

RESEARCH

Open Access



CrossMark

# Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing

Iris E. Jansen<sup>1,2†</sup>, Hui Ye<sup>3†</sup>, Sasja Heetveld<sup>1†</sup>, Marie C. Lechler<sup>1,4†</sup>, Helen Michels<sup>5</sup>, Renée I. Seinstra<sup>5</sup>, Steven J. Lubbe<sup>6,7</sup>, Valérie Drouet<sup>8</sup>, Suzanne Lesage<sup>8</sup>, Elisa Majounie<sup>9</sup>, J. Raphael Gibbs<sup>10</sup>, Mike A. Nalls<sup>10</sup>, Mina Ryten<sup>11,12</sup>, Juan A. Botia<sup>11</sup>, Jana Vandrovčová<sup>11</sup>, Javier Simon-Sánchez<sup>1,13</sup>, Melissa Castillo-Lizaro<sup>1</sup>, Patrizia Rizzu<sup>1</sup>, Cornelis Blauwendraat<sup>1</sup>, Amit K. Chouhan<sup>14</sup>, Yarong Li<sup>14</sup>, Puja Yogi<sup>14</sup>, Najaf Amin<sup>15</sup>, Cornelia M. van Duijn<sup>15</sup>, International Parkinson's Disease Genetics Consortium (IPGDC), Huw R. Morris<sup>6</sup>, Alexis Brice<sup>8,16</sup>, Andrew B. Singleton<sup>10</sup>, Della C. David<sup>1</sup>, Ellen A. Nollen<sup>5</sup>, Shushant Jain<sup>1</sup>, Joshua M. Shulman<sup>3,14,17,18\*</sup> and Peter Heutink<sup>1,2,13\*†</sup>

## 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 participants. To discover rare PD susceptibility variants, we performed WES in 1148 unrelated cases and 503 control participants. Candidate genes were subsequently validated for functions relevant to PD 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 four candidates enhanced α-synuclein-induced neurodegeneration in *Drosophila*. Based on complementary analyses in independent human datasets, five 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.

**Keywords:** Parkinson's disease, Genomics, Whole-exome sequencing, Loss-of-function, Rare variants, Functional screening, Mitochondria, Parkin, α-synuclein, Animal model

\* Correspondence: [Joshua.Shulman@bcm.edu](mailto:Joshua.Shulman@bcm.edu); [Peter.Heutink@dzne.de](mailto:Peter.Heutink@dzne.de)  
Joshua M. Shulman and Peter Heutink are co-senior authors.

Iris E. Jansen, Hui Ye, Sasja Heetveld and Marie C. Lechler are co-first authors.

†Equal contributors

<sup>3</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA

<sup>1</sup>German Center for Neurodegenerative Diseases (DZNE), Otfried-Müller-Str.  
23, Tübingen 72076, Germany

Full list of author information is available at the end of the article

## Background

Next-generation sequencing (NGS) approaches have recently accelerated the identification of variants responsible for familial Parkinson's disease (PD) [1–4]. While a positive family history is common in PD, large, multi-generational 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 [5, 6]. 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 (GWAS) of case/control cohorts [7]. The variants identified by GWAS have modest effect sizes and collectively fail to account for current estimates of PD heritability [8, 9]. Considering the above, it seems likely that additional less common alleles, with larger effect sizes, contribute to PD risk in the population and NGS is one promising approach to identify such alleles. Despite recent successes in other neurodegenerative disease with complex genetic etiologies, including Alzheimer's disease [10–12] and amyotrophic lateral sclerosis [13, 14], 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 [15]. Lastly, as PD is: (1) common (~1–3% prevalence); (2) strongly age-dependent; and (3) often preceded by a prolonged pre-symptomatic 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 [16], 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 1148 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 [17–19], 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

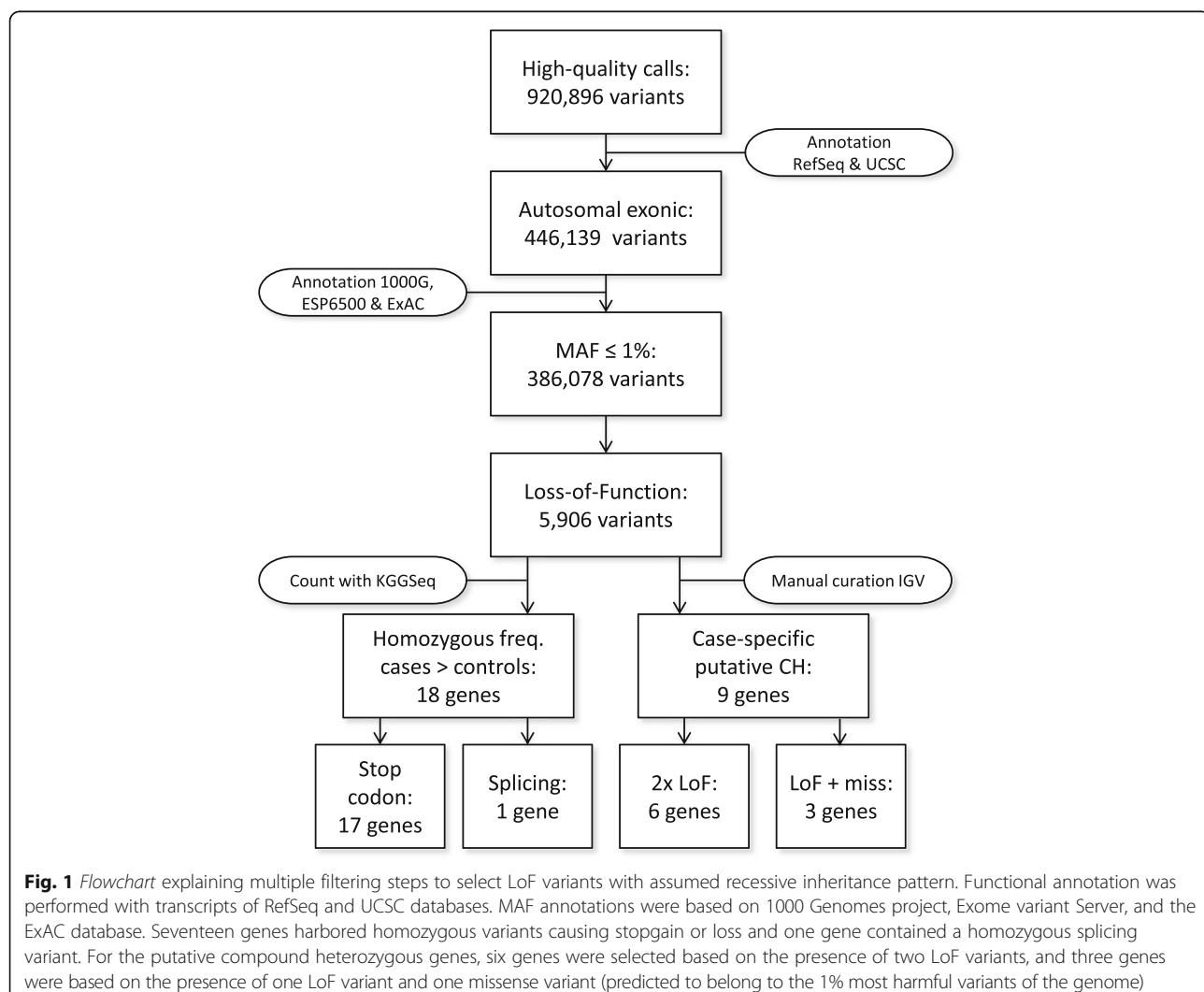
### Discovery of recessive LoF variants from PD exomes

A total of 920,896 variants (93.2% single nucleotide variants and 6.8% insertions and deletions) were called in a WES dataset of 1651 participants, including 1148 young-onset PD cases (average age of onset, 40.6 years; range, 5–56 years) and 503 control participants with European ancestry. As our cohort has an average age at onset of less than 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 two participants with homozygous exonic variants in *parkin* and *PINK1* (Additional file 1: Table S1). 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) [20], 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 changing missense alleles that were predicted to be

deleterious ( $\text{CADD} > 20$ ), consistent with a 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 nine genes harboring putative compound heterozygous variants—all predicted to cause a loss of gene function (Table 1). Approximately 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 other 26 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 four individual cases, along with one control participant, were homozygous for a variant (rs138279020) predicted to disrupt splicing. Although this variant is neither reported in ExAC nor was frequency information available from dbSNP, the MAF was 0.065 in our cohort ( $\text{MAF}_{\text{CASES}} = 0.073$ ,  $\text{MAF}_{\text{CONTROLS}} = 0.052$ ,  $p = 0.054$ ). Although relatively frequent in our control dataset ( $\text{MAF} > 1\%$ ), we have retained it among our results, based on three 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



**Fig. 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. MAF annotations were based on 1000 Genomes project, Exome variant Server, and the ExAC database. Seventeen genes harbored homozygous variants causing stopgain or loss and one gene contained a homozygous splicing variant. For the putative compound heterozygous genes, six genes were selected based on the presence of two LoF variants, and three 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				ExAC (EU)				Functional information				Validation	
Gene	Type	Chr:bp (hg19)	Ref/Alt all	dbsNP137	MAF	n Hz (freq)	Variant type	AA change	AA length	CADD	Functional	Genetic			
<i>ANKRD30A</i>	CH	10:37438759	G/T	x	x	x	stopgain	NM_052997:p.E487X	1341	35.0					
		10:37508648	A/AT	x	0.00031%	0	fs insertion	NM_052997:p.Q1280S	1341	24.2					
<i>ARSB</i>	HZ	5:78281383	T/TA	rs138279020	x	x	splicing	NM_198709:c.190-2insT	533	2.2	S			neuroX	
<i>c11orf21</i>	HZ	11:2321829	C/T	rs74048215	0.037%	0	stopgain	NM_001142946:p.W69X	178	35.0					
<i>CALML4</i>	HZ	15:68497597	G/A	rs11071990	0.75%	1 (0.00003%)	stopgain	NM_001031733:p.R40X	120	25.9					
<i>CAPS2</i>	HZ	12:75687045	C/CTATTGAT	x	0.68%	4 (0.00012%)	stopgain	NM_001286547:p.L321X	382	33.0					
<i>CD36</i>	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.R412S	472	25.9					
<i>COL6A5</i>	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.Y756M	2611	23.0					
<i>DJ53</i>	CH	13:73355096	G/A	rs201493043	0.0045%	0	stopgain	NM_0014953:p.Q92X	958	37.0	Mm				
		13:73333935	A/G	rs141067458	0.22%	0	stoploss	NM_0014953:p.X959Q	958	14.1					
<i>FAM71A</i>	HZ	1:212799882	A/T	rs143861655	0.80%	2 (0.00006%)	stopgain	NM_153606:p.K55X	594	36.0	Mm				
<i>FAM83A</i>	CH	8:124195352	G/T	rs148011353	0.23%	2 (0.00006%)	stopgain	NM_032899:p.G86X	434	26.2					
		8:124195506	T/G	x	x	x	missense	NM_032899:p.V137G	434	25.6					
<i>GH2</i>	HZ	17:61957946	C/T	rs1506668018	0.33%	1 (0.00003%)	stopgain	NM_022557:p.W214X	256	26.8					
		14:76644266	C/T	rs117516637	0.53%	1 (0.00003%)	stopgain	NM_017927:p.R362X	482	18.8	Mp,Mm				
<i>GPA1CH2L</i>	HZ	3:124303696	C/T	rs56407180	0.39%	0	stopgain	NM_007064:p.R10X	1289	37.0	Mp				
<i>KALRN</i>	HZ	6:39282816	G/A	rs138573996	0.44%	2 (0.00006%)	stopgain	NM_001135107:p.Q251X	322	9.5	Mm				
<i>KCNK16</i>	HZ	15:56736045	G/A	x	x	x	stopgain	NM_018365:p.Q232X	495	38.0	Mm				
<i>MNS1</i>	CH	15:567486830	G/A	x	0.003%	0	stopgain	NM_018365:c.C265Tp.Q89X	495	37.0	Mp				
		19:9236969	G/A	rs61751875	0.15%	1 (0.00003%)	stopgain	NM_001001958:p.R220X	312	35.0	Mm				
<i>OR7G3</i>	HZ	5:140229907	C/G	x	0.0046%	0	stopgain	NM_014005:p.Y609X	950	34.0	Mm				
		10:27688101	G/A	rs142646098	0.97%	6 (0.00018%)	stopgain	NM_001034842:p.R476X	767	34.0	Mp,Mm				
<i>PTPRH</i>	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					
<i>PZP</i>	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					
<i>SSPO</i>	HZ	7:149493519	C/T	rs57595625	0.011%	0	stopgain	NM_198455:p.Q2199X	5150	10.4					
		7:138341219	G/A	rs117871806	0.35%	0	stopgain	NM_174959:p.R18X	492	38.0	Mp,Mm				

**Table 1** Recessive LoF variants (*Continued*)

TCHH1	HZ	1:152058192	G/A	rs150014958	0.037%	0	stopgain	NM_001008536;p.Q656X	904	35.0	Mp
TMEM134	HZ	11:67235051	G/A	rs143199541	0.70%	3 (0.00009%)	stopgain	NM_001078650;p.R84X	195	34.0	S
UHRF1BP1L	HZ	12:100433523	T/A	x	x	x	stopgain	NM_015054;p.K1376X	1464	35.0	Mm
VPS13C	CH	15:62174851	C/A	x	x	x	stopgain	NM_017684;p.E3147X	3628	55.0	S,Mm
ZNF543	HZ	15:62305257	C/CTCTG	x	x	x	fs insertion	NM_017684;p.R226fs	3628	27.4	
				rs150392165	0.60%	1 (0.00003%)	stopgain	NM_213598;p.W68X	600	25.0	Mp,Mm

The current designation of rs138279020 is rs11424557

Type of affecting 2 alleles, CH putative compound heterozygote, HZ homozygote, Chrbp chromosome and base-pair position, hg19 reference build, Ref reference, Alt alternative, ExAC Exome Aggregation consortium, EU European, fs frameshift, AA change amino acid change for specified RefSeq transcript, MAF minor allele frequency, n HZ number of individuals with homozygous variant, AA length length of specified transcript in amino acids, PPMI Parkinson Progression Markers Initiative, CADD functional algorithm prediction (>20 = belongs to 1% most damaging variants of total genome), Mp mitochondrial Parkin translocation assay, Mm mitochondrial morphology assay, S α-synuclein assay

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 [21]. 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 LoF 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 [16] of predicted complete gene knock-outs in the Icelandic population, including 104,220 participants with imputed genotypes, based on whole genome sequencing from a subset of 2363 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 (Additional file 1: Table S2); however, the observed homozygote frequencies are not sufficiently high to confidently exclude them as possible PD genes and importantly, detailed phenotypic data are not publicly available for these participants. For example, 29 Icelandic participants 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 (Additional file 1: Table S2). 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 [22]. 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, ten were well conserved in the *C. elegans* genome and nine 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 Additional file 1: Table S3. 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* [23, 24] and mice [25, 26]. 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 [16].

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 LoF. 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: (1) a subset predicted to be deleterious (CADD > 20); and (2) 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 Additional file 1: Table S4, the rare variant aggregation association analyses provided further evidence in support of four 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 versus 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 [27–29]. Notably, *parkin* (*PARK2*), *DJ-1*, and *PINK1*, associated with autosomal recessive, juvenile-onset Parkinsonism, have roles in mitochondrial dynamics and quality control [30]. Specifically, Parkin is an E3 ubiquitin ligase and recruited selectively to dysfunctional mitochondria with a low membrane potential [31]. Further, the neurotoxicity of  $\alpha$ -synuclein, the primary constituent of Lewy body inclusions in PD, has also been linked to mitochondrial injury [32]. We therefore hypothesized that LoF in candidate genes identified from our analyses of WES, might similarly impact mitochondria, consistent with roles in PD susceptibility.

Therefore we quantified mitochondrial morphology after gene knockdown in BE(2)-M17 neuroblastoma cells by examining three parameters commonly used for quantification of mitochondrial morphology: mitochondrial number, axial length ratio, and roundness [33]. Cells transduced with the short hairpin RNA (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 (Fig. 2a, b) [34]. Knockdown of 13 genes show a significant effect on at least one of the three parameters (Additional file 1: Table S5 and Table S6 and Additional file 2: Figure S1). *GPATCH2L* shows the largest increase in mitochondrial

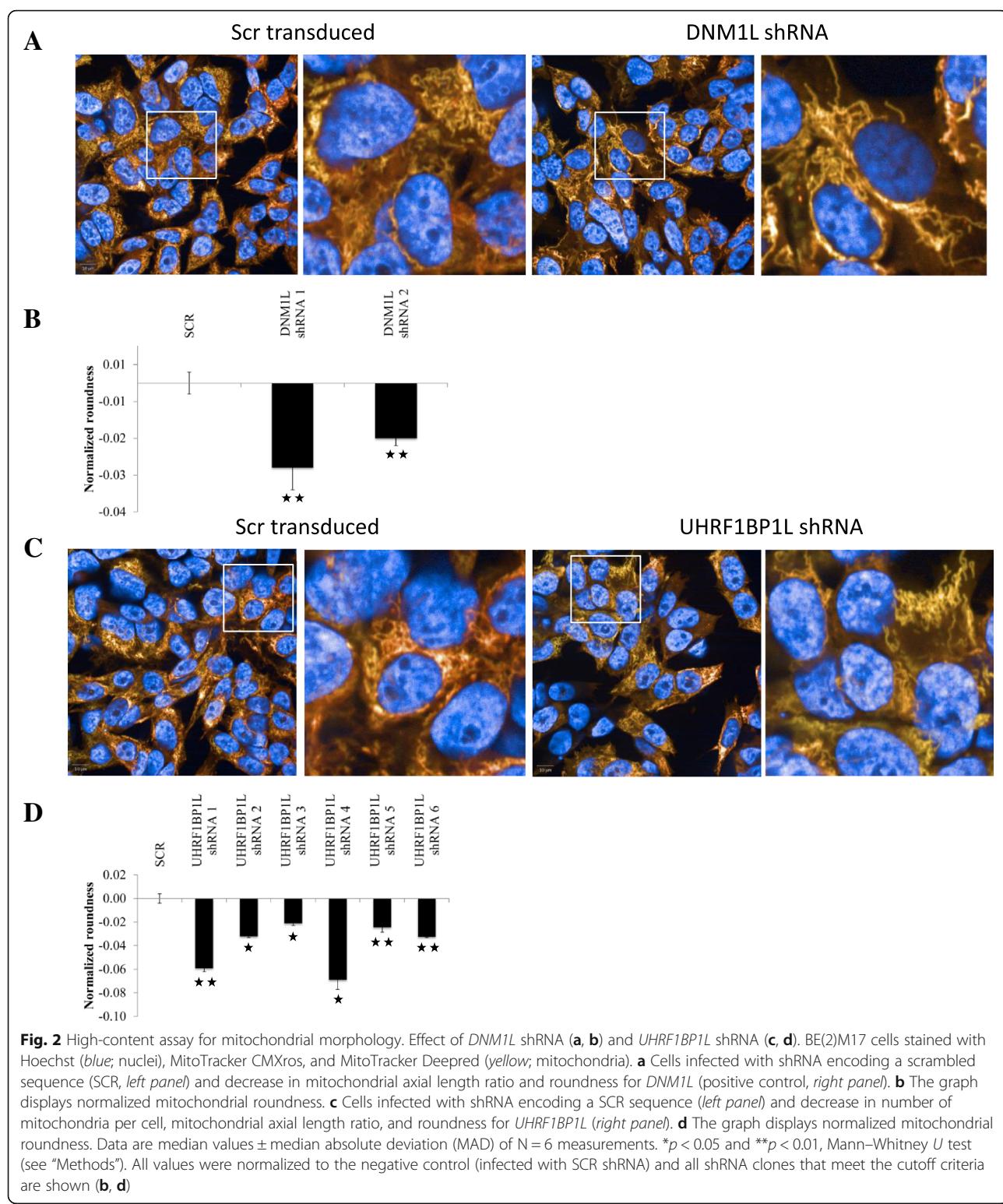
roundness, while *UHFR1BP1L* displays the largest decrease (Fig. 2c, d).

We also took advantage of a well-established Parkin translocation assay [31, 35–38] 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 (Fig. 3a, untreated) to the mitochondria (Fig. 3a, CCCP-SCR transduced) and this was PINK1-dependent (Fig. 3a, CCCP-PINK1 shRNA), which provides an internal, positive control in 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 (Fig. 3b). Based on stringent criteria (see “Methods”), six genes significantly modified Parkin translocation (Fig. 3c and d; Additional file 2: Figure S2; Additional file 1: Table S5 and Table S6), including four genes (*GPATCH2L*, *PTCHD3*, *SVOPL*, and *ZNF543*) with consistent activities in both the mitochondrial morphology and Parkin translocation assays.

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

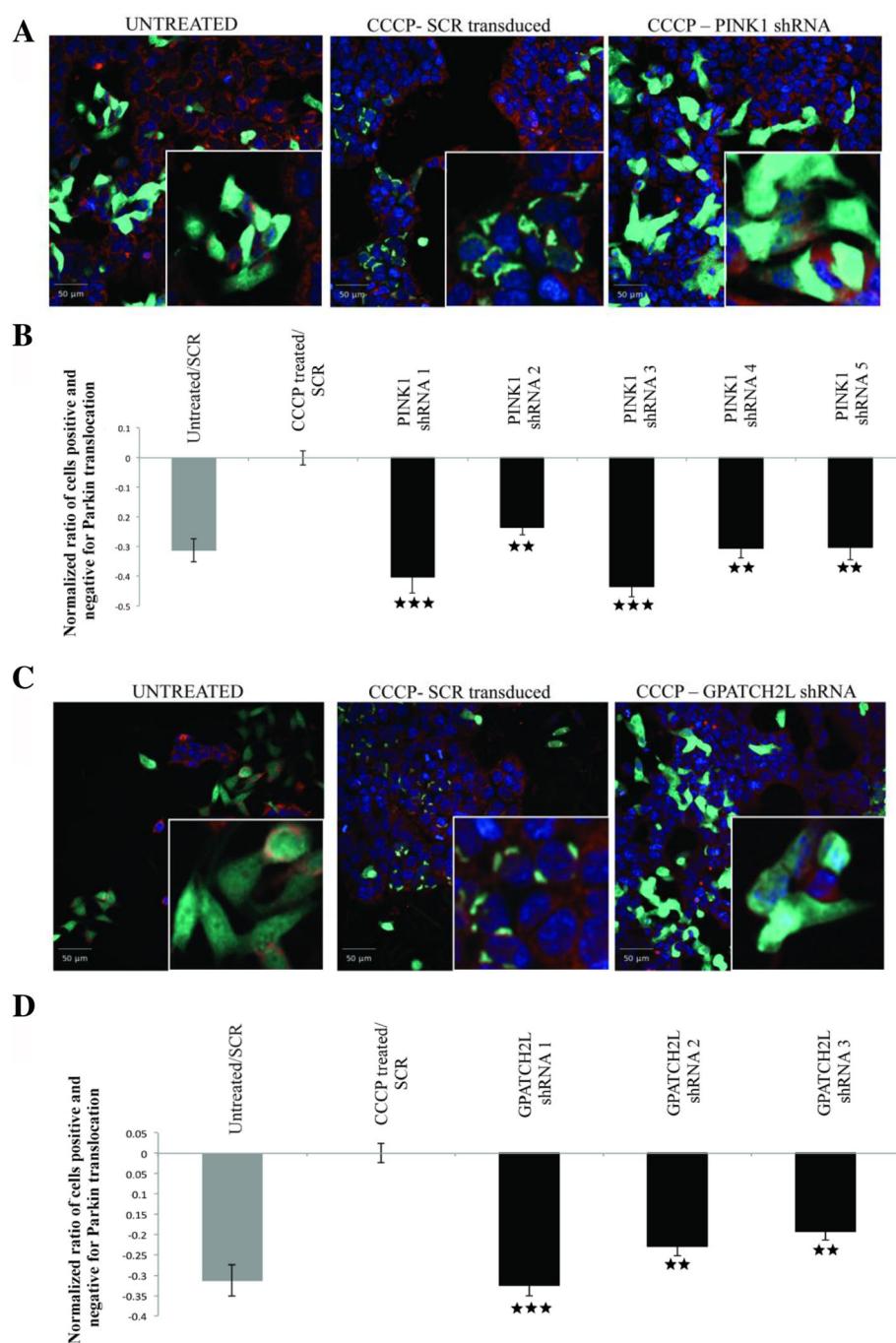
A wealth of evidence also supports a central role for  $\alpha$ -synuclein-mediated toxicity in PD pathogenesis.  $\alpha$ -synuclein aggregates, termed Lewy bodies, are the defining disease pathology and  $\alpha$ -synuclein gene (*SNCA*) mutations, locus multiplication, and promoter polymorphisms are associated with PD susceptibility [5]. Further, expression of  $\alpha$ -synuclein in numerous animal models including in the fruit fly [39–41], *Drosophila melanogaster*, recapitulates features of PD-related neurodegenerative pathology. Transgenic expression of  $\alpha$ -synuclein in the fly retina leads to neurotoxic changes [39] and is amenable for detection of genetic modifiers [42, 43]. Genetic manipulation of established PD susceptibility genes, including *PARK2* [44, 45] and *VPS35* [46], modulate  $\alpha$ -synuclein toxicity in transgenic flies, similar to findings in mammalian models [44, 47]. 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* (Additional file 1: Table S7). 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*) [48]. 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

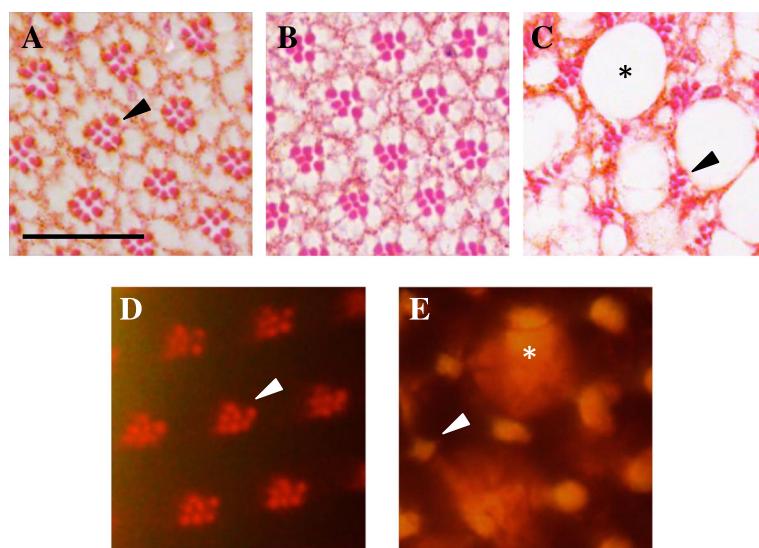


morphologically normal at 1 day (Fig. 4), but demonstrates age-dependent degeneration leading to progressive vacuolar changes, rhabdomere loss, and culminating with extensive tissue destruction by 30 days. At the 15-day time point selected for screening, only mild, if any, retinal

pathology is detectable on most histologic sections, consistent with a weakly penetrant degenerative phenotype following optical neutralization (mean penetrance  $\sim 25\%$ ) (Fig. 4). However, co-expression of RNAi targeting fly homologs of four candidate genes (*ARSB*, *TMEM134*,



**Fig. 3** High content assay for Parkin translocation. Effect of *PINK1* shRNA (**a, b**) and *GPATCH2L* shRNA (**c, d**). **a, 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 mitochondrial (right panel). **b, d** The graph displays the normalized ratio of cells positive for translocation and cells negative for parkin translocation. All values were normalized 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 cutoff criteria (see “Methods”) are shown



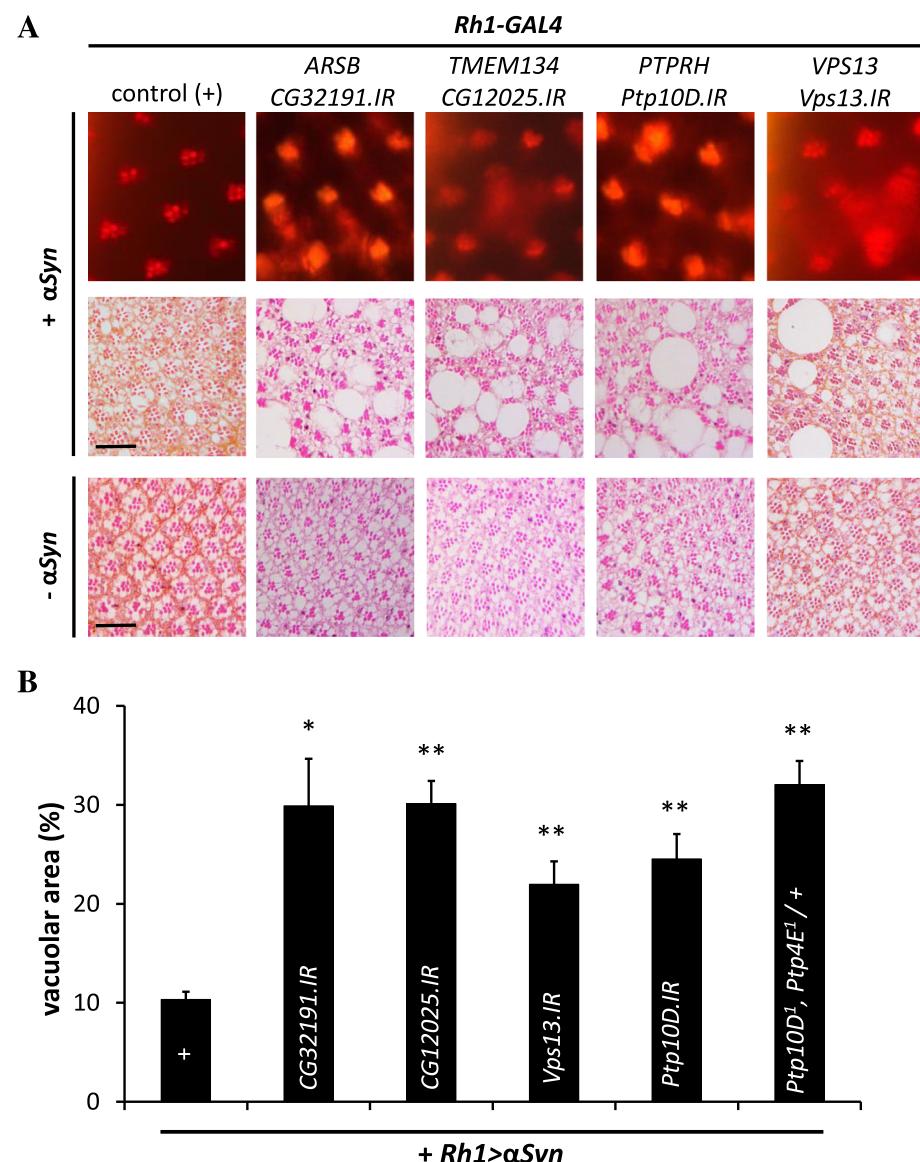
**Fig. 4**  $\alpha$ -synuclein-induced retinal degeneration and screening assays in *Drosophila* transgenic animals. Tangential sections through the fly retina stained with hematoxylin and eosin reveal the ordered ommatidial array in control animals (**a** *Rh1-GAL4* / +). Each ommatidium consists of a cluster of eight photoreceptive neurons (seven 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 one-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 Fig. 5. Scale bar: 20  $\mu$ m

*PTPRH*, and *VPS13C*) was observed to robustly enhance  $\alpha$ -synuclein-mediated neurodegeneration in the retina (mean penetrance ~ 75%; Additional file 1: Table S8).

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 (Fig. 5). Importantly, when 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 (Fig. 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 [5], western blot analyses 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 (Additional file 2: Figure S3). Thus, we hypothesize potential interactions with more downstream mechanisms of  $\alpha$ -synuclein neurotoxicity.

For 3 out of 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 [49]. However, no significant differences were detected in the  $\alpha$ -synuclein-induced locomotor phenotype observed in one-week-old worms following knockdown of these genes (Additional file 2: Figure S4). 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 four genes discovered to interact with  $\alpha$ -synuclein toxicity in *Drosophila*, we were able to obtain additional genetic reagents, including classical LoF alleles, for the two homologs of *PTPRH*: *Ptp10D* and *Ptp4E*. In our screen, two independent RNAi lines targeting *Ptp10D* robustly enhanced  $\alpha$ -synuclein toxicity, but only one of the two available lines for *Ptp4E* met our threshold criteria (Additional file 1: Table S8). 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 [50]. Consistent with this, we found that transheterozygosity for



**Fig. 5** PD gene candidates harboring LoF variants enhance α-synuclein toxicity in *Drosophila*. Conserved fly orthologs of human genes discovered from WES analysis were targeted with RNAi (IR) and screened for enhancement of α-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 α-synuclein pseudopupil phenotype and enhancers required consistent results for at least two independent RNAi lines (see Additional file 1: Table S8). Representative results from the primary screen are shown for controls (*Rh1-GAL4* / +; *UAS-α-synuclein* / +) and one IR line each for the implicated enhancers [Human Gene-Fly Ortholog (experimental genotype shown): *ARSB-CG32191* (*Rh1-GAL4* / +; *UAS-α-synuclein* / *UAS-CG32191.IR.v14294*), *TMEM134-CG12025* (*Rh1-GAL4* / *UAS-CG12025.IR.v104336*; *UAS-α-synuclein* / +), *PTPRH-Ptp10D* (*Rh1-GAL4* / *UAS-Ptp10D.IR.v1102*; *UAS-α-synuclein* / +), and *VPS13-Vps13* (*Rh1-GAL4* / *UAS-Vps13.IR.HMS02460*; *UAS-α-synuclein* / +)]. At the 15-day-old time point, *Rh1 > α-synuclein* causes a weakly-penetrant pseudopupil phenotype and mild histopathologic changes which are amenable to modifier screening (compare with Fig. 4, panels c and 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 α-synuclein co-expression (**a** bottom row, *Rh1-GAL4* / IR transgene). Scale bars: 20 μm. **b** Enhancement of α-synuclein-induced retinal degeneration was quantified based on the extent of vacuolar changes (area occupied by vacuoles / total retinal area). For quantification, three 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 Additional file 2: Figure S5). Statistical comparisons were made using unpaired t-tests. Error bars are based on Standard Error of the Mean. \* $p < 0.05$ ; \*\* $p < 0.01$ .

strong (null) alleles of both genes enhanced  $\alpha$ -synuclein-induced retinal degeneration (*Ptp4E<sup>I</sup>*, *Ptp10D<sup>I</sup>* / +; *Rh1-Gal4* / +; *UAS- $\alpha$ -synuclein* / +); whereas heterozygosity for either allele in isolation showed no significant enhancement (Fig. 5b and Additional file 2: Figure S5).

#### 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 [51]; (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 [52]; (3) an independent NeuroX exome array dataset [7, 53]; and (4) a large PD GWAS dataset [53]. Within the PPMI exome dataset, including 462 PD cases and 183 controls, evidence supporting replication was discovered for two genes, in which we identified the identical variants from the IPDGC discovery exome dataset (Additional file 1: Table S9). A PD case from PPMI carries the same homozygous stopgain variant (p.R362X) in *GPATCH2L* 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 patients, 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 [54] 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 1610 studied PD patients (1148 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 participant with sporadic PD diagnosed at the age of 28 years, was also observed in a single sporadic PD case from PPMI with an age of onset of 62 years. These *FAM83A* carriers presented with similar symptoms, including bradykinesia, rigidity, and resting tremor. In both datasets, the p.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 two PD

index cases and absent in controls. We identified a heterozygous missense variant (NM\_001127444:c.1176G > T:p.L392F) in *CD36* for three 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, 61–79 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 two 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 6801 individuals with PD and 5970 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 (Additional file 1: Table S10). 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 participants 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 (Additional file 1: Table S10).

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 1-Mb genomic window centered on each of the 27 candidate genes. Three loci (*VPS13C*, *PCDHA9*, and *TCHHL1*) showed evidence consistent with an association peak (Additional file 2: Figure S6). A genome-wide significant association at the *VPS13C* locus, was in fact recently reported [7]; 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 (Additional file 2: Figure S6), 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 [55] in which the SNP is located. The other two 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 seven genes (*GPATCH2L*, *FAM83A*, *CD36*, *UHRF1BP1L*, *PTPRH*, *ARSB*, and *VPS13C*) that were implicated by our WES analysis, of which five (*GPATCH2L*, *UHRF1BP1L*, *PTPRH*, *ARSB*, and *VPS13C*) are further validated based on functional evidence from PD-relevant experimental models.

#### 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 [56]. Of the 27 candidate genes, seven were not sufficiently expressed in substantia nigra on the basis of UKBEC. Except for *DIS3*, these genes were also expressed poorly in publicly available data of the Genotype-Tissue Expression (GTEx) project [56]. 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 (Additional file 1: Table S11 and Additional file 2: Figure S7). This approach highlighted three 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 *FBXO7* and enriched for oligodendrocyte markers ( $p = 8.69 \times 10^{-22}$ ). Importantly, the implicated modules were preserved ( $Z_{\text{summary}} > 10$ ) 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 five 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 cannot 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 27 genes. Evidence from human genetics and functional studies converge to most strongly implicate five 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 [56], 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 [28, 30]. 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 three 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 [20]. 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 [57]. *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 [58, 59]. 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.

Another promising gene, *UHRF1BP1L*, harbored a homozygous stopgain variant (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 [60].

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 [21] and this locus also harbors common PD susceptibility variants based on GWAS [7].

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 [61] and has also recently been implicated in mitochondrial dynamics [62], including interactions with Parkin [63]. 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*, four candidate genes from our WES analyses were implicated as LoF enhancers of α-synuclein neurotoxicity, which also has a central role in PD pathogenesis. We discuss the three 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 α-synuclein toxicity. In the single reported *VPS13C* PD case with a completed autopsy, neuropathological findings included abundant α-synuclein aggregates in both the brainstem and cortex [21]. Thus, *VPS13C* and associated endosomal sorting pathways (including *VPS35*) may represent a point of convergence for mitochondrial and α-synuclein-mediated PD mechanisms. Consistent with this, evidence for the impact of α-synuclein toxicity on mitochondria has recently emerged [28], including from studies in mammals [64].

In the IPDGC WES cohort, a single PD case was discovered with compound heterozygous LoF variants in *PTPRH* (p.Q887X and 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 [65, 66] 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 [67]. In *Drosophila*, RNAi-mediated knockdown of the conserved *PTPRH* ortholog, *Ptp10D*, enhanced α-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 [68]. More recent studies further implicated *Ptp10D* in neural-glial interactions during development of the central nervous system [69], 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 three other conserved fly *ARSB* homologs showed consistent interactions with  $\alpha$ -synuclein (Additional file 1: Table S7 and Table S8). In the IPDGC cohort, we discovered four 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 four 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 [70]. Substrate accumulation and associated cellular stress has been reported to induce markers of impaired autophagy and mitochondrial dysfunction in *ARSB* deficient fibroblasts from MPSVI patients, as in other lysosomal disorders [71, 72]. 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 [70, 73, 74]. Similar genotype–phenotype relationships have been documented for other lysosomal-storage disorders, including Gaucher's disease, which has established links with PD risk [75, 76].

While a full accounting is outside the scope of this study, at least one of the three 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 LoF screening using RNAi). 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 the 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, most of the variants implicated by the IPDGC WES cohort were represented at low frequencies within the largest available public database, ExAC [77, 78]; however, we have no information about potential PD manifestations in such individuals or even participant 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 false-positive findings. Importantly, experimental evidence of a genetic interaction with either mitochondrial dynamics or  $\alpha$ -synuclein-mediated neuronal injury in our screening assays cannot 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 [79, 80]. 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.

## Conclusions

We have identified five excellent PD gene candidates (*GPATCH2L*, *UHFR1BP1L*, *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 [21] 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 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

### Genetic analyses

#### Whole-exome sequencing

WES was performed on 1148 PD cases and 503 neurologically healthy controls of European descent. All participants 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 Additional file 2: Figure S8.

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, CA, USA) (62 Mb target); Roche (Basel, Switzerland) Nimblegen SeqCap (44.1 Mb target); and Agilent (Santa Clara, CA, USA) SureSelect (37.6 Mb target), which captured 96%, 81%, and 71% of the targeted exome at least ten times, respectively (Additional file 1: Table S12). Exome libraries were sequenced on a HiSeq 2000 (Illumina, San Diego, CA, USA). The Burrows Wheeler Aligner MEM v0.7.9.a [81] 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 [82], 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 [83] 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 [84].

#### Variant identification in IPDGC WES dataset

Considering the worldwide prevalence of 0.041% for PD in the age range of 40–49 years [20], 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 cutoff, 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* [85], the most frequent mutation is an exon 3 deletion, which has been identified in 16.4% of YOPD patients [86]. Using ANNOVAR [83], all variants were annotated with MAF information of ESP6500si (European American population) [87], 1000 Genomes Project (European population of April 2012 version) [88], and the ExAC browser (non-Finnish European population) [77, 78]. 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 [89] 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-site 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 GATKVQSR and hard filtering. Quality predictions based on the ExAC database are more adequate, as it includes ~37× 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 [90] database and with a CADD score > 20 (predicted to belong to the 1% most deleterious variants of the total genome), indicating a pathogenic effect. The latter two 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 [91] 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 [92] 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 1540

samples (1062 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 a 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 [83] 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; and (2) missense variants, including amino-acid changing and LoF variants.

As suggested by SKAT, we selected a MAF cutoff 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 [93] (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 PPMI [51]. 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 [83] and KGGSeq [89].

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

The southwest of the Netherlands contains a recently isolated population which is part of the GRIP program [52]. 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 two 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 [53] 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 6801 PD cases and 5970 neurologically healthy controls remained with high-quality genotype data for 178,779 variants. Based on the sample size, the MAF cutoff was 0.0063.

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

Approximately 70% of the participants included in this study have also been included in previous published GWAS [7, 94, 95]. 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 upstream and downstream of the gene of interest using the recent PD meta-analysis through pdgene.org [7]. Significant associations and suggestive *p* values < 1e-4 were considered. To understand the underlying linkage disequilibrium structure, LocusZoom [96] 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) [97]. 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 eigen-genes as centroids. The first GCN is based on 19,152 genes from 65 substantia nigra control brains from the UKBEC consortium. The gene expression profiles are based on Affymetrix Exon 1.0 ST Arrays [98]. The second GCN is based on 63 samples from the same tissue, GTEx [56] V6 gene RPKM values. Genes were filtered with a RPKM based cutoff of 0.2 and missingness < 30%

resulting in the analysis of 18,363 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 [99] 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 [100]. 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 [101].

#### **Human cellular screen**

##### **shRNA virus production**

Bacterial glycerol stocks containing the shRNA vectors (Sigma, St. Louis, MO, USA; TRC1 and 1.5) were grown overnight in Luria-Bertani media containing 100 µg/mL of ampicillin (Sigma-Aldrich, St. Louis, MO, USA). We selected at least five shRNA clones per gene. Endotoxin-free shRNA plasmids were extracted according to the manufacturer's protocol (Zymo, Irvine, CA, USA; ZR Plasmid Miniprep Classic kit). Lentivirus was produced as follows: HEK293T packaging cells were seeded at a density of  $4 \times 10^{45}$ /mL (100 µL per well) in cell culture media, Optimem (Invitrogen, Carlsbad, CA, USA) containing 10% fetal bovine serum (FBS) in 96-well tissue culture plates. Cells were incubated for 24 h (37 °C, 5% CO<sub>2</sub>). Each well was subsequently transfected with 100 ng of shRNA plasmid, 90 ng of packaging plasmid (pCMV-dr8.74psPAX2), and 10 ng of envelope plasmid (VSV-G/pMD2.G) combined with 0.6 µL of FugeneHD (Promega, Madison, WI, USA) in a total volume of 10 µL. Transfection efficiency was monitored using the pkLO.1 GFP plasmid (Sigma, St. Louis, MO, USA) and had to be greater than 90%. Sixteen hours after transfection, media was refreshed and supernatant harvested after a further 24 h. Virus was stored at -80 °C.

To ensure successful lentivirus production, HEK293T cells were plated out at a density of  $2 \times 10^{45}$ /mL (100 µL per well) in Optimem containing 10% FBS and 15 µg/mL of protamine sulfate (Sigma, St. Louis, MO, USA). Cells were infected with 10 µL, 25 µL, and 50 µL of

lentivirus. The following day, media was refreshed with media containing 2.5 µg/mL of puromycin. After a further three 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 293 T (ATCC® CRL-3216™) cell lines were obtained from the American Type Culture Collection (Manassas, VA, USA). BE(2)-M17 cell lines were cultured in Dulbecco's Modified Eagle/Nutrient Mixture F-12 Medium (DMEM/F-12) with GlutaMAX (Invitrogen, Carlsbad, CA, USA) supplemented with 10% FBS, 1× non-essential amino acids (NEAA), and 1% Penicillin/Streptomycin. HEK 293 T cells were cultured in Opti-MEM (Invitrogen, Carlsbad, CA, USA) containing 10% FBS and 1× NEAA. All cell lines were routinely tested for mycoplasma contamination. For lentivirus infection, 25 µL of the lentivirus was added to each well of a 96-well plates and protamine sulfate was added at a final concentration of 1 µg/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 two different assays:

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

For the Parkin translocation assay BE(2)-M17 cells were also utilized. The PLVX inducible vector (Clontech, Mountain View, CA, USA) 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 ng/mL Puromycin, 200 µg/mL G418, and 1 µg/mL of doxycycline. BE(2)-M17 cells were expanded and plated at a density of  $7.5 \times 10^4$ /mL (100 µL per well) in 96-well black CellCarrier plates (PerkinElmer, Waltham,

MA, USA) pre-pipetted with 25 µL 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 100 nM Mito-Tracker DeepRed (Molecular Probes, Eugene, OR, USA) and 1 µg/mL Hoechst. After 20 min, media was refreshed with media containing 15 µM Carbonyl cyanide m-chlorophenyl hydrazone (CCCP). Cells were incubated for 2 h 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 60× water immersion objective lens for improved resolution. Nuclei were imaged utilizing the 405 nm laser, Mitotracker CMXros utilizing the 561 nm laser, and mitotracker DeepRed utilizing the 640nm laser. For the translocation assay, a total of 60 fields per well were taken using a 20× objective lens. Nuclei were imaged utilizing the 405 nm laser, Parkin-GFP utilizing the 488 nm laser, and mitotracker DeepRed utilizing the 640 nm laser.

Images were stored and analyzed by the Columbus Image Data storage (PerkinElmer, Waltham, MA, USA). 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 were 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 six 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 [102]. Effects were considered significant when the SSMD\* normalized effect of shRNA treatment was greater than or less than 4 or -4 and at least two 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 [103], 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, WI, USA) 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, CA, USA) according to the manufacturer's specifications. Quantitative polymerase chain reaction (PCR) was carried out in triplicates on a ViiA7 real-time PCR system using SYBR Green PCR master mix (Life Technologies, Carlsbad, CA, USA) and 0.04 μM specific primer pairs for all targets. For multiple exons, gene 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 two animal models; *C. elegans* and *Drosophila*. The DRSC Integrated Ortholog Prediction Tool (DIOPT) [104] 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 two 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 α-synuclein transgenic flies with codon-optimization for *Drosophila* (*UAS-α-synuclein* line #7), were recently described [48] and are available from the Bloomington Stock Center (Bloomington, IN, USA). 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 Additional file 1: Table S8. The GAL4-UAS system [105] was used for ectopic co-expression of both the α-synuclein and RNAi transgene. The *Rh1-Gal4* driver line (second-chromosome insertion) has been previously described [48, 106]. 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 h of eclosion and aged 15 days. To examine for potential α-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>I</sup>*; (2) *w, Ptp10D<sup>I</sup>*; (3) *yw, Ptp4E<sup>I</sup>, Ptp10D<sup>I</sup> / FM7C*. All experimental results were quantified and photographed in female animals.

##### 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 40× objective on a Leica (Wetzlar, Germany) DM6000B

light microscope. Eyes from at least four animals were examined per genotype (at least eight retinae). 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 retinae, showing evidence of either reduced rhabdomere numbers or altered refraction of light indicative of vacuolar changes, by the total number of retinae examined. For identification of genetic enhancers, we required two independent RNAi lines targeting non-overlapping sequences with 50% or greater degenerate retinae observed using the pseudopupil assay. Following our initial screen of two 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-day-old animals were fixed in 8% glutaraldehyde and embedded in paraffin. Tangential (3  $\mu\text{m}$ ) retinal sections were cut using a Leica Microtome (RM2245) and stained with hematoxylin and eosin. Retinae from at least three 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 40 $\times$  objective from well-oriented, intact tangential sections at a depth in which the retina achieves maximal diameter. Using ImageJ software [107], we recorded the area occupied by all vacuoles with a diameter greater than 4  $\mu\text{m}$  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, 1:1000, BD Transduction Laboratories, San Diego, CA, USA).

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

All strains were maintained as described previously [108]. For this study, the worm strains N2 (wild-type), CF512 (*fer-15(b26)II; fem-1(hc17)III*), and OW40 (*zgIs15[P(unc-54)::α-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 [109]. RNAi clones targeting the genes of interest (9/27; Additional file 1: Table S3) 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)*) [110]. 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 1 h 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's 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 α-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-β-D-galactoside (IPTG, 15 mg/L) and 50  $\mu\text{g}/\text{mL}$  ampicillin (plates for RNAi treatment). Plates were seeded with RNAi bacteria. Prior to the experiment, the plates were kept at room temperature for two 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 s, after which the number of body bends was

counted for another 30 s. 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 three independent experiments and the relative body bends of one representative experiment is shown.

## Additional files

**Additional file 1:** Includes all 12 additional tables with every table on a separate spreadsheet. The file format is an Excel spreadsheet. (XLSX 183 kb)

**Additional file 2:** Includes all eight additional figures with every figure on a separate slide with the exception of Additional file 1: Figure S1 (divided over two slides). The file formats are PowerPoint (.pptx) and pdf. (PDF 2504 kb)

## Abbreviations

ESP6500: Exome Sequencing Project v. 6500; ExAC: Exome Aggregation Consortium; FBS: Fetal bovine serum; GCN: Gene co-expression network; GO: Gene ontology; GRIP: Genetic Research in Isolated Population; GTEx: The Genotype-Tissue Expression; GWAS: Genome-wide association study; indels: Insertions/deletions; IPDGC: International Parkinson's Disease Genomics Consortium; IR: Interfering RNA; LoF: Loss of function; MAF: Minor allele frequency; MPS VI: Mucopolysaccharidosis type VI; NGM: Nematode Growth medium; PD: Parkinson's disease; PPMI: Parkinson Progression Markers Initiative; RNAi: RNA-interference; SNVs: Single nucleotide variants; SSMD: Strictly standardized median difference; UKBEC: United Kingdom Brain Expression Consortium; WES: Whole-exome sequencing; WGCNA: Weighted gene co-expression network analysis; YOPD: Young-onset Parkinson's disease

## Acknowledgements

We would like to thank all the participants 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. 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.

### IPDGC consortium members and affiliations:

Mike A Nalls (Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA), Vincent Plagnol (UCL Genetics Institute, London, UK), Dena G Hernandez (Laboratory of Neurogenetics, National Institute on Aging; and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Manu Sharma (Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry and Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen Germany), Una-Marie Sheerin (Department of Molecular Neuroscience, UCL Institute of Neurology), Mohamad Saad (INSERM U563, CPTP, Toulouse, France; and Paul Sabatier University, Toulouse, France), Javier Simón-Sánchez (Genetics and Epigenetics of Neurodegeneration, German Center for Neurodegenerative Diseases (DZNE)-Tübingen and Hertie Institute for Clinical Brain Research (HIH)), Claudia Schulte (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research), Suzanne Lesage (Sorbonne Université, UPMC Univ Paris 06, UM 1127, ICM; Inserm, U 1127, ICM; Cnrs, UMR 7225, ICM; ICM, Paris), Sigurlaug Sveinbjörnsdóttir (Department of Neurology, Landspítali University Hospital, Reykjavík, Iceland; Department of Neurology, MEHT Broomfield Hospital, Chelmsford, Essex, UK; and Queen Mary College, University of London, London, UK), Sampath Arapalli (Laboratory of Neurogenetics, National Institute on Aging), Roger Barker (Department of Neurology, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK), Yoav Ben-Shlomo (School of Social and Community Medicine, University of Bristol), Henk W Berendse (Department of Neurology and Alzheimer Center, VU University Medical Center), Daniela Berg (Department for Neurodegenerative

Diseases, Hertie Institute for Clinical Brain Research and DZNE, German Center for Neurodegenerative diseases), Kailash Bhatia (Department of Motor Neuroscience, UCL Institute of Neurology), Rob M A de Bie (Department of Neurology, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands), Alessandro Biffi (Center for Human Genetic Research and Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; and Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA), Bas Bloem (Department of Neurology, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands), Zoltan Bochdanovits (Department of Clinical Genetics, Section of Medical Genomics, VU University Medical Centre), Michael Bonin (Department of Medical Genetics, Institute of Human Genetics, University of Tübingen, Tübingen, Germany), Jose M Bras (Department of Molecular Neuroscience, UCL Institute of Neurology), Kathrin Brockmann (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research and DZNE, German Center for Neurodegenerative diseases), Janet Brooks (Laboratory of Neurogenetics, National Institute on Aging), David J Burn (Newcastle University Clinical Ageing Research Unit, Campus for Ageing and Vitality, Newcastle upon Tyne, UK), Elisa Majounie (Laboratory of Neurogenetics, National Institute on Aging), Steven Lubbe (Department of Clinical Neuroscience, UCL Institute of Neurology, London, UK), Iris E Jansen (Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands, and German Center for Neurodegenerative diseases (DZNE), Tübingen, Germany), Ryan Price (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Aude Nicolas (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Gavin Charlesworth (Department of Molecular Neuroscience, UCL Institute of Neurology), Codrin Lungu (National Institutes of Health Parkinson Clinic, NINDS, National Institutes of Health), Honglei Chen (Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, NC, USA), Patrick F Chinnery (Neurology M4104, The Medical School, Framlington Place, Newcastle upon Tyne, UK), Sean Chong (Laboratory of Neurogenetics, National Institute on Aging), Carl E Clarke (School of Clinical and Experimental Medicine, University of Birmingham, Birmingham, UK; and Department of Neurology, City Hospital, Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, UK), Mark R Cookson (Laboratory of Neurogenetics, National Institute on Aging), J Mark Cooper (Department of Clinical Neurosciences, UCL Institute of Neurology), Jean Christophe Corvol (Sorbonne Université, UPMC Univ Paris 06, UM 1127, ICM; Inserm, U 1127, ICM; Cnrs, UMR 7225, ICM; ICM, Paris; and INSERM CIC-9503, Hôpital Pitié-Salpêtrière, Paris, France), Carl Counsell (University of Aberdeen, Division of Applied Health Sciences, Population Health Section, Aberdeen, UK), Philippe Damier (CHU Nantes, CIC0004, Service de Neurologie, Nantes, France), Jean-François Dartigues (INSERM U897, Université Victor Segalen, Bordeaux, France), Panos Deloukas (Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK), Günther Deuschl (Klinik für Neurologie, Universitätsklinikum Schleswig-Holstein, Campus Kiel), Christian-Albrechts-Universität Kiel, Kiel, Germany), David T Dexter (Parkinson's Disease Research Group, Faculty of Medicine, Imperial College London, London, UK), Karin D van Dijk (Department of Neurology and Alzheimer Center, VU University Medical Center), Allissa Dillman (Laboratory of Neurogenetics, National Institute on Aging), Frank Durif (Service de Neurologie, Hôpital Gabriel Montpied, Clermont-Ferrand, France), Alexandra Dürr (Sorbonne Université, UPMC Univ Paris 06, UM 1127, ICM; Inserm, U 1127, ICM; Cnrs, UMR 7225, ICM; ICM, Paris; and AP-HP, Pitié-Salpêtrière Hospital, Département de Génétique et Cytogénétique), Sarah Edkins (Wellcome Trust Sanger Institute), Jonathan R Evans (Cambridge Centre for Brain Repair, Cambridge, UK), Thomas Foltyne (UCL Institute of Neurology), Jing Dong (Epidemiology Branch, National Institute of Environmental Health Sciences), Michelle Gardner (Department of Molecular Neuroscience, UCL Institute of Neurology), J Raphael Gibbs (Laboratory of Neurogenetics, National Institute on Aging; and Department of Molecular Neuroscience, UCL Institute of Neurology), Alison Goate (Department of Psychiatry, Department of Neurology, Washington University School of Medicine, MI, USA), Emma Gray (Wellcome Trust Sanger Institute), Rita Guerreiro (Department of Molecular Neuroscience, UCL Institute of Neurology), Clare Harris (University of Aberdeen), Jacobus J van Hilten (Department of Neurology, Leiden University Medical Center, Leiden, Netherlands), Albert Hofman (Department of Epidemiology, Erasmus University Medical Center, Rotterdam, Netherlands), Albert Hollenbeck (AARP, Washington DC, USA), Janice Holton (Queen Square Brain Bank for Neurological Disorders, UCL Institute of Neurology), Michele Hu (Department of Clinical Neurology, John Radcliffe

Hospital, Oxford, UK), Xuemei Huang (Departments of Neurology, Radiology, Neurosurgery, Pharmacology, Kinesiology, and Bioengineering, Pennsylvania State University– Milton S Hershey Medical Center, Hershey, PA, USA), Isabel Wurster (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research and German Center for Neurodegenerative diseases), Walter Mätzler (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research and German Center for Neurodegenerative diseases), Gavin Hudson (Neurology M4104, The Medical School, Newcastle upon Tyne, UK), Sarah E Hunt (Wellcome Trust Sanger Institute), Johanna Huttenlocher (deCODE genetics), Thomas Illig (Institute of Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany), Pálmi V Jónsson (Department of Geriatrics, Landspítali University Hospital, Reykjavík, Iceland), Jean-Charles Lambert (INSERM U744, Lille, France; and Institut Pasteur de Lille, Université de Lille Nord, Lille, France), Cordelia Langford (Cambridge Centre for Brain Repair), Andrew Lees (Queen Square Brain Bank for Neurological Disorders), Peter Lichtner (Institute of Human Genetics, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany), Patricia Limousin (Institute of Neurology, Sobell Department, Unit of Functional Neurosurgery, London, UK), Grisel Lopez (Section on Molecular Neurogenetics, Medical Genetics Branch, NHGRI, National Institutes of Health), Delia Lorenz (Klinik für Neurologie, Universitätsklinikum Schleswig-Holstein), Codrin Lungu (National Institutes of Health Parkinson Clinic, NINDS, National Institutes of Health), Alasdair McNeill (Department of Clinical Neurosciences, UCL Institute of Neurology), Catriona Moorby (School of Clinical and Experimental Medicine, University of Birmingham), Matthew Moore (Laboratory of Neurogenetics, National Institute on Aging), Huw R Morris (National Hospital for Neurology and Neurosurgery, University College London, London, UK), Karen E Morrison (School of Clinical and Experimental Medicine, University of Birmingham; and Neurosciences Department, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK), Valentina Escott-Price (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK), Ese Mudanohwo (Neurogenetics Unit, UCL Institute of Neurology and National Hospital for Neurology and Neurosurgery), Sean S O'Sullivan (Queen Square Brain Bank for Neurological Disorders), Justin Pearson (MRC Centre for Neuropsychiatric Genetics and Genomics), Joel S Perlmutter (Department of Neurology, Radiology, and Neurobiology at Washington University, St Louis), Hjörvar Pétursson (deCODE genetics; and Department of Medical Genetics, Institute of Human Genetics, University of Tübingen), Pierre Pollak (Service de Neurologie, CHU de Grenoble, Grenoble, France), Bart Post (Department of Neurology, Radboud University Nijmegen Medical Centre), Simon Potter (Wellcome Trust Sanger Institute), Bernard Ravina (Translational Neurology, Biogen Idec, MA, USA), Tamás Revesz (Queen Square Brain Bank for Neurological Disorders), Olaf Riess (Department of Medical Genetics, Institute of Human Genetics, University of Tübingen), Fernando Rivadeneira (Departments of Epidemiology and Internal Medicine, Erasmus University Medical Center), Patrizia Rizzu (Department of Clinical Genetics, Section of Medical Genomics, VU University Medical Centre), Mina Ryten (Department of Molecular Neuroscience, UCL Institute of Neurology), Stephen Sawcer (University of Cambridge, Department of Clinical Neurosciences, Addenbrooke's hospital, Cambridge, UK), Anthony Schapira (Department of Clinical Neurosciences, UCL Institute of Neurology), Hans Scheffer (Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands), Karen Shaw (Queen Square Brain Bank for Neurological Disorders), Ira Shoulson (Department of Neurology, University of Rochester, Rochester, NY, USA), Joshua Shulman (Departments of Neurology, Molecular and Human Genetics, and Neuroscience, Baylor College of Medicine and The Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital), Ellen Sidransky (Section on Molecular Neurogenetics, Medical Genetics Branch, NHGRI), Colin Smith (Department of Pathology, University of Edinburgh, Edinburgh, UK), Chris C A Spencer (Wellcome Trust Centre for Human Genetics, Oxford, UK), Hreinn Stefánsson (deCODE genetics), Francesco Bettella (deCODE genetics), Joanna D Stockton (School of Clinical and Experimental Medicine), Amy Strange (Wellcome Trust Centre for Human Genetics), Kevin Talbot (University of Oxford, Department of Clinical Neurology, John Radcliffe Hospital, Oxford, UK), Carlie M Tanner (Clinical Research Department, The Parkinson's Institute and Clinical Center, Sunnyvale, CA, USA), Avazeh Tashakkori-Ghanbaria (Wellcome Trust Sanger Institute), François Tison

(Service de Neurologie, Hôpital HautLévèque, Pessac, France), Daniah Trabzuni (Department of Molecular Neuroscience, UCL Institute of Neurology), Bryan J Traynor (Laboratory of Neurogenetics, National Institute on Aging), André G Uitterlinden (Departments of Epidemiology and Internal Medicine, Erasmus University Medical Center), Daan Velseboer (Department of Neurology, Academic Medical Center), Marie Vidailhet (Sorbonne Université, UPMC Univ Paris 06, UM 1127, ICM; Inserm, U 1127, ICM; Cnrs, UMR 7225, ICM; ICM, Paris), Robert Walker (Department of Pathology, University of Edinburgh), Bart van de Warrenburg (Department of Neurology, Radboud University Nijmegen Medical Centre), Mirdhu Wickremaratchi (Department of Neurology, Cardiff University, Cardiff, UK), Nigel Williams (MRC Centre for Neuropsychiatric Genetics and Genomics), Caroline H Williams-Gray (Department of Neurology, Addenbrooke's Hospital), Sophie Winder-Rhodes (Department of Psychiatry and Medical Research Council and Wellcome Trust Behavioural and Clinical Neurosciences Institute, University of Cambridge), Kári Stefánsson (deCODE genetics), Maria Martinez (INSERM UMR 1043; and Paul Sabatier University), Nicholas W Wood (UCL Genetics Institute; and Department of Molecular Neuroscience, UCL Institute of Neurology), John Hardy (Department of Molecular Neuroscience, UCL Institute of Neurology), Peter Heutink (Department of Clinical Genetics, Section of Medical Genomics, VU University Medical Centre), Alexis Brice (Sorbonne Université, UPMC Univ Paris 06, UM 1127, ICM; Inserm, U 1127, ICM; Cnrs, UMR 7225, ICM; ICM, Paris; AP-HP, Hôpital de la Salpêtrière, Département de Génétique et Cytogénétique), Thomas Gasser (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, and DZNE, German Center for Neurodegenerative Diseases), Andrew B Singleton (Laboratory of Neurogenetics, National Institute on Aging).

## Funding

Funding for this study was provided by the Prinses Beatrix Spierfonds (IEJ, PH); the EU joint Program-Neurodegenerative Diseases (JPND): COURAGE-PD (PH, VD, SL, AB); the Federal Ministry of Education and Research Germany (BMBF): MitoPD (PH); NIH grants (K08AG034290, R21NS089854, R01AG033193, U01AG046161, C06RR029965, R01NS037167, R01CA141668, P50NS071674) (JMS, ABS); the Alzheimer's Association (JMS); the American Federation for Aging Research (JMS); Huffington Foundation (JMS); the Robert and Renee Belfer Family Foundation (JMS); the Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital (JMS); a Career Award for Medical Scientists from the Burroughs Wellcome Fund (JMS); the Wellcome Trust under awards 076113, 085475, and 090355 (SL, HM), Parkinson's UK (grants 8047, J-0804, F1002, and F-1201) (SL, HM); the Medical Research Council (G0700943 and G1100643) (SL, HM); the France-Parkinson Association (VD, SL, AB); the Roger de Spoelberch Foundation (R12123DD) (VD, SL, AB); the French Academy of Sciences (VD, SL, AB); the French program "Investissements d'avenir" (ANR-10-IAIHU-06) (VD, SL, AB); the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services (Z01 AG000957) (MN, JRG, ABS), the National Institute of Neurological Disorders and Stroke (NINDS) (Z01-AG000949-02), and the National Institute of Environmental Health Sciences (Z01-ES101986). Department of Defense (award W81XWH-09-2-0128); The Michael J Fox Foundation for Parkinson's Research; American Parkinson Disease Association (APDA); Barnes Jewish Hospital Foundation; Greater St Louis Chapter of the APDA; Hersenstichting Nederland; the German National Genome Network (NGFNplus number 01GS08134, German Ministry for Education and Research); the German Federal Ministry of Education and Research (NGFN 01GR0468, PopGen); 01EW0908 in the frame of ERA-NET NEURON and Helmholtz Alliance Mental Health in an Ageing Society (HA-215); European Community Framework Programme 7, People Programme; IAPP on novel genetic and phenotypic markers of Parkinson's disease, and Essential Tremor (MarkMD), contract number PIAP-GA-2008-230596 MarkMD. PPMI – a public–private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Avid, Biogen, Bristol-Myers Squibb, Covance, GE Healthcare, Gentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, Servier, Teva, UCB, and Golub Capital. The Pathology and Histology Core at Baylor College of Medicine is supported by NIH grant P30CA125123. The DIGPD cohort was sponsored by the Assistance Publique Hôpitaux de Paris and founded by the French clinical research hospital program-PHRC (code AOR08010). This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD, USA (<https://hpc.nih.gov/systems/>) and DNA panels, samples, and clinical data from the National Institute of

Neurological Disorders and Stroke Human Genetics Resource Center DNA and Cell Line Repository. The TRIP at Harvard Medical School, which provided fly stocks, is supported by R01GM084947.

The relevant funding bodies (above) did not participate in the design of the study; collection, analysis, or interpretation of data; or in drafting the manuscript.

### Availability of data and materials

Based on individual participant consents and institutional ethics committee approval (above), we are permitted to release WES data for 766 cases and 395 controls (70.3% of the IPDGC discovery cohort) into the public domain. These data are deposited in the National Institutes of Health database of Genotypes and Phenotypes (dbGaP; accession phs001103.v1.p2 (National Institute of Neurological Disorders and Stroke (NINDS) Exome Sequencing in Parkinson's Disease)) and the European Genome-phenome Archive (EGA; accession EGAS00001002103 (Whole-exome sequencing of Dutch Parkinson's disease patients), EGAS00001002110 (North American Brain Expression Consortium (NABEC) Exome Sequencing), EGAS00001002113 (Exome sequencing of United Kingdom Brain Expression Consortium samples), EGAS00001002156 (Whole-exome sequencing of Parkinson's disease patients from the United Kingdom)). For the remaining 382 cases and 108 controls within the IPDGC WES dataset, consent does not allow for public release of the data; investigators wishing to access these datasets should apply through the IPDGC webpage (<http://pdgenetics.org/resources>). Additional file 1: Table S13 provides a complete overview of the public availability of the IPDGC discovery cohort. The NeuroXdata have been deposited in dbGaP (accession phs000918.v1.p1 (International Parkinson's Disease Genomics Consortium (IPDGC), NeuroX Dataset)). PD GWAS summary statistics are accessible through <http://www.pdgene.org>, which includes a straightforward search on chromosomal region, gene name, or SNP identifier. We used the summary statistics of the pdgene.org website. For example, focusing on VPS13C (one of the genes we describe in the context of PD GWAS loci), we used the 1 MB region around VPS13C (chr15:61144588–63352664) as input for the search field, and copied and pasted the variants with *p* values below 1e-4 from the website in a text file. Then we converted the layout in a locus zoom format, so we were able to plot the GWAS variants together with our WES variants. Data from the PPMI WES cohort are available at <http://www.ppmi-info.org/access-data-specimens/download-data/>. Subsequent to approval of data access by PPMI, the PPMI WES dataset can be downloaded through an account login (DOWNLOAD > Genetic Data > Exome Sequencing > Data). All *Drosophila* strains used in this work are already available from public stock centers (Bloomington Drosophila Stock Center and Vienna Drosophila RNAi Center).

### Authors' contributions

IE: Conception/design, data analysis/acquisition, data interpretation, and writing manuscript. HY: Conception/design, data analysis/acquisition, data interpretation, and writing manuscript. SH: Data analysis/acquisition, data interpretation, and writing manuscript. ML: Data analysis/acquisition, data interpretation, and writing manuscript. HM: Data acquisition and revising manuscript. RS: Data acquisition and revising manuscript. SJL: Data acquisition and revising manuscript. VD: Data acquisition and revising manuscript. SL: Data acquisition and revising manuscript. EM: Data analysis/acquisition and revising manuscript. JRG: Data analysis/acquisition and revising manuscript. MAN: Data analysis/acquisition and revising manuscript. MR: Data analysis/acquisition and revising manuscript. JAB: Data analysis/acquisition and writing/revising manuscript. JV: Data analysis/acquisition and revising manuscript. JSS: Data analysis/acquisition and revising manuscript. MCL: Data analysis/acquisition and revising manuscript. PR: Data analysis/acquisition and revising manuscript. CB: Data analysis and revising manuscript. AKC: Conception/design, data acquisition and revising manuscript. YL: Data acquisition and revising manuscript. PY: Data acquisition and revising manuscript. International Parkinson's Disease Genetics Consortium: Data acquisition and revising manuscript. HRM: Data acquisition and revising manuscript. AB: Data acquisition and revising manuscript. ABS: Data acquisition and revising manuscript. DCD: Data acquisition and revising manuscript. EAN: Data acquisition and revising manuscript. SJ: Data analysis/acquisition, data interpretation, and writing manuscript. JMS: Conception/design, data interpretation, and writing manuscript. PH: Conception/design, data interpretation, and writing manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Ethics approval and consent to participate

All human participants contributing clinical data and genetic samples to this study provided written informed consent, subject to oversight by the relevant institutional review boards and ethical committees at each study site, including the NIH Office of Human Subjects Research (protocols: 4813; 12250), NIA IRB (protocol: 03-AG-N39), National Hospital for Neurology and Neurosurgery ethics committee (protocol: 10/H07166/3), Wales Research Ethics Committee 3 (protocols: 14/WA/1179; 05MRE09/58), INSERM Comité de Protection des Personnes (protocols: 00001072; RBM 03-48), CMO Regio Arnhem-Nijmegen (protocol: 2005/190) and LUMC Commissie Medische Ethiek (protocol: P03.140). All experimental methods pertaining to human participants are in accordance with the Helsinki Declaration.

### Author details

<sup>1</sup>German Center for Neurodegenerative Diseases (DZNE), Otfried-Müller-Str. 23, Tübingen 72076, Germany. <sup>2</sup>Department of Clinical Genetics, VU University Medical Center, Amsterdam 1081HZ, The Netherlands.

<sup>3</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. <sup>4</sup>Graduate School of Cellular & Molecular Neuroscience, Tübingen 72074, Germany. <sup>5</sup>European Research Institute for the Biology of Aging, University of Groningen, University Medical Centre Groningen, Groningen 9700AD, The Netherlands. <sup>6</sup>Department of Clinical Neuroscience, UCL Institute of Neurology, London, UK. <sup>7</sup>Northwestern University Feinberg School of Medicine, Ken and Ruth Davee Department of Neurology, Chicago, IL, USA. <sup>8</sup>Inserm U1127, CNRS UMR7225, Sorbonne Universités, UPMC Univ Paris 06, UMR\_S1127, Institut du Cerveau et de la Moelle épinière, Paris, France. <sup>9</sup>Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK. <sup>10</sup>Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA. <sup>11</sup>Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK. <sup>12</sup>Department of Medical & Molecular Genetics, King's College London, London, UK. <sup>13</sup>Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany. <sup>14</sup>Department of Neurology, Baylor College of Medicine, Houston, TX, USA. <sup>15</sup>Genetic Epidemiology Unit, Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands. <sup>16</sup>Assistance Publique Hôpitaux de Paris, Hôpital de la Salpêtrière, Département de Génétique et Cytogénétique, Paris, France. <sup>17</sup>Department of Neuroscience and Program in Developmental Biology, Baylor College of Medicine, Houston, TX, USA. <sup>18</sup>Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, 1250 Moursund St., N.1150, Houston, TX 77030, USA.

Received: 17 June 2016 Accepted: 3 January 2017

Published online: 30 January 2017

### References

- Zimprich A, Benet-Pages A, Struhal W, Graf E, Eck SH, Offman MN, et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *Am J Hum Genet*. 2011;89:168–75.
- Vilarino-Guell C, Widner C, Ross OA, Dachsel JC, Kachergus JM, Lincoln SJ, et al. VPS35 mutations in Parkinson disease. *Am J Hum Genet*. 2011;89:162–7.
- Funayama M, Ohe K, Amo T, Furuya N, Yamaguchi J, Saiki S, et al. CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. *Lancet Neurol*. 2015;14:274–82.
- Farlow JL, Robak LA, Hetrick K, Bowling K, Boerwinkle E, Coban-Akdemir ZH, et al. Whole-exome sequencing in familial Parkinson disease. *JAMA Neurol*. 2016;73:68–75.
- Shulman JM, De Jager PL, Feany MB. Parkinson's disease: genetics and pathogenesis. *Annu Rev Pathol*. 2011;6:193–222.
- Trinh J, Farrer M. Advances in the genetics of Parkinson disease. *Nat Rev Neurol*. 2013;9:445–54.
- Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet*. 2014;46:989–93.
- Hamza TH, Payami H. The heritability of risk and age at onset of Parkinson's disease after accounting for known genetic risk factors. *J Hum Genet*. 2010;55:241–3.

9. Keller MF, Saad M, Bras J, Bettella F, Nicolaou N, Simon-Sanchez J, et al. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum Mol Genet.* 2012;21:4996–5009.
10. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature.* 2012;488:96–9.
11. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med.* 2013;368:107–16.
12. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogeava E, Majounie E, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med.* 2013;368:117–27.
13. Smith BN, Ticiczi N, Fallini C, Gkazi AS, Topp S, Kenna KP, et al. Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron.* 2014;84:324–31.
14. Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science.* 2015;347:1436–41.
15. Moutsianas L, Agarwala V, Fuchsberger C, Flannick J, Rivas MA, Gaulton KJ, et al. The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. *PLoS Genet.* 2015;11:e1005165.
16. Sulem P, Helgason H, Oddson A, Stefansson H, Gudjonsson SA, Zink F, et al. Identification of a large set of rare complete human knockouts. *Nat Genet.* 2015;47:448–52.
17. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature.* 1998;392:605–8.
18. Bonifati V, Rizzu P, Squitieri F, Krieger E, Vanacore N, van Swieten JC, et al. DJ-1(PARK7), a novel gene for autosomal recessive, early onset parkinsonism. *Neurology.* 2003;61:159–60.
19. Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, Gispert S, et al. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science.* 2004;304:1158–60.
20. Pringsheim T, Jette N, Frolkis A, Steeves TD. The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Mov Disord.* 2014;29:1583–90.
21. Lesage S, Drouet V, Majounie E, Deramecourt V, Jacoupy M, Nicolas A, et al. Loss of VPS13C function in autosomal-recessive Parkinsonism causes mitochondrial dysfunction and increases PINK1/Parkin-dependent mitophagy. *Am J Hum Genet.* 2016;98:500–13.
22. Ghahramani Seno MM, Kwan BY, Lee-Ng KK, Moessner R, Lionel AC, Marshall CR, et al. Human PTCHD3 nulls: rare copy number and sequence variants suggest a non-essential gene. *BMC Med Genet.* 2011;12:45.
23. Newsome TP, Schmidt S, Dietzl G, Keleman K, Asling B, Debant A, et al. Trio combines with dock to regulate Pak activity during photoreceptor axon pathfinding in *Drosophila*. *Cell.* 2000;101:283–94.
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;8:580–93.
25. Ma XM, Kiraly DD, Gaier ED, Wang Y, Kim EJ, Levine ES, et al. Kalirin-7 is required for synaptic structure and function. *J Neurosci.* 2008;28:12368–82.
26. Mandela P, Yankova M, Conti LH, Ma XM, Grady J, Eipper BA, et al. Kalrn plays key roles within and outside of the nervous system. *BMC Neurosci.* 2012;13:136.
27. Greenamyre JT, Hastings TG. Biomedicine. Parkinson's—divergent causes, convergent mechanisms. *Science.* 2004;304:1120–2.
28. Haelterman NA, Yoon WH, Sandoval H, Jaiswal M, Shulman JM, Bellen HJ. A mitocentric view of Parkinson's disease. *Annu Rev Neurosci.* 2014;37:137–59.
29. Pickrell AM, Youle RJ. The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. *Neuron.* 2015;85:257–73.
30. Cookson MR. Parkinsonism due to mutations in PINK1, parkin, and DJ-1 and oxidative stress and mitochondrial pathways. *Cold Spring Harb Perspect Med.* 2012;2:a009415.
31. Narendra D, Tanaka A, Suen DF, Youle RJ. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *J Cell Biol.* 2008;183:795–803.
32. Kamp F, Exner N, Lutz AK, Wender N, Hegermann J, Brunner B, et al. Inhibition of mitochondrial fusion by alpha-synuclein is rescued by PINK1, Parkin and DJ-1. *EMBO J.* 2010;29:3571–89.
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 A.* 2006;69:1–12.
34. Chang CR, Blackstone C. Dynamic regulation of mitochondrial fission through modification of the dynamin-related protein Drp1. *Ann N Y Acad Sci.* 2010;1201:34–9.
35. Narendra DP, Jin SM, Tanaka A, Suen DF, Gautier CA, Shen J, et al. PINK1 is selectively stabilized on impaired mitochondria to activate Parkin. *PLoS Biol.* 2010;8:e1000298.
36. Vives-Bauza C, Zhou C, Huang Y, Cui M, de Vries RL, Kim J, et al. PINK1-dependent recruitment of Parkin to mitochondria in mitophagy. *Proc Natl Acad Sci U S A.* 2010;107:378–83.
37. Geisler S, Holmstrom KM, Skujat D, Fiesel FC, Rothfuss OC, Kahle PJ, et al. PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nat Cell Biol.* 2010;12:119–31.
38. Vincow ES, Merrihew G, Thomas RE, Shulman NJ, Beyer RP, MacCoss MJ, et al. The PINK1-Parkin pathway promotes both mitophagy and selective respiratory chain turnover in vivo. *Proc Natl Acad Sci U S A.* 2013;110:6400–5.
39. Feany MB, Bender WW. A *Drosophila* model of Parkinson's disease. *Nature.* 2000;404:394–8.
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.* 2002;295:865–8.
41. MacLeod DA, Rhinn H, Kuwahara T, Zolin A, Di Paolo G, McCabe BD, et al. RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. *Neuron.* 2013;77:425–39.
42. Chen L, Feany MB. Alpha-synuclein phosphorylation controls neurotoxicity and inclusion formation in a *Drosophila* model of Parkinson disease. *Nat Neurosci.* 2005;8:657–63.
43. Cullen V, Lindfors M, Ng J, Paetau A, Swinton E, Kolodziej P, et al. Cathepsin D expression level affects alpha-synuclein processing, aggregation, and toxicity in vivo. *Mol Brain.* 2009;2:5.
44. Petrucci L, O'Farrell C, Lockhart PJ, Baptista M, Kehoe K, Vink L, et al. Parkin protects against the toxicity associated with mutant alpha-synuclein: proteasome dysfunction selectively affects catecholaminergic neurons. *Neuron.* 2002;36:1007–19.
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;37:911–24.
46. Miura E, Hasegawa T, Konno M, Suzuki M, Sugeno N, Fujikake N, et al. VPS35 dysfunction impairs lysosomal degradation of alpha-synuclein and exacerbates neurotoxicity in a *Drosophila* model of Parkinson's disease. *Neurobiol Dis.* 2014;71:1–13.
47. Dhungel N, Eleuteri S, Li LB, Kramer NJ, Chartron JW, Spencer B, et al. Parkinson's disease genes VPS35 and EIF4G1 interact genetically and converge on alpha-synuclein. *Neuron.* 2015;85:76–87.
48. Chouhan AK, Guo C, Hsieh Y-C, Ye H, Senturk M, Zuo Z, et al. Uncoupling neuronal death and dysfunction in *Drosophila* models of neurodegenerative disease. *Acta Neuropathol Comm.* 2016;4:62.
49. van Ham TJ, Holmberg MA, van der Goot AT, Teuling E, Garcia-Arenzibia M, Kim HE, et al. Identification of MOAG-4/SERF as a regulator of age-related proteotoxicity. *Cell.* 2010;142:601–12.
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 Dev.* 2008;3:.
51. Parkinson Progression Marker Initiative. The Parkinson Progression Marker Initiative (PPMI). *Prog Neurobiol.* 2011;95:629–35.
52. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS. The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet.* 2005;69:288–95.
53. Nalls MA, Bras J, Hernandez DG, Keller MF, Majounie E, Renton AE, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. *Neurobiol Aging.* 2015;36(1605):e7–e12.
54. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010;42:565–9.
55. Xu J, Bai J, Zhang X, Lv Y, Gong Y, Liu L, et al. A comprehensive overview of lncRNA annotation resources. *Brief Bioinform.* 2016. doi:10.1093/bib/bbw015.
56. GTEx-Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;348:648–60.
57. Aravind L, Koonin EV. G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends Biochem Sci.* 1999;24:342–4.

58. Lin ML, Fukukawa C, Park JH, Naito K, Kijima K, Shimo A, et al. Involvement of G-patch domain containing 2 overexpression in breast carcinogenesis. *Cancer Sci.* 2009;100:1443–50.
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. *Mol Med Rep.* 2015;11:1252–7.
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.* 2010;11:688–705.
61. Wang S, Bellen HJ. The retromer complex in development and disease. *Development.* 2015;142:2392–6.
62. Wang W, Wang X, Fujioka H, Hoppel C, Whone AL, Caldwell MA, et al. Parkinson's disease-associated mutant VPS35 causes mitochondrial dysfunction by recycling DLP1 complexes. *Nat Med.* 2016;22:54–63.
63. Song P, Trajkovic K, Tsunemi T, Krainc D. Parkin modulates endosomal organization and function of the endo-lysosomal pathway. *J Neurosci.* 2016;36:2425–37.
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. *J Neurosci.* 2015;35:890–905.
65. Matuzaki T, Suzuki T, Uchida T, Inazawa J, Ariyama T, Matsuda K, et al. Molecular cloning of a human transmembrane-type protein tyrosine phosphatase and its expression in gastrointestinal cancers. *J Biol Chem.* 1994;269:2075–81.
66. Matuzaki 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. *Cell Signal.* 2010;22:1811–7.
67. Takahashi H, Craig AM. Protein tyrosine phosphatases PTPdelta, PTPsigma, and LAR: presynaptic hubs for synapse organization. *Trends Neurosci.* 2013;36:522–34.
68. Qian M, Pan G, Sun L, Feng C, Xie Z, Tully T, et al. Receptor-like tyrosine phosphatase PTP1D is required for long-term memory in *Drosophila*. *J Neurosci.* 2007;27:4396–402.
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;78:183–26.
70. Valayannopoulos V, Nicely H, Harmatz P, Turbeville S. Mucopolysaccharidosis VI. *Orphanet J Rare Dis.* 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;24.
72. Lieberman AP, Puertollano R, Raben N, Slaugenhaupt S, Walkley SU, Ballabio A. Autophagy in lysosomal storage disorders. *Autophagy.* 2012;8:719–30.
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. *Mol Genet Metab.* 2005;85:236–8.
74. Karageorgos L, Brooks DA, Pollard A, Melville EL, Hein LK, Clements PR, et al. Mutational analysis of 105 mucopolysaccharidosis type VI patients. *Hum Mutat.* 2007;28:897–903.
75. Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, Barbosa ER, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med.* 2009;361:1651–61.
76. Sidransky E, Lopez G. The link between the GBA gene and parkinsonism. *Lancet Neurol.* 2012;11:986–98.
77. Exome Aggregation Consortium (ExAC) C, MA. <http://exac.broadinstitute.org> [April 2015].
78. Lek M, Karczewski K, Minikel E, Samocha K, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536:285–91.
79. Dawson TM, Ko HS, Dawson VL. Genetic animal models of Parkinson's disease. *Neuron.* 2010;66:646–61.
80. Shulman JM. *Drosophila* and experimental neurology in the post-genomic era. *Exp Neurol.* 2015;274:4–13.
81. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60.
82. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43:491–8.
83. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38:e164.
84. Kircher M, Witten DM, Jain P, O'Rourke BJ, Cooper GM. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46:310–5.
85. Klein C, Westenberger A. Genetics of Parkinson's disease. *Cold Spring Harb Perspect Med.* 2012;2:a008888.
86. Grunewald A, Kasten M, Ziegler A, Klein C. Next-generation phenotyping using the parkin example: time to catch up with genetics. *JAMA Neurol.* 2013;70:1186–91.
87. Exome Variant Server NGESPE, Seattle, WA. <http://evs.gs.washington.edu/EVS/> [September 2013 and April 2015].
88. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491:56–65.
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 Res.* 2012;40:e53.
90. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigelski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29:308–11.
91. Robinson JT, Thorvaldsdottir H, Winckler W, Guttmann M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29:24–6.
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. *Am J Hum Genet.* 2013;92:841–53.
93. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
94. Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet.* 2009;41:1308–12.
95. International Parkinson's Disease Genomics Consortium, Wellcome Trust Case Control Consortium. A two-stage meta-analysis identifies several new loci for Parkinson's disease. *PLoS Genet.* 2011;7:e1002142.
96. Pruij RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010;26:2336–7.
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, Walker R, Smith C, Bras J, et al. Insights into TREM2 biology by network analysis of human brain gene expression data. *Neurobiol Aging.* 2013;34:2699–714.
99. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. gProfiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* 2007;35:W193–200.
100. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? *PLoS Comput Biol.* 2011;7:e1001057.
101. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504.
102. Zhang XD. Illustration of SSMD, z score, SSMD\*, z\* score, and t statistic for hit selection in RNAi high-throughput screens. *J Biomol Screen.* 2011;16:775–85.
103. Zhang JH, Chung TD, Oldenburg KR. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen.* 1999;4:67–73.
104. Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, 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.* 1993;118:401–15.
106. Xiong B, Bayat V, Jaiswal M, Zhang K, Sandoval H, Chang WL, et al. Crag is a GEF for Rab11 required for rhodopsin trafficking and maintenance of adult photoreceptor cells. *PLoS Biol.* 2012;10:e1001438.
107. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012;9:671–5.
108. Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics.* 1974;77:71–94.
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 Genet.* 2005;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;161:1101–12.