disord.* **33**, 1591–1600 (2018).
- Blauwendraat, C. et al. Parkinson disease age of onset GWAS: defining heritability, genetic loci and a-synuclein mechanisms. (2018). <https://doi.org/10.1101/424010>
- Huang, Y. et al. SNCA Gene, but Not MAPT, influences onset age of Parkinson's disease in Chinese and Australians. *BioMed. Res. Int.* **2015**, 1–6 (2015).
- 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
- Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
- Keller, M. F. et al. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum. Mol. Genet.* **21**, 4996–5009 (2012).
- Wirdefeldt, K., Gatz, M., Reynolds, C. A., Prescott, C. A. & Pedersen, N. L. Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiol. Aging* **32**, 1923.e1–8 (2011).
- Gasser, T. Genetics of Parkinson's disease. *Curr. Opin. Neurol.* **18**, 363–369 (2005).
- Wickremaratne, M. M. et al. Prevalence and age of onset of Parkinson's disease in Cardiff: a community based cross sectional study and meta-analysis. *J. Neurol. Neurosurg. Psychiatry* **80**, 805–807 (2009).
- Porter, B., Macfarlane, R., Unwin, N. & Walker, R. The prevalence of Parkinson's disease in an area of North Tyneside in the north-east of England. *Neuroepidemiology* **26**, 156–161 (2006).
- Viechtbauer, W. Conducting meta-analyses in R with the meta for Package. *J. Stat. Softw.* **36**, 1–48 (2010).
- Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2014).
- Broer, L. et al. GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *J. Gerontol. A Biol. Sci. Med. Sci.* **70**, 110–118 (2015).
- Carithers, L. J. et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv. Biobank* **13**, 311–319 (2015).

52. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, 1724–1735 (2007).
53. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
54. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
55. Fabregat, A. et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
56. 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).
57. Qi, T. et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282 (2018).
58. Westra, H.-J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).



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) 2019

INTERNATIONAL PARKINSON'S DISEASE GENOMICS CONSORTIUM (IPDGC)

A. Noyce¹³, A. Tucci¹⁴, B. Middlehurst¹, D. Kia¹⁵, M. Tan¹⁶, H. Houlden¹⁴, H. R. Morris¹⁶, H. Plun-Favreau¹⁴, P. Holmans¹⁷, J. Hardy¹⁴, D. Trabzunj^{14,18}, J. Bras¹⁹, K. Mok¹⁴, K. Kinghorn²⁰, N. Wood¹⁵, P. Lewis²¹, R. Guerreiro^{14,19}, R. Loring²², L. R'Bibo¹⁴, M. Rizig¹⁴, V. Escott-Price^{23,24}, V. Chelban¹⁴, T. Foltynie⁶, N. Williams²⁴, A. Brice²⁵, F. Danjou²⁵, S. Lesage²⁵, M. Martinez²⁶, A. Giri^{27,28}, C. Schulte^{27,28}, K. Brockmann^{27,28}, J. Simón-Sánchez^{27,28}, P. Heutink^{27,28}, P. Rizzu²⁸, M. Sharma²⁹, T. Gasser^{27,28}, A. Nicolas², M. Cookson², F. Faghri^{2,30}, D. Hernandez², J. Shulman^{31,32}, L. Robak³³, S. Lubbe³⁴, S. Finkbeiner^{35,36,37}, N. Mencacci³⁸, C. Lungu³⁹, S. Scholz⁴⁰, X. Reed², H. Leonard², G. Rouleau⁷, L. Krohan⁴¹, J. van Hilten⁴², J. Marinus⁴², A. Adames-Gómez⁴³, M. Aguilar⁴⁴, I. Alvarez⁴⁴, V. Alvarez⁴⁴, F. Javier Barrero⁴⁶, J. Bergareche Yarza⁴⁷, I. Bernal-Bernal⁴³, M. Blazquez⁴⁵, M. Bonilla-Toribio Bernal⁴³, M. Boungiorno⁴⁴, Dolores Buiza-Rueda⁴³, A. Cámara⁴⁸, M. Carcel⁴⁴, F. Carrillo⁴³, M. Carrión-Claro⁴³, D. Cerdan⁴⁹, J. Clarimón^{50,51}, Y. Compta⁴⁸, M. Diez-Fairen⁴⁴, O. Dols-Icardo^{50,51}, J. Duarte⁴⁹, R. I. Duran⁵², F. Escamilla-Sevilla⁵³, M. Ezquerro⁴⁸, M. Fernández⁴⁸, R. Fernández-Santiago⁴⁸, C. García⁴⁵, P. García-Ruiz⁵⁴, P. Gómez-Garre⁴³, M. Gomez Heredia⁵⁵, I. Gonzalez-Aramburu⁵⁶, A. Gorostidi Pagola⁵⁷, J. Hoenicka⁵⁸, J. Infante^{51,56}, S. Jesús⁴³, A. Jimenez-Escrig⁵⁹, J. Kulisevsky^{51,60}, M. Labrador-Espinosa⁴³, J. Lopez-Sendon⁵⁹, A. López de Munain Arregui⁵⁹, D. Macias⁴³, I. Martínez Torres⁶¹, J. Marín^{51,60}, M. Jose Marti⁴⁸, J. Martínez-Castrillo⁵⁹, C. Méndez-del-Barrio⁴³, M. Menéndez González⁴³, A. Mínguez⁵³, P. Mir⁴³, E. Mondragon Rezola⁵⁷, E. Muñoz⁴⁸, J. Pagonabarraga^{51,60}, P. Pastor⁴⁴, F. Perez Errazquin⁵⁵, T. Periñán-Tocino⁴³, J. Ruiz-Martínez⁵⁷, C. Ruz⁵², A. Sanchez Rodriguez⁵⁶, M. Sierra⁵⁶, E. Suarez-Sanmartin⁴, C. Taberner⁵⁹, J. Pablo Tartari⁴⁴, C. Tejera-Parrado⁴³, E. Tolosa⁴⁸, F. Valdeoriola⁴⁸, L. Vargas-González⁴³, L. Vela⁶², F. Vives⁵², A. Zimprich⁶³, L. Pihlstrom⁶⁴, P. Taba⁶⁵, K. Majamaa^{66,67}, A. Siitonen⁶⁶, N. Okubadejo⁶⁸ and O. Ojo⁶⁸

¹³Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, QMUL, London, UK; ¹⁴Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK; ¹⁵UCL Genetics Institute; and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK; ¹⁶Department of Clinical Neuroscience, University College London, London, UK; ¹⁷Biostatistics & Bioinformatics Unit, Institute of Psychological Medicine and Clinical Neuroscience, MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff, UK; ¹⁸Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh 11211, Saudi Arabia; ¹⁹UK Dementia Research Institute at UCL and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK; ²⁰Institute of Healthy Ageing, University College London, London, UK; ²¹University of Reading, Reading, UK; ²²University College London, London, UK; ²³MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff, UK; ²⁴Cardiff University School of Medicine, Cardiff, UK; ²⁵Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS, UMR 7225, Sorbonne Universités, UPMC University Paris 06, UMR S 1127, AP-HP, Pitié-Salpêtrière Hospital, Paris, France; ²⁶INSERM UMR 1220; and Paul Sabatier University, Toulouse, France; ²⁷Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany; ²⁸DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany; ²⁹Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tübingen, Tübingen, Germany; ³⁰Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA; ³¹Departments of Neurology, Neuroscience, and Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA; ³²Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA; ³³Baylor College of Medicine, Houston, TX, USA; ³⁴Ken and Ruth Davee Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; ³⁵Departments of Neurology and Physiology, University of California, San Francisco, CA, USA; ³⁶Gladstone Institute of Neurological Disease, San Francisco, CA, USA; ³⁷TauBe/Koret Center for Neurodegenerative Disease Research, San Francisco, CA, USA; ³⁸Northwestern University Feinberg School of Medicine, Chicago, IL, USA; ³⁹National Institutes of Health Division of Clinical Research, NINDS, National Institutes of Health, Bethesda, MD, USA; ⁴⁰Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA; ⁴¹Department of Human Genetics, McGill University, Montréal, QC H3A 0G4, Canada; ⁴²Department of Neurology, Leiden University Medical Center, Leiden, Netherlands; ⁴³Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain; ⁴⁴Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona, Spain; ⁴⁵Hospital Universitario Central de Asturias, Oviedo, Spain; ⁴⁶Hospital Universitario Parque Tecnológico de la Salud, Granada, Spain; ⁴⁷Instituto de Investigación Sanitaria Biodonostia, San Sebastián, Spain; ⁴⁸Hospital Clinic de Barcelona, Barcelona, Spain; ⁴⁹Hospital General de Segovia, Segovia, Spain; ⁵⁰Memory Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain; ⁵¹Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain; ⁵²Centro de Investigación Biomédica, Universidad de Granada, Granada, Spain; ⁵³Hospital Universitario Virgen de las Nieves, Instituto de Investigación Biosanitaria de Granada, Granada, Spain; ⁵⁴Instituto de Investigación Sanitaria Fundación Jiménez Díaz, Madrid, Spain; ⁵⁵Hospital Universitario Virgen de la Victoria, Malaga, Spain; ⁵⁶Hospital Universitario Marqués de Valdecilla-IDIVAL, Santander, Spain; ⁵⁷Instituto de Investigación Sanitaria Biodonostia, San Sebastián, Spain; ⁵⁸Institut de Recerca Sant Joan de Déu, Barcelona, Spain; ⁵⁹Hospital Universitario Ramón y Cajal Madrid, Madrid, Spain; ⁶⁰Movement Disorders Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain; ⁶¹Department of Neurology, Instituto de Investigación Sanitaria La Fe, Hospital Universitario y Politécnico La Fe, Valencia, Spain; ⁶²Department of Neurology, Hospital Universitario Fundación Alcorcón, Madrid, Spain; ⁶³Department of Neurology, Medical University of Vienna, Vienna, Austria; ⁶⁴Department of Neurology, Oslo University Hospital, Oslo, Norway; ⁶⁵Department of Neurology and Neurosurgery, University of Tartu, Tartu, Estonia; ⁶⁶Institute of Clinical Medicine, Department of Neurology, University of Oulu, Oulu, Finland; ⁶⁷Department of Neurology and Medical Research Center, Oulu University Hospital, Oulu, Finland and ⁶⁸University of Lagos, Yaba, Lagos State, Nigeria