

AWARD NUMBER: W81XWH-17-1-0249

TITLE: Early-Onset Parkinson's Disease Is a Mitochondrial Disease: A Nigral Mitochondrial Cytopathy

PRINCIPAL INVESTIGATOR: Owen A. Ross, PhD

CONTRACTING ORGANIZATION: Mayo Clinic Jacksonville
Jacksonville, Florida 32224

REPORT DATE: Sept 2019

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE Sept 2019		2. REPORT TYPE Annual report		3. DATES COVERED 08/15/2018 - 08/14/2019	
4. TITLE AND SUBTITLE Early-Onset Parkinson's Disease Is a Mitochondrial Disease: A Nigral Mitochondrial Cytopathy				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-17-1-0249	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Owen A. Ross, PhD E-Mail:				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND Mayo Clinic Jacksonville Jacksonville, Florida 32224				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Loss of function mutations in the genes encoding PINK1 and Parkin results in early-onset forms of Parkinson's disease (EOPD). Both enzymes are functionally linked and together direct a neuroprotective mitochondrial quality control (mitoQC) ensuring elimination of damaged organelles from cells via the autophagy-lysosome system (i.e. mitophagy), which is lost in EOPD. Given the complexity of this pathway and the general missing heritability in EOPD, it is highly likely that additional genes regulating this pathway may also be found mutated in EOPD. The overarching goals of this project are to 1) identify high confidence genetic modifiers of the PINK1/Parkin pathway by a two-tiered functional screening (overlay of genome-wide siRNA and miRNA screens) in cells, 2) to identify the underlying genetic variation and characterize the EOPD genome (whole-genome-sequencing of patients), as well as 3) to determine the pathogenicity of these novel EOPD sequence variants in functional readout studies. Using this combined functional genetics approach we will determine the regulation of mitophagy as well as the genetic architecture of EOPD.					
15. SUBJECT TERMS early-onset Parkinson's disease, mitochondrial quality control, mitophagy, PINK1, Parkin, functional genomic screening, genetic architecture					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Unclassified	18. NUMBER OF PAGES 9	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code)

Table of Contents

	<u>Page</u>
1. Introduction.....	4
2. Keywords.....	4
3. Accomplishments.....	4
4. Impact.....	9
5. Changes/Problems.....	10
6. Products, Inventions, Patent Applications, and/or Licenses.....	10
7. Participants & Other Collaborating Organizations.....	10
8. Special Reporting Requirements.....	12
9. Appendices.....	13

1. **INTRODUCTION:** Loss of function mutations in the genes encoding PINK1 and Parkin results in early-onset forms of Parkinson's disease (EOPD). Both enzymes are functionally linked and together direct a neuroprotective mitochondrial quality control (mitoQC) ensuring elimination of damaged organelles from cells via the autophagy-lysosome system (i.e. mitophagy), which is lost in EOPD. Given the complexity of this pathway and the general missing heritability in EOPD, it is highly likely that additional genes regulating this pathway may also be found mutated in EOPD. The overarching goals of this project are to 1) identify high confidence genetic modifiers of the PINK1/Parkin pathway by a two-tiered functional screening (overlay of genome-wide siRNA and miRNA screens) in cells, 2) to identify the underlying genetic variation and characterize the EOPD genome (whole-genome-sequencing of patients), as well as 3) to determine the pathogenicity of these novel EOPD sequence variants in functional readout studies. Using this combined functional genetics approach we will determine the regulation of mitophagy as well as the genetic architecture of EOPD.
2. **KEYWORDS:** early-onset Parkinson's disease, mitochondrial quality control, mitophagy, PINK1, Parkin, functional genomic screening
3. **ACCOMPLISHMENTS:**
 - **What were the major goals of the project?**
Major Task 1: Nomination of mitoQC candidate genes by an accelerated, two-tiered functional screen and processing through bioinformatics resource/filtering strategy – Month 1-18

Major Task 2: Whole-Genome sequencing in patients with EOPD and nomination of disease genes/variants – Month 1-36

Major Task 3: Validation of high-confidence mitoQC/EOPD genes and dysfunctions of sequence variants on molecular, cellular, and organismal level – Month 6-36
 - **What was accomplished under these goals?**
Towards accomplishing major goals of the project, we first completed the regulatory review and approval by Mayo Clinic IRB and the HRPO (subtask 1.1); and this has been renewed for Year 2 (Appendix 1). We have continued to screen newly collected early-onset patients with Parkinson's disease to identify those that are negative for *PINK1* and *PARKIN* mutations, and have prioritized these subjects for whole-genome sequencing. In Year 1, we identified all early-onset patients for which we have collected fibroblast biopsies and grown the cell lines and extracted DNA. In Year 2 we sequenced these cell lines and are analyzing the sequence variants identified. We plan to functionally assess the influence of variants identified within these subjects on the efficiency of the mitophagy pathway working with Dr. Springer (Co-PI). We currently have whole-genome sequencing for 95 patients with early-onset disease.

Known Parkinsonism Genes

Variants within the well-known and published parkinsonism genes (*SNCA*, *PINK1*, *PRKN*, *LRRK2*, *PARK7*, *VPS13C* and *GBA*) were filtered to identify individuals with variants in these genes. A total of 13,272 variants within those genes were present in the sequences (3'-UTR variants: 94, 5'-UTR variants: 12, intron variants : 13,070, missense variants: 60, splice-region variants: 7, synonymous: 27, stop gained: 1).

From the Mayo sequenced individuals, 36 carry coding or loss of function variants in one or more of the listed genes (Table 1). Two individuals carried homozygous common *PRKN* variants (PRKNp.Ser167Asn and PRKNp.Arg444Asn) and were not considered the

pathogenic cause; one individual carries two novel heterozygous mutations in *VPS13C* (*VPS13C*p.Ala1687Val and *VPS13C*p.Leu216 Val); one individual carries the published PD risk variant *LRRK2*p.Arg1628Pro, and five individuals carry the known PD risk factor *GBAp*.Leu483Pro and this can influence earlier onset of disease.

Other interesting findings include a patient who is heterozygous to novel variants in the previously published gene *SYNJ1* (p.Arg1328Pro and p.Val1405_Leu1406insAsnThr) who also carries a variant in *PARK7* (p.Val20Ala). See the below table for more details (Table1).

Table 1. Individuals carrying variants in previously published EOPD genes (highlighted ones probably have the disease explained)

Sample ID	Gender	Age of onset (years)	Race	Family History	Known PD genes variants
1487-1	F	40	White	Y	<i>GBAp</i> .Leu483Pro het
1622-1	F	35	White	Y	<i>SYNJ1</i> p.Arg1328Pro het; <i>SYNJ1</i> p.Val1405_Leu1406insAsnThr het <i>PARK7</i> p. Val20ala het
3021-1	M	33	White	Y	<i>GBAp</i> .Leu483Pro het
3021-10	F	50	White	Y	<i>GBAp</i> .Leu483Pro het
5541-1	F	34	White	Y	<i>GBAp</i> .Leu483Pro het
6634-1	M	37	White	N	<i>PINK1</i> p.Val184Met het <i>GBAp</i> .Asp448His het
6636-1	F	43	White		<i>PRKN</i> p.Met192Leu het
6743-1	F	44	White		<i>VPS13C</i> p.Ala1687Val het; <i>VPS13C</i> p.Leu216Val het
552-1	F	39	White	Y	<i>PINK1</i> p.Gln115Leu het
707-3	M	34	White	Y	<i>PINK1</i> p.Gln115Leu het
1200-1	F	40	White	Y	<i>GBAp</i> .Glu365Lys het; <i>GBAp</i> .Asp179His het
1256-1	M	43	White	N	<i>VPS13C</i> p.Arg470His het
1263-1	M	40	White	AD	<i>PRKN</i> p.Ser167Asn het <i>VPS13C</i> p.Ile2789Thr het; <i>VPS13C</i> p.Ile1132Val het; <i>VPS13C</i> p.Arg153His het
1308-1	M	41	White	Y	<i>GBAp</i> .Asn409Ser <i>PRKN</i> p.Ser167Asn homozygous
1560-1	F	37	White	Y	<i>GBAp</i> .Glu365Lys het
5167-1	F	37	White	N	<i>PINK1</i> p.Gln115Leu het
5274-1	F	37	White	Y	<i>GBAp</i> .Arg202Ter het <i>PINK1</i> p.Gly411Ser het
5311-1	M	36	White	No	<i>PRKN</i> p.Arg256Cys het
5400-1	M	34	White	Y	<i>VPS13C</i> splicing c.684+7T>C het
5536-1	F	40	White	N	<i>GBAp</i> .Ala483Arg het
5779-1	M	45	White	Y	<i>GBAp</i> .Ala495Pro het ; <i>GBAp</i> .483Pro het
5891-1	F	41	White	N	<i>GBAp</i> .Glu365Lys het
5980-1	M	36	White	Y	<i>PRKN</i> p.Arg256Cys het
6179-1	M	38	White	Y	<i>PRKN</i> p.Arg275Gln het
6375-1	M	45	White	Y	<i>GBAp</i> .Glu365Lys het
6560-1	F	36	White	Y	<i>LRRK2</i> p.Gly2019Ser het
7200-1	F	47	White	N	<i>PINK1</i> p.Gln115Leu het

Novel and rare Loss-of-function variants

A total of 3,890 Loss of function (LoF) variants were detected, of which 694 are singletons and 1,461 are rare (<0.01 or absent from GnomAD database). Some of those LoF variants with a CADD score >36 are shown (>20 is predicted to be in the top 1% of deleterious variants) are listed on Table 2.

Table 2. Rare (<0.01 allele frequency in GnomAD) or singleton loss of function variants

ID	Sex	Onset (years)	Gene	aa change	CADD	Freq GnomAD
767-3	M	35	<i>BCAT2</i>	p.Tyr200Ter	37	<0.01
1487-1	F	40	<i>CNTN6</i>	p.Glu924Ter	49	
1651-2	F	43	<i>CPNE5</i>	p.Glu266Ter	45	
1487-1	F	40	<i>DOCK6</i>	p.Gln1518Ter	38	
1951-1	M	29	<i>FMO4</i>	p.Lys42Ter	38	<0.01
597-1	M	43	<i>IFIH1</i>	p.Glu627Ter	38	<0.01
1202-1	M	39	<i>MMP21</i>	p.Leu546Ter	39	
5136-1	F	34	<i>NMNAT3</i>	p.Trp184Ter	41	<0.01
5704-1	M	40	<i>PCDHB7</i>	p.Glu632Ter	42	<0.01
767-3	M	35	<i>PLSCR5</i>	p.Arg129Ter	42	<0.01
5948-1	M	41	<i>SPAG1</i>	p.Gln672Ter	38	<0.01
1487-1	F	40	<i>TAOK1</i>	p.Arg698Ter	39	
1951-1	M	29	<i>TMEM45A</i>	p.Arg36Ter	37	<0.01
1695-1	F	38	<i>TSFM</i>	p.Leu207Ter	39	<0.01
5619-1	F	37	<i>ULK4</i>	p.Arg862Ter	52	<0.01

Novel and rare non-synonymous variants

Lists of variants that were absent or have low allele frequency (<0.01) in the public database GnomAD and CADD scores >25 were generated for each individual. We attempted to identify potential detrimental rare homozygous variants or compound heterozygous. A total of 42,091 missense variants were detected and 5,675 variants have CADD scores > 25 and some examples those variants are highlighted on Table 3.

Table 3. Examples of rare missense variants (allele freq < 0.01 or singletons) with CADD scores > 30

ID	Sex	Onset (y)	Gene	aa change	CADD	Freq GnomAD
5944-1	M	41	<i>CSMD3</i>	p.Pro1233Leu	33	na
6323-1	M	45	<i>FOXO6</i>	p.Lys190Thr	32	na
5167-1	F	37	<i>GMPR2</i>	p.Gly201Val	35	<0.01
552-1	F	39	<i>KCNT1</i>	p.Arg910Gln	35	<0.01
5676-1	M	41	<i>RPS27A</i>	p.Ser57Pro	33	<0.01
5255-1	M	40	<i>TOMM5</i>	p.Pro9Arg	33	<0.01
5980-1	M	36	<i>TRPM3</i>	p.Ser453Leu	33	<0.01

Novel or rare variants in mitochondrial quality control genes

We extracted variants within the genes of the mitochondrial quality control pathways to feed into functional studies to assess pathogenicity and clinical relevance. A list with 1,346 genes with the GO term “mitophagy” was obtained from Genecards, and a total of 3,116 are missense and 171 are LoF (includes frameshift, splicing, stop gain and stop loss variants) variants in 713 of those genes were detected. A few examples of variants within genes that may be involved in the mitophagy pathway are summarized on Table 4.

Table 4. Rare variants (<0.01 in GnomAD) within genes that may participate on the mitophagy pathway (CADD scores > 20)

ID	Sex	Onset (years)	Gene	aa change	CADD	Freq GnomAD
501-1	M	37	<i>TRPM2</i>	p.Leu1034Phe	26.2	<0.01
1308-1	M	41	<i>TRPM2</i>	p.Gln953Ter	58	<0.01
6636-1	F	43	<i>TRPM2</i>	p.Arg411Trp	26.7	<0.01
1200-1	F	40	<i>TRAP1</i>	p.Arg47Ter	35	<0.01
5256-1	F	35	<i>SLC25A13</i>	p.Gln312Ter	35	<0.01
5834-1	M	48	<i>MFN2</i>	p.Thr362Met	26.4	<0.01
1776-1	F	43	<i>USP30</i>	p.Ser171Leu	26.3	<0.01
5430-1	F	43	<i>ATG13</i>	p.Asp490Asn	26.8	<0.01
5676-1	M	41	<i>CLEC16A</i>	p.Arg860Cys	25.9	<0.01
6179-1	M	38	<i>TOMM40</i>	p.Gly24Glu	23.1	<0.01
6026-1	F	42	<i>VPS13D</i>	p.Ala171Gly	24.6	<0.01
1064-1	M	39	<i>VPS13D</i>	p.Arg3267Trp	27.3	<0.01

Investigating Copy Number Variation (CNV)

We have also prioritized assessing copy number variation (CNV) within the early-onset sequence data set and are working with our bioinformatics colleagues (Dr. Ren) to optimize the pipelines and adapt those to work with multiple programs. CNVs were called using the following CNV detection software: Lumpy, CNVnator, PatternCNV and Wandy and are currently being analyzed. Large structural variants and CNVs detection and annotation require a different and more complex approach than SNVs. There were no CNVs disrupting EOPD gene *PINK1*, and all of the rare CNVs disrupting the *PRKN* gene were successfully detected by all algorithms, with slight differences on breakpoint coordinates.

Each rare CNV event was visually inspected using IGV for validation. Efforts to verify the overlap of the common CNVs among the different software, and to calculate their allele frequencies are underway. This task is much more challenging because the breakpoint coordinates from the same CNV event can be variable from sample to sample in addition to algorithm differences. The CNV annotation database SCAN and the software package SVTtyper and intansv for R are being used to assist on the annotation and frequency calls of these inconsistent events.

Table 5 lists examples of CNV calls within genic regions that may disrupt function or increase gene dosage All CNVs on Table 5 were called using Lumpy. The CNV calls from PatternCNV, CNVnator, Lumpy and Wandy are currently being analyzed.

Table 5. Example of CNVs detected in the Mayo EOPD cohort. CNVs with ID numbers have been previously detected and may be common.

CNV coordinates (GRCh37)	CNV size (bp)	CNV type	Genes	CNV ID
chr1:2886954-3845268	958,314	Deletion	<i>DFFB MEGF6 TP73 KIAA0562 ARHGEF16 WDR8 KIAA0495 LRRC47 PRDM16 TPRG1L ACTRT2 CCDC27 C1orf174</i>	NA
chr2:96823988-97075783	251,795	Deletion	<i>CNNM3 ANKRD39 FAM178B SEMA4C ANKRD23 LOC643445 LOC100128957 LOC100132375</i>	NA
chr3:14444330-14561629	117,299	Deletion	<i>GRIP2</i>	NA
chr6:157691300-164676192	6,984,892	Duplication/ Translocation	<i>ACAT2 IGF2R LPA MAS1 MAP3K4 PARK2 PLG SLC22A1 SLC22A3 SLC22A2 SOD2 TCP1 DYNLT1 EZR SYNJ2 QKI WTAP MRPL18 SNX9 AGPAT4 TULP4 TMEM181 ZDHHC14 RIPPLY2 LPAL2 RSPH3 FNDC1 SERAC1</i>	CNVR3123.1 CNVR3126_full CNVR3130.1 CNVR3135.1 CNVR3136_full CNVR3137.1 CNVR3141.1 CNVR3142.1 CNVR3143.1 CNVR3144.1 CNVR3150.1
chr7:72729346-72738350	9,004	Deletion	<i>DNAJC30</i>	NA
chr21:14338129-16526867	2,188,738	Duplication	<i>HSPA13 NRIP1 USP25 VDACC2P RBM11 POLR2CP SAMSN1 LIPI ABCC13 CYCSP42 LOC388813 C21orf126 LOC100128341....</i>	CNVR7955.1 CNVR7956_full

Non-coding variation in EOPD genes

We also attempted to identify variants that may affect differential gene expression of mitophagy genes such as *PINK1* and *PRKN*. The variants were scored and filtered according to multiple resources. ENCODE was used to determine the genic regions with are predicted to harbor open chromatin in multiple different tissues. Intronic variants located within open chromatin regions are more likely to have a regulatory function in comparison to intronic variants found outside of open chromatin regions. In addition to the ENCODE regions, ATAC-seq data from mice dopaminergic neurons in forebrain (FB) and midbrain (MB) was used to predict regulatory regions specific to dopaminergic neurons by similarity. *PINK1*, for example, is predicted to have 2 main regulatory regions in mice dopaminergic neurons, and 10 cis-regulatory regions according to ENCODE.

The scoring algorithm CADD was used to identify variants with high conservation metrics and protein level scores. Variants with a CADD score higher than 20 are more likely to affect protein sequence and function. However, intronic and synonymous variants with CADD scores higher than 10 may indicate a possible regulatory role based on protein binding motif proximity. Regulomedb is a manually curated scoring algorithm that is based on eQTL evidence, chipseq, TF binding motifs and DNAase peaks. Regulomedb scores equal to 1a have many sources of regulatory evidence, whereas scores of 4 and higher are less likely to affects transcription factor binding or other known functional protein binding motifs. Unlike CADD, Regulomedb does not use sequence conservation as a criteria, so the scores may be discrepant and be carefully interpreted. The GnomAD database was used to access the allele frequency of each variant across the general population. Other resources such as Haploreg, GWAVA, BRAINEAC, GTEx, ORegAnno, Enhancer Atlas, SYNAPSE, FANTOM and ConSite were used to identify known eQTLs and fine tune the top candidate variants.

We plan to identify further key variants that are involved in the development of early-onset PD and play a role in the dysfunction of the mitoQC pathway. These target genes/variants will be assessed for functional effects with the co-PI, Dr. Springer. We will identify other dataset that can add to and expand our sample size focusing on early-onset PD. We have already identified and obtained data from the Michael J. Fox Foundation PPMI cohort and will actively seek other data that may be informative (e.g. PDBP, UK biobank). Increasing the sample size will give us more power to genetically identify disease-relevant genes. We will use a cohort of 1000 whole-genome sequences that we have collected from the Mayo Clinic Biobank cohort. In addition, we have obtained whole-genome sequence data from the Parkinson's Progression Markers Initiative (PPMI) cohort supported by the Michael J. Fox Foundation and are analyzing data to use as a replication cohort for our sequence data. This series and other sequence data we are generating will also allow us to rapidly assess the role of variants/genes in the more frequent late-onset forms of Parkinson's disease that we identify in our studies.

- **What opportunities for training and professional development has the project provided?**

Nothing to Report.

- **How were the results disseminated to communities of interest?**

In Year 2 of the award we published for papers relating to genetic factors that may affect early-onset forms of parkinsonism or altering the penetrance of variation and therefore the age-at-onset of clinical PD signs. McClymont et al. (2018) and Berge-Seidl et al. (2019) investigated genetic factors that may result in an earlier onset of disease. Two other related papers were also published on the role of mtDNA in PD and TREM2 in MSA (early-onset parkinsonism).

We also presented some of the EOPD data at the Mayo Clinic APDA PD research symposium for patients and at a symposium for The Parkinson Association of Greater Daytona Beach (PAGDB) patient group.

- **What do you plan to do during the next reporting period to accomplish the goals?**

We will complete the whole-genome sequencing of 150 early-onset patients with Parkinson's disease. We will extract variants that are predicted deleterious in genes that are established within the mitophagy/autophagy pathways and assess these against our in-house control whole genome sequence data and also through additional datasets we have obtained, e.g. PPMI. We will then plan to work with Dr. Springer to assess the functional impact of the variants within his cellular model of mitophagy. We will employ novel long-read sequence technologies to assess inheritance patterns and haplotype structure for EOPD patients and try to discern complex genetic regions and identify possible repeat disorders/causes of EOPD. The ultimate goal is to identify novel genes for EOPD and clarifying the role of the mitochondrial quality control pathways in disease pathogenesis.

4. **IMPACT:** *Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project relative to:*

- **What was the impact on the development of the principal discipline(s) of the project?**

Nothing to Report.

- **What was the impact on other disciplines?**

Nothing to Report.

- **What was the impact on technology transfer?**

Nothing to Report.

- **What was the impact on society beyond science and technology?**

Nothing to Report.

5. CHANGES/PROBLEMS:

▪ **Changes in approach and reasons for change**

Nothing to Report

▪ **Actual or anticipated problems or delays and actions or plans to resolve them**

We encountered an initial delay of >4 months due to IRB and HRPO approval. This had been solved end by the beginning of 2018. In addition, we have been attempting to establish collaboration with The American Genome Center at the Uniformed Services University in Bethesda. The cost of whole-genome sequencing would be reduced if we are able to perform the whole-genome sequencing there. We have been trying to complete a material transfer agreement and work with Dr. Clifton Dalgard, Ph.D. who is Associate Professor and Director of the Laboratory Core (Collaborative Health Initiative Research Program; CHIRP). These discussions are still on-going and we hope we can establish a productive collaboration. We have begun sequencing with Hudson Alpha as the provider until such time as we can use The American Genome Center at the Uniformed Services University. The time line has only been marginally affected by the above and we are confident to be able to complete all sequencing as scheduled with either center. Update: We have now been using the Mayo Genomics core for the whole-genome sequencing and they have recently acquired a new long-read system that will reduce the charges for this innovative approach (using PacBio Sequel II System).

▪ **Changes that had a significant impact on expenditures**

Nothing to Report

▪ **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

Nothing to Report

6. PRODUCTS:

Nothing to Report at this point.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

▪ **What individuals have worked on the project?**

Name:	<i>Owen A. Ross, PhD</i>
Project Role:	<i>PI</i>
Researcher Identifier (e.g. ORCID ID):	<i>0000-0003-4813-756X</i>
Nearest person month worked:	<i>2.0</i>
Contribution to Project:	<i>Together with the co-PI Dr. Springer, Dr. Ross has supervised the project, collected all regulatory material and ensured all necessary steps towards completion of the milestones</i>
Funding Support:	

Name:	<i>Ana Kolicheski, PhD</i>
Project Role:	<i>Research Fellow</i>
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	<i>3.0</i>
Contribution to Project:	<i>Dr. Kolicheski has coordinated the analysis of the next-generation sequence data in consultation with biostatisticians and bioinformaticians.</i>
Funding Support:	
Name:	<i>Ronald Walton, BSc</i>
Project Role:	<i>Laboratory Technician</i>
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	<i>2.0</i>
Contribution to Project:	<i>Mr. Walton processed and organized all the samples from patients with early-onset Parkinson's disease. He completed the gene screening of PINK1 and PARKIN to identify and exclude carriers of known mutations prior to selection for whole-genome sequencing. He prepares and is responsible for shipping samples for sequencing, retrieval of data and facilitating bioinformatics processing.</i>
Funding Support:	
Name:	<i>Alexandra Soto, MSc</i>
Project Role:	<i>Laboratory Technician</i>
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	<i>4.0</i>
Contribution to Project:	<i>Ms. Soto oversaw additional screening of the EOPD patients for atypical forms of early-onset parkinsonism and is establishing the new protocol for analysis of long-read whole genome sequencing.</i>
Funding Support:	
Name:	<i>Yingxue Ren, PhD</i>
Project Role:	<i>Bioinformatician</i>

Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	1.0
Contribution to Project:	Ms. Ren is working with Ms. Soto in establishing the new protocol for analysis of long-read whole genome sequencing.
Funding Support:	

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

Owen A. Ross, PhD

Changes in Active Support:

New

Title: Molecular Mechanisms of PINK1-PRKN directed mitochondrial quality control

Grant number: RF1 NS085070

Committed Time: 2.40

Supporting Agency: National Institute of Neurological Disorders and Stroke

Contracting/Grants Officer: Beth-Anne Sieber, Program Officer

Performance Period: 09/01/2019-05/31/2024

Level of Funding: \$360,754 (current annual direct costs)

Goals & Specific Aims: To elucidate the modifying roles of heterozygous PINK1-PRKN mutations in Parkinson's disease and related Lewy body disorders

Role: Co-I

Ended

- **What other organizations were involved as partners?**

"Nothing to Report."

8. SPECIAL REPORTING REQUIREMENTS

- **COLLABORATIVE AWARDS:**

For an update on Major Task 1 see report from the co-PI Dr. Springer.

- **QUAD CHARTS:**

See appendix

9. APPENDICES: *Quad chart, publications.*

Early-Onset Parkinson's Disease Is a Mitochondrial Disease: A Nigral Mitochondrial Cytopathy

PR160606P1

W81XWH-17-1-0249



PI: Owen A. Ross, PhD

Org: Mayo Clinic Jacksonville

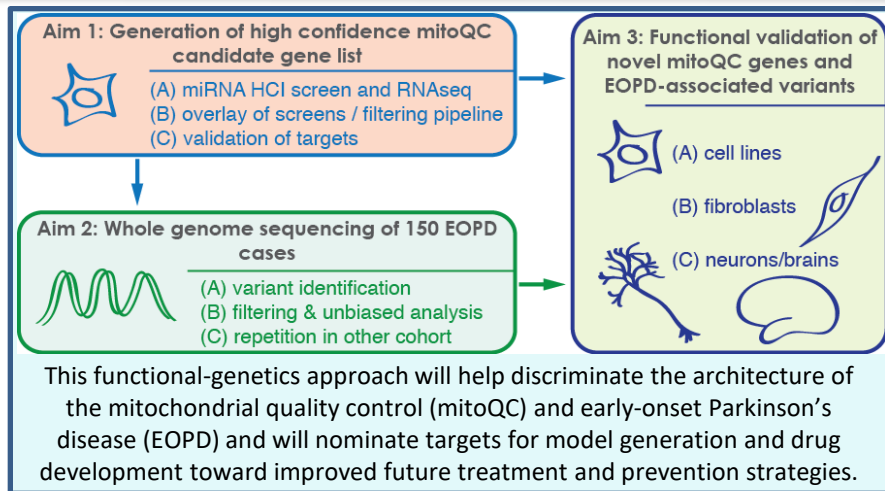
Award Amount: \$1,176,917

Study/Product Aim(s)

- Specific Aim 1: To identify high-confidence genetic modifiers of PINK1/PARK2-directed mitochondrial quality control (mitoQC)
- Specific Aim 2: To identify the underlying genetic variation and characterize the early-onset Parkinson's disease (EOPD) genome
- Specific Aim 3: To determine pathogenicity of novel EOPD sequence variants in functional readout studies

Approach

We hypothesize that EOPD is a mitochondrial disease and that its genetic causes cluster around loss of mitoQC functions resulting in failure to safely dispose of damaged organelles. Our overarching goal is to delineate this pathway and the disease relevance of individual key players and their variants towards rationalized biomarker and drug development. This will be achieved through combining whole-genome-sequencing data from EOPD patients with functional genetic screening of genes/variants.



Accomplishment: Complete regulatory review and approval by Mayo Clinic IRB and by HRPO.

Timeline and Cost

Activities	CY	17	18	19	20
Aim 1: Functional screening					
Aim 2: WGS & analysis					
Aim 3: Validation & pathogenicity					
Estimated Budget (\$K)		\$196	\$392	\$392	\$196

Goals/Milestones

CY17 Goal – Regulatory review and roll-out of study

✓ Complete regulatory review and approval by HRPO

✓ Whole-genome sequencing of 50 EOPD patients

CY18 Goals – First-tier filtering of genomic variants

✓ Whole-genome sequencing of 100 EOPD patients

✓ Critical variant identification of mitoQC candidates

CY19 Goal – Second-tier filtering of genomic variants

□ Whole-genome sequencing of 150 EOPD patients

✓ Critical variant identification with informed and unbiased strategies

CY20 Goal – Third-tier filtering and replication of candidate variants

□ Pathogenic variant identification in EOPD/mitoQC genes

□ Variant replication in additional cases of EOPD, LOPD & controls

□ **Comments/Challenges/Issues/Concerns**

• If timelines change, comment here.

• If off by more than one quarter in spending, comment here.

Budget Expenditure to Date

Projected Expenditure:

Actual Expenditure: \$404,120.74

Parkinson-Associated *SNCA* Enhancer Variants Revealed by Open Chromatin in Mouse Dopamine Neurons

Sarah A. McClymont,¹ Paul W. Hook,¹ Alexandra I. Soto,² Xylena Reed,¹ William D. Law,¹ Samuel J. Kerans,¹ Eric L. Waite,¹ Nicole J. Briceno,¹ Joey F. Thole,¹ Michael G. Heckman,³ Nancy N. Diehl,³ Zbigniew K. Wszolek,⁴ Cedric D. Moore,⁵ Heng Zhu,⁵ Jennifer A. Akiyama,⁶ Diane E. Dickel,⁶ Axel Visel,^{6,7,8} Len A. Pennacchio,^{6,7,9} Owen A. Ross,^{2,10,11} Michael A. Beer,^{1,12} and Andrew S. McCallion^{1,13,14,*}

The progressive loss of midbrain (MB) dopaminergic (DA) neurons defines the motor features of Parkinson disease (PD), and modulation of risk by common variants in PD has been well established through genome-wide association studies (GWASs). We acquired open chromatin signatures of purified embryonic mouse MB DA neurons because we anticipated that a fraction of PD-associated genetic variation might mediate the variants' effects within this neuronal population. Correlation with >2,300 putative enhancers assayed in mice revealed enrichment for MB cis-regulatory elements (CREs), and these data were reinforced by transgenic analyses of six additional sequences in zebrafish and mice. One CRE, within intron 4 of the familial PD gene *SNCA*, directed reporter expression in catecholaminergic neurons from transgenic mice and zebrafish. Sequencing of this CRE in 986 individuals with PD and 992 controls revealed two common variants associated with elevated PD risk. To assess potential mechanisms of action, we screened >16,000 proteins for DNA binding capacity and identified a subset whose binding is impacted by these enhancer variants. Additional genotyping across the *SNCA* locus identified a single PD-associated haplotype, containing the minor alleles of both of the aforementioned PD-risk variants. Our work posits a model for how common variation at *SNCA* might modulate PD risk and highlights the value of cell-context-dependent guided searches for functional non-coding variation.

Introduction

Parkinson disease (PD) is a common progressive neurodegenerative disorder characterized by preferential and extensive degeneration of dopaminergic (DA) neurons in the *substantia nigra*.^{1,2} This loss of midbrain (MB) DA neurons disrupts the nigrostriatal pathway and results in the movement phenotypes observed in PD. Although this disorder affects approximately 1% of people over 70 years old worldwide,³ the mechanisms underlying genetic risk of sporadic PD in the population remain largely unknown. Familial cases of PD with known pathogenic mutations are better understood but account for ≤10% of PD cases.⁴

The α -synuclein gene (*SNCA*) is commonly disrupted in familial PD through missense mutations predicted to promote misfolding^{5–7} or genomic multiplications, resulting in an overexpression paradigm.⁸ The *SNCA* locus has also been shown by genome-wide association studies (GWASs) to harbor common variants, modulating risk of sporadic PD.⁹ In the same way, common variants at more than 40 additional loci have been implicated in PD,¹⁰ but the causal variants which are responsible for elevating risk,

and the genes they modulate in doing so, remain largely undetermined.

That most GWAS-implicated variants are non-coding¹¹ is a major source of this uncertainty, obstructing the identification of: (1) the causative variant at a locus; (2) the context in which a variant is acting; and (3) the mechanism by which a variant asserts its effect on disease risk.

GWASs are inherently biologically agnostic, and their exploitation of linkage disequilibrium (LD) structure frequently results in the implication of many variants at a particular locus, but no one variant is prioritized over those in LD. One method to prioritize non-coding variants is to examine the chromatin status at that locus.^{11–13} Accessible chromatin is more likely to be functional, and variants therein might impact that activity more so than those variants residing in inaccessible chromatin. Recent studies have prioritized neuropsychiatric variants through examination of the chromatin status of iPSC-derived neurons or post-mortem whole brain tissues.^{14,15} However, chromatin accessibility is dynamic and often varies across cell types and developmental time; therefore, understanding and isolating the *in vivo* cellular context in which

¹McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; ²Department of Neuroscience, Mayo Clinic, Jacksonville, FL 32224, USA; ³Division of Biomedical Statistics and Informatics, Mayo Clinic, Jacksonville, FL 32224, USA; ⁴Department of Neurology, Mayo Clinic, Jacksonville, FL 32224, USA; ⁵Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; ⁶Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; ⁷US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA; ⁸School of Natural Sciences, University of California, Merced, CA 95343, USA; ⁹Comparative Biochemistry Program, University of California, Berkeley, CA 94720, USA; ¹⁰Neuroscience Track, Mayo Graduate School, Jacksonville, FL 32224, USA; ¹¹Department of Clinical Genomics, Mayo Clinic, Jacksonville, FL 32224, USA; ¹²Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; ¹³Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; ¹⁴Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA

*Correspondence: andy@jhmi.edu
<https://doi.org/10.1016/j.ajhg.2018.10.018>

© 2018 American Society of Human Genetics.

variation acts is critical to increasing our ability to prioritize variants and query their methods of action.^{11,16–18}

By exploiting the preferential vulnerability of MB DA neurons in PD, we have prioritized DA neurons as the biological context in which a fraction of PD-associated variants are likely to act. DA neurons in other brain regions, such as the forebrain (FB), provide a related substrate that is less vulnerable to loss in PD. We sought to use chromatin data from *ex vivo* populations of DA neurons to investigate the contributions of non-coding variation to PD risk. To maximize the specificity of the biological context, we generated chromatin signatures for purified mouse MB and FB DA neurons. We examined the resulting regulatory regions for their ability to direct *in vivo* reporter expression and developed a regulatory sequence vocabulary specific to DA neurons. In doing so, we identified a novel MB DA regulatory element that falls within intron 4 of *SNCA* and demonstrated its ability to direct reporter expression in catecholaminergic neurons from transgenic mice and zebrafish, confirming it to be an enhancer. Furthermore, this enhancer harbors two common variants, falling within a haplotype that we determine to be associated with PD risk. We demonstrate that these enhancer variants impact protein binding, and we propose a model for how the variants and the haplotype at large contribute to *SNCA* regulatory control. This work illustrates the power of cell-context-dependent guided searches for the identification of disease-associated and functional non-coding variation.

Material and Methods

Animal Husbandry

Tg(Th-EGFP)DJ76Gsat mice (Th-EGFP) were generated by the GENSAT project¹⁹ and purchased through the Mutant Mouse Resource and Research Centers Repository. Colony-maintenance matings were between hemizygous male Th-EGFP mice and female Swiss Webster (SW) mice, obtained from Charles River Laboratories. This same mating scheme was used for establishing timed matings and thus generating litters for assay; the day on which the vaginal plug was observed was E0.5. Adult AB zebrafish lines were maintained in system water according to standard methods.²⁰ All work involving mice and zebrafish (husbandry, colony maintenance, procedures, and euthanasia) were reviewed and pre-approved by the institutional care and use committee.

Neural Dissociation and Fluorescence-Activated Cell Sorting

Pregnant SW mice were euthanized at E15.5, and the embryos were removed and immediately placed in chilled Eagle's Minimum Essential Medium (EMEM) on ice. The embryos were decapitated, and the brains were removed into Hank's Balanced Salt Solution without Mg^{2+} and Ca^{2+} (HBSS w/o) on ice. Under a fluorescent microscope, EGFP+ brains were identified and microdissected to yield the desired MB and FB regions. Microdissected regions were placed in fresh HBSS w/o on ice and pooled per litter for dissociation.

Pooled brain regions were dissociated via the Papain Dissociation System (Worthington Biochemical Corporation). The tissue

was dissociated in the papain solution for 30 min at 37°C, and gentle trituration was performed every 10 min with a sterile Pasteur pipette. After dissociation, cells were passed through a 40 μ m cell strainer into a 50 mL conical tube, centrifuged for 5 min at 300 $\times$ g, resuspended in albumin-inhibitor solution containing DNase, applied to a discontinuous density gradient, and centrifuged for 6 min at 70 $\times$ g. The resulting cell pellet was resuspended in HBSS with Mg^{2+} and Ca^{2+} and submitted to fluorescence-activated cell sorting (FACS). Aliquots of 50,000 EGFP+ cells were sorted directly into 300 μ L HBSS with Mg^{2+} and Ca^{2+} and 10% FBS for ATAC-seq. Aliquots containing $\geq 50,000$ EGFP+ cells were sorted into kit-provided lysis buffer for RNA-seq. This procedure was repeated such that a single aliquot of cells from each region per litter was submitted to either ATAC-seq or bulk RNA-seq, repeated three (ATAC-seq) or four (RNA-seq) times for each region.

ATAC-seq Library Preparation and Quantification

ATAC-seq library preparation generally followed the steps as set out in the original ATAC-seq paper,²¹ with minor modifications. Aliquots of 50,000 EGFP+ cells were centrifuged for 5 min at 4°C and 500 $\times$ g, washed with 50 μ L of chilled PBS, and centrifuged again for 5 mins at 4°C and 500 $\times$ g. Next, the cell pellet was resuspended in lysis buffer, as set out in the protocol, and cells were left to lyse for 5 min at 4°C before being centrifuged for 10 min at 4°C at 500 $\times$ g. The resulting nuclei pellet was tagged, as written, using the transposase from the Nextera DNA Library Preparation Kit. After transposition, DNA was purified with the MinElute Reaction Clean-up Kit (QIAGEN) and eluted in 10 μ L elution buffer.

The libraries were amplified according to the original ATAC-seq protocol.²¹ The qPCR surveillance steps were modified such that the additional number of cycles of amplification were calculated as $1/4$ maximum intensity, so as to limit PCR duplication rates in the final libraries. The amplified libraries were purified with Ampure XP beads (Beckman Coulter) according to the Nextera DNA Library Prep Protocol Guide. The libraries were quantified with the Qubit dsDNA High Sensitivity Assay (Invitrogen) in combination with the High Sensitivity DNA Assay (Agilent) on the Agilent 2100 Bioanalyzer.

ATAC-seq Sequencing, Alignment, and Peak Calling

Individual ATAC-seq libraries were sequenced on the Illumina MiSeq to a minimum depth of 15 million, 2 $\times$ 75 bp reads per library. A single MB ATAC-seq library was sequenced on the Illumina HiSeq in rapid run mode with 2 $\times$ 100 bp reads to a depth of ≥ 350 million paired-end reads.

The quality of sequencing was evaluated with FastQC (v0.11.2). Reads were aligned to mm9 with Bowtie2²² (v2.2.5) under –local mode. Reads aligning to the mitochondrial genome; unknown and random chromosomes; and PCR duplicates were removed prior to peak calling (SAMtools²³). Peaks were called on individual libraries and on a concatenated file combining all MB or all FB (“joint”) libraries via MACS2²⁴ (v2.1.1.20160309) “callpeak” with the following options: –nomodel –nolambda –B –f BAMPE –gsize mm –keep-dup all. Peaks overlapping blacklisted regions that were called by ENCODE and that were in the original ATAC-seq paper were removed.^{21,25} Peaks were examined for their genomic distribution via CEAS in the Cistrome pipeline.^{26,27} The fragment lengths were extracted from the SAM files and plotted with a custom script. Mouse (mm9) transcriptional start site

(TSS) coordinates were extracted from the UCSC Genome Browser,²⁸ and deepTools²⁹ was used for quantifying the pileup of reads over TSSs.

RNA-seq Library Preparation and Quantification

Total RNA was extracted with the Purelink RNA Micro Kit (Invitrogen). After FACS isolation into kit-provided lysis buffer, samples were homogenized, and RNA extraction proceeded according to the manufacturer's recommendations. Total RNA integrity was determined with the RNA Pico Kit (Agilent) on the Agilent 2100 Bioanalyzer. RNA samples were sent to the Sidney Kimmel Comprehensive Cancer Center Next Generation Sequencing Core at Johns Hopkins for library preparation with the Ovation RNA-Seq System V2 (Nugen) and for sequencing.

RNA-seq Sequencing, Alignment, and Transcript Quantification

The libraries were pooled and sequenced on Illumina's HiSeq 2500 in rapid-run mode with 2×100 bp reads to an average depth of >90 million reads per library. The quality of sequencing was evaluated via FastQC. FASTQ files were aligned to mm9 with HISAT2³⁰ (v2.0.1-beta) with `-dta` specified.

Aligned reads from individual samples were quantified against a reference transcriptome with the Rsubread package^{31–33} (v1.22.3) function “featureCounts” with the following options: `isPairedEnd = TRUE`, `requireBothEndsMapped = TRUE`, `isGTFAnnotationFile = TRUE`, and `useMetaFeature = TRUE`. The GENCODE vM9 GTF was downloaded³⁴ (date: March 30, 2016) and lifted over from the mm10 genome to the mm9 genome with CrossMap (v0.2.2) under default parameters.³⁵ This was used for quantification, in which gene-level raw counts were converted to RPKM (reads per kilobase of transcript per million) values and means for each region were calculated.

Relationship between RNA-seq and ATAC-seq

The 1,000 most highly expressed genes and the 1,000 least highly expressed genes ($\text{RPKM} \geq 1$) in both the MB and FB were identified, and their transcriptional start sites (Ensembl) were extracted from the UCSC Table Browser.²⁸ Intervals of 1, 10, and 100 kb surrounding these TSSs were intersected with the ATAC-seq libraries, and the overlap quantified³⁶ and plotted. These same TSSs were provided to deepTools,²⁹ and the ATAC-seq signal over these most highly- and least highly-expressed genes was quantified and plotted. Additionally, the 1,000 highest and lowest ATAC-seq peaks (by *q* value) were extracted, and the expression of the nearest gene was quantified and plotted as a final metric to relate the RNA-seq and ATAC-seq datasets.

cDNA Synthesis and RT-qPCR for DA Neuron Markers

RNA was extracted with the RNeasy Mini Kit (QIAGEN) after 50,000 cells were sorted directly into Buffer RLT. Aliquots of 50,000 non-fluorescing cells were also collected and processed in parallel. 100 ng of each RNA sample was submitted to first-strand cDNA synthesis with the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen) according to the oligo(dT) method.

Primers (Table S1) were designed with Primer-BLAST³⁷ under default parameters, except the requirement for exon-exon junction spanning was specified. qPCR was performed with Power SYBR Green Master Mix (Applied Biosystems). Reactions were run in triplicate under default SYBR Green Standard cycle specifications on the Viia7 Real-Time PCR System (Applied Biosystems). Relative

quantification followed the $2^{-\Delta\Delta CT}$ method, and results were normalized to *Actb* in the EGFP– aliquot of cells for each region.

Correlation Analysis between Regions and within Replicates

Peaks from all six ATAC-seq libraries and the two “joint” ATAC-seq libraries were concatenated together, sorted on the basis of chromosomal location, merged into a unified peak set,³⁶ and converted to Simplified Annotation Format. Reads from each BAM file overlapping this unified peak set were quantified with the Rsubread package's “featureCounts” command, with the following options: `isPairedEnd = TRUE` and `requireBothEndsMapped = FALSE`. Read counts were normalized for each library via conditional quantile normalization;³⁸ library size, peak length, and peak GC content were accounted for. Pearson correlation coefficients were calculated from this normalized count matrix and visualized with `corplot`, `RColorBrewer`, and `LSD` (see Web Resources).

Sequence-Constraint Analysis

Average phastCons³⁹ were calculated for the “joint” peak file for both the MB and FB libraries with Cistrome.²⁷ Beforehand, peaks with overlap of exons or promoters (defined here as $\pm 2,000$ bp from the transcriptional start site) were removed. The exon and promoter BED files were downloaded from the UCSC Table Browser²⁸ (Mouse genome; mm9 assembly; Genes and Gene Predictions; RefSeq Genes track using the table `refGene`).

Gene Ontology of Nearest Expressed Gene

The Genomic Regions Enrichment of Annotations Tool⁴⁰ (GREAT; v3.0.0) predicted the gene ontology (GO) term enrichment in the catalogs. Beforehand, peaks were processed so that (1) peaks overlapping commonly open regions would be removed; (2) the top 20,000 peaks would be selected; (3) peaks would overlap the nearest expressed gene's transcriptional start site (TSS).

First, regions that are commonly open were defined as those regions of the genome that are open in >30% of ENCODE DNase hypersensitivity site (DHS) assays in mouse tissues. These ubiquitously open regions were removed from the peak files. Next, so that the binomial distribution for calculating enrichment was still valid, we limited the number of regions considered by GREAT by only submitting the top 20,000 peaks on the basis of *q* value.

Finally, in order to limit ourselves to the nearest expressed gene, we supplied a list of the TSSs of the nearest expressed genes, considering only those genes that are used by GREAT.

Only genes that are in this list and for which $\text{RPKM} > 1$ were considered to be expressed. The expressed gene nearest to each of the top 20,000 peaks was identified. Each peak is associated with its nearest expressed gene, and to ensure that GREAT only considered these nearest genes for analysis, we submitted these nearest expressed genes' TSSs as a proxy for each peak. We submitted these proxy peaks to GREAT by using the NCBI build 37 (mm9) assembly under the “whole-genome background regions” setting; we used the single nearest gene as the association rule and included curated regulatory domains.

Quantification of Overlap between CRE Catalogs and the VISTA Enhancer Browser

All elements tested *in vivo* were downloaded from the VISTA Enhancer Browser on September 4, 2016. These regions were stratified into those annotated as positive and those annotated as negative. BED coordinates of these regions were extracted and

intersected with the ATAC-seq catalogs. Positive regions were further stratified into those with annotations for only forebrain, only midbrain, or only hindbrain; combinations of regions (“multiple regions”); or all three regions (“whole brain”). These categories comprised the “neuronal” category. Other regions that were annotated as positive but not driving expression in any of those three regions were placed in the “non-neuronal” category.

Testing Five Putative CREs for *In Vivo* Reporter Activity

Prioritized regions were amplified (Table S1) from human genomic DNA via PCR and cloned into either pENTR for mouse *lacZ* assays (Invitrogen) or pDONR221 for zebrafish assays (Invitrogen). Sequences were validated, and regions were cloned via LR cloning (Invitrogen) into either an *hsp68-lacZ* vector or pXIG vector, with a TdTomato cassette in place of GFP.

Generation of transgenic mice and E11.5 embryo-staining were performed as previously described^{41–43} with FVB strain mice. Embryos expressing the *lacZ* reporter gene were scored and annotated for their expression patterns by multiple curators. For a construct to be considered positive, a minimum of three embryos per construct were required to demonstrate reporter activity in the same tissue. Mouse transient transgenic assays were approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee.

Generation of transgenic zebrafish was performed as previously described⁴⁴ in AB zebrafish. Reporter expression patterns were evaluated at both 3 and 5 days post-fertilization (dpf). For a construct to be considered as positive, $\geq 25\%$ of mosaic embryos had to display reporter activity in one or more anatomical structures. Positive zebrafish were quantified for reporter activity in five anatomical regions (forebrain, midbrain, hindbrain, amacrine cells, and the spinal cord).

Development of Regulatory Vocabulary

We applied the machine-learning algorithm gkm-SVM⁴⁵ to our MB and FB catalogs under default settings. We trained gkm-SVM on the sequences underlying the summits ± 250 bp of non-ubiquitously open, top 10,000 peaks by q value versus five negative sets, which were matched for GC content, length, and repeat content. Weights across all five tests were averaged for all 10-mers.

All 10-mers with weight ≥ 1.50 were clustered on the basis of sequence similarity via Starcode;⁴⁶ sphere clustering was specified, and the distance was set to 3. clustalOmega⁴⁷ aligned the sequences within these clusters, and MEME,⁴⁸ under default parameters except for -dna -maxw 12, generated position weight matrices (PWMs) for these aligned clusters. Tomtom,⁴⁹ querying the Jolma 2013, JASPAR Core 2014, and Uniprobe mouse databases, identified the top transcription factors corresponding to these PWMs, under default parameters except for -no-ssc -min-overlap 5 -eval -thresh 10.0.

We used the same procedure to identify transcription factors specifically conveying regulatory potential in the MB library relative to the FB library; except during gkm-SVM training, the positive set was specified as the top 10,000 non-ubiquitously open MB summits, and the negative set was specified to be the top 10,000 non-ubiquitously open FB summits, both ± 250 bp.

Transcription-Factor Footprinting

CENTIPEDe⁵⁰ was used for footprint identification. Sequences underlying the deeply sequenced MB library peaks, less those that were ubiquitously open, were extracted. FIMO,⁵¹ with options

-text -parse-genomic-coord, identified all locations underlying ATAC-seq peaks of the motifs identified above. Additionally, conservation data from 30-way vertebrate phastCons were considered in the CENTIPEDe calculations; for each PWM site, those with a mean conservation score greater than 0.9 were considered. Finally, the BAM file read-end coordinates were adjusted in response to the shift resulting from the transposase insertion.⁵² As such, reads were adjusted +4 bp on the positive strand and -5 bp on the negative strand as performed in the original ATAC-seq method.²¹

Genome-wide Read Pileup over Predicted Motif Sites

FIMO, as above, was used for genome-wide identification of all coordinates of the identified motifs. The deepTools²⁹ “bamCoverage” tool was run, under default conditions, for conversion of the deeply sequenced MB library BAM to bigwig format. After this, a matrix file was generated with “computeMatrix,” specified options were -referencePoint center -b 1000 -a 1000 -bs 50. Finally, “plotHeatmap” was used for generating plots indicating ATAC-seq read pileup over predicted motif sites.

Intersection of CRE Catalogs and PD-Associated GWAS Variants

Lead SNPs from two of the most recent meta-analyses^{9,10} were submitted to rAggr (see Web Resources), and SNPs in LD were identified (1000 Genomes, phase 3, EUR populations; minimum minor-allele frequency [MAF] = 0.05, $r^2 \geq 0.8$; maximum distance = 5,000 kb). These variants were intersected³⁶ with the CRE catalogs after they were lifted over to hg19 coordinates, and the overlap was extracted and quantified.

In Vivo Validation of the MB-Specific Enhancer

The MB-specific peak was amplified (Table S1) from human genomic DNA via PCR and cloned into pCR8 via TA cloning (Invitrogen). Sequences were validated, and regions were cloned into either an *hsp68-lacZ* vector or a modified pXIG vector, with a TdTomato cassette in place of GFP, via LR cloning (Invitrogen).

For zebrafish transgenesis, the modified pXIG vector was injected into 1- to 2-cell-stage embryos as previously described⁴⁴ in AB zebrafish. TdTomato reporter expression was assayed at 72 hr post-fertilization (hpf) and 5 dpf; mosaic embryos positive for TdTomato expression were selected and raised to adulthood, and founders were identified. The progeny of founders were screened at 72 hpf for reporter activity. For mouse transgenesis, the generated *hsp68-lacZ* vector was purified in a double CsCl gradient (Lofstrand Labs), and stable mouse transgenesis was performed in C57BL/6 mice by Cyagen Biosciences. Multiple founder lines were generated. For *lacZ* staining, embryos were collected at E12.5, and mouse brains were isolated at E15.5, P7, P30, and P574. Brains were roughly sectioned in 1 mm sections at P7 and P30. The animals were perfused at P574, and fixed brains were sectioned (200 μ m) with a vibratome. Specimens were subsequently fixed for 2 hr on ice in 1% formaldehyde, 0.2% glutaraldehyde, and 0.02% Igepal CA-630 in PBS. After fixation, tissues were permeabilized over the course of three 15 min washes in 2 mM MgCl₂ and 0.02% Igepal CA-630 in PBS at room temperature. Embryos and/or tissues were incubated overnight at 37°C in a staining solution containing 320 μ g/mL X-Gal in N,N-dimethyl formamide, 12 mM K-ferricyanide, 12 mM K-ferrocyanide, 0.002% Igepal CA-630, and 4 mM MgCl₂ in PBS. Specimens were washed twice for 30 min each time in 0.2% Igepal CA-630 in PBS and

finally stored in 4% formaldehyde, 100 mM sodium phosphate, and 10% methanol.

Individuals with and without PD-Sequencing and Genotyping at *SNCA*

986 individuals with PD and 992 controls who were all seen at the Mayo Clinic in Jacksonville, FL were sequenced across the putative enhancer and genotyped for 25 variants across the *SNCA* locus. The variants chosen for genotyping were those identified during the sequencing of the enhancer and a subset of those identified by Guella et al.⁵³ For PD-affected individuals, median age at blood draw was 69 years (range: 28–97 years), median age at PD onset was 67 years (range: 28–97 years), and 631 subjects (64.0%) were male. Median age at blood draw for control individuals was 67 years (range: 18–92 years), and 415 subjects (41.8%) were male. Individuals with PD were diagnosed according to standard clinical criteria.⁵⁴ All subjects are unrelated, non-Hispanic Caucasians of European descent. The Mayo Clinic Institutional Review Board approved the study, and all subjects provided written informed consent.

Genomic DNA was extracted from whole blood with the Auto-gen FlexStar. Sanger sequencing of the enhancer region was performed bidirectionally with the ABI 3730xl DNA analyzer (Applied Biosystems) according to standard protocols. Sequence data were analyzed with SeqScape (v2.5; Applied Biosystems). Statistical analyses were performed with both SAS and R (see [Web Resources](#)). Of the variants identified within the enhancer, only those with a MAF greater than 5% were evaluated for association with PD in single-variant analysis. Associations between individual variants and PD were evaluated with logistic regression models, adjusted for both sex and age at blood draw, and, when variants were considered, under an additive model (i.e., the effect of each additional minor allele was evaluated). Odds ratios and 95% confidence intervals were estimated. A Bonferroni correction for multiple testing was utilized in single-variant analysis because of the four common variants that were evaluated for association with PD, after which p values ≤ 0.0125 were considered as statistically significant.

Genotyping of the 25 SNPs across the *SNCA* locus was performed with the iPLEX Gold protocol on the MassARRAY System and analyzed in TYPER 4.0 software (Agena Bioscience). For the 25 SNPs genotyped across the *SNCA* locus, all genotype call rates were $>95\%$, and there was no evidence for departure from Hardy-Weinberg equilibrium (all χ^2 p values >0.05 after Bonferroni correction). Haplotype frequencies in cases and controls were estimated with the haplo.stats package function “haplo.group” (see [Web Resources](#)). The haplo.score function was used to perform score tests for association evaluating the relationships between haplotypes and the risk of PD.⁵⁵ Tests were adjusted for both sex and age at blood draw, haplotypes occurring in less than 1% of subjects were excluded, and only individuals with no missing genotype calls for any variants were included. A Bonferroni correction for multiple testing was applied as a result of the 12 different common haplotypes that were observed and tested for association with PD risk, after which p values ≤ 0.0042 were considered as statistically significant.

LD structure and r^2 values at the *SNCA* locus in the 1000 Genomes EUR population were extracted from LDlink⁵⁶ with the LDmatrix tool and plotted with R. The chromatin structure at *SNCA* was extracted from the 3D Genome Browser, and POLR2A binding in MCF-7 cells was examined at the *SNCA* promoter.

Protein Array Testing Differential Binding

HuProt v3.1 human proteome microarrays (Grace Bio-Labs) containing $>16,000$ unique proteins representing 12,586 genes (CDI laboratories)⁵⁸ were blocked with 25 mM HEPES (pH 8.0), 50 mM potassium glutamate, 8 mM MgCl_2 , 3 mM dithiothreitol (DTT), 10% glycerol, 0.1% Triton X-100, and 3% BSA on an orbital shaker at 4°C for ≥ 3 hr. Allele-specific protein-DNA binding interactions were identified through dye-swap competition of major and minor alleles labeled with either Cy3 or Cy5. DNA fragments for dbSNP: rs2737024 and dbSNP: rs2583959 were synthesized such that each allele was flanked by 15 nucleotides of the upstream and downstream sequence and there was a common priming site at the 3' end ([Table S1](#)).

We created the dsDNA fragments by separately annealing a primer containing a Cy3 or Cy5 label and adding Klenow (New England Biolabs) with dNTP to fill in the complementary strand for each allele.⁵⁹ Cy3-labeled major allele was mixed with Cy5-labeled minor allele (each at 40 nM) in $1\times$ hybridization buffer (10 mM Tris-Cl [pH 8.0], 50 mM KCl, 1 mM MgCl_2 , 1 mM DTT, 5% glycerol, 10 μM ZnCl_2 , and 3 mg/mL BSA) and added to an array; dyes were then swapped for each allele, and the mixture was then added to a second array. DNA was allowed to bind overnight at 4°C on an orbital shaker with protection from light. Chips were washed once with cold $1\times$ Tris-buffered saline-Triton solution (0.1% Triton X-100) for 5 min at 4°C , rinsed, and dried in the centrifuge. Cy5 and Cy3 images were taken separately on a Genepix 4000B scanner and, after alignment to the GAL file, individual spot intensities were extracted with Genepix Pro software.

Allele-specific interactions were identified through dye-swap analysis. The ratio of major to minor allele binding was calculated from the duplicate spot-average, median-foreground signal for each protein according to the following equation:

$$\log_2 \sqrt{\frac{\text{Cy3}_{\text{major}} * \text{Cy5}_{\text{major}}}{\text{Cy3}_{\text{minor}} * \text{Cy5}_{\text{minor}}}}$$

Mean intensity was calculated from the average foreground signal for the Cy3 and Cy5 channels of the major and minor alleles. MA plots were made for each allele using the calculated mean intensity and the log ratio of the major to the minor allele.

Results

ATAC-seq Identifies Open Chromatin in MB and FB DA Neurons

To identify open chromatin regions (OCRs) in DA neurons, we performed ATAC-seq²¹ on $\sim 50,000$ FACS-isolated cells (per replicate) from microdissected regions of the MB and FB of embryonic day 15.5 (E15.5) Tg(Th-EGFP)DJ76Gsat BAC transgenic mice⁶⁰ ([Figure 1A](#)). EGFP, expressed under the control of the tyrosine hydroxylase (Th) locus, labels catecholaminergic neurons (i.e.: DA, noradrenergic, and adrenergic neurons) in this mouse line. To confirm capture of the corresponding catecholaminergic neurons, we performed RT-qPCR on the isolated reporter-labeled cells and established them to be enriched for DA neuronal markers in comparison to unlabeled populations from the same dissected tissues ([Figure S1](#)).

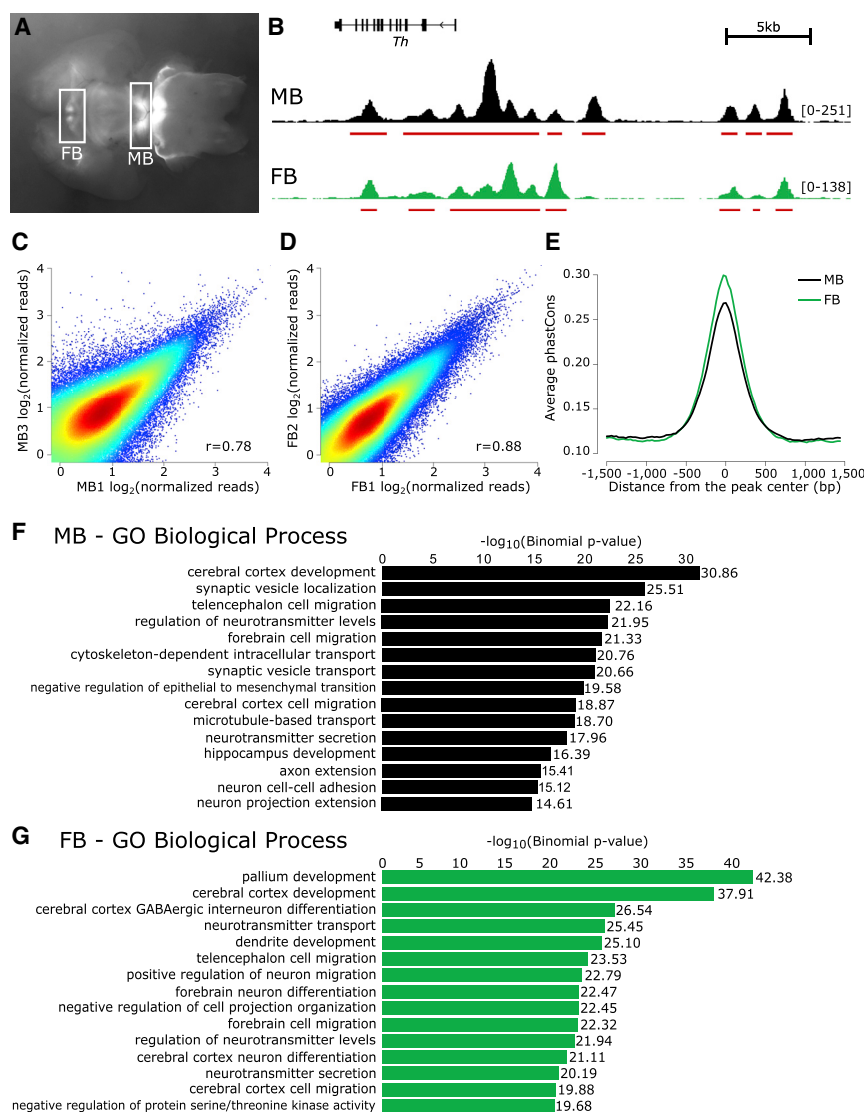


Figure 1. Preliminary Validation of ATAC-seq Catalogs Generated from Ex Vivo DA Neurons

(A) The midbrain (MB) and forebrain (FB) of E15.5 brains from Tg(Th-EGFP) DJ76Gsat mice are microdissected, dissociated, and isolated by FACS.

(B) Read pileup and called peaks for the MB and FB libraries at the *Th* locus.

(C and D) Chromatin accessibility, genome-wide, is correlated between replicates.

(E) The sequences underlying MB and FB peaks display a high degree of evolutionary sequence constraint as measured by phastCons scores.

(F and G) For both the MB and FB, gene ontology terms of the expressed genes nearest to each peak reflect the neuronal origin and function of these catalogs.

To evaluate the ATAC-seq libraries, we examined *in silico* quality-control measures (Figure S2); evaluated the called peaks and read pileups with the Integrative Genomics Viewer (IGV);^{61,62} and quantified the correlation between brain regions and within replicates. A representative browser trace at the *Th* locus in both MB and FB libraries is presented in Figure 1B. Replicates are well correlated: MB library replicates have an average correlation of 0.72 (Figure 1C), and FB replicates are more highly correlated, at $r = 0.86$ (Figure 1D). Given the robust correlation between replicates, we pooled all reads from the same brain region and called peaks on this unified set to increase our power to detect regions of open chromatin. As a result, we identified 104,217 regions of open chromatin in the MB DA neurons and 87,862 regions in the FB. MB and FB libraries are moderately well correlated (average $r = 0.64$; Figure S3), and approximately 60% of MB OCRs are also represented in the FB libraries.

To assess these catalogs for characteristics of functionality, we examined the sequence constraint underlying the called

regions of open chromatin but excluded peaks that overlap promoters. Promoters are typically accessible,⁶³ and thus we aimed to reduce the inflation that affects sequence conservation as a result of highly conserved promoter-overlapping ATAC-seq peaks. Despite removal of these highly conserved peaks, we found that the degree of sequence constraint underlying open-chromatin peaks was high in comparison to background constraint (Figure 1E). The fact that elements in these libraries of putative cis-regulatory elements (CREs) are constrained highlights their probable functional significance.

To further examine the OCR catalogs for biological relevance, we explored the GO terms of nearby genes. Although CREs are not restricted to acting solely on the nearest gene, this restriction is often used as a proxy in the absence of other information. To bolster our predictions, we have also generated bulk RNA-seq data on these same populations of sorted cells (Figure S4) and used these data to examine the GO terms of the nearest expressed gene (RPKM ≥ 1). Although still imperfect, implementing this method as a proxy for function results in GO terms that are enriched for neuronal functions in both MB and FB OCR catalogs (Figures 1F and 1G). Thus, we establish that these OCR catalogs are enriched for putative CREs, probably directing the expression of genes with key roles in neuronal biology.

Candidate Regulatory Regions Are Capable of Directing Expression *In Vivo*

Although sequence conservation and GO suggest that our OCR catalogs are enriched for functional elements, both of these metrics are indirect surrogates for true measures of

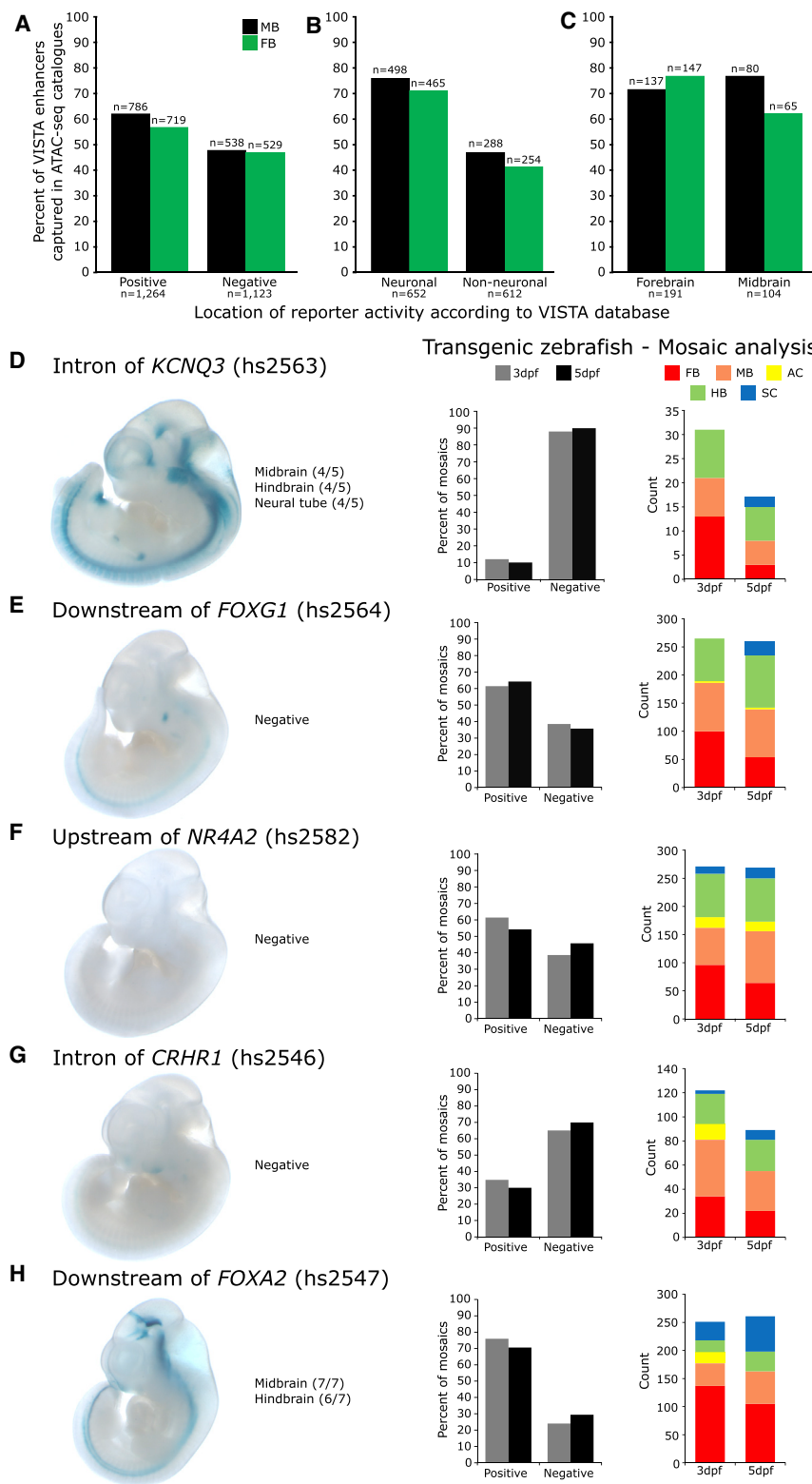


Figure 2. Validation of the Putative CRE Catalogs In Vivo

(A) Of the elements annotated in VISTA as having enhancer activity, 62% and 56% of these are represented in the MB and FB catalogs, respectively.

(B) An abundance of open-chromatin regions in the MB and FB catalogs overlap confirmed neuronal enhancers ($\geq 70\%$).

(C) Neuronal enhancers were stratified by the anatomical domains in which they are active; those that are reported active in the MB and FB are enriched in our MB and FB catalogs, respectively.

(D–H) Testing five prioritized putative CREs *in vivo* identifies five neuronal enhancers.

(D) A putative CRE in intron 1 of *KCNQ3* directs expression in the midbrain, hindbrain, and neural tube of E11.5 *lacZ* reporter mice. It fails to direct expression in a transgenic zebrafish assay at either 3 or 5 days post fertilization (dpf); reporter expression is present in $\leq 25\%$ of mosaics.

(E, F, and G) Putative CREs downstream of *FOXG1*, upstream of *NR4A2*, and in an intron of *CRHR1* fail to direct expression in transgenic mice; however, they direct robust neuronal appropriate expression in transgenic zebrafish reporter assays (scored for expression in MB, FB, amacrine cells [ACs], hindbrain [HB], and spinal cord [SC]).

(H) A putative CRE downstream of *FOXA2* directs neuronal expression in both transgenic mice and zebrafish assays. n mosaic zebrafish scored: ≥ 141 for 3 dpf, ≥ 119 for 5 dpf. All constructs have since been deposited in the VISTA database under the supplied hs numbers.

and that have been cataloged in the VISTA Enhancer Browser⁶⁴ (accessed September 4, 2016). Overlap between our catalogs and all 2,387 VISTA Enhancer Browser elements, which were scored for their ability to direct *lacZ* reporter expression in E11.5 mice, was quantified (Table S2). Of the 1,264 VISTA elements identified as enhancers, 786 were present in the MB catalog, and 719 were present in the FB catalog (Figure 2A). We examined the overlap of the FB and MB catalogs with enhancers that have been demonstrated to direct expression in either non-neuronal or neuronal tissues, and we observed that 42%–47%

function. To more directly measure the biological relevance of the catalogs and to identify enhancers, we assessed the capability of the candidate CREs to direct expression *in vivo*.

We took advantage of the large repository of elements that have already been tested in *lacZ* reporter assays *in vivo*

of enhancers reported to direct expression in non-neuronal tissues are present in the catalogs. In contrast, 71%–76% of enhancers that direct expression in one or more regions of the brain overlapped the FB and MB catalogs (Figure 2B), confirming an abundance of brain enhancers in our catalogs. Stratifying these confirmed

neuronal enhancers on the basis of their expression patterns in VISTA, we observed an abundance of MB-specific enhancers in our MB catalog and an abundance of FB-specific enhancers in our FB catalog; 77% of MB- and FB-specific enhancers in VISTA were captured in our MB and FB catalogs, respectively (Figure 2C). Collectively, these data establish that our region-specific OCR catalogs capture region-specific, active CREs with high efficiency.

To extend our assessment of the biological activity of sequences within these OCR catalogs, we focused on an additional five candidate CREs not already tested in the VISTA browser and evaluated their ability to act as enhancers in *lacZ* reporter mice and in transgenic zebrafish TdTomato reporter assays. All five regions were represented by robust peaks in both the MB and FB catalogs (Figure S5). Two regions, one in the first intron of *Kcnq3* and the other downstream of *Foxg1*, were additionally prioritized via H3K27ac ChIP-seq from a variety of tissues taken from E11.5 and E15.5 embryonic mice because we sought to limit our selection to candidate enhancers predicted to have neuronal-specific activity. The remaining three candidate CREs were selected on the basis of their proximity to genes important in DA neuron biology. We selected sequences at *Foxa2* and *Nr4a2* because both are key transcription factors (TFs) in the development and maintenance of DA neurons.^{65–68} We selected the final region, located in an intron of *Crhr1*, because this locus has been implicated in PD by GWAS,⁹ and our group has recently prioritized this gene as a candidate for PD risk.⁶⁹ All selected sequences were lifted over to hg19, a process which facilitated the identification and assay of their corresponding human sequence intervals.

When tested in transgenic reporter mice at E11.5 (Figure S5), two of the five regions (those near *KCNQ3* and *FOXA2*) were validated as enhancers (Figures 2D and 2H). Recognizing that a disparity exists between the developmental time at which we generated the catalogs (E15.5) and the time at which the mice were assayed (E11.5), and that this disparity might compromise validation rates, we also assayed each sequence across multiple time points in zebrafish. All assayed regions except the region at *KCNQ3* directed reporter expression in mosaic transgenic zebrafish (Figures 2E–2H). All five regions displayed enhancer activity *in vivo* in neuronal tissues in one or both transgenic assays. Our transgenic animal experiments corroborate the results of the retrospective VISTA Enhancer Browser intersection, implying that our OCR catalogs are biologically active and enriched for sequences capable of driving neural expression *in vivo*.

Candidate CREs Are Enriched for TF Motifs Active in DA Neurons

To identify sequence modules (kmers) predicted to contribute regulatory activity of putative CREs in our catalogs, we applied the machine learning algorithm gkm-SVM.⁴⁵ The resulting regulatory vocabularies of kmers had high predictive power ($\text{auROC}_{\text{MB}} = 0.915$, $\text{auROC}_{\text{FB}} = 0.927$). We rank-ordered and collapsed related kmers to

reveal motifs enriched in the OCRs and their corresponding TFs (Figures 3A, 3E, 3I, and 3M). In the MB, the four most enriched motifs correspond to RFX1, FOXA2, ASCL2, and NR4A2. Given the degeneracy of binding motifs within TF families, we consulted the bulk RNA-seq data for each of the implicated TF families and examined the relative expression levels to prioritize which TFs are most likely producing the observed motif enrichments (Figures 3B, 3F, 3J, and 3N). For example, the reported DNA binding domain is highly conserved between RFX family members, and as a result, the predicted sequence motif for each is highly similar;^{70,71} thus, we must use other means to identify which family member is likely to be acting in these cells. Although no member of the RFX family has been canonically associated with MB DA neurons, we expect *Rfx3* and *Rfx7*, as the two most highly-expressed *Rfx* genes, to probably be active in MB DA neurons and be driving this motif enrichment (Figure 3B). FOXA1 and, especially, FOXA2 are both known to DA neuron biology^{65,72} and both are highly expressed in the MB DA neurons (Figure 3F). Regarding enrichment for the ASCL family, ASCL1 is known to be involved in DA neuron biogenesis⁷³ and is more highly expressed than any other TF in the family (Figure 3J). Finally, NR4A2 is both canonically associated with DA neurons and required for their development;⁶⁸ we observe it to be highly expressed in MB DA neurons (Figure 3N). Examining the sequences underlying the OCR catalogs, we identified TF families known and unknown to DA neuron biology and further refined the TF associations by using expression data.

We also examined the qualities that differentiate MB CREs from FB CREs by examining the sequences underlying MB-specific and FB-specific regions. We developed a vocabulary that discriminates MB and FB regions with high predictive power ($\text{auROC} = 0.926$) and identified kmers enriched in MB-specific peaks where the top corresponding TFs are FOXA1 and/or FOXA2 and NR4A2 (Figure S6). We confirmed this MB bias by again considering the bulk RNA-seq for genes encoding these proteins. As expected, these TFs are more highly expressed in the MB where *Nr4a2* is present at 12-fold higher levels than in the FB (135 RPKM in the MB versus 11 RPKM in the FB) and *Foxa1* and *Foxa2* are not expressed in the FB but are present in the MB (*Foxa1*: 28 RPKM, *Foxa2*: 7 RPKM). Not only do we identify FOXA1 and/or FOXA2 and NR4A2 as more active in MB DA neurons than in the FB, but we also did so solely by comparing their role in the vocabulary of MB-specific OCRs versus FB-specific OCRs.

In a parallel strategy to identify TFs actively engaging the DNA in MB DA neurons, we performed TF footprinting in a single, deeply sequenced MB ATAC-seq library. In doing so, we confirm that two of the TFs prioritized by gkm-SVM leave robust footprints. The motif corresponding to RFX-binding results in a dearth of cuts directly over predicted binding sites (Figure 3C). The same can be seen to a lesser extent for the motif corresponding to FOXA1 and/or FOXA2 (Figure 3G). By contrast, motifs corresponding to

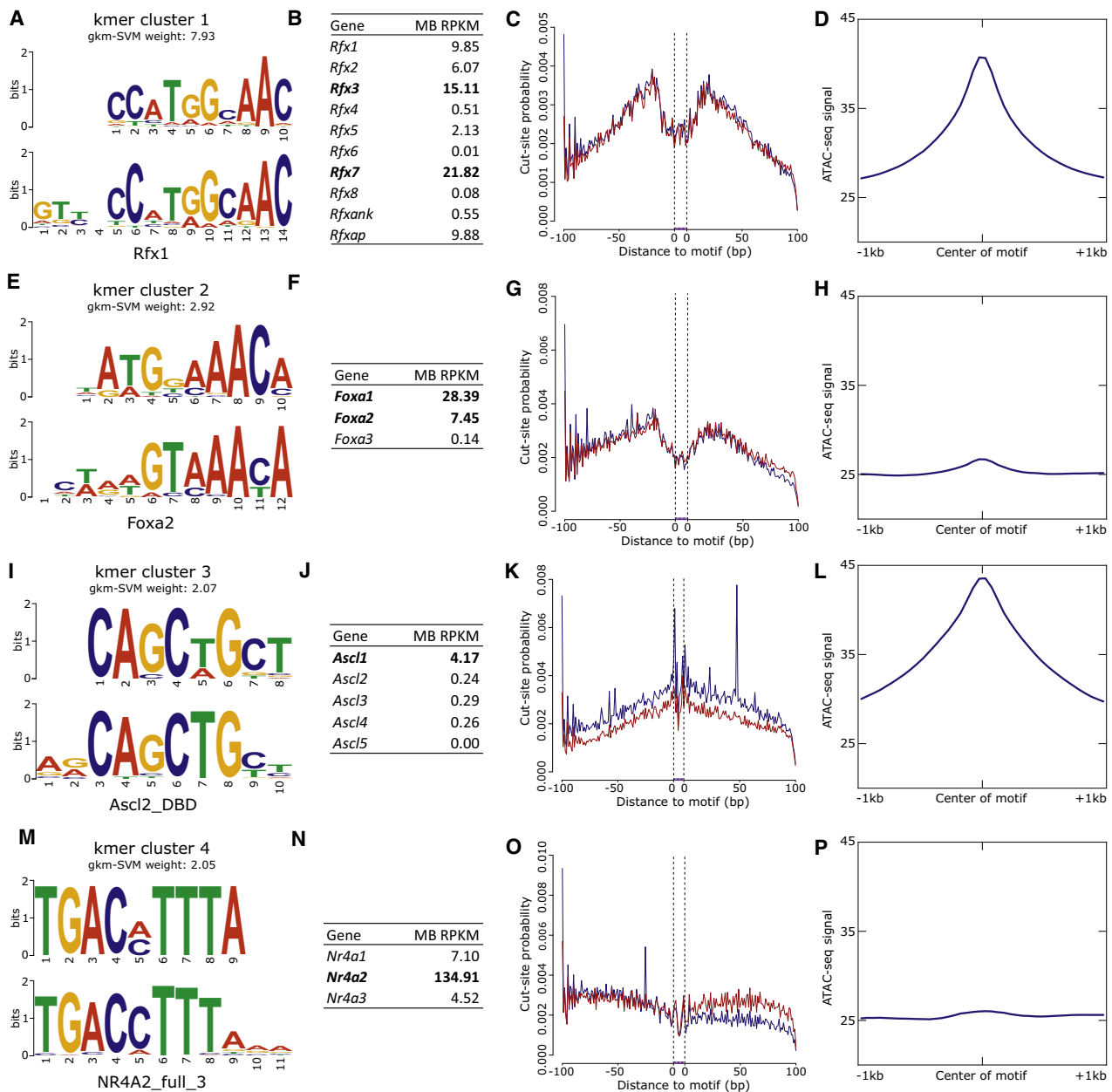


Figure 3. Identification of Transcription Factors (TFs) Important to DA Neurons

(A) The kmer predicted to have the greatest regulatory potential underlying MB ATAC-seq peaks corresponds to the RFX family of TFs. (B, C, and D) RNA-seq quantification in these same cells indicates this enrichment is likely due to RFX3 or RFX7 activity. Examining the ATAC-seq signal over predicted binding sites reveals a robust TF footprint (C) and a general enrichment of reads overlapping RFX sites genome-wide (D).

(E–H) Similarly, a kmer corresponding to FOXA1 and/or FOXA2 has similar evidence for the activity of one or both of these TFs.

(I–L) The third-ranked motif most likely corresponds to ASCL1, and although it fails to leave a robust TF footprint (K), there is clear enrichment of ATAC-seq signal overlapping genome-wide predicted ASCL1 binding sites (L).

(M–P) NR4A2, canonically associated with DA neuron biology, is identified as a highly expressed TF that probably contributes to the regulatory potential of the putative CREs; however, it fails to leave a TF footprint in the cut-site patterns around predicted motif sites (O) and is only mildly enriched for ATAC-seq reads over its predicted binding sites (P).

ASCL1 or NR4A2 fail to leave a robust mark on the chromatin availability (Figures 3K and 3O). It has been noted that nuclear receptors, such as NR4A2, only transiently interact with DNA,⁷⁴ and as a result, it could be that the short DNA residence time fails to result in a robust footprint detectable by transposition. These footprinting data

substantiate the claim that the RFX family of TFs and FOXA1 and/or FOXA2 are active in MB DA neuron CREs.

We confirmed that these sequences are indeed enriched in the catalogs by examining the pileup of reads overlapping all genome-wide predicted motif binding sites for each motif identified by gkm-SVM. We see an abundance

of reads over predicted binding sites of all four motifs (Figures 3D, 3H, 3L, and 3P); the strongest enrichment overlaps RFX and ASCL1 motif sites (Figures 3D and 3L). Despite the less robust footprint generated at the ASCL1 sites, this TF clearly underlies a larger-than-expected proportion of OCRs in the MB catalog. Integrating a support vector machine-learning algorithm, as applied to the sequences underlying OCRs, with footprinting analysis in the same chromatin substrate both powerfully identifies TFs that are important for DA neuron biology and suggests the RFX family of TFs, FOXA1 and FOXA2, ASCL1, and NR4A2 are actively influencing gene expression in the MB DA neurons.

A Candidate CRE in Intron 4 of *SNCA* Is Associated with PD Risk

Having established the biological robustness of the OCR catalog, we moved to exploit these data to investigate how non-coding variation therein might be contributing to PD risk. We established two complementary strategies. First, we sought to globally examine the overlap of PD GWAS SNPs^{9,10} and those in LD ($r^2 > 0.8$) with our DA OCR catalogs. In doing so, we identified 129 unique PD-associated variants that occur at 20 GWAS-associated loci and that are present in one or both of our OCR catalogs (34 specifically overlap the MB catalog, 14 specifically overlap the FB catalog, and 81 overlap both; Table S3).

Second, we examined the chromatin landscape surrounding familial PD genes by focusing on genes with no obvious overlaps in the first strategy. In doing this, we turned our attention to the *SNCA* locus. Despite the fact that this locus is the most significant hit in PD GWASs,^{9,10} the LD structure surrounding the lead SNP (dbSNP: rs356182) is such that no variants in LD are apparent at our r^2 cut-off, and the lead SNP itself is not overlapped by either our MB or FB catalog. Given α -synuclein's established role in PD pathogenesis and the strength of GWAS signal at *SNCA*, we prioritized this locus for a closer, more targeted, inspection.

We first noted that *Snca* expression differs significantly between the MB and FB DA neurons in our bulk RNA-seq analysis (Figure 4B). Examining the chromatin accessibility at the *Snca* locus, we found that the MB and FB were largely the same with the exception of one robust peak in intron 4 (mm9: chr6: 60,742,503–60,744,726) that is present in the MB but completely absent in the FB (Figure 4A). DNase hypersensitivity site (DHS) linkage^{57,63} suggests that this putative CRE interacts with the *SNCA* promoter. Given the MB specificity of this putative CRE and indications that it interacts with the *SNCA* promoter, we suspected that this region might be a driving force behind the MB-specific expression of *Snca*.

To test this hypothesis, we assayed whether, when lifted over to hg19, the central portion of this putative CRE (chr4: 90,721,063–90,722,122), is capable of directing appropriate reporter expression in transgenic zebrafish and mouse reporter assays. Stable transgenesis of zebrafish

indicates that this CRE directs reporter expression at 72 hpf in the locus coeruleus, a key population of catecholaminergic neurons preferentially degenerated in PD,⁷⁵ and along the catecholaminergic tract through the hindbrain, which is largely composed of DA neurons⁷⁶ (Figure 4C). Additionally, we observe reporter expression throughout the diencephalic catecholaminergic cluster with projections to the subpallium, which is analogous to mammalian dopaminergic projections from the ventral midbrain to the striatum.⁷⁷ Reporter expression in these transgenic zebrafish is largely consistent with an enhancer active in catecholaminergic populations.

To further evaluate this CRE in a mammalian system, we generated *lacZ* reporter mice and examined reporter activity across developmental time. Whole-mount E12.5 reporter mice indicate this enhancer directs exquisitely restricted expression in Th+ populations, including the dorsal root ganglia, extending into the sympathetic chain and throughout the cranial nerves (particularly the trigeminal). Additional diffuse staining is noted throughout the MB and FB (Figure 4D). Specifically examining the brains of *lacZ* animals at E15.5, we identified reporter expression in the MB and hypothalamus, as well as strong expression through the amygdala and piriform cortex and along the anterior portion of the sympathetic chain (Figure 4E); we see similar reporter patterns at P7 (Figure 4F). At P30, we detect reporter activity in the amygdala, hypothalamus, thalamus, periaqueductal gray area, brain stem, and importantly, in the *substantia nigra* and ventral tegmental area (Figure 4G). By contrast, in aged *lacZ* reporter mice (574 days old [$\sim$ 19 months]), we detect strong reporter expression only in the brain stem and observe weak reporter expression in the amygdala (Figure 4H). Collectively, the regions in which we detect reporter activity reflect those compromised in PD; Lewy bodies (aggregates of α -synuclein) have been detected in the locus coeruleus, sympathetic chain, amygdala, hypothalamus, ventral tegmental area, and periaqueductal gray area of PD-affected individuals.^{78–82} Critically, the preferential degradation of the *substantia nigra* is the pathological hallmark of PD progression.² This enhancer directs region-specific appropriate expression throughout development in key locations in concordance with *SNCA* activity in PD pathogenesis.

After confirming this CRE's regulatory activity in brain regions associated with PD, we next inspected this sequence for PD-associated variation. We sequenced across this interval in 986 individuals with PD and 992 controls and identified 14 variants (Table S4); four of these variants had a MAF greater than 5% and were common and present in both affected individuals and controls. Of these, two tightly linked variants ($r^2 = 0.934$; Table S5), dbSNP: rs2737024 (OR = 1.25, 95% CI = 1.09–1.44, p value = 0.002) and dbSNP: rs2583959 (OR = 1.22, 95% CI = 1.06–1.40, p value = 0.005), were significantly associated with PD (Table 1). These data support a role for variation within the enhancer in conferring PD risk.

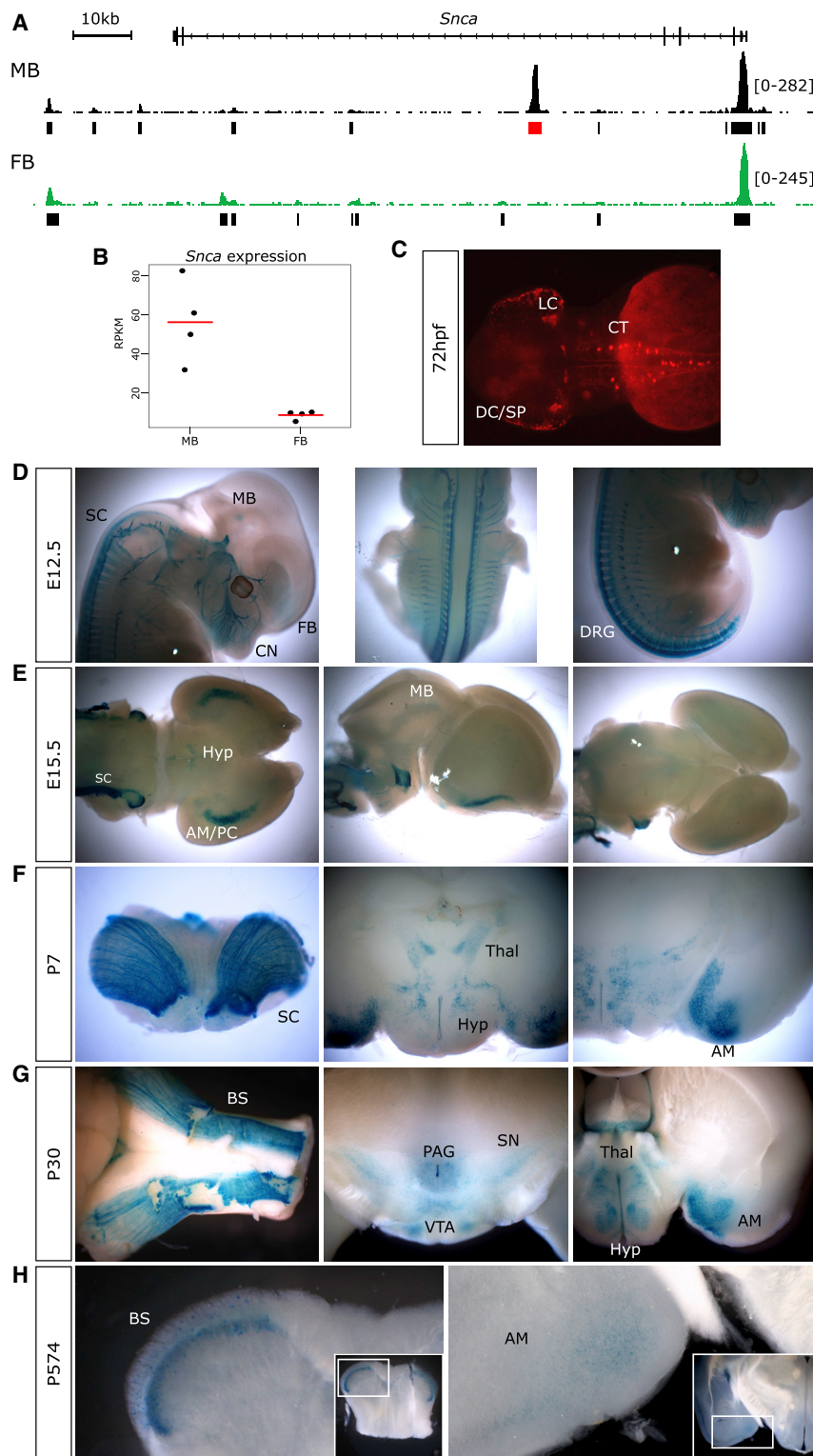


Figure 4. A MB-Specific Enhancer Directs Expression in Catecholaminergic Populations of Neurons Known to Parkinson Disease Biology

(A) An IGV track indicating the location of the MB-specific region of open chromatin located in intron 4 of *Snca*.

(B) *Snca* is differentially expressed between the MB and FB DA neurons. The red bar is the mean expression of the four replicates (black dots).

(C) At 72 hpf, stable transgenic zebrafish reporter assays indicate this putative CRE is capable of directing reporter expression in key catecholaminergic neuronal populations, including the locus coeruleus (LC), the catecholaminergic tract (CT) of the hindbrain, the diencephalic cluster (DC), and the subpallium (SP), into which the DC projects.

(D–H) Further studies in *lacZ* reporter assays in embryonic (E) and post-natal (P) mice indicate dynamic enhancer usage across developmental time.

(D) This enhancer directs expression throughout the MB, FB, dorsal root ganglia (DRG), sympathetic chain (SC), and cranial nerves (CN) of E12.5 mice.

(E) By E15.5, reporter expression is observed in the amygdala and/or piriform cortex (AM/PC), sympathetic chain, MB, and hypothalamus (Hyp).

(F) Patterns of reporter expression at P7 reflect those seen at E15.5.

(G) Reporter activity is observed at P30 in the amygdala; hypothalamus and thalamus (Thal); brain stem (BS); *substantia nigra* (SN); ventral tegmental area (VTA); and the periaqueductal gray area (PAG).

(H) In aged mice (P574), reporter expression is detected robustly in the brain stem and faintly in the amygdala.

pressed at appreciable levels in both MB and FB DA neurons (Figure 5D). Of the remaining four proteins, three (PEG10, SNRPA, and CHMP5) demonstrate an increased binding affinity for the minor risk allele over the major allele; this direction of effect is consistent with the overexpression paradigm by which *SNCA* confers PD risk.⁸ Interestingly, out of those proteins we identified, CHMP5 is the only one whose binding affinity is impacted by variant dbSNP: rs2583959, and our group has

To assess how these variants might impact enhancer function and thus PD risk, we assayed differential protein binding at these variants for >16,000 proteins.⁵⁸ In doing so, we identified five proteins whose binding is robustly impacted by these implicated variants: NOVA1, APOBEC3C, PEG10, SNRPA, and CHMP5 (Figures 5A–5C). Of these, all except APOBEC3C (RPKM ≤ 1) are ex-

recently implicated one of its family members, CHMP7, in conferring PD risk,⁶⁹ perhaps indicating a role for this family of proteins in PD. Although no single protein stands out, the increased affinity of proteins expressed in DA neurons for the risk alleles of the identified enhancer variants is consistent with a potential mechanistic contribution to *SNCA* expression and, therefore, to PD risk.

Table 1. Two Tightly Linked SNPs within the Enhancer are Significantly Associated with PD Risk

Variant	MA	MAF in PD-Affected Individuals (n = 986)	MAF in Controls (n = 992)	Association with PD	
				OR (95% CI)	p Value
rs7684892	A	0.063	0.069	0.93 (0.72–1.20)	0.562
rs17016188	C	0.082	0.061	1.35 (1.04–1.75)	0.023
rs2583959	G	0.317	0.271	1.22 (1.06–1.40)	0.005*
rs2737024	G	0.319	0.270	1.25 (1.09–1.44)	0.002*

Abbreviations are as follows: MA = minor allele; MAF = minor-allele frequency; OR = odds ratio; and CI = confidence interval.

Only variants with MAF > 0.05 were considered.

ORs, 95% CIs, and p values result from additive logistic regression models adjusted for sex and age at blood draw. p values ≤ 0.0125 were considered as statistically significant after a Bonferroni correction for multiple testing (*).

Finally, we set out to refine the haplotype structure and understand how this identified variation might be interacting with other variants at this locus. A panel of common variants had previously been genotyped across SNCA, and PD-associated haplotypes were identified.⁵³ After genotyping our PD-affected individuals and controls for a subset of this panel of variants (in addition to all enhancer-associated variants identified by sequencing [Table S6]), we identified a single haplotype that was significantly associated with PD (p value = 0.003) and that had a higher observed frequency in PD-affected individuals (28.3%) than in controls (23.4%; Table 2). This haplotype implicates some of the same variants as those in Guella et al.⁵³ (dbSNP: rs356220, dbSNP: rs737029) but also implicates dbSNP: rs356225 and dbSNP: rs356168, as well as the two enhancer-associated variants. Additionally, within the 1000 Genomes data, we observe that moderate LD structure exists between the lead GWAS variant (dbSNP: rs356182) and the enhancer variants ($r^2 = 0.418$, $D' = 0.745$) in the general European population. Despite the moderate LD, the risk allele of dbSNP: rs356182 falls in our identified PD-associated haplotype 94% of the time. Thus, it is likely that at least part of the risk captured by dbSNP: rs356182 can be attributed to these enhancer variants and the implicated haplotype reported here. Furthermore, this does not preclude additional variants from being present and contributing to the risk captured by the lead SNP because the dbSNP: rs356182 risk allele can occur in the absence of the enhancer-associated variants (i.e.: ~31% of EUR individuals with the dbSNP: rs356182 risk allele do not carry the risk alleles of the PD-associated enhancer variants). A schematic depiction of the variants, open-chromatin regions, chromatin interactions,^{57,63} and LD structure at this locus is presented in Figure 6. Of the variants, including dbSNP: rs356182, whose minor alleles define this PD-associated haplotype, only the two enhancer-associated variants and dbSNP: rs2737029 are identified as eQTLs for SNCA expression in any tissue in the GTEx database (Figure S7). Collectively, these data identify a catecholaminergic enhancer harboring common variation that is part of a larger haplotype associated with PD risk, likely by modulating SNCA activity.

Discussion

The identification and prioritization of biologically pertinent non-coding variation associated with disease remains challenging. Recent studies by our and other groups have emphasized the importance of cellular context in the identification of sequences harboring biologically pertinent variation and the genes they regulate. To this end, we used chromatin signatures from *ex vivo* isolated DA neurons to reveal biologically active sequences that harbor non-coding variation contributing to PD risk. We generated robust OCR catalogs for both MB and FB DA neurons, confirmed their capacity to act as enhancers, identified motifs that confer their regulatory potential, and notably, identified two variants that are located within a MB-specific enhancer and are associated with an increase in PD risk.

In contrast to strategies predicated solely on dissection of post-mortem tissues or on the differentiation of cultured cells, we leveraged the use of transgenic reporter mice to specifically isolate Th-expressing neurons from discrete neuro-anatomical (FB and MB) domains. Although our approach assays a more refined population of DA neurons than would be achieved via gross dissection, recent single-cell RNA-seq analyses of these same cells make clear that even within these highly restricted MB and FB populations there exist two primary cellular phenotypes.⁶⁹ The “homogenous” MB and FB populations each are comprised of an immature neuroblast population and a more mature, domain-specific, post-mitotic population of DA neurons. As a result, our OCR catalogs capture the chromatin accessibility from both of these states. These catalogs are demonstrably biologically relevant for our purposes, but future studies requiring even greater homogeneity might wish to consider single-cell ATAC-seq to refine these domains further.⁸³

Through *in silico* validation of the catalogs, we established that they are enriched for both sequence constraint and biological relevance in a manner consistent with function and their FB or MB origin. Furthermore, these sequences are frequently domain-appropriate enhancers, and each catalog captures a large fraction (77%) of previously validated MB and FB enhancers. Although more regions are shown to direct neuronal expression than to direct negative or non-neuronal expression, it is

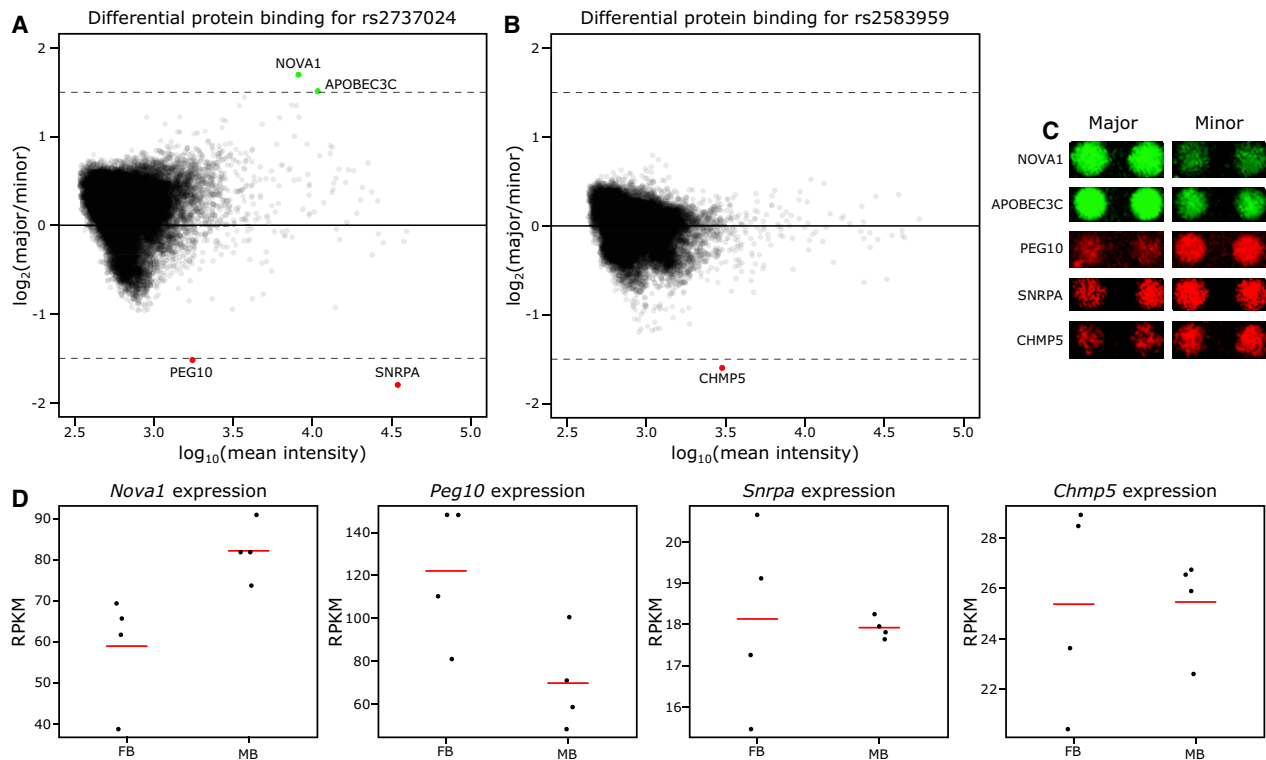


Figure 5. Identification of Proteins Whose Binding Is Impacted by the Implicated PD-risk SNPs

(A and B) MA plots for both dbSNP: rs2737024 and dbSNP: rs2583959 indicate the magnitude of the effect of the minor and major allele on binding. Cut-off for differential binding: $\log_2(\text{major/minor}) \geq 1.5$ or ≤ -1.5 .

(A) NOVA1 and APOBEC3C (green circles) bind at dbSNP: rs2737024 with greater affinity for the major allele, but PEG10 and SNRPA (red circles) have a greater affinity for the minor allele.

(B) CHMP5 (red circle) has a greater affinity for the minor allele of dbSNP: rs2583959.

(C) Representative images of the protein binding for each of the differentially bound proteins.

(D) Expression analysis in the MB and FB DA neurons for each of the differentially bound proteins indicates *Nova1*, *Peg10*, *Snrpa*, and *Chmp5* to be highly expressed in these populations, yet none of the *Apobec* family member genes are expressed (RPKM ≤ 1 , data not shown). The red bar is the mean expression of the four replicates (black dots).

interesting to note that almost half of the sequences previously documented as not directing expression *in vivo* are also represented in one or both of our catalogs.

Given the frequently dynamic nature of CRE activity, this overlap with negative regions most likely results from temporal differences in these assays. Our data indicate that these regions are accessible at E15.5, but the *lacZ* reporter assays were carried out at E11.5; regions that have been annotated as negative at E11.5 might be active at later time points and, as such, appear in our catalogs. As we moved from these unbiased functional comparisons to more highly selected ones, the potential impact of temporal differences became more pronounced. In mouse transgenic reporter assays, two of five assayed putative CREs directed detectable expression of *lacZ* in neuronal populations. Consistent with the temporally dynamic nature of CREs, when we tested these same regions in zebrafish across multiple developmental time points, we observed that four of the five sequences acted as neuronal enhancers.

By examining the sequence composition underlying the ATAC-seq peaks, we illuminate powerful vocabularies for both FB and MB DA neuron transcriptional regulatory con-

trol. The machine-learning algorithm gkm-SVM prioritized four transcription factor families (RFX, FOXA1/2, NR4A2, ASCL1/2) as conveying significant regulatory potential in the CRE catalogs. Of these, the RFX family had not previously been implicated in DA neuron biology. Although several of the RFX family members have been annotated as having expression in the cerebellum or fetal brain,⁷⁰ a role specifically in MB DA neurons has not previously been appreciated. By contrast, NR4A2 is canonically associated with MB DA neurons,^{67,68} is highly expressed in this population (139 RPKM), and was prioritized as having TF-conferring regulatory potential in these cells; however, TF footprinting fails to provide evidence supporting its activity. We postulate that this lack of a footprint might reflect the transient DNA-binding dynamics of NR4A2. Transcription factors with short DNA residence times often fail to reveal footprints, and nuclear receptors, such as NR4A2, have markedly transient DNA interactions.⁷⁴

Taken collectively, these data establish a robust biological platform from which PD-associated variation can be evaluated. To this end, an obvious candidate to interrogate was an apparent MB-specific open-chromatin domain within

Table 2. A Single Haplotype, Containing the Minor Alleles of the Implicated SNPs, Is Significantly Associated with PD Risk

		rs356220	rs356225	rs3857057	rs356168	rs10018362	rs2737029	rs62306323	rs2737024	rs2583959	rs17016188	rs7684892	rs7689942	Frequency in PD- Affected Individuals	Frequency in Controls	p Value
Haplotypes spanning the <i>SNCA</i> locus	1	C	C	A	G	T	T	C	A	C	T	G	C	0.015	0.027	0.012
	2	C	G	A	A	T	T	T	A	C	T	G	C	0.092	0.110	0.029
	3	C	C	A	G	C	C	C	A	C	T	A	T	0.039	0.048	0.184
	4	T	C	A	G	T	T	C	A	C	T	G	C	0.037	0.040	0.480
	5	C	G	A	A	T	T	C	A	C	T	G	C	0.380	0.397	0.593
	6	T	C	A	G	T	T	T	A	C	T	G	C	0.010	0.011	0.698
	7	C	G	A	A	T	C	C	G	G	T	G	C	0.009	0.012	0.944
	8	C	C	A	G	T	C	C	G	G	T	G	C	0.016	0.015	0.768
	9	T	C	G	G	C	C	C	A	C	T	A	T	0.021	0.017	0.360
	10	T	C	G	G	T	C	C	A	C	C	G	C	0.014	0.009	0.189
	11	T	C	G	G	C	C	C	A	C	C	G	C	0.057	0.044	0.124
	12	T	C	A	G	T	C	C	G	G	T	G	C	0.283	0.234	0.003*

Only haplotypes with frequency ≥ 0.01 were considered. Black boxes indicate the minor allele in Europeans.

p values result from score tests for association that were performed under an additive model and adjusted for sex and age at blood draw.

p values ≤ 0.0042 were considered as statistically significant after application of a Bonferroni correction for multiple testing (*).

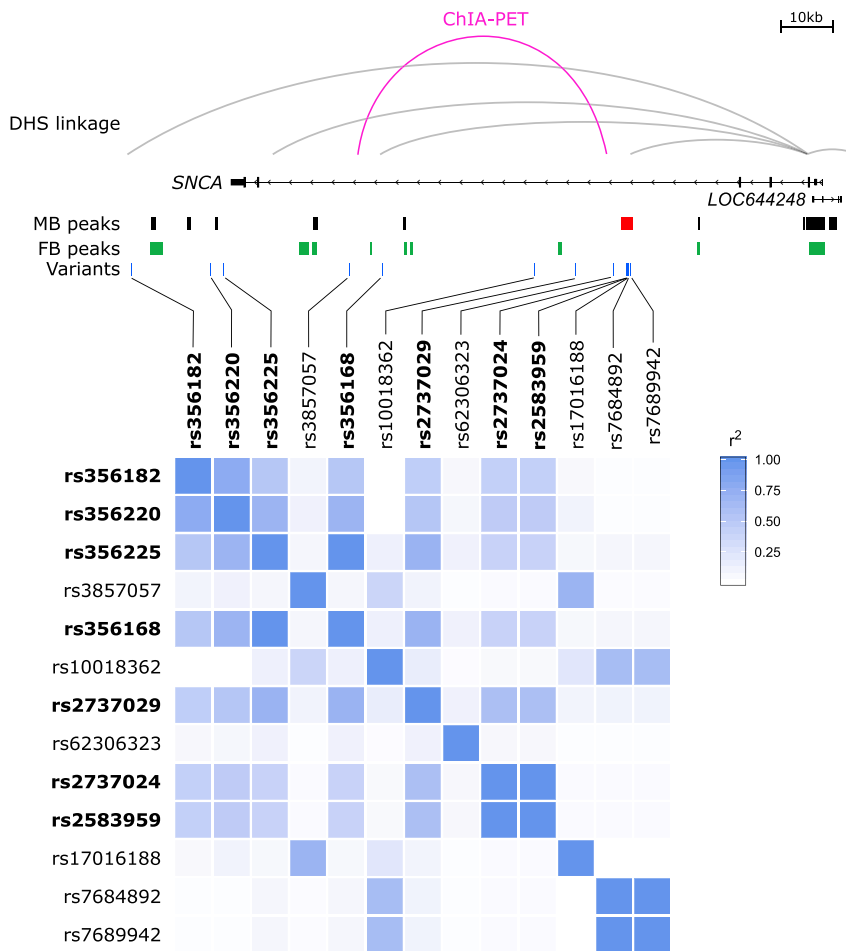


Figure 6. A Schematic of the Chromatin Interactions, LD Structure, Variation, and Open Chromatin at the SNCA Locus
Publicly available DNase hypersensitivity site (DHS) linkage analysis suggests that the promoter of SNCA possibly interacts with the identified MB-specific enhancer, the lead GWAS variant (dbSNP: rs356182), and a previously functionally validated variant (dbSNP: rs356168). ChIA-PET data suggest that the MB-specific enhancer might interact with variant dbSNP: rs356168. Open-chromatin data from DA neurons do not overlap with any variants at this locus or haplotype other than at the MB-specific enhancer. LD analysis at this locus indicates that despite the low LD structure between the lead GWAS variant (dbSNP: rs356182) and the enhancer-associated variants (dbSNP: rs2737024 and dbSNP: rs2583959), the variants are in the same haplotype. Therefore, the GWAS signal might, at least in part, be flagging the identified enhancer-associated variants.

intron 4 of the known PD-associated gene SNCA. We assayed the activity of this putative CRE in zebrafish and across the life course of mice and found it to be active in key catecholaminergic structures (e.g.: the *substantia nigra* and locus coeruleus) injured in PD, from mid-gestation until at least P30. Thereafter, the utilization of this enhancer in the brain is diminished and by late life appears restricted to the brainstem and amygdala. By the time of clinical presentation, individuals diagnosed with PD have already lost a significant proportion ($\geq 30\%$) of their nigral DA neurons;^{2,84} the observed biology of this CRE is consistent with a progressive pathogenic influence that begins early in life and renders these populations preferentially vulnerable to loss over an extensive period of time.

Sequencing this interval in PD-affected individuals and controls revealed two common variants (dbSNP: rs2737024 and dbSNP: rs2583959), individually associated with an increased risk of PD. After testing these variants for their effect on protein binding, we identified five proteins whose binding is affected, three of which (PEG10, SNRPA, and CHMP5) display greater affinity for the risk allele. Furthermore, we identified a larger haplotype containing these variants; this haplotype is also significantly associated with PD risk. Although none of the other SNPs in this haplotype overlap with CREs identified in the DA

neuron catalogs, there is significant functional evidence of the activity and contribution to PD risk by variant dbSNP: rs356168.⁸⁵ The same DHS correlation analysis^{57,63} that suggests an interaction between the SNCA promoter and our identified CRE also suggests an interaction between the SNCA promoter and the dbSNP: rs356168 variant. Additionally, ChIA-PET data^{25,57} indicate that sequences encompassing this variant might interact with our enhancer, suggesting a potential cooperative mode of action; Gupta and colleagues⁸⁶ recently proposed that such a paradigm takes place at the *EDN1* locus. We propose that the variants within the enhancer, independently or in concert with other variation within the identified haplotype, might act throughout an individual's lifespan to render key populations of catecholaminergic neurons vulnerable and thus increase PD risk in individuals harboring this variation.

This work emphasizes the value of biologically informed, cell-context-dependent guided searches for the identification of disease-associated and functional non-coding variation. Given the extent of non-coding GWAS-identified variation, the need for strategies to prioritize variants for functional follow-up is greater than ever. Here, we generate chromatin accessibility data from purified populations of DA neurons to generate catalogs of putative CREs. We have demonstrated how these data can be used to reveal non-coding variation contributing to PD risk; focusing on a single region of open chromatin at the SNCA locus, we uncover PD-associated variation therein and propose a model through which this sequence can contribute to normal DA neuronal biology and PD risk. There remains a plethora of information still to be explored in these catalogs, either through further

single-locus investigations or through massively parallel assays. For example, our MB DA-neuron OCR catalog overlaps variants at 20 of 49 (41%) PD-associated loci,^{9,10} all of which can be investigated further for the mechanisms by which they impact PD risk. Our work establishes a powerful paradigm, leveraging transgenic model systems to systematically generate cell-type-specific chromatin accessibility data and reveal disease-associated variation, in a manner that can be progressively guided by improved biological understanding.

Accession Numbers

ATAC-sequencing, RNA-sequencing and related data will be available at the Gene Expression Omnibus (GEO) under the accession number GSE122450.

Supplemental Data

Supplemental Data include seven figures and six tables and can be found with this article online at <https://doi.org/10.1016/j.ajhg.2018.10.018>.

Acknowledgments

This research undertaken at Johns Hopkins University School of Medicine was supported in part by awards from the NIH (NS62972 and MH106522 to A.S.M.; GM111514 to H.Z.; HG007348 to M.A.B.). The Mayo Clinic collection was supported in part by Morris K. Udall Center of Excellence in Parkinson's Disease Research (P50 NS072187), American Parkinson Disease Association Center, and The Mangurian Foundation for Lewy Body Research. O.A.R. is supported by NS078086 and NS10069 (NIH), W81XWH-17-1-0249 (Department of Defense), The Michael J. Fox Foundation, and The Little Family Foundation. Z.K.W. is supported by the Mayo Clinic Center for Regenerative Medicine; the Mayo Clinic Center for Individualized Medicine; the Mayo Clinic Neuroscience Focused Research Team (Cecilia and Dan Carmichael Family Foundation; and the James C. and Sarah K. Kennedy Fund for Neurodegenerative Disease Research at Mayo Clinic in Florida); a gift from Carl Edward Bolch, Jr., and Susan Bass Bolch; The Sol Goldman Charitable Trust; and Donald G. and Jodi P. Heeringa. Research conducted at the E.O. Lawrence Berkeley National Laboratory was performed under U.S. Department of Energy contract DE-AC02-05CH11231, University of California, and was supported by HG003988 (NIH) to L.A.P.

Declaration of Interests

The authors declare no competing interests.

Received: July 11, 2018

Accepted: October 17, 2018

Published: November 29, 2018

Web Resources

The 3D Genome Browser, <http://promoter.bx.psu.edu/hi-c/index.html>

Cistrome, <http://cistrome.org/ap/>

FastQC, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

The Genomic Regions Enrichment of Annotations Tool (GREAT), <http://great.stanford.edu>

LDlink, <https://ldlink.nci.nih.gov/>

The List of Genes and Their TSSs Used by GREAT: <http://bejerano.stanford.edu/help/download/attachments/2752609/mm9.great3.0.genes.txt>

The UCSC Table Browser, <http://genome.ucsc.edu>

R: A Language and Environment for Statistical Computing, <https://www.R-project.org/>

R *Corrplot* package, <https://github.com/taiyun/corrplot>

R *haplo.stats* package, <https://CRAN.R-project.org/package=haplo.stats>

R *LSD* package, <https://CRAN.R-project.org/package=LSD>

R *RColorBrewer* package, <https://CRAN.R-project.org/package=RColorBrewer>

rAggr, <http://raggr.usc.edu/>

VISTA Enhancer Browser, <https://enhancer.lbl.gov>

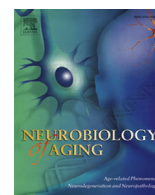
References

1. Ma, S.Y., R  ytt  , M., Rinne, J.O., Collan, Y., and Rinne, U.K. (1997). Correlation between neuromorphometry in the substantia nigra and clinical features in Parkinson's disease using disector counts. *J. Neurol. Sci.* *151*, 83–87.
2. Fearnley, J.M., and Lees, A.J. (1991). Ageing and Parkinson's disease: Substantia nigra regional selectivity. *Brain* *114*, 2283–2301.
3. Pringsheim, T., Jette, N., Frolkis, A., and Steeves, T.D.L. (2014). The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Mov. Disord.* *29*, 1583–1590.
4. Thomas, B., and Beal, M.F. (2007). Parkinson's disease. *Hum. Mol. Genet.* *16 Spec No. 2*, R183–R194.
5. Zarranz, J.J., Alegre, J., G  mez-Esteban, J.C., Lezcano, E., Ros, R., Ampuero, I., Vidal, L., Hoenicka, J., Rodr  guez, O., Atar  s, B., et al. (2004). The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann. Neurol.* *55*, 164–173.
6. Kr  ger, R., Kuhn, W., M  ller, T., Voitalla, D., Graeber, M., K  sel, S., Przuntek, H., Epplen, J.T., Sch  ls, L., and Riess, O. (1998). Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nat. Genet.* *18*, 106–108.
7. Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., et al. (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* *276*, 2045–2047.
8. Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R., et al. (2003). alpha-Synuclein locus triplication causes Parkinson's disease. *Science* *302*, 841.
9. Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., et al.; International Parkinson's Disease Genomics Consortium (IPDGC); Parkinson's Study Group (PSG) Parkinson's Research: The Organized GENetics Initiative (PROGENI); 23andMe; GenePD; NeuroGenetics Research Consortium (NGRC); Hussman Institute of Human Genomics (HIHG); Ashkenazi Jewish Dataset Investigator; Cohorts for Health and Aging Research in Genetic Epidemiology (CHARGE); North American Brain Expression Consortium (NABEC); United Kingdom Brain Expression Consortium (UKBEC); Greek Parkinson's Disease Consortium; and Alzheimer Genetic Analysis Group (2014). Large-scale meta-analysis of

- genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* 46, 989–993.
10. Chang, D., Nalls, M.A., Hallgrímsdóttir, I.B., Hunkapiller, J., van der Brug, M., Cai, F., Kerchner, G.A., Ayalon, G., Bingol, B., Sheng, M., et al.; International Parkinson's Disease Genomics Consortium; and 23andMe Research Team (2017). A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* 49, 1511–1516.
 11. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195.
 12. Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., and Snyder, M. (2012). Linking disease associations with regulatory information in the human genome. *Genome Res.* 22, 1748–1759.
 13. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49.
 14. Forrest, M.P., Zhang, H., Moy, W., McGowan, H., Leites, C., Di-onisio, L.E., Xu, Z., Shi, J., Sanders, A.R., Greenleaf, W.J., et al. (2017). Open chromatin profiling in hiPSC-derived neurons prioritizes functional noncoding psychiatric risk variants and highlights neurodevelopmental loci. *Cell Stem Cell* 21, 305–318.e8.
 15. Fullard, J.F., Giambartolomei, C., Hauberg, M.E., Xu, K., Voloudakis, G., Shao, Z., Bare, C., Dudley, J.T., Mattheisen, M., Robakis, N.K., et al. (2017). Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci. *Hum. Mol. Genet.* 26, 1942–1951.
 16. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell* 169, 1177–1186.
 17. Lee, D., Gorkin, D.U., Baker, M., Strober, B.J., Asoni, A.L., McCallion, A.S., and Beer, M.A. (2015). A method to predict the impact of regulatory variants from DNA sequence. *Nat. Genet.* 47, 955–961.
 18. Praetorius, C., Grill, C., Stacey, S.N., Metcalf, A.M., Gorkin, D.U., Robinson, K.C., Van Otterloo, E., Kim, R.S.Q., Bergsteinsdottir, K., Ogmundsdottir, M.H., et al. (2013). A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell* 155, 1022–1033.
 19. Gong, S., Zheng, C., Doughty, M.L., Losos, K., Didkovsky, N., Schambra, U.B., Nowak, N.J., Joyner, A., Leblanc, G., Hatten, M.E., and Heintz, N. (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* 425, 917–925.
 20. Westerfeld, M. (2007). *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio)* (Eugene: Univ. Oregon Press).
 21. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218.
 22. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
 23. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
 24. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.
 25. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
 26. Shin, H., Liu, T., Manrai, A.K., and Liu, X.S. (2009). CEAS: cis-regulatory element annotation system. *Bioinformatics* 25, 2605–2606.
 27. Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y., et al. (2011). Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.* 12, R83.
 28. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 32, D493–D496.
 29. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44 (W1), W160–5.
 30. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360.
 31. Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* 41, e108.
 32. Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121.
 33. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5, R80.
 34. Mudge, J.M., and Harrow, J. (2015). Creating reference gene annotation for the mouse C57BL6/J genome assembly. *Mamm. Genome* 26, 366–378.
 35. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., and Wang, L. (2014). CrossMap: A versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* 30, 1006–1007.
 36. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
 37. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13, 134.
 38. Hansen, K.D., Irizarry, R.A., and Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* 13, 204–216.
 39. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050.

40. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 495–501.
41. Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444, 499–502.
42. Poulin, F., Nobrega, M.A., Plajzer-Frick, I., Holt, A., Afzal, V., Rubin, E.M., and Pennacchio, L.A. (2005). In vivo characterization of a vertebrate ultraconserved enhancer. *Genomics* 85, 774–781.
43. Kothary, R., Clapoff, S., Darling, S., Perry, M.D., Moran, L.A., and Rossant, J. (1989). Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* 105, 707–714.
44. Fisher, S., Grice, E.A., Vinton, R.M., Bessling, S.L., Urasaki, A., Kawakami, K., and McCallion, A.S. (2006). Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat. Protoc.* 1, 1297–1305.
45. Ghandi, M., Lee, D., Mohammad-Noori, M., and Beer, M.A. (2014). Enhanced regulatory sequence prediction using gapped k-mer features. *PLoS Comput. Biol.* 10, e1003711.
46. Zorita, E., Cuscó, P., and Filion, G.J. (2015). Starcode: Sequence clustering based on all-pairs search. *Bioinformatics* 31, 1913–1919.
47. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.
48. Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
49. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. *Genome Biol.* 8, R24.
50. Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* 21, 447–455.
51. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018.
52. Adey, A., Morrison, H.G., Asan, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., and Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* 11, R119.
53. Guella, I., Evans, D.M., Szu-Tu, C., Nosova, E., Bortnick, S.F., Goldman, J.G., Dalrymple-Alford, J.C., Geurtsen, G.J., Litvan, I., Ross, O.A., et al.; SNCA Cognition Study Group (2016). α -synuclein genetic variability: A biomarker for dementia in Parkinson disease. *Ann. Neurol.* 79, 991–999.
54. Hughes, A.J., Daniel, S.E., Kilford, L., and Lees, A.J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J. Neurol. Neurosurg. Psychiatry* 55, 181–184.
55. Schaid, D.J., Rowland, C.M., Tines, D.E., Jacobson, R.M., and Poland, G.A. (2002). Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am. J. Hum. Genet.* 70, 425–434.
56. Machiela, M.J., and Chanock, S.J. (2015). LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 31, 3555–3557.
57. Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., et al. (2018). The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* 19, 151.
58. Jeong, J.S., Jiang, L., Albino, E., Marrero, J., Rho, H.S., Hu, J., Hu, S., Vera, C., Bayron-Poueymiroy, D., Rivera-Pacheco, Z.A., et al. (2012). Rapid identification of monospecific monoclonal antibodies using a human proteome microarray. *Mol. Cell. Proteomics* 11.
59. Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C., et al. (2013). DNA methylation presents distinct binding sites for human transcription factors. *eLife* 2, e00726.
60. Heintz, N. (2004). Gene expression nervous system atlas (GENSAT). *Nat. Neurosci.* 7, 483.
61. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
62. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192.
63. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. *Nature* 489, 75–82.
64. Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser—A database of tissue-specific human enhancers. *Nucleic Acids Res.* 35, D88–D92.
65. Stott, S.R.W., Metzakopian, E., Lin, W., Kaestner, K.H., Hen, R., and Ang, S.-L. (2013). Foxa1 and foxa2 are required for the maintenance of dopaminergic properties in ventral midbrain neurons at late embryonic stages. *J. Neurosci.* 33, 8022–8034.
66. Arenas, E. (2008). Foxa2: The rise and fall of dopamine neurons. *Cell Stem Cell* 2, 110–112.
67. Prakash, N., and Wurst, W. (2006). Development of dopaminergic neurons in the mammalian brain. *Cell. Mol. Life Sci.* 63, 187–206.
68. Smits, S.M., Ponnio, T., Conneely, O.M., Burbach, J.P.H., and Smidt, M.P. (2003). Involvement of Nurr1 in specifying the neurotransmitter identity of ventral midbrain dopaminergic neurons. *Eur. J. Neurosci.* 18, 1731–1738.
69. Hook, P.W., McClymont, S.A., Cannon, G.H., Law, W.D., Morton, A.J., Goff, L.A., and McCallion, A.S. (2018). Single-cell RNA-Seq of mouse dopaminergic neurons informs candidate gene selection for sporadic Parkinson disease. *Am. J. Hum. Genet.* 102, 427–446.
70. Sugiaman-Trapman, D., Vitezic, M., Jouhilahti, E.-M., Mathelier, A., Lauter, G., Misra, S., Daub, C.O., Kere, J., and Swoboda, P. (2018). Characterization of the human RFX transcription factor family by regulatory and target gene analysis. *BMC Genomics* 19, 181.
71. Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403, 916–921.

72. Kittappa, R., Chang, W.W., Awatramani, R.B., and McKay, R.D.G. (2007). The *foxa2* gene controls the birth and spontaneous degeneration of dopamine neurons in old age. *PLoS Biol.* 5, e325.
73. Caiazzo, M., Dell'Anno, M.T., Dvoretzkova, E., Lazarevic, D., Taverna, S., Leo, D., Sotnikova, T.D., Menegon, A., Roncaglia, P., Colciago, G., et al. (2011). Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature* 476, 224–227.
74. Sung, M.-H., Guertin, M.J., Baek, S., and Hager, G.L. (2014). DNase footprint signatures are dictated by factor dynamics and DNA sequence. *Mol. Cell* 56, 275–285.
75. Zarow, C., Lyness, S.A., Mortimer, J.A., and Chui, H.C. (2003). Neuronal loss is greater in the locus coeruleus than nucleus basalis and substantia nigra in Alzheimer and Parkinson diseases. *Arch. Neurol.* 60, 337–341.
76. Kastnerhuber, E., Kratochwil, C.F., Ryu, S., Schweitzer, J., and Driever, W. (2010). Genetic dissection of dopaminergic and noradrenergic contributions to catecholaminergic tracts in early larval zebrafish. *J. Comp. Neurol.* 518, 439–458.
77. Rink, E., and Wullmann, M.F. (2001). The teleostean (zebrafish) dopaminergic system ascending to the subpallium (striatum) is located in the basal diencephalon (posterior tuberculum). *Brain Res.* 889, 316–330.
78. Seidel, K., Mahlke, J., Siswanto, S., Krüger, R., Heinsen, H., Auburger, G., Bouzrou, M., Grinberg, L.T., Wicht, H., Korf, H.-W., et al. (2015). The brainstem pathologies of Parkinson's disease and dementia with Lewy bodies. *Brain Pathol.* 25, 121–135.
79. Wakabayashi, K., Mori, F., Tanji, K., Orimo, S., and Takahashi, H. (2010). Involvement of the peripheral nervous system in synucleinopathies, tauopathies and other neurodegenerative proteinopathies of the brain. *Acta Neuropathol.* 120, 1–12.
80. Wakabayashi, K., and Takahashi, H. (1997). Neuropathology of autonomic nervous system in Parkinson's disease. *Eur. Neurol.* 38 (Suppl 2), 2–7.
81. Braak, H., Braak, E., Yilmazer, D., de Vos, R.A., Jansen, E.N., Bohl, J., and Jellinger, K. (1994). Amygdala pathology in Parkinson's disease. *Acta Neuropathol.* 88, 493–500.
82. Langston, J.W., and Forno, L.S. (1978). The hypothalamus in Parkinson disease. *Ann. Neurol.* 3, 129–133.
83. Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D.U., Zhang, Y., Sos, B.C., Afzal, V., Dickel, D.E., et al. (2018). Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* 21, 432–439.
84. Greffard, S., Verny, M., Bonnet, A.-M., Beinis, J.-Y., Gallinari, C., Meaume, S., Piette, F., Hauw, J.-J., and Duyckaerts, C. (2006). Motor score of the Unified Parkinson Disease Rating Scale as a good predictor of Lewy body-associated neuronal loss in the substantia nigra. *Arch. Neurol.* 63, 584–588.
85. Soldner, F., Stelzer, Y., Shivalila, C.S., Abraham, B.J., Latourelle, J.C., Barrasa, M.I., Goldmann, J., Myers, R.H., Young, R.A., and Jaenisch, R. (2016). Parkinson-associated risk variant in distal enhancer of α -synuclein modulates target gene expression. *Nature* 533, 95–99.
86. Gupta, R.M., Hadaya, J., Trehan, A., Zekavat, S.M., Roselli, C., Klarin, D., Emdin, C.A., Hilvering, C.R.E., Bianchi, V., Mueller, C., et al. (2017). A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. *Cell* 170, 522–533.e15.



No evidence for *DNM3* as genetic modifier of age at onset in idiopathic Parkinson's disease

Victoria Berge-Seidl^{a,b,*}, Lasse Pihlstrøm^a, Zbigniew K. Wszolek^c, Owen A. Ross^d, Mathias Toft^{a,b}

^a Department of Neurology, Oslo University Hospital, Oslo, Norway

^b Faculty of Medicine, University of Oslo, Oslo, Norway

^c Department of Neurology, Mayo Clinic, Jacksonville, FL, USA

^d Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA



ARTICLE INFO

Article history:

Received 9 November 2017

Received in revised form 14 September 2018

Accepted 15 September 2018

Available online 22 September 2018

Keywords:

Parkinson's disease

Age at onset

DNM3

ABSTRACT

Parkinson's disease (PD) is a disorder with highly variable clinical phenotype. The identification of genetic variants modifying age at onset and other traits is of great interest because it may provide insight into disease mechanisms and potential therapeutic targets. A variant in the *DNM3* gene (rs2421947) has been reported as a genetic modifier of age at onset in *LRRK2*-associated PD. To test the possible effect of genetic variation in *DNM3* on age at onset in idiopathic PD, we examined rs2421947 in a total of 5918 patients with PD from seven data sets. We also assessed the potential effect of all common variants in the *DNM3* locus. There was no significant association between rs2421947 and age at onset in any of the individual studies. Meta-analysis of the seven studies was nonsignificant and the between-study heterogeneity was minimal. No other common variants within the *DNM3* locus affected age at onset. In conclusion, we find no evidence of an association between *DNM3* variants and age at onset in idiopathic PD.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Parkinson's disease (PD) is a common neurodegenerative disorder with a complex etiology. A small proportion of patients with PD have a monogenic form of the disease with highly penetrant mutations following an autosomal dominant or recessive inheritance pattern. However, most PD cases are idiopathic, presumably caused by the combined action of multiple genetic variants in interplay with epigenetic, environmental, and stochastic factors (Lill, 2016). To date, genome-wide association studies (GWASs) have linked more than 40 risk loci to PD susceptibility (Chang et al., 2017).

In addition to affecting the risk of disease development, genetic variants may also affect the clinical phenotype once the disease has manifested. The clinical heterogeneity of PD is characterized by a marked variation in the pattern and progression of motor, cognitive, and other nonmotor symptoms. Both rare monogenic mutations and common genetic variants have been shown to contribute to the

clinical heterogeneity of PD (Pihlstrøm et al., 2016; Puschmann, 2013; Winder-Rhodes et al., 2013).

There is a broad range of age at onset in PD, varying between debut in early adulthood to patients reaching the 8th and 9th decade of life before the onset of motor symptoms. Several studies have investigated the effect of PD risk loci on the onset age. Cumulative genetic risk scores calculated across PD risk loci have been shown to have a small, but consistent, effect on age at onset (Escott-Price et al., 2015; Lill et al., 2015; Nalls et al., 2015; Pihlstrøm and Toft, 2015). In addition, risk loci having the greatest effect in PD GWAS meta-analysis (*GBA*, *SNCA*, *MAPT*, and *TMEM175*) (Nalls et al., 2014) are reported to individually be associated with age at onset (Brockmann et al., 2013; Davis et al., 2016; Lill et al., 2015; Nalls et al., 2015).

Linking common variation to age at onset represents an interesting step toward a better understanding of how genetics affects the PD phenotype. In a recent genome-wide study of genetic modifiers of age at onset in leucine-rich repeat kinase 2 (*LRRK2*) p.G2019S carriers, a *DNM3* haplotype tagged by rs2421947 was identified (Trinh et al., 2016). This *LRRK2* mutation is the most frequent genetic cause of PD in many populations, estimated to have a frequency of 1% in white North

* Corresponding author at: Department of Neurology, Oslo University Hospital, P.O. Box 4950 Nydalen, N-0424 Oslo, Norway. Tel.: +47 23079022.

E-mail address: victoria.berge@medisin.uio.no (V. Berge-Seidl).

American and as high as 39% in North African Arab patients (Healy et al., 2008).

LRRK2 mutations cause an autosomal dominant form of PD often segregating in families, while GWASs provide consistent evidence that common variation at this locus also modulates disease risk. Because *LRRK2* is part of the genetic background for idiopathic PD, variants that modulate age at onset in *LRRK2* parkinsonism may also exert an effect in a much wider group of patients. Herein we report analyses of data from seven studies of PD from Europe and North America to determine associations between the *DNM3* rs2421947 variant and age at onset of idiopathic PD. To study the potential effect of other *DNM3* variants, we also performed a complete assessment of common variation in the gene locus.

2. Methods

2.1. Study populations

We analyzed individual-level genotypes from seven different data sets. Samples originated from genetic studies of PD from Oslo University Hospital and Mayo Clinic Jacksonville. The remaining five data sets were publicly available and selected due to available individual genotype information in PD patients with a reported age at onset. The following four data sets were accessed from dbGaP: 1) CIDR: Genome Wide Association Study in Familial Parkinson Disease (Accession number: phs000126.v1.p1), 2) Mayo-Perlegen LEAPS (Linked Efforts to Accelerate Parkinson's Solutions) Collaboration (Accession number: phs000048.v1.p1), 3) National Institute of Neurological Disorders and Stroke (NINDS) Genome-Wide Genotyping in Parkinson's Disease (Accession number: phs000089.v3.p2), 4) Genome-Wide Association Study of Parkinson Disease: Genes and Environment performed by the NeuroGenetics Research Consortium (NGRC) (Accession number: phs000196.v2.p1). The last data set is made available by the Parkinson's Progression Markers Initiative (<http://www.ppmi-info.org>).

All patients have been examined by a neurologist. The Oslo and Mayo Clinic patients were diagnosed according to the revised UKPDSBB criteria. Inclusion criteria for the other studies have previously been described (Hamza et al., 2010; Maraganore et al., 2005; Nalls et al., 2016; Pankratz et al., 2009; Simon-Sanchez et al., 2009). Age at onset is either reported as age at symptom onset or age at diagnosis. In the current analysis, patients with PD reporting other than Caucasian non-Hispanic ethnicity have been excluded along with *LRRK2* p.G2019S carriers that were identified by imputation or had previously been genotyped. A large subset of the Oslo and Mayo Clinic patients was sequenced for genes causing Mendelian forms of PD, and mutation carriers were excluded from the analysis. In addition, no known carriers of Mendelian PD mutations in the publicly available data sets were included in the analysis. Demographic characteristics are summarized in Table 1. All participants gave written informed consent. Sample and data collection at each study site was approved by local ethics

committees. The study was approved by the Regional Committee for Medical Research Ethics (Oslo, Norway).

2.2. Genotyping and quality control

Mayo Clinic patients were genotyped for rs2421947 using a TaqMan assay. A subset of genotypes was validated by Sanger sequencing with complete concordance. For the other studies genome-wide genotypes were available. Oslo samples were genotyped using the Illumina Infinium OmniExpress v.1.1 array. Pre-imputation quality filtering included filtering out variants with genotype rate <0.95 or Hardy-Weinberg equilibrium $p < 10^{-6}$ and removal of individuals with call rate <0.95, excess heterozygosity >4 standard deviations from mean, and evidence of cryptic relatedness or sex-check failure. Population outliers were excluded from analysis after inspection of principal component analysis plot (>2.5 standard deviation from mean). Details of genotyping methods and data quality assessments for the publicly available GWASs and the Parkinson's Progression Markers Initiative study are described in previous publications (Hamza et al., 2010; Maraganore et al., 2005; Nalls et al., 2016; Pankratz et al., 2009; Simon-Sanchez et al., 2009).

Common, pruned, genotyped variants (minor allele frequency >0.05 and $r^2 < 0.5$) were used to calculate principal components for each of the six genome-wide data sets. For all genome-wide data sets, imputation was performed using the Michigan imputation server (Das et al., 2016) with reference data from the Haplotype Reference Consortium (McCarthy et al., 2016) setting a quality cutoff of $r^2 > 0.3$ for variants included in the analysis. A set of common, pruned variants from each imputed data set was merged to assess cryptic relatedness across studies. Duplicates and related samples were removed. The final sample sets included in analysis after quality control comprise a total of 5918 patients with PD (Table 1).

2.3. Statistical analyses

First we tested all seven studies individually for association between the *DNM3* rs2421947 variant and age at onset under an additive linear regression model. In an alternative binary analysis, age at onset was dichotomized by the median onset calculated across all seven data sets (61 years of age) and logistic regression was used to test for association with the *DNM3* rs2421947 variant within each data set. In both regression analyses, sex and the first five principal components were used as covariates in the genome-wide data sets, while sex was the single covariate in analysis of the Mayo Clinic study. Association analyses were performed in the PLINK 1.9 software (<https://www.cog-genomics.org/plink/1.9/>) (Chang et al., 2015). Inverse-variance, fixed-effects meta-analysis of the seven studies was conducted using the Genome-wide Association Meta-Analysis software (Magi and Morris, 2010). Between-study heterogeneity was assessed using Cochran's Q test and Higgins's I^2 statistic. Owing to the increased burden of recessive

Table 1
Demographic characteristics of study samples

Study	Population	Patients with PD	% Male	Age at onset $\pm$ SD	Genotyping method
Oslo	Norway	472	64	55.9 $\pm$ 11.2	Illumina Infinium OmniExpress v.1.1.
Mayo	USA	987	64	65.5 $\pm$ 11.8	Taqman assay
CIDR	North America, Europe and Australia	823	59	62.0 $\pm$ 10.7	Illumina HumanCNV370 BeadChip
LEAPS	USA	439	62	60.9 $\pm$ 11.1	Perlegen DNA chip (85k SNP markers)
NINDS	USA	912	60	58.4 $\pm$ 13.2	Illumina HumanHap550 BeadChip
NGRC	USA	1971	68	58.4 $\pm$ 11.9	Illumina HumanOmni1_Quad
PPMI	USA and Europe	314	68	59.7 $\pm$ 9.8	Illumina NeuroX

Key: CIDR, Center for Inherited Disease Research; LEAPS, Linked Efforts to Accelerate Parkinson's Solutions; NINDS, National Institute of Neurological Disorders and Stroke; NGRC, NeuroGenetics Research Consortium; PPMI, Parkinson's Progression Markers Initiative; PD, Parkinson's disease; SD, standard deviation.

disease-causing mutations in early-onset PD (Puschmann, 2013), we repeated the linear regression analysis excluding all patients with an age at onset <40 years of age.

Next, common variation within *DNM3* as well as 100kb upstream and downstream of the gene was analyzed by linear regression in the six genome-wide data sets. Variants with a minor allele frequency below 0.01 in each data set were excluded from analysis. Association analysis, both within the individual data sets and meta-analysis, were performed as described for *DNM3* rs2421947. To estimate the degree of multiple testing, we generated a combined, pruned data set using a cutoff of linkage disequilibrium $> r^2 = 0.5$, leaving 226 independent variants. Adjusting for 226 independent tests by Bonferroni correction, a p -value < 0.0002 was considered significant. Power calculations were performed with the function `pwr.f2.test` in the R package `pwr` (version 1.2–1; <https://cran.r-project.org/web/packages/pwr/index.html>).

3. Results

We found no significant association between rs2421947 and age at onset in any of the individual studies, both when analyzed as a quantitative trait and in the alternative binary analysis. The frequency of the alternative allele G of rs2421947 was similar in all the seven studies, varying between 54% and 56%. Meta-analysis of the seven studies analyzed as a quantitative trait was nonsignificant ($p = 0.55$), and the between-study heterogeneity was minimal ($I^2 = 0\%$, $p = 0.77$). Results from linear regression analysis of the individual studies and meta-analysis are shown in Fig. 1. We obtained similar results when we excluded patients with an age at onset <40 years from the analysis. Meta-analysis of age at onset as a binary trait was also nonsignificant (OR = 1.03, 95% CI = 0.96–1.11, $p = 0.44$) and between-study heterogeneity was minimal ($I^2 = 0\%$, $p = 0.85$).

Next, we analyzed all common variants within the *DNM3* gene and the flanking genomic region. Variants covered in at least four of the six analyzed genome-wide data sets were included in the meta-analysis, constituting 1932 variants. None of the meta-analyzed variants had a significant association with age at onset when corrected for multiple testing. Results from the extended *DNM3* analysis are provided in Supplementary Table 1. Tests for statistical heterogeneity indicated that heterogeneity was low ($I^2 < 50\%$) for the vast majority (97%) of the meta-analyzed variants. Studies of the combined impact of PD risk loci on age at onset report a phenotypic variance explained by the calculated genetic risk score of 0.6% and 0.7%, with the signal mostly being driven by two individually age at

onset-associated variants (Lill et al., 2015; Nalls et al., 2015). Assuming a variance explained of $\sim 0.5\%$ for the tested variant, we have a power of over 99% for the primary analysis of rs2421947 ($N = 5918$) and a power of 89.5% to achieve a p -value of 0.0002 ($N = 4931$) in the extended *DNM3* analysis.

The effect of rs2421947 on age at onset was also assessed in the LRRK2 p.G2019S carriers ($N = 39$) that had been excluded from the aforementioned analyses, although this test was limited by a small sample size. Association with age at onset analyzed as a quantitative trait, with sex and data set as covariables, was nonsignificant for rs2421947, with the trend for direction of effect reversed relative to the original report by Trinh et al., 2016 (effect allele = G, $\beta = 1.79$, 95% CI = -3.31 – 6.89 , $p = 0.50$). Trinh et al. report a correlation between rs2421947 genotype and *DNM3* mRNA levels in striatal brain tissue (Trinh et al., 2016). We explored the Genotype-Tissue Expression (GTEx) Portal (version 7; <https://www.gtexportal.org/home/>) and found that significant expression quantitative trait loci (eQTLs) for *DNM3* are reported in cerebellar hemisphere and cerebellum, but not in any of the other brain regions examined by the GTEx project. Interestingly, rs2421947 and variants in high linkage disequilibrium ($r^2 \geq 0.6$) are not reported as significant eQTLs for *DNM3* in any brain tissue.

4. Discussion

In this study, we found no evidence for a modifying effect of rs2421947 or other common *DNM3* variants on age at onset in idiopathic PD. A *DNM3* haplotype tagged by rs2421947 was identified by Trinh et al. as a modifier of age at onset in LRRK2 p.G2019S carriers. They reported that the median age at onset of *DNM3* GG homozygotes was 12.5 years younger than that of CC homozygotes. This is a large difference in the onset age compared with the effect of other variants associated with age at onset in PD, and could be meaningful in the clinical setting. As shown by the 95% confidence interval of our primary analysis, it is highly unlikely that we did not detect an effect altering the onset of PD more than a few months per G allele.

We performed a meta-analysis including a total of 5918 patients with PD, and the high number of analyzed individuals is a strength of our study. However, by including cases from different study sites with variations in study design, heterogeneity may be introduced in meta-analysis of the genetic data. We assessed this and found low heterogeneity. We accounted for population substructure within the individual studies by including five eigenvectors in the regression model. The meta-analyzed studies use mostly self-

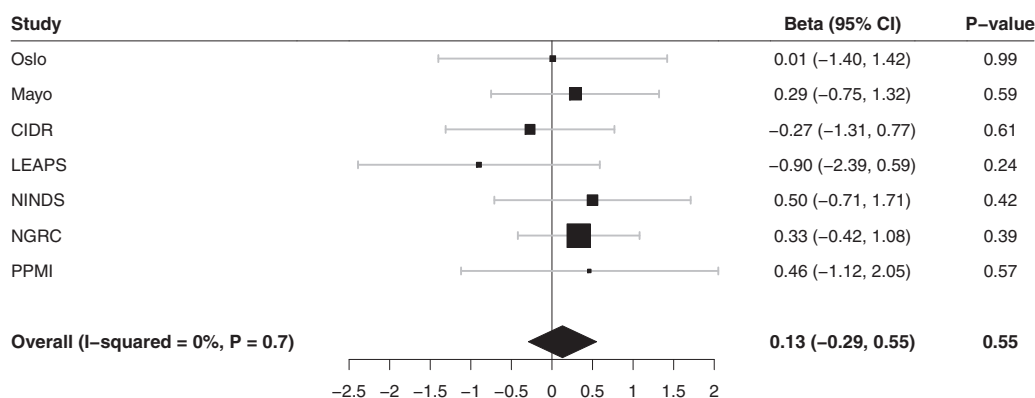


Fig. 1. Study-specific and meta-analysis results for the *DNM3* variant rs2421947. Forest plot showing the effect of rs2421947 on age at onset in idiopathic Parkinson's disease in individual studies and meta-analysis. The effect size of the G allele is given as a beta estimate with a 95% confidence interval (CI). The size of the squares indicates the size of the data sets. Abbreviations: CIDR, Center for Inherited Disease Research; LEAPS, Linked Efforts to Accelerate Parkinson's Solutions; NINDS, National Institute of Neurological Disorders and Stroke; NGRC, NeuroGenetics Research Consortium; PPMI, Parkinson's Progression Markers Initiative.

reported symptom onset. Age at onset is subjective and may be prone to recall bias. Nevertheless, the reliability of self- and family-reported age at onset compared with medical records is high and all three methods have been regarded as valid (Reider et al., 2003).

A recent study of LRRK2 p.G2019S carriers in the Spanish population did not find any association between *DNM3* rs2421947 and age at onset of PD (Fernandez-Santiago et al., 2018). Linkage patterns vary among populations, and the possibility that disease-relevant variation could be tagged by different genetic markers in Europeans/North Americans as compared with the Arab-Berber population studied by Trinh et al. prompted us to extend our analysis to all imputed common variants across the *DNM3* locus. The genomic region flanking the *DNM3* gene is included in our analysis to cover potential regulatory variants, although this increases the multiple testing burden. On the other hand, regulatory variants affecting *DNM3* expression may reside in an even more distal part of the genome not covered in our analysis. eQTL data could be used to identify potential regulatory variants and reduce the number of tests to adjust for. However, the currently available databases are incomplete because eQTLs, in addition to depending on tissue and cell type, also may vary between different physiological conditions (Albert and Kruglyak, 2015).

Despite recent efforts to elucidate the genetic architecture behind age at onset and other clinical characteristics of PD, the vast majority of genetic variation affecting PD phenotypes remains unexplained. Many studies have limited their analysis to known risk loci of PD, while attempts at identifying novel genetic modifiers of age at onset have proven challenging. Genetic modifiers of age at onset may be limited to subgroups of patients carrying specific mutations or susceptibility variants. Variations in the *MAPT* gene have been found to be associated with age at onset in patients with PD carrying a *LRRK2* mutation (Gan-Or et al., 2012; Golub et al., 2009). A recent GWAS of age at onset analyzed PD patients with and without a family history of the disease separately. No significant association was found in those without a family history of PD, while two signals were detected in individuals reporting to have a first- or second-degree relative with PD. Both these signals mapped to gene regions that are not known to affect PD risk (Hill-Burns et al., 2016).

Discovering genetic modifiers of phenotype in PD is important because it may provide insight into disease mechanisms and help in identifying potential therapeutic targets. *DNM3* has previously not been identified by GWASs of disease risk or onset age (Chang et al., 2017; Hill-Burns et al., 2016; Latourelle et al., 2009; Nalls et al., 2014). We found no association between common *DNM3* variants and age at onset in idiopathic PD, but the possible contribution of rare variants within this genetic locus cannot be excluded. Variability within the *DNM3* locus may be a specific modifier of *LRRK2* parkinsonism, although this has yet to be replicated in independent studies. *DNM3* encodes the protein dynamin-3 which is highly expressed in neurons (Raimondi et al., 2011). The *LRRK2* protein has been shown to interact with dynamin-3 and other members of the dynamin GTPase superfamily that regulate membrane dynamics important for endocytosis and mitochondrial morphology (Stafa et al., 2014). Trinh et al. report a correlation between rs2421947 genotype and *DNM3* mRNA levels in striatal tissue, but such an association is not observed in the GTEx portal. There are several methodological differences between these analyses and additional evidence is needed before conclusions regarding rs2421947 as an eQTL for *DNM3* in the brain can be drawn.

Further insight into disease mechanisms may be gained by examining complex genetic interactions. A novel epistatic interaction between two genetic variants was reported by a recent study incorporating genetic, molecular, and clinical data into models to predict motor progression in PD (Latourelle et al., 2017). Larger

patient cohorts with comprehensive characterization of disease phenotype will benefit future studies of how genetics affect clinical heterogeneity.

Disclosure statement

The authors report no conflicts of interest concerning the research related to the article.

Acknowledgements

The authors thank all study participants, investigators, and funding partners for their contribution to the data analyzed in the current publication. The funding organizations had no role in study design, data collection and analysis, decision to publish, or preparation of the article. VBS and MT are funded by a grant from the South-Eastern Norway Regional Health Authority. MT is also supported by the Research Council of Norway. LP is supported by the Norwegian Health Association. Mayo Clinic is a Morris K. Udall Parkinson's Disease Research Center of Excellence (NINDS P50 #NS072187 to OR and ZKW); ZKW is also funded by Mayo Clinic Center for Individualized Medicine, Mayo Clinic Neuroscience Focused Research Team (Cecilia and Dan Carmichael Family Foundation, and the James C. and Sarah K. Kennedy Fund for Neurodegenerative Disease Research at Mayo Clinic in Florida), the gift from Carl Edward Bolch, Jr. and Susan Bass Bolch, the Sol Goldman Charitable Trust, and Donald G. and Jodi P. Heeringa. OR is supported by NINDS R01 NS078086, Department of Defense award W81XWH-16-PRMRP-IIRA, Michael J. Fox Foundation. The analyses presented in the current publication are based on the use of study data downloaded from the dbGaP website, under dbGaP accession phs000126.v1.p1, phs000048.v1.p1, phs000089.v3.p2 and phs000196.v2.p1. Data were also obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners including AbbVie, Avid Radiopharmaceuticals, Biogen, BioLegend, Bristol-Myers Squibb, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, Sanofi Genzyme, Servier, Takeda, Teva and UCB.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.neurobiolaging.2018.09.022>.

References

- Albert, F.W., Kruglyak, L., 2015. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* 16, 197–212.
- Brockmann, K., Schulte, C., Hauser, A.K., Lichtner, P., Huber, H., Maetzel, W., Berg, D., Gasser, T., 2013. SNCA: major genetic modifier of age at onset of Parkinson's disease. *Mov. Disord.* 28, 1217–1221.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Giga-Science* 4, 7.
- Chang, D., Nalls, M.A., Hallgrimsdottir, I.B., Hunkapiller, J., van der Brug, M., Cai, F., Kerchner, G.A., Ayalon, G., Bingol, B., Sheng, M., Hinds, D., Behrens, T.W., Singleton, A.B., Bhangale, T.R., Graham, R.R., 2017. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* 49, 1511–1516.
- Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.J., Chew, E.Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.R., Iacono, W.G., Swaroop, A., Scott, L.J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G.R., Fuchsberger, C., 2016. Next-generation genotype imputation service and methods. *Nat. Genet.* 48, 1284–1287.

- Davis, A.A., Andruska, K.M., Benitez, B.A., Racette, B.A., Perlmuter, J.S., Cruchaga, C., 2016. Variants in GBA, SNCA, and MAPT influence Parkinson disease risk, age at onset, and progression. *Neurobiol. Aging* 37, 209.e201–209.e207.
- Escott-Price, V., Nalls, M.A., Morris, H.R., Lubbe, S., Brice, A., Gasser, T., Heutink, P., Wood, N.W., Hardy, J., Singleton, A.B., Williams, N.M., 2015. Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann. Neurol.* 77, 582–591.
- Fernandez-Santiago, R., Garrido, A., Infante, J., Gonzalez-Aramburu, I., Sierra, M., Fernandez, M., Valdeoriola, F., Munoz, E., Compta, Y., Marti, M.J., Rios, J., Tolosa, E., Ezquerro, M., 2018. alpha-synuclein (SNCA) but not dynamin 3 (DNM3) influences age at onset of leucine-rich repeat kinase 2 (LRRK2) Parkinson's disease in Spain. *Mov. Disord.* 33, 637–641.
- Gan-Or, Z., Bar-Shira, A., Mirelman, A., Gurevich, T., Giladi, N., Orr-Urtreger, A., 2012. The age at motor symptoms onset in LRRK2-associated Parkinson's disease is affected by a variation in the MAPT locus: a possible interaction. *J. Mol. Neurosci.* 46, 541–544.
- Golub, Y., Berg, D., Calne, D.B., Pfeiffer, R.F., Uitti, R.J., Stoessl, A.J., Wszolek, Z.K., Farrer, M.J., Mueller, J.C., Gasser, T., Fuchs, J., 2009. Genetic factors influencing age at onset in LRRK2-linked Parkinson disease. *Parkinsonism Relat. Disord.* 15, 539–541.
- Hamza, T.H., Zabetian, C.P., Tenesa, A., Laederach, A., Montimurro, J., Yearout, D., Kay, D.M., Doherty, K.F., Paschall, J., Pugh, E., Kusel, V.I., Collura, R., Roberts, J., Griffith, A., Samii, A., Scott, W.K., Nutt, J., Factor, S.A., Payami, H., 2010. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat. Genet.* 42, 781–785.
- Healy, D.G., Falchi, M., O'Sullivan, S.S., Bonifati, V., Durr, A., Bressman, S., Brice, A., Aasly, J., Zabetian, C.P., Goldwurm, S., Ferreira, J.J., Tolosa, E., Kay, D.M., Klein, C., Williams, D.R., Marras, C., Lang, A.E., Wszolek, Z.K., Berciano, J., Schapira, A.H., Lynch, T., Bhatia, K.P., Gasser, T., Lees, A.J., Wood, N.W., 2008. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol.* 7, 583–590.
- Hill-Burns, E.M., Ross, O.A., Wissemann, W.T., Soto-Ortolaza, A.I., Zarepari, S., Suda, J., Lynch, T., Wszolek, Z.K., Silburn, P.A., Mellick, G.D., Ritz, B., Scherzer, C.R., Zabetian, C.P., Factor, S.A., Breheny, P.J., Payami, H., 2016. Identification of genetic modifiers of age-at-onset for familial Parkinson's disease. *Hum. Mol. Genet.* 25, 3849–3862.
- Latourelle, J.C., Beste, M.T., Hadzi, T.C., Miller, R.E., Oppenheim, J.N., Valko, M.P., Wuest, D.M., Church, B.W., Khalil, I.G., Hayete, B., Venuto, C.S., 2017. Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's disease: a longitudinal cohort study and validation. *Lancet Neurol.* 16, 908–916.
- Latourelle, J.C., Pankratz, N., Dumitriu, A., Wilk, J.B., Goldwurm, S., Pezzoli, G., Mariani, C.B., DeStefano, A.L., Halter, C., Gusella, J.F., Nichols, W.C., Myers, R.H., Foroud, T., 2009. Genomewide association study for onset age in Parkinson disease. *BMC Med. Genet.* 10, 98.
- Lill, C.M., 2016. Genetics of Parkinson's disease. *Mol. Cell. Probes* 30, 386–396.
- Lill, C.M., Hansen, J., Olsen, J.H., Binder, H., Ritz, B., Bertram, L., 2015. Impact of Parkinson's disease risk loci on age at onset. *Mov. Disord.* 30, 847–850.
- Magi, R., Morris, A.P., 2010. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* 11, 288.
- Maraganore, D.M., de Andrade, M., Lesnick, T.G., Strain, K.J., Farrer, M.J., Rocca, W.A., Pant, P.V., Frazer, K.A., Cox, D.R., Ballinger, D.G., 2005. High-resolution whole-genome association study of Parkinson disease. *Am. J. Hum. Genet.* 77, 685–693.
- McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L.J., Zhang, H., Mahajan, A., Veldink, J., Peters, U., Pato, C., van Duijn, C.M., Gillies, C.E., Gandin, I., Mezzavilla, M., Gilly, A., Cocca, M., Traglia, M., Angius, A., Barrett, J.C., Boomsma, D., Branham, K., Breen, G., Brummett, C.M., Busonero, F., Campbell, H., Chan, A., Chen, S., Chew, E., Collins, F.S., Corbin, L.J., Smith, G.D., Dedoussis, G., Dorr, M., Farmaki, A.E., Ferrucci, L., Forer, L., Fraser, R.M., Gabriel, S., Levy, S., Groop, L., Harrison, T., Hattersley, A., Holmen, O.L., Hveem, K., Kretzler, M., Lee, J.C., McGue, M., Meitinger, T., Melzer, D., Min, J.L., Mohlke, K.L., Vincent, J.B., Nauck, M., Nickerson, D., Palotie, A., Pato, M., Pirastu, N., McInnis, M., Richards, J.B., Sala, C., Salomaa, V., Schlessinger, D., Schoenherr, S., Slagboom, P.E., Small, K., Spector, T., Stambolian, D., Tuke, M., Tuomilehto, J., Van den Berg, L.H., Van Rheenen, W., Volker, U., Wijmenga, C., Toniolo, D., Zeggini, E., Gasparini, P., Sampson, M.G., Wilson, J.F., Frayling, T., de Bakker, P.I., Swertz, M.A., McCarrall, S., Kooperberg, C., Dekker, A., Altshuler, D., Willer, C., Iacono, W., Ripatti, S., Soranzo, N., Walter, K., Swaroop, A., Cucca, F., Anderson, C.A., Myers, R.M., Boehnke, M., McCarthy, M.I., Durbin, R., 2016. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 48, 1279–1283.
- Nalls, M.A., Escott-Price, V., Williams, N.M., Lubbe, S., Keller, M.F., Morris, H.R., Singleton, A.B., 2015. Genetic risk and age in Parkinson's disease: continuum not stratum. *Mov. Disord.* 30, 850–854.
- Nalls, M.A., Keller, M.F., Hernandez, D.G., Chen, L., Stone, D.J., Singleton, A.B., 2016. Baseline genetic associations in the Parkinson's progression markers initiative (PPMI). *Mov. Disord.* 31, 79–85.
- Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., Schulte, C., Keller, M.F., Arepalli, S., Letson, C., Edsall, C., Stefansson, H., Liu, X., Pliner, H., Lee, J.H., Cheng, R., Ikram, M.A., Ioannidis, J.P., Hadjigeorgiou, G.M., Bis, J.C., Martinez, M., Perlmutter, J.S., Goate, A., Marder, K., Fiske, B., Sutherland, M., Xiromerisiou, G., Myers, R.H., Clark, L.N., Stefansson, K., Hardy, J.A., Heutink, P., Chen, H., Wood, N.W., Houlden, H., Payami, H., Brice, A., Scott, W.K., Gasser, T., Bertram, L., Eriksson, N., Foroud, T., Singleton, A.B., 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* 46, 989–993.
- Pankratz, N., Wilk, J.B., Latourelle, J.C., DeStefano, A.L., Halter, C., Pugh, E.W., Doherty, K.F., Gusella, J.F., Nichols, W.C., Foroud, T., Myers, R.H., 2009. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum. Genet.* 124, 593–605.
- Pihlstrom, L., Morset, K.R., Grimstad, E., Vitelli, V., Toft, M., 2016. A cumulative genetic risk score predicts progression in Parkinson's disease. *Mov. Disord.* 31, 487–490.
- Pihlstrom, L., Toft, M., 2015. Cumulative genetic risk and age at onset in Parkinson's disease. *Mov. Disord.* 30, 1712–1713.
- Puschmann, A., 2013. Monogenic Parkinson's disease and parkinsonism: clinical phenotypes and frequencies of known mutations. *Parkinsonism Relat. Disord.* 19, 407–415.
- Raimondi, A., Ferguson, S.M., Lou, X., Armbruster, M., Paradise, S., Giovedi, S., Messa, M., Kono, N., Takasaki, J., Cappello, V., O'Toole, E., Ryan, T.A., De Camilli, P., 2011. Overlapping role of dynamin isoforms in synaptic vesicle endocytosis. *Neuron* 70, 1100–1114.
- Reider, C.R., Halter, C.A., Castelluccio, P.F., Oakes, D., Nichols, W.C., Foroud, T., 2003. Reliability of reported age at onset for Parkinson's disease. *Mov. Disord.* 18, 275–279.
- Simon-Sanchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G., Kruger, R., Federoff, M., Klein, C., Goate, A., Perlmutter, J., Bonin, M., Nalls, M.A., Illig, T., Gieger, C., Houlden, H., Steffens, M., Okun, M.S., Racette, B.A., Cookson, M.R., Foote, K.D., Fernandez, H.H., Traynor, B.J., Schreiber, S., Arepalli, S., Zonozi, R., Gwinn, K., van der Brug, M., Lopez, G., Chanock, S.J., Schatzkin, A., Park, Y., Hollenbeck, A., Gao, J., Huang, X., Wood, N.W., Lorenz, D., Deuschl, G., Chen, H., Riess, O., Hardy, J.A., Singleton, A.B., Gasser, T., 2009. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.* 41, 1308–1312.
- Stafa, K., Tsika, E., Moser, R., Musso, A., Glauser, L., Jones, A., Biskup, S., Xiong, Y., Bandopadhyay, R., Dawson, V.L., Dawson, T.M., Moore, D.J., 2014. Functional interaction of Parkinson's disease-associated LRRK2 with members of the dynamin GTPase superfamily. *Hum. Mol. Genet.* 23, 2055–2077.
- Trinh, J., Gustavsson, E.K., Vilarino-Guell, C., Bortnick, S., Latourelle, J., McKenzie, M.B., Tu, C.S., Nosova, E., Khinda, J., Milnerwood, A., Lesage, S., Brice, A., Tazir, M., Aasly, J.O., Parkkinen, L., Haytural, H., Foroud, T., Myers, R.H., Sassi, S.B., Hentati, E., Nabli, F., Farhat, E., Amouri, R., Hentati, F., Farrer, M.J., 2016. DNM3 and genetic modifiers of age of onset in LRRK2 Gly2019Ser parkinsonism: a genome-wide linkage and association study. *Lancet Neurol.* 15, 1248–1256.
- Winder-Rhodes, S.E., Evans, J.R., Ban, M., Mason, S.L., Williams-Gray, C.H., Foltynie, T., Duran, R., Mencacci, N.E., Sawcer, S.J., Barker, R.A., 2013. Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. *Brain* 136 (Pt 2), 392–399.

Association study between multiple system atrophy and TREM2 p.R47H

Kotaro Ogaki, MD, PhD, Michael G. Heckman, MS, Shunsuke Koga, MD, PhD, Yuka A. Martens, PhD, Catherine Labbé, PhD, Oswaldo Lorenzo-Betancor, MD, PhD, Ronald L. Walton, BSc, Alexandra I. Soto, BSc, Emily R. Vargas, MPH, Shinsuke Fujioka, MD, Ryan J. Uitti, MD, Jay A. van Gerpen, MD, William P. Cheshire, MD, Steven G. Younkin, MD, PhD, Zbigniew K. Wszolek, MD, Phillip A. Low, MD, Wolfgang Singer, MD, Guojun Bu, PhD, Dennis W. Dickson, MD, and Owen A. Ross, PhD

Correspondence
Dr. Ross
ross.owen@mayo.edu

Neurol Genet 2018;4:e257. doi:10.1212/NXG.0000000000000257

Abstract

Objective

The triggering receptor expressed on myeloid cells 2 (TREM2) p.R47H substitution (rs75932628) is a risk factor for Alzheimer disease (AD) but has not been well studied in relation to the risk of multiple system atrophy (MSA); the aim of this study was to evaluate the association between the TREM2 p.R47H variant and the risk of MSA.

Methods

A total of 168 patients with pathologically confirmed MSA, 89 patients with clinically diagnosed MSA, and 1,695 controls were included. TREM2 p.R47H was genotyped and assessed for association with MSA. Positive results in the Taqman genotyping assay were confirmed by Sanger sequencing. The primary comparison involved patients with pathologically confirmed MSA and controls due to the definitive MSA diagnosis in the pathologically confirmed series.

Results

We identified TREM2 p.R47H in 3 patients with pathologically confirmed MSA (1.79%), 1 patient with clinically diagnosed MSA (1.12%), and 7 controls (0.41%). Minimal AD pathology was observed for the pathologically confirmed MSA p.R47H carriers. For the primary comparison of patients with pathologically confirmed MSA and controls, risk of disease was significantly higher for p.R47H carriers (odds ratio [OR]: 4.39, $p = 0.033$). When supplementing the 168 pathologically confirmed patients with the 89 clinically diagnosed and examining the combined MSA series, the association with TREM2 p.R47H remained significant (OR: 3.81, $p = 0.034$).

Conclusions

Our preliminary results suggest that the TREM2 p.R47H substitution may be a risk factor for MSA, implying a link to neuroinflammatory processes, especially microglial activation. Validation of this finding will be important, given our relatively small sample size; meta-analytic approaches will be needed to better define the role of this variant in MSA.

From the Department of Neuroscience (K.O., S.K., Y.A.M., C.L., O.L.-B., R.L.W., A.I.S., S.F., S.G.Y., G.B., D.W.D., O.A.R.), Mayo Clinic, Jacksonville, FL; Department of Neurology (K.O.), Juntendo University Shizuoka Hospital, Izunokunishi, Shizuoka, Japan; Division of Biomedical Statistics and Informatics (M.G.H., E.R.V.), Mayo Clinic, Jacksonville, FL; Department of Neurology (R.J.U., J.A.v.G., W.P.C., Z.K.W.), Mayo Clinic, Jacksonville, FL; Department of Neurology (P.A.L., W.S.), Mayo Clinic, Rochester, MN; Mayo Graduate School (O.A.R.), Neurobiology of Disease, Jacksonville, FL; and Department of Clinical Genomics (O.A.R.), Jacksonville, FL.

Funding information and disclosures are provided at the end of the article. Full disclosure form information provided by the authors is available with the full text of this article at Neurology.org/NG.

The Article Processing Charge was funded by the authors.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Glossary

AD = Alzheimer disease; **FTD** = frontotemporal dementia; **MSA** = multiple system atrophy; **MSA-C** = MSA with predominant olivopontocerebellar involvement; **MSA-P** = MSA with predominant striatonigral involvement; **OR** = odds ratio; **PD** = Parkinson disease; **TREM2** = triggering receptor expressed on myeloid cells 2.

Multiple system atrophy (MSA) is a rare, rapidly progressing, adult-onset neurodegenerative disease that is clinically characterized by parkinsonism, autonomic failure, pyramidal symptoms, and cerebellar ataxia and pathologically defined by the presence of α -synuclein-positive glial cytoplasmic inclusions.¹ Uncovering genetic risk factors for MSA has proven difficult because of the challenge of accumulating a reasonable sample size and potential for clinical misdiagnosis. Nonetheless, studies have nominated several genes that may be associated with the risk of MSA.^{1,2} The first genome-wide association study of MSA was recently completed, and although no genome-wide significant associations were identified, suggestive associations were identified involving variants in *MAPT*, *FBXO47*, *ELOVL7*, and *EDN1*.³ Clearly, much remains to be understood regarding the genetics of MSA, and replication of previously nominated MSA susceptibility loci remains crucial.

A relatively rare substitution p.R47H (rs75932628) in the triggering receptor expressed on myeloid cells 2 (TREM2) protein is now a well-established risk factor for Alzheimer disease (AD) and has been suggested to play a role in other neurodegenerative disorders such as frontotemporal dementia (FTD) and Parkinson disease (PD).⁴ Of interest, mutations in the *TREM2* gene were originally reported as a genetic cause of Nasu-Hakola disease, a rare leukodystrophy characterized by progressive presenile dementia and bone cysts.⁴ Given the important role of microglia in neuroinflammation, we evaluated the association between the TREM2 p.R47H variant and the risk of MSA using a predominantly neuropathologically confirmed MSA series.

Methods

Study participants

This study included 168 patients with pathologically confirmed MSA, 89 patients with clinically diagnosed MSA, and 1,695 healthy controls. The pathologically confirmed MSA series consisted of all cases obtained from the Mayo Clinic brain bank for neurodegenerative disorders in Jacksonville, FL, between 1998 and 2017. Pathologically confirmed MSA was diagnosed by a single neuropathologist (D.W.D.) as previously described.⁵ Immunohistochemistry for α -synuclein (NACP, 1:3,000 rabbit polyclonal) was performed on sections of the basal forebrain, striatum, midbrain, pons, medulla, and cerebellum to establish a neuropathologic diagnosis. MSA cases were pathologically subclassified as MSA with predominant striatonigral involvement (MSA-P), MSA with predominant olivopontocerebellar involvement (MSA-C), and MSA with equally severe involvement of striatonigral and

olivopontocerebellar systems (MSA-mixed).⁵ One case was not assigned pathologic subtype because only the pons was available for histology.

The clinically diagnosed patients (N = 81 probable MSA, N = 8 possible MSA) were seen at the Mayo Clinic in Rochester, MN (N = 35), or Jacksonville, FL (N = 54), where diagnosis was made according to the criteria of Gilman et al.⁶ Clinically diagnosed MSA cases were free of notable cognitive symptoms that might suggest the presence of an overlapping dementia or other neurodegenerative conditions. Pathologic assessment and clinical diagnoses were made without the knowledge of the TREM2 p.R47H genotype. Controls were also from the Mayo Clinic in Rochester, MN (N = 991), or Jacksonville, FL (N = 704). Controls were free of neurologic symptoms and without a family history of movement disorders. All participants were unrelated non-Hispanic Caucasians. The primary comparison of TREM2 p.R47H involved the patients with pathologically confirmed MSA and controls. In secondary analysis, we combined the patients with pathologically confirmed MSA and patients with clinically diagnosed MSA and compared this combined MSA series with the controls. Table 1 shows characteristics of patients with MSA and controls.

Standard protocol approvals, registrations, and patient consents

The Mayo Clinic IRB approved the use of human participants in this study.

Genetic analysis

Genomic DNA was extracted from brain tissue or peripheral blood monocytes using standard protocols. Genotyping of TREM2 exon 2 variant rs75932628 (NM_018965.3: c.140G>A, p.R47H) was performed using a custom TaqMan Allelic Discrimination Assay on an ABI 7900HT Fast Real-Time PCR system (Applied Bio-systems, Foster City, CA) according to the manufacturer's instructions (sequences available upon request). Genotype calls were made using SDS 2.2.2 software. Positive or ambiguous results in the TaqMan assay were also confirmed and resolved via Sanger sequencing. There was no evidence of a departure from Hardy-Weinberg equilibrium in controls ($p = 0.93$).

Statistical analysis

We evaluated the association between TREM2 p.R47H and MSA using logistic regression models. Unadjusted models were first examined. Subsequently, although it is uncertain whether adjusting for any variable is reasonable, given the rare nature of this variant, we did also adjust our logistic regression models for age and sex in a sensitivity analysis. Odds ratios (ORs) and 95%

Table 1 Characteristics of patients with MSA and controls

Variable	Patients with pathologically confirmed MSA (N = 168)	Patients with clinically diagnosed MSA (N = 89)	Controls (N = 1,695)
Age at onset in patients with MSA or age at blood draw in controls (y)	58 (39, 88)	64 (46, 83)	76 (30, 88)
Age at death (y)	66 (47, 91)	NA	NA
Sex			
Male	96 (57.1%)	53 (59.6%)	786 (46.4%)
Female	72 (42.9%)	36 (40.4%)	909 (53.6%)
MSA subtype			
MSA-P	73 (43.5%)	62 (69.7%)	NA
MSA-C	27 (16.1%)	26 (29.2%)	NA
Mixed ^a	67 (39.9%)	1 (1.1%)	NA
Unclassified ^b	1 (0.6%)	0 (0.0%)	NA
Braak NFT stage			
0	29 (17.8%)	NA	NA
I	61 (37.4%)	NA	NA
II	49 (30.1%)	NA	NA
III	16 (9.8%)	NA	NA
IV	5 (3.1%)	NA	NA
V	2 (1.2%)	NA	NA
VI	1 (0.6%)	NA	NA
Thal amyloid phase			
0	88 (55.7%)	NA	NA
1	33 (20.9%)	NA	NA
2	18 (11.4%)	NA	NA
3	9 (5.7%)	NA	NA
4	8 (5.1%)	NA	NA
5	2 (1.3%)	NA	NA

Abbreviations: MSA = multiple system atrophy; MSA-C = MSA with predominant olivopontocerebellar involvement; MSA-P = MSA with predominant striatonigral involvement; NA = not applicable.

The sample median (minimum, maximum) is given for age. Information was unavailable regarding age at onset (N = 49), Braak NFT stage (N = 5), and Thal amyloid phase (N = 10) for the patients with pathologically confirmed MSA.

^a MSA with equally severe involvement of striatonigral and olivopontocerebellar systems was classified as MSA-mixed.⁵

^b One patient with MSA was considered to be "unclassified" because only the pons was available for histologic assessment; therefore, we could not diagnose this patient as MSA-C or MSA-mixed.

confidence intervals were estimated. $p \leq 0.05$ was considered statistically significant. All statistical analyses were performed using SAS (version 9.4; SAS Institute, Inc., Cary, NC).

Results

An evaluation of the association between TREM2 p.R47H and MSA is shown in table 2. The p.R47H variant was observed in 3 patients with pathologically confirmed MSA (1.79%). Of these 3 p.R47H carriers, neuropathologic examination was available for 2 and revealed MSA subtypes of MSA-P and MSA-C, Braak neurofibrillary tangle stages of I and III, and Thal amyloid phases of 0 and 0. For the 89 patients with clinically diagnosed MSA, there was 1 p.R47H carrier (1.12%, from Mayo Jacksonville), and this patient's MSA subtype was probable MSA-P. There were 7 carriers of p.R47H in the controls (0.41%, N = 4 Mayo Jacksonville, N = 3 Mayo Rochester), corresponding to a minor allele frequency of 0.21%, which is similar to the frequency reported in the ExAC database (Cambridge, MA [exac.broad-institute.org]) for this population (0.26%).⁷ For the primary comparison of patients with pathologically confirmed MSA and controls, risk of MSA was significantly higher for p.R47H carriers (OR = 4.39, $p = 0.033$). When combining the pathologically confirmed and clinically diagnosed MSA series, the risk of MSA (vs controls) was also significantly higher (OR = 3.81, $p = 0.034$). In a sensitivity analysis adjusting for age and sex, the findings regarding the patients with pathologically confirmed MSA (OR = 3.55, $p = 0.076$) and the combined MSA series (OR = 3.17, $p = 0.078$) remained relatively consistent though weakened slightly and were no longer quite significant.

Discussion

TREM2 p.R47H is a well-established risk factor for AD⁴; however, results have been mixed for FTD and PD,⁴ while studies involving amyotrophic lateral sclerosis and dementia with Lewy bodies mostly point to a lack of association.⁴ Of interest, the results of our current study suggest that the TREM2 p.R47H variant may be associated with an increased risk of MSA. With a notably higher frequency in our combined MSA series (1.6%) compared with controls (0.4%), we observed a nearly 4-fold increased risk of MSA in p.R47H carriers. It is important that this finding was strongest in our pathologically confirmed MSA series. The absence of notable AD pathology in pathologically confirmed MSA p.R47H carriers indicates that the observed association was not driven by such pathology. The association was attenuated slightly when adjusting for age and sex; however, those results should not be overinterpreted because it is uncertain whether attempting any adjustment is reasonable, given the low frequency of p.R47H. These preliminary findings suggest that TREM2 p.R47H may play a role in susceptibility to MSA and that further study is warranted.

To the best of our knowledge, only 1 association study between TREM2 p.R47H and MSA has been conducted to date. In a study of 407 Chinese patients with clinically diagnosed MSA and 869 controls, the p.R47H substitution was observed in 1 MSA patient and no controls ($p = 0.14$).⁸ This negative result is not surprising because TREM2 p.R47H is extremely rare in East Asians.

Table 2 Association between TREM2 p.R47H and MSA

Disease group	N	No. (%) with TREM2 p.R47H	Unadjusted analysis		Adjusting for age and sex	
			OR (95% CI)	p Value	OR (95% CI)	p Value
Controls	1,695	7 (0.41)	1.00 (reference)	NA	1.00 (reference)	NA
Patients with pathologically confirmed MSA	168	3 (1.79)	4.39 (1.12–17.12)	0.033	3.55 (0.88–14.38)	0.076
All patients with MSA (pathologically confirmed and clinically diagnosed)	257	4 (1.56)	3.81 (1.11–13.12)	0.034	3.17 (0.88–11.45)	0.078

Abbreviations: CI = confidence interval; MSA = multiple system atrophy; NA = not applicable; OR = odds ratio; TREM2 = triggering receptor expressed on myeloid cells.

ORs, 95% CIs, and *p* values result from logistic regression models.

For age, the age that was adjusted for was age at death for patients with pathologically confirmed MSA, age at MSA onset for patients with clinically diagnosed MSA, and age at blood draw for controls.

How the TREM2 p.R47H variant could increase risk of MSA is an important topic for further study. MSA is regarded as an oligodendrogliaopathy, but the mechanism of demyelination remains unclear. While demyelination depends primarily on oligodendrocytes, microglia also contribute to myelination and myelin homeostasis.⁹ In the CNS, TREM2 is predominantly expressed in microglia, and it has been shown to play critical roles in phagocytic clearance of apoptotic cells and disease-associated molecules and modulating microglial immune response.¹⁰ A recent study showed that TREM2 is required for myelin debris removal and remyelination after cuprizone-induced demyelination.¹¹ We hypothesize that TREM2 p.R47H, a loss-of-function mutation, leads to impaired clearance of myelin debris and aberrant microglial activation, which result in increased risk of MSA.

Our results provide evidence that the TREM2 p.R47H substitution may be a genetic susceptibility factor for MSA. Our study is limited by the relatively small number of patients with pathologically confirmed MSA, and therefore, validation will be crucial. Recently, the clinical diagnostic accuracy of MSA was reported to be 62%, further underscoring the importance of using neuropathologically confirmed MSA cases in research studies.¹² We hope that these initial promising results will form the basis for future examinations of TREM2 p.R47H in MSA.

Author contributions

K. Ogaki: drafting/revising the manuscript; study concept and design; and interpretation of data. M. G. Heckman and S. Koga: drafting/revising the manuscript and interpretation of data. Y.A. Martens: drafting/revising the manuscript. C. Labbé and O. Lorenzo-Betancor: revising the manuscript. R.L. Walton, A.I. Soto, and E.R. Vargas: revising the manuscript and interpretation of data. S. Fujioka: revising the manuscript. R.J. Uitti, J. Van Gerpen, and W.P. Cheshire: revising the manuscript and providing material for this study. S.G. Younkin: revising the manuscript. Z.K. Wszolek: revising the manuscript, providing material and funding for this study, and IRB committee approval. P.A. Low and W. Singer: revising

the manuscript and providing material for this study. G. Bu: revising the manuscript. D.W. Dickson: drafting/revising the manuscript, providing material and funding for this study, and IRB committee approval. O.A. Ross: drafting/revising the manuscript; study concept and design; and interpretation of data.

Acknowledgment

The authors thank all those who have contributed to our research, particularly the patients and families who donated Brain and DNA samples for this work.

Study funding

K. Ogaki is supported by a research grant from the NAITO Foundation, JSPS KAKENHI Grant Number JP17K14966 and a grant from Institute for Environmental and Gender-Specific Medicine, Juntendo University. R.J. Uitti, O.A. Ross, Z.K. Wszolek, and D.W. Dickson are partially supported by the NIH/NINDS P50 NS072187. O.A. Ross is supported by the NINDS R01# NS078086, U54 NS100693, DOD W81XWH-17-1-0249, and in part by the Michael J. Fox Foundation for Parkinson's Research. C. Labbé is the recipient of a F.R.S.Q. fellowship. Z.K. Wszolek receives research support from NIH/NIA (primary), and NIH/NINDS (secondary) 1U01AG045390-01A1, Mayo Clinic Center for Regenerative Medicine, Mayo Clinic Neuroscience Focused Research Team (Cecilia and Dan Carmichael Family Foundation, and the James C. and Sarah K. Kennedy Fund for Neurodegenerative Disease Research at Mayo Clinic in Florida), and is also partially supported by a gift from Carl Edward Bolch, Jr, and Susan Bass Bolch, The Sol Goldman Charitable Trust, and Donald G. and Jodi P. Heeringa. P.A. Low receives research support from the Cure MSA Foundation. Samples included in this study were clinical patients or brain donors to the brain bank at Mayo Clinic in Jacksonville, which is supported by the CurePSP Society for Progressive Supranuclear Palsy and Udall Center for Excellence in Parkinson Research (P50 NS072187), APDA Center for Advanced Research, an Alzheimer's disease Research Center grant (NIA P50 AG16574), and NINDS R01 NS078086.

Disclosure

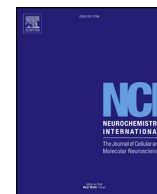
K. Ogaki receives JSPS KAKENHI Grant Number JP17K14966 and a grant from Institute for Environmental & Gender-Specific Medicine, Juntendo University. M.G. Heckman serves as the editorial board member of *Parkinsonism & Related Disorders*. S. Koga and Y.A. Martens report no disclosures. C. Labbé receives a Fonds de recherche du Québec-Santé postdoctoral fellowship. O. Lorenzo-Betancor, R.L. Walton, A.I. Soto, and E. Vargas report no disclosures. S. Fujioka receives JSPS KAKENHI Grant Number JP15K19501 and is an editorial board member of *Parkinsonism & Related Disorders*. R.J. Uitti serves as an associate editor of *Neurology*. J. Van Gerpen report no disclosures. W.P. Cheshire serves in the editorial boards of *Parkinsonism & Related Disorders* and *Autonomic Neuroscience* and as an associate editor of *Clinical Autonomic Research*. S.G. Younkin reports no disclosures. Z.K. Wszolek receives research support from P50 NS072187, NIH/NIA (primary), and NIH/NINDS (secondary) 1U01AG045390-01A1, Mayo Clinic Center for Regenerative Medicine, Mayo Clinic Neuroscience Focused Research Team (Cecilia and Dan Carmichael Family Foundation, and the James C. and Sarah K. Kennedy Fund for Neurodegenerative Disease Research at Mayo Clinic in Florida), the gift from Carl Edward Bolch, Jr., and Susan Bass Bolch, The Sol Goldman Charitable Trust, and Donald G. and Jodi P. Heeringa; serves as coeditor-in-chief of *Parkinsonism & Related Disorders* and an associate editor of the *European Journal of Neurology*; is on the editorial boards of *Neurologia i Neurochirurgia Polska*, *European Journal of Clinical and Experimental Medicine*, *Clinical and Experimental Medical Letters*, and *Wiadomosci Lekarskie*; holds and has contractual rights for receipt of future royalty payments from patents for “A Novel Polynucleotide Involved in Heritable Parkinson’s Disease”; and receives royalties from publishing *Parkinsonism & Related Disorders* (Elsevier, 2016, 2017) and the *European Journal of Neurology* (Wiley Blackwell, 2016, 2017). P.A. Low receives research support from the NIH (P01 NS44233, U54 NS065736, R01 NS092625, and UL1 TR000135), FDA (R01 FD004789), Cure MSA Foundation, and Mayo Funds and

serves as a clinical editor of *Autonomic Neuroscience*. W. Singer reports no disclosures. G. Bu received support from P50AG016574, RF1AG051504, R01AG027924, R01AG035355, R01AG046205, and P01NS074969 and a grant from the Cure Alzheimer’s fund. D.W. Dickson receives support from P50-AG016574, P50-NS072187, P01-AG003949, and CurePSP: Foundation for PSP|CBD and Related Disorders. D.W. Dickson is an editorial board member of *Acta Neuropathologica*, *Annals of Neurology*, *Brain*, *Brain Pathology*, and *Neuropathology* and is editor-in-chief of the *American Journal of Neurodegenerative Disease* and *International Journal of Clinical and Experimental Pathology*. O.A. Ross received support from R01 NS078086, P50-NS072187, and U54 NS100693 and the Michael J. Fox Foundation. O.A. Ross is an editorial board member of the *American Journal of Neurodegenerative Disease*, *Neurologia i Neurochirurgia Polska*, and *Molecular Neurodegeneration*. Full disclosure form information provided by the authors is available with the full text of this article at Neurology.org/NG.

Received January 23, 2018. Accepted in final form May 23, 2018.

References

1. Ahmed Z, Asi YT, Sailer A, et al. The neuropathology, pathophysiology and genetics of multiple system atrophy. *Neuropathol Appl Neurobiol* 2012;38:4–24.
2. Scholz SW, Bras J. Genetics underlying atypical parkinsonism and related neurodegenerative disorders. *Int J Mol Sci* 2015;16:24629–24655.
3. Sailer A, Scholz SW, Nalls MA, et al. A genome-wide association study in multiple system atrophy. *Neurology* 2016;87:1591–1598.
4. Jay TR, von Saucken VE, Landreth GE. TREM2 in neurodegenerative diseases. *Mol Neurodegener* 2017;12:56.
5. Koga S, Parks A, Uitti RJ, et al. Profile of cognitive impairment and underlying pathology in multiple system atrophy. *Mov Disord* 2017;32:405–413.
6. Gilman S, Wenning GK, Low PA, et al. Second consensus statement on the diagnosis of multiple system atrophy. *Neurology* 2008;71:670–676.
7. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–291.
8. Chen Y, Chen X, Guo X, et al. Assessment of TREM2 rs75932628 association with Parkinson’s disease and multiple system atrophy in a Chinese population. *Neurol Sci* 2015;36:1903–1906.
9. Simons M, Lyons DA. Axonal selection and myelin sheath generation in the central nervous system. *Curr Opin Cell Biol* 2013;25:512–519.
10. Painter MM, Atagi Y, Liu CC, et al. TREM2 in CNS homeostasis and neurodegenerative disease. *Mol Neurodegener* 2015;10:43.
11. Poliani PL, Wang Y, Fontana E, et al. TREM2 sustains microglial expansion during aging and response to demyelination. *J Clin Invest* 2015;125:2161–2170.
12. Koga S, Aoki N, Uitti RJ, et al. When DLB, PD, and PSP masquerade as MSA: an autopsy study of 134 patients. *Neurology* 2015;85:404–412.



The unresolved role of mitochondrial DNA in Parkinson's disease: An overview of published studies, their limitations, and future prospects

Amica C. Müller-Nedebock^a, Rebecca R. Brennan^b, Marianne Venter^c, Ilse S. Pienaar^{d,e}, Francois H. van der Westhuizen^c, Joanna L. Elson^{c,f}, Owen A. Ross^{b,g,h}, Soraya Bardien^{a,*}

^a Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa

^b Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA

^c Human Metabolomics, North-West University, Potchefstroom, South Africa

^d School of Life Sciences, University of Sussex, Falmer, BN1 9PH, United Kingdom

^e Centre for Neuroinflammation and Neurodegeneration, Imperial College London, London, United Kingdom

^f Institute of Genetic Medicine, Newcastle University, Newcastle Upon Tyne, United Kingdom

^g Department of Clinical Genomics, Mayo Clinic, Jacksonville, FL, USA

^h School of Medicine and Medical Science, University College Dublin, Dublin, Ireland

ARTICLE INFO

Keywords:

Parkinson's disease
Mitochondrial DNA
Mitochondrial haplogroups
Somatic mtDNA variation
Homoplasmic mtDNA variation
mtDNA depletion

ABSTRACT

Parkinson's disease (PD), a progressive neurodegenerative disorder, has long been associated with mitochondrial dysfunction in both sporadic and familial forms of the disease. Mitochondria are crucial for maintaining cellular homeostasis, and their dysfunction is detrimental to dopaminergic neurons. These neurons are highly dependent on mitochondrial adenosine triphosphate (ATP) and degenerate in PD. Mitochondria contain their own genomes (mtDNA). The role of mtDNA has been investigated in PD on the premise that it encodes vital components of the ATP-generating oxidative phosphorylation (OXPHOS) complexes and accumulates somatic variation with age. However, the association between mtDNA variation and PD remains controversial. Herein, we provide an overview of previously published studies on the role of inherited as well as somatic (acquired) mtDNA changes in PD including point mutations, deletions and depletion. We outline limitations of previous investigations and the difficulties associated with studying mtDNA, which have left its role unresolved in the context of PD. Lastly, we highlight the potential for further research in this field and provide suggestions for future studies. Overall, the mitochondrial genome is indispensable for proper cellular function and its contribution to PD requires further, more extensive investigation.

1. The mitochondrial genome

Mitochondria are primarily responsible for generating adenosine triphosphate (ATP) via oxidative phosphorylation (OXPHOS). Besides nuclei, they are the only other cellular organelles harbouring their own genome (mtDNA). mtDNA is a compact (16 569 bp), circular, double-stranded genome, comprised of 37 genes including 13 which encode essential subunits of the OXPHOS enzymes (Fig. 1) (Andrews et al., 1999). More recently mtDNA has been reported to additionally encode two short open reading frames (sORFs) within the ribosomal RNA genes; *MT-RNR1* and *MT-RNR2* (Lee et al., 2015). These sORFs can be translated into mitochondrial-derived peptides (MDPs) which have important functions including regulating nuclear gene expression in response to metabolic stress (Kim et al., 2018).

Mitochondrial genomes differ to their nuclear counterparts with regards to their replication, repair and inheritance mechanisms. Unlike nuclear DNA (nDNA), mtDNA is primarily maternally inherited, although rare cases of biparental inheritance have recently been reported (Luo et al., 2018). Moreover, it exists as multiple copies (known as polyploidy) within individual mitochondria. Hundreds to thousands of mtDNA copies can exist per mitochondria and thus per cell, depending on the energetic demand of the cell (Miller et al., 2003). Compared to nDNA, mtDNA is highly susceptible to accumulating errors due to the lack of protective histones and its close proximity to damaging by-products of OXPHOS e.g. reactive oxygen species (ROS) (Khrapko et al., 1997). In addition, reduced fidelity of the mtDNA polymerase, mtDNA polymerase gamma (POLG1), has been suggested to contribute to a higher mutational rate of mtDNA (Song et al., 2005). As a result, most

* Corresponding author. Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Stellenbosch University, PO Box 241, Cape Town, 8000, South Africa.

E-mail address: sbardien@sun.ac.za (S. Bardien).

<https://doi.org/10.1016/j.neuint.2019.104495>

Received 16 April 2019; Received in revised form 27 May 2019; Accepted 21 June 2019

Available online 21 June 2019

0197-0186/ © 2019 Elsevier Ltd. All rights reserved.

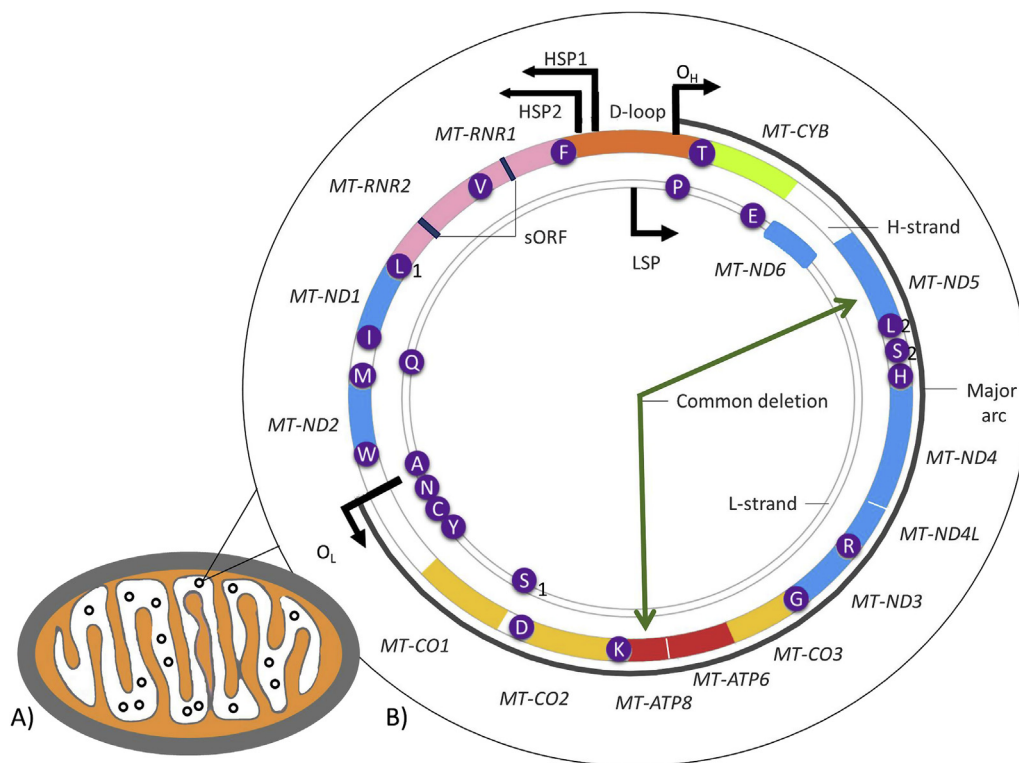


Fig. 1. The mitochondrion and structure of human mitochondrial DNA A) Basic structure of a mitochondrion containing multiple copies of mitochondrial DNA (mtDNA) represented as black circles. B) Structure of mtDNA: a 16 569 bp circular, double-stranded molecule consisting of a guanine-rich heavy strand (H-strand) and a light strand (L-strand), rich in cytosine. Approximately 93% of the genome is coding, having only one significant non-coding control region which includes the displacement loop (D-loop). This stretch of DNA contains the origin of H-strand replication (OH), the H-strand transcriptional promoters (HSP1 and HSP2) as well as the L-strand promoter (LSP). The DNA contains 37 primarily contiguous genes, nine on the L-strand and 28 on the H-strand. Twenty-four genes encode mature RNA products: one 12 S rRNA (small ribosomal subunit) and one 16 S rRNA (large ribosomal subunit), and 22 mitochondrial tRNAs (genes highlighted in purple). The remaining 13 mtDNA genes encode polypeptide components of the electron transport chain (ETC) involved in energy production via oxidative phosphorylation (OXPHOS). These include

seven subunits (MT-ND1,2,3,4L, 5,6) of complex I (genes highlighted in blue), one subunit (MT-CYB) of complex III (gene highlighted in light green), three subunits (MT-CO1,2, and 3) of complex IV (genes highlighted in yellow), and two subunits (MT-ATP6 and 8) of complex V (genes highlighted in red). mtDNA also encodes two mitochondrial-derived peptides, MOTS-c and Humanin, as short open reading frames (sORF) in the MT-RNR1 and MT-RNR2 genes (highlighted in pink) respectively. Deletions in mtDNA frequently occur in the major arc (indicated by the dark grey line) between the heavy and light strand origins of replication (OH and OL respectively). Among these deletions is the 'common deletion': a 4977-base-pair deletion located between the MT-ATP8 and MT-ND5 genes as indicated by the dark green arrows.

individuals harbour low levels (< 1%) of inherited and/or acquired mtDNA variants (Ye et al., 2014a). Most copies of mtDNA in a cell are identical, which is referred to as *homoplasmy*. A mix of both mutated and wild-type mtDNA is referred to as *heteroplasmy*. Levels of heteroplasmy can change over time through mechanisms of relaxed replication in postmitotic tissue and random segregation of mtDNA during cell division in mitotic tissues (Chinnery and Samuels, 1999). Although, the process of mammalian mtDNA replication is not yet completely understood, mtDNA is known to replicate independently of the cell cycle - undergoing life-long replication in both proliferating and post-mitotic cells (Yasukawa and Kang, 2018). Tissue-specific differences in mtDNA maintenance, replication, and expression exist (Herbers et al., 2019) which, together with the unique features of mtDNA, particularly polyploidy and heteroplasmy, pose a significant challenge for studying their role in diseases including Parkinson's disease (PD).

PD is a neurodegenerative movement disorder with a complex aetiology comprising both environmental and genetic factors. Progressive dopaminergic (DA) neuronal loss in the Substantia Nigra pars compacta (SNpc) is characteristic of PD, but pathological mechanisms are poorly understood. Despite decades of research and definite links to mitochondrial dysfunction in disease susceptibility and progression, the role of mtDNA in risk and pathogenesis of sporadic PD (sPD) remains equivocal. This review aims to appraise the literature to determine the reason(s) for this. We discuss results from published studies which have investigated mtDNA changes in PD and highlight inconsistencies between studies which have left the role of mtDNA in PD largely unresolved. We also draw attention to factors which complicate the study of mtDNA in PD and may have contributed to conflicting results. In doing so, we provide suggestions to guide and possibly improve future research in this field.

2. Parkinson's disease and mitochondrial dysfunction

The DA neuronal loss in PD results in an array of classical motor symptoms including bradykinesia, tremor and rigidity (Gelb et al., 1999). These, together with various non-motor symptoms (e.g. depression, anxiety and insomnia), arising from the loss of additional neuronal populations, significantly compromise patients' quality of life. A pathological hallmark of PD are Lewy bodies (LB); intracellular inclusions of the aggregate-prone, alpha-synuclein protein (Spillantini et al., 1997). Only approximately 10% of cases have familial PD (Elbaz et al., 1999) and around 30% of these have monogenic PD i.e. have mutations in nDNA-encoded genes (Kumar et al., 2011). Notably, most PD cases are sporadic and have an unknown aetiology of disease as well as an absence of a family history. However, both familial and sPD have been linked to mitochondrial dysfunction (Liu et al., 2009; Parker et al., 1989).

Although the generation of ATP is the most recognised function of mitochondria, the organelles also participate in critical processes which help ensure cell viability, including lipid biosynthesis, calcium homeostasis and apoptosis and even play diverse roles in immune response (Bulua et al., 2011; West et al., 2011). Mitochondrial dysfunction therefore impinges on a wide spectrum of cellular functions. This dysfunction is multifactorial in origin and can arise from impaired mitochondrial biogenesis, altered mitochondrial dynamics (e.g. fission and fusion), impaired mitochondrial quality control (mitophagy), compromised OXPHOS and calcium imbalances (Park et al., 2018).

Early evidence that mitochondrial dysfunction may play a central role in sPD pathogenesis resulted from the unintentional exposure of humans to the drug 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridin (MPTP), specifically its active derivative MPP⁺, causing parkinsonian symptoms (Langston et al., 1983; Davis et al., 1979). Primate models

subsequently reported SNpc DA cell loss and other parkinsonism symptoms including akinesia, tremor and rigidity in monkeys treated with a form of MPTP (Burns et al., 1983). Similarly, MPP + mouse models exhibited DA neuronal loss (Heikkilä et al., 1984; Wallace et al., 1984) and showed that MPP + inhibits complex I activity, thereby interfering with OXPHOS (Heikkilä et al., 1985; Mizuno et al., 1987). As such, an energy crisis was suggested as an important mechanism underlying neuronal degeneration leading to PD symptoms (Heikkilä et al., 1985; Mizuno et al., 1987). Subsequently, in the late 1980's and early 1990's, defects of complex I were observed in various tissues of SPD patients, including skeletal muscle (Shoffner et al., 1991; Bindoff et al., 1991; Blin et al., 1994; Nakagawa-Hattori et al., 1992; Cardellach et al., 1993), the SNpc (Janetzky et al., 1994; Schapira et al., 1989) and platelets (Haas et al., 1995; Benecke et al., 1993; Yoshino et al., 1992; Krige et al., 1992). As a consequence of these initial studies, it was hypothesised that the observed complex I defect may be derived via the mitochondrial genome (Parker et al., 1989). Moreover, genes linked to early-onset forms of PD (e.g. *PINK1*, *PRKN* and *DJ-1*), encode proteins which participate in- or mediate some form of mitochondrial function, regulation or quality control (Park et al., 2018). For instance, as reviewed elsewhere (Trinh and Farrer, 2013), the proteins *PINK1* and *Parkin* are involved in mitophagy, consequently, loss of function mutations in nDNA encoding these proteins affect mitochondrial biogenesis and induction of autophagy.

3. Early evidence for mtDNA involvement in PD from cybrids

In addition to nDNA mutations contributing to familial PD onset, mtDNA variation has been linked to SPD, as demonstrated through cybrid (cytoplasmic hybrid) cell line studies (Table 1) (Swerdlow, 2012). Briefly, p0 cells which are devoid of mtDNA but contain identical nDNA, are fused with mtDNA-containing platelets from either PD patients or controls. Following multiple cellular replication cycles, the cybrids generated share the same environmental and nDNA background, differing only with regards to their mitochondrial genomes (Ghosh et al., 1999; Swerdlow et al., 1996). Therefore, differences observed between cell lines with differing mtDNA content, from either a patient or control, are thought to be solely attributable to the mtDNA itself (Ghosh et al., 1999; Swerdlow et al., 1996; Esteves et al., 2008). As a result, p0 cybrid cells have provided a useful tool to investigate the role of mtDNA in cellular health and in disease.

Albeit an effective model to study mtDNA background *in vitro*, cybrid studies have limitations which contribute to discrepancies between studies. One of the most notable is the frequent use of tumour cell lines exhibiting aneuploidy or other forms of nuclear genomic instability (Wilkins et al., 2014). Despite potential limitations, SPD cybrid studies

have consistently provided insight into the role which mtDNA background has on mitochondrial morphology and function between cell lines from patients and controls (Table 1).

Notably, emerging evidence implicates mitochondrial genomes in modulating nuclear gene expression by means of signalling pathways (Kenney et al., 2014; Vivian et al., 2017; Jandova et al., 2012). These include the integrated stress response, retrograde (mitochondria-nucleus) communication, proteostasis regulation and mitonuclear feed-back signalling (Quiros et al., 2016). With mitochondrial biogenesis, maintenance and other mitochondrial processes heavily depending on nDNA gene expression (Dominy and Puigserver, 1500), mtDNA may influence these processes and could for instance account for variable biogenesis between SPD cybrids with differing mtDNA donors (Keeney et al., 2009).

Importantly, SPD cybrids have exhibited several features of mitochondrial dysfunction including decreased complex I activity, increased ROS, morphologically abnormal mitochondria and decreased maximum respiratory capacity, linking mtDNA to SPD (Table 1). These findings all highlight that mtDNA can substantially influence important mitochondrial processes and physiology. Consequently, cybrid studies spawned investigations aimed at identifying mtDNA variation which could account for mitochondrial abnormalities observed in SPD cybrids.

4. Studies investigating homoplasmic mtDNA variation

4.1. Early mtDNA sequencing studies

In the 1990's and early 2000's, a few studies sequenced the complex I genes of PD patients and controls to identify possible mtDNA variants associated with the disorder (Smigrodzki et al., 2004; Kösel et al., 1998; Richter et al., 2002; Simon et al., 2000; Kirchner et al., 2000). Standard Sanger sequencing techniques available during this time were primarily limited to identifying homoplasmic or high-frequency heteroplasmic variants (Howell et al., 2005). A few studies sequenced whole mtDNA (Ikebe et al., 1995; Ozawa et al., 1991), whilst others sequenced all of the mtDNA tRNA (Simon et al., 2000; Grasbon-Frodl et al., 1999; Brown et al., 1996) and rRNA genes (Brown et al., 1996). Furthermore, some employed restriction fragment length polymorphism (RFLP) methods to investigate whether certain mtDNA variants (e.g. m.3397A > G in *MT-ND1*) associate with PD (Bandmann et al., 1997; Mayr-Wohlfart et al., 1997).

A handful of studies identified 'novel' variants in PD patients (Kösel et al., 1998; Richter et al., 2002; Grasbon-Frodl et al., 1999; Brown et al., 1996) and reported variants at an increased frequency in cases compared to controls (Kirchner et al., 2000; Mayr-Wohlfart et al., 1997). However, their findings may have been coincidental. This is

Table 1
Features of mitochondrial dysfunction observed in sporadic Parkinson's disease cybrids.

Phenotype	References
Altered mitochondrial ultrastructure: e.g. rounded, swollen mitochondria; mitochondrial matrix with few or disrupted cristae; presence of intramitochondrial inclusions	(Esteves et al., 2008, 2009; Swerdlow et al., 1998; Trimmer et al., 2000)
Decreased complex I activity	(Swerdlow et al., 1996; Esteves et al., 2008, 2009, 2010a; Gu et al., 1998)
Mitochondrial depolarisation	(Esteves et al., 2008)
Reduced mitochondrial ability to buffer cytosolic calcium/altered calcium homeostasis	(Esteves et al., 2010b; Sheehan et al., 1997)
Reduced ATP levels	(Esteves et al., 2008, 2009, 2010c)
Apoptosis: altered levels of apoptosis-related proteins	(Esteves et al., 2008; Onyango et al., 2005a, 2005b)
Reduced mtDNA content	(Keeney et al., 2009; Borland et al., 2009)
Microtubule depolymerisation	(Esteves et al., 2009, 2010c)
Increased alpha-synuclein oligomerization	(Esteves et al., 2009, 2010b, 2010c)
Existence of fibrillar and vesicular inclusions (cybrid Lewy bodies)	(Esteves et al., 2009; Trimmer et al., 2004)
Reduced respiratory reserve capacity	(Esteves et al., 2010a)
Increased toxin susceptibility	(Swerdlow et al., 1996; Esteves et al., 2008)
Significantly reduced mitochondrial axonal transport	(Borland et al., 2009)
Increased ROS production	(Swerdlow et al., 1996; Esteves et al., 2009)

ATP = Adenosine triphosphate; mtDNA = Mitochondrial DNA; ROS = Reactive oxygen species

because the sample sizes were often too small to draw statistically powerful conclusions. Additionally, sequencing was still in its infancy, making it more likely to identify false positives due to the limited number of known variants which had been reported. In support of this statement, most of these ‘novel’ variants have now been reported in individuals without PD on the MITOMAP database (Lott et al., 2013; <https://www.mitomap.org/MITOMAP>), although not at high frequencies (Supplementary Table 1). Based on these early sequencing studies, no homoplasmic nor high-frequency heteroplasmic mtDNA variants were conclusively implicated in PD pathogenesis or risk.

4.2. Haplogroup association studies

The sequential accumulation of new variation in maternal lineages over thousands of years has generated stable mtDNA sequence variation which is homoplasmic. Subsets of such stable mtDNA variation define haplogroups which are primarily restricted to particular populations and geographic areas (Torroni et al., 2006). Details of variants which define individual mtDNA haplogroups are recorded on PhyloTree (Van Oven and Kayser, 2009); <http://www.phylotree.org/>.

Over the past two decades, the associations between mtDNA haplogroups and PD risk have been extensively studied using the ‘haplogroup association approach’, particularly in populations of European ancestry (Table 2 and Supplementary Table 2). This approach associates haplogroups with a disease phenotype and suggests that one or more common population variants may modify disease risk. As a result of these studies, multiple European haplogroups including J, K, U, and some super-haplogroups (e.g. UK and JT) have been associated with a reduced risk of PD (Gaweda-Walerych et al., 2008; Georgiou et al., 2017; Ghezzi et al., 2005; Hudson et al., 2013; Khusnutdinova et al., 2008; Mehta et al., 2009; Pyle et al., 2005; Van Der Walt et al., 2003). However, a number of additional studies were unable to replicate these findings and found no significant association between PD and haplogroups or common haplogroup variants (Mehta et al., 2009; Fachal et al., 2015; Huerta et al., 2005; Latsoudis et al., 2008; Ross et al., 2003; Autere et al., 2004) (Table 2 and Supplementary Table 2). For instance, Mehta et al. (Mehta et al., 2009) were unable to replicate previous associations between PD risk and haplogroups J and K in a large Australian cohort of European ancestry. Additionally, Van Der Walt et al. (Van Der Walt et al., 2003) and Huerta et al. (Huerta et al., 2005) identified the *MT-ND3* single nucleotide polymorphism (SNP) m.10398A > G to be protective against PD, whilst multiple other studies could not replicate this finding (Gaweda-Walerych et al., 2008; Ghezzi et al., 2005; Hudson et al., 2013; Latsoudis et al., 2008). Asian haplogroup association studies have also produced somewhat inconsistent results (Table 2 and Supplementary Table 2). For instance, although Wu et al. (Wu et al., 2018) and Liou et al. (Liou et al., 2016) both reported the Asian haplogroup B5 to be protective against PD, studies done by Chen et al. (Chen et al., 2007) and Chen et al. (Chen et al., 2015) reported no significant overall association between any of the common Asian haplogroups and PD risk.

Overall, the evident variability in results and lack of reproducibility across haplogroup studies has made it difficult to conclude whether common mtDNA population variants contribute to PD risk.

4.2.1. Limitations of past haplogroup association studies and possible sources of inter-study variability

To improve the design of future studies, it is important to identify all possible factors which are introducing variability into, and between studies in order to limit false positive or erroneous findings. In the following section we therefore highlight possible methodological concerns which may contribute to discrepancies between study findings and should be addressed in future work.

4.2.1.1. Population stratification. mtDNA haplogroup association studies are often confounded by high levels of population

stratification, referring to systematic differences in allele frequencies between subpopulations in a population, due to different ancestry (Hellwege et al., 2017). This is because mtDNA itself is more prone to population substructure than nDNA, given that its effective population size is four times smaller than that of nDNA due to its uni-parental inheritance. Even when cases and controls are well matched in other regards (e.g. age, sex) substructure can be a confounding factor which may result in false positives. A study done by Otaegui et al. (Otaegui et al., 2004) with a cohort of Spanish PD patients, from different origins (Basque and non-Basque), and non-Basque controls, highlighted the importance of conducting case-control association studies in ethnically homogeneous populations (Table 2 and Supplementary Table 2). The authors found no association between PD and the m.10398A > G SNP ($P = 0.088$) when comparing all cases to controls but observed a significant overrepresentation of this SNP specifically in PD patients of Basque origin when compared to the non-Basque control population ($P = 0.0221$).

4.2.1.2. Replication cohorts and multiple testing. Several case-control association studies, including some of the ‘PD haplogroup’ studies listed in Table 2, lack replication cohorts (Salas and Elson, 2015). Such replication cohorts or even a second control group (e.g. as used by Ghezzi et al. (Ghezzi et al., 2005), Pyle et al. (Pyle et al., 2005) and Chen et al. (Chen et al., 2015)) can be useful for detecting population substructure which affects the validity of results if not taken into consideration (Salas et al., 2009).

Additionally, multiple PD haplogroup association studies failed to correct their P-values/significance levels for multiple testing (Table 2 and Supplementary Table 2). This may also have resulted in false positive findings (Pardo-Seco et al., 2013).

4.2.1.3. Power. Samuels et al. (Samuels et al., 2006) demonstrated that when using the haplogroup association hypothesis, very large sample sizes of cases and controls are necessary in order to reliably detect an association between haplogroups and complex human disease. To demonstrate this, the authors gave the example that studies with 500 cases and 500 controls have 90% power to detect a greater than 35% change in the frequency of the common haplogroup H. Therefore, several of the PD haplogroup studies are underpowered (Table 2 and Supplementary Table 2). Additionally, such underpowered haplogroup association studies are frequently analysed in ways which violate the assumptions of the statistical tests used, thereby inflating Type 1 error (Salas and Elson, 2015).

4.2.1.4. Pseudo-haplogroups. Pseudo-haplogroups are usually composed of low frequency haplogroups present in the study cohort, clustered together in ways which cannot be justified by phylogenetic knowledge (Fachal et al., 2015). Interestingly, this artificial clustering of haplogroups is commonly observed in neurological association studies including those done by Pyle et al. (Pyle et al., 2005), Ghezzi et al. (Ghezzi et al., 2005), Gaweda-Walerych et al. (Gaweda-Walerych et al., 2008) and Latsoudis et al. (Latsoudis et al., 2008) on PD risk. The creation of such pseudo-haplogroups results in erroneous inferences and false positive findings (Fachal et al., 2015). This is because pseudo-haplogroups are not biologically meaningful; they do not share a set of variants characterising an exclusive phylogenetic branch. To demonstrate this, Fachal et al. (Fachal et al., 2015) artificially merged haplogroups U and V of a migraine cohort from Cantabria. This resulted in the UV pseudo-haplogroup being significantly associated with migraines ($P = 0.0268$) even though it did not make phylogenetic sense. Thus, emphasising the risk of false associations when pseudo-haplogroups are used.

In summary, PD haplogroup studies have reported inconsistent findings. Evidently, there are many methodological concerns around haplogroup studies which may have influenced the findings of past studies. These concerns should carefully be addressed in future.

Table 2
Summary of mtDNA haplogroup association studies in Parkinson's disease

Ethnicity (ancestry)	Study participants	N	Multiple-test correction	Estimates of statistical power determined	Outcome/ Main findings	Reference
American (European)	Cases Controls	609 340	NS	NS	Haplogroups J (P = 0,02) and K (P = 0,02) associated with reduced PD risk. SNP m.10398A > G was reported to be protective against PD (P = 0,0001). SNP m.9055G > A in <i>MT-ATP6</i> reduced PD risk in women (P = 0,03). SNP m.13708G > A reduced PD risk in individuals > 70 years of age (P = 0,010).	Van Der Walt et al. (2003)
Irish (European)	Cases Controls	90 129	Yes (Bonferroni correction)	NS	No significant association between haplogroup J and PD found. SNP m.4216T > C, in linkage with mtDNA TJ cluster, was associated with increased PD risk (P = 0,014).	Ross et al.(2003)
Finnish (European)	Cases (PD) Cases (PDD) Controls	210 28 104	NS	NS	Supercluster JTWIX was associated with an increased risk of PD (P = 0,27) and PD with dementia (PDD) (P = 0,18).	Autere et al.(2004)
Spanish from Asturias, Northern Spain (European)	Cases Controls	271 230	Yes (Bonferroni correction)	NS	No Haplogroups were associated with PD risk. SNP m.4336T > C associated with increased PD risk only in females (P = 0,011). SNP m.10398A > G was protective against PD (P = 0,009).	Huerta et al.(2005)
English (European)	Cases Control group 1 Control group 2 (Birth cohort) Control group 3 (Post-mortem brain tissue from AD patients)	455 269 178 185	NS	NS	Haplogroups J and K not significantly associated with reduced PD risk. UKJT cluster reduced PD risk (P < 0,0001).	Pyle et al. (2005)
Italian (European)	Cases Control group 1 (CT1) Control group 2 (CT2)	620 1486 509	NS	NS	Haplogroup K (P = 0,048) but not J associated with reduced PD risk. SNP m.10398A > G did not significantly alter PD risk.	Ghezzi et al.(2005)
Taiwanese (Asian)	Cases Controls	416 372	Yes (Bonferroni correction)	NS	No haplogroups associated with PD risk.	Chen et al.(2007)
Polish from central, South and North-Western Poland (European)	Cases Controls Cases for gender stratified haplogroup J analysis Controls Additional cases for haplogroup K analysis Additional controls for haplogroup K analysis	241 277 304 316 91 137	NS	Yes	Haplogroup J (P = 0,0014) associated with reduced PD risk in males (after stratification by gender). Sub-lineages U4 + U5a1 + K + J1c + J2 also reduced PD risk (P = 0,027). SNP m.10398A > G did not significantly alter PD risk.	Gawęda-Walerych et al. (2008)
Greek from Crete (European)	Cases Controls	224 383	NS	NS	No haplogroups associated with PD risk. SNP m.10398A > G did not significantly alter PD risk.	Latsoudis et al.(2008)
Russian, Tatar (European)	Cases Controls	157 183	NS	NS	Haplogroup H associated with an increased PD risk (P = 0,0001). UK cluster associated with a decreased PD risk (P = 0,003).	Khusnutdinova et al. (2008)
Australian from New South Wales and Queensland (European)	Cases Controls	890 3491	NS	Yes	No significant associations between PD risk and haplogroups J and K found, nor the pooled UJKT haplogroup cluster.	Mehta et al. (2009)
English (European)	Cases (Discovery phase) Controls (Discovery phase) Cases (Replication phase) Controls (Replication phase) Cases (Meta-Analysis) Controls (Meta-Analysis)	1719 2889 851 2717 6140 13280	NS	Yes	Association study: No association between haplogroups and PD nor SNP m.10398A > G and PD. Super-haplogroup JT was associated with a protective effect against PD (P = 0,0354). Mitochondrial variants m.2158T > C (discovery: P = 0,024; replication: P = 0,0245) and m.11251A > G (discovery: P = 0,0292; replication: P = 0,0012) were associated with a reduced risk of PD in both the discovery and replication cohorts. Meta-analysis: Haplogroups J (P = 0,0122), K (P = 0,00363), T (P = 0,0245) and super-haplogroup JT (P = 0,000584) associated with reduced PD risk. Increased PD risk was associated with cluster HV (P = 0,00364).	Hudson et al. (2013)
Spanish from Pamplona, North-East Spain (European)	Cases Controls	478 394	Yes (NS)	NS	No haplogroups were associated with PD risk	Fachal et al. (2015)
Spanish from Santiago de	Cases Controls	305 293				

(continued on next page)

Table 2 (continued)

Ethnicity (ancestry)	Study participants	N	Multiple-test correction	Estimates of statistical power determined	Outcome/ Main findings	Reference
Compostela, North-West Spain (European)	Cases (Total)	279	NS	NS	Overall, no association between haplogroups and PD. Haplogroup B (P = 0,004) associated with a lower risk for EOPD in individuals younger than 50 years (after age stratification). Haplogroup D associated with a higher risk of PD (P = 0,033) and Haplogroup B was associated with a lower risk of PD (P = 0,018) in individuals younger than 50.	Chen et al.(2015)
Han Chinese from Southern China (Asian)	Cases EOPD (< 50 years of age)	63				
	Cases LOPD (> 50 years of age)	216				
	Controls (Total)	510				
	Control team 1	118				
	Control team 2	392				
Han Chinese from Northern China (Asian)	Cases	322	Yes (Bonferroni correction)	NS	SNP m.10398A > G (P = 0,001) significantly associated with increased PD risk in females (P = 0,0036).	Chu et al.(2015)
Taiwanese (Asian)	Controls	332	Yes (Bonferroni correction)	NS	Haplogroup B5 associated with a reduced PD risk (P = 0,002).	Liou et al. (2016)
Greek from Cyprus (European)	Cases	725	Yes (Bonferroni correction)	NS	Haplogroup U associated with reduced PD risk (P = 0,03), supercluster LMN (P = 0,01) and cluster N(xR) (P = 0,006) were significantly protective against PD in females.	Georgiou et al.(2017)
	Controls	744	Yes (Bonferroni correction)	NS	D-loop-sequencing: SNPs m.151T > C (P = 0,023), m.189G > A (P = 0,03), m.16086C > T (P = 0,007 and m.16271C > T (P = 0,0497) were associated with increased PD risk. SNPs m.318C > T and m.16134T > C (P = 0,022) were associated with decreased PD risk.	Wu et al. (2018)
Han Chinese from East China (Asian)	Cases	500	NS	NS	Haplogroup A5 (P = 0,039) was associated with increased PD risk. Haplogroup B5 (P = 0,068) was associated with a reduced disease risk.	
	Controls	505			Meta-analysis: Haplogroup B5, but not B4, was protective against PD (P = 0,0003). Haplogroup G not associated with PD (P = 0,09).	

AD = Alzheimer's disease; EOPD = Early onset Parkinson's disease; LOPD = Late onset Parkinson's disease; mtDNA = Mitochondrial DNA; N = sample size; NA = Not applicable; NS = Not specified; PDD = Parkinson's disease with dementia; PD = Parkinson's disease

5. Somatic mtDNA changes in PD

Somatic mtDNA variation, including point mutations and deletions, accumulate over time in post-mitotic tissue, including the ageing human brain (Lin et al., 2002; Bender et al., 2006; Arnheim and Cortopassi, 1992). This gives rise to a mixed population of mutant and wild-type mtDNA molecules i.e. heteroplasmy. The acquired variants may additionally clonally expand in individual cells through relaxed replication and random intracellular drift (Elson et al., 2001). During relaxed replication, mtDNA molecules are randomly selected for replication. Thus, the proportion of mtDNA heteroplasmy in a cell can significantly increase over time if mutant mtDNA molecules are replicated more frequently by chance than the wild-type mitochondrial genomes (Elson et al., 2001). When the proportion of mutant mtDNA molecules with harmful mtDNA changes exceeds a critical threshold level, a cellular defect in OXPHOS will arise (Chinnery and Hudson, 2013). Consequently, a number of studies have investigated high levels of heteroplasmic mtDNA point mutations and mtDNA deletions in the post-mortem brain tissue of PD patients. Both deletions and point mutations in these studies were considered somatic rather than germline.

5.1. Somatic mtDNA point mutations and deletions

With the advent of sequencing technologies sensitive enough to detect low-level heteroplasmies (< 10%), a handful of studies sequenced mtDNA in PD cases and controls extracted from post-mortem brains (Table 3 and Supplementary Table 3). They hypothesised that multiple, individually rare mtDNA point mutations in either the entire mitochondrial genome, or subset of mtDNA genes, could collectively constitute a high variant burden in the brains of PD patients. This burden would ultimately lead to neuronal loss. However, only two studies reported significantly higher burdens of point mutations in the brains of PD cases compared to controls (Coxhead et al., 2016; Lin et al., 2012). Others reported no significant difference between case and control groups (Table 3 and Supplementary Table 3) (Smigrodzki et al., 2004; Dölle et al., 2016; Simon et al., 2004; Wei et al., 2017).

In addition, acquired mtDNA deletions have also been suggested to play an important part in the selective neuronal loss in the ageing and PD brain (Bender et al., 2006; Kraytsberg et al., 2006; Reeve et al., 2008; Nido et al., 2018). Mechanisms, such as breakpoints in tandem repeats of mtDNA or replication errors (Nissanka et al., 2019) give rise to these deletions (Reeve et al., 2008; Nido et al., 2018). Notably, DA metabolism has been shown to drive the generation of mtDNA deletions in *in vitro* and *in vivo* models (Neuhaus et al., 2013). DA is readily oxidised, leading to ROS and neurotoxic quinone production (Halliwell and Gutteridge, 1984). ROS, such as H₂O₂, have been shown to cause single and double stranded breaks in mtDNA *in vitro* (Shokolenko et al., 2009). As such, SNpc DA neurons appear to be particularly susceptible to accumulation of mtDNA deletions and most of these deletions are located between the heavy and light strand origins of replication (the major arc; Fig. 1) (Bender et al., 2008).

The 'common deletion' is a 4977-base-pair deletion located between the *MT-ATP8* and *MT-ND5* genes, spanning four genes coding for complex I (*MT-ND3*, *MT-ND4*, *MT-ND4L* and *MT-ND5*) (Fig. 1). It is thought to bring about a complex I defect, thus an energy crisis (Ikebe et al., 1990). The frequency of mtDNA carrying this deletion in PD cases compared to controls has been investigated using Southern blotting, competitive- and kinetic-PCR techniques as well as *in situ* hybridisation techniques (Ikebe et al., 1990; Kösel et al., 1997; Mann et al., 1992; Zhang et al., 2002; Ozawa et al., 1990). Despite two studies reporting a greater burden of this deletion in patients compared to controls (Ikebe et al., 1990; Ozawa et al., 1990), others could not replicate these findings (Kösel et al., 1997; Mann et al., 1992; Zhang et al., 2002). Instead, they reported no difference between cases and controls, suggesting that the common deletion may be a result of the natural ageing process rather than disease pathogenesis. Gu et al. (Gu et al., 2002)

later reported that the number of mtDNA deletions (not limited to the common deletion) was increased significantly in the SN homogenate of PD patients compared to age-matched controls, patients with multiple system atrophy, Dementia with Lewy Bodies, and those with Alzheimer's disease.

Additional studies compared the total burden of mtDNA carrying deletions in the major arc, between brains of PD cases and controls using quantitative PCR (qPCR) (Bender et al., 2006; Dölle et al., 2016; Grünwald et al., 2016; Bury et al., 2017). Such qPCR methods typically measure the ratio of a gene not commonly deleted (e.g. *MT-ND1*) in mtDNA to one which is frequently deleted (e.g. *MT-ND4*) using either relative or absolute quantification methods. The latter requiring a standard curve. Two studies reported significantly higher burdens of mtDNA deletions in individual DA SNpc ($P = 0.004$) (Dölle et al., 2016) and pedunculopontine cholinergic neurons ($P < 1 \times 10^{-4}$) (Bury et al., 2017) of PD cases compared to controls. The others reported no significant differences in deletion levels between SN neurons of cases and aged control groups (Bender et al., 2006; Grünwald et al., 2016).

Based on the evidence outlined above, it is reasonable to argue for or against a role of somatic point mutations and/or deletions in PD pathogenesis. However, variability between study designs (but also small sample sizes) makes it difficult to compare findings across studies and makes it challenging to draw a reliable conclusion regarding the contribution of somatic mtDNA deletions and point mutations to PD.

5.2. Changes in mtDNA copy number

In addition to possible point mutations and deletions in mtDNA molecules, mtDNA depletion should also be considered as a possible predisposing factor for PD. The importance of adequate mtDNA in cells is highlighted by mtDNA depletion syndromes – a group of autosomal recessive disorders characterised by severe mtDNA depletion which results in impaired energy production (Suomalainen and Isohanni, 2010). Since mtDNA encodes essential OXPHOS complex components, a sufficient amount of mtDNA is required for the production of these OXPHOS subunits and to manage cellular energy demands (El-Hattab and Scaglia, 2013). Therefore, there has also been growing interest in understanding mtDNA levels (mtDNA copy number) in PD.

Frequently, studies investigating mtDNA copy number have used qPCR methods to quantify the ratio of *MT-ND1* to a nuclear house-keeping gene. Recent studies have reported significant differences in mtDNA copy number values between PD patients and controls in the SN, with lower mtDNA copy numbers observed in the patients (Dölle et al., 2016; Grünwald et al., 2016; Pyle et al., 2016). Notably, in one study mtDNA copy number in DA neurons was observed to increase with age in controls (Dölle et al., 2016). Therefore, despite the age-related accumulation of mtDNA deletions, the pool of wild-type mtDNA was maintained. In contrast to the controls, no compensatory increase in mtDNA copy number was observed in patients, resulting in relative depletion of wild-type mtDNA in the SNpc. This was suggested to ultimately result in respiratory deficiency and neuronal loss. Bury et al. (Bury et al., 2017), who observed increased levels of both mtDNA deletions and mtDNA copy number in pedunculopontine cholinergic neurons of PD cases compared to controls, suggested that different neurochemical cell types and brain regions in PD patients may differ in response to mtDNA deletion accumulation. However, reasons for the lack of possible compensatory mechanisms in the SNpc of the PD brain remain unclear. Hence, further research is warranted on how mtDNA levels are regulated in different cell-types.

mtDNA depletion has also been recognised as a potential biomarker for PD detection because significant reductions in mtDNA copy number have not only been observed in the SN of PD patients (Dölle et al., 2016; Grünwald et al., 2016; Pyle et al., 2016) but also in peripheral blood (Pyle et al., 2016; Gui et al., 2015) and cerebrospinal fluid (Pyle et al., 2015). Overall, reports on mtDNA depletion in PD appear promising, although further research in this field is required, particularly with

Table 3
Summary of mtDNA sequencing studies on somatic variation in Parkinson's disease

Study participants	N	Brain region	Variants considered	Pathogenicity scoring	Method	mtDNA region sequenced	Outcome/ Main findings	Reference
Cases	8	FCtx (Tissue homogenate)	All point mutations; G:C to T:A and T:A to G:C transversions	NS	PCR-cloning-sequencing strategy	<i>MT-ND4</i>	No significant difference in heteroplasmic point mutation burdens between cases and controls.	Simon et al. (2004)
Controls (< 10 years old)	2	SN (Tissue homogenate)						
Controls (12-24 years old)	6							
Controls (24 years old)	7							
Controls (Elderly)	8							
Cases	3							
Cases (Elderly)	4							
Controls (< 10 years old)	10							
Controls (12-24 years old)								
Controls (Elderly)								
Cases	6	FCtx (Tissue homogenate)	Nonsynonymous point mutations	Yes	PCR-cloning-sequencing strategy	<i>MT-ND1</i> ; <i>MT-ND2</i> ; <i>MT-ND3</i> ; <i>MT-ND4L</i> ; <i>MT-ND4</i> ; <i>MT-ND5</i> ; <i>MT-ND6</i>	No significant difference in heteroplasmic point mutation burdens between cases and controls.	Smigrodzki et al. (2004)
Controls	6							
Cases (Early PD + ILBD)	9	SN (Isolated dopaminergic neurons; Glial cells)	All point mutations; GT/CA transversions	NS	PCR-cloning-sequencing strategy	<i>MT-ND5</i> ; D-loop	Significantly elevated levels of heteroplasmic point mutations in neurons of early PD + ILBD cases compared to controls ($P < 0.0001$) and late stage PD cases ($P = 0.0006$). No significant difference in mtDNA point mutation levels in SN glia from early PD + ILBD cases compared to controls ($P = 0.73$). Significantly higher heteroplasmic variant burden in the SN ($P = 0.012$) and FCtx ($P = 0.005$) of PD patients compared to controls.	Lin et al. (2012)
Cases (Late PD)	8							
Controls	23							
Cases	114	SN (Tissue homogenate)	All point mutations; nonsynonymous point mutations	Yes (MutPred scoring)	NGS (Illumina sequencing)	Whole mtDNA		Coxhead et al. (2016)
Controls	34	FCtx (Tissue homogenate)						
Cases	125							
Controls	30							
Cases	10	SNpc (Single dopaminergic neurons-184 in total)	SNVs; transitions and transversions; deletions	NS	Ultra-deep NGS (Illumina sequencing)	Two rRNA (<i>MT-RNR1</i> , <i>MT-RNR2</i>); 10 tRNA (<i>MT-TV</i> , <i>MT-TLL1</i> , <i>MT-TI</i> , <i>MT-TQ</i> , <i>MT-TM</i> , <i>MT-TW</i> , <i>MT-TA</i> , <i>MT-TN</i> , <i>MT-TC</i> , <i>MT-TY</i>); two Complex I genes (<i>MT-ND1</i> , <i>MT-ND2</i>)	Significantly higher levels of mtDNA deletions ($P = 0.004$), but not point mutations, in patients compared to controls. Significant mtDNA depletion in PD cases compared to controls ($P = 0.006$).	Dölle et al. (2016)
Controls	10							
Cases (DLB-PD)	89	Cerebellum; Cerebral Cortex; Other brain regions (Tissue homogenate)	All point mutations; nonsynonymous point mutations	Yes (MutPred and Polyphen-2)	NGS (Illumina sequencing)	Whole mtDNA (Extracted from whole exome sequencing data)	No significant difference in heteroplasmic point mutation burdens between cases and controls.	Wei et al. (2017)
Controls	351							
Controls (Young)	110							

DLB-PD = Dementia with Lewy Bodies or Parkinson's Disease; FCtx = Frontal cortex; ILBD = Incidental Lewy body disease; mtDNA = Mitochondrial DNA; N = Sample size; NA = Not applicable; NGS = Next Generation Sequencing; NS = Not specified; SN = Substantia Nigra; SNVs = Single nucleotide variants; PD = Parkinson's disease

regards to the mechanisms underlying mtDNA depletion and its utility as a biomarker.

5.3. Challenges of studying somatic mtDNA changes in PD

Studying acquired mtDNA changes in PD is particularly challenging, rendering the role of mtDNA in PD unresolved. Many factors can substantially influence study findings and can contribute to inconsistent reports such as those detailed above. In the following section we highlight important methodological and biological factors to consider when studying somatic mtDNA changes in PD case-control studies.

5.3.1. Techniques and technologies

Technologically, many well-established molecular techniques used to study nDNA variation may not be suitable for the study of mtDNA. A review by Moraes et al. (Moraes et al., 2003) drew attention to techniques including Southern blotting, PCR amplification, and RFLP, and reviewed their respective advantages and disadvantages when analysing heteroplasmic mtDNA point mutations and deletions. The authors highlighted that these techniques can yield misinterpreted results for mtDNA given the genome's unique properties of heteroplasmy and polyploidy. Thus, the use of Southern blotting and different PCR techniques in earlier mtDNA deletion studies could account for inconsistent findings. For instance, Ikebe et al. (Ikebe et al., 1990) and Ozawa et al. (Ozawa et al., 1990), employed PCR and kinetic-PCR techniques respectively, and reported significantly higher mtDNA common deletion levels in cases than in controls. In contrast Mann et al. (Mann et al., 1992) and Kösel et al. (Kösel et al., 1997) who employed Southern blotting and competitive PCR techniques respectively found no significant increase of this deletion in cases compared to controls. Newer techniques used to quantify mtDNA deletions also have disadvantages. For instance, the relative quantification of levels of *MT-ND1* to *MT-ND4* fails to quantify rarer mtDNA deletions which extend into the minor arc and may not reflect the true extent of deleted mtDNA.

Moreover, the ability to confidently detect and quantify low levels of point mutation heteroplasms in mtDNA varies between sequencing technologies (Pareek et al., 2011; Greaves et al., 2009). Some technologies such as Sanger sequencing are not sensitive enough to detect individually-rare mtDNA variants below 15% heteroplasmy, thereby potentially underestimating heteroplasmic variant levels (Zhang et al., 2012). Others, such as post-PCR cloning strategies, are at risk of overestimating heteroplasmic variant loads because of DNA polymerase enzyme transcriptional errors which are indistinguishable from true variants (Greaves et al., 2009). As a result, sequencing data generated using such technologies may not be truly representative of the heteroplasmic mtDNA levels in a sample and have likely contributed to some inconsistent findings between published studies (Table 3).

When employing newer, next generation sequencing approaches (NGS), sequencing depth affects the level of heteroplasmy detected (Ye et al., 2014b). This level varies between NGS approaches. Ultra-deep sequencing of the mtDNA, with a depth of several tens of thousands is required to confidently detect very low levels of heteroplasms (~1–10%). Additionally, technical artefacts and sequencing errors need to be controlled for in NGS to achieve accuracy of quantification and detection specificity (Ye et al., 2014b). To control for these errors and artefacts, NGS studies should integrate effective quality-control criteria. Web-based servers such as mtDNA-Server (<https://mtdna-server.uibk.ac.at>) are available specifically for the analysis of mtDNA NGS data. The workflow includes several quality control metrics, identification of artefacts and contamination, heteroplasmy detection, as well as variant annotation (Weissensteiner et al., 2016). The use of such servers could aid in standardising heteroplasmy detection and quantification across studies.

Moreover, as mentioned, mtDNA in post-mitotic tissue is likely to accumulate deletions with age (Kraytsberg et al., 2006). Consequently, mtDNA amplification in aged, post-mortem brains, prior to sequencing,

could also influence results of sequencing studies. For instance, amplifying mtDNA in two fragments which overlap the deletion-prone major arc may result in mtDNA sequence data which was unrepresentative of the entire mtDNA data population. This is because mtDNA carrying deletions in the primer binding sites may not be amplified given that primers may not bind and not amplify the mtDNA effectively.

Finally, predicting the functional impact of mtDNA variants in sequencing studies is important. This is because the levels of variation accumulating with age in controls could be similar to those accumulating in PD patients but may collectively be less pathogenic. It should be stressed that studies should use *in silico* tools specifically developed for mtDNA variation since those developed for nDNA may not be suitable to predict the pathogenicity of mtDNA variants (Bris et al., 2018). Examples of mtDNA tools include APOGEE (Castellana et al., 2017), MToolBox (Calabrese et al., 2014) and MitoTIP (Sonney et al., 2017). Tools like eKLIPse (Goudenège et al., 2007) are also available for identifying and quantifying mtDNA deletions (and other rearrangements) and should be considered for future studies.

5.3.2. Tissue and brain region

There are tissue specific differences in mtDNA heteroplasmy accumulation (Samuels et al., 2013) and ‘hot-spot’ regions of heteroplasmy accumulation in mtDNA (Li et al., 2010; Stoneking, 2000). As a result, the brain tissue and mtDNA region selected for examination in post-mortem studies will influence results. As detailed in Table 3, published studies did not all sequence the same mtDNA region, nor examined DNA from the same brain region, which could explain conflicting results. Although multiple regions of the brain are affected in PD, the loss of DA neurons in the pathologically affected SNpc is the most extensive (Brichta and Greengard, 2014). The SNpc, in comparison to other brain regions, may therefore be expected to show the greatest difference in mtDNA changes between cases and controls, should such changes contribute to neuronal loss in PD, as hypothesised.

Depending on cellular energy demand, the number of mitochondria per cell and the copies of mtDNA per mitochondria can differ substantially between cell types in whole tissue (Veltri et al., 1990). As such, whole blood samples frequently used for mtDNA copy number assays may be biased by cellular composition. For instance, samples with more granulocytes than lymphocytes, may show lower mtDNA copy numbers than ones with a higher proportion of lymphocytes (Pyle et al., 2010). This is because granulocytes have been observed to have fewer mitochondria, and thus less mtDNA, than lymphocytes (Pyle et al., 2010).

Moreover, mtDNA deletion and point mutation studies using tissue homogenate rather than single cells are also likely biased by the cellular composition of samples. mtDNA changes identified in tissue homogenate may not reflect the true changes in mtDNA between cases and controls. For instance, although both glial cells and neurons acquire mtDNA variants, somatic variants accumulate preferentially in neurons (Cantuti-Castelvetri et al., 2005). In support of this, Lin et al. (Lin et al., 2012) reported that overall levels of mtDNA point mutations were similar in SN glia of early PD and incidental Lewy body disease (ILBD) cases (Braak stage 3) and controls. However, they were significantly elevated in the single SN neurons of early PD + ILBD cases compared to controls (Table 3). As such, a homogenised tissue sample consisting of predominantly glial cells may have less mtDNA variation than one which consists predominantly of neuronal cells. Given that post-mortem brains of PD patients typically represent advanced stages of the disease and that up to 98% of DA SNpc neurons can be lost during advanced PD (Damier et al., 1999), mtDNA molecules examined in PD brain homogenate most likely originate primarily from surrounding glial cells.

Moreover, although studying individual cells is advantageous to avoid the bias of tissue composition, it can still be problematic. The surviving neurons in post-mortem brain tissue of patients with advanced PD may have accumulated fewer somatic mtDNA changes

allowing the neurons to persist. On the other hand, already degenerated neurons are likely to have accumulated very high levels of detrimental, somatic errors. Consequently, the surviving neurons of late-stage PD may not differ significantly to those of aged controls with regards to mtDNA changes. Therefore, the pathological stage of disease of the patients from whom the tissue samples originated needs to be accounted for (Simon et al., 2017). Examining neurons from early-stage (~Braak stage 3) post-mortem tissue may be the best strategy.

6. Future studies

To date the role of mtDNA in PD has not yet been resolved largely due to the difficulty of studying mtDNA variation but also 'poor' study design. However, there is potential for improvement in this field. Future studies need to consider the limitations of previous studies and the challenges associated with studying mtDNA to improve future study designs. For instance, small sample sizes are a significant limitation and can be overcome with new collaborations (Button et al., 2013). Studies should also consider using newer, more standardised approaches, such as the variant load approach, which potentially circumvent pitfalls of past investigations. Alternatively, future studies could branch into studying nDNA variation in conjunction with mtDNA changes. These and additional options are discussed below.

6.1. Homoplasmic mtDNA variation: considering functional studies and the collective role of rare population variants

Several authors have highlighted the methodological deficiencies of haplogroup association studies, some of which have also been outlined in this paper (Fachal et al., 2015; Herrnstadt and Howell, 2004; Raule et al., 2007). Yet, the methods used in these studies have not changed significantly over the past 20 years.

The current 'haplogroup association' approach uses only a few mtDNA SNPs. This makes it a cheap and relatively simple method to perform, that doesn't require more high throughput technologies such as whole exome or complete mitochondrial genome sequencing and avoids complex bioinformatics analyses. We suggest that results from future haplogroup studies should ideally be validated using functional investigations (e.g. cybrid or histochemical studies) (Giannoccaro et al., 2017) although this might be challenging given that association genetics is based on very subtle effects. Nonetheless, additional work assessing mitochondrial homeostasis regulation and function (e.g. measuring ROS production, cellular respiration etc.) can provide functional evidence to support the genetic associations made. For example, Liou et al. (Liou et al., 2016) were able to validate their findings using cybrids. In their study, cybrid results revealed that the B5 haplogroup cybrid, harbouring the m.8584G > A/m.10398A > G variants showed more resistance to rotenone than the B4 cybrid lacking these variants. Briefly, rotenone is a pesticide and a complex I inhibitor, implicated in PD pathogenesis (Betarbet et al., 2000). Furthermore, low ROS production and low apoptosis rates were also observed in the B5 cybrid (Liou et al., 2016), supporting the hypothesis that the B5 haplogroup variants have a protective effect against PD. Another recent study reported that haplogroup K1 increased mtDNA copy number and demonstrated a resistance to rotenone (Strobbe et al., 2018). These findings support haplogroup studies which associated haplogroup K with a PD protective effect (Ghezzi et al., 2005; Hudson et al., 2013), and highlight the value of such studies.

Although haplogroup association studies can be improved with such additional functional work, or by using replication cohorts, correcting for multiple testing, and avoiding pseudo-haplogroup construction, the current discrepancies between studies also highlight the need for newer and more standardised models.

Future studies should study homoplasmic, rare population variants which are mildly deleterious and could collectively contribute to disease risk in an individual, instead of only investigating haplogroup

defining variants in PD risk. This hypothesis, known as the variant load hypothesis, distances itself from the study of haplogroups (Pienaar et al., 2017). The most up-to-date version of the variant load hypothesis excludes variants predicted to be likely benign from the analysis, as many of these variants are common polymorphisms. Such common variants are more likely to be subject to population stratification, thus if left in the analysis they might be the cause of false positive associations. The variant load approach additionally condenses the likely impact of an individual's mtDNA variation into a numerical value on a continuous scale rather than in the form of a letter as done in haplogroup studies. Consequently, more powerful parametric statistics can be applied and fewer comparisons are needed, thus granting this approach greater statistical power with smaller sample sizes than the traditional haplogroup association method (Venter et al., 2017). For instance, a recent study with a total sample size of 82 employed the variant load approach to investigate the role of mtDNA in oxidative stress and inflammation (Venter et al., 2019). This study had 80% power to detect a correlation (with moderate effect size) between markers of oxidative stress and inflammation and variant loads. Given the decreasing cost of NGS and the growing availability of pipelines which simplify mtDNA data analyses, whole mtDNA sequencing data can be used to test the variant load hypothesis.

6.2. Somatic mtDNA variation and mtDNA depletion

Studying the role of somatic mtDNA variation in PD is challenging especially since the availability of PD post-mortem brains is limited and the pathological stage of disease may vary. Moreover, not all molecular techniques used to study nDNA variation are suitable for mtDNA given the unique features of mitochondrial genomes (polyploidy and heteroplasmy). Future studies should therefore carefully consider which molecular techniques would be most suitable for the purpose of their investigation to ensure results are reliable. Moreover, somatic mtDNA studies should correct for tissue cellular composition bias where possible. A combination of the strategies employed in the herein reviewed studies could be best to study somatic mtDNA variation in PD, particularly mtDNA point mutations. For instance, ultra-deep sequencing technologies could be used to sequence mtDNA molecules of individual, early-stage PD and control SNpc neurons.

mtDNA depletion in PD is an emerging biomarker of disease pathology and may be more promising to understand PD than somatic mtDNA variation given that mtDNA depletion studies have reported few discrepancies (Dölle et al., 2016; Grünewald et al., 2016; Pyle et al., 2015, 2016; Gui et al., 2015). We suggest future studies should try to replicate these reports of mtDNA depletion in PD but also correct for cellular composition when possible, to make results more reliable and reproducible (Guyatt et al., 2019).

Future studies should also place greater focus on studying nDNA variation, in combination with mtDNA changes, which may underlie potentially high mtDNA variant burdens in PD brains, and/or mtDNA depletion. For instance, Gui et al. (Gui et al., 2015) reported that PD patients with a particular *POLG1* genotype in combination with additional *POLG1* variants, had significantly lower numbers of mtDNA than those without such variants. This suggested that *POLG1* variation may contribute to mtDNA depletion in PD cases. *POLG1* encodes a subunit of the DNA polymerase which replicates mtDNA. mtDNA mutator mice with mutations in nDNA encoding mtDNA replication components, including *Polg1*, have additionally demonstrated an accelerated accumulation of mtDNA point mutations and deletions, highlighting a potential role for nDNA underlying mtDNA changes (Edgar et al., 2009; Trifunovic et al., 2004, 2005; Vermulst et al., 2008). As a result of these mtDNA changes, the mutator mice demonstrate a progressive respiratory chain dysfunction as seen in PD patients and premature aging phenotypes (e.g. hair loss, osteoporosis, and progressive hearing loss) (Trifunovic et al., 2004). These studies again implicate somatic mtDNA changes, resulting from nDNA mutations, in PD which has been

hypothesised to be a form of accelerated aging (Barbeau, 1984; Collier et al., 2017). Mitochondrial transcription factor A (TFAM) is an essential DNA-binding protein required for transcription and maintenance of mtDNA. Knockout mice with a deleted *Tfam* gene demonstrate a respiratory chain deficiency and reduced mtDNA expression in DA neurons (Ekstrand et al., 2007). These mice also exhibit parkinsonism typical phenotypes including progressive motor function impairment and DA neuronal loss (Ekstrand et al., 2007) implicating defective mtDNA maintenance in PD, brought about by nDNA changes. Notably, components of the nDNA-encoded, mtDNA replication machinery including mitochondrial genome maintenance exonuclease 1 (MGME1), POLG1, and mtDNA replicative helicase (TWINKLE) have recently been implicated in degradation of damaged mtDNA (Peeva et al., 2018). Thus, nDNA variation in genes encoding these factors is of particular interest for further study given their dual role mtDNA synthesis and degradation (Nissanka et al., 2019).

6.3. Mitochondrial-derived peptides (MDPs)

Recently, MDPs, encoded as sORFs, in mtDNA have been identified as important regulatory peptides of metabolic activity and stress response (Lee et al., 2013, 2015; Kim et al., 2017). The discovery of MDPs underscores the potential existence of more sORFs hidden in mtDNA. Recent findings have implicated the mitochondrial open reading frame of the 12S rRNA-c (MOTS-c), a 16-amino-acid MDP encoded within the *MT-RNR1* gene, in the regulation of nuclear gene expression in response to metabolic stress (Kim et al., 2018). These findings suggest that nuclear and mitochondrial genomes cross regulate each other but they also highlight the potential of MDP involvement in disease (Lee, 2019). Further work needs to be conducted to better understand the role of MDPs and the mechanisms regulating their expression in PD.

7. Concluding remarks

PD is an immensely complex disorder of multifactorial origin, but one with clear links to mitochondria and their genomes. Early evidence directly implicates mtDNA of sPD patients in altering mitochondrial structure and function. However, despite a multitude of studies investigating mtDNA in PD, the contribution of the mitochondrial genome to this disorder remains elusive.

To date, studies investigating homoplasmic and heteroplasmic mtDNA changes in PD have produced conflicting results. In almost all of these studies, low statistical power resulting from small sample sizes is a significant limitation, which likely contributed to some discrepant findings. Most mtDNA haplogroup association studies additionally suffered from methodological deficiencies (e.g. population stratification, lack of replication cohorts and correction for multiple testing) which could primarily contribute to such discrepancies. On the other hand, innate difficulties of studying somatic mtDNA changes (e.g. heteroplasmy and tissue specific differences in mtDNA) along with different methodological techniques applied have produced conflicting results in somatic mtDNA studies which either provide evidence for or against a role of somatic mtDNA changes in PD.

Consequently, a role for the involvement of mtDNA in PD cannot at this stage be ruled out and needs to be further investigated in a more critical and systematic way. Future studies should take into account limitations of past studies reviewed here to improve current study designs. Alternatively, they should consider the use of more standardised approaches such as the variant load approach. The limitations, challenges, and future prospects of studying mtDNA in PD are summarised in [Supplementary Fig. 1](#). Although not caveat-free, the study of mtDNA depletion appears most promising and we propose that further research in this area is required, particularly regarding the mechanisms and possible nDNA variation underlying mtDNA depletion in PD patients.

Albeit small, relative to the nuclear genome, the mitochondrial genome is an integral component for the production and functioning of

mitochondria, and overall eukaryotic cell homeostasis. The role of mtDNA therefore should not be underestimated, and future adequately powered and well-designed studies exploring current (mtDNA variation and levels) and new avenues (including MDPs) are needed to fully assess the magnitude of its contribution to PD pathogenesis.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

The authors are supported by the National Research Foundation of South Africa (Grant Numbers: 106052), the South African Medical Research Council (Self-Initiated Research Grant), and Stellenbosch University. This work was also supported by an NIH grant (P50NS072187) and DOD award (W81XWH-17-1-0249) to OAR, the Mayo Clinic investigators are supported in part by a Lewy Body Dementia Association (LBDA) Research Center of Excellence, an American Parkinson Disease Association Center for Advanced Research and the Michael J. Fox Foundation.

We also acknowledge the support of the NRF-DST Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuint.2019.104495>.

References

- Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N., 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23 (2), 147.
- Arnheim, N., Cortopassi, G., 1992. Deleterious mitochondrial DNA mutations accumulate in aging human tissues. *Mutat. Res. - DNAging* 275 (3–6), 157–167.
- Autere, J., Moilanen, J.S., Finnilä, S., Soininen, H., Mannermaa, A., Hartikainen, P., Hallikainen, M., Majamaa, K., 2004. Mitochondrial DNA polymorphisms as risk factors for Parkinson's disease and Parkinson's disease dementia. *Hum. Genet.* 115 (1), 29–35.
- Bandmann, O., Sweeney, M., Daniel, S., Marsden, C., Wood, N., 1997. Mitochondrial DNA polymorphisms in pathologically proven Parkinson's disease. *J. Neurol.* 244 (4), 262–265.
- Barbeau, A., 1984. Etiology of Parkinson's disease: a research strategy. *Can. J. Neurol. Sci.* 11 (1), 24–28.
- Bender, A., Krishnan, K.J., Morris, C.M., Taylor, G.A., Reeve, A.K., Perry, R.H., Jaros, E., Hersheson, J.S., Betts, J., Klopstock, T., 2006. High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat. Genet.* 38 (5), 515.
- Bender, A., Schwarzkopf, R.-M., McMillan, A., Krishnan, K.J., Rieder, G., Neumann, M., Elstner, M., Turnbull, D.M., Klopstock, T., 2008. Dopaminergic midbrain neurons are the prime target for mitochondrial DNA deletions. *J. Neurol.* 255 (8), 1231–1235.
- Benecke, R., Strümpfer, P., Weiss, H., 1993. Electron transfer complexes I and IV of platelets are abnormal in Parkinson's disease but normal in Parkinson-plus syndromes. *Brain* 116 (6), 1451–1463.
- Betarbet, R., Sherer, T.B., MacKenzie, G., Garcia-Osuna, M., Panov, A.V., Greenamyre, J.T., 2000. Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nat. Neurosci.* 3 (12), 1301.
- Bindoff, L., Birch-Machin, M., Cartledge, N., Parker Jr., W., Turnbull, D., 1991. Respiratory chain abnormalities in skeletal muscle from patients with Parkinson's disease. *J. Neurol. Sci.* 104 (2), 203–208.
- Blin, O., Desnuelle, C., Rascol, O., Borg, M., Saint Paul, H.P., Azulay, J., Bille, F., Figarella, D., Coulom, F., Pellissier, J., 1994. Mitochondrial respiratory failure in skeletal muscle from patients with Parkinson's disease and multiple system atrophy. *J. Neurol. Sci.* 125 (1), 95–101.
- Borland, M.K., Mohanakumar, K., Rubinstein, J.D., Keeney, P.M., Xie, J., Capaldi, R., Dunham, L.D., Trimmer, P.A., Bennett Jr., J.P., 2009. Relationships among molecular genetic and respiratory properties of Parkinson's disease cybrid cells show similarities to Parkinson's brain tissues. *Biochim. Biophys. Acta (BBA) - Mol. Basis Dis.* 1792 (1), 68–74.
- Brichta, L., Greengard, P., 2014. Molecular determinants of selective dopaminergic vulnerability in Parkinson's disease: an update. *Front. Neuroanat.* 8, 152.

- Bris, C., Goudenege, D., Desquiret-Dumas, V., Charif, M., Colin, E., Bonneau, D., Bonneau, P., Lemaers, G., Reynier, P., Procaccio, V., 2018. Bioinformatics tools and databases to assess the pathogenicity of mitochondrial DNA variants in the field of next generation sequencing. *Front. Genet.* 9, 632.
- Brown, M.D., Shoffner, J.M., Kim, Y.L., Jun, A.S., Graham, B.H., Cabell, M.F., Gurley, D.S., Wallace, D.C., 1996. Mitochondrial DNA sequence analysis of four Alzheimer's and Parkinson's disease patients. *Am. J. Med. Genet.* 61 (3), 283–289.
- Bulua, A.C., Simon, A., Maddipati, R., Pelletier, M., Park, H., Kim, K.-Y., Sack, M.N., Kastner, D.L., Siegel, R.M., 2011. Mitochondrial reactive oxygen species promote production of proinflammatory cytokines and are elevated in TNFR1-associated periodic syndrome (TRAPS). *J. Exp. Med.* 208 (3), 519–533.
- Burns, R.S., Chieh, C.C., Markey, S.P., Ebert, M.H., Jacobowitz, D.M., Kopin, I.J., 1983. A primate model of parkinsonism: selective destruction of dopaminergic neurons in the pars compacta of the substantia nigra by N-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine. *Proc. Natl. Acad. Sci. Unit. States Am.* 80 (14), 4546–4550.
- Bury, A.G., Pyle, A., Elson, J.L., Greaves, L., Morris, C.M., Hudson, G., Pienaar, I.S., 2017. Mitochondrial DNA changes in pedunculopontine cholinergic neurons in Parkinson disease. *Ann. Neurol.* 82 (6), 1016–1021.
- Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14 (5), 365.
- Calabrese, C., Simone, D., Diroma, M.A., Santorsola, M., Gutta, C., Gasparre, G., Picardi, E., Pesole, G., Attimonelli, M., 2014. MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. *Bioinformatics* 30 (21), 3115–3117.
- Cantuti-Castelvetri, I., Lin, M.T., Zheng, K., Keller-McGandy, C.E., Betensky, R.A., Johns, D.R., Beal, M.F., Standaert, D.G., Simon, D.K., 2005. Somatic mitochondrial DNA mutations in single neurons and glia. *Neurobiol. Aging* 26 (10), 1343–1355.
- Cardellach, F., Martí, M., Fernandez-Sola, J., Marin, C., Hoek, J., Tolosa, E., Urbano-Marquez, A., 1993. Mitochondrial respiratory chain activity in skeletal muscle from patients with Parkinson's disease. *Neurology* 43 (11), 2258–2258.
- Castellana, S., Fusilli, C., Mazzoccoli, G., Biagini, T., Capocefalo, D., Carella, M., Vescovi, A.L., Mazza, T., 2017. High-confidence assessment of functional impact of human mitochondrial non-synonymous genome variations by APOGEE. *PLoS Comput. Biol.* 13 (6), e1005628.
- Chen, C., Kuan, C., Lee-Chen, G.-J., Wu, Y., 2007. Mitochondrial DNA polymorphisms and the risk of Parkinson's disease in Taiwan. *J. Neural Transm.* 114 (8), 1017.
- Chen, Y.-F., Chen, W.-J., Lin, X.-Z., Zhang, Q.-J., Cai, J.-P., Liou, C.-W., Wang, N., 2015. Mitochondrial DNA haplogroups and the risk of sporadic Parkinson's disease in han Chinese. *Chin. Med. J.* 128 (13), 1748.
- Chinnery, P.F., Hudson, G., 2013. Mitochondrial genetics. *Br. Med. Bull.* 106 (1), 135–159.
- Chinnery, P.F., Samuels, D.C., 1999. Relaxed replication of mtDNA: a model with implications for the expression of disease. *Am. J. Hum. Genet.* 64 (4), 1158–1165.
- Chu, Q., Luo, X., Zhan, X., Ren, Y., Pang, H., 2015. Female genetic distribution bias in mitochondrial genome observed in Parkinson's Disease patients in northern China. *Sci. Rep.* 5, 17170–17170.
- Collier, T.J., Kanaan, N.M., Kordower, J.H., 2017. Aging and Parkinson's disease: different sides of the same coin? *Mov. Disord.* 32 (7), 983–990.
- Coxhead, J., Kurzawa-Akanbi, M., Hussain, R., Pyle, A., Chinnery, P., Hudson, G., 2016. Somatic mtDNA variation is an important component of Parkinson's disease. *Neurobiol. Aging* 38, 217.e1–217.e6.
- Damier, P., Hirsch, E., Agid, Y., Graybiel, A., 1999. The substantia nigra of the human brain: II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. *Brain* 122 (8), 1437–1448.
- Davis, G.C., Williams, A.C., Markey, S.P., Ebert, M.H., Caine, E.D., Reichert, C.M., Kopin, I.J., 1979. Chronic parkinsonism secondary to intravenous injection of meperidine analogues. *Psychiatr. Res.* 1 (3), 249–254.
- Dölle, C., Flönes, I., Nido, G.S., Miletic, H., Osuagwu, N., Kristoffersen, S., Lilleng, P.K., Larsen, J.P., Tysnes, O.-B., Haugarvoll, K., 2016. Defective mitochondrial DNA homeostasis in the substantia nigra in Parkinson disease. *Nat. Commun.* 7, 13548.
- Dominy, J.E. and P. Puigserver, Mitochondrial biogenesis through activation of nuclear signaling proteins. *Cold Spring Harb. Perspect. Biol.* 5(7): p. a015008.
- Edgar, D., Shabalina, I., Camara, Y., Wredenberg, A., Calvaruso, M.A., Nijtmans, L., Nedergaard, J., Cannon, B., Larsson, N.-G., Trifunovic, A., 2009. Random point mutations with major effects on protein-coding genes are the driving force behind premature aging in mtDNA mutator mice. *Cell Metabol.* 10 (2), 131–138.
- Ekstrand, M.I., Terzioglu, M., Galter, D., Zhu, S., Hofstetter, C., Lindqvist, E., Thams, S., Bergstrand, A., Hansson, F.S., Trifunovic, A., 2007. Progressive parkinsonism in mice with respiratory-chain-deficient dopamine neurons. *Proc. Natl. Acad. Sci. Unit. States Am.* 104 (4), 1325–1330.
- El-Hattab, A.W., Scaglia, F., 2013. Mitochondrial DNA depletion syndromes: review and updates of genetic basis, manifestations, and therapeutic options. *Neurother.* 10 (2), 186–198.
- Elbaz, A., Grigoletto, F., Baldereschi, M., Breteler, M., Manubens-Bertran, J., Lopez-Pousa, S., Dartigues, J., Alperovitch, A., Tzourio, C., Rocca, W., 1999. Familial aggregation of Parkinson's disease A population-based case-control study in Europe. *Neurology* 52 (9), 1876–1876.
- Elson, J., Samuels, D., Turnbull, D., Chinnery, P., 2001. Random intracellular drift explains the clonal expansion of mitochondrial DNA mutations with age. *Am. J. Hum. Genet.* 68 (3), 802–806.
- Esteves, A.R.F., Domingues, A.F., Ferreira, I.L., Januário, C., Swerdlow, R.H., Oliveira, C.R., Cardoso, S.M., 2008. Mitochondrial function in Parkinson's disease cybrids containing an nt2 neuron-like nuclear background. *Mitochondrion* 8 (3), 219–228.
- Esteves, A.R., Arduíno, D.M., Swerdlow, R.H., Oliveira, C.R., Cardoso, S.M., 2009. Oxidative stress involvement in α -synuclein oligomerization in Parkinson's disease cybrids. *Antioxidants Redox Signal.* 11 (3), 439–448.
- Esteves, A.R., Lu, J., Rodova, M., Onyango, I., Lezi, E., Dubinsky, R., Lyons, K.E., Pahwa, R., Burns, J.M., Cardoso, S.M., 2010a. Mitochondrial respiration and respiration-associated proteins in cell lines created through Parkinson's subject mitochondrial transfer. *J. Neurochem.* 113 (3), 674–682.
- Esteves, A.R., Arduíno, D.M., Swerdlow, R.H., Oliveira, C.R., Cardoso, S.M., 2010b. Dysfunctional mitochondria uphold calpain activation: contribution to Parkinson's disease pathology. *Neurobiol. Dis.* 37 (3), 723–730.
- Esteves, A.R., Arduíno, D.M., Swerdlow, R.H., Oliveira, C., Cardoso, S.M., 2010c. Microtubule depolymerization potentiates α -synuclein oligomerization. *Front. Aging Neurosci.* 1, 5.
- Fachal, L., Mosquera-Miguel, A., Pastor, P., Ortega-Cubero, S., Lorenzo, E., Oterino-Durán, A., Toriello, M., Quintán, B., Camiña-Tato, M., Sesar, A., 2015. No evidence of association between common European mitochondrial DNA variants in Alzheimer, Parkinson, and migraine in the Spanish population. *Am. J. Med. Genet. Part B: Neuropsychiatric Genetics* 168 (1), 54–65.
- Gawda-Walerych, K., Maruszak, A., Safranow, K., Bialecka, M., Klodowska-Duda, G., Czyżewski, K., Slawek, J., Rudzinska, M., Styczynska, M., Opala, G., 2008. Mitochondrial DNA haplogroups and subhaplogroups are associated with Parkinson's disease risk in a Polish PD cohort. *J. Neural Transm.* 115 (11), 1521–1526.
- Gelb, D.J., Oliver, E., Gilman, S., 1999. Diagnostic criteria for Parkinson disease. *Arch. Neurol.* 56 (1), 33–39.
- Georgiou, A., Demetriou, C.A., Heraclides, A., Christou, Y.P., Leonidou, E., Loukaides, P., Yiasoumi, E., Panagiotou, D., Manoli, P., Thomson, P., 2017. Mitochondrial superclusters influence age of onset of Parkinson's disease in a gender specific manner in the Cypriot population: a case-control study. *PLoS One* 12 (9), e0183444.
- Ghezzi, D., Marelli, C., Achilli, A., Goldwurm, S., Pezzoli, G., Barone, P., Pellicchia, M.T., Stanzione, P., Brusa, L., Bentivoglio, A.R., 2005. Mitochondrial DNA haplogroup K is associated with a lower risk of Parkinson's disease in Italians. *Eur. J. Hum. Genet.* 13 (6), 748.
- Ghosh, S.S., Swerdlow, R.H., Miller, S.W., Sheeman, B., PARKER Jr., W.D., Davis, R.E., 1999. Use of cytoplasmic hybrid cell lines for elucidating the role of mitochondrial dysfunction in Alzheimer's disease and Parkinson's disease. *Ann. N. Y. Acad. Sci.* 893 (1), 176–191.
- Giannoccaro, Maria Pia, Morgia, Chiara La, Rizzo, Giovanni, Carelli, Valerio, 2017. Mitochondrial DNA and primary mitochondrial dysfunction in Parkinson's disease. *Mov. Disord.* 32 (3), 346–363. <https://doi.org/10.1002/mds.26966>.
- Goudenege, D., Bris, C., Hoffmann, V., Desquiret-Dumas, V., Jardel, C., Rucheton, B., Bannwarth, S., Paquis-Flucklinger, V., Lebre, A.S., Colin, E., 2017. eKLPse: a sensitive tool for the detection and quantification of mitochondrial DNA deletions from next-generation sequencing data. *Genet. Med.* 21 (6).
- Grasbon-Frodl, E.M., Kösel, S., Sprinzl, M., von Eitzen, U., Mehraein, P., Graeber, M.B., 1999. Two novel point mutations of mitochondrial tRNA genes in histologically confirmed Parkinson disease. *Neurogenetics* 2 (2), 121–127.
- Greaves, L.C., Beadle, N.E., Taylor, G.A., Commane, D., Mathers, J.C., Khrapko, K., Turnbull, D.M., 2009. Quantification of mitochondrial DNA mutation load. *Aging Cell* 8 (5), 566–572.
- Grünwald, A., Rygiel, K.A., Hepplewhite, P.D., Morris, C.M., Picard, M., Turnbull, D.M., 2016. Mitochondrial DNA depletion in respiratory chain-deficient Parkinson disease neurons. *Ann. Neurol.* 79 (3), 366–378.
- Gu, M., Cooper, J., Taanman, J., Schapira, A., 1998. Mitochondrial DNA transmission of the mitochondrial defect in Parkinson's disease. *Ann. Neurol.* 44 (2), 177–186.
- Gu, G., Reyes, P.F., Golden, G.T., Woltjer, R.L., Hulette, C., Montine, T.J., Zhang, J., 2002. Mitochondrial DNA deletions/rearrangements in Parkinson disease and related neurodegenerative disorders. *J. Neuropathol. Exp. Neurol.* 61 (7), 634–639.
- Gui, Y.-X., Xu, Z.-P., Lv, W., Zhao, J.-J., Hu, X.-Y., 2015. Evidence for polymerase gamma, POLG1 variation in reduced mitochondrial DNA copy number in Parkinson's disease. *Park. Relat. Disord.* 21 (3), 282–286.
- Guyatt, A.L., Brennan, R.R., Burrows, K., Guthrie, P.A., Ascione, R., Ring, S.M., Gaunt, T.R., Pyle, A., Cordell, H.J., Lawlor, D.A., 2019. A genome-wide association study of mitochondrial DNA copy number in two population-based cohorts. *Hum. Genom.* 13 (1), 6.
- Haas, R.H., Nasirian, F., Nakano, K., Ward, D., Pay, M., Hill, R., Shults, C.W., 1995. Low platelet mitochondrial complex I and complex II/III activity in early untreated Parkinson's disease. *Ann. Neurol.* 37 (6), 714–722.
- Halliwell, B., Gutteridge, J., 1984. Role of iron in oxygen radical reactions. *Methods Enzymol.* 105, 47.
- Heikkilä, R.E., Hess, A., Duvoisin, R.C., 1984. Dopaminergic neurotoxicity of 1-methyl-4-phenyl-1, 2, 5, 6-tetrahydropyridine in mice. *Science* 224 (4656), 1451–1453.
- Heikkilä, R.E., Nicklas, W.J., Vyas, I., Duvoisin, R.C., 1985. Dopaminergic toxicity of rotenone and the 1-methyl-4-phenylpyridinium ion after their stereotaxic administration to rats: implication for the mechanism of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine toxicity. *Neurosci. Lett.* 62 (3), 389–394.
- Hellwege, J.N., Keaton, J.M., Giri, A., Gao, X., Velez Edwards, D.R., Edwards, T.L., 2017. Population stratification in genetic association studies. *Curr. Protoc. Hum. Genet.* 95 (1), 1.22. 1–12. 22. 23.
- Herbers, E., Kékäläinen, N.J., Hangas, A., Pohjoismäki, J.L., Goffart, S., 2019. Tissue specific differences in mitochondrial DNA maintenance and expression. *Mitochondrion* 44, 85–92.
- Herrnstadt, C., Howell, N., 2004. An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders. *Mitochondrion* 4 (5), 791–798.
- Howell, N., Elson, J.L., Chinnery, P.F., Turnbull, D.M., 2005. mtDNA mutations and common neurodegenerative disorders. *Trends Genet.* 21 (11), 583–586.
- Hudson, G., Nalls, M., Evans, J.R., Breen, D.P., Winder-Rhodes, S., Morrison, K.E., Morris, H.R., Williams-Gray, C.H., Barker, R.A., Singleton, A.B., 2013. Two-stage association

- study and meta-analysis of mitochondrial DNA variants in Parkinson disease. *Neurology* 80 (22), 2042–2048.
- Huerta, C., Castro, M.G., Coto, E., Blázquez, M., Ribacoba, R., Guisasaola, L.M., Salvador, C., Martínez, C., Lahoz, C.H., Alvarez, V., 2005. Mitochondrial DNA polymorphisms and risk of Parkinson's disease in Spanish population. *J. Neurol. Sci.* 236 (1), 49–54.
- Ikebe, S.-i., Tanaka, M., Ohno, K., Sato, W., Hattori, K., Kondo, T., Mizuno, Y., Ozawa, T., 1990. Increase of deleted mitochondrial DNA in the striatum in Parkinson's disease and senescence. *Biochem. Biophys. Res. Commun.* 170 (3), 1044–1048.
- Ikebe, S.-i., Tanaka, M., Ozawa, T., 1995. Point mutations of mitochondrial genome in Parkinson's disease. *Mol. Brain Res.* 28 (2), 281–295.
- Jandova, J., Janda, J., Sligh, J.E., 2012. Changes in mitochondrial DNA alter expression of nuclear encoded genes associated with tumorigenesis. *Exp. Cell Res.* 318 (17), 2215–2225.
- Janetzky, B., Hauck, S., Youdim, M.B., Riederer, P., Jellinger, K., Pantucek, F., Zo, R., Boissl, K.W., Reichmann, H., 1994. Unaltered aconitase activity, but decreased complex I activity in substantia nigra pars compacta of patients with Parkinson's disease. *Neurosci. Lett.* 169 (1–2), 126–128.
- Keeney, P.M., Dunham, L.D., Quigley, C.K., Morton, S.L., Bergquist, K.E., Bennett Jr., J.P., 2009. Cybrid models of Parkinson's disease show variable mitochondrial biogenesis and genotype-respiration relationships. *Exp. Neurol.* 220 (2), 374–382.
- Kenney, M.C., Chwa, M., Atilano, S.R., Falatounzadeh, P., Ramirez, C., Malik, D., Tarek, M., del Carpio, J.C., Nesburn, A.B., Boyer, D.S., 2014. Molecular and bioenergetic differences between cells with African versus European inherited mitochondrial DNA haplogroups: implications for population susceptibility to diseases. *Biochim. Biophys. Acta (BBA) - Mol. Basis Dis.* 1842 (2), 208–219.
- Khrapko, K., Collier, H.A., André, P.C., Li, X.-C., Hanekamp, J.S., Thilly, W.G., 1997. Mitochondrial mutational spectra in human cells and tissues. *Proc. Natl. Acad. Sci. Unit. States Am.* 94 (25), 13798–13803.
- Khusnutdinova, E., Gilyazova, I., Ruiz-Pesini, E., Derbeneva, O., Khusainova, R., Khidiyatova, I., Magzhanov, R., Wallace, D.C., 2008. A mitochondrial etiology of neurodegenerative diseases: evidence from Parkinson's disease. *Ann. N. Y. Acad. Sci.* 1147 (1), 1–20.
- Kim, S.J., Xiao, J., Wan, J., Cohen, P., Yen, K., 2017. Mitochondrially derived peptides as novel regulators of metabolism. *J. Physiol.* 595 (21), 6613–6621.
- Kim, K.H., Son, J.M., Benayoun, B.A., Lee, C., 2018. The mitochondrial-encoded peptide MOTS-c translocates to the nucleus to regulate nuclear gene expression in response to metabolic stress. *Cell Metabol.* 28 (3), 516–524 e7.
- Kirchner, S.C., Hallagan, S.E., Farin, F.M., Dilley, J., Costa-Mallen, P., Smith-Weller, T., Franklin, G.M., Swanson, P.D., Checkoway, H., 2000. Mitochondrial ND1 sequence analysis and association of the T4216C mutation with Parkinson's disease. *Neurotoxicology* 21 (4), 441–445.
- Kösel, S., Egensperger, R., Schnopp, N.M., Graeber, M.B., 1997. The 'common deletion' is not increased in parkinsonian substantia nigra as shown by competitive polymerase chain reaction. *Mov. Disord. Off. J. Mov. Disord. Soc.* 12 (5), 639–645.
- Kösel, S., Grasbon-Frodl, E.M., Mautsch, U., Egensperger, R., von Eitzen, U., Frishman, D., Hofmann, S., Gerbitz, K.-D., Mehraein, P., Graeber, M.B., 1998. Novel mutations of mitochondrial complex I in pathologically proven Parkinson disease. *Neurogenetics* 1 (3), 197–204.
- Kraytberg, Y., Kudryavtseva, E., McKee, A.C., Geula, C., Kowall, N.W., Khrapko, K., 2006. Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. *Nat. Genet.* 38 (5), 518.
- Krige, D., Carroll, M.T., Cooper, J.M., Marsden, C.D., Schapira, A.H., 1992. Platelet mitochondrial function in Parkinson's disease. *Ann. Neurol.* 32 (6), 782–788.
- Kumar, K.R., Djarmati-Westenberger, A., Grünewald, A., 2011. Genetics of Parkinson's disease. *Semin. Neurol.* 31 (05), 433–440.
- Langston, J.W., Ballard, P., Tetrud, J.W., Irwin, I., 1983. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. *Science* 219 (4587), 979–980.
- Latsoudis, H., Spanaki, C., Chlouverakis, G., Plaitakis, A., 2008. Mitochondrial DNA polymorphisms and haplogroups in Parkinson's disease and control individuals with a similar genetic background. *J. Hum. Genet.* 53 (4), 349–356.
- Lee, Changhan, 2019. Nuclear transcriptional regulation by mitochondrial-encoded MOTS-c. *Molecular & Cellular Oncology* 6, 2. <https://doi.org/10.1080/23723556.2018.1549464>.
- Lee, C., Yen, K., Cohen, P., 2013. Humanin: a harbinger of mitochondrial-derived peptides? *Trends Endocrinol. Metabol.* 24 (5), 222–228.
- Lee, C., Zeng, J., Drew, B.G., Sallam, T., Martin-Montalvo, A., Wan, J., Kim, S.-J., Mehta, H., Hevener, A.L., de Cabo, R., 2015. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metabol.* 21 (3), 443–454.
- Li, M., Schönberg, A., Schaefer, M., Schroeder, R., Nasidze, I., Stoneking, M., 2010. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am. J. Hum. Genet.* 87 (2), 237–249.
- Lin, M.T., Simon, D.K., Ahn, C.H., Kim, L.M., Beal, M.F., 2002. High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain. *Hum. Mol. Genet.* 11 (2), 133–145.
- Lin, M.T., Cantuti-Castelvetri, I., Zheng, K., Jackson, K.E., Tan, Y.B., Arzberger, T., Lees, A.J., Betensky, R.A., Beal, M.F., Simon, D.K., 2012. Somatic mitochondrial DNA mutations in early Parkinson and incidental Lewy body disease. *Ann. Neurol.* 71 (6), 850–854.
- Liou, C., Chuang, J., Chen, J., Tiao, M., Wang, P., Huang, S., Huang, T., Lee, W., Weng, S., Huang, P., 2016. Mitochondrial DNA variants as genetic risk factors for Parkinson disease. *Eur. J. Neurol.* 23 (8), 1289–1300.
- Liu, W., Vives-Bauza, C., Yamamoto, A., Tan, Y., Li, Y., Magrané, J., Stavarache, M.A., Shaffer, S., Chang, S., Kapliot, M.G., 2009. PINK1 defect causes mitochondrial dysfunction, proteasomal deficit and α -synuclein aggregation in cell culture models of Parkinson's disease. *PLoS One* 4 (2), e4597.
- Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., Procaccio, V., Wallace, D.C., 2013. mtDNA variation and analysis using MITOMAP and MITOMASTER. *Curr. Protoc. Bioinf.* 1.23. 1–1.23. 26.
- Luo, S., Valencia, C.A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., Wang, X., Li, Z., Dell, S., Brown, J., 2018. Biparental inheritance of mitochondrial DNA in humans. *Proc. Natl. Acad. Sci. Unit. States Am.* 115 (51), 13039–13044.
- Mann, V.M., Cooper, J.M., Schapira, A.H., 1992. Quantitation of a mitochondrial DNA deletion in Parkinson's disease. *FEBS Lett.* 299 (3), 218–222.
- Mayr-Wohlfart, U., Rödel, G., Henneberg, A., 1997. Mitochondrial tRNA(Gln) and tRNA(Thr) gene variants in Parkinson's disease. *Eur. J. Med. Res.* 2, 111–113.
- Mehta, P., Mellick, G.D., Rowe, D.B., Halliday, G.M., Jones, M.M., Manwaring, N., Vandebona, H., Silburn, P.A., Wang, J.J., Mitchell, P., 2009. Mitochondrial DNA haplogroups J and K are not protective for Parkinson's disease in the Australian community. *Mov. Disord.* 24 (2), 290–292.
- Miller, F.J., Rosenfeldt, F.L., Zhang, C., Linnane, A.W., Nagley, P., 2003. Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. *Nucleic Acids Res.* 31 (11) e61–e61.
- Mizuno, Y., Suzuki, K., Sone, N., Saitoh, T., 1987. Inhibition of ATP synthesis by 1-methyl-4-phenylpyridinium ion (MPP+) in isolated mitochondria from mouse brains. *Neurosci. Lett.* 81 (1–2), 204–208.
- Moraes, C.T., Atencio, D.P., Oca-Cossio, J., Diaz, F., 2003. Techniques and pitfalls in the detection of pathogenic mitochondrial DNA mutations. *J. Mol. Diagn.* 5 (4), 197–208.
- Nakagawa-Hattori, Y., Yoshino, H., Kondo, T., Mizuno, Y., Horai, S., 1992. Is Parkinson's disease a mitochondrial disorder? *J. Neurol. Sci.* 107 (1), 29–33.
- Neuhaus, J.F.G., Baris, O.R., Hess, S., Moser, N., Schröder, H., Chinta, S.J., Andersen, J.K., Kloppenburg, P., Wiesner, R.J., 2013. Catecholamine metabolism drives generation of mitochondrial DNA deletions in dopaminergic neurons. *Brain* 137 (2), 354–365.
- Nido, G.S., Dölle, C., Flönes, I., Tuppen, H.A., Alves, G., Tysnes, O.-B., Haugarvoll, K., Tzoulis, C., 2018. Ultradeep mapping of neuronal mitochondrial deletions in Parkinson's disease. *Neurobiol. Aging* 63, 120–127.
- Nissanka, Nadee, 2019. Michal Minczuk, Carlos T. Moraes, Mechanisms of Mitochondrial DNA Deletion Formation. *Trends in Genetics* 35 (3), 235–244. ISSN 0168-9525. <https://doi.org/10.1016/j.tig.2019.01.001>.
- Onyango, I.G., Tuttle, J.B., Bennett Jr., J.P., 2005a. Activation of p38 and N-acetyl-cysteine-sensitive c-Jun NH2-terminal kinase signaling cascades is required for induction of apoptosis in Parkinson's disease cybrids. *Mol. Cell. Neurosci.* 28 (3), 452–461.
- Onyango, I.G., Tuttle, J.B., Bennett Jr., J.P., 2005b. Brain-derived growth factor and glial cell line-derived growth factor use distinct intracellular signaling pathways to protect PD cybrids from H2O2-induced neuronal death. *Neurobiol. Dis.* 20 (1), 141–154.
- Otaegui, D., Paisan, C., Saenz, A., Marti, I., Ribate, M., Marti-Masso, J., Perez-Tur, J., de Munain, A.L., 2004. Mitochondrial polymorphisms in Parkinson's disease. *Neurosci. Lett.* 370 (2–3), 171–174.
- Ozawa, T., Tanaka, M., Ikebe, S.-i., Ohno, K., Kondo, T., Mizuno, Y., 1990. Quantitative determination of deleted mitochondrial DNA relative to normal DNA in parkinsonian striatum by a kinetic PCR analysis. *Biochem. Biophys. Res. Commun.* 172 (2), 483–489.
- Ozawa, T., Tanaka, M., Ino, H., Ohno, K., Sano, T., Wada, Y., Yoneda, M., Tanno, Y., Miyatake, T., Tanaka, T., 1991. Distinct clustering of point mutations in mitochondrial DNA among patients with mitochondrial encephalomyopathies and with Parkinson's disease. *Biochem. Biophys. Res. Commun.* 176 (2), 938–946.
- Pardo-Seco, J., Amigo, J., González-Manteiga, W., Salas, A., 2013. A generalized model to estimate the statistical power in mitochondrial disease studies involving 2 × k tables. *PLoS One* 8 (9), e73567.
- Pareek, C.S., Smoczynski, R., Tretyn, A., 2011. Sequencing technologies and genome sequencing. *J. Appl. Genet.* 52 (4), 413–435.
- Park, J.-S., Davis, R.L., Sue, C.M., 2018. Mitochondrial dysfunction in Parkinson's disease: new mechanistic insights and therapeutic perspectives. *Curr. Neurol. Neurosci. Rep.* 18 (5), 21.
- Parker Jr., W.D., Boyson, S.J., Parks, J.K., 1989. Abnormalities of the electron transport chain in idiopathic Parkinson's disease. *Ann. Neurol. Off. J. Am. Neurol. Assoc. Child Neurol. Soc.* 26 (6), 719–723.
- Peeva, V., Blei, D., Trombly, G., Corsi, S., Szukszo, M.J., Rebelo-Guimar, P., Gammage, P.A., Kudin, A.P., Becker, C., Altmüller, J., 2018. Linear mitochondrial DNA is rapidly degraded by components of the replication machinery. *Nat. Commun.* 9.
- Pienaar, I.S., Howell, N., Elson, J.L., 2017. MutPred mutational load analysis shows mildly deleterious mitochondrial DNA variants are not more prevalent in Alzheimer's patients, but may be under-represented in healthy older individuals. *Mitochondrion* 34, 141–146.
- Pyle, A., Poltynic, T., Tiangyou, W., Lambert, C., Keers, S.M., Allcock, L.M., Davison, J., Lewis, S.J., Perry, R.H., Barker, R., 2005. Mitochondrial DNA haplogroup cluster UKJ7 reduces the risk of PD. *Ann. Neurol.* 57 (4), 564–567.
- Pyle, A., Burn, D.J., Gordon, C., Swan, C., Chinnery, P.F., Baudouin, S.V., 2010. Fall in circulating mononuclear cell mitochondrial DNA content in human sepsis. *Intensive Care Med.* 36 (6), 956–962.
- Pyle, A., Brennan, R., Kurzawa-Akanbi, M., Yarnall, A., Thouin, A., Mollenhauer, B., Burn, D., Chinnery, P.F., Hudson, G., 2015. Reduced cerebrospinal fluid mitochondrial DNA is a biomarker for early-stage Parkinson's disease. *Ann. Neurol.* 78 (6), 1000–1004.
- Pyle, A., Anurgha, H., Kurzawa-Akanbi, M., Yarnall, A., Burn, D., Hudson, G., 2016. Reduced mitochondrial DNA copy number is a biomarker of Parkinson's disease. *Neurobiol. Aging* 38, 216. e7–216. e10.
- Quiros, P.M., Mottis, A., Auwerx, J., 2016. Mitonuclear communication in homeostasis and stress. *Nat. Rev. Mol. Cell Biol.* 17 (4), 213.
- Raule, N., Sevin, F., Santoro, A., Altília, S., Franceschi, C., 2007. Association studies on human mitochondrial DNA: methodological aspects and results in the most common

- age-related diseases. *Mitochondrion* 7 (1), 29–38.
- Reeve, A.K., Krishnan, K.J., Elson, J.L., Morris, C.M., Bender, A., Lightowlers, R.N., Turnbull, D.M., 2008. Nature of mitochondrial DNA deletions in substantia nigra neurons. *Am. J. Hum. Genet.* 82 (1), 228–235.
- Richter, G., Sonnenschein, A., Grünewald, T., Reichmann, H., Janetzky, B., 2002. Novel mitochondrial DNA mutations in Parkinson's disease. *J. Neural Transm.* 109 (5–6), 721–729.
- Ross, O.A., McCormack, R., Maxwell, L.D., Duguid, R.A., Quinn, D.J., Barnett, Y.A., Rea, I.M., El-Agnaf, O.M., Gibson, J.M., Wallace, A., 2003. mt4216C variant in linkage with the mtDNA TJ cluster may confer a susceptibility to mitochondrial dysfunction resulting in an increased risk of Parkinson's disease in the Irish. *Exp. Gerontol.* 38 (4), 397–405.
- Salas, A., Elson, J.L., 2015. Mitochondrial DNA as a risk factor for false positives in case-control association studies. *J. Genet. Genomics* 42 (4), 169–172.
- Salas, A., Fachal, L., Marcos-Alonso, S., Vega, A., Martín-Torres, F., ESIGEM, G.d.i., 2009. Investigating the role of mitochondrial haplogroups in genetic predisposition to meningococcal disease. *PLoS One* 4 (12), e8347.
- Samuels, D.C., Carothers, A.D., Horton, R., Chinnery, P.F., 2006. The power to detect disease associations with mitochondrial DNA haplogroups. *Am. J. Hum. Genet.* 78 (4), 713–720.
- Samuels, D.C., Li, C., Li, B., Song, Z., Torstenson, E., Clay, H.B., Rokas, A., Thornton-Wells, T.A., Moore, J.H., Hughes, T.M., 2013. Recurrent tissue-specific mtDNA mutations are common in humans. *PLoS Genet.* 9 (11), e1003929.
- Schapira, A., Cooper, J., Dexter, D., Jenner, P., Clark, J., Marsden, C., 1989. Mitochondrial complex I deficiency in Parkinson's disease. *The Lancet* 333 (8649), 1269.
- Sheehan, J., Swerdlow, R., Parker, W., Miller, S., Davis, R., Tuttle, J., 1997. Altered calcium homeostasis in cells transformed by mitochondria from individuals with Parkinson's disease. *J. Neurochem.* 68 (3), 1221–1233.
- Shoffner, J.M., Watts, R.L., Juncos, J.L., Torroni, A., Wallace, D.C., 1991. Mitochondrial oxidative phosphorylation defects in Parkinson's disease. *Ann. Neurol.* 30 (3), 332–339.
- Shokolenko, I., Venediktova, N., Bochkareva, A., Wilson, G., Alexeyev, M., 2009. Oxidative stress induces degradation of mitochondrial DNA. *Nucleic Acids Res.* 37 (8), 2539.
- Simon, D., Mayeux, R., Marder, K., Kowall, N., Beal, M., Johns, D., 2000. Mitochondrial DNA mutations in complex I and tRNA genes in Parkinson's disease. *Neurology* 54 (3), 703–703.
- Simon, D.K., Lin, M.T., Zheng, L., Liu, G.-J., Ahn, C.H., Kim, L.M., Mauck, W.M., Twu, F., Beal, M.F., Johns, D.R., 2004. Somatic mitochondrial DNA mutations in cortex and substantia nigra in aging and Parkinson's disease. *Neurobiol. Aging* 25 (1), 71–81.
- Simon, D.K., Matott, J.C., Espinosa, J., Abraham, N.A., 2017. Mitochondrial DNA mutations in Parkinson's disease brain. *Acta Neuropathol. Commun.* 5 (1), 33.
- Smigrodzki, R., Parks, J., Parker, W.D., 2004. High frequency of mitochondrial complex I mutations in Parkinson's disease and aging. *Neurobiol. Aging* 25 (10), 1273–1281.
- Song, S., Pursell, Z.F., Copeland, W.C., Longley, M.J., Kunkel, T.A., Mathews, C.K., 2005. DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity. *Proc. Natl. Acad. Sci. Unit. States Am.* 102 (14), 4990–4995.
- Sonney, S., Leipzig, J., Lott, M.T., Zhang, S., Procaccio, V., Wallace, D.C., Sondheimer, N., 2017. Predicting the pathogenicity of novel variants in mitochondrial tRNA with MitoTIP. *PLoS Comput. Biol.* 13 (12), e1005867.
- Spillantini, M.G., Schmidt, M.L., Lee, V.M.-Y., Trojanowski, J.Q., Jakes, R., Goedert, M., 1997. α -Synuclein in Lewy bodies. *Nature* 388 (6645), 839.
- Stoneking, M., 2000. Hypervariable sites in the mtDNA control region are mutational hotspots. *Am. J. Hum. Genet.* 67 (4), 1029–1032.
- Strobbe, D., Caporali, L., Iommarini, L., Maresca, A., Montopoli, M., Martinuzzi, A., Achilli, A., Olivieri, A., Torroni, A., Carelli, V., 2018. Haplogroup J mitogenomes are the most sensitive to the pesticide rotenone: relevance for human diseases. *Neurobiol. Dis.* 114, 129–139.
- Suomalainen, A., Isohanni, P., 2010. Mitochondrial DNA depletion syndromes—many genes, common mechanisms. *Neuromuscul. Disord.* 20 (7), 429–437.
- Swerdlow, R.H., 2012. Does mitochondrial DNA play a role in Parkinson's disease? A review of cybrid and other supportive evidence. *Antioxidants Redox Signal.* 16 (9), 950–964.
- Swerdlow, R.H., Parks, J.K., Miller, S.W., Davis, R.E., Tuttle, J.B., Trimmer, P.A., Sheehan, J.P., Bennett, J.P., Parker, W.D., 1996. Origin and functional consequences of the complex I defect in Parkinson's disease. *Ann. Neurol.* 40 (4), 663–671.
- Swerdlow, R.H., Parks, J.K., Davis, J.N., Cassarino, D.S., Trimmer, P.A., Currie, L.J., Dougherty, J., Bridges, W.S., Bennett, J.P., Wooten, G.F., 1998. Matrilineal inheritance of complex I dysfunction in a multigenerational Parkinson's disease family. *Ann. Neurol.* 44 (6), 873–881.
- Torroni, A., Achilli, A., Macaulay, V., Richards, M., Bandelt, H.-J., 2006. Harvesting the fruit of the human mtDNA tree. *Trends Genet.* 22 (6), 339–345.
- Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J.N., Rovio, A.T., Bruder, C.E., Bohlooly-Y, M., Gidlöf, S., Oldfors, A., Wibom, R., 2004. Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* 429 (6990), 417.
- Trifunovic, A., Hansson, A., Wredenberg, A., Rovio, A.T., Dufour, E., Khvorostov, I., Spelbrink, J.N., Wibom, R., Jacobs, H.T., Larsson, N.-G., 2005. Somatic mtDNA mutations cause aging phenotypes without affecting reactive oxygen species production. *Proc. Natl. Acad. Sci. Unit. States Am.* 102 (50), 17993–17998.
- Trimmer, P.A., Swerdlow, R.H., Parks, J.K., Keeney, P., Bennett Jr., J.P., Miller, S.W., Davis, R.E., Parker Jr., W.D., 2000. Abnormal mitochondrial morphology in sporadic Parkinson's and Alzheimer's disease cybrid cell lines. *Exp. Neurol.* 162 (1), 37–50.
- Trimmer, P.A., Borland, M.K., Keeney, P.M., Bennett, J.P., Parker, W.D., 2004. Parkinson's disease transgenic mitochondrial cybrids generate Lewy inclusion bodies. *J. Neurochem.* 88 (4), 800–812.
- Trinh, J., Farrer, M., 2013. Advances in the genetics of Parkinson disease. *Nat. Rev. Neurol.* 9 (8), 445.
- Van Der Walt, J.M., Nicodemus, K.K., Martin, E.R., Scott, W.K., Nance, M.A., Watts, R.L., Hubble, J.P., Haines, J.L., Koller, W.C., Lyons, K., 2003. Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. *Am. J. Hum. Genet.* 72 (4), 804–811.
- Van Oven, M., Kayser, M., 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* 30 (2).
- Veltri, K.L., Espiritu, M., Singh, G., 1990. Distinct genomic copy number in mitochondria of different mammalian organs. *J. Cell. Physiol.* 143 (1), 160–164.
- Venter, M., Malan, L., Van Dyk, E., Elson, J.L., Van der Westhuizen, F.H., 2017. Using MutPred derived mtDNA load scores to evaluate mtDNA variation in hypertension and diabetes in a two-population cohort: the SABPA study. *J. Genet. Genomics* 44 (3), 139–149.
- Venter, M., Malan, L., Elson, J.L., Van der Westhuizen, F.H., 2019. Implementing a new variant load model to investigate the role of mtDNA in oxidative stress and inflammation in a bi-ethnic cohort: the SABPA study. *Mitochondrial DNA Part A* 30 (3), 440–447.
- Vermulst, M., Wanagat, J., Kujoth, G.C., Bielas, J.H., Rabinovitch, P.S., Prolla, T.A., Loeb, L.A., 2008. DNA deletions and clonal mutations drive premature aging in mitochondrial mutator mice. *Nat. Genet.* 40 (4), 392.
- Vivian, C.J., Brinker, A.E., Graw, S., Koestler, D.C., Legendre, C., Gooden, G.C., Salhia, B., Welch, D.R., 2017. Mitochondrial genomic backgrounds affect nuclear DNA methylation and gene expression. *Cancer Res.* 77 (22), 6202–6214.
- Wallace, R.A., Boldry, R., Schmittgen, T., Miller, D., Uretsky, N., 1984. Effect of 1-methyl-4-phenyl-1, 2, 3, 6 tetrahydropyridine (MPTP) on monoamine neurotransmitters in mouse brain & heart. *Life Sci.* 35 (3), 285–291.
- Wei, W., Keogh, M.J., Wilson, L., Coxhead, J., Ryan, S., Rollinson, S., Griffin, H., Kurzawa-Akanbi, M., Santibanez-Koref, M., Talbot, K., 2017. Mitochondrial DNA point mutations and relative copy number in 1363 disease and control human brains. *Acta Neuropathol. Commun.* 5 (1), 13.
- Weissensteiner, H., Forer, L., Fuchsberger, C., Schöpf, B., Kloss-Brandstätter, A., Specht, G., Kronenberg, F., Schönherr, S., 2016. mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Res.* 44 (W1), W64–W69.
- West, A.P., Brodsky, I.E., Rahner, C., Woo, D.K., Erdjument-Bromage, H., Tempst, P., Walsh, M.C., Choi, Y., Shadel, G.S., Ghosh, S., 2011. TLR signalling augments macrophage bactericidal activity through mitochondrial ROS. *Nature* 472 (7344), 476.
- Wilkins, H.M., Carl, S.M., Swerdlow, R.H., 2014. Cytoplasmic hybrid (cybrid) cell lines as a practical model for mitochondrialopathies. *Redox Biol.* 2, 619–631.
- Wu, H.-M., Li, T., Wang, Z.-F., Huang, S.-S., Shao, Z.-Q., Wang, K., Zhong, H.-Q., Chen, S.-F., Zhang, X., Zhu, J.-H., 2018. Mitochondrial DNA variants modulate genetic susceptibility to Parkinson's disease in Han Chinese. *Neurobiol. Dis.* 114, 17–23.
- Yasukawa, Takehiro, September 2018. Dongchon Kang, An overview of mammalian mitochondrial DNA replication mechanisms. *J. Biochem.* 164 (3), 183–193. <https://doi.org/10.1093/jb/mvy058>.
- Ye, K., Lu, J., Ma, F., Keinan, A., Gu, Z., 2014a. Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proc. Natl. Acad. Sci. Unit. States Am.* 111 (29), 10654–10659.
- Ye, F., Samuels, D.C., Clark, T., Guo, Y., 2014b. High-throughput sequencing in mitochondrial DNA research. *Mitochondrion* 17, 157–163.
- Yoshino, H., Nakagawa-Hattori, Y., Kondo, T., Mizuno, Y., 1992. Mitochondrial complex I and II activities of lymphocytes and platelets in Parkinson's disease. *J. Neural Transm.* 4 (1), 27–34.
- Zhang, J., Montine, T., Smith, M., Siedlak, S., Gu, G., Robertson, D., Perry, G., 2002. The mitochondrial common deletion in Parkinson's disease and related movement disorders. *Park. Relat. Disord.* 8 (3), 165–170.
- Zhang, W., Cui, H., Wong, L.-J.C., 2012. Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing. *Clin. Chem.* 58 (9), 1322–1331.