


RESEARCH ARTICLE

Open Access

Identification of sixteen novel candidate genes for late onset Parkinson's disease



Alessandro Gialluisi^{1†}, Mafalda Giovanna Reccia^{1†}, Nicola Modugno¹, Teresa Nutile², Alessia Lombardi¹, Luca Giovanni Di Giovannantonio², Sara Pietracupa¹, Daniela Ruggiero^{1,2}, Simona Scala¹, Stefano Gambardella^{1,3}, International Parkinson's Disease Genomics Consortium (IPDGC), Licia Iacoviello^{1,4}, Fernando Gianfrancesco², Dario Acampora², Maurizio D'Esposito², Antonio Simeone², Marina Ciullo^{1,2} and Teresa Esposito^{1,2*} 

Abstract

Background: Parkinson's disease (PD) is a neurodegenerative movement disorder affecting 1–5% of the general population for which neither effective cure nor early diagnostic tools are available that could tackle the pathology in the early phase. Here we report a multi-stage procedure to identify candidate genes likely involved in the etiopathogenesis of PD.

Methods: The study includes a discovery stage based on the analysis of whole exome data from 26 dominant late onset PD families, a validation analysis performed on 1542 independent PD patients and 706 controls from different cohorts and the assessment of polygenic variants load in the Italian cohort (394 unrelated patients and 203 controls).

Results: Family-based approach identified 28 disrupting variants in 26 candidate genes for PD including *PARK2*, *PINK1*, *DJ-1(PARK7)*, *LRRK2*, *HTRA2*, *FBXO7*, *EIF4G1*, *DNAJC6*, *DNAJC13*, *SNCAIP*, *AIMP2*, *CHMP1A*, *GIPC1*, *HMOX2*, *HSPA8*, *IMMT*, *KIF21B*, *KIF24*, *MAN2C1*, *RHOT2*, *SLC25A39*, *SPTBN1*, *TMEM175*, *TOMM22*, *TVP23A* and *ZSCAN21*. Sixteen of them have not been associated to PD before, were expressed in mesencephalon and were involved in pathways potentially deregulated in PD. Mutation analysis in independent cohorts disclosed a significant excess of highly deleterious variants in cases ($p = 0.0001$), supporting their role in PD.

Moreover, we demonstrated that the co-inheritance of multiple rare variants (≥ 2) in the 26 genes may predict PD occurrence in about 20% of patients, both familial and sporadic cases, with high specificity ($> 93\%$; $p = 4.4 \times 10^{-5}$). Moreover, our data highlight the fact that the genetic landmarks of late onset PD does not systematically differ between sporadic and familial forms, especially in the case of small nuclear families and underline the importance of rare variants in the genetics of sporadic PD.

Furthermore, patients carrying multiple rare variants showed higher risk of manifesting dyskinesia induced by levodopa treatment.

* Correspondence: teresa.esposito@igb.cnr.it

[†]Alessandro Gialluisi and Mafalda Giovanna Reccia contributed equally to this work.

¹IRCCS Istituto Neurologico Mediterraneo Neuromed, Pozzilli, Isernia, Italy

²National Research Council, Institute of Genetics and Biophysics "Adriano Buzzati-Traverso", Naples, Italy

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



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions: Besides confirming the extreme genetic heterogeneity of PD, these data provide novel insights into the genetic of the disease and may be relevant for its prediction, diagnosis and treatment.

Keywords: Late onset Parkinson's disease, Whole exome sequencing, Novel candidate genes for Parkinson's disease, Rare variant burden analysis

Background

Parkinson's disease (PD) is a neurodegenerative movement disorder characterized by the loss of mesodiencephalic dopaminergic (mdDA) neurons of the substantia nigra pars compacta (SNpc), in association with the presence of Lewy bodies in some surviving neurons [1, 2]. mdDA neurons play a crucial role in the control of motor, sensory-motor and motivated behavior [3]. Their degeneration causes the characteristic symptoms of PD, which include resting tremor, bradykinesia, rigidity, postural instability and a variety of other motor and non-motor symptoms [2]. It has been demonstrated that disease's symptoms are evident when patients have already lost 50–60% of their DA neurons, suggesting the need for early diagnosis of the disease [2]. Most PD cases are sporadic, with unknown etiology. Approximately 5–10% of PD cases appear to have monogenic forms of inheritance including autosomal dominant (*SNCA*, *LRRK2*, and *VPS35*) and recessive (*PARK2*, *PINK1*, *DJ1*) forms [4–6]. However, many additional loci associated to PD have recently been identified, even though for most of them the molecular function is still unknown [5]. Moreover, there is a high genetic heterogeneity at the basis of PD, with many different rare mutations usually detected only in a single family or in small populations [5]. Despite extensive ongoing studies, the pathogenic mechanisms underpinning PD remain largely elusive [6]. Moreover, although the Next Generation Sequencing (NGS) methods are widespread also in the diagnostic field, their application to PD is still limited to a small number of patients with a clear family history and to a small number of causative mutations, due to the high genetic heterogeneity associated with the disease and to the difficulty in interpreting test results.

In this study, we performed a multi-stage procedure aimed at the identification of potential PD-causative variants and then at demonstrating that the co-inheritance of multiple rare variants in Mendelian genes may increase the risk of PD in a non-Mendelian fashion, and that this may influence the occurrence of the disease and the manifestation of clinical signs.

Patients and methods

Study participants

PIB cohort

Twenty-three PD families with supposedly dominant transmission (Fig. 1, families numbered 1 through 23)

were recruited at the “Parkinson Institute Biobank,” member of the Telethon Network of Genetic Biobank (biobanknetwork.telethon.it/) (hereafter-called PIB cohort). Forty-seven PD patients (siblings or parent-offspring pairs) from these families underwent whole-exome sequencing (WES) analysis to identify novel PD candidate genes. Forty PD patients (of which only a small amount of DNA was available), from the same 23 families, were used for segregation analysis to confirm the variants identified through WES approach. All PD subjects (age at onset 61.7; Standard Deviation (SD) \pm 8.57) were evaluated by neurologists of the Parkinson Study Group, according to published diagnostic criteria (Table S1) [7]. Written informed consent was obtained from all participants.

MNI cohort

Three PD families with supposedly dominant transmission were recruited at the IRCCS Mediterranean Neurological Institute (MNI) in Pozzilli (hereafter called MNI cohort) (Fig. 1; families numbered 24 through 26). Two PD cases for each family underwent WES analysis to identify novel PD candidate genes. Segregation analysis to confirm WES results was performed on PD cases from the same families (Fig. 1).

Three hundred ninety-four independent and unrelated PD patients (243 males; 172 familiar and 222 sporadic cases), were used as validation cohort, as well as to assess the polygenic inheritance of rare variants as risk factor for PD occurrence and for endophenotypes manifestation (Table S1) [8]. The project was approved by the ethical committees of IRCCS Neuromed and written informed consent was signed by all the participants. All subjects were of European ancestry and were evaluated by qualified neurologist of the Parkinson Centre from June 2015 to December 2017, with a thorough protocol comprising neurological examination and evaluation of non-motor domains. Information about family history, demographic characteristics, anamnesis, and pharmacological therapy was also collected [8, 9]. The mean age at diagnosis was 58.26 years (SD 9.74) [8, 9].

The Movement Disorder Society revised version of the Unified Parkinson's Disease Rating Scale Part III (18 items, maximum score 72; hereafter called UPDRS) [10] was used to assess clinical motor symptoms. These included language, facial expressions, tremor, rigidity, agility in movements, stability, gait and bradykinesia. Cognitive abilities were tested through an Italian

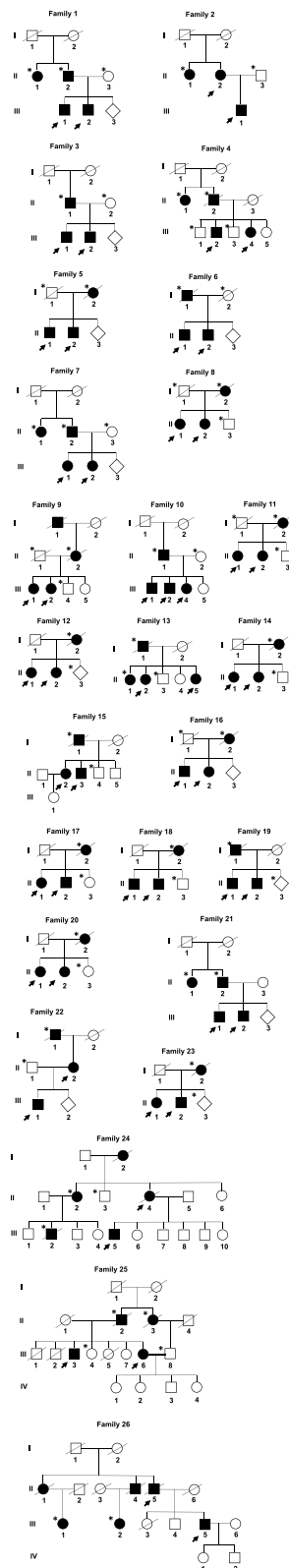


Fig. 1 Graphical representation of the 26 families used as discovery cohort. Pedigrees of the 26 Italian families with supposedly dominant PD forms. Affected individuals are indicated with dark symbols. Patients who underwent WES analysis are shown with dark arrows. Additional family members for which DNA was available are indicated with an asterisk. Only affected subjects were used for segregation analysis. For each family, the rare deleterious variants reported in Table 1 were carried by the affected family members

validated version of the Montreal Cognitive Assessment (MoCA) [11]. Cognitive domains assessed include short-term memory (5 points); visuospatial abilities via clock drawing (3 points), and a cube copy task (1 point); executive functioning via an adaptation of Trail Making Test Part B (1 point), phonemic fluency (1 point), and verbal abstraction (2 points); attention, concentration, and working memory via target detection (1 point), serial subtraction (2 points), digits forward and backward (1 point each); language via confrontation naming with low-familiarity animals (3 points), and repetition of complex sentences (2 points); and orientation to time and place (6 points). The total score was given by the sum of these domains, and then divided by the maximum score that could be obtained (30 points). Where one or more domains could not be tested (e.g. visuospatial tasks, due to unavailability of optical devices), the corresponding score was subtracted from the total score obtainable.

Non-motor symptoms were assessed through an Italian validated version of Non Motor Symptoms Scale (NMMS) for Parkinson Disease [12]. This scale tests 9 items, including cardiovascular domain, sleep/fatigue, mood/cognition, perceptual problems/hallucinations, attention/memory, gastrointestinal, urinary, sexual function, and ability to taste or smell. For each item, both severity and frequency of symptoms is measured, so that the scale accounts for both aspects. Here, the sleep domain was slightly modified by adding a further question on the occurrence of vivid dreams. This question was treated as all the others, i.e. the severity of impairment was scored from 0 (no symptoms) to 3 (severe impairment), and the frequency of impairment was scored from 0 (less than once a week) to 4 (daily impairment), then the total score of the sub-item was computed as the product of severity by frequency, and added to the scores of the other sub-items. In the endophenotypes analysis we used UPDRS score < 30 (corresponding to a mild phenotype: 1–2 of Hoehn and Yahr (HY) scale) and ≥ 30 (corresponding to a severe phenotype: 3–5 HY scale) [13], while in non-motor symptoms we used as cutoff ≤ 54 and > 54 (9 items \times 2 (mild impairment) \times 3 (weekly impairment)). Statistical

significance was set to $\alpha = 0.01$, correcting for five phenotypes tested.

Italian control cohort

As general Italian population, we downloaded whole genome data of 107 samples of the Tuscan Italians (TSI) population of the 1000 Genome Project (phase 3 release) [14] and extracted the exome regions covered by the same capture kit used in our sequencing experiments (SureSelect All Exome kit v6; Agilent® Technologies, Santa Clara, CA, USA).

Moreover, 96 neurological controls, including 38 from MNI and 58 from the Moli-sani genetic biobank (mean age 77 years; SD 5.4; 45 women) [15], underwent targeted resequencing (NGS-TR) of the panel of 26 PD genes identified in this study.

International cohort

WES data of 1148 young-onset unrelated PD cases (average age at onset 40.6 years; range 35–56 years) and 503 control participants of European ancestry from the International Parkinson's Disease Genomics Consortium (IPDGC) were used to replicate the observations made in the Italian cohort. Details on this dataset and on the QC carried out on these samples are reported elsewhere [16].

Next generation sequencing (NGS)

Genomic DNA was isolated from peripheral blood lymphocytes by PAX gene Blood DNA Midi Kit (QIAGEN, Hilden, Germany).

Whole-exome sequencing (WES)

WES was performed on 47 affected individuals (23 families) from PIB cohort, 6 affected individuals (3 families) and 106 independent and unrelated PD cases from MNI cohort. Exonic regions were enriched using the SureSelect All Exome kit v6 (Agilent® Technologies, Santa Clara, CA, USA) based on DNA fragmentation and capture. Exomes were barcoded and sequenced at Helmholtz Zentrum, München, Germany, using the Illumina® HiSeq2000 platform.

Targeted resequencing (NGS-TR)

NGS-TR was performed on 288 independent and unrelated PD cases from MNI cohort and on 96 Italian healthy individuals. Probes specific for 100 genes including the 26-targeted genes, identified in this study, and GBA were designed with Nimble Design software (Roche Diagnostics, Mannheim Germany). Targeting regions were enriched using the SeqCap kit based on DNA fragmentation (Kapa Hyper plus kit) and capture (Roche Diagnostics, Mannheim Germany). Targeted regions were barcoded and sequenced on MiSeq platform (Illumina, San Diego, CA, US).

Quality control and variant annotation

The alignments of the 100-bp paired-end reads to the human reference genome was performed by using the Burrows Wheeler Aligner (BWA) MEM v0.7.5 [17]. After removal of duplicate reads through Picard Mark-Duplicates command (with standard options), we called the single nucleotide variants (SNVs) and insertions/deletions (indels) for all samples using HaplotypeCaller (BP_RESOLUTION option) and GenotypeGVCFs in Genome Analysis Toolkit (GATK) v3.5-0-g36282e4, following the manufacturer best practice guidelines (available at <https://software.broadinstitute.org/gatk/best-practices/>) [18]. Variants with Minor Allele Count (MAC) = 0, number of alternative alleles $\neq 2$ and call rate < 95% were also filtered out, as well as samples with identical-by-descent (IBD) sharing and sex mismatches, and samples with call rate < 90%. Similarly, samples were checked for absence of outliers in terms of genome-wide homozygosity, of number of singleton variants and of genetic ancestry (through Multidimensional Scaling Analysis in PLINK) [19]. Variants passing quality control were annotated to genes (within 10 kb from transcription start/stop site) through ANNOVAR [20]. Variant annotation contained information concerning variant type, Minor Allele Frequency (MAF) in the general population, and predictions of the variant's effect on gene function. MAF was annotated in NHLBI GO Exome Sequencing Project ESP6500si-v2 (European American and African American population), 1000 Genomes Project (AFR [African], AMR [Admixed American], EAS [East Asian], EUR [European], SAS [South Asian], Exome Aggregation Consortium (ExAC) (EUR, non-Finish European population [NFE], AFR, SAS, EAS and AMR). SIFT, PolyPhen2 and Combined Annotation Dependent Depletion (CADD) were used to assess the deleterious effects of the identified variants.

To discover novel PD causative variants we analyzed WES data from 53 PD patients (mean age at onset 61.71, SD 8.57) belonging to 26 late onset PD families with supposedly dominant transmission (Fig. 1), by using a multi-step filtering approach as reported in our previous publications [21, 22]. Prioritization of candidate genes was made through segregation analysis. STRING database analysis was used to identify network of PD-interacting genes products [23].

Multiple ligation dependent probe amplification (MLPA)

The commercially available kit P051-P052 (MRC-Holland, Amsterdam, Netherlands) was used for the multiplex dosage of exons for the following genes: DJ1 (4 probes in P051), SNCA (5 probes in P051, 1 probe in P052), PARKIN (12 probes in P051, 12 in P052), LRRK2 (8 probes in P052), PINK1 (8 probes in P051). The

MLPA was performed on DNA from the 26 families who underwent WES analyses.

Polygenic variant load in unrelated Italian individuals

We identified 26 PD candidate genes in this study (*AIMP2*, *CHMP1A*, *GIPC1*, *HMOX2*, *HSPA8*, *IMMT*, *KIF21B*, *KIF24*, *MAN2C1*, *RHOT2*, *SLC25A39*, *SPTBN1*, *TMEM175*, *TOMM22*, *TVP23A*, *ZSCAN21*, *PARK2*, *PINK1*, *DJ-1*, *LRRK2*, *HTRA2*, *FBXO7*, *EIF4G1*, *DNAJC6*, *DNAJC13*, *SNCAIP*). To assess the co-existence of multiple rare variants in these 26 genes we counted the number of the rare exonic variants in each individual. The protocol used a $MAF \leq 0.001$ which was chosen to avoid the loss of heterozygous pathogenic mutations in recessive genes such as *PARK2*, *PINK1* and *DJI*. The pipeline included:

- Annotation and selection of all the exonic variants;
- Exclusion of synonymous variants;
- Exclusion of variants with alternative allele frequency > 0.001 ;
- Count of the variants detected per subject in the set of 26 candidate genes.
- Statistical analysis

The co-inheritance of rare variants ($MAF \leq 0.001$) was investigated by counting the number of variants laying in the 26 genes, for each sample analyzed, both cases and controls. Fisher's Exact tests were performed to test differences in the number of cases and controls carrying 0, 1 or ≥ 2 variants. Odds Ratios (ORs) and their corresponding 95% Confidence Intervals (95% CIs) were also estimated in R [24]. Bonferroni correction for two variant load contrasts was applied, resulting in a corrected $\alpha = 0.025$.

Then an analysis of Receiver Operating Characteristic (ROC) curve was performed to establish the accuracy of the test.

Expression studies

The expression profile of the 16 novel PD candidate genes *AIMP2*, *CHMP1A*, *GIPC1*, *HMOX2*, *HSPA8*, *IMMT*, *KIF21B*, *KIF24*, *MAN2C1*, *RHOT2*, *SLC25A39*, *SPTBN1*, *TMEM175*, *TOMM22*, *TVP23A*, *ZSCAN21* was evaluated on total RNAs from the mesencephalon of adult mice (post-natal day (P) 45). RNA was isolated by using TRIreagent® (Sigma-Aldrich, Saint Louis, MO, USA) protocol. 2 μ g of total RNA were reverse transcribed with the Superscript III-First strand kit (Thermo Fisher Scientific, Waltham, MA, USA). Quantitative PCR (qPCR) reactions were performed in triplicate, using gene specific primers (Table S2) and ITaq Universal Sybr Green Supermix (Bio-Rad, Hercules, CA, USA) following the manufacturer's directions. Results were normalized versus the

expression of the glyceraldehydes-3-phosphate dehydrogenase (*Gapdh*) gene. *Pitx3* was used as an example of mdDA neurons restricted gene expression marker. Standard deviation was calculated by using data of three different experiments.

Immunohistochemistry assays

Immunohistochemistry was performed on paraformaldehyde-fixed, wax-included brains as described previously [25]. Mouse and rat adult brains were obtained from healthy animals sacrificed in accordance with the recommendations of the European Commission. All the procedures related to animal treatments were approved by Ethic-Scientific Committee for Animal Experiments and Italian Health Ministry. Formalin-fixed and paraffin-embedded (FFPE) Substantia Nigra tissue sections from human adult normal brain were purchased at BioChain (Newark, CA, USA). Rabbit antibodies were directed against SLC25A39 (Abcam, Cambridge, UK, Ab-105,683; dilution 1:80) and Tyrosine Hydroxylase (TH) (Chemicon International, Temecula, CA, USA, ab152; dilution 1:600). Mouse antibodies were directed against GIPC1 (Santa Cruz Biotechnology, Dallas, TX, USA, Sc-271,822; dilution 1:80), TOMM22 (Abcam, Cambridge, UK, Ab-57,523; dilution 1:80), ZSCAN21 (Novus Biologicals, Centennial, CO, USA, NBP2-45443; dilution 1:80), HSPA8 (Santa Cruz Biotechnology, Dallas, TX, USA, Sc-59,570; 1:100), TH (Chemicon International, Temecula, CA, USA, MAB318; dilution 1:500).

URLs

STRING (<https://string-db.org/>); 1000 Genomes Project phase 3 release (<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>); Human Gene Mutation Database (HGMD, <http://www.hgmd.cf.ac.uk/ac/index.php>; Jun-2017 version); Parkinson Disease Mutation Database (PDMutDB, <https://www.molgen.vib-ua.be/PDMutDB/>; Jun-2017 version); NHLBI GO Exome Sequencing Project ESP6500si-v2 (<http://evs.gs.washington.edu/EVS/>); 1000 Genomes Project (<http://www.1000genomes.org/>); Exome Aggregation Consortium (ExAC) (<http://exac.broadinstitute.org/>); SIFT (http://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html); PolyPhen2 (<http://genetics.bwh.harvard.edu/pph2/>); Combined Annotation Dependent Depletion (CADD) (cadd.gs.washington.edu/); R (<http://www.R-project.org/>).

Results

Identification of novel candidate genes for Parkinson's disease

The study aimed at the identification of genetic risk profiles for PD. The discovery cohort included 53 PD patients belonging to 26 families showing a dominant mode of inheritance of the disease (Fig. 1). As validation cohort, we

used 1542 independent PD cases and 706 controls from different data sets of European ancestry (see [Patients and Methods](#) for details) (Table S1) [8, 9, 16, 26].

We first analysed the WES data (without applying any filter) of the 53 PD patients belonging to 26 families to search for mutations in genes already known to be associated to dominant form of PD such as, *SNCA*, *VPS35* and *LRRK2*. Subsequently, we searched for large insertion/deletions in *SNCA* and *LRRK2* genes by using MLPA approach. All annotated variants were surveyed into the Parkinson's disease Mutation Database (<https://grenada.lumc.nl/LOVD2/TPI/home.php>) to confirm their pathogenicity. One out of the 26 analyzed families carried a pathogenic mutation in *LRRK2* gene (c.G4322A, p.R1441H).

To search for novel candidate-disease genes we used a multistep filtering approach. We selected splicing variants (regardless of their exonic or intronic position; minor allele frequency (MAF) ≤ 0.001) and exonic disruptive variants (MAF ≤ 0.001 ; CADD phred score ≥ 20 ; excluding synonymous changes), shared between affected relatives within each of the 26 families (siblings or parent-offspring pairs), assuming a dominant mode of transmission of the disease. The variants were selected assuming a MAF ≤ 0.001 in public exome databases including ESP6500si-v2, 1000 Genomes Project and ExAC. The cut off was set to 0.001, taking in account the prevalence of rare dominant diseases (1–2/10,000) and the MAF of the most frequent mutation associated with late onset PD p.(G2019S) in the *LRRK2* gene, also known as rs34637584, which shows a MAF ranging from 0.001 to 0.0002 in different databases. Retained variants (6488) were filtered for CADD phred score ≥ 20 (3712 remaining variants), corresponding to those variants which are predicted to be amongst the 1% most deleterious variants in the genome. Subsequently we compared data of the affected relatives of each family to select shared variants. This analysis identified an average number of 10 variants shared between PD relatives per family. Segregation analysis was performed in additional affected relatives belonging to the same 26 PD families (1–3 individuals per family based on DNA availability) (Fig. 1). Considering the late age at onset of the disease unaffected family members were not investigated. This analysis disclosed 28 rare disruptive variants (23 non-synonymous, 2 stop-gain, 1 frameshift, 2 non-frameshift deletions) laying in 26 genes, which were shared among familial PD cases in 18 out of the 26 analyzed families (Table 1). Among these, in 10 families we found single heterozygous deleterious variants in a single gene segregating with PD phenotype, supporting a dominant model of inheritance. Instead, we identified 2 variants in 6 families and 3 variants in 2 families in different genes segregating with PD phenotype suggesting a polygenic

model of inheritance. Unfortunately, in the remaining eight families we were not able to identify potential causative variants, probably due to the highly stringent criteria used in the pipeline of analysis. Seven (*PARK2*, *PINK1*, *DJ-1(PARK7)*, *LRRK2*, *FBXO7*, *DNAJC6*, *DNAJC13*) out of the 26 identified genes were already known as causing PD, including familial recessive forms, while three (*HTRA2*, *EIF4G1*, *SNCAIP*) genes were PD candidate genes/at risk factors for which a clear link with the disease is not completely established [5, 6]. Moreover, it is interesting to note that in several families, single mutations in recessive genes such as *PARK2*, *PINK1*, *PARK7*, were co-inherited with variants in other candidate genes, suggesting that they might play a role as risk factors in the heterozygous status. The presence of a second mutation in these recessive genes including large deletions/insertions was excluded through MLPA approach. Sixteen out of the 26 genes analyzed were novel PD candidate genes involved in pathways potentially deregulated in PD such as mitochondrial metabolism and oxidative stress (*AIMP2*, *HMOX2*, *IMMT*, *MAN2C1*, *RHOT2*, *SLC25A39*, *TOMM22*) [27–35], vesicular trafficking, microtubule dynamics, autophagy (*CHMP1A*, *GIPCI*, *HSPA8*, *KIF21B*, *KIF24*, *SPTBN1*, *TMEM175*, *TVP23A*) [36–42] or in *SNCA* gene expression (*ZSCAN21*) (Fig. 2a) [43, 44]. STRING database analysis showed that nine out of the 16 novel genes (*AIMP2*, *GIPCI*, *HSPA8*, *IMMT*, *RHOT2*, *SPTBN1*, *TMEM175*, *TOMM22*, *ZSCAN21*) encoded for proteins interacting with known PD genes (Fig. 2a and b) [39, 45].

Only very recently, *IMMT*, *ZSCAN21* and *TMEM175* have been associated to PD [43, 46, 47] further supporting the discovery pipeline employed in this study.

To further support the genetic involvement of the 16 novel identified genes in PD pathogenesis, we extended the mutational analysis to 1542 independent PD patients (including MNI (394 cases) and IPDGC (1148 cases)) and 706 controls from different cohorts of European ancestry, to detect further disruptive variants within these genes and to compare the total load of these variants between cases and controls. Overall data identified 256 different variants (MAF ≤ 0.001 ; CADD phred score ≥ 20), of which 170 were present only in cases, 61 only in controls and 25 were shared between cases and controls. Remarkably, the 170 different deleterious variants (153 non-synonymous, 5 stop-gain, 6 frameshifts, 3 non-frameshift deletions and 3 splicing) were identified in 90 Italian patients (MNI cohort) and 153 patients of the IPDGC cohort, confirming the extreme genetic heterogeneity at the basis of PD (Table S3). None of these variants was found in 706 healthy control subjects, including both IPDGC and Italian population controls (namely MNI healthy subjects and Tuscans

Table 1 Rare disruptive variants (MAF \leq 0.001, CADD phred \geq 20) in the 26 PD candidate genes identified in the discovery families

CHR	Genomic position (hg19)	dbSNP	Gene	RefSeq	Nucleotide Change	AA Change	Exonic Function	CI	MAF Max in public data sets	CADD phred	Family
16	723,544	rs143816083	RHOT2	NM_138769	c.G1795T	p.G599W	NSV	N	0.0001	25	1
3	184,039,223	rs746291399	EIF4G1	NM_198241	c.C851G	p.S284C	NSV	N	0.00004	25	2
1	8,022,928	rs142405016	PARK7	NM_007262	c.G83A	p.R28Q	NSV		0.0002	35	5
7	6,062,997	rs749728733	AIMP2	NM_006303	c.T638C	p.I213T	NSV	N	0.00002	27	7
2	86,408,450	rs201861204	IMMT	NM_001100169	c.C91T	p.R31C	NSV	N	0.0002	26	7
1	20,971,042	rs74315358	PINK1	NM_032409	c.G836A	p.R279H	NSV	P	0.00004	26	7
16	89,717,990		CHMP1A	NM_002768	c.C92T	p.A31V	NSV	N	NA	23	9
2	74,757,823		HTRA2	NM_013247	c.A586G	p.N196D	NSV	N	NA	23	9
3	132,222,165		DNAJC13	NM_015268	c.C4824G	p.S1608R	NSV	N	NA	20	10
17	42,398,033		SLC25A39	NM_001143780	c.C758A	p.T253N	NSV	N	NA	21	10
2	54,895,594	rs200093475	SPTBN1	NM_003128	c.C6983A	p.T2328N	NSV	N	0.00004	27	12
22	39,078,021	rs200476832	TOMM22	NM_020243	c.C38T	p.P13L	NSV	N	0.0002	25	12
19	14,589,354		GIPC1	NM_202470	c.875dupA	p.K292fs	Fs_ins	N	NA	NA	13
6	162,394,354	rs114974496	PARK2	NM_004562	c.C714G	p.C238W	NSV	P	0.00004	25	13
5	121,787,252	rs751412863	SNCAIP	NM_005460	c.A2710G	p.K904E	NSV	N	0.00004	24	14
12	40,704,237	rs34995376	LRRK2	NM_198578	c.G4322A	p.R1441H	NSV	P	0.00001	27	14
7	99,654,828		ZSCAN21	NM_145914	c.C199T	p.R67W	NSV	N	NA	24	16
9	34,290,328	rs372797268	KIF24	NM_194313	c.T971C	p.I324T	NSV	N	0.0001	26	17
16	10,911,991		TVP23A	NM_001079512	c.G58T	p.E20X	stopgain	N	NA	38	18
1	65,858,474		DNAJC6	NM_001256864	c.C1829T	p.P610L	NSV	N	NA	23	19
15	75,652,021		MAN2C1	NM_001256494	c.1882_1887del	p.628_629del	NFs_del	N	NA	NA	19
15	75,652,016		MAN2C1	NM_001256494	c.1890_1892del	p.630_631del	NFs_del	N	NA	NA	19
22	32,870,999		FBXO7	NM_012179	c.C10T	p.R4W	NSV	N	NA	33	20
16	4,557,822	rs772245870	HMOX2	NM_001286271	c.C226T	p.R76W	NSV	N	0.00001	34	20
1	200,957,957	rs751952433	KIF21B	NM_001252100	c.G3235A	p.A1079T	NSV	N	0.00002	29	21
11	122,931,465		HSPA8	NM_006597	c.G247A	p.V83I	NSV	N	NA	21	24
4	947,062		TMEM175	NM_032326	c.C547T	p.R183X	stopgain	N	NA	34	25
1	200,974,537		KIF21B	NM_001252100	c.C631T	p.R211C	NSV	N	NA	34	26

CHR Chromosome, hg19 human genome build to which these variants are annotated, dbSNP reference number in SNP database, ref. seq reference number of the gene transcript, AA Change Amino acid change, CI Clinical interpretation, P Pathogenic, UP Uncertain pathogenicity, N Novel, PD Parkinson's disease, IPDGC International Parkinson's Disease Genetics Consortium, CADD phred Combined Annotation Dependent Depletion, Fs_ins Frame shift insertion, NFs_del Non Frame shift deletion, NSV Non-synonymous variant, MAF Minor Allele Frequency, NA Not Annotated, MAF max in public datasets: highest allelic frequency annotated in public databases including 1000 Genomes Project (AFR. AMR. EAS. EUR. SAS), ExAC browser (NFE. AFR. SAS. EAS and AMR), ESP6500si-v2 (European American and African American population)

(TSI) “pseudo-controls “of the 1000 Genomes Project) [14]. Interestingly, significant enrichment of variants in these 16 genes was observed in patients compared to controls (243 patients (15.7%) vs 69 controls (9.7%); OR = 1.73 [1.3–2.29]; $p = 0.0001$ $\chi^2 = 14.01$).

Noteworthy, all the affected residues identified in PD patients occupy functionally important amino acid positions, which are highly conserved amongst vertebrates (Fig. 3; Fig. S1). In summary, although functional data are necessary to substantiate the causative role of the variants identified, these results suggest a potential involvement of the candidate genes in PD etiopathogenesis and indicate an extreme genetic heterogeneity of mutations in PD patients.

Expression analysis in human, mouse and rat DA neurons

Expression analysis through quantitative PCR (qPCR) assays showed that the 16 novel PD genes were all transcribed in the mesencephalon of adult mice at post-natal day (P) 45 (Fig. 4). Then we studied through immunohistochemistry experiments whether five representative genes (TOMM22, GIPC1, ZSCAN21, SLC25A39 and HSPA8) were co-expressed with TH in adult DA neurons of the substantia nigra (SN) and ventral tegmental area (VTA). We first analyzed adult human SN neurons and we found that TH⁺ neurons co-expressed all the five genes (Fig. 5a and Fig. S2a). This analysis was repeated in P45 mouse (Fig. 5b-g and Fig. S2b-d) and P60 rat (Fig. S3) brains. In mouse mdDA

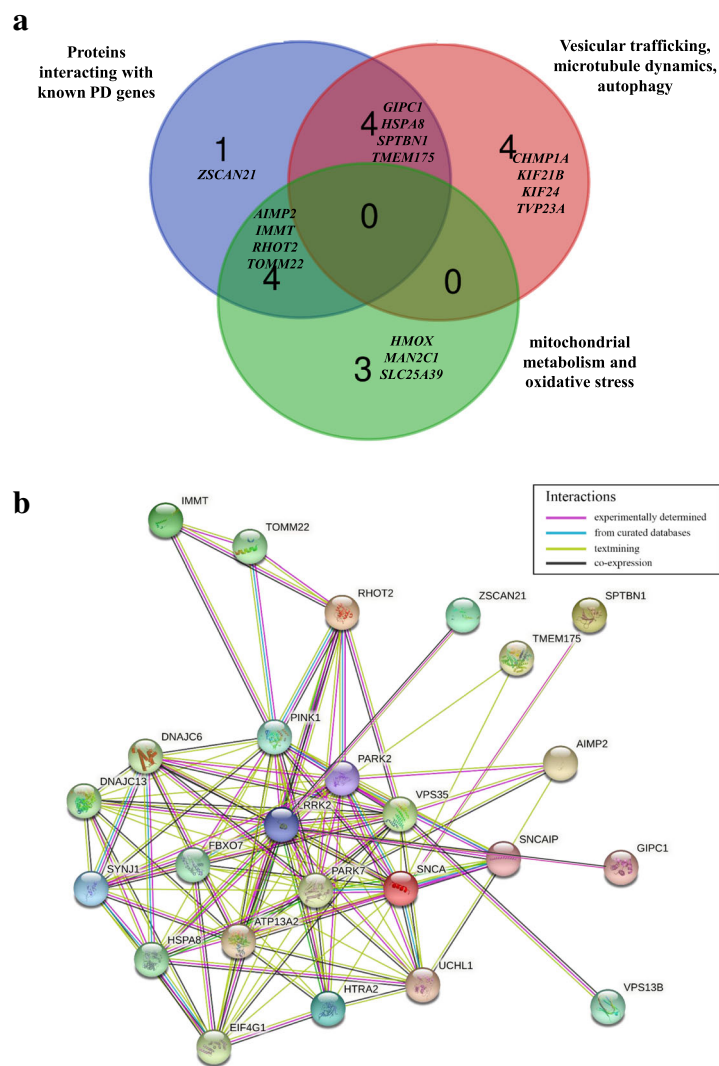


Fig. 2 Venn diagram and STRING analysis representing the logical relations between the 16 novel gene products and the network of PD-interacting genes products. **a** Venn diagram showing the 16 novel identified genes involved across different interconnected pathways: interactors of PD related genes, genes involved in mitochondrial metabolism and those involved in vesicular trafficking and autophagy. **b** Schematic representation of the network of interactions of the 16 novel PD gene products with known PD-associated genes. STRING database analysis identified 9 genes interacting with PD-associated genes

neurons belonging to the anterior SN located in the pre-tectum (Fig. 5b, c and Fig. S2b) or to the SN and VTA of mesencephalic origin (Fig. 5d-g and Fig. S2c, d) showed that the expression of TOMM22, GIPCI, ZSCAN21, SLC25A39 and HSPA8 colocalized with most of the TH⁺ neurons (Fig. 5b-g and Fig. S2b-d). A similar result was observed when this expression analysis was performed in rat SN and VTA neurons of mesencephalic origin (Fig. S3). These data indicate that at least five out of the 16 novel PD genes were expressed in mdDA neurons supporting the possibility that they may be cell-autonomously involved to provide mdDA neurons with functional features potentially controlling identity and/or

survival and/or DA neurotransmission. Therefore, together with previous data, this expression analysis further supports the possibility that mutations identified in the 16 novel PD candidate genes may potentially predispose for PD occurrence.

Polygenic load analysis of rare exonic variants in familial and sporadic cases

Based on the observation that 8 out of the 18 families of the discovery cohort carried two or more variants in the 26 PD genes (Table 1), we hypothesized that the genetic combination of multiple rare variants in these genes may represent a specific genetic hallmark of PD risk. To test

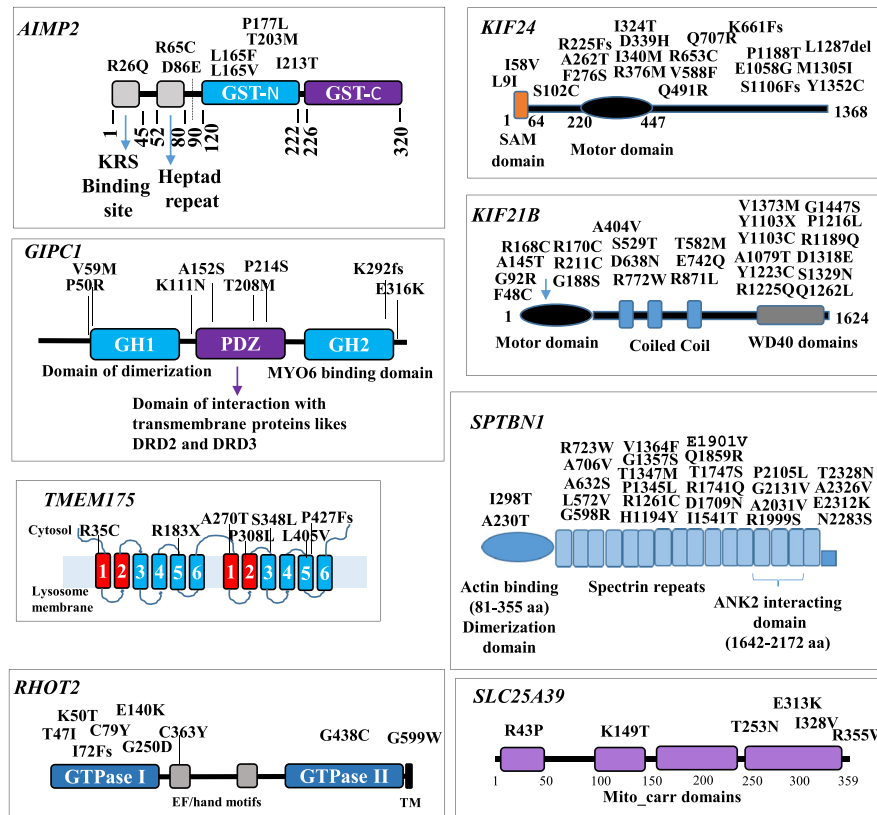


Fig. 3 Graphical representation of the protein domains and mapped mutations. For those genes whose protein structure was available (*AIMP2*, *GIPC1*, *KIF24*, *KIF21B*, *RHOT2*, *SLC25A39*, *SPTBN1*, *TMEM175*), we report the protein domains and mapped amino acid changes identified in the cohort of 1542 patients. Amino acid changes are referred to by their single letter code

this hypothesis, we analyzed in 394 unrelated and independent PD patients (including 172 familial and 222 sporadic cases from MNI data set) and in 203 controls (MNI, Moli-sani, and TSI data sets) [8, 9, 14, 15] the co-inheritance (per individual) of rare (MAF < 0.001) exonic (excluding synonymous) and splicing variants detected in the panel of the 26 genes (details in [Patients and Methods](#)). For this analysis, we selected rare variants based on a frequency threshold (MAF ≤ 0.001), without considering their degree of deleteriousness. This analysis identified 220 different variants among cases and 65 in the controls; only 18 of them were shared between case and controls (Table S4). The co-inheritance of these variants is reported in Table 2 and Fig. 6a. We observed that, approximately 17% of the PD patients carried two or more variants (cases 17.3% vs controls 6.8%; OR = 3.3 [1.8–6.7]; $p = 4.4 \times 10^{-5}$) (Table 2). Remarkably, sporadic cases showed a significant distribution within the same class (sporadic cases 13.9% vs controls 6.8%, OR = 2.6 [1.3–5.1]; $p = 0.005$) (Table 2). These differences remained statistically significant after Bonferroni correction for multiple testing of two contrasts (0 vs 1 and 0 vs ≥2 variants load classes) ($\alpha = 0.025$). This finding

corroborates the polygenic view of PD risk for both sporadic and familial cases. To measure the accuracy of the test, we applied the ROC curve and found a maximum sensitivity of 0.50 (based on the presence of at least one variant (specificity 0.66)) and a maximum specificity > 0.93 (based on the co-occurrence of two or more variants (sensitivity of 0.17)) (AUC = 0.59, SE = ± 0.020, $p = 4.42 \times 10^{-6}$) (Fig. 6b). Overall, the presence of two or more disruptive variants in the 26 candidate genes represented the best tradeoff between sensitivity (0.17) and specificity (0.93). Moreover, considering that is well established that rare variants in *LRRK2*, *PINK1*, *PARK2* and *PARK7* genes have a higher prevalence in PD patients compared to healthy individuals [48–52], we analyzed the presence of at least one rare variant (MAF < 0.001) in patients and controls taking into account solely the 16 novel candidate genes. This analysis confirmed an excess of patients carrying rare variants in the 16 novel genes (cases 31.6% vs controls 23.1%; $p = 0.014$), as also reported in the gene identification paragraph for rare deleterious variants (MAF ≤ 0.001; CADD phred score ≥ 20) selected in the 16 novel genes in cases and controls (243 (15.7%) vs 69 (9.7%); OR = 1.73 [1.3–

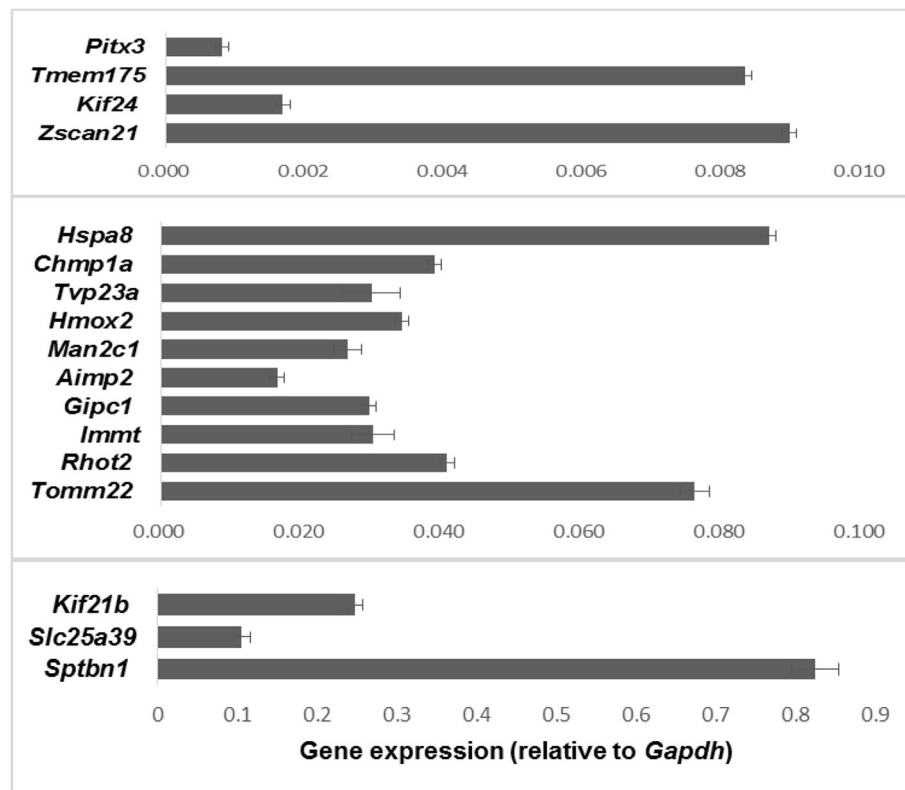


Fig. 4 Expression analysis of the novel identified Parkinson's disease genes. qPCR experiment showed that the 16 novel genes were expressed in mesencephalon of P45 adult mice. *Pitx3* was used as an example of mdDA neurons restricted gene expression. *Gapdh* was used to normalize the results. Data were shown as absolute values with standard deviation calculated from three different experiments. Genes were grouped on the basis of their expression level in three different scales

2.29]; $p = 0.0001$ $\chi^2 = 14.01$), further supporting their role into PD etiopathogenesis.

Moreover, the pool of rare variants identified in the PD cases was different from those detected in controls (Table S4). This suggests that, besides the load of rare variants (≥ 2), also the nature of the mutation is important in predisposing/causing PD.

Thus, the availability of a detailed database including variants detected only in PD cases or only in controls might be relevant to improve the operational characteristics of our predictive protocol in the general population.

Analysis of *GBA* gene

To test the involvement of *GBA* gene, that is the most important genetic risk factor for PD, alone and in combination with the 26 PD genes, we analyzed all *GBA* exonic variants annotated in the 26 late onset PD families and in the Italian validation cohort (394 independent PD cases and 203 controls) (Table S5). In the discovery cohort the *GBA* variant c.G1093A p.E365K (E326K, previous nomenclature) was present in the two affected siblings of family 7, co-inherited with single heterozygous

variants in *AIMP2*, *IMMT* and *PINK1* genes (Table 1), while the *GBA* variant c.C1223T p.T408M (T369M, previous nomenclature) was present in one of the two siblings of family 16 who also carried the c.G4322A p.R1441H mutation in *LRRK2* gene and the deleterious variant c.A2710G p.K904E in *SNCAIP*. In the independent cohort of PD cases and controls we found a significant distribution of *GBA* variants (42 cases (10.6%) vs 8 controls (3.9%); $p = 0.002$, OR = 2.91 [1.34–6.32]).

Polygenic load analysis including multiple rare variants in the 26 genes as well as rare pathogenic variants in *GBA* gene showed that, approximately 20% of the PD patients carried two or more variants (cases 20.5% vs controls 7.2%; OR = 3.59 [1.97–6.90]; $p = 3.4 \times 10^{-6}$) (Table 2). Remarkably, the increase in significance was mainly attributable to sporadic cases (sporadic cases increase from 13.9 to 18.5% with *GBA* mutations) (Table 2).

Analysis of PD endophenotypes

To investigate the relation of the polygenic variant load in the selected 26 candidate genes with PD endophenotypes, including both motor and non-motor symptoms, we analyzed the cohort of 394 PD-MNI unrelated

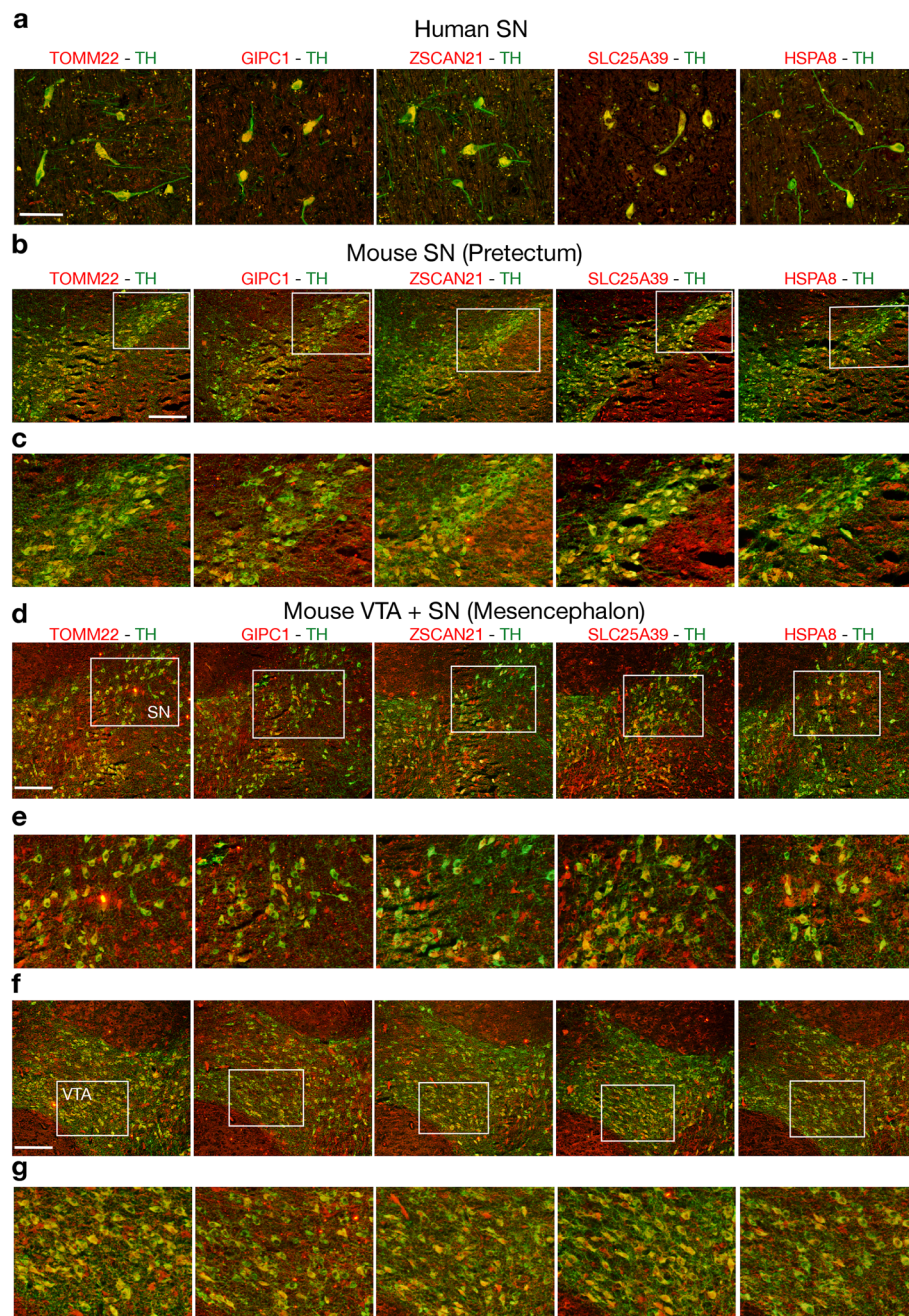


Fig. 5 Expression analysis of TOMM22, GIPC1, ZSCAN21, SLC25A39 and HSPA8 in human and mouse DA neurons. **a-g** Immunohistochemistry experiments performed on human (**a**) and P45 mouse (**b-g**) adult brain sections showed that TOMM22, GIPC1, ZSCAN21, SLC25A39 and HSPA8 were co-expressed with TH in DA neurons of the human SN (**a**) and in mouse DA neurons of the SN (**b-e**) and VTA (**f, g**). Images in (**c, e, g**) correspond to magnification of the boxed area reported in (**b, d, f**). Scale bars correspond to 100 μ m. Abbreviations: SN stands for substantia nigra and VTA for ventral tegmental area

patients mentioned before, for which clinical assessments were available (see [Patients and Methods](#) for details on the scales used). More specifically, we tested how the presence of 0, 1 \geq 2 rare variants in the 26 PD genes affected PD age at onset (AAO), motor symptoms (UPDRS scale) [10, 13], non-motor symptoms (NMS

scales) [12] and presence of L-dopa induced Dyskinesia (LID), a common motor side effect of levodopa therapy in PD patients [8, 9]. Overall data show that the selected genes might influence preferentially LID occurrence, although the contrast would not survive correction for multiple testing of five phenotypes (p 0.038; Fig. 6c;

Table 2 Polygenic variant load in the Italian cohort

Series	N variants	PD cases (n = 394)	FPD (n = 172)	SPD (n = 222)	Cnt. (n = 203)	p	OR	95%CI
MAF ≤ 0.001	0	197 (50.0%)	82 (47.6%)	115 (51.8%)	135 (66.5%)	baseline		
	1	129 (32.7%)	53 (30.8%)	76 (34.2%)	54 (26.6%)	0.014	1.6	1.1–2.4
	≥2	68 (17.3%)	37 (21.6%)	31 (13.9%)	14 (6.8%)	4.4 × 10 ⁻⁵	3.3	1.8–6.7
MAF ≤ 0.001 including variants in GBA gene	0	182 (46.2%)	79 (45.9%)	103 (46.4%)	128 (63.0%)	baseline		
	1	130 (33.2%)	53 (30.8%)	77 (35.1%)	59 (29.0%)	0.028	1.55	1.04–2.32
	≥2	82 (20.5%)	40 (23.2%)	42 (18.5%)	16 (7.2%)	3.4 × 10 ⁻⁶	3.59	1.97–6.90

MAF Minor allele frequency, PD Parkinson's disease, FPD Familial PD cases, SPD Sporadic PD cases, Cnt Controls from TSI, Tuscan Italian population, MNI cohort and Moli-sani genetic biobank, p p-value calculated with Fisher Exact Probability Test, OR Odds Ratio, CI Confidence interval

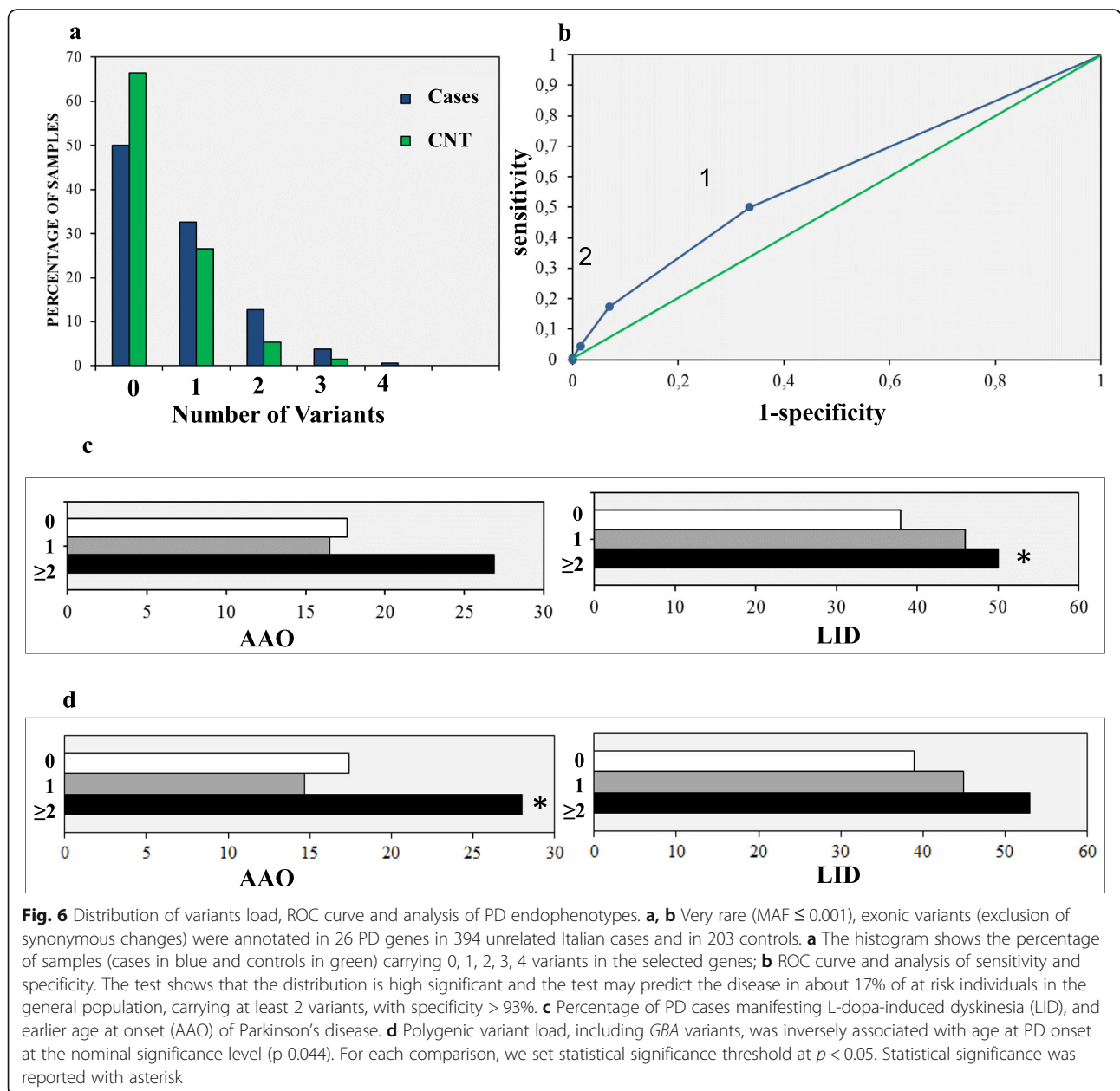


Table S6A). When we took into account also GBA variants, this contrast was not significant anymore, while variant load was inversely associated with age at PD onset at the nominal significance level (p 0.044; Table S6B; Fig. 6d).

Discussion

In this study, we discovered a set of 16 genes worthy of future studies to confirm their role in onset or progression in PD. We identified 170 rare and deleterious variants, absent in healthy individuals, in 15.7% of PD patients (243 PD cases out of 1542) supporting the extreme genetic heterogeneity of mutations in PD.

Although additional studies are needed to confirm the functional role of the novel identified genes in PD etio-pathogenesis, a number of published studies support this hypothesis.

In particular, the transcription factor ZSCAN21 may directly control *SNCA* gene expression [28, 37], and the c.C199T p.(R67W) mutation identified in this study affects a highly conserved arginine residue necessary for the formation of the ZSCAN21 functional homodimer [53]. *TMEM175* is a lysosomal K^+ channel that is required to maintain membrane potential and pH stability in lysosomes as well as to control α -synuclein aggregation [42, 54]. These functions may be affected by knock down of the gene *TMEM175* [42]. Although *TMEM175* was previously identified in a genome wide association study of PD [47], detrimental mutations have not been identified before this study.

RHOT2 (MIRO2), IMMT and TOMM22 directly interact with PINK-1 and Parkin and influence both mitochondrial homeostasis and vulnerability of DA cells to toxins [27–31, 55]. The aminoacyl-tRNA synthetase complex interacting multifunctional protein-2 (AIMP2) is a parkin substrate and its overexpression leads to a selective, age-dependent, progressive loss of DA neurons via activation of poly (ADP-ribose) polymerase-1 (PARP1) [35]. *GIPC1* interacts with LRRK2 in DA cells in vitro and in the mouse ventral mesencephalon in vivo [45]. The *Drosophila* homolog of *GIPC1* is expressed in the central brain of adult flies and its reduced expression significantly affects locomotive activity and longevity [56]. *SLC25A39* is a mitochondrial carrier of manganese that promotes neuronal survival in *Drosophila* [34]. β II-spectrin, encoded by *SPTBN1*, is required for axon growth and axonal transport of synaptic cargo [57]. Moreover, it specifically interacts with α -synuclein and its overexpression induces neurite formation [39]. The chaperone-mediated autophagy (CMA) protein HSPA8/HSC70 (heat shock protein 8) is essential in post-mitotic neurons for diluting toxic intracellular components including the disassembling of α -synuclein amyloid fibrils [58]. *KIF21B* is a processive kinesin whose depletion

alters neuronal dendritic tree branching and spine formation [37]. Noteworthy, similarly to *Kif21b* knockout mice, which exhibit learning and memory deficits [37], our patients carrying *KIF21B* mutations (Table S3) showed mild cognitive impairment assessed with Montreal Cognitive Assessment (MoCA) test (mean 23 SD 2). *KIF24* is a centriolar kinesin that specifically regulates ciliogenesis by dynamically controlling microtubule polymerization [38]. Single Nucleotide Polymorphisms (SNPs) located in the *KIF24* gene have been associated to Frontotemporal Lobar Degeneration (FTLD) [59]. Interestingly, the patients of the MNI cohort carrying mutations affecting the motor domain of *KIF24* protein (Fig. 3) showed mild-severe cognitive impairment (MoCA score ranging from 13 to 23).

HMOX exerts both pro- and anti-oxidant effects, which may influence the pathogenesis or progression of Parkinson's disease [32, 60]. α -Mannosidase (*MAN2C1*) is the enzyme responsible for the partial demannosylation occurring in the cytosol. *Man2c1*-deficient mice revealed neuronal and glial degeneration [61]. Charged multivesicular body protein 1A (*CHMP1A*) is a member of the ESCRT-III (endosomal sorting complex required for transport-III) complex which localizes to the nuclear matrix and regulates chromatin structure [36, 62]. *CHMP1A* regulates proliferation of neuronal progenitor cells [36, 62]. *TVP23A* is a novel gene reported in gene database (Gene ID: 780776) as involved in intracellular vesicular transport.

Previous studies identified the p.G2019S mutation in *LRRK2* as a major causative event in 1–4% of PD patients, depending on the analyzed population [63]. Within our cohort of patients, the p.G2019S mutation in *LRRK2* was detected in 8 out of 394 PD patients (2%). Moreover, other studies also correlated the severity but not the occurrence of the PD to the cumulative effect of *LRRK2* mutations and rare variants in other PD related genes (e.g. GBA) [26, 64–66].

In this study, we highlight the polygenic nature of both sporadic and familial forms of PD, especially when the familial forms come from small nuclear families. We identified a panel of 26 candidate genes carrying likely pathogenic variants and observed the presence of multiple rare variants (≥ 2) in these 26 genes in about 17% of PD patients, including familial and sporadic cases. This percentage increased to 20% of PD patients when we included rare pathogenic variants in *GBA*. Remarkably, this increase was mainly attributable to sporadic cases.

This means that the molecular diagnosis of PD should always take in account the analysis of multiple candidate genes. However, although the evidence reported here looks very promising, at this stage of advancement of our study we cannot offer any improvement in

practicing medicine, until further confirmatory studies and functional validations are warranted. Undoubtedly, the extreme genetic heterogeneity observed in PD patients, everyone genetically and biologically unique, carrying different genetic combinations of rare variants not shared by the majority of PD patients, could cause great uncertainty in the interpretation of the results. Nevertheless, at a molecular level, these findings offer a rational support to address additional genetic studies testing this protocol in larger cohort of PD patients and controls. If these data were confirmed we could consider the possibility of using this protocol to identify, with high specificity, subjects potentially at risk of developing PD by counting per individual the number of rare variants in the panel of candidate genes.

Moreover, our data also underline the importance of rare variants in the genetics of sporadic PD and corroborate the proposed approach as complementary to genome wide association studies (GWAS) and to classical co-segregation analyses of NGS data to identify variants implicated in PD.

The present study also reports, for the first time, statistical evidence that the polygenic variant load in the PD candidate genes analyzed is associated with the occurrence of L-dopa induced Dyskinesia, an important motor side effect of PD treatment. Although it is tempting to speculate about the translational relevance of this finding, caution is suggested in the interpretation since the association was only nominally significant and disappeared when GBA variants were included in the analysis. Interestingly, in the latter setting a high variant burden was associated with earlier PD onset, in line with previous evidence [67, 68].

Conclusions

Our findings encourage further studies to improve the predictive/diagnostic power of this protocol for PD and lead us to surmise that this strategy may hypothetically be extended to other polygenic disorders with largely unknown genetic etiology, such as dyslexia [69].

Therefore, in the future studies on PD genetic aetiology should be directed towards analyzing the polygenic burden of rare variants with presumably large effect size, and validating them at the functional level.

Abbreviations

PD: Parkinson's disease; LOPD: Late onset Parkinson's disease; GWAS: Genome wide association studies; mdDA: Mesencephalic-diencephalic DA; SNpc: Substantia nigra pars compacta; PIB: Parkinson Institute Biobank; WES: Whole exome sequencing; MNI: Mediterranean Neurological Institute; NGS: Next Generation Sequencing; TR: Targeted Re-sequencing; SNVs: Single nucleotide variants; MAF: Minor Allele Frequency; CADD: Combined Annotation Dependent Depletion; TSI: Tuscan Italians; qPCR: Quantitative PCR; ROC: Receiver Operating Characteristic; FPD: Familial PD; SPD: Sporadic PD

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13024-021-00455-2>.

Additional file 1: Figure 1S. Evolutionary conservation study of the likely pathogenic variants identified in the 16 novel Parkinson's disease candidate genes. The affected residues are reported in bold underlined. **Figure 2S.** Single channel images of human and mouse immunostainings shown in Fig. 5. **(a-d)** Single channel images showing expression of TH in combination with TOMM22, GIPC1, ZSCAN21, SLC25A39 and HSPA8 in human SN neurons **(a)** and in pretecal SN **(b)** mesencephalic SN **(c)** and VTA **(d)** mouse neurons. Images in **(a)** correspond to those reported in Fig. 5a; images in **(b-d)** correspond to those reported in Fig. 5c, e, g, respectively. **Figure 3S.** Expression analysis of TOMM22, GIPC1, ZSCAN21, SLC25A39 and HSPA8 in rat DA neurons. **a-e** Immunohistochemistry experiments performed on P60 rat brain sections showed that TOMM22, GIPC1, ZSCAN21, SLC25A39 and HSPA8 were co-expressed with TH in mesencephalic DA neurons of the SN **(b, c)** and VTA **(d, e)**. Images in **(c, e)** correspond to magnification of the boxed area reported in **(b, d)**. Sections in **(a)** are low magnification images showing only TH immunostaining and corresponding to the same sections reported in **(b-e)**. Scale bars correspond to 100 μ m. Abbreviations: SN stands for substantia nigra and VTA for ventral tegmental area. In **(c, e)** are shown merge and single channel images. **Table S1.** Cohort description. **Table S2.** List of primers used for gene expression analysis by quantitative PCR (qPCR). **Table S3.** Rare disruptive variants (MAF \leq 0.001, CADD phred \geq 20) identified in the 16 novel PD candidate genes in the validation cohorts. **Table S4.** Rare variants (MAF \leq 0.001) in the 26 PD candidate genes in affected and healthy individuals of the Italian cohort. **Table S5.** GBA variants annotated in Italian cohort of patients and controls.

Acknowledgements

We thank the participating patients and their families. We are grateful to Dr. Stefano Goldwurm and the "Parkinson Institute Biobank," member of the Telethon Network of Genetic Biobank (biobanknetwork.telethon.it/) and the Moli-sani Biobank at Neuromed Biobanking Centre, member of BBMRI-Italy. **IPDGC consortium members and affiliations:** **United Kingdom:** Alastair J Noyce (Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, QMUL, London, UK and Department of Molecular Neuroscience, UCL, London, UK), Rauan Kaiyrzhanov (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Ben Middlehurst (Institute of Translational Medicine, University of Liverpool, Liverpool, UK), Demis A Kia (UCL Genetics Institute; and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Manuela Tan (Department of Clinical Neuroscience, University College London, London, UK), Henry Houlden (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Huw R Morris (Department of Clinical Neuroscience, University College London, London, UK), Helene Plun-Favreau (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Peter Holmans (Biostatistics & Bioinformatics Unit, Institute of Psychological Medicine and Clinical Neuroscience, MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff, UK), John Hardy (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Daniah Trabzuni (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK; Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, 11211 Saudi Arabia), John Quinn (Institute of Translational Medicine, University of Liverpool, Liverpool, UK), Vivien Bubb (Institute of Translational Medicine, University of Liverpool, Liverpool, UK), Kin Y Mok (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Kerri J. Kinghorn (Institute of Healthy Ageing, University College London, London, UK), Kimberley Billingsley (Institute of Translational Medicine, University of Liverpool, Liverpool, UK), Nicholas W Wood (UCL Genetics Institute; and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Patrick Lewis (University of Reading, Reading, UK), Sebastian Schreglmann (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Ruth Lovering (University College London, London, UK), Lea R'Bibo (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Claudia Manzoni (University of Reading, Reading, UK), Mie Rizig (Department of Molecular Neuroscience, UCL Institute of

Neurology, London, UK), Mina Ryten (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Sebastian Guelfi (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Valentina Escott-Price (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK), Viorica Chelban (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Thomas Foltynie (UCL Institute of Neurology, London, UK), Nigel Williams (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff, UK), Karen E. Morrison (Faculty of Medicine, University of Southampton, UK), Carl Clarke (University of Birmingham, Birmingham, UK and Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, UK).

France: Alexis Brice (Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS, UMR 7225, Sorbonne Universités, UPMC University Paris 06, UMR S 1127, AP-HP, Pitié-Salpêtrière Hospital, Paris, France), Fabrice Danjou (Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS, UMR 7225, Sorbonne Universités, UPMC University Paris 06, UMR S 1127, AP-HP, Pitié-Salpêtrière Hospital, Paris, France), Suzanne Lesage (Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS, UMR 7225, Sorbonne Universités, UPMC University Paris 06, UMR S 1127, AP-HP, Pitié-Salpêtrière Hospital, Paris, France), Jean-Christophe Corvol (Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS, UMR 7225, Sorbonne Universités, UPMC University Paris 06, UMR S 1127, Centre d'Investigation Clinique Pitié Neurosciences CIC-1422, AP-HP, Pitié-Salpêtrière Hospital, Paris, France), María Martínez (INSERM UMR 1220; and Paul Sabatier University, Toulouse, France), **Germany:** Claudia Schulte (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany), Kathrin Brockmann (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany), Javier Simón-Sánchez (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany), Peter Heutink (DZNE, German Center for Neurodegenerative Diseases and Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany), Patrizia Rizzu (DZNE, German Center for Neurodegenerative Diseases), Manu Sharma (Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tübingen, Germany), Thomas Gasser (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany),

United States of America: Mark R Cookson (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, USA), Sara Bandres-Ciga (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Cornelis Blauwendraat (National Institute on Aging and National Institute of Neurological Disorders and Stroke, USA), David W. Craig (Department of Translational Genomics, Keck School of Medicine, University of Southern California, Los Angeles, USA), Derek Narendra (Inherited Movement Disorders Unit, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA), Faraz Faghri (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, USA; Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA), J Raphael Gibbs (Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA), Dena G. Hernandez (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Kendall Van Keuren-Jensen (Neurogenetics Division, TGen, Phoenix, AZ USA), Joshua M. Shulman (Departments of Neurology, Neuroscience, and Molecular & Human Genetics, Baylor College of Medicine, Houston, Texas, USA; Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas, USA), Hirotaka Iwaki (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Hampton L. Leonard (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Mike A. Nalls (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, USA; CEO/Consultant Data Tecnica International, Glen Echo, MD, USA), Laurie Robak (Baylor College of Medicine, Houston, Texas, USA), Jose Bras (Center for Neurodegenerative Science, Van Andel Research Institute, Grand Rapids, Michigan, USA), Rita Guerreiro (Center for Neurodegenerative Science, Van Andel Research Institute, Grand Rapids, Michigan, USA), Steven Lubbe (Ken and Ruth Davee Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA), Steven Finkbeiner (Departments of Neurology and Physiology, University of California, San Francisco; Gladstone Institute of

Neurological Disease; Taube/Koret Center for Neurodegenerative Disease Research, San Francisco, CA, USA), Niccolo E. Mencacci (Northwestern University Feinberg School of Medicine, Chicago, IL, USA), Codrin Lungu (National Institutes of Health Division of Clinical Research, NINDS, National Institutes of Health, Bethesda, MD, USA), Andrew B Singleton (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Sonja W. Scholz (Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA), Xylena Reed (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Roy N. Alcalay (Department of Neurology, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY, USA), Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY, USA).

Canada: Ziv Gan-Or (Montreal Neurological Institute and Hospital, Department of Neurology & Neurosurgery, Department of Human Genetics, McGill University, Montréal, QC, H3A 0G4, Canada), Guy A. Rouleau (Montreal Neurological Institute and Hospital, Department of Neurology & Neurosurgery, Department of Human Genetics, McGill University, Montréal, QC, H3A 0G4, Canada), Lynne Krohn (Montreal Neurological Institute and Hospital, Department of Neurology & Neurosurgery, Department of Human Genetics, McGill University, Montréal, QC, H3A 0G4, Canada), Lynne Krohn (Montreal Neurological Institute and Hospital, Department of Neurology & Neurosurgery, Department of Human Genetics, McGill University, Montréal, QC, H3A 0G4, Canada),

The Netherlands: Jacobus J van Hilten (Department of Neurology, Leiden University Medical Center, Leiden, Netherlands), Johan Marinus (Department of Neurology, Leiden University Medical Center, Leiden, Netherlands).

Spain: Astrid D. Adames-Gómez (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Sevilla), Miquel Aguilar (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona), Ignacio Alvarez (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona), Victoria Alvarez (Hospital Universitario Central de Asturias, Oviedo), Francisco Javier Barrero (Hospital Universitario San Cecilio de Granada, Universidad de Granada), Jesús Alberto Bergareche Yarza (Instituto de Investigación Sanitaria Bionostia, San Sebastián), Inmaculada Bernal-Bernal (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Sevilla), Marta Blazquez (Hospital Universitario Central de Asturias, Oviedo), Marta Bonilla-Toribio (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Sevilla), Juan A. Botía (Universidad de Murcia, Murcia), María Teresa Boungiorno (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona), Dolores Buiza-Rueda (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Sevilla), Ana Cámara (Hospital Clinic de Barcelona), Fátima Carrillo (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Sevilla), Mario Carrión-Claro (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Sevilla), Debora Cerdan (Hospital General de Segovia, Segovia), Jordi Clarimón (Memory Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), Madrid), Yaroslau Compta (Hospital Clinic de Barcelona), Monica Diez-Fairen (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona), Oriol Dols-Icardo (Memory Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), Madrid), Jacinto Duarte (Hospital General de Segovia, Segovia), Raquel Duran (Centro de Investigación Biomédica, Universidad de Granada, Granada), Francisco Escamilla-Sevilla (Hospital Universitario Virgen de las Nieves, Instituto de Investigación Biosanitaria de Granada, Granada), Mario Ezquerro (Hospital Clinic de Barcelona), Cici Feliz (Departamento de Neurología, Instituto de Investigación Sanitaria Fundación Jiménez Díaz, Madrid, Spain), Manel Fernández (Hospital Clinic de Barcelona), Rubén Fernández-Santiago (Hospital Clinic de Barcelona), Ciara Garcia (Hospital Universitario Central de Asturias, Oviedo), Pedro García-Ruiz (Instituto de Investigación Sanitaria Fundación Jiménez Díaz, Madrid), Pilar Gómez-Garre (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Sevilla), María Jose Gomez Heredia (Hospital Universitario Virgen de la Victoria, Malaga), Isabel Gonzalez-Aramburu (Hospital Universitario Marqués de Valdecilla-IDIVAL,

Santander), Ana Gorostidi Pagola (Instituto de Investigación Sanitaria Biodonostia, San Sebastián), Janet Hoernicka (Institut de Recerca Sant Joan de Déu, Barcelona), Jon Infante (Hospital Universitario Marqués de Valdecilla-IDIVAL and University of Cantabria, Santander, and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED)), Silvia Jesús (Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Adriano Jimenez-Escrib (Hospital Universitario Ramón y Cajal, Madrid), Jaime Kulisevsky (Movement Disorders Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED)), Miguel A. Labrador-Espinosa (Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Jose Luis Lopez-Sendon (Hospital Universitario Ramón y Cajal, Madrid), Adolfo López de Munain Arregui (Instituto de Investigación Sanitaria Biodonostia, San Sebastián), Daniel Macías (Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Irene Martínez Torres (Department of Neurology, Instituto de Investigación Sanitaria La Fe, Hospital Universitario y Politécnico La Fe, Valencia), Juan Marín (Movement Disorders Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED)), María Jose Martí (Hospital Clinic Barcelona), Juan Carlos Martínez-Castrillo (Instituto Ramón y Cajal de Investigación Sanitaria, Hospital Universitario Ramón y Cajal, Madrid), Carlota Méndez-del-Barrio (Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Manuel Menéndez González (Hospital Universitario Central de Asturias, Oviedo), Marina Mata (Department of Neurology, Hospital Universitario Infanta Sofía, Madrid, Spain) Adolfo Mir-guez (Hospital Universitario Virgen de las Nieves, Granada, Instituto de Investigación Biosanitaria de Granada), Pablo Mir (Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Elisabet Mondragon Rezola (Instituto de Investigación Sanitaria Biodonostia, San Sebastián), Esteban Muñoz (Hospital Clinic Barcelona), Javier Pagonabarraga (Movement Disorders Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED)), Pau Pastor (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona), Francisco Perez Errazquin (Hospital Universitario Virgen de la Victoria, Malaga), Teresa Perrián-Tocino (Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Javier Ruiz-Martínez (Hospital Universitario Donostia, Instituto de Investigación Sanitaria Biodonostia, San Sebastián), Clara Ruz (Centro de Investigación Biomédica, Universidad de Granada, Granada), Antonio Sanchez Rodriguez (Hospital Universitario Marqués de Valdecilla-IDIVAL, Santander), María Sierra (Hospital Universitario Marqués de Valdecilla-IDIVAL, Santander), Esther Suarez-Sanmartin (Hospital Universitario Central de Asturias, Oviedo), Cesar Tabernero (Hospital General de Segovia, Segovia), Juan Pablo Tartari (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona), Cristina Tejera-Parrado (Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Eduard Tolosa (Hospital Clinic Barcelona), Francesc Valldeoriola (Hospital Clinic Barcelona), Laura Vargas-González (Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Lydia Vela (Department of Neurology, Hospital Universitario Fundación Alcorcón, Madrid), Francisco Vives (Centro de Investigación Biomédica, Universidad de Granada, Granada).

Austria: Alexander Zimprich (Department of Neurology, Medical University of Vienna, Austria).

Norway: Lasse Pihlstrom (Department of Neurology, Oslo University Hospital, Oslo, Norway), Mathias Toft (Department of Neurology and Institute of Clinical Medicine, Oslo University Hospital, Oslo, Norway).

Estonia: Sulev Koks (Department of Pathophysiology, University of Tartu, Tartu, Estonia; Department of Reproductive Biology, Estonian University of Life Sciences, Tartu, Estonia), Pille Taba (Department of Neurology and Neurosurgery, University of Tartu, Tartu, Estonia).

Israel: Sharon Hassin-Baer (The Movement Disorders Institute, Department of Neurology and Sagol Neuroscience Center, Chaim Sheba Medical Center, Tel-Hashomer, 5262101, Ramat Gan, Israel, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel).

Finland: Kari Majamaa (Institute of Clinical Medicine, Department of Neurology, University of Oulu, Oulu, Finland; Department of Neurology and

Medical Research Center, Oulu University Hospital, Oulu, Finland), Ari Siitonen (Institute of Clinical Medicine, Department of Neurology, University of Oulu, Oulu, Finland; Department of Neurology and Medical Research Center, Oulu University Hospital, Oulu, Finland).

Nigeria: Njideka U. Okubadejo (University of Lagos, Lagos State, Nigeria),

Oluwadamilola O. Ojo (University of Lagos, Lagos State, Nigeria),

Kazakhstan: Coordinator - Rauan Kaiyrzhanov (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Chingiz Shashkin (Kazakh National Medical University named after Asfendiyarov, Almaty, Kazakhstan), Nazira Zharkynbekova (South Kazakhstan Medical Academy, Shymkent, Kazakhstan), Vadim Akhmetzhanov (Astana Medical University, Astana Kazakhstan), Akbota Aitkulova (National Center for Biotechnology, Astana, Kazakhstan; Al-Farabi Kazakh national university. Almaty Kazakhstan), Elena Zholdybayeva (National Center for Biotechnology, Astana, Kazakhstan), Zharkyn Zharmukhanov (National Center for Biotechnology, Astana, Kazakhstan), Gulnaz Kaishybayeva (Scientific and practical center "Institute of neurology named after Smagul Kaishybayev", Almaty, Kazakhstan), Altynay Karimova (Scientific and practical center "Institute of neurology named after Smagul Kaishybayev", Almaty, Kazakhstan), Dinara Sadykova (Astana Medical University, Astana, Kazakhstan).

Authors' contributions

TE, MC, MDE and AS designed the study. NM and SP recruited the patients, performed clinical study and collected the data. MGR and AL performed sample management and bio-banking. TN realized the clinical database. AG, MGR and TE performed the analysis of WES data. TE, SS and SG performed the NGS-TR analysis. DR performed the ROC analysis. The International Parkinson's Disease Genetics Consortium (IPDGC) provided validation data set. LI and FG provided controls subjects. MGR performed the mutation analysis of the patients and the RT-qPCR, experiments. LGDG, DA and AS performed immunohistochemistry experiments. AG, MGR, and TE, performed statistical analysis. TE and AS wrote the manuscript. All authors revised the final manuscript. The author(s) read and approved the final manuscript.

Funding

This work was supported by Italian Ministry of Health grant n. RF 2019–12370224 to TE, Current Research to TE and grant n. GR2016–02362247 to DR. This work was also supported by Italian Ministry of Economic Development (M.I.S.E.) project n. F/0009/00X26. AG was supported by Fondazione Umberto Veronesi. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Written informed consent was obtained from all participants. The ethical board of the IRCCS Neuromed approved the study protocol (N°9/2015).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹IRCCS Istituto Neurologico Mediterraneo Neuromed, Pozzilli, Isernia, Italy. ²National Research Council, Institute of Genetics and Biophysics "Adriano Buzzati-Traverso", Naples, Italy. ³Department of Biomolecular Science, University of Urbino Carlo Bò, Urbino, Italy. ⁴Research Center in Epidemiology and Preventive Medicine (EPIMED), Department of Medicine and Surgery, University of Insubria, Varese, Italy.

Received: 9 February 2021 Accepted: 6 May 2021

Published online: 21 June 2021

References

- Shulman JM, De Jager PL, Feany MB. Parkinson's disease: genetics and pathogenesis. *Annu Rev Pathol Mech Dis.* 2011;6(1):193–222. <https://doi.org/10.1146/annurev-pathol-011110-130242>.
- Langston JW, Schüle B, Rees L, Nichols RJ, Barlow C. Multisystem Lewy body disease and the other parkinsonian disorders. *Nat Genet.* 2015;47(12):1378–84. <https://doi.org/10.1038/ng.3454>.
- Jellinger KA. The pathology of Parkinson's disease. *Adv Neurol.* 2001;86:55–72.
- Kalinderi K, Bostantjopoulou S, Fidani L. The genetic background of Parkinson's disease: current progress and future prospects. *Acta Neurol Scand.* 2016;134(5):314–26. <https://doi.org/10.1111/ane.12563>.
- Verstraeten A, Theuns J, Van Broeckhoven C. Progress in unraveling the genetic etiology of Parkinson disease in a genomic era. *Trends Genet.* 2015;31(3):140–9. <https://doi.org/10.1016/j.tig.2015.01.004>.
- Martin I, Dawson VL, Dawson TM. Recent advances in the genetics of Parkinson's disease. *Annu Rev Genomics Hum Genet.* 2011;12(1):301–25. <https://doi.org/10.1146/annurev-genom-082410-101440>.
- Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord.* 2015;30(12):1591–601. <https://doi.org/10.1002/mds.26424>.
- Gialluisi A, Recchia MG, Tirozzi A, Nutile T, Lombardi A, De Sanctis C, et al. Whole exome sequencing study of Parkinson's disease and related endophenotypes in the Italian population. *Front Neurol.* 2020. <https://doi.org/10.3389/fneur.2019.01362>.
- Tirozzi A, Modugno N, Palomba NP, Feresse R, Lombardi A, Olivola E, et al. Analysis of genetic and non-genetic predictors of levodopa induced dyskinesia in Parkinson's disease. *Front Pharmacol.* 2021;12. <https://doi.org/10.3389/fphar.2021.640603>.
- Hentz JG, Mehta SH, Shill HA, Driver-Dunckley E, Beach TG, Adler CH. Simplified conversion method for unified Parkinson's disease rating scale motor examinations. *Mov Disord.* 2015;30(14):1967–70. <https://doi.org/10.1002/mds.26435>.
- Conti S, Bonazzi S, Laiacona M, Masina M, Coralli MV. Montreal cognitive assessment (MoCA)-Italian version: regression based norms and equivalent scores. *Neurol Sci.* 2015;36(2):209–14. <https://doi.org/10.1007/s10072-014-1921-3>.
- Cova I, Di Battista ME, Vanacore N, Papi CP, Alampì G, Rubino A, et al. Validation of the Italian version of the non motor symptoms scale for Parkinson's disease. *Park Relat Disord.* 2017;34:38–42. <https://doi.org/10.1016/j.parkrel.2016.10.020>.
- Shulman LM, Gruber-Baldini AL, Anderson KE, Vaughan CG, Reich SG, Fishman PS, et al. The evolution of disability in Parkinson disease. *Mov Disord.* 2008;23(6):790–6. <https://doi.org/10.1002/mds.21879>.
- The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015;526(7571):68–74. <https://doi.org/10.1038/nature15393>.
- Iacoviello L, Bonanni A, Costanzo S, De Curtis A, Di Castelnuovo A, Olivieri M, et al. The Moli-Sani project, a randomized, prospective cohort study in the Molise region in Italy; design, rationale and objectives. *IJPH.* 2007;4(2):110–8.
- Jansen IE, Ye H, Heetveld S, Lechler MC, Michels H, Seinstra RI, et al. Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing. *Genome Biol.* 2017;30:22.
- Li H, Durbin R. Fast and accurate short read alignment with burrows wheeler transform. *Bioinformatics.* 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
- 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(5):491–8. <https://doi.org/10.1038/ng.806>.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 2015;4(1):7. <https://doi.org/10.1186/s13742-015-0047-8>.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164. <https://doi.org/10.1093/nar/gkq603>.
- Sampaolo S, Napolitano F, Tirozzi A, Recchia MG, Lombardi L, Farina O, et al. Identification of the first dominant mutation of LAMA5 gene causing a complex multisystem syndrome due to dysfunction of the extracellular matrix. *J Med Genet.* 2017;54(10):710–20. <https://doi.org/10.1136/jmedgenet-2017-104555>.
- Napolitano F, Terracciano C, Bruno G, De Blasiis P, Lombardi L, Gialluisi A, et al. Novel autophagic vacuolar myopathies: phenotype and genotype features. *Neuropathol Appl Neurobiol.* 2021;00:1–15. <https://doi.org/10.1111/nan.12690>.
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRIPED v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019 Jan;47(D1):D607–13. <https://doi.org/10.1093/nar/gky1131>.
- R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014. <http://www.R-project.org/>
- Di Salvio M, Di Giovannantonio LG, Acampora D, Prosperi R, Omodei D, Prakash N, et al. Otx2 controls neuron subtype identity in ventral tegmental area and antagonizes vulnerability to MPTP. *Nat Neurosci.* 2010;13(12):1481–8. <https://doi.org/10.1038/nn.2661>.
- Lubbe SJ, Escott-Price V, Gibbs JR, Nalls MA, Bras J, Price TR, et al. Additional rare variant analysis in Parkinson's disease cases with and without known pathogenic mutations: evidence for oligogenic inheritance. *Hum Mol Genet.* 2016;25(24):5483–9. <https://doi.org/10.1093/hmg/ddw348>.
- Akabane S, Uno M, Tani N, Shimazaki S, Ebara N, Kato H, et al. PKA regulates PINK1 stability and Parkin recruitment to damaged mitochondria through phosphorylation of MIC60. *Mol Cell.* 2016;62(3):371–84. <https://doi.org/10.1016/j.molcel.2016.03.037>.
- John GB, Shang Y, Li L, Renken C, Mannella CA, Selker JM, et al. The mitochondrial inner membrane protein mitofilin controls cristae morphology. *Mol Biol Cell.* 2005;16(3):1543–54. <https://doi.org/10.1091/mbc.e04-08-0697>.
- Wang X, Winter D, Ashrafi G, Schlehe J, Wong YL, Selkoe D, et al. PINK1 and Parkin target Miro for phosphorylation and degradation to arrest mitochondrial motility. *Cell.* 2011;147(4):893–906. <https://doi.org/10.1016/j.cell.2011.10.018>.
- López-Doménech G, Covill-Cooke C, Ivankovic D, Halff EF, Sheehan DF, Norkett R, et al. Miro proteins coordinate microtubule- and actin-dependent mitochondrial transport and distribution. *EMBO J.* 2018;37(3):321–36. <https://doi.org/10.15252/emboj.201696380>.
- Bertolin G, Ferrando-Miguel R, Jacoupy M, Traver S, Grenier K, Greene AW, et al. The TOMM machinery is a molecular switch in PINK1 and PARK2/PARKIN-dependent mitochondrial clearance. *Autophagy.* 2013;9(11):1801–17. <https://doi.org/10.4161/aut.25884>.
- Ayuso P, Martínez C, Lorenzo-Betancor O, Pastor P, Luengo A, Jiménez-Jiménez FJ, et al. A polymorphism located at an ATG transcription start site of the heme oxygenase-2 gene is associated with classical Parkinson's disease. *Pharmacogenet Genomics.* 2011;21(9):565–71. <https://doi.org/10.1097/FPC.0b013e328328348f729>.
- Wang L, Suzuki T. Dual functions for cytosolic α -mannosidase (Man2C1): its down-regulation causes mitochondria-dependent apoptosis independently of its α -mannosidase activity. *J Biol Chem.* 2013;288(17):11887–96. <https://doi.org/10.1074/jbc.M112.425702>.
- Slabbaert JR, Kuenen S, Swerts J, Maes I, Uytterhoeven V, Kasprzowicz J, et al. Shawn, the Drosophila homolog of SLC25A39/40, is a mitochondrial carrier that promotes neuronal survival. *J Neurosci.* 2016;36(6):1914–29. <https://doi.org/10.1523/JNEUROSCI.3432-15.2016>.
- Lee Y, Karuppagounder SS, Shin JH, Lee YI, Ko HS, Swing D, et al. Parthanatos mediates AIMP2-activated age-dependent dopaminergic neuronal loss. *Nat Neurosci.* 2013;16(10):1392–400. <https://doi.org/10.1038/nn.3500>.
- Howard TL, Stauffer DR, Degnin CR, Hollenberg SM. CHMP1 functions as a member of a newly defined family of vesicle trafficking proteins. *J Cell Sci.* 2001;114(13):2395–404. <https://doi.org/10.1242/jcs.114.13.2395>.
- Muhia M, Thies E, Labonté D, Ghirelli AE, Gromova KV, Xompero F, et al. The Kinesin KIF21B regulates microtubule dynamics and is essential for neuronal morphology, synapse function, and learning and memory. *Cell Rep.* 2016;15(5):968–77. <https://doi.org/10.1016/j.celrep.2016.03.086>.
- Kobayashi T, Tsang WY, Li J, Lane W, Dynlacht BD. Centriolar kinesin Kif24 interacts with CP110 to remodel microtubules and regulate ciliogenesis. *Cell.* 2011;145(6):914–25. <https://doi.org/10.1016/j.cell.2011.04.028>.
- Lee HJ, Lee K, Im H. α -Synuclein modulates neurite outgrowth by interacting with SPTBN1. *Biochem Biophys Res Commun.* 2012;424(3):497–502. <https://doi.org/10.1016/j.bbrc.2012.06.144>.

40. Shang G, Brautigam CA, Chen R, Lu D, Torres-Vázquez J, Zhang X. Structure analyses reveal a regulated oligomerization mechanism of the PlexinD1/GIPC/myosin VI complex. *Elife*. 2017;24:6.
41. Fontaine SN, Zheng D, Sabbagh JJ, Martin MD, Chaput D, Darling A, et al. DnaJ/Hsc70 chaperone complexes control the extracellular release of neurodegenerative-associated proteins. *EMBO J*. 2016;35(14):1537–49. <https://doi.org/10.15252/embj.201593489>.
42. Jinn S, Drolet RE, Cramer PE, Wong AH, Toolan DM, Gretzula CA, et al. TMEM175 deficiency impairs lysosomal and mitochondrial function and increases α -synuclein aggregation. *Proc Natl Acad Sci U S A*. 2017;114(9):2389–94. <https://doi.org/10.1073/pnas.1616332114>.
43. Lassot I, Mora S, Lesage S, Zieba BA, Coque E, Condroyer C, et al. The E3 ubiquitin ligases TRIM17 and TRIM41 modulate α -Synuclein expression by regulating ZSCAN21. *Cell Rep*. 2018;25(9):2484–96. <https://doi.org/10.1016/j.celrep.2018.11.002>.
44. Dermentzaki G, Paschalidis N, Politis PK, Stefanis L. Complex effects of the ZSCAN21 transcription factor on transcriptional regulation of α -Synuclein in primary neuronal cultures and in vivo. *J Biol Chem*. 2016;291(16):8756–72. <https://doi.org/10.1074/jbc.M115.704973>.
45. Salašová A, Yokota C, Potěšil D, Zdráhal Z, Bryja V, Arenas E. A proteomic analysis of LRRK2 binding partners reveals interactions with multiple signaling components of the WNT/PCP pathway. *Mol Neurodegener*. 2017;12(1):54. <https://doi.org/10.1186/s13024-017-0193-9>.
46. Tsai PI, Lin CH, Hsieh CH, Papakyriakos AM, Kim MJ, Napolioni V, et al. PINK1 phosphorylates MITC60/Mitofilin to control structural plasticity of mitochondrial crista junctions. *Mol Cell*. 2018;69(5):744–56. <https://doi.org/10.1016/j.molcel.2018.01.026>.
47. 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;56(9):1–7.
48. Abou-Sleiman PM, Muqit MM, McDonald NQ, Yang YX, Gandhi S, Healy DG, et al. A heterozygous effect for PINK1 mutations in Parkinson's disease? *Ann Neurol*. 2006;60(4):414–9. <https://doi.org/10.1002/ana.20960>.
49. Djarmati A, Hedrich K, Svetel M, Lohnau T, Schwinger E, Romac S, et al. Heterozygous PINK1 mutations: a susceptibility factor for Parkinson disease? *Mov Disord*. 2006;21(9):1526–30. <https://doi.org/10.1002/mds.20977>.
50. Nuytemans K, Theuns J, Cruts M, Van Broeckhoven C. Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update. *Hum Mutat*. 2010;31(7):763–80. <https://doi.org/10.1002/humu.21277>.
51. Sun M, Latourelle JC, Wooten GF, Lew MF, Klein C, Shill HA, et al. Influence of heterozygosity for parkin mutation on onset age in familial Parkinson disease: the GenePD study. *Arch Neurol*. 2006;63(6):826–32. <https://doi.org/10.1001/archneur.63.6.826>.
52. Klein C, Djarmati A, Hedrich K, Schäfer N, Scaglione C, Marchese R, et al. PINK1, Parkin, and DJ-1 mutations in Italian patients with early-onset parkinsonism. *Eur J Hum Genet*. 2005;13(9):1086–93. <https://doi.org/10.1038/sj.ejhg.5201455>.
53. Rimsa V, Eadsforth TC, Hunter WN. Structure of the SCAN domain of human paternally expressed gene 3 protein. *PLoS One*. 2013;8(7):e69538. <https://doi.org/10.1371/journal.pone.0069538>.
54. Lee C, Guo J, Zeng W, Kim S, She J, Cang C, et al. The lysosomal potassium channel TMEM175 adopts a novel tetrameric architecture. *Nature*. 2017;547(7664):472–5. <https://doi.org/10.1038/nature23269>.
55. Ren Z, Yang N, Ji C, Zheng J, Wang T, Liu Y, et al. Neuroprotective effects of 5-(4-hydroxy-3-dimethoxybenzylidene)-thiazolidinone in MPTP induced parkinsonism model in mice. *Neuropharmacology*. 2015;93:e218.
56. Kim J, Lee S, Ko S, Kim-Ha J. dGIPC is required for the locomotive activity and longevity in *Drosophila*. *Biochem Biophys Res Commun*. 2010;402(3):565–70. <https://doi.org/10.1016/j.bbrc.2010.10.095>.
57. Lorenzo DN, Badesa A, Zhou R, Mohler PJ, Zhuang X, Bennett V. β II-spectrin promotes mouse brain connectivity through stabilizing axonal plasma membranes and enabling axonal organelle transport. *Proc Natl Acad Sci U S A*. 2019;116(31):15686–95. <https://doi.org/10.1073/pnas.1820649116>.
58. Gao X, Carroni M, Nussbaum-Krammer C, Mogk A, Nillegodda NB, Szlachcic A, et al. Human Hsp70 Disaggregase reverses Parkinson's-linked α -Synuclein amyloid fibrils. *Mol Cell*. 2015;59(5):781–93. <https://doi.org/10.1016/j.molcel.2015.07.012>.
59. Venturilli E, Villa C, Fenoglio C, Clerici F, Marcone A, Benussi L, et al. Is KIF24 a genetic risk factor for Frontotemporal lobar degeneration? *Neurosci Lett*. 2010;482(3):240–4. <https://doi.org/10.1016/j.neulet.2010.07.047>.
60. Rytter SW, Tyrrell RM. The heme synthesis and degradation pathways: role in oxidant sensitivity. Heme oxygenase has both pro and antioxidant properties. *Free Radic Biol Med*. 2000;28(2):289–309. [https://doi.org/10.1016/S0891-5849\(99\)00223-3](https://doi.org/10.1016/S0891-5849(99)00223-3).
61. Paciotti S, Persichetti E, Klein K, Tasegian A, Duvet S, Hartmann D, et al. Accumulation of free oligosaccharides and tissue damage in cytosolic α -mannosidase (Man2c1)-deficient mice. *J Biol Chem*. 2014;289(14):9611–22. <https://doi.org/10.1074/jbc.M114.550509>.
62. Mochida GH, Ganesh VS, de Michelena MI, Dias H, Atabay KD, Kathrein KL, et al. CHMP1A encodes an essential regulator of BMI1-INK4A in cerebellar development. *Nat Genet*. 2012;44(11):1260–4. <https://doi.org/10.1038/ng.2425>.
63. Healy DG, Falchi M, O'Sullivan SS, Bonifati V, Durr A, Bressman S, et al. Rhesotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol*. 2008;7(7):583–90. [https://doi.org/10.1016/S1474-4422\(08\)70117-0](https://doi.org/10.1016/S1474-4422(08)70117-0).
64. Benitez BA, Davis AA, Jin SC, Ibanez L, Ortega-Cubero S, Pastor P, et al. Resequencing analysis of five Mendelian genes and the top genes from genome-wide association studies in Parkinson's disease. *Mol Neurodegener*. 2016;11(1):29. <https://doi.org/10.1186/s13024-016-0097-0>.
65. Spitz M, Pereira JS, Nicareta DH, Abreu Gde M, Bastos EF, Seixas TL, et al. Association of LRRK2 and GBA mutations in a Brazilian family with Parkinson's disease. *Parkinsonism Relat Disord*. 2015;21(7):825–6. <https://doi.org/10.1016/j.parkrel.2015.03.029>.
66. Escott-Price V, International Parkinson's disease genomics consortium, Nalls MA, Morris HR, Lubbe S, Brice A, et al. Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann Neurol*. 2015;77(4):582–91. <https://doi.org/10.1002/ana.24335>.
67. Blauwendraat C, Reed X, Krohn L, Heilbron K, Bandres-Ciga S, Tan M, et al. Genetic modifiers of risk and age at onset in GBA associated Parkinson's disease and Lewy body dementia. *Brain*. 2020;143(1):234–48. <https://doi.org/10.1093/brain/awz350>.
68. Gan-Or Z, Liong C, Alcalay RN. GBA-associated Parkinson's disease and other Synucleinopathies. *Curr Neurol Neurosci Rep*. 2018;18(8):44. <https://doi.org/10.1007/s11910-018-0860-4>.
69. Gialluisi A, Andlauer TFM, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, et al. Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. *Transl Psychiatry*. 2019;9(1):77. <https://doi.org/10.1038/s41398-019-0402-0>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

