

**FHS PUBLIC ACCESS**

Author manuscript

JAMA Neurol. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

JAMA Neurol. 2016 January ; 73(1): 68–75. doi:10.1001/jamaneurol.2015.3266.**Whole-Exome Sequencing in Familial Parkinson Disease**

Janice L. Farlow, PhD, Laurie A. Robak, MD, PhD, Kurt Hetrick, MS, Kevin Bowling, PhD, Eric Boerwinkle, PhD, Zeynep H. Coban-Akdemir, PhD, Tomasz Gambin, PhD, Richard A. Gibbs, PhD, Shen Gu, PhD, Preti Jain, PhD, Joseph Jankovic, MD, Shalini Jhangiani, MS, Kaveeta Kaw, BS, Dongbing Lai, PhD, Hai Lin, BS, Hua Ling, PhD, Yunlong Liu, PhD, James R. Lupski, MD, PhD, DSc, Donna Muzny, MS, Paula Porter, MPhil, Elizabeth Pugh, PhD, Janson White, BS, Kimberly Doheny, PhD, Richard M. Myers, PhD, Joshua M. Shulman, MD, PhD, and Tatiana Foroud, PhD

Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis (Farlow, Lai, Liu, Foroud); Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas (Robak, Coban-Akdemir, Gambin, Gu, Kaw, Lupski, Porter, White, Shulman); Department of Pediatrics, Baylor College of Medicine, Houston, Texas (Robak, Lupski); Department of Pediatrics, Texas Children's Hospital, Houston (Robak, Lupski); Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston (Robak, Kaw, Porter, Shulman); Center for Inherited Disease Research, The Johns Hopkins University, Baltimore, Maryland (Hetrick, Ling, Pugh, Doheny); HudsonAlpha Institute for Biotechnology, Huntsville, Alabama (Bowling, Jain, Myers); Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas (Boerwinkle, Gibbs, Jhangiani, Lupski, Muzny); Human Genetics Center, University of Texas Health Science Center, Houston (Boerwinkle); Department of Pediatrics, Columbia University Medical Center, New York, New York (Jain); Department of Neurology, Baylor College of Medicine, Houston, Texas (Jankovic, Shulman); Department of BioHealth Informatics, Indiana University School of Informatics and Computing, Indianapolis (Lin); Department of Neuroscience, Baylor College of Medicine, Houston, Texas (Shulman)

Abstract

IMPORTANCE—Parkinson disease (PD) is a progressive neurodegenerative disease for which susceptibility is linked to genetic and environmental risk factors.

OBJECTIVE—To identify genetic variants contributing to disease risk in familial PD.

DESIGN, SETTING, AND PARTICIPANTS—A 2-stage study design that included a discovery cohort of families with PD and a replication cohort of familial probands was used. In the discovery cohort, rare exonic variants that segregated in multiple affected individuals in a family

Corresponding Author: Tatiana Foroud, PhD, Department of Medical and Molecular Genetics, Indiana University School of Medicine, 410 W 10th St, Health Information and Translational Sciences, Ste 4000, Indianapolis, IN 46202, (tforoud@iu.edu).

Author Contributions: Drs Shulman and Foroud are co–senior authors and had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Farlow, Robak, Liu, Shulman, Foroud.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Farlow, Robak, Shulman, Foroud.

Critical revision of the manuscript for important intellectual content: All authors.

Conflict of Interest Disclosures: None reported.

and were predicted to be conserved or damaging were retained. Genes with retained variants were prioritized if expressed in the brain and located within PD-relevant pathways. Genes in which prioritized variants were observed in at least 4 families were selected as candidate genes for replication in the replication cohort. The setting was among individuals with familial PD enrolled from academic movement disorder specialty clinics across the United States. All participants had a family history of PD.

MAIN OUTCOMES AND MEASURES—Identification of genes containing rare, likely deleterious, genetic variants in individuals with familial PD using a 2-stage exome sequencing study design.

RESULTS—The 93 individuals from 32 families in the discovery cohort (49.5% [46 of 93] female) had a mean (SD) age at onset of 61.8 (10.0) years. The 49 individuals with familial PD in the replication cohort (32.6% [16 of 49] female) had a mean (SD) age at onset of 50.1 (15.7) years. Discovery cohort recruitment dates were 1999 to 2009, and replication cohort recruitment dates were 2003 to 2014. Data analysis dates were 2011 to 2015. Three genes containing a total of 13 rare and potentially damaging variants were prioritized in the discovery cohort. Two of these genes (*TNK2* and *TNR*) also had rare variants that were predicted to be damaging in the replication cohort. All 9 variants identified in the 2 replicated genes in 12 families across the discovery and replication cohorts were confirmed via Sanger sequencing.

CONCLUSIONS AND RELEVANCE—*TNK2* and *TNR* harbored rare, likely deleterious, variants in individuals having familial PD, with similar findings in an independent cohort. To our knowledge, these genes have not been previously associated with PD, although they have been linked to critical neuronal functions. Further studies are required to confirm a potential role for these genes in the pathogenesis of PD.

Parkinson disease (PD) is a progressive neurodegenerative disease for which susceptibility is linked to genetic and environmental risk factors. Linkage studies have previously identified rare mutations responsible for PD in large, multiplex families,^{1–5} and whole-exome sequencing (WES) has been used successfully in such pedigrees more recently.^{6,7} It is recognized that PD is genetically heterogeneous, and many additional genes remain to be discovered, particularly in families with strong disease aggregation. Most important, several genes initially identified in familial PD have subsequently been demonstrated to have substantial contributions to sporadic PD without known family history (eg, *LRRK2* G2019S [OMIM 609007], *GBA* N307A [OMIM 606463], and the occurrence of rare and common variant alleles at *SNCA* [OMIM 163890]). Therefore, elucidation of rare alleles with strong effects on disease risk in families can have important implications for our understanding of the genetic architecture of PD in the general population.

Whole-exome sequencing yields more than 20 000 exonic single-nucleotide variants (SNVs) per individual,⁸ requiring a strategy to narrow the number of variants. In pedigrees characterized by potential autosomal dominant inheritance, filtering strategies based on segregation have facilitated identification of causal variants. However, this approach requires large, multigenerational pedigrees with available genetic samples and clinical characterization. By contrast, small to moderately sized families with PD are less informative for segregation analyses, leaving many variants after using bioinformatic filters

to predict damaging alleles. Even in well-defined cases of mendelian PD (with *SNCA* or *LRRK2*), there are instances in which not all affected family members carry the mutation (ie, intrafamilial heterogeneity).^{3,9,10} As recently suggested in amyotrophic lateral sclerosis,¹¹ oligogenic inheritance (in which multiple rare alleles contribute to individual risk) may also have a role in PD susceptibility. Therefore, some of the successful analytic strategies developed for simple mendelian disorders may need to be adapted for continued successful gene discovery in complex genetic disorders such as PD.

To date, WES in PD has been reported in studies involving one or a few families^{6,7,12–15} or in candidate gene investigations.¹⁶ We applied an innovative 2-stage study designed to address some challenges that are inherent in gene discovery in common, complex disorders. The first stage used exomes from a discovery cohort of multiplex families with PD. Each family was examined for segregating candidate variants, allowing for intrafamilial heterogeneity. We then prioritized genes identified across multiple families, allowing for allelic heterogeneity. In the second stage, we analyzed the most promising genes in an independent replication cohort of sequenced probands with familial PD. We identified the subset of candidate genes containing rare, potentially functional variants that may contribute to disease risk (Figure 1).

Methods

Discovery Cohort Participants

The study protocol was approved by the Indiana University Institutional Review Board and by the ethics boards of all study sites. Families with at least 1 pair of living siblings diagnosed as having PD were evaluated by Parkinson Study Group movement disorder neurologists. Written informed consent was obtained from all participants. Validated checklists^{17,18} were used to assign clinical diagnosis of PD. Individuals classified as having verified PD met United Kingdom Parkinson's Disease Brain Bank criteria,¹⁹ modified only to allow for positive family history. Individuals having no signs of a movement disorder were considered to have no evidence of PD. The remaining individuals were classified as having nonverified PD. These individuals had evidence of a movement disorder but failed to meet all inclusion criteria or met at least 1 exclusion criterion. All brain autopsies completed at the time of sequencing (n = 5) confirmed the diagnosis of PD. Peripheral blood was obtained from all individuals who provided written informed consent. Clinical evaluations or biospecimens from unaffected family members were not available as part of this study.

Discovery Samples

Whole-exome sequencing and annotation of identified variants (eMethods in the Supplement) were performed in 32 families with the largest number of verified PD cases without another segregating neurological disorder and without a known causative PD mutation in *LRRK2* or *PARKIN* (OMIM 602544). Among the 32 families, 90 individuals with verified PD underwent sequencing. An additional 3 individuals initially classified as having nonverified PD were also included as affected cases. Two had neuropathological confirmation of PD. The third met all clinical inclusion criteria (including onset after age 20 years, bradykinesia, persistent asymmetry, and diagnosis by a movement disorders

neurologist) and had significant supporting criteria (including rigidity, postural instability, a resting tremor, disease progression, and a positive response to levodopa) but met the sole exclusion criterion of having concomitant Alzheimer disease and sensory deficits. Of the 32 families (eFigures 1, 2, 3, and 4 in the Supplement), 6 had 2 cases sequenced, 23 had 3 cases sequenced, and 3 had 4 cases sequenced.

Filtering

Variants were retained (Figure 1) if they (1) were predicted to be SNVs or insertions/deletions (indels) in an exonic or splicing region based on 1 or more gene databases (eMethods in the Supplement), (2) had an allele frequency of less than 3% in European American populations in the annotated public and internal frequency databases, (3) were predicted to be damaging by at least 1 in silico protein functional and structural effect prediction program (eMethods in the Supplement) or were located in a highly conserved region, and (4) segregated with at least 2 PD cases in the same family. Genes were retained if they (1) were in a relevant Gene Ontology (GO) category,^{20,21} (2) were expressed in the brain, and (3) had retained variants that were observed in at least 2 cases in at least 4 families.

Replication Cohort and Variant Confirmation

The prioritized genes were examined in WES (eMethods in the Supplement) from a replication cohort of 49 unrelated individuals with familial PD. All individuals were diagnosed as having PD based on examination by movement disorders neurologists and reported at least 1 first-degree relative diagnosed as having PD. The study was approved by the Baylor College of Medicine Institutional Review Board and written informed consent was obtained from all participants. Compared with the discovery pipeline, a more stringent allele frequency filter (<1% in European American populations from 1000 Genomes and ESP) was used.^{22,23} Potentially deleterious and highly conserved variants were identified using SIFT, PolyPhen-2, MutPred, and GERP^{24–27} Variants present in genes prioritized from the discovery analysis were extracted.

All variants in replicated genes identified in the discovery and replication cohorts were reviewed in the Exome Aggregation Consortium (ExAC) (<http://exac.broadinstitute.org>) and were confirmed using targeted polymerase chain reaction and Sanger sequencing. Variants were annotated for Combined Annotation Dependent Depletion (CADD) (<http://cadd.gs.washington.edu>),²⁸ in which C-scores of at least 10 and at least 20 correspond to the 10% and 1% most deleterious substitutions in the genome, respectively. Genes were annotated for residual variation intolerance score (RVIS) percentiles, in which lower percentiles correspond to genes that are most intolerant of functional mutations.²⁹

Simulation

Simulated WES data sets using ExAC allele frequencies were generated and interrogated with the identical discovery and replication filtering pipeline described above, except for the use of segregation within families (eMethods in the Supplement) because the ExAC database consists of unrelated individuals and therefore does not contain data regarding segregation. A simulation was performed with data sets from 150 000 randomly chosen SNVs to parallel

the number of variants identified in our families with PD, and a more conservative simulation was conducted with 250 000 SNVs to establish a range of *P* values.

Results

Discovery Cohort

Clinical characteristics of 93 individuals from the 32 families in the discovery cohort are summarized in Table 1. Pedigrees and sequencing quality control metrics are provided in eFigures 1, 2, 3, and 4 and the eMethods in the Supplement.

Application of the Genome Analysis Toolkit³⁰ (<https://www.broadinstitute.org/gatk/>) quality filters resulted in 149 055 SNVs and 9378 indels across all samples (22 188–28 230 variants per sample) (Figure 1). Nonsynonymous SNVs or indels within an exon having an allelic frequency of less than 3% were retained. After removing variants that were predicted to be benign by all 4 protein prediction programs (eMethods in the Supplement) and were not in a highly conserved region, approximately 10% of the original variants remained. We next considered each family independently and filtered based on segregation, requiring that candidate variants must be shared by at least 2 affected individuals, allowing for allelic heterogeneity within a family, as has been seen with other genes associated with PD.^{3,9,10} Single-nucleotide variants identified in the filtered genes with GO annotation are listed in eTable 1 in the Supplement.

We performed an integrated analysis across our discovery sample to identify genes with damaging alleles in at least 4 families. Because many variants remained (eTable 1 in the Supplement), we further restricted the list of prioritized genes based on established gene expression in the brain and a priori participation in biologic categories strongly implicated in PD. This strategy identified 3 genes (*TNK2* [OMIM 606994], *TNR* [OMIM 601995], and *TOPORS* [OMIM 609507]), each of which had filtered variants observed in both the Center for Inherited Disease Research and HudsonAlpha Institute for Biotechnology sets of families.

Replication Cohort

Clinical characteristics of the 49 probands with familial PD in the replication cohort are summarized in Table 1. Sequencing metrics are described in the eMethods in the Supplement.

Rare variants predicted to be damaging in the 3 genes prioritized in the discovery analysis were extracted from the replication WES data. Two genes (*TNK2* and *TNR*) that harbored variants of interest (Replication Cohort and Variant Confirmation subsection of the Methods section) in the discovery cohort were also found to have distinct variants of interest in the replication cohort (Table 2).

In total, the 2 genes were observed to harbor 9 distinct potentially functionally relevant variants (Table 3). All 9 variants were confirmed by targeted polymerase chain reaction and Sanger sequencing in all relevant samples (eTable 2 in the Supplement). Genic intolerance RVIS percentiles²⁹ and CADD C-scores²⁸ were computed to characterize the potential effect

of functional mutations at the gene level and variant level, respectively. The 2 genes had a mean (SD) RVIS of -0.63 (0.13) and a mean (SD) percentile of 17.2% (4.1%). The calculated RVIS percentiles reflect purifying selection or probable greater intolerance for mutations within the gene than most genes. For example, *TNK2* has an RVIS of -0.72 and a percentile of 14.3%, placing it among the 14.3% most intolerant of genes. In contrast, genes known to cause autosomal dominant PD (ie, *SNCA*, *LRRK2*, and *VPS35* [OMIM 601501]) have a mean (SD) RVIS of -0.65 (0.50) and a mean (SD) percentile of 21.5% (0.2%).

The mean (SD) CADD C-score for the 9 variants was 23.9 (6.4). All 9 variants are predicted to be within the 1% most deleterious variants in the genome, with the exceptions of the *TNK2* pA977V variant and the *TNR* pT166A variant, both of which still fall within the predicted 10% most deleterious variants.

To estimate the significance of our findings, we examined many simulated discovery and replication data sets and determined the likelihood of similar observations by chance. After application of the filtering pipeline to the 10 000 data sets produced from 150 000 randomly chosen ExAC SNVs, one data set yielded 2 genes ($P < .001$), and 163 data sets yielded 1 gene each ($P = .02$). To provide a more conservative estimate of statistical likelihood, 10 000 data sets were then simulated from 250 000 randomly chosen ExAC SNVs. Of the 10 000 data sets, 9 yielded 2 genes ($P < .001$), and 340 yielded 1 gene ($P = .03$) after applying the filtering pipeline.

Discussion

Using WES in discovery and replication cohorts of individuals with familial PD, we detected 9 likely deleterious, rare exonic variants in 2 genes (*TNK2* and *TNR*) that may have a role in PD susceptibility (Table 3). All variants were heterozygous, consistent with dominant inheritance and the pedigree structures (Figure 2), suggesting that the disease phenotype results from a gain-of-function, haploinsufficiency, or dominant-negative mechanism. The candidate variants identified may contribute to PD risk in 12 families from our study, including one family in which variants in both genes are cosegregating. To our knowledge, neither *TNK2* nor *TNR* has previously been implicated in PD susceptibility from genetic investigation of large pedigrees or from genome-wide association studies in population cohorts.

TNK2 encodes for a nonreceptor tyrosine kinase (activated *CDC42* kinase 1) that is important for cell growth, survival, and migration. Findings from some studies^{31–33} suggest that *TNK2* is involved in synaptic function and plasticity, and the results of a recent study³⁴ propose that *TNK2* mutations may cause autosomal recessive infantile-onset epilepsy. Other studies^{35,36} have established links between the *TNK2* protein and the epidermal growth factor receptor (*EGFR* [OMIM 131550]). In the discovery and replication cohorts, 4 unique rare nonsynonymous *TNK2* variants were identified. Of the 5 families in the discovery cohort that had 2 or more members who shared a candidate *TNK2* variant, 2 families showed complete segregation (ie, all sequenced members of the family carried the variant of interest) (Figure 2). None of the family members of the probands from the replication cohort could be assessed for inheritance of variants of interest. One of the *TNK2* variants identified

in this study (pV363A) is found in the *EGFR* inhibitor Mig-6 domain (IPR021619 and PF11555). Binding of Mig-6 to the kinase domain of *EGFR* inactivates the receptor, which suggests that this domain in the *TNK2* protein may also be important for appropriate regulation of its function.

TNR, or tenascin R, encodes an extracellular matrix glycoprotein that is only found in the central nervous system.³⁷ Tenascin R is thought to be involved in neurite growth, neural cell adhesion, and sodium channel functioning.^{38,39} Of the 6 unique variants prioritized in *TNR*, 5 were found only in the discovery cohort as rare nonsynonymous variants. One variant (pR578X) was found solely in the replication cohort and results in addition of a stop site at position 578 of a 1358 amino acid protein. This variant, along with one other variant (pT592A), is found in the fibronectin type 3 domain (IPR003961) of the protein, which is important for cell surface binding.

Given the prevalence, late onset, and incomplete penetrance of PD, we expect that PD susceptibility alleles are likely observed at low frequencies within public databases. Therefore, we used a conservative 3% minor allele frequency filter in the discovery phase to retain these variants. In the replication phase, which uses familial samples but does not have the advantage of examining allele sharing within each family, we used a more stringent 1% minor allele frequency filter to gather additional evidence for the genes nominated from the discovery analysis. The final variants identified (Table 3) were not observed (n = 2) or were observed at low frequencies (n = 7) in the ExAC database. This finding is consistent with the expectation that causative alleles may still be observed within public databases, especially given that the ExAC database includes individuals as young as 18 years, at which age clinical manifestations of PD are unlikely.

Unlike previous studies^{6,7,13,40} focused on a single large pedigree or on extensive data sets of unrelated individuals, our blended approach leveraged a well-characterized set of moderately sized families and an additional set of unrelated familial probands. Also, a major advantage of this study is that both cohorts included only individuals with familial PD. Families with multiple affected members are more likely to be enriched for causative, moderately rare variants having a modest or large effect size.

Our experimental design contrasts with recent efforts that sequenced large pedigrees to identify variants with fully penetrant effects responsible for strictly mendelian PD. This category of variants appears to account for rare causes of PD,^{6,7,13} and heterogeneity has been observed even in these families.^{3,16,17} Our study design allows for detection of such mutations but also permits discovery of rare variants with intermediate penetrance such as *LRRK2* G2019S⁴¹ and mutations in *GBA* (OMIM 606463).⁴² Because 10% to 20% of patients with PD report having at least 1 first-degree relative affected by PD,^{43–45} it is likely that variants of intermediate penetrance remain a major contributor to PD heritability. Experimental designs that allow for exploration of intrafamilial and interfamilial heterogeneity are particularly important for studying PD. To limit false-positive results that are inherent in a prioritization scheme that requires incomplete allelic segregation in families, we used a 2-stage study design to increase the likelihood that the candidate genes identified in this study are involved in the etiology of PD.

One pedigree from our study (family C in Figure 2) illustrates possible intrafamilial heterogeneity involving cosegregation of variants in both *TNK2* and *TNR*. All sequenced family members carry at least 1 of 2 variants (*TNK2* pR877H and *TNR* pT592A). Individual 4 is heterozygous for both variants and had a younger age at onset than his siblings (49 vs 64 years), each of whom carries only a single variant in either gene. In other pedigrees (families P and AD in Figure 2), the cause of PD in individuals not carrying the identified variant may also be due to another unidentified gene or environmental insult. These pedigrees illustrate possible intrafamily oligogenic inheritance or phenocopies that would likely be missed by family-based sequencing study designs based on monogenic and completely penetrant inheritance models.

It is notable that the 2 genes implicated by our studies, *TNK2* and *TNR*, were each more than 1000 amino acids in length, raising potential concern that they would be discovered to harbor rare damaging variation by chance because of their large size. However, using simulated data sets generated from ExAC allele frequencies, we determined that it was statistically unlikely to identify 1 gene ($P = .02$ to $P = .03$) or 2 genes ($P < .001$) using our filtering pipeline by chance. The sizes of genes identified in simulations ranged from 463 to 3144 amino acids, and many genes with modest size were also captured. Families are not included in ExAC data, so our simulations were by nature more conservative. In addition, although our discovery sample yielded 150 000 SNVs, we also simulated data sets using 250 000 randomly chosen SNVs for a more conservative estimate of statistical likelihood. Therefore, we conclude that *TNK2* and *TNR* are unlikely to be chance findings but rather are due to enrichment of rare functional variants associated with PD. Nevertheless, it will be important for these genes to be examined in additional replication samples to definitively establish their potential contributions to PD risk.

Exome sequencing by design misses possibly important variation in intronic and regulatory regions. The use of the GO filter to focus on pathways of interest might have excluded important genes that were poorly annotated or were in pathways thus far not associated with PD. The GO filter (Figure 1) narrowed the number of variants under consideration from 6635 to 228 SNVs, ultimately prioritizing 13 SNVs (3 genes) for further study. Had the GO filter not been applied, the 6635 SNVs would have been narrowed only to 300 SNVs (87 genes) using the across-families filter. Future studies with larger sample sizes could use formal gene set enrichment analysis to bypass the potential limitation of relying on prespecified pathways for variant filtering.

Conclusions

We used a 2-stage strategy to identify and replicate genes that may harbor rare variants contributing to PD susceptibility. Both the discovery and replication samples were composed of individuals with familial PD, who may be more likely to segregate rare variants of larger effect on disease risk. The 2 genes nominated in this study (*TNK2* and *TNR*) warrant further evaluation for their potential role in the pathogenesis of PD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding/Support: Funding for this study was provided by grants R01NS037167, 1X01HG006236, C06RR029965, and U54HG006542 (Dr Lupski) and grant U54HG003273 (Dr Gibbs) from the National Institutes of Health. Sequencing services were provided by the Center for Inherited Disease Research, which is funded by federal contract number HHSN268201100011I from the National Institutes of Health to The Johns Hopkins University. Dr Farlow was supported by grant T32 GM077229 from the Indiana University Medical Scientist Training Program and by grant TLI TR000162 from the National Institutes of Health, National Center for Advancing Translational Sciences Clinical and Translational Sciences. Dr Robak was supported by medical genetics training grant T32 GM07526-37 from the Baylor College of Medicine. Drs Bowling and Myers were supported by the HudsonAlpha Institute for Biotechnology for a portion of this work. Dr Jankovic was supported by The Michael J. Fox Foundation for Parkinson's Research and by the National Parkinson Foundation. Dr Shulman was supported by the Caroline Weiss Law Fund for Research in Molecular Medicine, by grant R21NS089854 from the National Institutes of Health, and by a Burroughs Wellcome Fund Career Award for Medical Scientists.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: Emily Young, BA, and Weidong Le, MD, PhD (Baylor College of Medicine) assisted with sample preparation and handling. No compensation was provided. We thank all the families who participated in this research. We also thank the Parkinson Study Group-PROGENI (Parkinson's Research: The Organized Genetics Initiative) investigators and coordinators (listed at the end of the Supplement).

REFERENCES

1. Bonifati V, Rizzu P, van Baren MJ, et al. Mutations in the *DJ-1* gene associated with autosomal recessive early-onset parkinsonism. *Science*. 2003; 299(5604):256–259. [PubMed: 12446870]
2. Zimprich A, Biskup S, Leitner P, et al. Mutations in *LRRK2* cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron*. 2004; 44(4):601–607. [PubMed: 15541309]
3. Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*. 1997; 276(5321):2045–2047. [PubMed: 9197268]
4. Kitada T, Asakawa S, Hattori N, et al. Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. *Nature*. 1998; 392(6676):605–608. [PubMed: 9560156]
5. Valente EM, Abou-Sleiman PM, Caputo V, et al. Hereditary early-onset Parkinson's disease caused by mutations in *PINK1*. *Science*. 2004; 304(5674):1158–1160. [PubMed: 15087508]
6. Vilariño-Güell C, Wider C, Ross OA, et al. *VPS35* mutations in Parkinson disease [published correction appears in *Am J Hum Genet*. 2011;89(2):347]. *Am J Hum Genet*. 2011; 89(1):162–167. [PubMed: 21763482]
7. Zimprich A, Benet-Pagès A, Struhlar W, et al. A mutation in *VPS35* encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *Am J Hum Genet*. 2011; 89(1):168–175. [PubMed: 21763483]
8. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 2011; 12(11):745–755. [PubMed: 21946919]
9. Nichols WC, Pankratz N, Hernandez D, et al. Parkinson Study Group-PROGENI Investigators. Genetic screening for a single common *LRRK2* mutation in familial Parkinson's disease. *Lancet*. 2005; 365(9457):410–412. [PubMed: 15680455]
10. Klein C, Chuang R, Marras C, Lang AE. The curious case of phenocopies in families with genetic Parkinson's disease. *Mov Disord*. 2011; 26(10):1793–1802. [PubMed: 21735483]
11. Cady J, Allred P, Bali T, et al. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. *Ann Neurol*. 2015; 77(1):100–113. [PubMed: 25382069]

12. Neri M, Bovolenta M, Scotton C, et al. Whole exome sequencing filtered by novel candidate genes as tool for gene discovery in a recessive family with Parkinson and ataxia. *Neuromuscul Disord*. 2012; 22(9–10):810–810.
13. Vilariño-Güell C, Rajput A, Milnerwood AJ, et al. *DNAJC13* mutations in Parkinson disease. *Hum Mol Genet*. 2014; 23(7):1794–1801. [PubMed: 24218364]
14. Edvardson S, Cinnamon Y, Ta-Shma A, et al. A deleterious mutation in *DNAJC6* encoding the neuronal-specific clathrin-uncoating co-chaperone auxilin, is associated with juvenile parkinsonism. *PLoS One*. 2012; 7(5):e36458. [PubMed: 22563501]
15. Köro lu Ç, Baysal L, Cetinkaya M, Karasoy H, Tolun A. *DNAJC6* is responsible for juvenile parkinsonism with phenotypic variability. *Parkinsonism Relat Disord*. 2013; 19(3):320–324. [PubMed: 23211418]
16. Nuytemans K, Bademci G, Inchausti V, et al. Whole exome sequencing of rare variants in *EIF4G1* and *VPS35* in Parkinson disease. *Neurology*. 2013; 80(11):982–989. [PubMed: 23408866]
17. Nichols WC, Pankratz N, Uniacke SK, et al. Linkage stratification and mutation analysis at the *parkin* locus identifies mutation positive Parkinson's disease families. *J Med Genet*. 2002; 39(7):489–492. [PubMed: 12114481]
18. Pankratz N, Nichols WC, Uniacke SK, et al. Parkinson Study Group. Genome screen to identify susceptibility genes for Parkinson disease in a sample without *parkin* mutations. *Am J Hum Genet*. 2002; 71(1):124–135. [PubMed: 12058349]
19. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry*. 1992; 55(3):181–184. [PubMed: 1564476]
20. Ashburner M, Ball CA, Blake JA, et al. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000; 25(1):25–29. [PubMed: 10802651]
21. Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res*. 2015; 43(database issue):D1049–D1056. [PubMed: 25428369]
22. 1000 Genomes. A deep catalog of human genetic variation. [Accessed October 20, 2015] <http://www.1000genomes.org/about>.
23. National Heart, Lung, and Blood Institute. NHLBI Grand Opportunity Exome Sequencing Project (ESP). [Accessed October 20, 2015] <https://esp.gs.washington.edu/drupal/>.
24. J. Craig Venter Institute. SIFT (Sorting Intolerant From Tolerant). [Accessed October 20, 2015] <http://sift.jcvi.org/>.
25. Prediction of Functional Effects of Human nsSNPS. PolyPhen-2 (Polymorphism Phenotyping v2). [Accessed October 20, 2015] <http://genetics.bwh.harvard.edu/pph2/>.
26. Mutation Tasting. MutTaster (Mutation Taster). [Accessed October 20, 2015] <http://www.mutationtaster.org/>.
27. Sidow Lab. GERP (Genome Evolutionary Rate Profiling). [Accessed October 20, 2015] <http://mendel.stanford.edu/SidowLab/downloads/gerp/>.
28. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014; 46(3):310–315. [PubMed: 24487276]
29. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet*. 2013; 9(8):e1003709. [PubMed: 23990802]
30. Broad Institute. Genome Analysis Toolkit (GATK). [Accessed October 20, 2015] <https://www.broadinstitute.org/gatk/>.
31. La Torre A, del Rio JA, Soriano E, Ureña JM. Expression pattern of ACK1 tyrosine kinase during brain development in the mouse. *Gene Expr Patterns*. 2006; 6(8):886–892. [PubMed: 16750431]
32. Ureña JM, La Torre A, Martínez A, et al. Expression, synaptic localization, and developmental regulation of Ack1/Pyk1, a cytoplasmic tyrosine kinase highly expressed in the developing and adult brain. *J Comp Neurol*. 2005; 490(2):119–132. [PubMed: 16052498]
33. La Torre A, del Mar Masdeu M, Cotrufo T, et al. A role for the tyrosine kinase ACK1 in neurotrophin signaling and neuronal extension and branching. *Cell Death Dis*. 2013; 4:e602. [PubMed: 23598414]

34. Hitomi Y, Heinzen EL, Donatello S, et al. Mutations in *TNFK2* in severe autosomal recessive infantile onset epilepsy. *Ann Neurol*. 2013; 74(3):496–501. [PubMed: 23686771]
35. Howlin J, Rosenkvist J, Andersson T. *TNFK2* preserves epidermal growth factor receptor expression on the cell surface and enhances migration and invasion of human breast cancer cells. *Breast Cancer Res*. 2008; 10(2):R36. [PubMed: 18435854]
36. Galisteo ML, Yang Y, Ureña J, Schlessinger J. Activation of the nonreceptor protein tyrosine kinase Ack by multiple extracellular stimuli. *Proc Natl Acad Sci U S A*. 2006; 103(26):9796–9801. [PubMed: 16777958]
37. Pesheva P, Spiess E, Schachner M. J1-160 and J1-180 are oligodendrocyte-secreted nonpermissive substrates for cell adhesion. *J Cell Biol*. 1989; 109(4, pt 1):1765–1778. [PubMed: 2477380]
38. Jakovcevski I, Miljkovic D, Schachner M, Andjus PR. Tenascins and inflammation in disorders of the nervous system. *Amino Acids*. 2013; 44(4):1115–1127. [PubMed: 23269478]
39. Anlar B, Gunel-Ozcan A. Tenascin-R: role in the central nervous system. *Int J Biochem Cell Biol*. 2012; 44(9):1385–1389. [PubMed: 22634605]
40. Quadri M, Yang X, Cossu G, et al. An exome study of Parkinson's disease in Sardinia, a Mediterranean genetic isolate. *Neurogenetics*. 2015; 16(1):55–64. [PubMed: 25294124]
41. Healy DG, Falchi M, O'Sullivan SS, et al. International *LRRK2* Consortium. Phenotype, genotype, and worldwide genetic penetrance of *LRRK2*-associated Parkinson's disease: a case-control study. *Lancet Neurol*. 2008; 7(7):583–590. [PubMed: 18539534]
42. Alcalay RN, Dinur T, Quinn T, et al. Comparison of Parkinson risk in Ashkenazi Jewish patients with Gaucher disease and *GBA* heterozygotes. *JAMA Neurol*. 2014; 71(6):752–757. [PubMed: 24756352]
43. McDonnell SK, Schaid DJ, Elbaz A, et al. Complex segregation analysis of Parkinson's disease: the Mayo Clinic Family Study. *Ann Neurol*. 2006; 59(5):788–795. [PubMed: 16634030]
44. Kurz M, Alves G, Aarsland D, Larsen JP. Familial Parkinson's disease: a community-based study. *Eur J Neurol*. 2003; 10(2):159–163. [PubMed: 12603291]
45. Elbaz A, Grigoletto F, Baldereschi M, et al. EUROPARKINSON Study Group. Familial aggregation of Parkinson's disease: a population-based case-control study in Europe. *Neurology*. 1999; 52(9):1876–1882. [PubMed: 10371537]

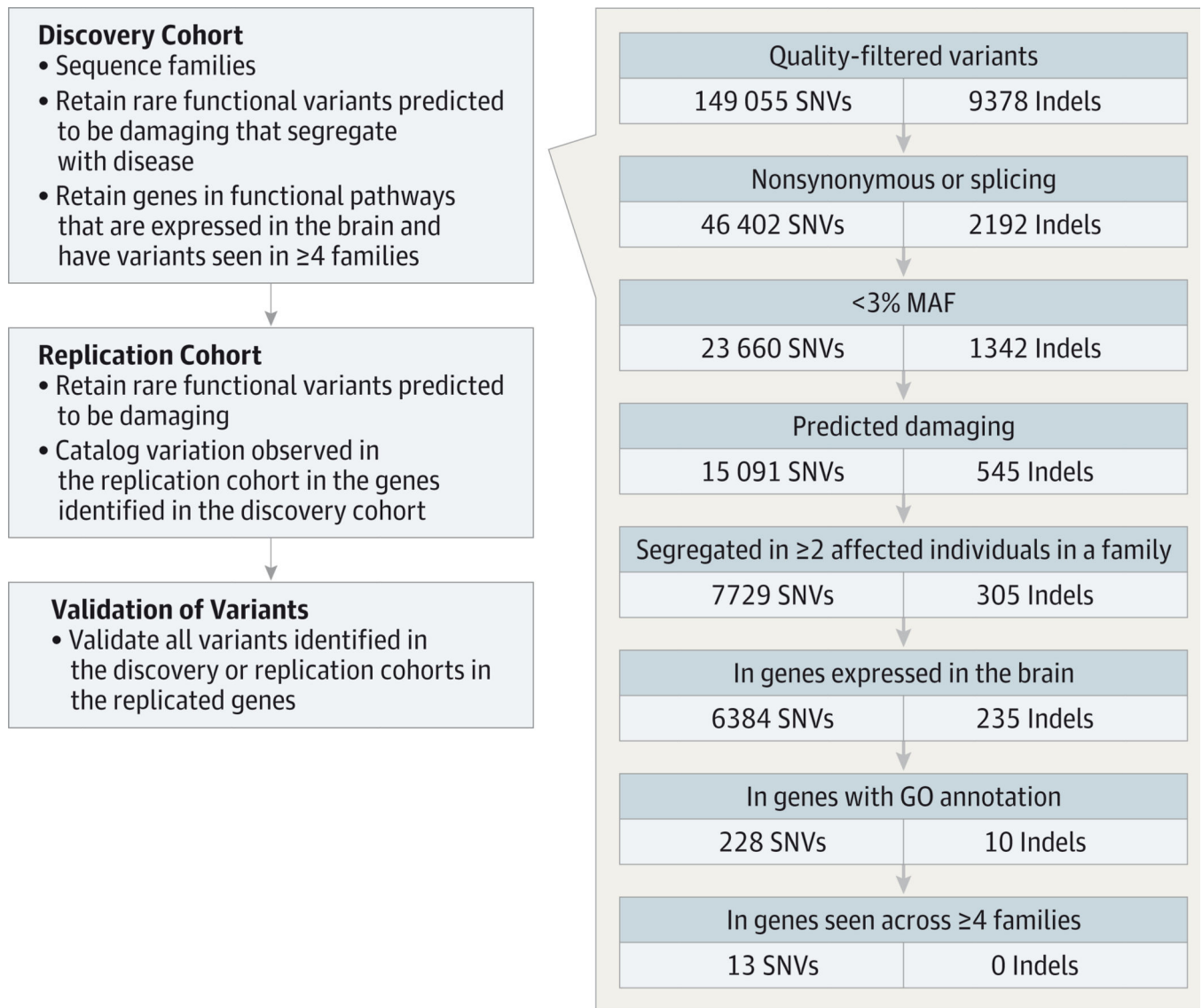


Figure 1. Study Design and Discovery Cohort Variant Filtering

Shown is the overall design for the 2-stage study and variant counts at each stage of filtering. GO indicates Gene Ontology (annotations listed in the Filtering subsection of the Methods section); Indels, insertions or deletions; MAF, minor allele frequency; and SNVs, single-nucleotide variants.

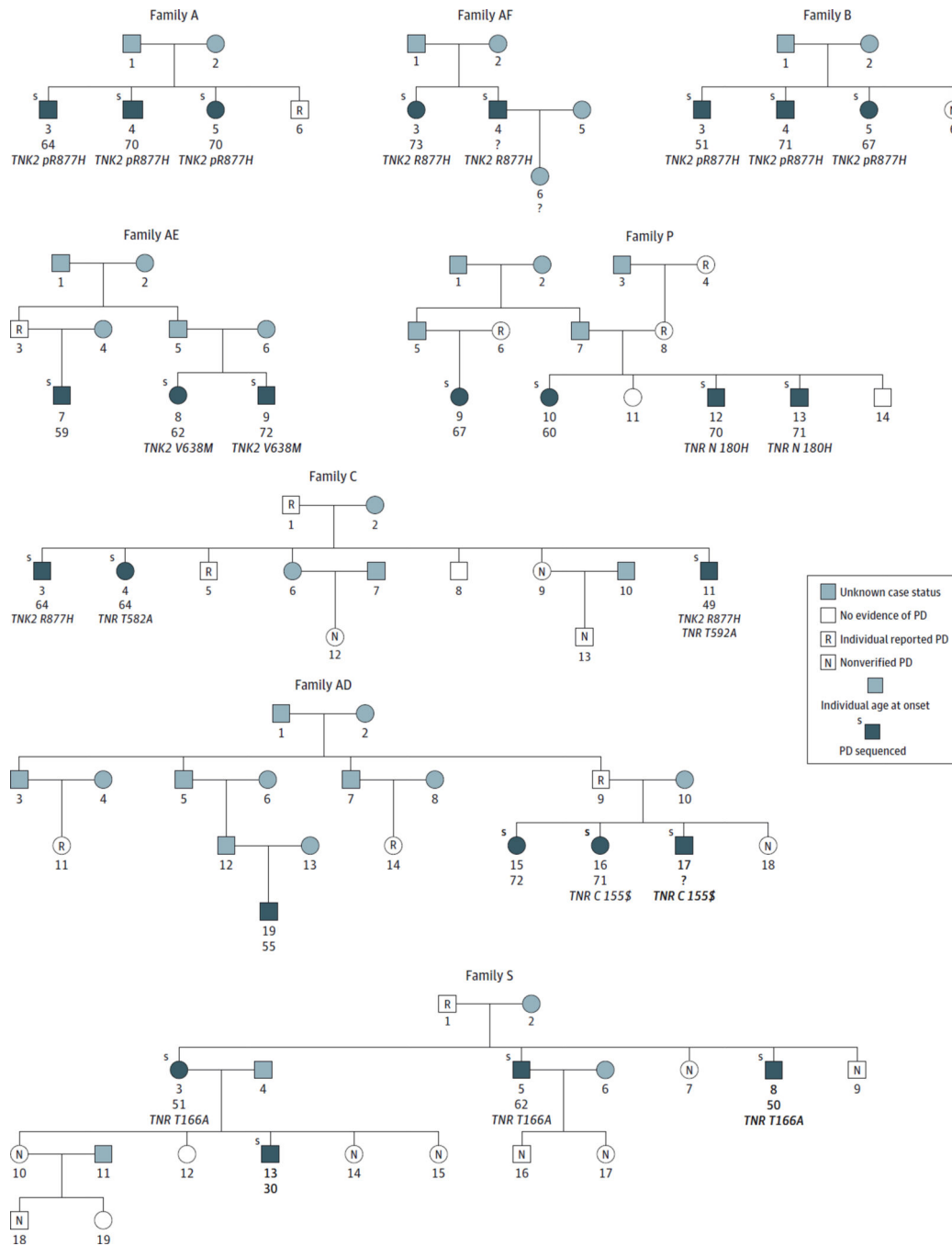


Figure 2. Pedigrees Segregating *TNK2* and *TNR* Variants
 PD indicates Parkinson disease. A square indicates a male individual; a circle, a female individual. Below each symbol is a subject number. For sequenced, affected individuals, the age at disease onset is below the subject number. A question mark indicates an unknown age at onset.

Table 1

Clinical Characteristics of Individuals With Parkinson Disease in the Discovery and Replication Cohorts

Clinical Characteristic	Discovery Cohort (n = 93)	Replication Cohort (n = 49)
No. of families	32	49
Female sex, No. (%)	46 (49.5)	16 (32.6)
Male sex, No. (%)	47 (50.5)	33 (67.3)
Age at onset, mean (SD), y	61.8 (10.0) ^a	50.1 (15.7) ^b
Race/ethnicity		
Non-Hispanic European American	90	37
Hispanic descent	...	8
Non-Hispanic Asian	3	3
Middle Eastern descent	...	1

Ellipses indicate not reported.

^aData not available for 2 of 93 individuals.

^bData not available for 6 of 49 individuals.

Table 2
Parkinson Disease Candidate Genes Identified Through Whole-Exome Sequencing

Gene Symbol	Gene Name	Map Location	Gene Ontology	Transcript Size, Base Pair	Genic Intolerance, RVIS Percentile ^a	No. of Variants/No. of Families in the Discovery Cohort	No. of Variants/Families in the Replication Cohort
<i>TNK2</i>	Tyrosine kinase nonreceptor 2	3q29	Protein ubiquitination	4476	14.3	2/5	2/2
<i>TNR</i>	Tenascin R (restrictin, janusin)	1q24	Neuron-neuron synaptic transmission, regulation of synaptic transmission	5190	20.0	4/4	1/1

Abbreviation: RVIS, residual variation intolerance score.

^a A lower percentile corresponds to more mutation-intolerant genes.

Table 3

Nine Variants Identified in the Candidate Genes

Chromosome	Position	Reference Allele	Alternate Allele	ExAC Total Frequency ^a	Exonic Prediction	Amino Acid Change	CADD C-Score	No. of Discovery Families/ No. of Replication Probands
<i>TNK2</i>								
3	195594092	G	A	5.83e-05	Nonsynonymous SNV	NM_005781:exon13:cC2930T;pA977V	17.8	0/1
3	195594494	C	T	0.015	Nonsynonymous SNV	NM_005781:cG2630A;pR877H	21.7	4/0
3	195595212	C	T	0.0026	Nonsynonymous SNV	NM_005781:exon12:cG1912A;pV638M	29.3	1/0
3	195605390	A	G	2.45e-05	Nonsynonymous SNV	NM_005781:exon8:cT1088C;pY363A	25.9	0/1
<i>TNR</i>								
1	175355171	T	C	NA	Nonsynonymous SNV	NM_003285:exon8:cA1774G;pT592A	24.2	1/0
1	175355213	G	A	NA	Stopgain SNV	NM_003285:exon8:cC1732T;pR578X	36.0	0/1
1	175372714	T	G	0.0043	Nonsynonymous SNV	NM_003285:cA538C;pN180H	24.5	1/0
1	175375355	T	C	0.0044	Nonsynonymous SNV	NM_003285:cA496G;pT166A	13.6	1/0
1	175375388	A	T	7.33e-05	Nonsynonymous SNV	NM_003285:cT463A;pC155S	22.0	1/0

Abbreviations: CADD, Combined Annotation Dependent Depletion; ExAC, Exome Aggregation Consortium; NA, not applicable in the ExAC; SNV, single-nucleotide variant.

^aTotal frequency in Exome Aggregation Consortium.