

# VU Research Portal

Cracking the code-ing sequence for Parkinson's disease

Jansen, I.E.

2017

**document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

***citation for published version (APA)***

Jansen, I. E. (2017). Cracking the code-ing sequence for Parkinson's disease. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

A dense grid of nucleotide sequences (A, C, G, T) arranged in a grid-like pattern. The sequences are organized into vertical columns. Each column contains multiple lines of text, representing a sequence alignment or a multi-sequence comparison. The background is a dark gray color with the text in a light, monospaced font.

# CHAPTER 6

## DISCOVERY AND FUNCTIONAL PRIORITIZATION OF PARKINSON'S DISEASE CANDIDATE GENES FROM LARGE- SCALE WHOLE EXOME SEQUENCING

This chapter has been published as: Jansen IE\*, Ye H\*, Heetveld S\*, Lechler M\*, Michels H, Seinstra RI, Lubbe SJ, Drouet V, Lesage S, Majounie E, Gibbs JR, Nalls MA, Ryten M, Botia JA, Vandrovcova J, Simon-Sanchez J, Castillo-Lizardo M, Rizzu P, Blauwendraat C, Chouhan AK, Li Y, Yogi P, Amin N, Van Duijn CM, International Parkinson's Disease Genetics Consortium, Morris HR, Brice A, Singleton AB, David DC, Nollen EA, Jain S, Shulman JM, Heutink P. Discovery and systematic functional validation of genes from large-scale whole exome sequencing in Parkinson's disease. *Genome Biology*, 2017; 18(1):22.

\* These authors contributed equally to this study

## ABSTRACT

**Background:** Whole-exome sequencing (WES) has been successful in identifying genes that cause familial Parkinson's disease (PD). However, until now this approach has not been deployed to study large cohorts of unrelated subjects. To discover rare PD susceptibility variants, we performed WES in 1,148 unrelated cases and 503 control subjects. Candidate genes were subsequently validated for functions relevant to Parkinson's disease based on parallel RNA-interference (RNAi) screens in human cell culture, and *Drosophila* and *C. elegans* models.

**Results:** Assuming autosomal recessive inheritance, we identify 27 genes that have homozygous or compound heterozygous loss-of-function variants in PD cases. Definitive replication and confirmation of these findings were hindered by potential heterogeneity and by the rarity of the implicated alleles. We therefore looked for potential genetic interactions with established PD mechanisms. Following RNAi-mediated knockdown, 15 of the genes modulated mitochondrial dynamics in human neuronal cultures, and 4 candidates enhanced  $\alpha$ -synuclein induced neurodegeneration in *Drosophila*. Based on complementary analyses in independent human datasets, 5 functionally validated genes—*GPATCH2L*, *UHRF1BP1L*, *PTPRH*, *ARSB* and *VPS13C*—also showed evidence consistent with genetic replication.

**Conclusions:** By integrating human genetic and functional evidence, we identify several PD susceptibility gene candidates for further investigation. Our approach highlights a powerful experimental strategy with broad applicability for future studies of disorders with complex genetic etiologies.

## INTRODUCTION

Next generation sequencing approaches have recently accelerated the identification of variants responsible for familial Parkinson's disease (PD).<sup>1-4</sup> While a positive family history is common in PD, large, multigenerational pedigrees, especially with available DNA and clinical evaluations remain exceptional, hindering progress in unraveling the genetic underpinnings. Importantly, several genes initially discovered to cause PD in families, such as *LRRK2*, *GBA* and *PARK2/parkin*, were subsequently discovered with surprisingly high frequency in "sporadic" PD cohorts.<sup>5,6</sup> To date, large population samples of individuals with PD have primarily contributed to the discovery of common variant susceptibility loci, based on genome-wide association studies of case/control cohorts.<sup>7</sup> The variants identified by genome-wide association studies (GWAS) have modest effect sizes and collectively fail to account for current estimates of PD heritability.<sup>8,9</sup> Considering the above, it seems likely that additional less common alleles, with larger effect sizes contribute to PD risk in the population, and next generation sequencing is one promising approach to identify such alleles. Despite recent successes in other neurodegenerative disease with complex genetic etiologies, including Alzheimer's disease<sup>10-12</sup> and amyotrophic lateral sclerosis,<sup>13,14</sup> sequencing has yet to be deployed in large, unrelated PD case/control samples for rare variant discovery.

The successful discovery of rare variant risk alleles in population-based PD samples faces a number of potential challenges. Perhaps most importantly, analyses of rare variants in large family pedigrees is greatly facilitated by segregation analysis which is not possible in cohorts of unrelated individuals, leading to an increased number of candidate variants to consider. Assumptions of a recessive inheritance model and the application of stringent filters, such as consideration of only strongly damaging, loss-of-function (LoF) variants, is one potential solution, but this is likely to miss many important variants, including dominantly-acting alleles. Further, PD is characterized by extensive genic and allelic heterogeneity and extremely large cohorts may be required to document sufficient numbers of cases to facilitate meaningful statistical comparisons.<sup>15</sup> Lastly, as PD is (i) common (~1-3% prevalence), (ii) strongly age-dependent, and (iii) often preceded by a prolonged presymptomatic or minimally symptomatic phase, we may expect to find truly pathogenic rare variants, including those with large effect sizes, in "control" cohorts of adults (due to unrecognized or early disease stages with minimal symptoms). Therefore, given the occurrence of rare variants, including potentially damaging variants in most genomes of presumably healthy individuals,<sup>16</sup> it may be difficult to identify genes/variants that truly cause disease. Importantly, recent advances in cellular and animal models, along with improved understanding of PD pathogenesis, enable an integrated approach, in which variant discovery, is coupled with a functional screening pipeline for prioritization of those genes worthy of more intensive study.

In this collaborative study of the International Parkinson's Disease Genomics Consortium (IPDGC), we report the results of whole exome sequencing (WES) in 1,148 PD cases, the largest such cohort examined to date. Consistent with the younger age of PD onset in this cohort, which is often associated with a recessive inheritance,<sup>17-19</sup> and to prioritize candidate genes/variants for initial investigation, our analysis focuses on genes with homozygous or compound heterozygous LoF variants. We further couple the human genetic studies with functional screening in mammalian cell culture and invertebrate animal models, successfully identifying those candidate genes showing interactions with established PD mechanisms, including mitochondrial dynamics and  $\alpha$ -synuclein mediated neurodegeneration. Although no sufficiently powered exome dataset was available for definitive replication, human genetic validation was undertaken in several independent datasets. Our integrated approach identifies five strong candidate PD susceptibility genes worthy of further investigation, and exemplifies a powerful strategy with potential broad applicability to the follow up of future rare variant studies in PD and other neurologic disorders with complex genetic etiologies.

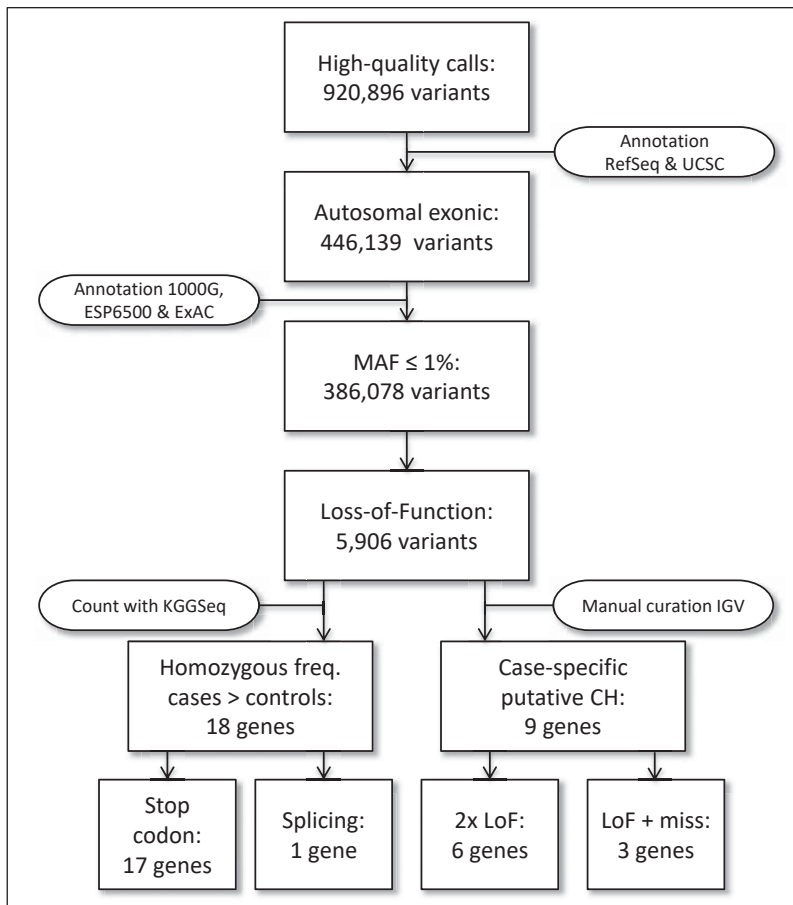
## RESULTS

### Discovery of recessive loss-of-function variants from PD exomes

A total of 920,896 variants (93.2% single nucleotide variants and 6.8% insertions and deletions) were called in a WES dataset of 1,651 subjects, including 1,148 young-onset PD cases (average age of onset 40.6 years, range 5-56 years) and 503 control subjects with European ancestry. As our cohort has an average age at onset below 45 years, we focused our search on homozygous and putative compound heterozygous variants, consistent with a recessive inheritance model. Although most PD cases were prescreened for mutations in established PD genes, we identified 2 subjects with homozygous exonic variants in *parkin* and *PINK1* (Supplementary Table 1). In order to identify novel PD gene candidates we focused on variants that are rare in control populations. Considering the worldwide prevalence for PD (0.041% in individuals aged 40-49 years),<sup>20</sup> we used a minor allele frequency (MAF) threshold of 1%, and only considered LoF variants causing a premature stop codon or splicing site mutations (see Methods). When co-occurring with a heterozygous LoF variant, we also considered rare, heterozygous amino-acid compound heterozygous recessive genotype.

Figure 1 displays each variant filtering step along with the corresponding numbers of implicated variants. Following Sanger sequencing confirmation, we identified a total of 27 candidate genes, 18 genes encompassing homozygous variants and 9 genes harboring putative compound heterozygous variants, all predicted to cause a loss of gene function (Table 1). ~17% of the variants are absent in public allele frequency databases (1000 Genomes Project (1000G), Exome Sequencing Project v. 6500 (ESP6500) or Exome

Aggregation Consortium (ExAC)) and therefore implicated to be novel. Except in the case of *ARSB*, the 26 other genes harbor LoF variants in only a single case, consistent with the hypothesis that novel recessive PD alleles may consist of many rare, “private” mutations. Four PD cases in our cohort were identified with a LoF variant in the *ARSB* gene, in which mutations have previously been linked with the recessive lysosomal storage disorder, MPS VI (also called Maroteaux-Lamy Syndrome). All 4 individual cases, along with 1 control subject, were homozygous for a variant (rs138279020) predicted to disrupt splicing.



**Figure 1.** Flowchart explaining multiple filtering steps to select LoF variants with assumed recessive inheritance pattern. Functional annotation was performed with transcripts of RefSeq and UCSC databases. Minor allele frequency annotations were based on 1000 Genomes project, Exome variant Server and the ExAC database. Seventeen genes harbored homozygous variants causing stopgain or loss and 1 gene contained a homozygous splicing variant. For the putative compound heterozygous genes, 6 genes were selected based on the presence of 2 times a LoF variant and 3 genes were based on the presence of one LoF variant and one missense variant predicted to belong to the 1% most harmful variants of the genome.

Table 1. Recessive LoF variants

WES results		Variant characteristics				Functional information				Validation	
Gene	Type	Chr:bp (hg19)	Ref/Alt all	ExAC (EU)		Variant Type	AAchange	AA length	CADD	Func-tional	Genetic
				maf	n HZ (freq.)						
ANKRD30A	CH	10:37438759	G/T	x	x	stopgain	NM_052997:p.E487X	1341	35.0		
		10:37508648	A/AT	x	0.0031%	0	fs insertion	NM_052997:p.Q1280fs	1341	24.2	
ARSB	HZ	5:78281383	T/TA	rs138279020	x	x	splicing	NM_198709:c.190-2inst	533	2.2	S
<i>c11orf21</i>	HZ	11:2321829	C/T	rs74048215	0.037%	0	stopgain	NM_001142946:p.W69X	178	35.0	neuroX
CALML4	HZ	15:68497597	G/A	rs11071990	0.75%	1 (0.00003%)	stopgain	NM_001031733:p.R40X	120	25.9	
CAPS2	HZ	12:75687045	C/CTATT-GAT	x	0.68%	4 (0.00012%)	stopgain	NM_001286547:p.L321X	382	33.0	
CD36	CH	7:80285955	C/T	x	0.0075%	0	stopgain	NM_001127444:p.Q74X	472	26.5	GRIP
		7:80302705	C/T	x	0.011%	0	missense	NM_001127444:p.P412S	472	25.9	
COL6A5	CH	3:130132401	C/T	rs190283135	0.12%	0	stopgain	NM_001278298:p.Q1559X	2611	38.0	
		3:130107827	G/A	x	0.32%	0	missense	NM_001278298:p.V756M	2611	23.0	
DIS3	CH	13:73355096	G/A	rs201493043	0.0045%	0	stopgain	NM_014953:p.Q92X	958	37.0	Mm
		13:73333935	A/G	rs141067458	0.22%	0	stoploss	NM_014953:p.X959Q	958	14.1	
FAM71A	HZ	1:212799882	A/T	rs143861665	0.80%	2 (0.00006%)	stopgain	NM_153606:p.K555X	594	36.0	Mm
FAM83A	CH	8:124195352	G/T	rs148011353	0.23%	2 (0.00006%)	stopgain	NM_032899:p.G86X	434	26.2	PPMI
		8:124195506	T/G	x	x	x	missense	NM_032899:p.V137G	434	25.6	
GH2	HZ	17:61957946	C/T	rs150668018	0.33%	1 (0.00003%)	stopgain	NM_022557:p.W214X	256	26.8	
GPATCH2L	HZ	14:76644266	C/T	rs117516637	0.53%	1 (0.00003%)	stopgain	NM_017972:p.R362X	482	18.8	Mp,Mm PPMI
KALRN	HZ	3:124303696	C/T	rs56407180	0.39%	0	stopgain	NM_007064:p.R10X	1289	37.0	Mp
KCNK16	HZ	6:39282816	G/A	rs138573996	0.44%	2 (0.00006%)	stopgain	NM_001135107:p.Q251X	322	9.5	Mm
MNS1	CH	15:56736045	G/A	x	x	x	stopgain	NM_018365:p.Q232X	495	38.0	Mm
		15:56748680	G/A	x	0.003%	0	stopgain	NM_018365:c.C265T:p.Q89X	495	37.0	



OR7G3	HZ	19:9236969	G/A	rs61751875	0.15%	1 (0.00003%)	stopgain	NM_001001958:p.R220X	312	35.0	Mm
PCDHA9	HZ	5:140229907	C/G	x	0.0046%	0	stopgain	NM_014005:p.Y609X	950	34.0	Mm
PTCHD3	HZ	10:27688101	G/A	rs142646098	0.97%	6 (0.00018%)	stopgain	NM_001034842:p.R476X	767	34.0	Mp,Mm
PTPRH	CH	19:55697712	G/A	rs147881000	0.39%	1 (0.00003%)	stopgain	NM_002842:p.Q887X	1115	36.0	S,Mm neuroX
		19:55716715	C/A	rs201517965	0.0030%	0	stopgain	NM_002842:p.E200X	1115	26.0	
PZP	CH	12:9321534	G/A	rs145240281	0.85%	4 (0.00012%)	stopgain	NM_002864:p.R680X	1482	32.0	
		12:9333626	G/A	rs117889746	0.021%	0	stopgain	NM_002864:p.Q598X	1482	36.0	
SSPO	HZ	7:149493519	C/T	rs57595625	0.011%	0	stopgain	NM_198455:p.Q2199X	5150	10.4	
SVOPL	HZ	7:138341219	G/A	rs117871806	0.35%	0	stopgain	NM_174959:p.R18X	492	38.0	Mp,Mm
TCHHL1	HZ	1:152058192	G/A	rs150014958	0.037%	0	stopgain	NM_001008536:p.Q656X	904	35.0	Mp,Mm
TMEM134	HZ	11:67235051	G/A	rs143199541	0.70%	3 (0.00009%)	stopgain	NM_001078650:p.R84X	195	34.0	S
UHRF1B-P1L	HZ	12:100433523	T/A	x	x	x	stopgain	NM_015054:p.K1376X	1464	35.0	Mm neuroX
VPS13C	CH	15:62174851	C/A	x	x	x	stopgain	NM_017684:p.E3147X	3628	55.0	S,Mm neuroX ,GWAS
ZNF543	HZ	15:62305257	C/CTCTG	x	x	x	fs insertion	NM_017684:p.R226fs	3628	27.4	
		19:57838058	G/A	rs150392165	0.60%	1 (0.00003%)	stopgain	NM_213598:p.W68X	600	25.0	Mp,Mm

Type = type of affecting 2 alleles; CH = putative compound heterozygote; HZ = homozygote; Chr:bp = chromosome and base-pair position; hg19 = reference build; Ref = reference; Alt = alternative; ExAC = Exome Aggregation consortium; EU = European; fs = frameshift; Achange = amino acid change for specified RefSeq transcript; maf = minor allele frequency; n HZ = number of individuals with homozygous variant; AAlength = length of specified transcript in amino acids; CADD = functional algorithm prediction (>20 = belongs to 1% most damaging variants of total genome), Mp = mitochondrial Parkin translocation assay; Mm = mitochondrial morphology assay; S =  $\alpha$ -synuclein assay. The current designation of rs138279020 is rs11424557.

Although this variant is neither reported in ExAC nor was frequency information available from dbSNP, the MAF was 0.065 in our cohort ( $MAF_{\text{CASES}}=0.073$ ,  $MAF_{\text{CONTROLS}}=0.052$ ,  $p=0.054$ ). Although relatively frequent in our control dataset ( $MAF>1\%$ ), we have retained it among our results, based on 3 considerations. First, information was not present in dbSNP, ExAC or ESP6500, which was the basis for applying this frequency filter in all other cases. Second, at least one of the homozygous individuals had clinical manifestations consistent with MPS VI, supporting potential pathogenicity of this allele (see Discussion). Lastly, as detailed below, our functional studies identify links between manipulation of ARSB and cellular/organismal phenotypes consistent with a potential role in PD.

Of note, while the analyses of the IPDGC WES dataset and subsequent work described here were in progress, an independent family-based sequencing study identified *VPS13C* as a cause of autosomal recessive parkinsonism.<sup>21</sup> Although the single IPDGC subject with compound heterozygous *VPS13C* LoF alleles was published as a replicate case in that work, we retained it among the 27 candidates described here, since it was independently carried forward for all analyses detailed below.

### Tolerability of gene loss-of-function in humans and animal models

The “tolerability” of recessive LoF genotypes has important implications for understanding the genetic basis of adult-onset, age-influenced disorders such as PD. As most of the identified homozygous and putative compound heterozygous LoF genotypes are based on a single individual, we also examined for their occurrence in a large, recently published study<sup>16</sup> of predicted complete gene knock-outs in the Icelandic population, including 104,220 subjects with imputed genotypes, based on whole genome sequencing from a subset of 2,363 individuals. The Icelandic population is enriched for rare disease-causing mutations with a recessive inheritance pattern, given a strong founder effect and non-random mating patterns. Twelve of the variants that we identified are also present in the Icelandic study (Supplementary Table 2); however, the observed homozygote frequencies are not sufficiently high to confidently exclude them as possible PD genes and importantly, detailed phenotypic data is not publically available for these subjects. For example, 29 Icelandic subjects are reported homozygous for the identical *PTCHD3* stopgain variant (c.C1426T, p.R476X) as the single PD case in our WES study. However, this is only 0.028% of the total sample set and below the reported prevalence of young-onset PD (0.041%).

We additionally examined for the presence of other LoF variants with a recessive inheritance pattern in our implicated candidate genes (Supplementary Table 2). For a subset of genes, we indeed identified several variants with particularly high homozygote frequencies including *OR7G3* (9.16%), *SSPO* (9.38%) and *PTCHD3* (16.55%). This is consistent with prior reports describing a homozygous deletion covering *PTCHD3* in apparently healthy individuals, consistent with a non-essential role.<sup>22</sup> Assuming that the variants in *OR7G3*, *SSPO* and *PTCHD3* confer similar LoF to the alleles identified in our PD

WES data, their high variant frequency makes these genes unlikely to be highly penetrant PD risk loci.

Human genes harboring homozygous LoF variants—especially those observed recurrently in large population-based datasets, potentially identify genes that are dispensable for fetal and subsequent child development. Given the limited human phenotypic information available, we further investigated the potential tolerability for the implicated genes using a cross-species approach, performing systematic LoF analysis in the nematode, *C. elegans*. Out of the 27 candidate genes identified in our WES analysis, 10 were well conserved in the *C. elegans* genome, and 9 had readily available RNA-interference (RNAi) reagents for LoF screening (see Methods). Each gene was targeted for knockdown using RNAi and we assessed for developmental lethality and survival. The results of these studies, along with other LoF data from public databases, are available in Supplementary Table 3. Knockdown of homologs of *DIS3* (*dis-3*), *KALRN* (*unc-73*), and *PTCHD3* (*ptr-10*) resulted in developmental arrest and/or reduced survival in *C. elegans*. Notably, homologs of *KALRN* and *DIS3* are also associated with reduced viability following genetic disruption in both *Drosophila*<sup>23,24</sup> and mice.<sup>25, 26</sup> Thus, these results are potentially consistent with conserved, early and/or essential developmental roles for these genes and the absence of individuals harboring homozygous LoF variants in the Icelandic cohort.<sup>16</sup>

Since the human genome contains multiple gene paralogs for *KALRN* and *PTCHD3*, genetic redundancy might account for how LoF might be tolerated in humans but not in simple animal models. Alternatively, it is possible that the allelic variants implicated in our PD WES cohort and Icelandic study might not cause a complete LoF (*i.e.* genetic null) despite the algorithmic predictions, instead causing only a partial loss-of-function. Nevertheless, these cross-species comparisons suggest essential and early developmental roles for homologs of *PTCHD3*, *DIS3* and *KALRN*, and informing our consideration of potential contribution to adult-onset disorders, such as PD.

### Variant aggregation analyses

For the 27 genes implicated based on our primary analyses of homozygous or compound heterozygous LoF variants, we additionally considered evidence for the presence of other allelic variants conferring risk for PD in our cohort. We therefore performed burden analyses leveraging our IPDGC WES data, testing two nested classes of variants: (i) a subset predicted to be deleterious (CADD>20) and (ii) all amino-acid changing missense alleles. Rare variants (MAF < 0.018) were considered either selectively or in joint models with common variants (MAF > 0.018). As detailed in Supplementary Table 4, the rare variant aggregation association analyses provided further evidence in support of 4 candidate genes: *GH2*, *PTPRH*, *UHRF1BP1L* and *ZNF453*. Interestingly, the burden association at the *PTPRH* gene is further enhanced when common and rare variants are simultaneously modeled.

Our analyses of LoF variants in PD exomes identify a number of promising candidate genes. However, even though a positive family history was observed for almost 40% of the cases, segregation analysis of the variants in families is not feasible, as DNA samples are not available from additional family members. Further, since most of the genes implicated contribute to single or few cases, we are unable to perform meaningful statistical comparisons, based on the limited numbers of LoF variants identified by WES in cases vs. controls. As an alternative strategy we therefore deployed a combination of cell-based and model organism functional screens to define potential links between the 27 candidate genes (Table 1) and well-established mechanisms of PD susceptibility and pathogenesis, including (1) mitochondrial health and (2)  $\alpha$ -synuclein mediated toxicity.

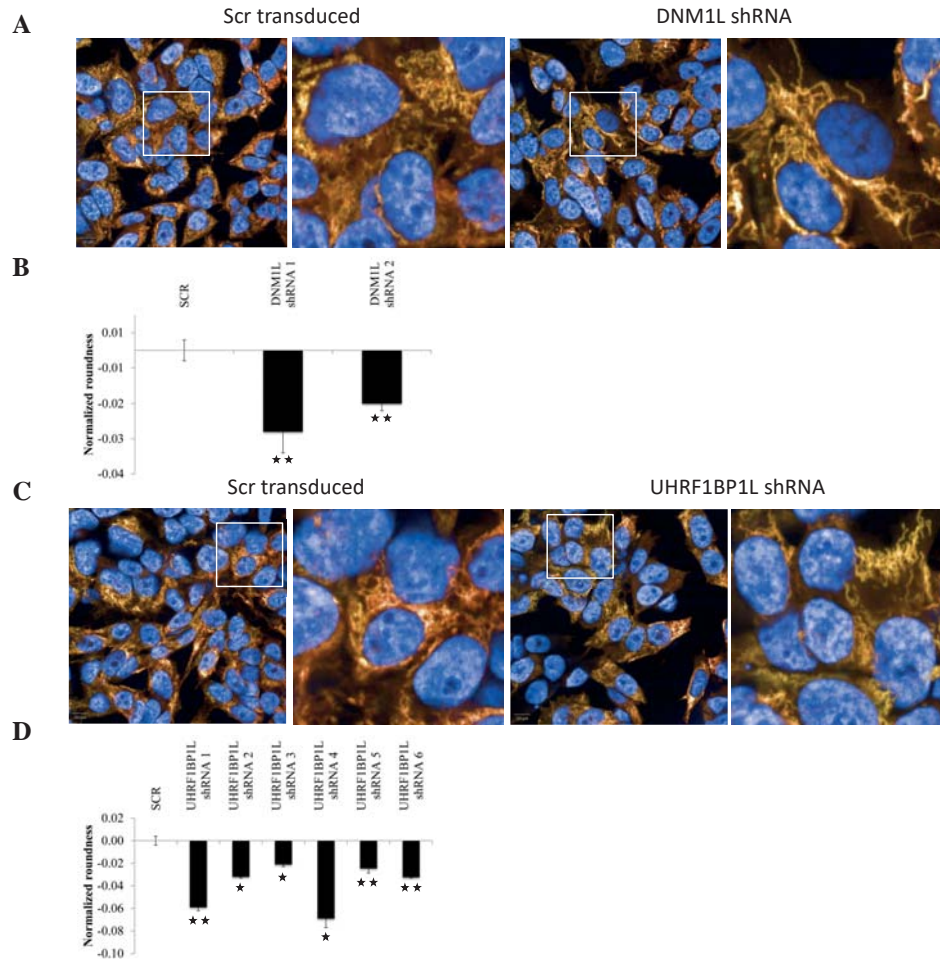
### Functional prioritization: mitochondrial health

Although the mechanism of neurodegeneration in PD remains incompletely defined and may be heterogeneous, mitochondrial dysfunction has been proposed to play an important role particularly in young onset PD.<sup>27-29</sup> Notably, *parkin* (*PARK2*), *DJ-1* and *PINK1*, associated with autosomal recessive, juvenile-onset Parkinsonism, have roles in mitochondrial dynamics and quality control.<sup>30</sup> Specifically, Parkin is an E3 ubiquitin ligase and recruited selectively to dysfunctional mitochondria with a low membrane potential.<sup>31</sup> Further, the neurotoxicity of  $\alpha$ -synuclein, the primary constituent of Lewy body inclusions in PD, has also been linked to mitochondrial injury.<sup>32</sup> We therefore hypothesized that LoF in candidate genes identified from our analyses of WES, might similarly impact mitochondria, consistent with roles in PD susceptibility.

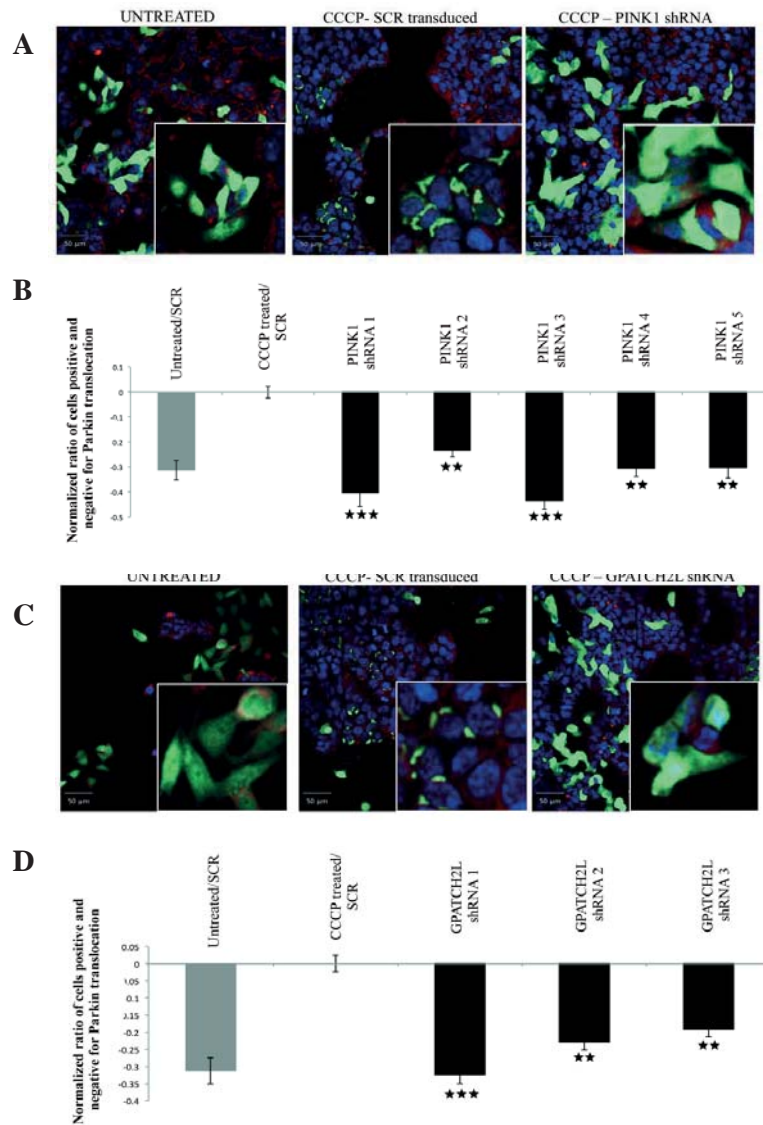
6

Therefore we quantified mitochondrial morphology after gene knockdown in BE(2)-M17 neuroblastoma cells by examining 3 parameters commonly used for quantification of mitochondrial morphology: mitochondrial number, axial length ratio, and roundness.<sup>33</sup> Cells transduced with the shRNA encoding a scrambled sequence were used for normalization and positive controls for mitochondrial morphology were included in each experiment. For example, knockdown of the mitochondrial fission gene dynamin 1-like (*DNM1L*), a positive control, results in elongated mitochondria and therefore decreases mitochondrial axial length ratio and roundness (Figure 2A,B).<sup>34</sup> Knockdown of 13 genes show a significant effect on at least one of the 3 parameters (Supplementary Table 5; Supplementary Figure 1). *GPATCH2L* shows the largest increase in mitochondrial roundness, while *UHRF1BP1L* displays the largest decrease (Figure 2C,D).

We also took advantage of a well-established Parkin translocation assay<sup>31, 35-38</sup> based on BE(2)-M17 human neuroblastoma cells stably expressing Parkin-GFP. As expected, upon exposure to the mitochondrial toxin and electron transport chain uncoupling reagent, CCCP, we observed robust translocation of Parkin-GFP from the cytoplasm (Figure 3A, untreated) to the mitochondria (Figure 3A, CCCP-SCR transduced), and this was PINK1-dependent (Figure 3A, CCCP-PINK1 shRNA), which provides an internal, positive control in



**Figure 2.** High-content assay for Mitochondrial Morphology. Effect of DNM1L shRNA (A and B) and UHRF1BP1L shRNA (C and D). BE(2)M17 cells stained with Hoechst (BLUE; Nuclei), MitoTracker CMXros and MitoTracker Deepred (YELLOW; mitochondria). A) Cells infected with shRNA encoding a scrambled sequence (SCR, left panel) and decrease in mitochondrial axial length ratio and roundness for DNM1L (positive control, right panel). B) The graph displays normalized mitochondrial roundness. C) Cells infected with shRNA encoding a SCR sequence (left panel) and decrease in number of mitochondria per cell, mitochondrial axial length ratio and roundness for UHRF1BP1L (right panel). D) The graph displays normalized mitochondrial roundness. Data are median values  $\pm$  median absolute deviation (MAD) of N=6 measurements. \*P < 0.05 and \*\* P < 0.01, Mann-Whitney U test (see methods). All values were normalized to the negative control (infected with SCR shRNA) and all shRNA clones that meet the cut-off criteria are shown (B and D). This figure is available at higher resolution through <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1147-9>.



6

**Figure 3.** High content assay for Parkin Translocation. Effect of PINK1 shRNA (A and B) and GPATCH2L shRNA (C and D). A and C) Cells are labeled for nuclei (BLUE; Hoechst), Parkin-GFP (GREEN), mitochondria (RED, Mitotracker Deepred). Untreated cells infected with shRNA encoding a scrambled sequence show absence of puncta (left panel). Cells infected with a scrambled sequence but treated with CCCP show a significant increase in puncta formation (middle panel). Infection of cells with shRNA targeting PINK1 or GPATCH2L prevents the accumulation of Parkin on mitochondria (right panel). B and D) The graph displays the normalised ratio of cells positive for translocation and cells negative for parkin translocation. All values were normalised to the negative control (CCCP treated infected with shRNA encoding a scrambled sequence). Data are median values  $\pm$  median absolute deviation (MAD) of N=6 measurements. \*  $P < 0.05$ , \*\*  $P < 0.01$  and \*\*\*  $P < 0.001$ , Mann-Whitney U test (see methods). All shRNA clones that meet the cut-off criteria (see methods) are shown. This figure is available at higher resolution through <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1147-9>.

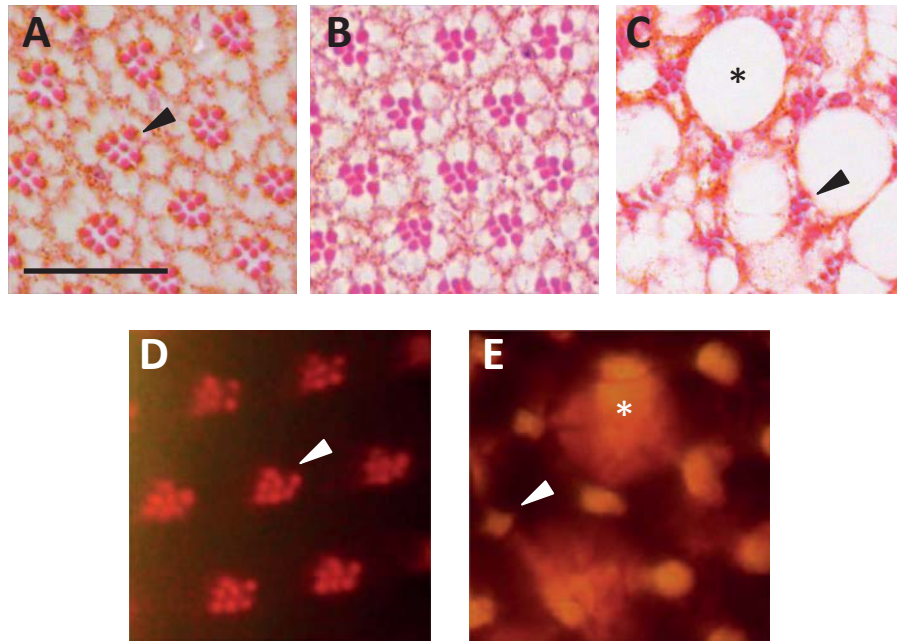
our assay. CCCP-induced Parkin accumulation was assessed by high-content microscopy and automated image analysis following systematic shRNA-knockdown of our 27 candidate genes (Figure 3B). Based on stringent criteria (see Methods), 6 genes significantly modified Parkin translocation (Figure 3C and 3D; Supplementary Figure 2; Supplementary Table 5), including 4 genes (*GPATCH2L*, *PTCHD3*, *SVOPL* and *ZNF543*) with consistent activities in both the mitochondrial morphology and Parkin translocation assays.

### Functional prioritization: $\alpha$ -Synuclein mediated toxicity

A wealth of evidence also supports a central role for  $\alpha$ -synuclein-mediated toxicity in PD pathogenesis.  $\alpha$ -synuclein aggregates, termed Lewy bodies, are the defining disease pathology, and both  $\alpha$ -synuclein gene (*SNCA*) mutations, locus multiplication, and promoter polymorphisms are associated with PD susceptibility.<sup>5</sup> Further, expression of  $\alpha$ -synuclein in numerous animal models including in the fruit fly,<sup>39-41</sup> *Drosophila melanogaster*, recapitulates features of PD-related neurodegenerative pathology. Transgenic expression of  $\alpha$ -synuclein in the fly retina leads to neurotoxic changes<sup>39</sup> and is amenable for detection of genetic modifiers.<sup>42, 43</sup> Genetic manipulation of established PD susceptibility genes, including *PARK2*<sup>44, 45</sup> and *VPS35*,<sup>46</sup> modulate  $\alpha$ -synuclein toxicity in transgenic flies, similar to findings in mammalian models.<sup>44, 47</sup> We therefore hypothesized that LoF in homologs of novel PD genes may similarly enhance  $\alpha$ -synuclein-induced retinal degeneration.

Out of the 27 candidate genes implicated by our WES analyses, 13 were well-conserved in *Drosophila* (Supplementary Table 6). Available RNAi stocks targeting each of the 18 fly homologs (some genes had multiple conserved paralogs) were crossed to flies in which the human  $\alpha$ -synuclein transgene was directed to adult photoreceptors using the *Rhodopsin1-GAL4* (*Rh1*) driver (*Rh1*> $\alpha$ -synuclein).<sup>48</sup> For rapid screening, retinal neurodegeneration was monitored using the optical neutralization technique which allows assessment of retinal tissue integrity in intact, unfixed heads. In *Rh1*> $\alpha$ -synuclein animals, the retina appears morphologically normal at 1 day (Figure 4), but demonstrates age-dependent degeneration leading to progressive vacuolar changes, rhabdomere loss, and culminating with extensive tissue destruction by 30 days. At the 15-day time point selected for screening, only mild, if any, retinal pathology is detectable on most histologic sections, consistent with a weakly-penetrant degenerative phenotype following optical neutralization (mean penetrance ~25%) (Figure 4). However, co-expression of RNAi targeting fly homologs of 4 candidate genes (*ARSB*, *TMEM134*, *PTPRH*, and *VPS13C*) was observed to robustly enhance  $\alpha$ -synuclein-mediated neurodegeneration in the retina (mean penetrance ~ 75%; Supplementary Table 7).

All candidate enhancers of  $\alpha$ -synuclein identified using the screening assay were further confirmed based on retinal histology, demonstrating accelerated pathologic changes with a significantly increased overall extent and severity of degeneration compared to *Rh1*> $\alpha$ -synuclein controls without RNAi transgenes present (Figure 5). Importantly, when



6

**Figure 4.**  $\alpha$ -synuclein-induced retinal degeneration and screening assays in *Drosophila* transgenic animals. Tangential sections through the fly retina stained with hematoxylin and eosin reveal the ordered ommatidial array in control animals (A, Rh1-GAL4 / +). Each ommatidia consists of a cluster of 8 photoreceptive neurons (7 visible at the level examined). The photoreceptors each contain a single rhabdomere, the specialized organelle subserving phototransduction, giving the ommatidia cluster its characteristic appearance (arrowhead). Expression of  $\alpha$ -synuclein in adult photoreceptors (B & C, Rh1-GAL4 / +; UAS- $\alpha$ -synuclein / +) causes age-dependent, progressive retinal degeneration. Compared to 1-day old Rh1> $\alpha$ -synuclein flies (B), histologic sections in 30-day old animals (C) demonstrate rhabdomere/cell loss and substantial vacuolar changes (asterisk). The pseudopupil preparation allows visualization of rhabdomeres (arrowhead) in intact, unfixed intact fly heads, permitting medium-throughput screening for progression of  $\alpha$ -synuclein-induced retinal pathology. Compared to controls (D, Rh1-GAL4 / +), in 30-day old  $\alpha$ -synuclein transgenic animals (E, Rh1-GAL4 / +; UAS- $\alpha$ -synuclein / +) rhabdomeres frequently appear indistinct (arrowhead) and vacuolar changes disrupt light refraction (asterisk). Representative control histology (A) and pseudopupil images (D) are shown for 15-day-old animals, the timepoint used for screening, in order to facilitate comparison with Figure 5. Scale bar: 20  $\mu$ m

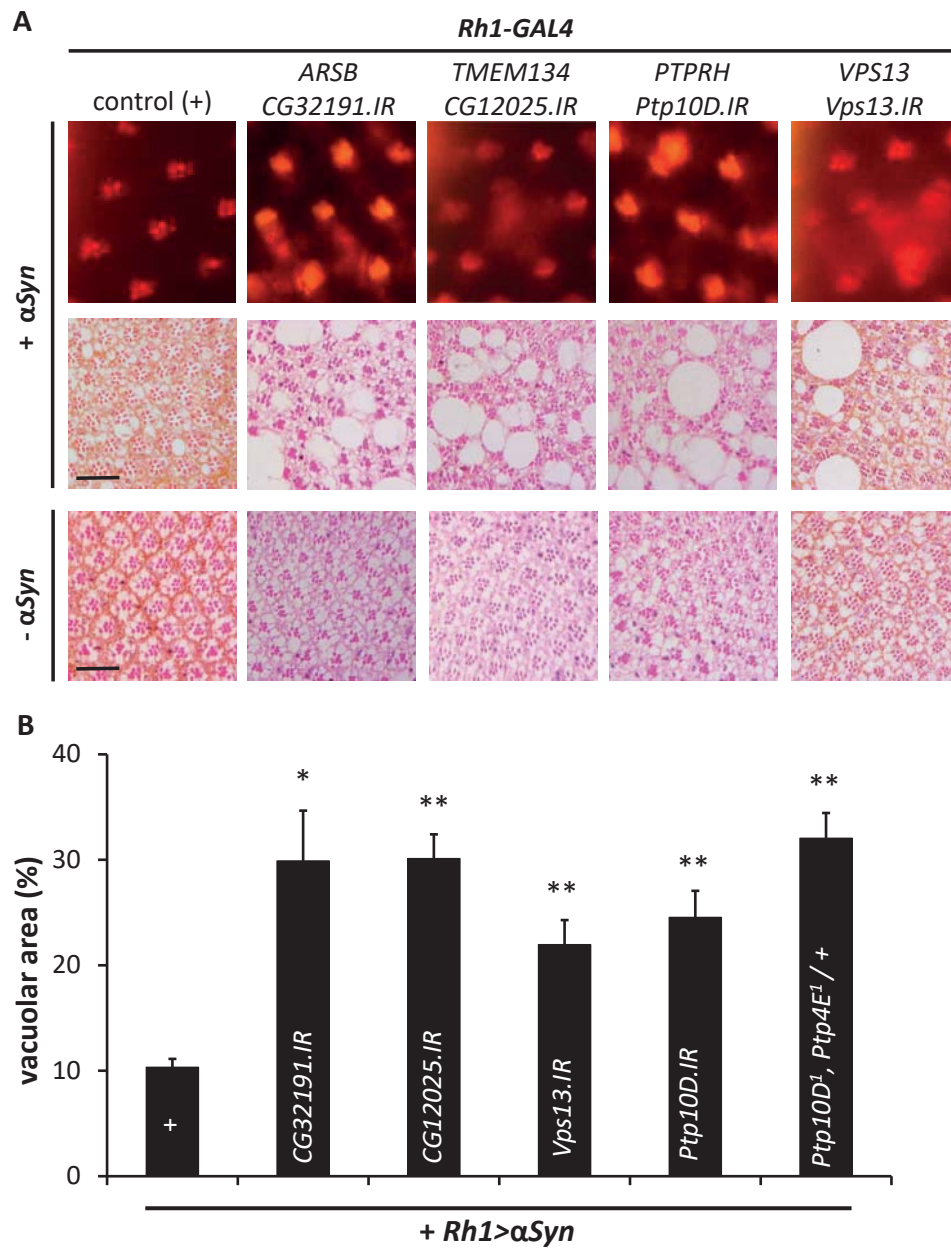


each of these genes were targeted under similar experimental conditions (*Rh1>RNAi*), but independent of  $\alpha$ -synuclein expression, we did not observe any significant retinal pathology in 15 day-old animals (Figure 5). Therefore, within the *Drosophila*  $\alpha$ -synuclein transgenic model system, the implicated LoF enhancers appear consistent with synergistic (non-additive) effects on  $\alpha$ -synuclein-mediated retinal degeneration. Since increased  $\alpha$ -synuclein expression levels are one important mechanism of PD susceptibility,<sup>5</sup> western blots were performed to determine whether any of the identified genetic enhancers alter  $\alpha$ -synuclein protein levels. However, following RNAi-mediated knockdown, none led to significant changes (Supplementary Figure 3). Thus, we hypothesize potential interactions with more downstream mechanisms of  $\alpha$ -synuclein neurotoxicity. For 3 out of the 4 candidate enhancers (*ARSB*, *VPS13C*, *PTPRH*), available siRNAs permitted additional testing of gene homologs as candidate modifiers in an established *C. elegans* model of  $\alpha$ -synuclein toxicity.<sup>49</sup> However, no significant differences were detected in the  $\alpha$ -synuclein-induced locomotor phenotype observed in 1 week old worms following knockdown of these genes (Supplementary Figure 4). We speculate that these contradictory results might stem from differences in assay sensitivity and/or tissue-specific toxic mechanisms as the fly and worm models are based on  $\alpha$ -synuclein expression in the retina versus muscle, respectively.

Of the 4 genes discovered to interact with  $\alpha$ -synuclein toxicity in *Drosophila*, we were able to obtain additional genetic reagents, including classical LoF alleles, for the 2 homologs of *PTPRH*, *Ptp10D* and *Ptp4E*. In our screen, 2 independent RNAi lines targeting *Ptp10D* robustly enhanced  $\alpha$ -synuclein toxicity, but only 1 of the 2 available lines for *Ptp4E* met our threshold criteria (Supplementary Table 7). Interestingly prior studies in *Drosophila* suggest that *Ptp10D* and *Ptp4E* are the result of a gene duplication event, and these genes show evidence of partial functional redundancy, including for nervous system phenotypes.<sup>50</sup> Consistent with this, we found that transheterozygosity for strong (null) alleles of both genes enhanced  $\alpha$ -synuclein-induced retinal degeneration (*Ptp4E<sup>1</sup>*, *Ptp10D<sup>1</sup> / +*; *Rh1-Gal4 / +*; *UAS- $\alpha$ -synuclein / +*); whereas heterozygosity for either allele in isolation showed no significant enhancement (Figure 5B and Supplementary Figure 5).

### Genetic replication of candidate PD genes from WES

We next evaluated our 27 gene candidates in additional available genetic datasets including (1) an independent exome sequencing dataset from the Parkinson Progression Markers Initiative (PPMI) project,<sup>51</sup> (2) a whole genome sequencing dataset including PD index cases of a Dutch genetic isolate belonging to the Genetic Research in Isolated Population (GRIP) program,<sup>52</sup> (3) an independent NeuroX exome array dataset,<sup>7,53</sup> and (4) a large PD GWAS dataset.<sup>53</sup> Within the PPMI exome dataset, including 462 PD cases and 183 controls, evidence supporting replication was discovered for 2 genes, in which we identified the identical variants from the IPDGC discovery exome dataset (Supplementary Table 8). A PD case from PPMI carries the same homozygous stopgain variant (p.R362X) in *GPATCH2L*



**Figure 5.** PD gene candidates harboring LoF variants enhance  $\alpha$ -synuclein toxicity in *Drosophila*. Conserved fly orthologs of human genes discovered from WES analysis were targeted with RNAi (IR) and screened for enhancement of  $\alpha$ -synuclein pathology using the pseudopupil assay (A, top row). For each line evaluated, the severity of retinal degeneration was scored based on penetrance of the  $\alpha$ -synuclein pseudopupil phenotype, and enhancers required consistent results for at least 2 independent RNAi lines (see Additional Table 7). For complete legend see supplemental data.

as observed for an IPDGC case. Although the age of onset differs 20 years between these two PD cases (47 and 68 years for the IPDGC and PPMI patient, respectively), they share similar asymmetric clinical symptoms at onset, which are characterized by resting tremor, bradykinesia and rigidity. Furthermore, both PD cases have a father diagnosed with PD, implying the variant to be highly penetrant. We excluded the possibility that these two PD cases might be related by computing pairwise genetic relationships<sup>54</sup> from common SNPs ( $MAF \geq 0.01$ ). No evidence of relatedness was observed ( $A_{jk} = -0.0018$ ). Based on ExAC, only one (0.003%) out of 32,647 European individuals has this same homozygous variant. The observation of two PD cases (0.12%) of our 1,610 studied PD patients (1,148 IPDGC WES plus 462 PPMI WES) with this *GPATCH2L* mutation is consistent with a 40-fold enrichment in our PD cohort. The second gene harboring an identical LoF variant is *FAM83A*. The p.G86X variant in *FAM83A*, detected within an IPDGC subject with sporadic PD diagnosed at the age of 28, was also observed in a single sporadic PD case from PPMI with an age of onset of 62 years. These *FAM83A* carriers presented similar symptoms, such as bradykinesia, rigidity and resting tremor. In both datasets, the G86X allele is predicted to be in *trans* with another variant: p.R347X or p.V137G in PPMI and IPDGC respectively.

The second genetic independent dataset that was investigated included a whole genome sequencing study (39 PD index cases and 19 controls) of a genetic GRIP isolate from the Netherlands, focusing on variants within our candidate genes that were present in at least 2 PD index cases and absent in controls. We identified a heterozygous missense variant (NM\_001127444:c.1176G>T: p.L392F) in *CD36* for 3 PD index cases. Although not consistent with a recessive inheritance model, this variant has not been observed in the 60,706 unrelated individuals of the ExAC database, suggesting potential enrichment in PD cases. These heterozygote variant carriers have a substantial higher age of onset (range 61 to 79 years) in comparison to the PD patient (age of onset: 38 years) with the putative compound heterozygous variant within the discovery WES dataset. This observation supports an additive model of pathogenicity, implying more severe disease onset when 2 alleles are affected. Further, *CD36* (p.L392F) is predicted to represent the top 1% most harmful variants within the genome (CADD-score=23.3). In the IPDGC discovery dataset, the discovered compound heterozygous variants, p.Q74X and p.P412S (Table 1), are also predicted to be strongly deleterious (CADD-scores of 26.5 and 25.9, respectively).

We next interrogated the independent IPDGC NeuroX dataset, including genotypes from 6,801 individuals with PD and 5,970 neurologically healthy controls. NeuroX is a genotyping array that includes pre-selected exonic variants, and is therefore not suitable to search for the identical recessive LoF variants implicated by our WES analyses. Instead, we examined the burden of multiple variant classes within the 27 candidate genes, following the same variant categories as for the original IPDGC WES dataset (Supplementary Table 9). When only considering variants predicted to be deleterious (CADD>20), an association is detected for *UHRF1BP1L* with PD risk ( $p = 0.005$ ).

This gene also shows an association with PD in the IPDGC WES dataset when performing a similar burden analysis considering missense variants (see above,  $p = 0.016$ ). Using the NeuroX dataset, we additionally confirmed the enrichment of rare *PTPRH* variants in subjects with PD (WES:  $p=0.034$ , NeuroX:  $p=0.045$ ). Furthermore, *VPS13C* and *ARSB* show significant associations to PD when considering the joint effect of all variants, both common and rare (Supplementary Table 9).

Leveraging available IPDGC GWAS data (13,708 cases / 95,282 controls) we next assessed for potential common variant association signals ( $p < 1 \times 10^{-4}$ ), using a 1Mb genomic window centered on each of the 27 candidate genes. Three loci (*VPS13C*, *PCDHA9*, and *TCHHL1*) showed evidence consistent with an association peak (Supplementary Figure 6). A genome wide significant association at the *VPS13C* locus, was in fact recently reported<sup>7</sup>; the best SNP (rs2414739,  $p = 3.59 \times 10^{-12}$ ) maps ~150 kb distal to *VPS13C*. Based on local patterns of linkage disequilibrium defined by Hapmap (Supplementary Figure 6), it is unlikely that rs2414739 is a proxy for p.E3147X or similar LoF variants in *VPS13C*; however, it might be possible that the SNP influences *VPS13C* expression by affecting the long non-coding RNA lnc-VPS13C-1<sup>55</sup> in which the SNP is located. The other 2 candidate association peaks, adjacent to *PCDHA9* and *TCHHL1*, are considerably weaker signals (rs349129 =  $1.40 \times 10^{-5}$  and rs7529535 =  $7.66 \times 10^{-5}$ , respectively) and given the distances (~500 kb) many other candidate genes are potentially implicated.

In sum, we identify additional genetic evidence consistent with replication for 7 genes (*GPATCH2L*, *FAM83A*, *CD36*, *UHRF1BP1L*, *PTPRH*, *ARSB* and *VPS13C*) that were implicated by our WES analysis, of which 5 (*GPATCH2L*, *UHRF1BP1L*, *PTPRH*, *ARSB* and *VPS13C*) are further validated based on functional evidence from PD-relevant experimental models.

## 6

### Transcriptomics-based functional exploration

Lastly, we examined each candidate gene from our WES analysis for co-expression with established PD susceptibility gene in expression networks derived from human substantia nigra, leveraging available data from the United Kingdom Brain Expression Consortium (UKBEC) and the Genotype-Tissue Expression project.<sup>56</sup> Of the 27 candidate genes, 7 were not sufficiently expressed in substantia nigra on the basis of UKBEC. Except for *DIS3*, these genes were also expressed poorly in publically available data of the Genotype-Tissue Expression (GTEx) project.<sup>56</sup> Consequently, expression values for these genes were not used for construction of the UKBEC gene co-expression network (GCN). The remaining 20 genes were assessed for co-expression with known Mendelian PD genes (*ATP13A2*, *FBXO7*, *LRRK2*, *PARK2*, *PARK7*, *PINK1*, *RAB39B*, *SNCA* and *VPS35*) using the UKBEC GCN (Supplementary Table 10; Supplementary Figure 7). This approach highlighted 3 genes (*UHRF1BP1L*, *GPATCH2L* and *PTPRH*), and the implicated networks were further interrogated based on gene set enrichment analysis using gene ontology (GO) terms to

denote potential functions. *UHRF1BP1L* was co-expressed with *SNCA*, *PINK1*, *GBA*, and *ATP13A2* in a network significantly enriched for genes with roles in synaptic transmission ( $p=2.27 \times 10^{-11}$ ) as well as astrocytic ( $p=8.18 \times 10^{-8}$ ) and dopaminergic neuronal markers ( $p=3.98 \times 10^{-46}$ ). *GPATCH2L* was co-expressed with *PARK7* in a network enriched for other neuronal genes ( $p=3.41 \times 10^{-12}$ ) with cellular roles in metabolism of macromolecules ( $p=3.82 \times 10^{-15}$ ). Lastly, *PTPRH* was assigned to a co-expression module including *FBX07* and enriched for oligodendrocyte markers ( $p=8.69 \times 10^{-22}$ ). Importantly, the implicated modules were preserved (Z.summary  $\geq 10$ ) in the independent GTEx dataset.

## DISCUSSION

We report the results from WES analysis in the largest PD cohort studied to date. Assuming a recessive inheritance model, we identified 27 candidate genes harboring rare homozygous or compound heterozygous LoF variants. With the exception of *ARSB*, we did not identify recurrent recessive alleles in more than a single PD case. This result—potentially consistent with a highly heterogeneous genetic etiology for PD—creates significant barriers for statistical confirmation and genetic replication of novel PD susceptibility loci. Additional genetic samples were not available for segregation analysis, and given the rarity and heterogeneity of the implicated alleles, definitive human genetic replication would likely require very large sample sizes, including many thousands of PD cases with either WES or gene resequencing. We therefore coupled our WES analyses with functional studies in both mammalian cells and experimental animal models, including *Drosophila* and *C. elegans*, in order to prioritize genes for future study. Our results highlight 15 out of the 27 gene candidates that interact with mitochondrial dynamics and 4 loci that enhance  $\alpha$ -synuclein-mediated neurodegeneration. As discussed below, while these results highlight a promising subset of genes with potential links to PD-relevant mechanisms, we can not exclude contributions from other implicated genes/variants. All of these data, including promising variants from the human genetic analyses and results of functional studies, will be a valuable resource for future investigations of PD genomics. Analyses of several other WES and complementary large-scale, genetic datasets provide additional evidence supporting replication for 7 out of the 27 genes. Evidence from human genetics and functional studies converge to most strongly implicate 5 gene candidates discussed below; however, further investigation will be required to definitively link each of these loci to PD susceptibility and elucidate the relevant mechanisms. Nearly all of these genes are robustly expressed in brain,<sup>56</sup> including the substantia nigra, thereby consistent with their implication in PD. A subset (*GPATCH2L*, *UHRF1BP1L* and *PTPRH*) are co-expressed with established Mendelian PD genes in the substantia nigra based on analyses of UKBEC and GTEx expression data. In sum, our results define several promising new susceptibility

loci candidates for further investigation, and illustrate a powerful, integrative discovery strategy for future, large-scale PD genomic studies.

Mitochondrial mechanisms have been strongly implicated in PD risk and pathogenesis.<sup>28, 30</sup> Following shRNA-mediated knockdown, 15 candidate recessive loci identified in our WES dataset showed effects on mitochondrial morphology and Parkin translocation to mitochondria in cell culture. We focus our initial discussion on 3 genes, *GPATCH2L*, *UHRF1BP1L*, and *VPS13C*, for which we discovered additional genetic evidence consistent with replication in independent cohorts. In the IPDGC cohort, a single PD case was identified with a homozygous stopgain variant (p.R362X) in *GPATCH2L*, and a second individual with the identical, rare genotype was discovered in PPMI. This variant is reported with a low frequency of 0.003% in ExAC. Although minimal clinical or demographic information is available within ExAC, this finding is compatible with population prevalence estimates for PD.<sup>20</sup> Nevertheless, genotyping of p.R362X in additional large PD case and control cohorts will be required to definitively establish an association with PD susceptibility. *GPATCH2L* knockdown both increased mitochondrial roundness and impaired Parkin translocation. The encoded protein, GPATCH2L, which has not previously been studied, contains a glycine-rich RNA-binding motif, the “G-patch” domain.<sup>57</sup> *GPATCH2*, a paralog of *GPATCH2L*, is upregulated in cancer cells, localizes to the nucleus where it interacts with RNA-processing machinery, and manipulation in culture alters cell proliferation.<sup>58,59</sup> Notably, *GPATCH2L* is non-conserved in either the *C. elegans* or *Drosophila* genomes, precluding study of this candidate in these models. While our results using cellular assays implicate *GPATCH2L* in mitochondrial quality control mechanisms, further follow-up studies in mammalian model systems will be needed to confirm a role in PD pathogenesis.

6

Another promising gene, *UHRF1BP1L*, harbored a homozygous stopgain variant (c.4126A>T, p.K1376X) in a single IPDGC case. This is a novel variant, based on its absence from the ExAC cohort. Additional support for *UHRF1BP1L* as a *bona fide* PD locus, comes from complementary analyses in both the IPDGC WES and NeuroX datasets, documenting a burden of rare missense and LoF variants in association with disease risk. In the UKBEC, *UHRF1BP1L* was associated with a substantia nigra co-expression module including both *SNCA* and *PINK1*, reinforcing potential links with established PD genetic mechanisms. Indeed, *UHRF1BP1L* knockdown cause sharply reduced mitochondrial numbers and altered morphology. Interestingly, *UHRF1BP1L* encodes a protein bearing an amino terminal homologous to yeast VPS13, and studies in cell culture provide support for a role in retrograde transport from the endosome to the trans-Golgi network.<sup>60</sup>

Notably, LoF in human *VPS13C* was also implicated by our analyses of IPDGC WES data, and knockdown disrupted mitochondrial morphology. Besides the single IPDGC case, several families with autosomal recessive early onset parkinsonism and dementia due to *VPS13C* were recently reported,<sup>21</sup> and this locus also harbors common

PD susceptibility variants based on GWAS.<sup>7</sup> Our findings of a potential mitochondrial role for *VPS13C*, agree with those of Lesage et al., who additionally reported that *VPS13C* localizes to the outer membrane of mitochondria and LoF was associated with reduced mitochondrial membrane potential, fragmentation, and increased Parkin-dependent mitophagy. Importantly, *VPS35*, which causes autosomal dominant, late-onset PD, is similarly involved in endosomal trafficking,<sup>61</sup> and has also recently been implicated in mitochondrial dynamics,<sup>62</sup> including interactions with Parkin.<sup>63</sup> Like *UHRF1BP1L*, *VPS13C* and *GPATCH2L* are expressed in the brain, including within the substantia nigra; however, additional work will be needed to define their functions, including potential interactions with other established disease genes (e.g. *VPS35*, *parkin*) and requirements for mitochondrial maintenance.

Based on functional screening in *Drosophila*, 4 candidate genes from our WES analyses were implicated as LoF enhancers of  $\alpha$ -synuclein neurotoxicity, which also has a central role in PD pathogenesis. We discuss the 3 genes (*VPS13C*, *PTPRH*, and *ARSB*) where additional human genetic evidence supports replication. Interestingly, besides its requirement for mitochondrial maintenance, RNAi-mediated knockdown of *Drosophila Vps13* enhanced  $\alpha$ -synuclein toxicity. In the single reported *VPS13C* PD case with a completed autopsy, neuropathological findings included abundant  $\alpha$ -synuclein aggregates in both the brainstem and cortex.<sup>21</sup> Thus, *VPS13C* and associated endosomal sorting pathways (including *VPS35*) may represent a point of convergence for mitochondrial and  $\alpha$ -synuclein-mediated PD mechanisms. Consistent with this, evidence for the impact of  $\alpha$ -synuclein toxicity on mitochondria has recently emerged,<sup>28</sup> including from studies in mammals.<sup>64</sup>

In the IPDGC WES cohort, a single PD case was discovered with compound heterozygous LoF variants in *PTPRH* (c.2659C>T, p.Q887X and c.598G>T, p. E200X). Both variants were also observed at low frequencies in the ExAC database (0.039% and 0.003%, respectively); however, they each met our pre-specified threshold of < 1% based on the population prevalence of PD. Encoding a receptor protein tyrosine phosphatase, *PTPRH* (also called *SAP-1*) was first discovered for its potential association with gastrointestinal cancers<sup>65, 66</sup> and remains poorly studied in the nervous system context. In studies of both vertebrates and invertebrates, receptor protein tyrosine phosphatases have been strongly implicated as key neural cell adhesion receptors, with roles in neurodevelopment and synaptic function, and other members of this family have been implicated in numerous neuropsychiatric disorders.<sup>67</sup> In *Drosophila*, RNAi-mediated knockdown of the conserved *PTPRH* ortholog, *Ptp10D*, enhanced  $\alpha$ -synuclein-triggered retinal degeneration; but was not associated with substantial neurotoxicity independent of  $\alpha$ -synuclein expression. *Ptp10D* mutant flies are also viable and fertile but demonstrate long-term memory deficits in behavioral assays.<sup>68</sup> More recent studies further implicated Ptp10D in neural-glial interactions during development of the central nervous system,<sup>69</sup> potentially

consistent with our findings that human *PTPRH* participates in a substantia nigra gene co-expression network strongly enriched for oligodendrocyte markers. Besides our discovery of homozygous LoF in *PTPRH*, further analyses of the IPDGC WES dataset, and the substantially larger, independent NeuroX cohort, implicate a burden of rare variants at this locus in association with PD susceptibility.

$\alpha$ -synuclein-induced neurodegeneration was also enhanced by knockdown of *CG32191*, a *Drosophila* homolog of *ARSB*. RNAi transgenic lines targeting 3 other conserved fly *ARSB* homologs showed consistent interactions with  $\alpha$ -synuclein (Supplementary Table 6 and 7). In the IPDGC cohort, we discovered 4 PD cases homozygous for a variant predicted to disrupt splicing of exons 1 and 2 in *ARSB*. Although the identified variant has not previously been documented in ExAC, we identified a single IPDGC control homozygote. Additional evidence supporting association of the *ARSB* gene with PD susceptibility comes from burden analysis in the independent NeuroX cohort. The surprisingly common *ARSB* splicing variant (rs138279020, MAF=0.065 in IPDGC) is a single nucleotide insertion allele within a poly-A repeat, which we speculate might lead to inefficient capture in prior WES and possibly explain the absence of this variant from ExAC and the 1000 Genomes project reference. All 4 PD cases in our data with the homozygous *ARSB* splicing variant were confirmed by Sanger sequencing. Intriguingly, mutations in *ARSB*, encoding the lysosomal enzyme Arylsulfatase B, are associated with the recessive lysosome disorder, Mucopolysaccharidosis type VI (MPS VI, also called Maroteaux-Lamy syndrome), in which the glycosaminoglycan, dermatan sulfate, accumulates causing skeletal dysplasia and other heterogeneous manifestations.<sup>70</sup> Substrate accumulation and associated cellular stress has been reported to induce markers of impaired autophagy and mitochondrial dysfunction in *ARSB* deficient fibroblasts from MPS VI patients, as in other lysosomal disorders.<sup>71, 72</sup> Importantly, Maroteaux-Lamy can be characterized by minimal or even absent clinical signs, leading to incidental discovery or diagnosis in adulthood, and such mild phenotypes have been suggested to accompany partial LoF with preserved low-level *ARSB* enzymatic activity.<sup>70, 73, 74</sup> Similar genotype-phenotype relationships have been documented for other lysosomal-storage disorders, including Gaucher disease, which has established links with PD risk.<sup>75, 76</sup> While a full accounting is outside the scope of this study, at least 1 of the 3 IPDGC cases for which records were available revealed clinical features potentially overlapping with MPS VI.

The strengths of our study include the largest PD WES discovery dataset assembled to date, complementary analyses in independent available cohorts to establish replication, and integration of promising human genetic findings with multiple functional assays relevant to PD mechanisms. Nevertheless, we also make note of several inherent limitations. In order to prioritize candidate genes for initial investigation, assumptions were made concerning the specific inheritance model (recessive) and stringent criteria were employed for variant filtering. In the future, it will be important to also consider



the possibility of dominantly-acting alleles; however, this substantially increases the number of variants to consider and also potentially complicates functional studies (i.e. compared with loss-of-function screening using RNA-interference). Our study design excluded consideration of many non-synonymous variants that could potentially cause loss- (or gain-) of gene function, along with certain non-truncating, frameshifting alleles (see Methods). Even with fairly stringent criteria for variant filtering and the assumption of recessive inheritance, we found evidence for substantial etiologic heterogeneity. Improved confidence for discovery of PD causal variants will likely come from PD WES cohorts with significantly enhanced sample sizes, as well as increased numbers of adult controls, including those with careful neurological assessments to exclude mild PD symptoms. Indeed, while most of the variants implicated by the IPDGC WES cohort were represented at low frequencies within the largest available public database, ExAC<sup>77, 78</sup>; however, we have no information about potential PD manifestations in such individuals or even subject age.

Since no single cellular or animal experimental model is expected to universally recapitulate all potential facets of disease biology, we note that the employed functional screening assays are potentially liable to false negative or positive findings. Importantly, experimental evidence of a genetic interaction with either mitochondrial dynamics or  $\alpha$ -synuclein mediated neuronal injury in our screening assays can not in isolation confirm a role in disease causation, but rather serves to prioritize genes for future investigation. Out of the 27 candidate genes implicated in the IPDGC WES discovery analysis, 14 were insufficiently conserved for follow-up in  $\alpha$ -synuclein transgenic flies. While simple animal models, including *Drosophila* or *C. elegans*, have made important contributions to our understanding of PD pathogenesis, selected mechanisms, such as the potential role of adaptive immunity or basal ganglia circuit dysfunction, cannot be addressed in invertebrates.<sup>79, 80</sup> We were unable to confirm our findings from *Drosophila* in a published *C. elegans* model of  $\alpha$ -synuclein toxicity. In the future, it will also be important to examine potential genetic interactions in other PD models, including *LRRK2* transgenic flies or those containing mutations in other PD loci, such as *VPS35* or *parkin*. While neuroblastoma cells offer the convenience of robust mitochondrial readouts, they are limited by their undifferentiated, transformed state distinct from that of postmitotic neurons. In the future, human induced pluripotent stem cells, including those derived from individuals with PD, can be differentiated into dopaminergic or other neuronal types and potentially deployed for functional screening strategies. Additionally, genome-editing technologies may facilitate systematic functional evaluation of candidate disease-associated variants of unknown significance.

## CONCLUSION

We have identified 5 excellent PD gene candidates (*GPATCH2L*, *UHRF1BP1L*, *PTPRH*, *ARSB* and *VPS13C*), harboring homozygous or compound heterozygous LoF variants in PD exomes, demonstrating functional interactions with mitochondrial and/or  $\alpha$ -synuclein-mediated mechanisms, and supported by evidence of replication in independent human datasets. The recent report<sup>21</sup> of additional PD families segregating LoF mutations in *VPS13C* along with other experiments supporting a role in mitochondrial mechanisms significantly strengthens the evidence in support of this gene in PD, and validates our overall approach. These loci are well-suited for further future efforts directed at human genetic replication and in-depth functional dissection. We also make available results, including findings from human genetic analyses and functional studies in most cases, on 22 other promising loci. These data will serve as a valuable reference for ongoing and future PD genetic studies. More broadly, our approach of integrating high-throughput sequencing in PD case/control cohorts with parallel systematic screening in cells and model organisms for functional prioritization exemplifies a powerful experimental strategy with great promise for future genomic studies of PD and other human disorders.

## METHODS

### 6

### Genetic analyses

#### *Whole exome sequencing*

Whole exome sequencing (WES) was performed on 1,148 PD cases and 503 neurologically healthy controls of European descent. All subjects provided written informed consent. Relevant local ethical committees for medical research approved participation in genetic studies. If PD patients were prescreened for known pathogenic mutations, they were excluded for exome sequencing when having such a variant. The cases were diagnosed with PD at a relatively young average age of 40.6 years (range 6-56 years), of which approximately 37% reported a positive family history. The neurologically healthy controls are on average 48.2 years of age (range 10-97 years). A more extensive overview of demographic information is reported in Supplementary Figure 8.

Due to improvements of the exome sequencing protocol over time, the exome sample libraries were prepared with different capture kits. For this study three different capture kits were used: Illumina TruSeq (San Diego, California) (62 Mb target), Roche (Basel, Switzerland) Nimblegen SeqCap (44.1 Mb target) and Agilent (Santa Clara, California) SureSelect (37.6 Mb target), which captured 96%, 81% and 71% of the targeted exome at least 10 times respectively (Supplementary Table 11). Exome libraries were sequenced on a HiSeq 2000 (Illumina, San Diego, California). The Burrows Wheeler Aligner MEM v0.7.9.a<sup>81</sup> was used to align the 100-bp paired-end reads to the human reference genome build

hg19. We called the single nucleotide variants (SNVs) and insertions/deletions (indels) for all samples simultaneously using Genome Analysis Toolkit (GATK) 3.x,<sup>82</sup> followed by the exclusion of low-quality variant calls not passing the default GATK filters. Individual genotypes were removed with genotype quality Phred-scores below 40. ANNOVAR<sup>83</sup> was applied to annotate the variants with information concerning variant type (valid annotations when Refseq in concordance with UCSC), MAF in the general population and predictions of the variant's effect on gene function, implementing CADD.<sup>84</sup>

#### *Variant identification in IPDGC WES dataset*

Considering the worldwide prevalence of 0.041% for PD in the age range of 40 to 49 years,<sup>20</sup> we selected rare variants with a MAF < 1% (corresponding to a homozygous frequency of 0.01%) in the European population. Because the specified 0.041% of the population with young-onset Parkinson's disease (YOPD) is not caused by one shared genetic factor, we expect a homozygous frequency of 0.01% to be an adequate cut-off, which would be able to determine variants present in approximately 25% of the YOPD population. As a comparison to the most common genetic cause of YOPD, *parkin*,<sup>85</sup> the most frequent mutation is an exon 3 deletion, which has been identified in 16.4% of YOPD patients.<sup>86</sup> Using ANNOVAR,<sup>83</sup> all variants were annotated with MAF information of ESP6500si (European American population),<sup>87</sup> 1000 Genomes Project (European population of April 2012 version)<sup>88</sup> and the Exome Aggregation Consortium (ExAC) browser (non-Finish European population).<sup>77, 78</sup> When no public allele frequency was available for homozygous variants, the in-house control dataset of 503 individuals was used as a reference for the general population. Homozygous variants were excluded when being common (> 1%) in controls or having a relative higher frequency in controls than in cases. KGGseq<sup>89</sup> was used to count the number of homozygous variants for the cases versus controls.

In addition to the population allele frequency filters, we only selected SNVs and indels affecting the position of the stop codon or located at a splice-site (within 2 bp of splicing junction), which are variants expected to result in a loss of gene function. As the aim of this study was to validate our approach to identify high promising PD candidate genes, rather than discovering all putative PD genes present within our WES dataset, we set a conservative selection criteria by only including frameshifts that caused an immediate stopcodon at the position of the indel. Splice-sites variants were only considered when being adjacently located to an exon that is coding for amino acids. As a final filter for the homozygous variants, we manually excluded variants that failed GATK VQSR and hard filtering. Quality predictions based on the ExAC database are more adequate, as it includes ~37x more samples than our dataset.

For the putative compound heterozygous mutations, both variants should be located within the same transcript and at least one allele should contain a LoF variant. The second variant could be 1) a LoF variant or 2) a missense variant that is absent in

dbSNP137<sup>90</sup> database and with a CADD-score above 20 (predicted to belong to the 1% most deleterious variants of the total genome), indicating a pathogenic effect. The latter 2 filter criteria should decrease the chance of including benign missense variants. The putative compound heterozygous variants were identified by scoring the number of variants per sample per gene with PSEQ (<https://atgu.mgh.harvard.edu/plinkseq/pseq.shtml>). The reads of variants located within approximately 200 base pairs were visualized in IGV<sup>91</sup> to judge the authenticity of the compound heterozygous variant. When the different variants are located on distinct alleles, the combination of variants was considered a true compound heterozygous mutation.

All recessive variants that remained after the filtering procedures were Sanger sequenced to confirm the variant calls generated by the exome pipeline.

#### *Variant aggregation analyses in the IPDGC WES dataset*

SKAT-c<sup>92</sup> was used to analyze the burden of coding variants for each identified gene. Both rare variants only and the joint effect of common and rare variants were tested. Because variant aggregation tests are prone to coverage differences, capture usage and population stratification, we performed a more stringent individual and variant QC, resulting in a reduced dataset of 1,540 samples (1,062 cases and 478 controls) covering 268,038 variants. Individuals were excluded when failing gender test, showing evidence of relatedness, having dubious heterozygosity/genotype calls or being a population outlier. Variants were removed when having a genotype missingness > 5%, a Hardy-Weinberg equilibrium  $P$ -value <  $1e^{-6}$  or a  $P$ -value for non-random missingness by phenotype <  $1e^{-5}$ . Variants were only considered for association analyses if located in region targeted by all different capture kits.

Benign variants have the potential to dilute a true association signal of the combined effect of functional variants in a gene. We therefore annotated variants with ANNOVAR<sup>83</sup> to group variants according to their type or predicted pathogenicity. Two subsets of variants were examined: 1) predicted pathogenic variants, including LoF variants and missense mutations that are predicted to be pathogenic by the CADD framework; 2) missense variants, including amino-acid changing and LoF variants.

As suggested by SKAT, we selected a MAF cut-off of 0.018, which is based on the total sample size and separates rare and common variants. Common variants (MAF > 0.018) were pruned using PLINK<sup>93</sup> (indep settings 50 5 1.5). Due to confounding factors (usage different capture kits and multiple CEU populations), 20 principle components, 10x coverage and gender were taken into account as covariates. Both a traditional one-sided burden (assuming all variants to have a harmful effect) and a two-sided SKAT test (allowing variants to be either damaging or protective) were performed. Empirical  $p$ -values were calculated by comparison of the nominal  $p$ -value to 10,000 permutations of affection status. Genes with an empirical  $p$ -value < 0.05 were considered to be significantly associated to PD.

#### *Genetic replication 1: Variant identification in PPMI WES dataset*

We obtained permission to access WES data generated by the Parkinson Progression Markers Initiative (PPMI).<sup>51</sup> After standard variant and individual QC, the dataset includes 477,512 variants for 462 PD cases and 183 neurologically healthy controls. A similar search for homozygous and putative compound heterozygous LoF variants, as described for the original IPDGC WES dataset, was applied for this second independent PPMI WES dataset by using ANNOVAR<sup>83</sup> and KGGSeq.<sup>89</sup>

#### *Genetic replication 2: GRIP genetic isolate*

The southwest of the Netherlands contains a recently isolated population which is part of the Genetic Research in Isolated Population (GRIP) program.<sup>52</sup> A total of 39 PD index cases and 19 controls of this isolate were subjected to whole genome-sequencing to explore the genetic factors underlying PD within this geographic region. Missense and LoF variants which were present in at least 2 index cases and a MAF < 0.1% in public databases (ExAC, 1000G dbSNP138 and ESP6500) were considered as potential PD variants. Genes harboring such variants were surveyed for overlap with our list of candidate genes.

#### *Genetic replication 3: Variant aggregation analyses in NeuroX*

We investigated the genetic burden of common and rare variants in these genes by using the independent NeuroX dataset, which is generated by a custom-made genotype array<sup>53</sup> using a backbone of ~240,000 standard Illumina Exome content as a basis with an additional ~24,000 variants that are suggested to be involved neurological diseases. The same procedures as described for the burden test in the IPDGC WES dataset were applied. After QC, a total of 6,801 PD cases and 5,970 neurologically healthy controls remained with high-quality genotype data for 178,779 variants. Based on the sample size the MAF cut-off was 0.0063.

#### *Genetic replication 4: Overlap PD risk loci*

Approximately 70% of the subjects included in this study have also been included in previous published genome-wide association studies (GWAS).<sup>7, 94, 95</sup> To explore the possibility that our candidate genes might also contain common risk variants increasing the risk to develop PD, next to the identified LoF variants with assumed high penetrance, we searched for GWAS loci within 1 Mb up- and downstream of the gene of interest using the recent PD meta-analysis through pdgene.org.<sup>7</sup> Significant associations and suggestive *p*-values < 1e-4 were considered. To understand the underlying linkage disequilibrium structure, LocusZoom<sup>96</sup> was applied to visualize the European 1000G recombination events for the candidate genes that were closely located to a GWAS locus.

### Gene co-expression analyses

We constructed gene co-expression networks (GCN) from two different substantia nigra datasets using the R software package, WGCNA (weighted gene co-expression network analysis).<sup>97</sup> This was followed by the same post-processing of WGCNA gene modules based on k-means: a heuristic to rearrange misplaced genes between modules using the number of modules detected by the standard WGCNA as k and the eigengenes as centroids. The first GCN is based on 19152 genes from 65 substantia nigra control brains from the UKBEC consortium. The gene expression profiles are based on Affymetrix Exon 1.0 ST Arrays<sup>98</sup> The second GCN is based on 63 samples, from the same tissue, GTEx<sup>56</sup> V6 gene RPKM values. Genes were filtered with a RPKM based cut-off of 0.2 and missingness < 30% resulting in the analysis of 18363 Ensembl genes. We corrected this gene expression dataset for the principal components significantly correlated with GTEx samples covariates using the Swamp R package. WGCNA gene modules were functionally annotated with gProfileR<sup>99</sup> R software package using GO database, accounting for multiple testing with gSCS's gProfiler test. Background genes used were all genes in the substantia nigra GCN. Cell type enrichment analysis was performed with the userListEnrichment function with brain specific enrichment, implemented in the WGCNA R package. Preservation analysis of UKBEC GCN in GTEx's substantia nigra profiles was performed with WGCNA's preservation analysis. Results are reported with the Z.summary statistic.<sup>100</sup> Graphical representation of the GCN subnetworks were constructed by using the 27 candidate genes and known PD genes (ATP13A2, FBXO7, LRRK2, PARK2, PARK7, PINK1, RAB39B, SNCA and VPS35) as seed genes. For each of these genes sequentially, in a round robin fashion, we added the gene with highest adjacency, based on TOM values, and the links this gene has with all the seed genes. We used Cytoscape 3.3 for display with a Kamada-kawai layout algorithm.<sup>101</sup>

6

### Human cellular screen

#### *shRNA virus production*

Bacterial glycerol stocks containing the shRNA vectors (Sigma (St. Louis, Missouri), TRC1 and 1.5) were grown overnight in Luria-Bertani media containing 100 µg/ml of ampicillin (Sigma-Aldrich, St. Louis, Missouri). We selected at least 5 shRNA clones per gene. Endotoxin free shRNA plasmids were extracted according to the manufacturer's protocol (Zymo (Irvine, California) ZR Plasmid Miniprep Classic kit). Lentivirus was produced as follows: HEK293T packaging cells were seeded at a density of  $4 \times 10^5$ /ml (100 µl per well) in cell culture media, Optimem (Invitrogen, Carlsbad, California) containing 10% fetal bovine serum (FBS) in 96-well tissue culture plates. Cells were incubated for 24 hours (37°C, 5% CO<sub>2</sub>). Each well was subsequently transfected with 100ng of shRNA plasmid, 90ng of packaging plasmid (pCMV-dr8.74psPAX2) and 10ng of envelope plasmid (VSV-G/pMD2.G) combined with 0.6 µl of FugeneHD (Promega, Madison, Wisconsin) in a total volume of 10 µl. Transfection efficiency was monitored using the pKLO.1 GFP plasmid (Sigma, St. Louis

Missouri) and had to be greater than 90%. Sixteen hours after transfection media was refreshed and supernatant harvested after a further 24 hours. Virus was stored at  $-80^{\circ}\text{C}$ .

To ensure successful lentivirus production, HEK293T cells were plated out at a density of  $2 \times 10^5/\text{ml}$  ( $100 \mu\text{l}$  per well) in Optimem containing 10% FBS and  $15 \mu\text{g}/\text{ml}$  of protamine sulphate (Sigma, St. Louis Missouri). Cells were infected with  $10 \mu\text{l}$ ,  $25 \mu\text{l}$  and  $50 \mu\text{l}$  of lentivirus. The following day, media was refreshed with media containing  $2.5 \mu\text{g}/\text{ml}$  of puromycin. After a further 3 days, plates were manually inspected to determine cell viability of each well. If more than 10% of the wells contained dead cells, lentiviral production for that plate was repeated.

#### *Neuroblastoma cell culture*

BE(2)-M17 (ATCC® CRL-2267™) and HEK 293T (ATCC® CRL-3216™) cell lines were obtained from the American Type Culture Collection (Manassas, Virginia). BE(2)-M17 cell lines were cultured in Dulbecco's Modified Eagle/Nutrient Mixture F-12 Medium (DMEM/F-12) with GlutaMAX (Invitrogen, Carlsbad, California) supplemented with 10% FBS, 1x non-essential amino acids (NEAA) and 1% Penicilin/Streptomycin. HEK 293T cells were cultured in OptiMEM (Invitrogen, Carlsbad, California) containing 10% FBS and 1x NEAA. All cell lines were routinely tested for mycoplasma contamination. For lentivirus infection,  $25 \mu\text{l}$  of the lentivirus was added to each well of a 96-well plates and protamine sulphate was added at a final concentration of  $1 \mu\text{g}/\text{ml}$  in each well of the 96-well plate. Specific wells on each lentiviral plate contained GFP expressing virus to ensure efficient transduction.

#### *Cell-based screening assays*

Four phenotypes were studied in 2 different assays:

Mitochondrial morphology<sup>33</sup> was examined in a single assay with BE(2)-M17 cells, which were expanded and plated at a density of  $5 \times 10^4/\text{ml}$  ( $100 \mu\text{l}$  per well) in 96- well black CellCarrier plates (PerkinElmer, Waltham, Massachusetts) pre-pipetted with  $25 \mu\text{l}$  of the lentivirus. On day 2, media was refreshed with DMEM/F12 (with 10% FBS) supplemented with  $2 \mu\text{g}/\text{mL}$  puromycin. On day 4 the cells were incubated with  $100\text{nM}$  MitoTracker Red CMXros,  $100 \text{nM}$  MitoTracker DeepRed (Molecular Probes) and  $1 \mu\text{g}/\text{mL}$  Hoechst for 20 minutes at room temperature. Media was refreshed and the cells were incubated for a further 2 hours before fixation with 4% paraformaldehyde (pH 7.3). We examined 3 parameters commonly used for quantification of mitochondrial morphology: mitochondrial number, axial length ratio, and roundness.

For the Parkin translocation assay also BE(2)-M17 cells were utilized. The PLVX inducible vector (Clontech, Mountain View, California) overexpressing C-terminally tagged Parkin-GFP was used to make polyclonal stable BE(2)-M17 cells. Stable cell lines were cultured in DMEM/F12 supplemented with 10% FBS, 1% NEAA, 1% P/S,  $250\text{ng}/\text{ml}$  Puromycin,  $200\mu\text{g}/\text{ml}$  G418 and  $1\mu\text{g}/\text{ml}$  of doxycycline. BE(2)-M17 cells were expanded

and plated at a density of  $7.5 \times 10^4$ /ml (100  $\mu$ l per well) in 96-well black CellCarrier plates (PerkinElmer, Waltham, Massachusetts) pre-pipetted with 25  $\mu$ l of the lentivirus. The following day, media was exchanged with media without doxycycline to induce the expression of Parkin-GFP. On day 5 the cells were incubated with 100nM MitoTracker DeepRed (Molecular Probes, Eugene, Oregon) and 1  $\mu$ g/mL Hoechst. After 20 minutes, media was refreshed with media containing 15  $\mu$ M Carbonyl cyanide m-chlorophenyl hydrazone (CCCP). Cells were incubated for 2 hours before fixation in 4% paraformaldehyde (pH 7.3).

#### *Image acquisition and analysis*

Image acquisition was carried out using the automated confocal imaging system, Cell Voyager CV7000 (Yokogawa, Tokyo, Japan). The mitochondrial morphology assay involved a total of 60 fields per well using a 60x water immersion objective lens for improved resolution. Nuclei were imaged utilizing the 405 nm laser and Mitotracker CMXros utilizing the 561 nm laser, mitotracker DeepRed utilizing the 640nm laser. For the translocation assay, a total of 60 fields per well were taken using a 20x objective lens. Nuclei were imaged utilizing the 405 nm laser and Parkin-GFP utilizing the 488 nm laser, mitotracker DeepRed utilizing the 640 nm laser.

Images were stored and analyzed by the Columbus Image Data storage (PerkinElmer, Waltham, Massachusetts). Image quality control: Only well-segmented interphase cells were included. Mitotic, apoptotic badly segmented and out-of-focus cells were excluded. Cells touching the border of the image were removed to avoid analysis of artificially cropped cells. All wells where the perturbation strongly decreased cell number were disregarded. Morphological characteristics and signal intensities were quantified and results exported to R package CellHTS2. To quantify mitochondrial morphology, the median mitochondrial number per object, roundness, axial length ratio and intensity of mitorackerCMXros (mitochondrial potential) were calculated.

To differentiate between CCCP-treated Parkin stable cell lines and untreated cells, the number of spots formed on mitochondria was calculated. Cells containing more than two spots were considered positive for Parkin translocation. The ratio of cells positive for translocation versus the number of cells negative for translocation was calculated per well to give a cell number independent measure of Parkin translocation. CCCP-treated cells transduced with a scrambled shRNA and CCCP-treated cells transduced with shRNA targeting *PINK1* were included on each plate. An average  $Z'$  of 0.61 was calculated for the entire screen, with a minimum Spearman's Rank correlation between replicates of 0.8.

Data from high content imaging assays were analyzed using the BioConductor CellHTS2 package for the R software environment (R version 2.11.1, BioConductor version 2.6). Data was normalized to negative controls on a per plate basis to minimize plate-to-plate variation. For the Parkin-translocation screen, negative controls were considered as



wells which had been transduced with lentivirus encoding a scrambled sequence and had been treated with CCCP. For the remaining screens, negative controls were considered as wells that had been transduced with lentivirus encoding a scrambled sequence.

#### *Statistical analysis*

For each of the shRNA screens, each assay plate was completed with 6 replicates to enable the detection of subtle effects and minimize false negatives. For each shRNA, Mann-Whitney U tests with false discovery rate (FDR) correction were performed and the robust strictly standardized median difference (SSMD\*) was calculated.<sup>102</sup> Effects were considered significant when the SSMD\* normalized effect of shRNA treatment was greater than or less than 4 or -4 and at least 2 independent clones per gene showed a significant effect. Seed sequences were manually inspected to ensure no common sequence.

For each assay, a positive control plate containing known modifiers of the phenotype in question was run in parallel to ensure the assay worked optimally. The robust Z-factor was calculated as previously described,<sup>103</sup> using the normalized values for the controls from all plates. For the mitochondrial assay, known regulators of mitochondrial fission or fusion were included. For the Parkin translocation assay, TOMM7 and PINK1 were used as positive controls.

#### *shRNA knockdown validation*

Cell culture and shRNA mediated knockdown were performed as described above. Cells were harvested for RNA isolation using the SV 96 Total RNA Isolation System (Promega, Madison, Wisconsin) according to the manufacturer's protocol. Total RNA primed with oligo dT (Qiagen, Hilden, Germany) was used for cDNA synthesis with Superscript III RT (Life Technologies, Carlsbad, California) according to the manufacturer's specifications. Quantitative PCR was carried out in triplicates on a ViiA7 real time PCR system using SYBR Green PCR master mix (Life Technologies, Carlsbad, California) and 0.04  $\mu$ M specific primer pairs for all targets. For multiple exons genes primers were designed to span exon-exon junctions or to be separated by one intron on the corresponding genomic DNA. Normalized relative quantities were calculated with HMBS as housekeeping gene by using the qbasePLUS software (Biogazelle, Gent, Belgium) and knockdown efficiencies per clone were calculated using scrambled control wells (N=3) as a reference.

### **Animal models**

#### *Orthologue selection*

The function of the candidate genes and their involvement in neurodegeneration was tested in 2 animal models; *C. elegans* and *Drosophila*. The DRSC Integrated Ortholog Prediction Tool (DIOPT)<sup>104</sup> was used to identify the conserved homologs of human genes in the nematode or fly genomes. Orthologues were defined based on a minimum unweighted

DIOPT score of 2, such that 2 independent bioinformatics algorithms were in agreement concerning the orthologue pairing. In cases where multiple genes were identified as potential orthologues for a given human gene, we carried forward all candidates with DIOPT scores greater than 3.

#### *Fly Stocks and Husbandry*

The human  $\alpha$ -synuclein transgenic flies with codon-optimization for *Drosophila* (UAS- $\alpha$ -synuclein line #7), were recently described,<sup>48</sup> and are available from the Bloomington Stock Center (Bloomington, Indiana). RNAi transgenic lines were obtained from the Vienna *Drosophila* RNAi Centre (Vienna, Austria) or from Bloomington for the Harvard Transgenic RNAi Project. All RNAi lines used for this study are detailed in Supplementary Figure 7. The GAL4-UAS system<sup>105</sup> was used for ectopic co-expression of both the  $\alpha$ -synuclein and RNAi transgene. The *Rh1-Gal4* driver line (second-chromosome insertion) has been previously described.<sup>48, 106</sup> For screening, individual RNAi (IR) lines or Canton S (as a control) were crossed to animals of the genotype: *Rh1-Gal4/CyO; UAS-Syn/TM6B*. All crosses were established at 18°C, and F1 experimental animals (*Rh1-Gal4 / UAS-IR; UAS-Syn / +* or *Rh1-Gal4 / +; UAS-Syn / UAS-IR*) were shifted to 25°C within 24 hours of eclosion and aged 15 days. To examine for potential  $\alpha$ -synuclein independent retinal degeneration, each *UAS-IR* transgenic line was separately crossed to *Rh1-Gal4*, using identical conditions. Based on the results of the primary RNAi screen, we also obtained from Bloomington available mutant alleles for the fly orthologues of *PTPRH*: *Ptp10D* and *Ptp4E*. The following additional stocks were used: (1) *w, Ptp4E<sup>1</sup>*; (2) *w, Ptp10D<sup>1</sup>*; (3) *yw, Ptp4E<sup>1</sup>, Ptp10D<sup>1</sup> / FM7C*. All experimental results were quantified and photographed in female animals.

6

#### *Characterization of Retinal Degeneration in Drosophila*

For optical neutralization (also known as the pseudopupil preparation), fly heads of 15 day-old animals were immersed in mineral oil and transilluminated using a 40x objective on a Leica (Wetzlar, Germany) DM6000B light microscope. Eyes from at least 4 animals were examined per genotype (at least 8 retinæ). All candidate modifier lines and controls were scored blinded by three independent examiners. The penetrance of degeneration caused by each RNAi line was calculated by dividing the number of abnormal retinæ, showing evidence of either reduced rhabdomere numbers or altered refraction of light indicative of vacuolar changes, by the total number of retinæ examined. For identification of genetic enhancers, we required two independent RNAi lines targeting non-overlapping sequences with 50% or greater degenerate retinæ observed using the pseudopupil assay. Following our initial screen of 2 RNAi lines targeting each of 18 fly gene homologs, additional RNAi lines and mutant strains were evaluated, where possible, for the most promising candidates. For each enhancer gene, the strongest RNAi line was independently re-tested for consistency using the pseudopupil assay, and retinal

histologic sections were also performed for further confirmation. To examine for potential  $\alpha$ -synuclein-independent retinal degeneration, the strongest RNAi modifier for each gene was separately crossed to *Rh1-Gal4*, and histologic sections were examined for 15-day old animals. For histology, fly heads from 15 days old animals were fixed in 8% glutaraldehyde and embedded in paraffin. Tangential (3  $\mu$ m) retinal sections were cut using a Leica Microtome (RM2245), and stained with hematoxylin and eosin. Retinae from at least 3 animals were examined and quantified per genotype. Enhancement of  $\alpha$ -synuclein-induced retinal degeneration was quantified based on the severity of retinal vacuolar changes seen in stained histologic sections. We examined representative photographs taken with a 40x objective from well-oriented, intact tangential sections at a depth in which the retina achieves maximal diameter. Using ImageJ software,<sup>107</sup> we recorded the area occupied by all vacuoles with diameter greater than 4  $\mu$ m, and divided by the total retinal area to compute a percentage. Statistical comparisons were implemented using a two-tailed student's t-test.  $\alpha$ -synuclein expression levels were determined by immunoblot (clone 42, BD Transduction Laboratories (San Diego, California), 1:1000).

#### *C.elegans media and strains*

All strains were maintained as described previously.<sup>108</sup> For this study the worm strains N2 (wildtype), CF512 (*fer-15(b26)II*; *fem-1(hc17)III*) and OW40 (*zgl15[P(unc-54):: $\alpha$ -synuclein::YFP]IV*) were used. Strains were grown at 20°C on Nematode Growth medium (NGM) seeded with *Escherichia coli* strain OP50. For each orthologue, one RNAi clone was selected to target the corresponding gene.

#### *Phenotype assays for basal phenotypes in C.elegans*

The systematic RNAi screen was carried out as described.<sup>109</sup> RNAi clones targeting the genes of interest (9 of the 27 – Supplementary Table 3) were obtained from the Vidal cDNA RNAi library or the Ahringer RNAi library. Bacteria expressing the empty vector L4440 were used as negative control. For the survival assay, we employed a sterile strain, CF512 (*fer-15(b26)*; *fem-1(hc17)*).<sup>110</sup> To induce sterility, eggs were collected and kept in M9 medium at 25°C overnight until they reached L1 arrest. Approximately 25 L1 worms were added to plates seeded with RNAi clones of interest and empty vector control and allowed to develop to adults at 25°C. At day 9 of adulthood at 25°C, when approximately half of the worms grown on control plates were dead, the survival of worms on RNAi plates was determined.

The offspring and developmental phenotypes were tested in a single assay. N2 worms were grown at 20°C until L4 stage on OP50 bacteria and then transferred to plates seeded with RNAi clones of interest and empty vector control. At day 2 of adulthood ten worms were put onto a new plate seeded with the same RNAi clone for one hour to produce progeny. The plates containing the progeny were kept at 20°C until the F1 generation of

the control worms reached L4 stage. The number and developmental phenotypes of the offspring were scored at the last time point using a dissecting microscope. A one-sided student t-test was used to determine the significant changes compared to controls. All counting was done in a blind fashion in which the identity of the samples was concealed and each experiment was performed in three biological replicates.

#### *Motility assay for $\alpha$ -synuclein toxicity model in *C. elegans**

Animals were age-synchronized by hypochlorite treatment, hatched overnight in M9 buffer and subsequently cultured on NGM containing isopropylthio- $\beta$ -D-galactoside (IPTG, 15mg/L) and 50 $\mu$ g/ml ampicillin (plates for RNAi treatment). Plates were seeded with RNAi bacteria. Prior to the experiment, the plates were kept at room temperature for 2 days to allow the production of dsRNA by the bacteria. On day 1 of adulthood (one day after larval stage L4), animals were transferred to RNAi plates containing 5-fluoro-2'-deoxy-uridine (FUDR) to prevent the offspring from growing. RNAi clones targeting C54D2.4 (*ARSB*), T08G11.1 (*VPS13C*) and F44G4.8 (*PTPRH*) were used from the Ahringer *C.elegans* RNAi library. All clones were verified by sequencing. RNAi clones for the *C.elegans* orthologue F21F3.7 (*TMEM134*) was not available.

Animals were scored at day 4 and day 8 of adulthood. Animals were placed in a drop of M9 and allowed to adjust for 30 seconds, after which the number of body bends was counted for another 30 seconds. Fifteen animals were scored per condition. Relative body bends were calculated by normalizing to control values. Error bars are showing the standard error of mean. Assays were repeated in 3 independent experiments and the relative body bends of one representative experiment is shown.

6

## ACKNOWLEDGMENTS

We would like to thank all of the subjects who donated their time and biological samples to be a part of this study. This study was supported by the UK Brain Expression Consortium (UKBEC), the French Parkinson's Disease Genetics Study (PDG) and the Drug Interaction with Genes in Parkinson's Disease (DIGPD) study. IJ was funded by the Prinses Beatrix Spierfonds. Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database ([www.ppmi-info.org/data](http://www.ppmi-info.org/data)). For up-to-date information on the study, visit [www.ppmi-info.org](http://www.ppmi-info.org). We also thank the Bloomington Drosophila stock center, the Vienna Drosophila RNAi Center, and the TRiP at Harvard Medical School for providing fly strains.

## REFERENCES

1. Zimprich A, Benet-Pages A, Struhal W, et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *American journal of human genetics* 2011; 89(1): 168-75.
2. Vilarino-Guell C, Wider C, Ross OA, et al. VPS35 mutations in Parkinson disease. *American journal of human genetics* 2011; 89(1): 162-7.
3. Funayama M, Ohe K, Amo T, et al. CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. *Lancet neurology* 2015.
4. Farlow JL, Robak LA, Hetrick K, et al. Whole-Exome Sequencing in Familial Parkinson Disease. *JAMA neurology* 2016; 73(1): 68-75.
5. Shulman JM, De Jager PL, Feany MB. Parkinson's disease: genetics and pathogenesis. *Annual review of pathology* 2011; 6: 193-222.
6. Trinh J, Farrer M. Advances in the genetics of Parkinson disease. *Nature reviews Neurology* 2013; 9(8): 445-54.
7. Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics* 2014; 46(9): 989-93.
8. Hamza TH, Payami H. The heritability of risk and age at onset of Parkinson's disease after accounting for known genetic risk factors. *Journal of human genetics* 2010; 55(4): 241-3.
9. Keller MF, Saad M, Bras J, et al. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Human molecular genetics* 2012; 21(22): 4996-5009.
10. Jonsson T, Atwal JK, Steinberg S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 2012; 488(7409): 96-9.
11. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *The New England journal of medicine* 2013; 368(2): 107-16.
12. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. *The New England journal of medicine* 2013; 368(2): 117-27.
13. Smith BN, Ticozzi N, Fallini C, et al. Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron* 2014; 84(2): 324-31.
14. Cirulli ET, Lasseigne BN, Petrovski S, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science (New York, NY)* 2015; 347(6229): 1436-41.
15. Moutsianas L, Agarwala V, Fuchsberger C, et al. The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. *PLoS genetics* 2015; 11(4): e1005165.
16. Sulem P, Helgason H, Oddson A, et al. Identification of a large set of rare complete human knockouts. *Nature genetics* 2015; 47(5): 448-52.
17. Kitada T, Asakawa S, Hattori N, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* 1998; 392(6676): 605-8.
18. Bonifati V, Rizzu P, Squitieri F, et al. DJ-1( PARK7), a novel gene for autosomal recessive, early onset parkinsonism. *Neurological sciences : official journal of*

- the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* 2003; 24(3): 159-60.
19. Valente EM, Abou-Sleiman PM, Caputo V, et al. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science (New York, NY)* 2004; 304(5674): 1158-60.
  20. Pringsheim T, Jette N, Frolkis A, Steeves TD. The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Movement disorders : official journal of the Movement Disorder Society* 2014; 29(13): 1583-90.
  21. Lesage S, Drouet V, Majounie E, et al. Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. *American journal of human genetics* 2016; 98(3): 500-13.
  22. Ghahramani Seno MM, Kwan BY, Lee-Ng KK, et al. Human PTCHD3 nulls: rare copy number and sequence variants suggest a non-essential gene. *BMC medical genetics* 2011; 12: 45.
  23. Newsome TP, Schmidt S, Dietzl G, et al. Trio combines with dock to regulate Pak activity during photoreceptor axon pathfinding in Drosophila. *Cell* 2000; 101(3): 283-94.
  24. Neumuller RA, Richter C, Fischer A, Novatchkova M, Neumuller KG, Knoblich JA. Genome-wide analysis of self-renewal in Drosophila neural stem cells by transgenic RNAi. *Cell stem cell* 2011; 8(5): 580-93.
  25. Ma XM, Kiraly DD, Gaier ED, et al. Kalirin-7 is required for synaptic structure and function. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2008; 28(47): 12368-82.
  26. Mandela P, Yankova M, Conti LH, et al. Kalrn plays key roles within and outside of the nervous system. *BMC neuroscience* 2012; 13: 136.
  27. Greenamyre JT, Hastings TG. Biomedicine. Parkinson's--divergent causes, convergent mechanisms. *Science (New York, NY)* 2004; 304(5674): 1120-2.
  28. Haelterman NA, Yoon WH, Sandoval H, Jaiswal M, Shulman JM, Bellen HJ. A mitocentric view of Parkinson's disease. *Annual review of neuroscience* 2014; 37: 137-59.
  29. Pickrell AM, Youle RJ. The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. *Neuron* 2015; 85(2): 257-73.
  30. Cookson MR. Parkinsonism due to mutations in PINK1, parkin, and DJ-1 and oxidative stress and mitochondrial pathways. *Cold Spring Harbor perspectives in medicine* 2012; 2(9): a009415.
  31. Narendra D, Tanaka A, Suen DF, Youle RJ. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *The Journal of cell biology* 2008; 183(5): 795-803.
  32. Kamp F, Exner N, Lutz AK, et al. Inhibition of mitochondrial fusion by alpha-synuclein is rescued by PINK1, Parkin and DJ-1. *The EMBO journal* 2010; 29(20): 3571-89.
  33. Koopman WJ, Visch HJ, Smeitink JA, Willems PH. Simultaneous quantitative measurement and automated analysis of mitochondrial morphology, mass, potential, and motility in living human skin fibroblasts. *Cytometry Part A : the journal of the International Society for Analytical Cytology* 2006; 69(1): 1-12.
  34. Chang CR, Blackstone C. Dynamic regulation of mitochondrial fission through modification of the dynamin-related protein Drp1. *Annals of the New York Academy of Sciences* 2010; 1201: 34-9.
  35. Narendra DP, Jin SM, Tanaka A, et al.

- PINK1 is selectively stabilized on impaired mitochondria to activate Parkin. *PLoS biology* 2010; 8(1): e1000298.
36. Vives-Bauza C, Zhou C, Huang Y, et al. PINK1-dependent recruitment of Parkin to mitochondria in mitophagy. *Proceedings of the National Academy of Sciences of the United States of America* 2010; 107(1): 378-83.
  37. Geisler S, Holmstrom KM, Skujat D, et al. PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nature cell biology* 2010; 12(2): 119-31.
  38. Vincow ES, Merrihew G, Thomas RE, et al. The PINK1-Parkin pathway promotes both mitophagy and selective respiratory chain turnover in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 2013; 110(16): 6400-5.
  39. Feany MB, Bender WW. A Drosophila model of Parkinson's disease. *Nature* 2000; 404(6776): 394-8.
  40. Auluck PK, Chan HY, Trojanowski JQ, Lee VM, Bonini NM. Chaperone suppression of alpha-synuclein toxicity in a Drosophila model for Parkinson's disease. *Science (New York, NY)* 2002; 295(5556): 865-8.
  41. MacLeod DA, Rhinn H, Kuwahara T, et al. RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. *Neuron* 2013; 77(3): 425-39.
  42. Chen L, Feany MB. Alpha-synuclein phosphorylation controls neurotoxicity and inclusion formation in a Drosophila model of Parkinson disease. *Nature neuroscience* 2005; 8(5): 657-63.
  43. Cullen V, Lindfors M, Ng J, et al. Cathepsin D expression level affects alpha-synuclein processing, aggregation, and toxicity in vivo. *Molecular brain* 2009; 2: 5.
  44. Petrucelli L, O'Farrell C, Lockhart PJ, et al. Parkin protects against the toxicity associated with mutant alpha-synuclein: proteasome dysfunction selectively affects catecholaminergic neurons. *Neuron* 2002; 36(6): 1007-19.
  45. Yang Y, Nishimura I, Imai Y, Takahashi R, Lu B. Parkin suppresses dopaminergic neuron-selective neurotoxicity induced by Pael-R in Drosophila. *Neuron* 2003; 37(6): 911-24.
  46. Miura E, Hasegawa T, Konno M, et al. VPS35 dysfunction impairs lysosomal degradation of alpha-synuclein and exacerbates neurotoxicity in a Drosophila model of Parkinson's disease. *Neurobiology of disease* 2014; 71: 1-13.
  47. Dhungel N, Eleuteri S, Li LB, et al. Parkinson's disease genes VPS35 and EIF4G1 interact genetically and converge on alpha-synuclein. *Neuron* 2015; 85(1): 76-87.
  48. Chouhan AK, Guo C, Hsieh Y-C, et al. Uncoupling neuronal death and dysfunction in Drosophila models of neurodegenerative disease. *Acta Neuropathol Comm* In press.
  49. van Ham TJ, Holmberg MA, van der Goot AT, et al. Identification of MOAG-4/SERF as a regulator of age-related proteotoxicity. *Cell* 2010; 142(4): 601-12.
  50. Jeon M, Nguyen H, Bahri S, Zinn K. Redundancy and compensation in axon guidance: genetic analysis of the Drosophila Ptp10D/Ptp4E receptor tyrosine phosphatase subfamily. *Neural development* 2008; 3: 3.
  51. Parkinson Progression Marker Initiative The Parkinson Progression Marker Initiative (PPMI). *Progress in neurobiology* 2011; 95(4): 629-35.
  52. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS. The effect of genetic drift in a young genetically isolated

- population. *Annals of human genetics* 2005; 69(Pt 3): 288-95.
53. Nalls MA, Bras J, Hernandez DG, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. *Neurobiology of aging* 2014.
  54. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics* 2010; 42(7): 565-9.
  55. Xu J, Bai J, Zhang X, et al. A comprehensive overview of lncRNA annotation resources. *Briefings in bioinformatics* 2016.
  56. GTEx-Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, NY)* 2015; 348(6235): 648-60.
  57. Aravind L, Koonin EV. G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends in biochemical sciences* 1999; 24(9): 342-4.
  58. Lin ML, Fukukawa C, Park JH, et al. Involvement of G-patch domain containing 2 overexpression in breast carcinogenesis. *Cancer science* 2009; 100(8): 1443-50.
  59. Hu F, Gou L, Liu Q, Zhang W, Luo M, Zhang X. G-patch domain containing 2, a gene highly expressed in testes, inhibits nuclear factor-kappaB and cell proliferation. *Molecular medicine reports* 2015; 11(2): 1252-7.
  60. Otto GP, Razi M, Morvan J, Stenner F, Tooze SA. A novel syntaxin 6-interacting protein, SHIP164, regulates syntaxin 6-dependent sorting from early endosomes. *Traffic (Copenhagen, Denmark)* 2010; 11(5): 688-705.
  61. Wang S, Bellen HJ. The retromer complex in development and disease. *Development (Cambridge, England)* 2015; 142(14): 2392-6.
  62. Wang W, Wang X, Fujioka H, et al. Parkinson's disease-associated mutant VPS35 causes mitochondrial dysfunction by recycling DLP1 complexes. *Nature medicine* 2016; 22(1): 54-63.
  63. Song P, Trajkovic K, Tsunemi T, Krainc D. Parkin Modulates Endosomal Organization and Function of the Endo-Lysosomal Pathway. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2016; 36(8): 2425-37.
  64. Chen L, Xie Z, Turkson S, Zhuang X. A53T human alpha-synuclein overexpression in transgenic mice induces pervasive mitochondria macroautophagy defects preceding dopamine neuron degeneration. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2015; 35(3): 890-905.
  65. Matozaki T, Suzuki T, Uchida T, et al. Molecular cloning of a human transmembrane-type protein tyrosine phosphatase and its expression in gastrointestinal cancers. *The Journal of biological chemistry* 1994; 269(3): 2075-81.
  66. Matozaki T, Murata Y, Mori M, Kotani T, Okazawa H, Ohnishi H. Expression, localization, and biological function of the R3 subtype of receptor-type protein tyrosine phosphatases in mammals. *Cellular signalling* 2010; 22(12): 1811-7.
  67. Takahashi H, Craig AM. Protein tyrosine phosphatases PTPdelta, PTPsigma, and LAR: presynaptic hubs for synapse organization. *Trends in neurosciences* 2013; 36(9): 522-34.
  68. Qian M, Pan G, Sun L, et al. Receptor-like tyrosine phosphatase PTP10D is required for long-term memory in Drosophila. *The Journal of neuroscience : the official journal of the Society for Neuroscience*



- 2007; 27(16): 4396-402.
69. Lee HK, Cording A, Vielmetter J, Zinn K. Interactions between a receptor tyrosine phosphatase and a cell surface ligand regulate axon guidance and glial-neuronal communication. *Neuron* 2013; 78(5): 813-26.
  70. Valayannopoulos V, Nicely H, Harmatz P, Turbeville S. Mucopolysaccharidosis VI. *Orphanet journal of rare diseases* 2010; 5: 5.
  71. Tessitore A, Pirozzi M, Auricchio A. Abnormal autophagy, ubiquitination, inflammation and apoptosis are dependent upon lysosomal storage and are useful biomarkers of mucopolysaccharidosis VI. *PathoGenetics* 2009; 2(1): 4.
  72. Lieberman AP, Puertollano R, Raben N, Slaugenhaupt S, Walkley SU, Ballabio A. Autophagy in lysosomal storage disorders. *Autophagy* 2012; 8(5): 719-30.
  73. Brooks DA, Gibson GJ, Karageorgos L, Hein LK, Robertson EF, Hopwood JJ. An index case for the attenuated end of the mucopolysaccharidosis type VI clinical spectrum. *Molecular genetics and metabolism* 2005; 85(3): 236-8.
  74. Karageorgos L, Brooks DA, Pollard A, et al. Mutational analysis of 105 mucopolysaccharidosis type VI patients. *Human mutation* 2007; 28(9): 897-903.
  75. Sidransky E, Nalls MA, Aasly JO, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *The New England journal of medicine* 2009; 361(17): 1651-61.
  76. Sidransky E, Lopez G. The link between the GBA gene and parkinsonism. *Lancet neurology* 2012; 11(11): 986-98.
  77. Exome Aggregation Consortium (ExAC) C, MA (URL: <http://exac.broadinstitute.org>) [April 2015]. 2015.
  78. Lek M, Karczewski K, Minikel E, et al. Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv* 2015.
  79. Dawson TM, Ko HS, Dawson VL. Genetic animal models of Parkinson's disease. *Neuron* 2010; 66(5): 646-61.
  80. Shulman JM. Drosophila and experimental neurology in the post-genomic era. *Experimental neurology* 2015.
  81. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 2009; 25(14): 1754-60.
  82. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* 2011; 43(5): 491-8.
  83. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* 2010; 38(16): e164.
  84. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM. A general framework for estimating the relative pathogenicity of human genetic variants. 2014; 46(3): 310-5.
  85. Klein C, Westenberger A. Genetics of Parkinson's disease. *Cold Spring Harbor perspectives in medicine* 2012; 2(1): a008888.
  86. Grunewald A, Kasten M, Ziegler A, Klein C. Next-generation phenotyping using the parkin example: time to catch up with genetics. *JAMA neurology* 2013; 70(9): 1186-91.
  87. Exome Variant Server NGSPE, Seattle, WA (URL: <http://evs.gs.washington.edu/EVS/>) [September 2013 and April 2015].
  88. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; 491(7422): 56-65.
  89. Li MX, Gui HS, Kwan JS, Bao SY, Sham PC. A

- comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic acids research* 2012; 40(7): e53.
90. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic acids research* 2001; 29(1): 308-11.
  91. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nature biotechnology* 2011; 29(1): 24-6.
  92. Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. *American journal of human genetics* 2013; 92(6): 841-53.
  93. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007; 81(3): 559-75.
  94. Simon-Sanchez J, Schulte C, Bras JM, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature genetics* 2009; 41(12): 1308-12.
  95. International Parkinson's Disease Genomics Consortium, Wellcome Trust Case Control Consortium2. A two-stage meta-analysis identifies several new loci for Parkinson's disease. *PLoS genetics* 2011; 7(6): e1002142.
  96. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics (Oxford, England)* 2010; 26(18): 2336-7.
  97. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* 2008; 9: 559.
  98. Forabosco P, Ramasamy A, Trabzuni D, et al. Insights into TREM2 biology by network analysis of human brain gene expression data. *Neurobiology of aging* 2013; 34(12): 2699-714.
  99. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic acids research* 2007; 35(Web Server issue): W193-200.
  100. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? *PLoS computational biology* 2011; 7(1): e1001057.
  101. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 2003; 13(11): 2498-504.
  102. Zhang XD. Illustration of SSMD, z score, SSMD\*, z\* score, and t statistic for hit selection in RNAi high-throughput screens. *Journal of biomolecular screening* 2011; 16(7): 775-85.
  103. Zhang JH, Chung TD, Oldenburg KR. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *Journal of biomolecular screening* 1999; 4(2): 67-73.
  104. Hu Y, Flockhart I, Vinayagam A, et al. An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC bioinformatics* 2011; 12: 357.
  105. Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development (Cambridge, England)* 1993; 118(2): 401-15.
  106. Xiong B, Bayat V, Jaiswal M, et al. Crag is a GEF for Rab11 required for rhodopsin trafficking and maintenance of adult photoreceptor cells. *PLoS biology* 2012; 10(12): e1001438.
  107. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image

- analysis. *Nature methods* 2012; 9(7): 671-5.
108. Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics* 1974; 77(1): 71-94.
109. Hansen M, Hsu AL, Dillin A, Kenyon C. New genes tied to endocrine, metabolic, and dietary regulation of lifespan from a *Caenorhabditis elegans* genomic RNAi screen. *PLoS genetics* 2005; 1(1): 119-28.
110. Garigan D, Hsu AL, Fraser AG, Kamath RS, Ahringer J, Kenyon C. Genetic analysis of tissue aging in *Caenorhabditis elegans*: a role for heat-shock factor and bacterial proliferation. *Genetics* 2002; 161(3): 1101-12.

## SUPPLEMENTAL DATA

*Complete legend for figure 5 in main body of the paper:*

**Figure 5.** PD gene candidates harboring LoF variants enhance  $\alpha$ -synuclein toxicity in *Drosophila*. Conserved fly orthologs of human genes discovered from WES analysis were targeted with RNAi (IR) and screened for enhancement of  $\alpha$ -synuclein pathology using the pseudopupil assay (A, top row). For each line evaluated, the severity of retinal degeneration was scored based on penetrance of the  $\alpha$ -synuclein pseudopupil phenotype, and enhancers required consistent results for at least 2 independent RNAi lines (see Additional Table 7). Representative results from the primary screen are shown for controls (Rh1-GAL4 / +; UAS- $\alpha$ -synuclein / +) and one IR line each for the implicated enhancers [Human Gene-Fly Ortholog (experimental genotype shown)]: ARSB-CG32191 (Rh1-GAL4 / +; UAS- $\alpha$ -synuclein / UAS-CG32191.IR.v14294), TMEM134-CG12025 (Rh1-GAL4 / UAS-CG12025.IR.v104336; UAS- $\alpha$ -synuclein / +), PTPRH-Ptp10D (Rh1-GAL4 / UAS-Ptp10D.IR.v1102; UAS- $\alpha$ -synuclein / +), and VPS13-Vps13 (Rh1-GAL4 / UAS-Vps13.IR.HMS02460; UAS- $\alpha$ -synuclein / +). At the 15-day old time point, Rh1> $\alpha$ -synuclein causes a weakly-penetrant pseudopupil phenotype and mild histopathologic changes which are amenable to modifier screening (compare with Figure 4, panels C & E). Enhancers identified in the primary screen were confirmed based on retinal histology (A, middle row), and demonstrated increased tissue destruction and disorganization. Activation of RNAi was not associated with any significant retinal degeneration in the absence of  $\alpha$ -synuclein co-expression (A, bottom row, Rh1-GAL4 / IR transgene). Scale bars: 20  $\mu$ m. (B) Enhancement of  $\alpha$ -synuclein-induced retinal degeneration was quantified based on the extent of vacuolar changes (area occupied by vacuoles / total retinal area). For quantification, 3 animals were examined per genotype. For PTPRH, additional confirmation was obtained by evaluating flies doubly-heterozygous for strong alleles of the paralogs Ptp10D and Ptp4e (see also Supplemental Figure 3). Statistical comparisons were made using unpaired t-tests. Error bars are based on Standard Error of the Mean. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ .

**Table 1.** Homozygous variants in known Mendelian PD genes.

	<b>PD patient 1</b>	<b>PD patient 2</b>
Gene	<i>PINK1</i>	<i>parkin</i>
Position (hg 19)	chr1:20972118	chr6:162864357
Variant	GCT GCT GCA del.	T del.
AAchange	del342-345	del52-
ExAC MAF	0	0.0003
Domain	Protein kinase domain	ubiquitin domain

AAchange = amino acid change; ExAC MAF = minor allele frequency in ExAC database; Domain = identified variant located within the specified functional domain.

6

**Table 2.** Human knock-outs overlap with Icelandic population.

Type of overlap	Gene	Chr	Pos (hg19)	Ref all	Alt all	Variant Type	AA change	Icelandic study									
								MAF EUR	MAF ExAC	MAF FIN	MAF	n CH	n HZ	% HZ	n de-ceased	high-est death	high-est age
	CAP52	12	75687045	C	CTATTT-GAT	stopgain	NM_001286547:p.L321X	0.0068	0.011	1.08	1	5	0.0048%	0	NA	74	
	COL6A5	3	130132401	C	T	stopgain	NM_001278298:p.Q1559X	0.0012	0	0.55	44	8	0.0077%	2	49	93	
	FAM71A	1	212799882	A	T	stopgain	NM_153606:p.K555X	0.008	0.0036	0.46	0	7	0.0067%	1	78	78	
	FAM83A	8	124195352	G	T	stopgain	NM_032899:p.G86X	0.0046	0.00061	0.51	0	2	0.0019%	0	NA	55	
	GH2	17	61957946	C	T	stopgain	NM_022557:p.W214X	0.0033	0.0024	0.34	0	3	0.0029%	2	76	83	
	GPA7CH2L	14	76644266	C	T	stopgain	NM_017972:p.R362X	0.0053	0.00030	0.44	0	1	0.0010%	1	84	84	
	KCNK16	6	39282816	G	A	stopgain	NM_001135107:p.Q251X	0.0044	0.0047	0.84	0	9	0.0086%	0	NA	87	
	PTCHD3	10	27688101	G	A	stopgain	NM_001034842:p.R476X	0.0097	0.0048	2.12	NA	29	0.028%	4	73	96	
	PTPRH	19	55697712	G	A	stopgain	NM_002842:p.O887X	0.0039	0.00045	0.13	1	0	0%	NA	NA	NA	
	PZP	12	9321534	G	A	stopgain	NM_002864:p.R680X	0.0085	0.010	0.85	0	3	0.0029%	0	NA	52	
	TMEM134	11	67235051	G	A	stopgain	NM_001078650:p.R84X	0.0070	0.0060	1.04	0	13	0.012%	3	76	102	
	ZNF543	19	57838058	G	A	stopgain	NM_213598:p.W68X	0.0060	0.00091	0.34	0	1	0.0010%	1	73	73	
	ANKRD30A	10	37419292	G	T	stopgain	NM_052997:p.E110X	0.018	0.0082	1.82	14	36	0.035%	9	61	93	
	ANKRD30A	10	37482182	AG	A	frameshift	NM_052997:p.A815fs	0.031	0.031	3.23	NA	107	0.10%	14	43	96	
	ANKRD30A	10	37508578	CTG	C	frameshift	NM_052997:p.E1258fs	0.0042	0.0029	0.46	14	4	0.0038%	0	NA	60	
	C11orf21	11	2323089	C	T	startloss	NM_001142946:p.M1I	0.016	0.0098	0.76	0	10	0.0096%	2	81	87	
	CAP52	12	75698657	C	T	splicing	NM_001286547:c.567-1G>A	x	x	0.10	1	0	0%	NA	NA	NA	
	CD36	7	80303460	ATAAG	A	frameshift	NM_001127444:p.473_473del	0.017	0.023	2.52	NA	70	0.067%	4	68	95	
	COL6A5	3	130107852	AC	A	stopgain	NM_001278298:p.Y764X	x	x	0.04	4	0	0%	NA	NA	NA	
	COL6A5	3	130116522	CCAGG-CATCT-TAGAAGA	C	frameshift	NM_001278298:p.1222_1227del	0.00012	0	0.15	13	0	0%	NA	NA	NA	
Other variants	COL6A5	3	130116524	AGG-CATCT-TAGAAG-ACA	A	frameshift	NM_001278298:p.1223_1228del	x	x	0.15	13	0	0%	NA	NA	NA	
	COL6A5	3	130159330	C	T	stopgain	NM_001278298:p.Q2050X	0.0079	0.022	0.89	36	5	0.0048%	1	80	80	
	COL6A5	3	130187662	G	T	stopgain	NM_001278298:p.E272X	0.012	0.0083	1.52	65	22	0.021%	3	55	92	
	FAM83A	8	124195217	GC	G	frameshift	NM_032899:p.A41fs	x	x	0.06	1	0	0%	NA	NA	NA	

FAM83A	8	124219662	C	T	stopgain	NM_032899.p.R347X	0.0023	0.0056	0.24	1	0	0%	NA	NA	NA
OR7G3	19	9236698	G	GATGGT	frameshift	NM_001001958p.S310fs	0.31	0.31	30.62	NA	9550	9.16%	1443	23	107
OR7G3	19	9236916	AG	A	frameshift	NM_001001958.p.A237fs	0.051	0.037	4.80	NA	244	0.23%	42	38	103
PTCHD3	10	27687225	A	G	stoploss	NM_001034842.p.X768Q	0.69	0.69	41.11	NA	17252	16.55%	2524	18	107
PTCHD3	10	27687759	CT	C	frameshift	NM_001034842.p.R589fs	0.016	0.0051	2.20	NA	49	0.047%	8	54	100
PTCHD3	10	27702256	G	GC	frameshift	NM_001034842.p.G308fs	0.35	0.31	29.93	NA	9827	9.43%	1477	26	107
PTPRH	19	55710079	T	TA	frameshift	NM_002842.p.Y541fs	0.000045	0	0.16	1	0	0%	NA	NA	NA
SSPO	7	149475918	TG	T	frameshift	NM_198455.p.L295fs	0.0072	0.0031	0.93	17	10	0.0096%	3	81	86
SSPO	7	149487608	T	G	splicing	NM_198455.c.4921+1T>G	0.13	.	7.35	NA	426	0.41%	69	47	97
SSPO	7	149503916	GGATA	G	frameshift	NM_198455.p.2915_2915del	0.24	0.28	15.25	NA	2374	2.28%	371	24	103
SSPO	7	149515188	TA	T	splicing	NM_198455.c.11577+2A>	0.38	0.39	31.81	NA	9771	9.38%	1438	23	107
SSPO	7	149516838	TC	T	frameshift	NM_198455.p.S4014fs	0.013	0	0.92	16	5	0.0048%	1	77	84

*n* CH = number of compound heterozygotes in a variant pair with MAF<2% for both variants; *n* HZ = number of observed homozygotes; *n* deceased = number of deceased homozygotes; earliest death = earliest reported death in years among homozygotes; highest age = highest reported age in years among homozygotes.

**Table 3.** C. elegans basal phenotype assays.

Human gene	Worm gene	DIOPT score	Development	Egg laying (p)	Survival rate	Wormbase
<i>ARSB</i>	<i>sul-3</i>	8				-
<i>CALML4</i>	<i>T09B4.4</i>	2				-
<i>CD36</i>	<i>scav-2</i>	5				cell death variant;slow growth
	<i>scav-1</i>	5				pathogen induced death increased
<i>DIS3</i>	<i>dis-3</i>	9	L1-L2	0.035	↓	slow growth;small;sterile progeny
<i>KALRN</i>	<i>unc-73</i>	7			↓	axon;egg laying;embryo lethal;movement
<i>PTCHD3</i>	<i>ptr-2</i>	7		0.0003		early larval;embryo lethal;lethal;movement
	<i>daf-6</i>	5		0.05		sensory neuroanatomy
	<i>ptr-10</i>	5			↓	body vacuole;early larval arrest
	<i>ptr-19</i>	5				-
<i>PTPRH</i>	<i>dep-1</i>	3				-
<i>TMEM134</i>	<i>(F21F3.7)</i>	6				-
<i>UHRF1BP1L</i>	<i>C44H4.4</i>	6				egg laying;embryo lethal;movement
<i>VPS13C</i>	<i>T08G11.1</i>	6				accumulated cell corpses

Development = worm shows developmental arrest during offspring screen; ↓ = worms live shorter; Egg laying (p) = decrease in number of offspring expressed in a mean p-value over the 3 repeats (significance level post Bonferonni correction = 0.0019). Orthologues with a minimum DIOPT score of 2 were included in *C. elegans* screen. In cases where multiple genes were identified as potential orthologues for a given human gene, we carried forward all candidates with DIOPT scores greater than 3. The worm gene in brackets was not screened as it was not included in RNAi library.

6

**Table 4.** Variant aggregation association results of IPDGC WES dataset.

Variant type	variant set	BURDEN		SKAT	
		p-value (emp.)	# var	p-value (emp.)	# var
Rare only	CADD	<i>ZNF543</i> : 0.015 (0.017)	2	<i>ZNF543</i> : 0.015 (0.021)	2
	missense	<i>GH2</i> : 0.026 (0.042)	4	<i>GH2</i> : 0.038 (0.019)	4
		<i>UHRF1BP1L</i> : 0.013 (0.016)	19	<i>PTPRH</i> : 0.034 (0.034)	40
Common and rare	missense			<i>PTPRH</i> : 0.013 (0.021)	46
				<i>ZNF543</i> : 0.042 (0.045)	16

BURDEN = one-sided burden test assuming all variants to have the same direction and magnitude of effect; SKAT = two-sided SKAT test allowing variants to have different directions and magnitudes of effect. P-value (emp.) = Nominal p-value (empirical p-value). Variant categories are only displayed if significant associations are present.



**Table 5.** Overview of the cellular screens for 15 hits.

Gene Symbol	Parkin Translocation	Mitochondrial Morphology		
	pos:neg	Mitochondria per cell	Axial length ratio	Roundness
<i>DIS3</i>	-	-24.280	-	-
<i>FAM71A</i>	-	-	-0.022	-0.027
<i>GPATCH2L</i>	-0.228	-	-	0.020
<i>KALRN</i>	-0.218	-	-	-
<i>KCNK16</i>	-	-	-0.023	-0.027
<i>MNS1</i>	-	-	-0.021	-
<i>OR7G3</i>	-	5.434	-0.019	0.007
<i>PCDHA9</i>	-	-	-0.019	-0.023
<i>PTCHD3</i>	-0.193	-	0.013	0.0180
<i>PTPRH</i>	-	-	0.002	-
<i>SVOPL</i>	0.204	-	0.013	-0.007
<i>TCHHL1</i>	-0.168	-	-	-
<i>UHRF1BP1L</i>	-	-24.585	-0.020	-0.0323
<i>VPS13C</i>	-	-	-0.022	-
<i>ZNF543</i>	-0.263	-2.557	-	-

The effects of knockdown on parkin translocation and mitochondrial morphology are given for 15 potential PD genes. Genes without an effect are not reported. Hits are defined meeting three criteria: FDR-corrected  $p$ -value  $< 0.05$ , SSMD\* of  $-4 \geq x \geq 4$  and at least 2 or more shRNA clones per gene. Gene symbols are indicated in the first column. For Parkin translocation, pos:neg indicates the normalized ratio of cells positive for translocation versus cells negative for translocation. A negative value indicates inhibition of Parkin translocation and a positive value indicates an enhancement of Parkin translocation. Mitochondria numbers are median values of significant shRNA clones per gene. Increase (positive) or decrease (negative) in mitochondria axial length ratio and roundness is based on median values of significant shRNA clones compared to scrambled control wells. Axial length ratio: ratio of the shorter to the longer axial lengths. Roundness: positive values indicate increase in mitochondrial roundness and negative values indicate increased branching. - : using above mentioned criteria, the gene was not hit for this particular gene/phenotype combination. shRNA validation indicates the percentage of knockdown for the strongest clone. ND = not determined. - = failed assay.

**Table 6.** Conserved *Drosophila* homologs for candidate genes.

Human	Fly	CG	DIOPT score	Algorithms
<i>SSPO</i>	Hml	CG7002	2	orthoMCL, RoundUp
<i>PTCHD3</i>	Ptr	CG11212	6	Compara, Homologene, Isobase, OrthoDB, orthoMCL, RoundUp
<i>ARSB</i>	CG7402	CG7402	8	Compara, Homologene, Inparanoid, Isobase, OrthoDB, orthoMCL, Phylome, Treefam
	CG7408	CG7408	6	Compara, Inparanoid, OrthoDB, orthoMCL, Phylome, Treefam
	CG32191	CG32191	6	Compara, Inparanoid, OrthoDB, orthoMCL, Phylome, Treefam
<i>CAPS2</i>	CG8646	CG8646	5	Compara, OrthoDB, Phylome, RoundUp, Treefam
	CG10126	CG10126	3	Compara, Isobase, Phylome
<i>KALRN</i>	trio	CG18214	7	Inparanoid, OMA, OrthoDB, orthoMCL, Phylome, RoundUp, Treefam
<i>TMEM134</i>	CG12025	CG12025	8	Compara, Inparanoid, Isobase, OrthoDB, orthoMCL, Phylome, RoundUp, Treefam
<i>UHRF1BP1L</i>	CG34126	CG34126	7	Compara, Homologene, Inparanoid, orthoMCL, Phylome, RoundUp, Treefam
<i>PTPRH</i>	Ptp10D	CG1817	3	OrthoDB, Phylome, Treefam
	Ptp4E	CG6899	3	OrthoDB, Phylome, Treefam
<i>CD36</i>	CG3829	CG3829	2	Isobase, Phylome
	Snmp2	CG7422	3	Compara, Phylome, Treefam
<i>VPS13C</i>	Vps13	CG2093	6	Compara, OrthoDB, orthoMCL, Phylome, RoundUp, Treefam
<i>MNS1</i>	CG7352	CG7352	2	orthoMCL, Phylome
<i>CALML4</i>	CG17272	CG17272	3	Isobase, OMA, RoundUp
<i>DIS3</i>	Dis3	CG6413	9	Compara, Homologene, Inparanoid, Isobase, OMA, OrthoDB, orthoMCL, Phylome, Treefam

**Table 7.**  $\alpha$ Syn-mediated neurodegeneration in the retina.

Category	Human	Fly	RNAi lines	Penetrance (%)		
Hits	<i>ARSB</i>	CG7402	v37302	75.0		
			v103947	0.0		
		CG7408	v8415	87.5		
			v107353	12.5		
		<u>CG32191</u>	v14294	100.0		
			v101578	50.0		
		CG8646	v38092	50.0		
			v110479	25.0		
		<i>TMEM134</i>	<u>CG12025</u>	v4098	62.5	
				v104336	87.5	
		<i>PTPRH</i>	<u>Ptp10D</u>	v1102*	87.5	
				v110443	50.0	
				v8010*	75.0	
				v1104*	100.0	
				Ptp4E	v1012*	100.0
					v27232	37.5
VPS13C	<u>Vps13</u>			HMS01715	75.0	
				HMS02460	87.5	
All others	<i>SSPO</i>	Hml	v37006	10.0		
			v108935	10.0		
	<i>PTCHD3</i>	Ptr	HM05184	0.0		
			CAPS2	CG10126	v44104	87.5
	v102435	44.4				
	<i>KALRN</i>	trio	v40137*	0.0		
			v40138*	87.5		
	<i>UHRF1BP1L</i>	CG34126	v26336	50.0		
			v107307	0.0		
	<i>CD36</i>	CG3829	v42872	50.0		
			v103492	44.4		
			Snmp2	v27997	62.5	
				v101136	37.5	
	<i>CALML4</i>	CG17272	v32874	62.5		
			v101784	0.0		
	<i>DIS3</i>	Dis3	v35090	20.0		
v108013			62.5			
<i>MNS1</i>	CG7352	v40636*	100.0			
		v107210*	50.0			

RNAi transgenic lines with penetrance  $\geq 50\%$  are shown in bold. For enhancers, we required evidence from 2 independent RNAi lines targeting non-overlapping sequences. \*Denotes RNAi targeting significantly overlapping or identical sequences within the gene transcript.

**Table 8.** Recessive LoF variants in PPMI exome dataset.

Gene	State	chr:bp (chr19)	ExAC (EUR)	ExAC # HZ	AAchange	Overlap IPDGC
<i>GPATCH2L</i>	HZ	14:76644266	0.53%	1	R362X	<b>yes (HZ)</b>
<i>FAM83A</i>	CH	8:124195352	0.46%	2	G86X	<b>yes (CH)</b>
		8:124219662	0.23%	x	R347X	<b>no</b>
<i>PSG9</i>	HZ	19:43763057	0.020%	1	D314fsdel	<b>no</b>

HZ = homozygous; CH = putative compound heterozygous; ExAC (EUR) = MAF in European population according to ExAC; ExAC # HZ = frequency of homozygous variants in European population according to ExAC; AAchange = amino acid change; fsdel = frameshift deletion; del = deletion.

**Table 9.** Variant aggregation association results of IPDGC neuroX dataset.

Variant type	variant set	BURDEN		SKAT	
		p-value (emp.)	# var	p-value (emp.)	# var
Rare only	missense	<i>UHRF1BP1</i> : 0.021 (0.021)	18	<i>UHRF1BP1</i> : 0.021 (0.019)	<b>18</b>
		<i>PTPRH</i> : 0.042 (0.045)	12	<i>ARSB</i> : 0.031 (0.029)	<b>11</b>
Common and rare	CADD	<i>VPS13C</i> : 0.023 (0.022)	18	<i>VPS13C</i> : 0.029 (0.030)	<b>18</b>
	missense	<i>UHRF1BP1</i> : 0.006 (0.005)	21	<i>UHRF1BP1</i> : 0.005 (0.005) <i>ARSB</i> : 0.039 (0.041)	<b>21</b> <b>14</b>

BURDEN = one-sided burden test assuming all variants to have the same direction and magnitude of effect; SKAT = two-sided SKAT test allowing variants to have different directions and magnitudes of effect. P-value (emp.) = Nominal p-value (empirical p-value). Variant categories are only displayed if significant associations present.

6

**Table 10 - part 1.** Gene co-expression analysis of candidate and established PD genes.

Gene symbol	Module	Module Size	Member-ship	SN UKBEC	SN GTEX	Preservation GTEX	Enrichment cell specific markers	Enrichment most relevant GO terms
<i>GPATCH2L</i>	brown	1081	0.46	Yes	Yes	12.37	Neuron in Human brain (Geschwind,2010) ( $p=3.413e-12$ ) Neuron in Cortex ( $p=4.286e-13$ )	Cellular macromolecule metabolic process ( $p=3.82e-15$ ) Macromolecule metabolic process ( $p=1.45e-14$ )
<i>PARK7</i>			0.517	Yes	Yes		Probably Neuron (Cahoy, 2008) ( $p=4.128e-09$ )	Cellular metabolic process ( $p=3.21e-14$ ) Organic substance metabolic process ( $p=1.47e-13$ ) Primary metabolic process ( $p=1.75e-13$ )
<i>PTPRH</i>	green	1461	0.638	Yes	Yes	31.36	Oligodendrocyte from conservative data set (Lein, 2007) ( $p=8.692e-22$ ) Oligodendrocytes in Human brain (Geschwind,2010) ( $p=5.19e-198$ ) Oligodendrocytes in Cortex ( $p=8.426e-107$ ) Definite Oligodendrocyte (Cahoy, 2008) ( $p=1.067e-37$ )	No significant enrichment
<i>FBX07</i>			0.644	Yes	Yes			
<i>UHRF1BP1L</i>	yellow	1973	0.821	Yes	Yes	53.77	Astrocytes in Cortex ( $p=8.182e-08$ ) Neuron in Human brain (Geschwind,2010) ( $p=3.984e-46$ ) Neuron in Cortex ( $p=1.937e-29$ ) Pyramidal Neurons in network from Sugino/Winden ( $p=1.145e-05$ ) Definite Neuron (Cahoy, 2008) ( $p=6.955e-19$ )	Transport ( $p=1.87e-16$ ) Establishment of localization ( $p=1.72e-15$ ) Single-organism transport ( $p=6.88e-14$ ) Single-organism localization ( $p=1.25e-13$ ) Synaptic transmission ( $p=2.27e-11$ )
<i>ATP13A2</i>			0.869	Yes	Yes			
<i>GBA</i>			0.676	Yes	Yes			
<i>PINK1</i>			0.536	Yes	Yes			
<i>SNCA</i>			0.68	Yes	Yes			

**Table 10 - part 2.** Gene co-expression analysis of candidate and established PD genes.

Gene symbol	Module	Module Size	Member-ship	SN UKBEC	SN GTEX	Preservation GTEX	Enrichment cell specific markers	Enrichment most relevant GO terms
ANKRD30A	blue	2564	0.627	Yes	Weak	1.4	Probably Neuron (Cahoy, 2008) ( $p=7.97e-100$ )	No significant enrichment
CALML4			0.618	Yes	Yes		CA1.Pyramidal ( $p=0.000316$ )	
CD36			0.659	Yes	Weak		Interneuron ( $p=1.829e-08$ )	
MNS1			0.471	Yes	Yes		Ependymal ( $p=0.001636$ )	
ARSB	cyan	1174	0.422	Yes	Yes	35.08	Interneuron ( $p=2.477e-09$ )	Mitochondrion organization ( $p=0.00135$ )
CAPS2			0.671	Yes	Yes			
FAM71A	turquoise	2969	0.602	Yes	Weak	0.69	No significant enrichment	Detection of stimulus involved in sensory perception ( $p=5.93e-09$ ), Detection of stimulus ( $p=7.3e-08$ ), Detection of chemical stimulus involved in sensory perception ( $p=1.41e-06$ ) Sensory perception of chemical stimulus ( $p=3.62e-06$ ) Detection of chemical stimulus ( $p=1.32e-05$ )
FAM83A			0.21	Yes	No			
KCNK16			0.787	Yes	Weak			
PTCHD3			0.66	Yes	Weak			
SSPO			0.673	Yes	Yes			
SVOPL			0.706	Yes	Weak			
TCHHL1			0.6	Yes	No			
KALRN	magenta	1121	0.378	Yes	Yes	14.73	Definite Neuron (Cahoy, 2008) ( $p=1.712e-06$ ) Probably Neuron (Cahoy, 2008) ( $p=2.094e-23$ )	Synaptic transmission ( $p=7.41e-11$ ) Cell-cell signaling ( $p=4.65e-09$ )

									CA1.Pyramidal ( $p=7.9e-05$ )	Cellular potassium ion transport ( $p=2.87e-06$ )
									S1.Pyramidal ( $p=9.005e-10$ )	Potassium ion transmembrane transport ( $p=2.87e-06$ )
<i>LRRK2</i>	pink	1294	0.253	Yes	Yes	22.59			Highly probable astrocytes (Cahoy, 2008) ( $p=1.056e-31$ )	Potassium ion transport ( $p=3.38e-06$ )
									Astrocytes in Cortex ( $p=7.603e-70$ )	Single-organism process ( $p=6.88e-13$ )
									ukbec-astrocytes ( $p=1.844e-12$ )	Single-organism cellular process ( $p=1.76e-11$ )
									Ependymal ( $p=3.117e-06$ )	Response to stimulus ( $p=4.15e-08$ )
									Astrocyte ( $p=4.442e-29$ )	Cellular response to stimulus ( $p=3.77e-07$ )
<i>PARK2</i>	salmon	1101	0.699	Yes	Yes	21.55			Neuron in Cortex ( $p=0.03054$ )	Biological_process ( $p=3.79e-07$ )
									Probably Neuron (Cahoy, 2008) ( $p=3.574e-14$ )	Intracellular transport ( $p=1.03e-08$ )
										Single-organism intracellular transport ( $p=2.59e-06$ )
										Cellular metabolic process ( $p=7.75e-06$ )
<i>PZP</i>	greenyel-low	1428	0.773	Yes	No	8.31				Cytoplasmic transport ( $p=4.95e-05$ )
<i>TMEM134</i>			0.62	Yes	Yes				Microglia in Human brain (Geschwind, 2010) ( $p=3.19e-25$ )	Intracellular protein transport ( $p=0.000106$ )
									Microglia (Type2) in Cortex ( $p=9.733e-05$ )	Anatomical structure formation involved in morphogenesis ( $p=1.96e-07$ )
									Endothelial ( $p=7.646e-15$ )	Circulatory system development ( $p=9.37e-05$ )
										Cardiovascular system development ( $p=9.37e-05$ )

**Table 10 - part 3.** Gene co-expression analysis of candidate and established PD genes.

Gene symbol	Module	Module Size	Member-ship	SN UKBEC	SN GTEX	Preservation GTEX	Enrichment cell specific markers	Enrichment most relevant GO terms
<i>VPS13C</i>	red	1442	0.283	Yes	Yes	18.75	Mural ( $p=0.02782$ ) Oligodendrocytes in Human brain (Geschwind,2010) ( $p=1.263e-20$ ). Oligodendrocytes in Cortex ( $p=4.048e-13$ ).	Blood vessel development ( $p=0.000177$ ) Vasculature development ( $p=0.000197$ ) Nucleic acid metabolic process ( $p=3.5e-09$ ) RNA metabolic process ( $p=5.38e-08$ )
<i>C11ORF21</i>	x	N/A	N/A	N/A	Weak	N/A	N/A	N/A
<i>COL6A5</i>	x	N/A	N/A	No	Weak	N/A	N/A	N/A
<i>DJS3</i>	x	N/A	N/A	N/A	Yes	N/A	N/A	N/A
<i>GH2</i>	x	N/A	N/A	No	No	N/A	N/A	N/A
<i>OR7G3</i>	x	N/A	N/A	N/A	No	N/A	N/A	N/A
<i>PCDHA9</i>	x	N/A	N/A	N/A	Weak	N/A	N/A	N/A
<i>ZNF543</i>	x	N/A	N/A	N/A	Weak	N/A	N/A	N/A

Established PD genes are in bold. Module size = number of genes in specified module; Memberschip = Pearson correlation of gene expression and the module eigengene which indicates the level of gene for the corresponding module (range: 0-1, 0=irrelevant, 1 = relevant). SN UKBEC = micro-array based gene expression in substantia nigra of UKBEC dataset; SN GTEX = gene expression in substantia nigra of GTEX dataset (yes=box plot mean >1, weak=box plot mean <1, no = box plot mean <0.5); Preservation GTEX = Module preservation score (>10=high evidence of replication, 2-10=moderate evidence of replication, <2=no evidence of replication). The references (cell specific markers column) indicate the source of the gene-set.



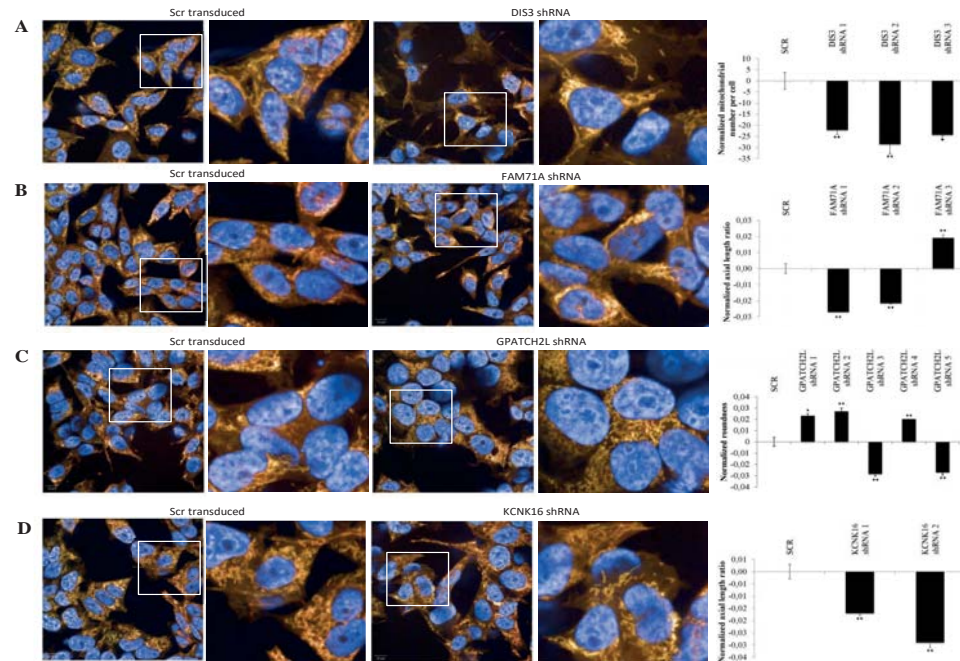
**Table 11.** Quality overview exome library preparation kits.

Capture type	Target Mb	n samples	# of reads	% aligned reads	% on target	% ≥10X depth	% ≥30X depth
Truseq	62	1163	95,852,037	97.7%	34.8%	96.2%	77.2%
Nimblegen	44.1	390	228,495,737	94.1%	31.4%	82.2%	71.5%
Agilent	37.6	6	91,853,688	98.0%	34.4%	70.7%	55.3%
mixed	-	88	230,110,307	95.1%	30.0%	97.2%	87.1%
NA	-	4	86,764,400	89.3%	37.8%	73.0%	60.0%

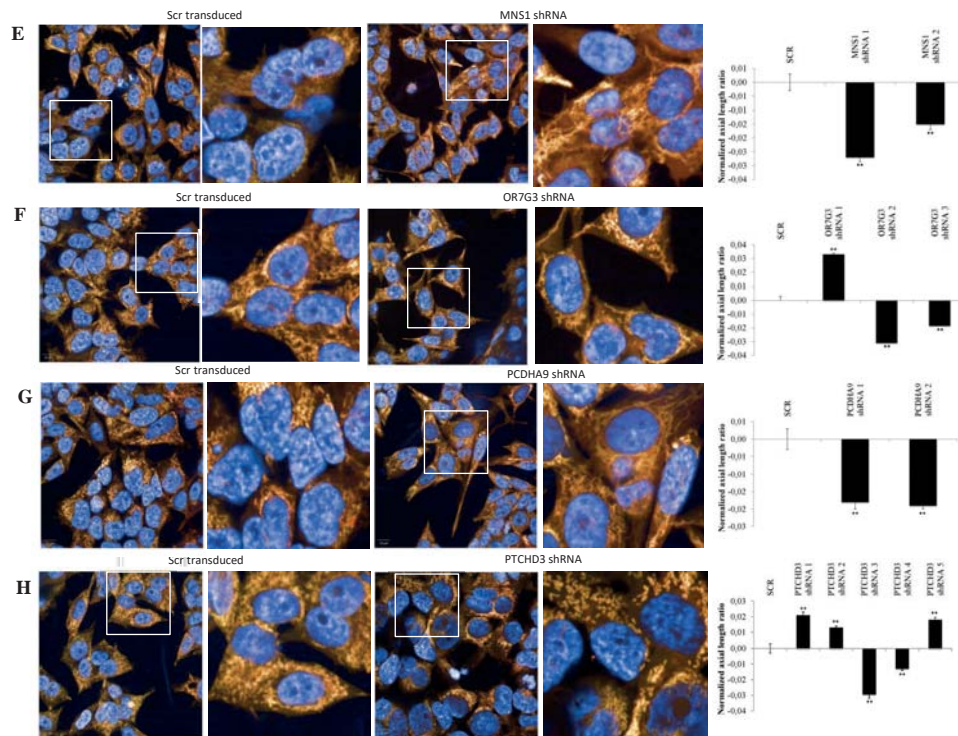
Target Mb = total size of regions that are targeted; mixed = samples of which bam files have been merged after being captured by different kits.

**Table 12.** Public availability of the IPDGC discovery dataset.

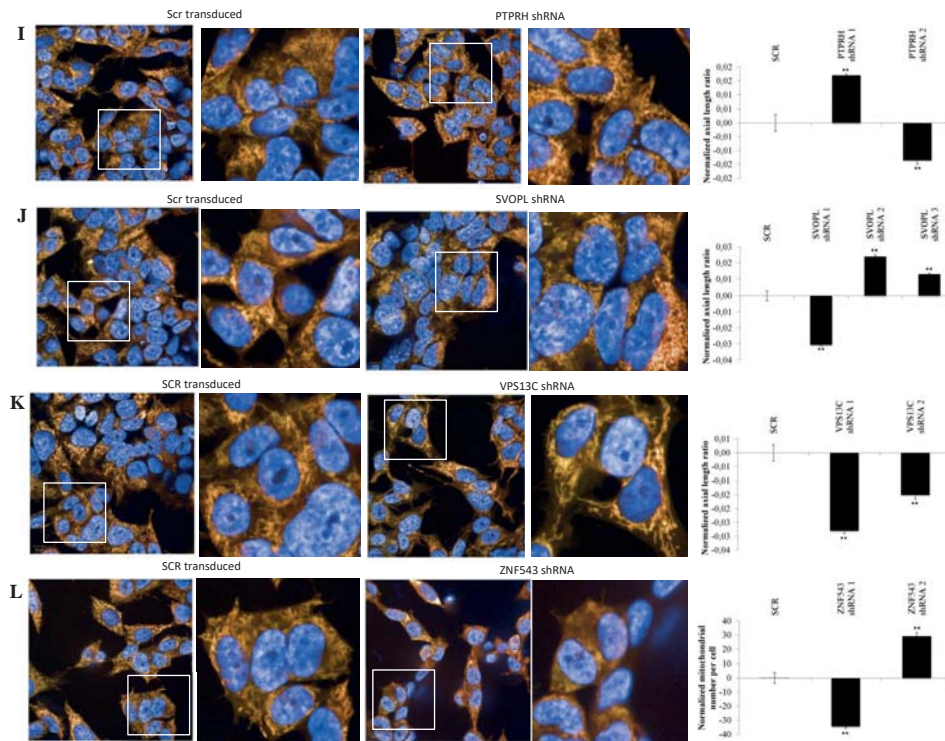
Cohort	n cases	n controls	Data accessible through:
NINDS US PD	534		dbGaP (phs001103.v1.p1)
US PD	6		contact Peter.Heutink@dzne.de
NABEC		324	dbGaP (phs000249) and EGA (EGAS00001002110)
UKBEC		71	EGA (EGAS00001002113)
UK PD	57		EGA
UK PD	126		contact Peter.Heutink@dzne.de
UK controls		108	contact Peter.Heutink@dzne.de
French PD	225		contact Peter.Heutink@dzne.de
Dutch PD	175		EGA (EGAS00001002103)
Dutch PD	25		contact Peter.Heutink@dzne.de



**Figure 1 - part 1.** High Content assay for Mitochondrial Morphology. Images show cells labeled with Hoechst (BLUE; Nuclei) and mitochondria (Mitotracker CMXRos and Mitotracker Deepred, YELLOW). Untreated cells are infected with shRNA with a scrambled sequence (left panel). Cells infected with shRNA targeting the specified gene show a decrease or increase in number/roundness/axial length ratio of mitochondria per cell (right panel). The graph displays normalised numbers of the specified parameters. Data are median values  $\pm$  median absolute deviation (MAD) of N=6 measurements. \* $P < 0.05$  and \*\* $P < 0.01$ , Mann-Whitney U test (see methods). All values were normalised to the negative control (infected with shRNA encoding a scrambled sequence). All shRNA clones that meet the cut-off criteria (see methods) are shown. A) *DIS3* shRNA decreases number of mitochondria per cell. B) *FAM71A* shRNA decreases mitochondrial axial length ratio and roundness. C) *GPATCH2L* shRNA increases mitochondrial roundness. D) *KCNK16* shRNA decreases mitochondrial axial length ratio and roundness. This figure is available at higher resolution through <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1147-9>.

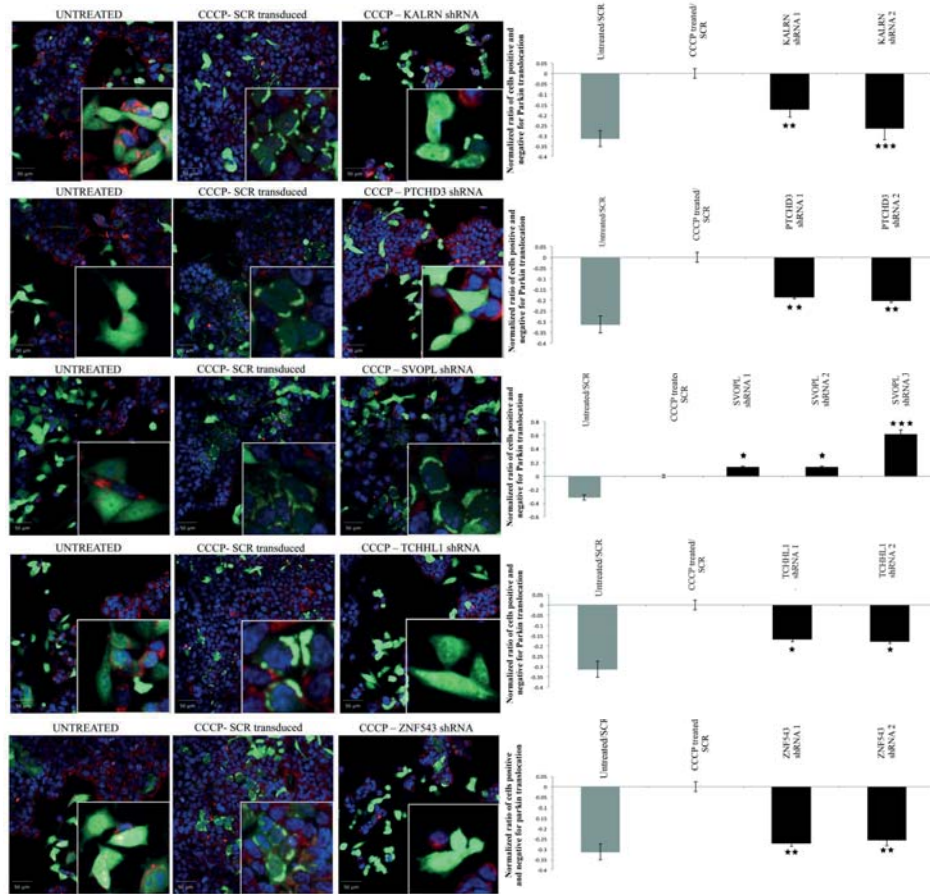


**Figure 1 - part 2.** High Content assay for Mitochondrial Morphology. Images show cells labeled with Hoechst (BLUE; Nuclei) and mitochondria (Mitotracker CMXRos and Mitotracker Deepred, YELLOW). Untreated cells are infected with shRNA with a scrambled sequence (left panel). Cells infected with shRNA targeting the specified gene show a decrease or increase in number/roundness/axial length ratio of mitochondria per cell (right panel). The graph displays normalised numbers of the specified parameters. Data are median values  $\pm$  median absolute deviation (MAD) of N=6 measurements. \* $P < 0.05$  and \*\* $P < 0.01$ , Mann-Whitney U test (see methods). All values were normalised to the negative control (infected with shRNA encoding a scrambled sequence). All shRNA clones that meet the cut-off criteria (see methods) are shown. E) *MNS1* shRNA decreases mitochondrial axial length ratio. F) *OR7G3* shRNA decreases mitochondrial axial length ratio and increases mitochondrial roundness and number of mitochondria per cell. G) *PCDHA9* shRNA decreases mitochondrial axial length ratio and roundness. H) *PTCHD3* shRNA increases mitochondrial axial length ratio and roundness. This figure is available at higher resolution through <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1147-9>.

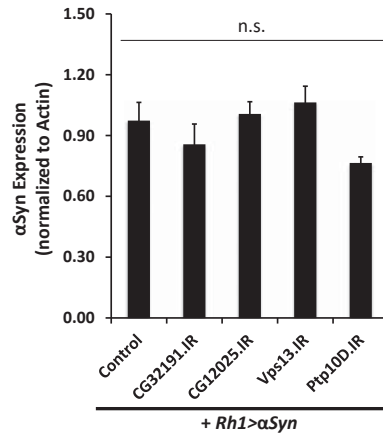


6

**Figure 1 - part 3.** High Content assay for Mitochondrial Morphology. Images show cells labeled with Hoechst (BLUE; Nuclei) and mitochondria (Mitotracker CMXRos and Mitotracker Deepred, YELLOW). Untreated cells are infected with shRNA with a scrambled sequence (left panel). Cells infected with shRNA targeting the specified gene show a decrease or increase in number/roundness/axial length ratio of mitochondria per cell (right panel). The graph displays normalised numbers of the specified parameters. Data are median values  $\pm$  median absolute deviation (MAD) of N=6 measurements. \* $P < 0.05$  and \*\* $P < 0.01$ , Mann-Whitney U test (see methods). All values were normalised to the negative control (infected with shRNA encoding a scrambled sequence). All shRNA clones that meet the cut-off criteria (see methods) are shown. I) *PTPRH* shRNA increases mitochondrial axial length ratio. J) *SVOPL* shRNA increases mitochondrial axial length ratio and decreases mitochondrial roundness. K) *VPS13C* shRNA decreases axial length ratio. L) *ZNF543* shRNA decreases number of mitochondria per cell. This figure is available at higher resolution through <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1147-9>.

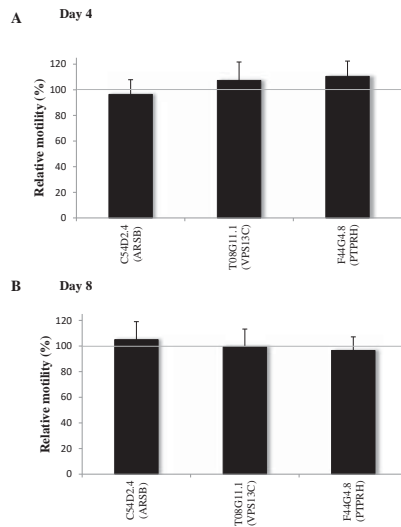


**Figure 2.** High content assay for Parkin Translocation. Effect of *KALRN* shRNA (A), *PTCHD3* shRNA (B), *SVOPL* shRNA (C), *TCHHL1* shRNA (D) and *ZNF543* shRNA (E). The images display cells labeled for nuclei (BLUE; Hoechst), Parkin-GFP (GREEN), mitochondria (RED, Mitotracker Deepred). Untreated cells infected with shRNA encoding a scrambled sequence show absence of puncta (left panel). Cells infected with a scrambled sequence but treated with CCCP show a significant increase in puncta formation (middle panel). Infection of cells with shRNA targeting *KALRN*, *PTCHD3*, *TCHHL1* and *ZNF543* prevents the accumulation of Parkin on mitochondrial (right panel), while targeting *SVOPL* promotes this accumulation. The graphs display the normalised ratio of cells positive for translocation and cells negative for Parkin translocation. All values were normalised to the negative control (CCCP treated infected with shRNA encoding a scrambled sequence). Data are median values  $\pm$  median absolute deviation (MAD) of N=6 measurements. \*  $P < 0.05$ , \*\*  $P < 0.01$  and \*\*\*  $P < 0.001$ , Mann-Whitney U test (see methods). All shRNA clones that meet the cut-off criteria (see methods) are shown.

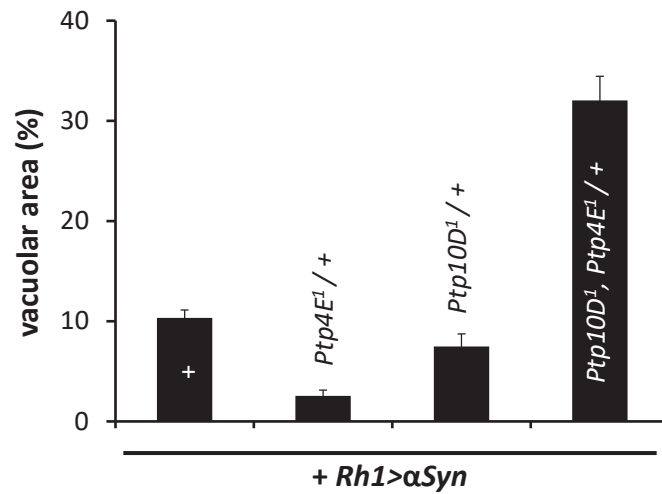
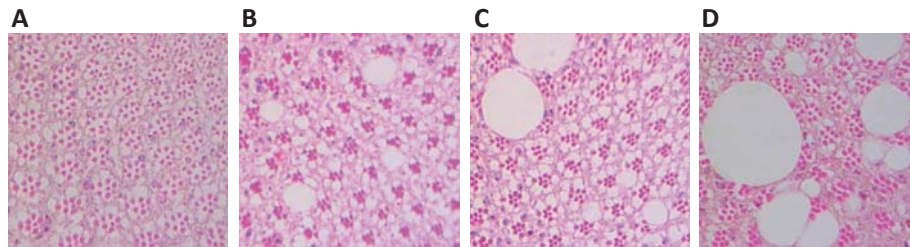


**Figure 3.** Enhancers of  $\alpha$ -synuclein toxicity do not significantly increase  $\alpha$ -synuclein protein levels. Western blots were performed on fly heads for *Rh1> $\alpha$ -synuclein* controls and following knockdown of each enhancer (same genotypes detailed in Figure 5 legend). Relative expression of  $\alpha$ -synuclein was determined based on normalization to Actin. n.s., non-significant.

6



**Figure 4.** Motility assay in *C. elegans* model overexpressing human  $\alpha$ -synuclein and YFP in the body wall muscle. Motility relative to the control is displayed for the time points day 4 and day 8. No significant decrease or increase of body movements observed after knockdown by shRNA of the corresponding orthologues.



6

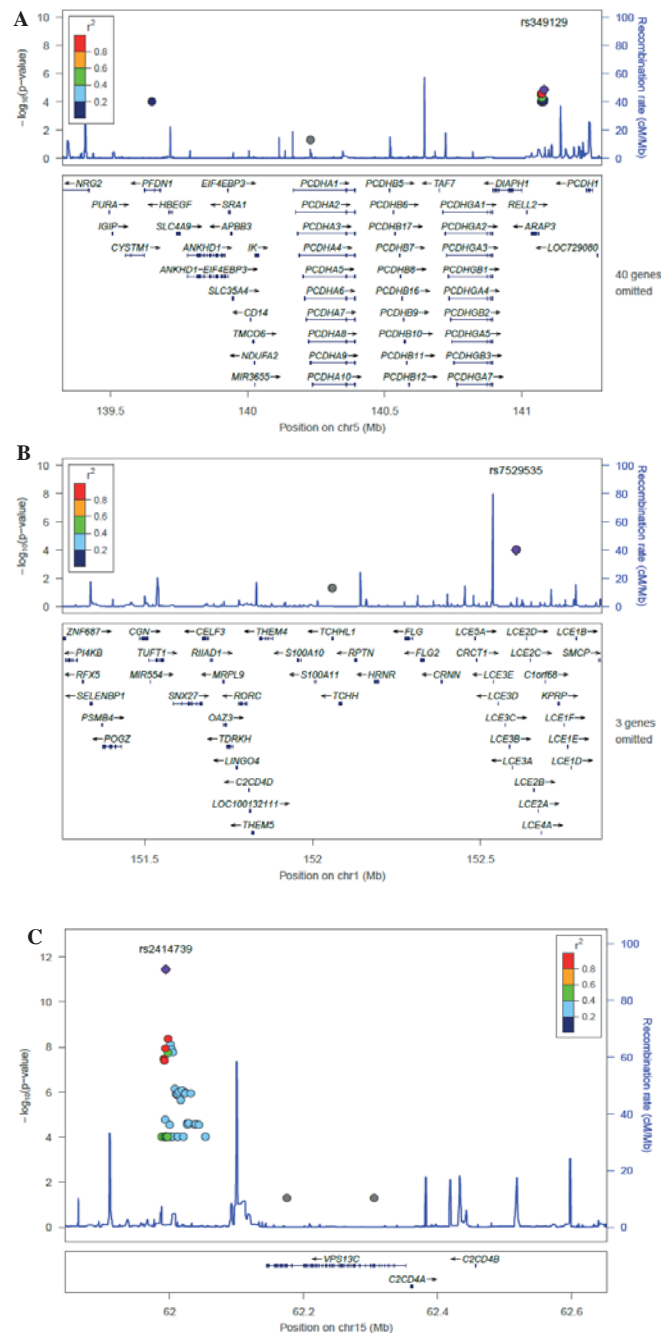
**Figure 5.** Enhancement of  $\alpha$ -synuclein toxicity by *Drosophila* homologs of *PTPRH*. Double-heterozygotes for strong mutant alleles in the paralogous genes, *Ptp10D* and *Ptp4E*, enhance  $\alpha$ -synuclein-induced retinal degeneration, whereas neither mutant allele enhances  $\alpha$ -synuclein toxicity on its own. .

(A) *Ptp10D<sup>1</sup>, Ptp4E<sup>1</sup>/+; Rh1-Gal4/+*

(B) *Ptp4E<sup>1</sup>/+; Rh1-Gal4/+; UAS-Syn/+*

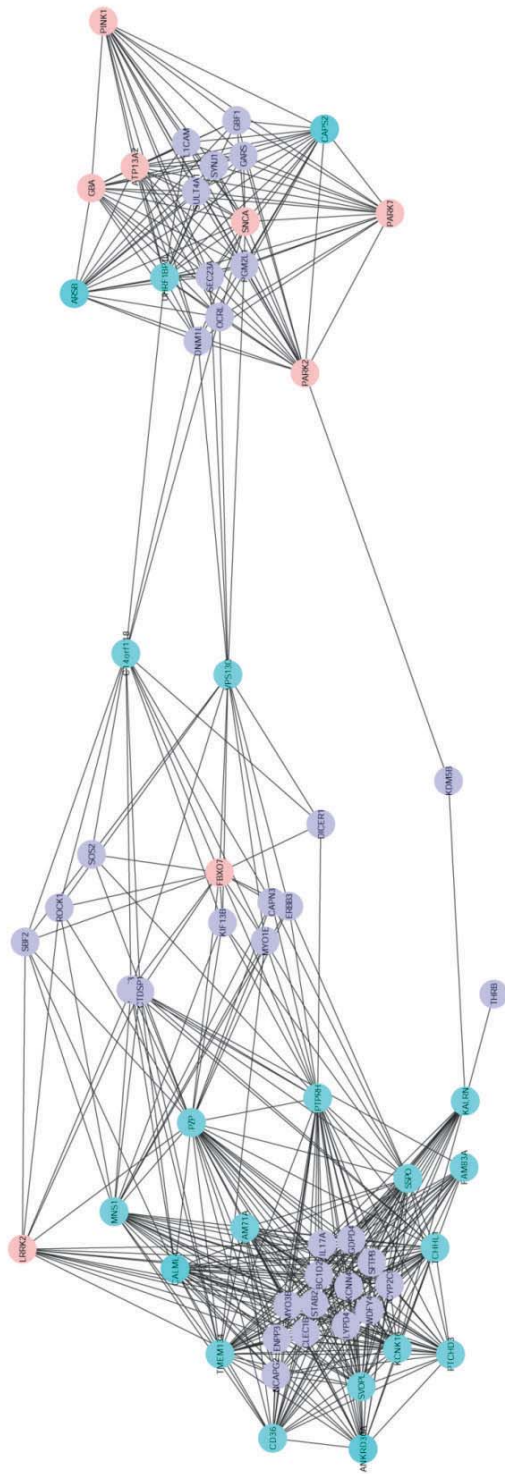
(C) *Ptp10D<sup>1</sup>/+; Rh1-Gal4/+; UAS-Syn/+*

(D) *Ptp10D<sup>1</sup>, Ptp4E<sup>1</sup>/+; Rh1-Gal4/+; UAS-Syn/+*

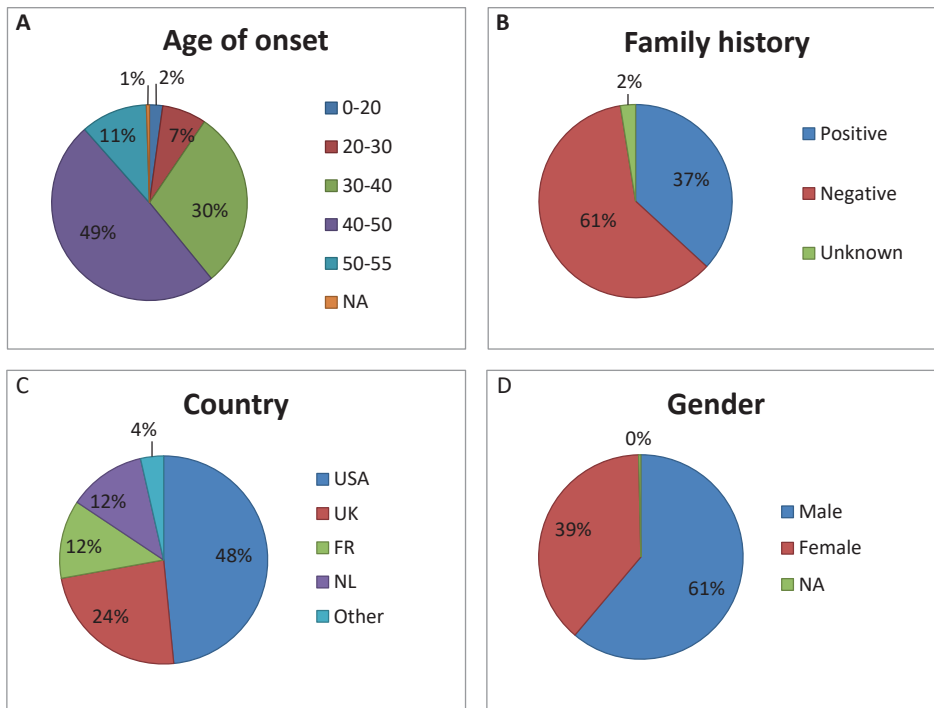


**Figure 6.** GWAS peak regions around *PCDHA9* (A), *TCHHL1* (B) and *VPS13C* (C) with underlying European LD-structure. The grey dots indicate the location of the LoF variant of the exome data. The colored dots indicate the positions and  $p$ -values of the GWAS hits.





**Figure 7.** Plot to demonstrate the co-expression of candidate genes with well-established PD genes. We generated a “bottom-up” plot using the 20 candidate genes expressed in *substantia nigra* and known PD genes as seeds and based on TOM values between genes (edges were filtered for an adjacency cut-off of TOM values < 0.01 to show the most relevant interactions) for the UKBEC GCN. Candidate genes are highlighted in light blue, PD genes in pink and context genes in purple. This plot demonstrate two main subnetworks, the one on the left centered around *FBXO7* and *LRRK2*, the one on the right centered around *ATP13A2*, *GBA*, *PARK2*, *PARK7*, *PINK1* and *SNCA*. The latter shows a tight interaction of *UHRF1BP1L* with the PD genes it clusters with within the yellow module, but also with *PARK2* and *PARK7*.



**Figure 8.** Cohort demographics. A: Age of onset distribution of cases expressed in years (NA identifies PD cases that were labeled as “YOPD” but lacked an age of onset); B: Family history distribution of cases; C: Country of origin distribution of all samples (USA=United States of America; UK=United Kingdom; NL=The Netherlands; FR= France); D: Gender distribution of all samples.

6

