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 FL 32224-1865

REPORT DATE: September 2020

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	**2. REPORT TYPE**	**3. DATES COVERED**
Sep 2020	Annual	15 Aug 2019 - 14 Aug 2020

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 ADDRESS(ES)
Mayo Clinic Jacksonville
4500 San Pablo Road
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**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	15	

Table of Contents

Page

1. **Introduction…………………………………………………………....4**
2. **Keywords………………………………………………………….....4**
3. **Accomplishments………..……………………………………….....4**
4. **Impact…………………………...…………………………………12**
5. **Changes/Problems...….………………………………………………12**
6. **Products, Inventions, Patent Applications, and/or Licenses..……..13**
7. **Participants & Other Collaborating Organizations………………13**
8. **Special Reporting Requirements……………………………........15**
9. **Appendices…………………………………………………………....**

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). In year 3, we currently have whole-genome sequencing for 95 patients with early-onset disease and we have a further 55 patient samples being processed. We are still focused on single nucleotide variation and copy number changes but are also focusing on non-coding variation. We are currently applying programs to look for repeat expansions and are assessing the variability across callers (ExpansionHunter, gangSTR and STRetch). The hypothesis is that non-coding repeat expansions could knockout or drastically reduce expression of a given allele and this may be another process by which recessive loss-of-function of a gene could result in early-onset forms of PD.

In Year 1-2 we focused on single nucleotide variation and copy number changes across our series of people with early-onset Parkinson's disease (EOPD). In year 3 given the COVID-19 pandemic research had slowed but we are beginning to implement a further analysis tool to the samples and data and look at non-coding DNA and possible expanded repeat sequences as a cause for loss of expression and phenotypic presentation of disease. In addition we have applied for access to a whole-genome sequence (WGS) data that is available through a number of consortia efforts including the Accelerating Medicines Partnership (AMP) for Parkinson's (PD; AMP-PD) datasets and Lewy body dementia sequencing cohorts.

Single nucleotide variation in 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* (VPS13Cp.Ala1687Val and VPS13Cp.Leu216 Val); one individual carries the published PD risk variant LRRK2p.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	*GBA*p.Leu483Pro het
1622-1	F	35	White	Y	*SYNJ1*p.Arg1328Pro het; SYNJ1p.Val1405_Leu1406insAsnThr het *PARK7*p. Val20ala het
3021-1	M	33	White	Y	*GBA*p.Leu483Pro het
3021-10	F	50	White	Y	*GBA*p.Leu483Pro het
5541-1	F	34	White	Y	*GBA*p.Leu483Pro het
6634-1	M	37	White	N	*PINK1*p.Val184Met het *GBA*p.Asp448His het
6636-1	F	43	White		*PRKN*p.Met192Leu het
6743-1	F	44	White		*VPS13C*p.Ala1687Val het; *VPS13C*p.Leu216Val het
552-1	F	39	White	Y	*PINK1*p.Gln115Leu het
707-3	M	34	White	Y	*PINK1*p.Gln115Leu het
1200-1	F	40	White	Y	*GBA*p.Glu365Lys het; *GBA*p.Asp179His het
1256-1	M	43	White	N	*VPS13C*p.Arg470His het
1263-1	M	40	White	AD	*PRKN*p.Ser167Asn het *VPS13C*p.Ile2789Thr het; *VPS13C*p.Ile1132Val het; *VPS13C*p.Arg153His het
1308-1	M	41	White	Y	*GBA*p.Asn409Ser

					*PRKN*p.Ser167Asn homozygous
1560-1	F	37	White	Y	*GBA*p.Glu365Lys het
5167-1	F	37	White	N	*PINK1*p.Gln115Leu het
5274-1	F	37	White	Y	*GBA*p.Arg202Ter het *PINK1*p.Gly411Ser het
5311-1	M	36	White	No	*PRKN*p.Arg256Cys het
5400-1	M	34	White	Y	*VPS13C* splicing c.684+7T>C het
5536-1	F	40	White	N	*GBA*p.Ala483Arg het
5779-1	M	45	White	Y	*GBA*p.Ala495Pro het ; *GBA*p.483Pro het
5891-1	F	41	White	N	*GBA*p.Glu365Lys het
5980-1	M	36	White	Y	*PRKN*p.Arg256Cys het
6179-1	M	38	White	Y	*PRKN*p.Arg275Gln het
6375-1	M	45	White	Y	*GBA*p.Glu365Lys het
6560-1	F	36	White	Y	*LRRK2*p.Gly2019Ser het
7200-1	F	47	White	N	*PINK1*p.Gln115Leu het

Novel and rare single nucleotide 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	*BCAT2*	p.Tyr200Ter	37	<0.01
1487-1	F	40	*CNTN6*	p.Glu924Ter	49	
1651-2	F	43	*CPNE5*	p.Glu266Ter	45	
1487-1	F	40	*DOCK6*	p.Gln1518Ter	38	
1951-1	M	29	*FMO4*	p.Lys42Ter	38	<0.01
597-1	M	43	*IFIH1*	p.Glu627Ter	38	<0.01
1202-1	M	39	*MMP21*	p.Leu546Ter	39	
5136-1	F	34	*NMNAT3*	p.Trp184Ter	41	<0.01
5704-1	M	40	*PCDHB7*	p.Glu632Ter	42	<0.01
767-3	M	35	*PLSCR5*	p.Arg129Ter	42	<0.01
5948-1	M	41	*SPAG1*	p.Gln672Ter	38	<0.01
1487-1	F	40	*TAOK1*	p.Arg698Ter	39	
1951-1	M	29	*TMEM45A*	p.Arg36Ter	37	<0.01
1695-1	F	38	*TSFM*	p.Leu207Ter	39	<0.01
5619-1	F	37	*ULK4*	p.Arg862Ter	52	<0.01

Novel and rare single nucleotide 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	*CSMD3*	p.Pro1233Leu	33	na
6323-1	M	45	*FOXO6*	p.Lys190Thr	32	na
5167-1	F	37	*GMPR2*	p.Gly201Val	35	<0.01
552-1	F	39	*KCNT1*	p.Arg910Gln	35	<0.01
5676-1	M	41	*RPS27A*	p.Ser57Pro	33	<0.01
5255-1	M	40	*TOMM5*	p.Pro9Arg	33	<0.01
5980-1	M	36	*TRPM3*	p.Ser453Leu	33	<0.01

Novel or rare single nucleotide 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	*TRPM2*	p.Leu1034Phe	26.2	<0.01
1308-1	M	41	*TRPM2*	p.Gln953Ter	58	<0.01
6636-1	F	43	*TRPM2*	p.Arg411Trp	26.7	<0.01
1200-1	F	40	*TRAP1*	p.Arg47Ter	35	<0.01
5256-1	F	35	*SLC25A13*	p.Gln312Ter	35	<0.01
5834-1	M	48	*MFN2*	p.Thr362Met	26.4	<0.01
1776-1	F	43	*USP30*	p.Ser171Leu	26.3	<0.01
5430-1	F	43	*ATG13*	p.Asp490Asn	26.8	<0.01
5676-1	M	41	*CLEC16A*	p.Arg860Cys	25.9	<0.01
6179-1	M	38	*TOMM40*	p.Gly24Glu	23.1	<0.01
6026-1	F	42	*VPS13D*	p.Ala171Gly	24.6	<0.01
1064-1	M	39	*VPS13D*	p.Arg3267Trp	27.3	<0.01

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 in figure 1, is predicted to have 2 main regulatory regions in mice dopaminergic neurons, and 10 cis-regulatory regions according to ENCODE.

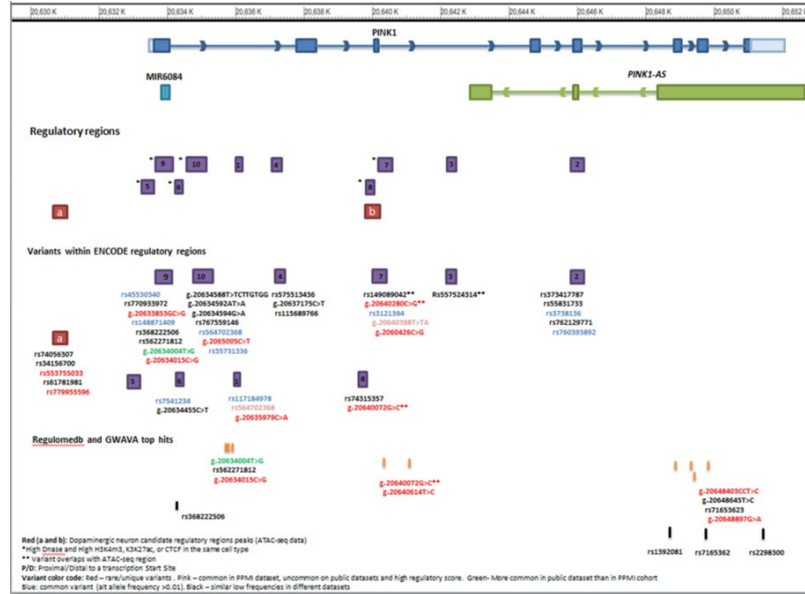


Figure 1. Regulatory regions and variation in the *PINK1*

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.

PINK1

- From the 48 "Regulatory" variants, 21 have HaploReg outputs

Query SNP: rs45530340 and variants with r^2 >= 0.8

chr	pos (hg38)	LD (r^2)	LD (D')	variant	Ref	Alt	AFR freq	AMR freq	ASN freq	EUR freq	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	NHGRI/EBI GWAS hits	GRASP QTL hits	Selected eQTL hits	GENCODE genes	dbSNP func annot
1	20626439	0.99	1	rs12035659	A	C	0.05	0.19	0.12	0.18		ESC	ESC, IPSC			7 altered motifs			9 hits	7kb 5' of PINK1	
1	20627365	0.89	1	rs72650843	T	G	0.05	0.19	0.12	0.19			4 tissues			NRSF,VDR			9 hits	6.1kb 5' of PINK1	
1	20628707	1	1	rs56350927	G	A	0.04	0.19	0.12	0.18						Lhx3			10 hits	4.7kb 5' of PINK1	
1	20629132	0.96	0.98	rs55671019	C	A	0.05	0.19	0.12	0.18						10 altered motifs			8 hits	4.3kb 5' of PINK1	
1	20633737	1	1	rs45530340	C	T	0.05	0.19	0.12	0.18		24 tissues		26 tissues	SIN3AK20,KAP1,ETS1				10 hits	PINK1	synonymous
1	20640679	1	1	rs72650846	G	A	0.05	0.19	0.12	0.18			ESC, BRN, MUS						9 hits	PINK1	intronic
1	20645418	0.92	0.99	rs56125389	G	C	0.01	0.17	0.12	0.17						CTCF,HNF4			8 hits	PINK1	intronic
1	20645811	0.92	0.99	rs67251778	15-mer	G	0.01	0.18	0.12	0.17						8 altered motifs			8 hits	PINK1	intronic
1	20645880	0.98	0.99	rs141853877	CG	C	0.01	0.18	0.12	0.18						6 altered motifs			1 hit	PINK1	intronic
1	20647293	0.85	0.94	rs111354972	G	A	0.01	0.18	0.12	0.18						4 altered motifs			9 hits	PINK1	intronic
1	20648309	0.86	0.93	rs2298299	A	G	0.01	0.18	0.12	0.18									11 hits	PINK1	intronic
1	20651106	0.85	0.93	rs8064	C	G	0.01	0.18	0.12	0.18			LIV		POL2,POL24H8	Maf,Msx-1,VDR		2 hits	9 hits	PINK1	3'-UTR
1	20651622	0.83	0.93	rs17406145	G	A	0.01	0.17	0.12	0.17						EBF,Ik-2			8 hits	PINK1-AS	
1	20658861	0.84	0.92	rs200300978	T	TC	0.01	0.18	0.12	0.18			9 tissues			6 altered motifs			8 hits	DDOST	intronic
1	20658862	0.8	0.91	rs55897342	T	C	0.01	0.18	0.12	0.18			9 tissues			7 altered motifs			7 hits	DDOST	intronic
1	20664780	0.82	0.92	rs56070322	G	A	0.05	0.18	0.12	0.17						7 altered motifs			8 hits	KIF17	intronic

Example of Haploreg output

Figure 2. Regulatory variation in the *PINK1*

Other resources such as Haploreg (see Figure 2), GWAVA, BRAINEAC, GTEx, ORegAnno, Enhancer Atlas, SYNAPSE, FANTOM and ConSite were used to identify known eQTLs and fine tune the top candidate variants.

Proteins bound in ChIP-Seq experiments (ENCODE Project Consortium, 2011)

Cell ID	Protein
GM12878	SIN3AK20
HEK293(b)	KAP1
K562	ETS1

Rs45530340 (0.2 allele freq) L63L

Hits from selected eQTL studies

Study ID	Paper Title	PMID	Tissue	Correlated gene	p-value
GTEx2015_v6	The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans	25954001	Adipose_Subcutaneous	CDA	1.01341696006182e-07
GTEx2015_v6	The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans	25954001	Adrenal_Gland	CDA	4.12713791475528e-07
GTEx2015_v6	The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans	25954001	Breast_Mammary_Tissue	CDA	9.17601897702916e-07
GTEx2015_v6	The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans	25954001	Cells_Transformed_fibroblasts	CDA	3.81623844673574e-11
GTEx2015_v6	The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans	25954001	Heart_Left_Ventricle	CDA	4.70848047242169e-06
GTEx2015_v6	The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans	25954001	Muscle_Skeletal	CDA	7.90801846252552e-10
GTEx2015_v6	The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans	25954001	Nerve_Tibial	CDA	3.15598604104796e-07
GTEx2015_v6	The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans	25954001	Skin_Sun_Exposed_Lower_leg	PINK1	7.60486251355997e-08
GTEx2015_v6	The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans	25954001	Thyroid	CDA	5.99015945711464e-10
Lappalainen2013	Transcriptome and genome sequencing uncovers functional variation in humans	24037378	Lymphoblastoid_EUR_exonlevel	ENSG00000158828.5_20975488_20975724	1.16249078848474e-05

E071	Brain	BRN.HIPP.MID	Brain Hippocampus Middle	1_TssA	1_TssA	H3K4me1_Enh	H3K4me3_Pro	H3K27ac_Enh		
E074	Brain	BRN.SUB.NIG	Brain Substantia Nigra	1_TssA	1_TssA		H3K4me3_Pro	H3K27ac_Enh	H3K9ac_Pro	
E068	Brain	BRN.ANT.CAUD	Brain Anterior Caudate	1_TssA	1_TssA		H3K4me3_Pro	H3K27ac_Enh	H3K9ac_Pro	
E069	Brain	BRN.CING.GYR	Brain Cingulate Gyrus	1_TssA	1_TssA		H3K4me3_Pro	H3K27ac_Enh	H3K9ac_Pro	
E072	Brain	BRN.INF.TMP	Brain Inferior Temporal Lobe	1_TssA	1_TssA		H3K4me3_Pro	H3K27ac_Enh	H3K9ac_Pro	
E067	Brain	BRN.ANG.GYR	Brain Angular Gyrus	1_TssA	1_TssA	H3K4me1_Enh	H3K4me3_Pro	H3K27ac_Enh	H3K9ac_Pro	
E073	Brain	BRN.DL.PRFRNTL.CRTX	Brain_Dorsolateral_Prefrontal_Cortex	1_TssA	1_TssA	H3K4me1_Enh	H3K4me3_Pro	H3K27ac_Enh	H3K9ac_Pro	
E070	Brain	BRN.GRM.MTRX	Brain Germinal Matrix	1_TssA	1_TssA		H3K4me3_Pro			
E082	Brain	BRN.FET.F	Fetal Brain Female	1_TssA	1_TssA		H3K4me3_Pro			
E081	Brain	BRN.FET.M	Fetal Brain Male		1_TssA	H3K4me1_Enh	H3K4me3_Pro			DNase

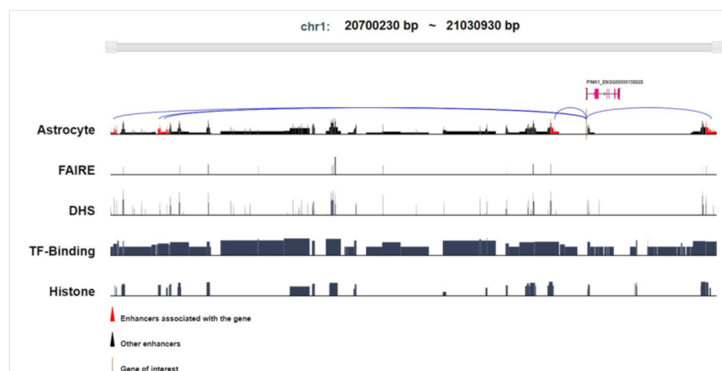
Figure 3. Tissue-specific regulatory variation in the *PINK1* gene

We have been using expression quantitative trait loci (eQTL) databases to help with the annotation across tissues. For example the *PINK1* variant, rs45530340 appears to regulate *PINK1* expression but only in sun exposed skin (Figure 3). We had identified this rare variant in a patient who harbored another single heterozygous *PINK1* mutation.

Another approach we have been using to use programs to predict functional regulatory motifs and this includes transcription factor binding, microRNA binding sites and distant gene expression enhancers. We are using cell-specific derived expression data to identify potentially functional non-coding variation that may act in a cell-specific manner. We have used enhancer atlas (http://www.enhanceratlas.org/) to predict cell-specific enhancers.

Figure 4. Distant expression enhancer for the *PINK1* gene

PINK1 astrocyte enhancer prediction



No enhancers associated with *PARK2* in astrocytes this dataset

Source: Enhancer Atlas

The importance of non-coding variation is becoming clearer but is more difficult to functionally test and given the cell-specific nature of the regulation of transcript expression has a higher degree of complexity. In addition to the non-coding variation we are also performing complementary studies to try to resolve cell-specific transcript expression that may highlight coding variation that is only observed in a specific-cell type.

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	*DFFB MEGF6 TP73 KIAA0562 ARHGEF16 WDR8 KIAA0495 LRRC47 PRDM16 TPRG1L ACTRT2 CCDC27 C1orf174*	NA
chr2:96823988-97075783	251,795	Deletion	*CNNM3 ANKRD39 FAM178B SEMA4C ANKRD23 LOC643445 LOC100128957 LOC100132375*	NA
chr3:14444330-14561629	117,299	Deletion	*GRIP2*	NA
chr6:157691300-164676192	6,984,892	Duplication/ Translocation	*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 ….*	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	*DNAJC30*	NA
chr21:14338129-16526867	2,188,738	Duplication	*HSPA13 NRIP1 USP25 VDAC2P RBM11 POLR2CP SAMSN1 LIPI ABCC13 CYCSP42 LOC388813 C21orf126 LOC100128341….*	CNVR7955.1 CNVR7956_full

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. We have also run CNV callers through the early-onset patients within the PPMI cohort to support our CNV studies. The figure shows the CNVs identified at the well-established EOPD locus, PRKN, across the PPMI patients and controls as well as the Mayo clinic Biobank controls.

In 2020, we have applied for access to the recently released Accelerating Medicines Partnership (AMP) for Parkinson's (PD; AMP-PD) datasets that includes the PPMI series mentioned above. This will give us access to over 1800 whole-genome sequences from people with PD and no known genetic cause. In addition, we have recently participated in the Lewy body dementia sequencing consortium which has given us access to over 750 whole genome sequences from pathologically-confirmed Lewy body disease from cases collected at the Mayo Clinic brain bank. These resources will allow us to screen genetic variation identified in our early-onset cases to rapidly assess the role of variants/genes in the more frequent late-onset forms of Parkinson's disease and Lewy body disorders that we identify in our studies.

Given the discussion on whether early-onset forms of disease are mitochondrial-centered or if synuclein pathology is still key these additional datasets will allow us to potentially screen novel early-onset genes against related neuropathologies to assess protein aggregation and mitochondrial dysfunction.

Repeat sequence variation in people with early-onset parkisnonism

A number of repeat sequences have been identified to cause neurodegenerative disease for example Huntington disease, the C9orf72 expanded hexanucleotide repeat is the most frequent genetic cause of frontotemporal dementia and amyotrophic lateral sclerosis (and has been implicated in PD), and the group of disorders known as the spinocerebellar ataxias (SCA) are frequently caused by DNA repeat sequences. Indeed, the clinical presentation for a number of SCA genes includes parkinsonism. As part of repeat analysis we screened our EOPD patients for the most common SCA repeats that can results in parkinsonism (SCA1, SCA2 and SCA3) but we are now using this data as an internal validation to apply genome-wide programs that will identify expanded repeats from WGS data. We are testing two programs to begin with, the first is from Illumina termed ExpansionHunter (Dolzhenko et al., 2020) and the second is GangSTR (Mousavi et al., 2019). We will filter our data against the same control series as we have used for the CNV analysis and the newly acquired AMP-PD cohorts.

- **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 3 of the award we published 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 (Tipton et al. (2020) and Wernick et al. (2020)). Deutschlander et al. (2020) investigated genetic factors that may result in an earlier onset of disease. Four other related papers were also published on the role of mtDNA in PSP/CBD, non-MAPT related genes on chr17 haplotype in sporadic PD and overlap in genetic forms of ataxia and MSA (an early-onset parkinsonism disorder). Early in the first quarter of 2020 I also presented some of the EOPD data at the annual Mayo Clinic APDA PD research symposium for patients and at the 2020 Symposium on Alzheimer's Disease and Related Dementias, at the Alvin A. Dubin Alzheimer's Resource Center in Ft. Myers, Florida.

- **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**

The COVID-19 pandemic has shut the lab down or reduced productivity for almost six months and we are slowly and carefully ramping back up following Mayo Clinic approved guidelines that include and promote social distancing, mask wearing and protection for staff. We have implemented a shift system (morning and afternoon) to minimize the number of people in the lab at any time and when work can be performed from home that option is promoted.

Ana Kolicheski, PhD (Research Fellow) on the project has taken a position at Washington University (November 2019) and I will need to replace her with another fellow. I have an active search to find a suitable candidate, however there is a hire freeze at Mayo Clinic due to the COVID-19 pandemic which is preventing recruiting a new Research Fellow at this time.

In addition, due to the COVID-19 pandemic the Genomics Core has also been closed ad not accepting samples for projects, this has slowed our data generation but we hope this will rapidly move forward over the coming months.

- **Changes that had a significant impact on expenditures**
  Given the COVID-19 pandemic and the needed acquisition of a research fellow spending was reduced over Year 3, however a no-cost extension has been approved and will allow for continued effort and spending on the grant.

- **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:	*Owen A. Ross, PhD*
Project Role:	*PI*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	*2.0*
Contribution to Project:	*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*
Funding Support:	
Name:	*Alexandra Soto, MSc*
Project Role:	*Laboratory Technician*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	*2.0*
Contribution to	*Ms. Soto oversaw additional screening of the EOPD patients for*

Project:	*atypical forms of early-onset parkinsonism and is establishing the new protocol for analysis of long-read whole genome sequencing.*
Funding Support:	
Name:	Yingxue Ren, PhD
Project Role:	Bioinformatician
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	2.0
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:	
Name:	
Project Role:	
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	
Contribution to Project:	
Funding Support:	
Name:	
Project Role:	
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	
Contribution to Project:	
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:
  <u>New</u>
  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: (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

  <u>Ended</u>

- **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

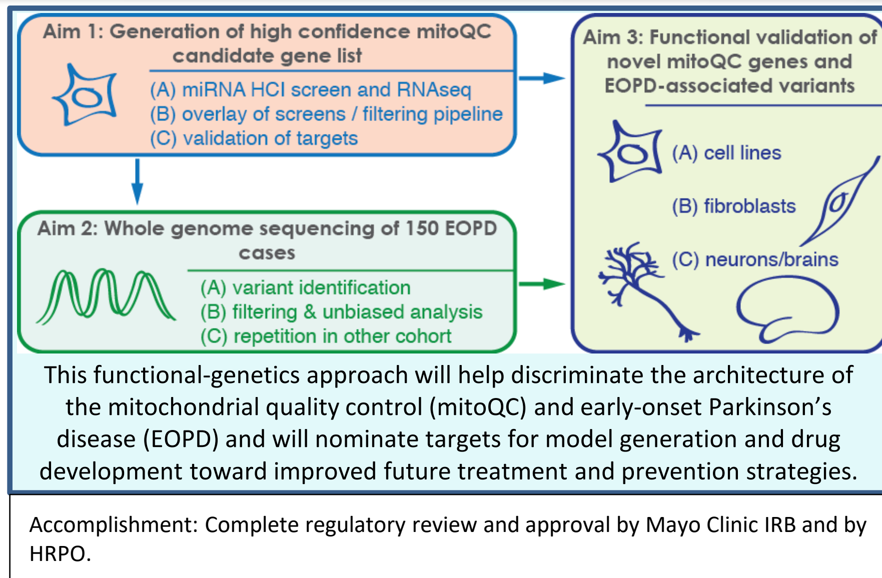
PI: Owen A. Ross, PhD **Org:** Mayo Clinic Jacksonville **Award Amount:** $1,176,917

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**

Updated: (09/12/2020)

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**

- Dr. Kolicheski (Fellow) left for a position at WashU.
- COVID-19 Pandemic closed the lab and core for much of Y3Q3+4
- 12month NCE approved

Budget Expenditure to Date

Actual Expenditure: $579,370