1 **Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease**

2

3 Laurie A. Robak, MD, PhD[1,2,*], Iris E. Jansen, PhD[3,4,*], Jeroen van Rooij, BSc[5,6,7], André G.

4 Uitterlinden, PhD[5,6,8], Robert Kraaij, PhD[5,6,8], Joseph Jankovic, MD[9], International Parkinson's

5 Disease Genomics Consortium (IPDGC), Peter Heutink, PhD[3,#], Joshua M. Shulman, MD,

6 PhD[1,2,9,10,#]

7

8 1. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston TX
9   USA
10 2. Jan and Dan Duncan Neurologic Research Institute, Texas Children's Hospital, Houston TX
11   USA
12 3. German Center for Neurodegenerative Diseases (DZNE), Tübingen 72076, Germany
13 4. Department of Clinical Genetics, VU University Medical Center, Amsterdam 1081HZ, The
14   Netherlands
15 5. Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands
16 6. Netherlands Consortium for Healthy Ageing (NCHA), Rotterdam, The Netherlands
17 7. Department of Neurology, Erasmus MC, Rotterdam, the Netherlands
18 8. Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands
19 9. Department of Neurology, Baylor College of Medicine, Houston, TX, USA
20 10. Department of Neuroscience and Program in Developmental Biology, Baylor College of
21   Medicine, Houston, TX, USA

22

23 * Co-first authors, contributed equally.

24

25 [#] Correspondence:
26 Joshua M. Shulman, MD, PhD
27 Jan and Dan Duncan Neurological Research Institute
28 1250 Moursund St., Suite N.1150
29 Houston, TX 77030
30 832 824-8976
31 Email: Joshua.Shulman@bcm.edu

32

33 Peter Heutink, PhD
34 German Center for Neurodegenerative Diseases (DZNE)
35 Otfried-Müller-Str. 23
36 72076 Tübingen
37 Germany
38 Email: Peter.Heutink@dzne.de

39

40 <u>Running Head:</u> Lysosomal storage disorder genes and Parkinson's disease

Abstract

Mutations in the *glucocerebrosidase* gene (*GBA*), which cause Gaucher disease, are also potent risk factors for Parkinson's disease. We examined whether a genetic burden of variants in other lysosomal storage disorder genes is more broadly associated with Parkinson's disease susceptibility. The sequence kernel association test was used to interrogate variant burden among 54 lysosomal storage disorder genes, leveraging whole exome sequencing data from 1,156 Parkinson's disease cases and 1,679 control subjects. We discovered a significant burden of rare, likely damaging lysosomal storage disorder gene variants in association with Parkinson's disease risk. The association signal was robust to the exclusion of *GBA*, and consistent results were obtained in 2 independent replication cohorts, including 436 cases and 169 controls with whole exome sequencing and an additional 6,713 cases and 5,964 controls with exome-wide genotyping. In secondary analyses designed to highlight the specific genes driving the aggregate signal, we confirmed associations at the *GBA* and *SMPD1* loci and newly implicate *CTSD, SLC17A5,* and *ASAH1* as candidate Parkinson's disease susceptibility genes. In our discovery cohort, the majority of Parkinson's disease cases (56%) have at least one putative damaging variant in a lysosomal storage disorder gene, and 21% carry multiple alleles. Our results highlight several promising new susceptibility loci and reinforce the importance of lysosomal mechanisms in Parkinson's disease pathogenesis. We suggest that multiple genetic hits may act in combination to degrade lysosomal function, enhancing Parkinson's disease susceptibility.

Key Words

Parkinson's disease

Lysosomal storage disorders

1 Genetics

2 Whole exome sequencing

3

Abbreviations

5 CADD   Combined Annotation Dependent Depletion

6 GBA   Glucocerebrosidase

7 IPDGC   International Parkinson's Disease Genomics Consortium

8 LSD   Lysosomal Storage Disorder

9 MAF   Minor allele frequency

10 PPMI   Parkinson's Progression Markers Initiative

11 QC   Quality control

12 RSX-1   Rotterdam Study exome dataset version 1

13 SKAT-O   Sequence kernel association test – optimized

14 WES   Whole exome sequencing

15

Introduction

17 Parkinson's disease is a common neurodegenerative disorder with evidence for a substantial

18 genetic etiology (Kalia and Lang, 2015). Studies in families as well as large population-based

19 cohorts have implicated more than 30 genes (Bras *et al.*, 2015; Singleton *et al.*, 2013;

20 Verstraeten *et al.*, 2015); however, the risk alleles identified to date explain only a fraction of

21 Parkinson's disease heritability estimates (Do *et al.*, 2011; Hamza and Paymi, 2010; Keller *et al.*,

22 2012), suggesting the involvement of additional loci. Beyond discovering the responsible genes,

23 a major challenge remains to understand the mechanisms by which these factors alter disease

1   onset and/or progression, including whether they act independently or interact within coherent

2   biologic pathways.

3   Substantial evidence highlights the importance of lysosomal mechanisms in Parkinson's

4   disease susceptibility and pathogenesis (Kalia and Lang, 2015; Moors *et al.*, 2016; Vekrellis *et*

5   *al.*, 2011; Wong and Krainc, 2016). Prior to its discovery as a Parkinson's disease risk locus, the

6   *glucocerebrosidase* gene, *GBA*, was known to cause Gaucher disease, an autosomal recessive

7   lysosomal storage disorder (LSD). Increased risk for Parkinson's disease in heterozygous carriers

8   of *GBA* loss-of-function alleles was first recognized in families of individuals with Gaucher

9   disease (Goker-Alpan *et al.*, 2004; Tayebi *et al.*, 2003). Follow-up studies in large, case-control

10 samples confirmed that heterozygous *GBA* variants confer at least a five-fold increased risk of

11 Parkinson's disease (Aharon-Peretz *et al.*, 2004; Sidransky *et al.*, 2009). *GBA* variants may also

12 modify Parkinson's disease clinical manifestations (Brockmann *et al.*, 2015; Clark *et al.*, 2007;

13 Davis *et al.*, 2016; Winder-Rhodes *et al.*, 2012), causing earlier age-of-onset, higher risk of

14 cognitive impairment, and accelerated progression. LSDs—of which there are more than 50—are

15 strictly Mendelian-inherited, metabolic disorders collectively caused by dysfunction in lysosomal

16 biogenesis or function, and similarly characterized by the abnormal accumulation of non-

17 degraded metabolites in the lysosome (Boustany, 2013; Filocamo and Morrone, 2011). The

18 strong genetic evidence linking Gaucher disease and Parkinson's disease risk leads to the

19 intriguing, generalized hypothesis that LSDs and Parkinson's disease may share a common

20 genetic mechanism. Other LSD genes have therefore become attractive candidate risk factors for

21 Parkinson's disease (Deng *et al.*, 2015; Shachar *et al.*, 2011). Several studies have consistently

22 supported a role for *SMPD1* (Clark *et al.*, 2015; Foo *et al.*, 2013; Gan-Or *et al.*, 2013; 2015; Wu

23 *et al.*, 2014), which causes Niemann-Pick disease, Type A/B. Initial reports evaluating other

LSD genes, including *NPC1*, *NPC2*, *MCOLN1, NAGLU* and *ARSB*, have either shown

conflicting results or await further replication (Clark *et al.*, 2015; Jansen *et al.*, 2017;

Kluenemann *et al.*, 2013; Winder-Rhodes *et al.*, 2012; Zech *et al.*, 2013). LSDs are individually

quite rare in populations of European ancestry, as are the known genetic variants established to

cause these disorders (Boustany, 2013; Filocamo and Morrone, 2011). However, with the

exception of *GBA*, most studies of LSD gene candidates have been small and therefore likely

underpowered to detect the effects of rare alleles or those with more modest effect sizes.

Genome-wide association studies in large Parkinson's disease case-control cohorts have

independently implicated more common risk alleles at another LSD gene, *SCARB2* (Do *et al.*,

2011; Nalls *et al.*, 2014), which encodes a membrane protein required for correct targeting of

glucocerebrosidase to the lysosome. Besides this growing genetic evidence, studies in cellular

and animal models also implicate the lysosome in the clearance of alpha-synuclein (Cuervo *et

al.*, 2004; H. J. Lee, 2004; Vogiatzi *et al.*, 2008), which aggregates to form Lewy body pathology

in Parkinson's disease. Reciprocally, alpha-synuclein disrupts neuronal vesicle trafficking and

lysosomal function (Cooper *et al.*, 2006; Mazzulli *et al.*, 2011; Moors *et al.*, 2016; Wong and

Krainc, 2016).

In this study, we leverage the largest Parkinson's disease whole exome sequencing

(WES) dataset currently available to systematically examine the overlap between genes

responsible for LSDs and Parkinson's disease. Our results reveal an aggregate burden for genetic

variants among 54 genes established to cause LSDs and suggest that many genes besides *GBA*

likely contribute to susceptibility for Parkinson's disease.

Materials and Methods

1 **Subjects**

2 Clinical and demographic features for our study cohorts, which have also been described in other

3 recent reports (Giri *et al.*, 2017; Jansen *et al.*, 2017), are shown in Supplemental Table 1. The

4 International Parkinson's Disease Genomics Consortium (IPDGC) WES discovery dataset used

5 for this study consists of 2,835 samples of Northern and Western European ancestry, including

6 1,156 Parkinson's disease cases and 1,679 controls not known to have Parkinson's disease.

7 Subjects were recruited from academic medical centers across the United States and Europe.

8 Cases were recruited at a mean age of 51.5 years (SD=11.5) and diagnosed with Parkinson's

9 disease at a mean age of 41.2 years (SD=10.8); 40.4% reported a positive family history. Control

10 subjects were on average 63.7 years of age (SD=17.1). 1,201 control exomes originated from the

11 Rotterdam Study exome dataset version 1 (RSX-1) (van Rooj *et al.*, 2017; Giri *et al.*, 2017). The

12 Rotterdam Study is a prospective population-based cohort study based in Rotterdam, the

13 Netherlands. WES was performed on DNA from participants from the RSX-I subcohort, enrolled

14 in 1990, with an average age at baseline of 68.6 (SD=8.6, 54.4% female) (Hofman *et al.*, 2015).

15 All IPDGC and RSX-1 subjects gave written informed consent for participation in genetic

16 research, which was approved by relevant oversight committees and institutional review boards.

17 Subjects with pathogenic variants in established Mendelian Parkinson's disease genes (*SNCA,*

18 *LRRK2, VPS35, PARK2/parkin, PARK7/DJ-1,* or *PINK1*) were excluded from analysis (Jansen *et*

19 *al.*, 2017). Following quality control filters, the Parkinson's Progression Markers Initiative

20 (PPMI) replication dataset (Parkinson Progression Marker Initiative, 2011) includes 436 cases

21 and 169 controls of Northwest European descent. Cases were recruited at a mean age of 61.7

22 years (SD 9.7) and diagnosed with Parkinson's disease at an average age of 59.8 years

23 (SD=10.0); 27.1% reported a positive family history. PPMI controls were an average of 61.8

1   years of age (SD=10.1) at the time of evaluation. Data used in the preparation of this article were

2   obtained from the PPMI database ([www.ppmi-info.org/data)](www.ppmi-info.org/data); for up-to-date information on the

3   study, visit www.ppmi-info.org. Samples analyzed for both the IPDGC and PPMI cohorts were

4   derived from whole blood. The NeuroX cohort has also been previously described in detail

5   (Jansen *et al.*, 2017; Nalls *et al.* 2015). A minority of subjects overlapping with the IPDGC WES

6   discovery sample were removed, such that the NeuroX replication cohort was a completely

7   independent sample, including 6,713 individuals with Parkinson's disease and 5,964 controls.

8   NeuroX cases were diagnosed at an average age of 61.6 (SD=12.4) and controls were evaluated

9   at an average age of 64.1 (SD=14.3).

10

11   **Sequencing/Genotyping and Quality Control**

12   Data generation and detailed quality control procedures for the IPDGC and RSX-1 samples has

13   recently been reported (Giri *et al.*, 2017; Jansen *et al.*, 2017; van Rooj *et al.*, 2017). WES was

14   performed using the Roche Nimblegen SeqCap v2 or Illumina exome capture kits to prepare

15   sample libraries, followed by paired-end sequencing with Illumina HiSeq2000. The generation of

16   the PPMI WES dataset are described elsewhere (www.ppmi-info.org). Although the datasets

17   originate from different consortia, the same algorithms were used for read processing. The

18   Burrows-Wheeler Aligner-MEM algorithm (Li and Durbin, 2010) was used for alignment of

19   sequencing reads to the human reference genome (hg19). Using Picard tools

20   (http://broadinstitute.github.io/picard), Binary Alignment/Map files were generated in a sorted

21   and indexed manner. Alignments were Base-Quality score recalibrated and indels realigned

22   using the Genome Analysis Toolkit (McKenna *et al.*, 2010) v3.3-0, after which single nucleotide

23   variants and small insertions/deletions were called with the HaplotypeCaller to one genomic

1   Variant Call Format file per individual. The IPDGC and RSX-1 WES datasets (hereafter referred

2   to as simply the IPDGC discovery dataset) were merged by joint variant calling from the

3   individual genomic Variant Call Format files. Variants that were not assigned with the standard

4   Genome Analysis Toolkit quality annotation 'PASS' were excluded for subsequent analyses.

5   94.4% and 98.0% of the IPDGC and PPMI exomes, respectively, achieved a minimum of 10x

6   coverage.

7   As previously described (Giri *et al.*, 2017; Jansen *et al.*, 2017), for individual quality

8   control, samples were excluded for ambiguous gender, deviating heterozygosity/genotype calls,

9   low genotype call rates, or cryptic relatedness following identity-by-descent analyses. Population

10   structure was further evaluated using multi-dimensional scaling component analysis based on

11   linkage disequilibrium-pruned, genome-wide common variant markers. Prior to these

12   calculations, our datasets were merged with available genotypes from 1000 Genomes Project

13   (1000GP) ancestry-based population samples, including African (AFR), East Asian (EAS),

14   European (EUR) and the Americas (AMR) (1000 Genomes Project Consortium 2012). Using the

15   European samples as a reference, population outliers were excluded, resulting in the removal of

16   39 or 9 individuals from the IPDGC and PPMI datasets, respectively. All remaining samples

17   cluster tightly with European ancestry subjects on multi-dimensional scaling plots (Supplemental

18   Figure 1). Genotype and variant quality control was accomplished by removal of low-quality

19   genotypes (Phred-scaled genotype quality score < 20, depth < 8) and variants with low call rates

20   or departure from Hardy-Weinberg equilibrium. Furthermore, for the IPDGC discovery dataset,

21   variants were only considered when located within the overlapping targeted regions of the

22   applied library preparation capture kits. Post-quality control procedures, a total of 462,946 and

23   192,421 variants were called for the IPDGC and PPMI datasets, respectively.

1   Data generation and quality control for the NeuroX cohort has also previously described

2 in detail (Jansen *et al.* 2017; Nalls *et al.*, 2015). NeuroX consists of 242,901 exonic variants from

3 the Illumina Infinium HumanExome BeadChip and 24,706 custom variants related to neurologic

4 disease. For individual quality control, as above, samples were excluded for gender ambiguity,

5 dubious heterozygosity/genotype calls, evidence of relatedness, or poor clustering on multi-

6 dimensional scaling plots (Supplemental Figure 1). We similarly excluded variants for low call

7 rates, departure from Hardy-Weinberg equilibrium, or for significant differences in missingness

8 rate between cases and controls. Post-quality control, we called 177,028 exonic variants from the

9 NeuroX dataset.

10   Where allowable based on individual consents and institutional review board approval,

11 the datasets used in this study, including WES and NeuroX data from the IPDGC, are publicly

12 available. Data availability is detailed in Jansen et al. 2017 and at http://pdgenetics.org/resources.

13 Data from PPMI is also available for download at http://www.ppmi-info.org/access-data-

14 specimens/download-data/.

15

16 **Variant Selection**

17 Our analyses initially considered 54 LSDs (Table 1), defined based on widely accepted clinical,

18 pathologic, and metabolic criteria (Amberger *et al.*, 2015; Boustany, 2013; Filocamo and

19 Morrone, 2011). All variants within the LSD gene set were extracted from the three datasets. For

20 the IPDGC WES dataset, no variants in the genes *CLN5* and *NEU1* passed the pre-specified

21 maximum missingness criteria of 15%, yielding 1,136 total exonic variants for consideration in

22 these analyses. In addition, there were no non-synonymous variants identified in *SUMF1*.

23 Variants were categorized in nested groups (Figure 1) including (1) nonsynonymous (n=760

1   variants in 51 genes), (2) likely damaging (n=596 variants in 51 genes), or (3) loss-of-function

2   (n=69 variants in 27 genes) (see Table 1 and Supplemental Table 2). Loss-of-function variants

3   included stop gain/loss, frameshift, and splicing mutations falling within two base pairs of exon-

4   intron junctions. Predictions of variant pathogenicity were obtained from ANNOVAR (Wang *et*

5   *al.*, 2010), based on the Combined Annotation Dependent Depletion (CADD) algorithm (v1.3,

6   http://cadd.gs.washington.edu) (Kircher *et al.*, 2014). CADD integrates predictions from

7   numerous bioinformatic algorithms into a single "C-score" and ranks all possible nucleotide

8   changes in the genome based on potential to disrupt gene/protein function. In accordance with

9   prior work (Amendola *et al.*, 2015), we selected a stringent CADD C-score≥12.37, representing

10 the top ~2% most damaging of all possible nucleotide changes in the genome—this subset is

11 enriched for known pathogenic alleles. For descriptive purposes, all putative damaging variants

12 within the IPDGC discovery cohort were further cross-referenced with ClinVar (Landrum *et al.*,

13 2016) to identify those previously established with pathogenicity for LSDs (Supplemental Table

14 3). For the PPMI cohort, no variants were called in *DNAJC5*, resulting in a dataset of 515 total

15 exonic variants, of which 256 variants from 49 genes were nonsynonymous and 187 variants in

16 47 genes met the CADD criteria for putative damaging changes (Supplemental Table 2). For the

17 NeuroX cohort, all genes in the 54-gene set were represented, resulting in 467 nonsynonymous

18 variants, of which 348 were classified as likely damaging (Supplemental Table 2). Within these

19 categories, variants were filtered based on two minor allele frequency (MAF) thresholds: (a)

20 <1% and (b) <3% (Figure 1). The latter, more relaxed frequency threshold is based on the

21 population prevalence (de Lau and Breteler, 2006; Pringsheim *et al.*, 2014) and known

22 incomplete penetrance of Parkinson's disease risk alleles (Anheim *et al.* 2012; Marder *et al.*,

23 2015; Rana *et al.* 2013; Trinh *et al.*, 2014). For a subset of individuals in the IPDGC (n=572) and

1   PPMI (n=566) WES cohorts, array-based genotyping data was also available, allowing us to

2   compute concordance rates for genotyping calls present in both datasets using 2 independent

3   assays (Supplemental Table 4). We observe complete concordance for *GBA* variants as well as

4   nearly perfect concordance (>>99%) for variant genotype calls in the full LSD gene set.

5

6   **Statistical Analysis**

7   The sequence kernel association test – optimal (SKAT-O) (S. Lee *et al.*, 2012; 2016) was

8   implemented in R using SKAT v1.0.9 to determine the difference in the aggregate burden of

9   rare LSD gene variants between Parkinson's disease cases and controls. SKAT-O aggregates

10   genetic information across defined genomic regions to test for associations. Covariates were

11   included to adjust analyses for gender and WES coverage (pre-quality control missingness).

12   Twenty multi-dimensional scaling components were also included to account for other possible

13   confounding factors (4 components for analyses of the NeuroX genotyping cohort). An

14   empirical *p*-value (*p*) was derived from the distribution of null results based on 10,000

15   permutation trials in which case/control assignment was randomized. As diagrammed in Figure

16   1, SKAT-O analysis was initially performed for the complete LSD gene set, considering each

17   class of variants defined based on frequency and functional characteristics. In order to adjust for

18   multiple comparisons, we applied the Bonferroni-Holm stepwise procedure to control for the

19   familywise error rate and establish a corrected statistical significance threshold and adjusted *p*-

20   value (p_{adj}) based on a significance level, *alpha*, of 0.05 (Holm 1979). For those categories with

21   a significant SKAT-O association in the full gene set, a secondary analysis was performed

22   excluding all *GBA* variants in order to confirm the involvement of additional genes. For

23   example, in the IPDGC discovery cohort, we adjusted for *k*=5 or 2 comparisons for the number

of variant categories evaluated in the primary and secondary analyses, respectively. Due to the nested variant categories (Figure 1) and the highly interdependent nature of the respective burden tests, we separately considered those results with an empirical SKAT-O p-value < 0.05, but not surviving the Bonferonni-Holm correction, as "suggestive". Unadjusted, empiric SKAT-O p-values for all gene set analyses are included in Supplemental Table 5. Lastly, in order to highlight those loci driving associations detected in the gene set, secondary analyses were also performed using SKAT-O to evaluate variants in each LSD gene independently. For these per gene analyses, which we considered exploratory due to limited statistical power (below), we report all findings with an empirical unadjusted p-value < 0.05.

To estimate statistical power, we performed 1,000 SKAT simulations of causal subregions within the discovery or replication datasets. We assumed a Parkinson's disease prevalence of 0.0041 and 0.0017 for the IPDGC and PPMI datasets, respectively, based on their distinct ages of onset (Supplemental Table 1) (Pringsheim *et al.*, 2014). For gene set simulations, subregion length was defined as the sum of individual LSD gene coding region lengths (169.5 kb or 170.4 in IPDGC and PPMI, respectively). For single gene simulations, the average gene length was used (3.5kb or 3.2 kb, respectively). The MAF cutoff for causal variants was set to 0.00035 (based on the frequency of rare *GBA* loss-of-function alleles in the IPDGC data set) or 0.03 for the rare or more common variant models, respectively, and penetrance was assumed to be either 100% or 10%. Because we predict that LSD gene variants associated with Parkinson's disease will have a damaging effect, all causal variants were assumed to have a positive coefficient (risk rather than protective alleles).

Results

1   Variants were extracted from 54 genes responsible for LSDs, defined based on widely accepted

2   criteria (Table 1), and filtered into nested categories based on 2 frequency thresholds and 3 tiers

3   of functional criteria (Figure 1A). Our overall analytic approach is diagrammed in Figure 1B. To

4   test our hypothesis that an aggregate burden of variants in the LSD gene set contributes to

5   Parkinson's disease, we first implemented SKAT-O within the IPDGC WES discovery cohort

6   (Table 2). Following adjustment for multiple comparisons (see Methods), significant associations

7   were detected for the LSD gene set considering either all non-synonymous variants (category 1b,

8   p_{adj}=0.014) or likely damaging variants (category 2b, p_{adj}=0.0055), when using the more relaxed

9   frequency threshold of MAF < 3%. When considering only the subset of rare (MAF < 1%)

10   nonsynonymous or likely damaging variants, the SKAT-O result was attenuated and no longer

11   significant (category 1a, p_{adj}=0.056 and category 2a, 0.066, respectively). No association was

12   observed when considering only loss-of-function alleles (category 3, p_{adj}=0.464), possibly due to

13   the relative paucity of such variants limiting statistical power (Supplemental Table 1). We next

14   repeated analyses with significant results, but excluding all *GBA* variants. As expected, the

15   strength of the associations was attenuated; however, both SKAT-O results including either all

16   nonsynonymous variants (MAF<3%) or the subset of likely damaging variants was robust to the

17   exclusion of *GBA* and remained significant (category 1b, p_{adj}=0.026 and category 2b,

18   p_{adj}=0.0198). Our results indicate that the association between variant burden and Parkinson's

19   disease risk in the IPDGC discovery cohort is mediated, at least in part, by the effects of LSD

20   genes other than *GBA,* an established Parkinson's disease susceptibility locus.

21   To replicate our findings, we leveraged two independent cohorts, including an additional

22   WES dataset from PPMI (436 Parkinson's disease cases and 169 controls) (Parkinson

23   Progression Marker Initiative, 2011) and the NeuroX exome-wide genotyping dataset from

1   IPDGC (6,713 Parkinson's disease cases and 5,964 controls) (Nalls *et al.*, 2015). We again

2   implemented SKAT-O to detect a potential variant burden in Parkinson's disease cases versus

3   controls. In the smaller PPMI replication cohort, we discovered suggestive evidence for an

4   excessive LSD variant burden in Parkinson's disease (Table 2); however, this finding was not

5   significant following adjustment for multiple comparisons (category 1a, p_{adj}=0.096). The

6   association signal—which appeared independent of *GBA* (Supplemental Table 5)—was detected

7   exclusively among rare alleles (MAF < 1%) and only when considering all non-synonymous

8   variants. It is possible that SKAT-O is sensitive to cohort differences between PPMI and the

9   IPDGC, including both sample size and pertinent demographic features (e.g. age of onset and

10   family history; Supplemental Table 1). However, in the substantially larger NeuroX dataset,

11   significant burden associations were detected for the same 2 variant categories implicated by

12   SKAT-O in the IPDGC discovery cohort (Table 2), despite the less comprehensive genotyping

13   coverage compared to WES. A major driver for the robust LSD gene set association in NeuroX

14   (category 1b, p_{adj}=0.0004 and category 2b, p_{adj}=0.0003) appears to be the more common

15   GBA^{E326K} variant (Freq$_{Cases}$=0.021, Freq$_{Controls}$=0.011), which has been reported to be associated

16   with Parkinson's disease risk in several large studies (Duran *et al.*, 2012; Pankratz *et al.*, 2012).

17   Importantly, consistent with our findings in the IPDGC discovery cohort, the LSD gene set

18   burden association for both of these variant categories remained significant in NeuroX following

19   exclusion of *GBA* (category 1b, p_{adj}=0.002 and category 2b, p_{adj}=0.020). When considering only

20   the subset of rare (MAF < 1%) variants in the NeuroX dataset, the SKAT-O result for the LSD

21   gene set was attenuated and no longer significant; although, the association in the

22   nonsynonymous variant group remained suggestive, and this association was independent of

23   *GBA* (Supplemental Table 5). In sum, based on analyses in three independent Parkinson's

1   disease case-control datasets, we demonstrate a burden of variants in LSD genes associated with

2   Parkinson's disease risk, and this signal is at least partially independent of *GBA*.

3   To determine which additional LSD genes/variants may be responsible for the observed

4   association with Parkinson's disease risk, we performed exploratory analyses using SKAT-O to

5   assess for potential contribution of variants within each gene considered independently. For these

6   analyses, we returned to the IPDGC discovery dataset, and again focused on likely damaging

7   variants, which showed the strongest association signal in our primary analysis (category 2b). In

8   these gene-based analyses, besides the expected result for *GBA* ($p = 0.0001$) and confirmation of

9   *SMPD1* ($p = 0.029$), we discover evidence of novel aggregate associations for variants in *CTSD*

10   ($p = 0.002$), *SLC17A5* ($p = 0.005$), and *ASAH1* ($p = 0.031$). The specific variants implicated for

11   each of these genes are included in Supplemental Table 3, along with all other putative damaging

12   variants considered in our full LSD gene set analysis. While our datasets are underpowered to

13   definitively assess the contributions of a particular rare variant in any single gene (see

14   Discussion), these results identify the most likely specific loci driving the aggregate LSD gene

15   set association signal detected in the IPDGC discovery sample.

16   Lastly, we examined the distribution of putative damaging LSD gene variants (MAF <

17   3%, category 2b) within the IPDGC WES cohort (Figure 2). Consistent with our finding of an

18   excessive variant burden in Parkinson's disease, the distribution of variants appeared modestly

19   right-skewed in cases. The average variant burden among IPDGC cases was 0.9 alleles per

20   individual, which was slightly higher than that seen in controls (0.8 alleles per individual). Given

21   their commonality, the majority of IPDGC cases (56%) have at least one putative damaging

22   variant in an LSD gene, and 21% carry multiple alleles. Notably, only 22 out of 1156 total

23   Parkinson's disease cases are homo- or hemizygous for putative damaging LSD variants

1 (Supplemental Table 6), suggesting that Mendelian recessive or X-linked inheritance may

2 contribute minimally to the overall burden association. As discussed further below, our findings

3 are consistent with a hypothetical model in which multiple LSD gene variants may interact to

4 influence Parkinson's disease risk.

5

6 **Discussion**

7 This study reveals an important connection between the genetic factors broadly responsible for

8 LSDs, which are predominantly pediatric Mendelian disorders, and Parkinson's disease, an

9 adult-onset neurodegenerative disorder with complex genetic etiology. Specifically, among 54

10 genes that cause LSDs, we find evidence for a burden of damaging alleles in association with

11 Parkinson's disease risk. This association persisted after excluding *GBA*, consistent with a

12 contribution from additional LSD genes. More than half of Parkinson's disease cases in our

13 cohort harbor one or more putative damaging variants among the LSD genes. Thus, our results

14 implicate several promising new Parkinson's disease susceptibility loci and reinforce the

15 importance of lysosomal mechanisms in Parkinson's disease pathogenesis.

16 The strengths of this study include a large Parkinson's disease case/control discovery

17 cohort as well as two independent datasets for replication of our findings. The IPDGC WES

18 discovery sample is characterized by younger-onset Parkinson's disease cases (mean age~41

19 years) and those with a positive family history, thereby enriching for individuals with a potential

20 genetic contribution. Recruitment of a substantially older IPDGC control group (mean age~64),

21 reduces the possibility of latent, unrecognized Parkinson's disease (i.e. with minimal or absent

22 symptoms), likely further increasing power for genetic discovery. By contrast, our PPMI and

23 NeuroX replication cohorts include older cases (mean age~62 years) and age-matched controls,

1   making them more broadly representative of the older adult population commonly affected by

2   Parkinson's disease. Consistent findings of an excessive LSD variant burden across these 3

3   datasets, especially the large NeuroX sample (n~12,677), strongly enhances the generalizability

4   of our conclusions. To minimize the possibility of population stratification, stringent quality

5   control filters were implemented to ensure a homogeneous European ancestry sample in all study

6   cohorts (Supplemental Figure 1). Nevertheless, it will also be important to examine other ethnic

7   populations in the future, especially those potentially enriched for LSD-causing variants due to

8   genetic bottlenecks.

9   Since our understanding of the characteristics of causal alleles—including in both

10   Parkinson's disease and LSDs—is incomplete, our initial analyses systematically considered

11   multiple variant classes binned into categories based on frequency and putative functional

12   impact. In the IPDGC and PPMI cohorts, WES offers comprehensive characterization of LSD

13   gene variants. By contrast, since the NeuroX data is restricted to those variants included on the

14   genotyping array, it is possible that many potential pathogenic variants would be missed.

15   Nevertheless, a total of 348 putative damaging variants were detected, including alleles for all

16   LSD genes (Supplemental Table 2). Importantly, the selected analytic tool, SKAT-O, is robust to

17   a wide frequency spectrum, including rare and more common alleles, and to variants with

18   different magnitudes and directions of effect (S. Lee *et al.*, 2012; 2016). Our results suggest that

19   consideration of likely damaging alleles based on bioinformatic predictions, including more

20   common LSD variants (MAF < 3%), appeared to offer optimal sensitivity for detection of a

21   significant aggregate variant association. Many of these variants are known to be pathogenic for

22   LSDs (Supplemental Table 3). For example, of the *GBA* variants considered in our analyses,

23   27% of those with annotations available in ClinVar (Landrum *et al.*, 2016) are rated as likely or

1   definitively pathogenic. Critically, the implementation of burden association tests for joint

2   consideration of LSD genes significantly improves statistical power over single gene and variant

3   tests (Zuk *et al.*, 2014). In populations of European ancestry similar to our study cohorts, loss-of-

4   function alleles, including those established to cause LSDs, are individually rare (Supplemental

5   Table 2), and based on post-hoc simulations (see Methods), we estimate poor power for

6   discovery of rare Parkinson's disease risk alleles at single loci. For example, assuming a rare

7   variant model (MAF = 0.035%, as for *GBA* loss-of-function alleles in our sample) and even

8   assuming full penetrance, the IPDGC discovery cohort has only 30% power to discover an

9   association for a single gene. However, a similar simulation considering the full set of 54 LSD

10 genes was fully powered (100%). Our consideration of higher frequency variants further

11 enhances power for both discovery and replication, especially when coupled with filtering based

12 on potential pathogenicity. For example, allowing for more common variants (MAF < 3%) and

13 assuming 10% of such alleles are causal, we estimate that the smaller PPMI cohort achieves 95%

14 power for replication of a gene set association, whereas negligible power (1%) is available for

15 interrogation of a single gene candidate. We anticipate that larger WES datasets will significantly

16 improve power, including for per gene analyses.

17   We also performed analyses in the IPDGC cohort to pinpoint the specific drivers from the

18 LSD gene set responsible for increasing Parkinson's disease risk. Our results (i) recapitulate the

19 established association with *GBA*, (ii) strengthen the emerging evidence in support of *SMPD1,*

20 and (iii) newly implicate *SLC17A5, ASAH1*, and *CTSD* as candidate Parkinson's disease

21 susceptibility genes. Recessive mutations in *SMPD1* cause Niemann-Pick type A/B disease and

22 this locus has been independently implicated in Parkinson's disease risk based on several

23 published studies (Clark *et al.*, 2015; Foo *et al.*, 2013; Gan-Or *et al.*, 2013; 2015; S. Lee *et al.*,

1   2012). While our analysis identified 21 candidate, putative damaging *SMPD1* risk alleles

2   (Supplemental Table 3), most appear distinct from those reported in other studies of Parkinson's

3   disease. One notable exception, *SMPD1p.L304P* (also referred to as p.L302P), was previously

4   implicated in a study of Ashkenazi Jewish subjects (Gan-Or *et al.*, 2013). Another non-

5   synonymous variant, p.P332L implicated in the IPDGC sample is at the same amino acid

6   position as a different substitution, p.P332R, that was previously implicated in a Chinese

7   Parkinson's disease cohort (Foo *et al.*, 2013). Among the novel candidate genes, *SLC17A5,*

8   *ASAH1*, and *CTSD*, most of the implicated variants are rare (MAF<1%). Only 2 of these variants

9   (*rs16883930* and *rs141068211* in *SLC17A5* and *ASAH1*, respectively) are present in the 1000

10   Genomes reference (The 1000 Genomes Project Consortium, 2012), having been previously

11   examined in genome-wide scans, and both were non-associated with Parkinson's disease risk

12   (p>0.05) based on available data (Lill *et al.*, 2012). Mutations in *SLC17A5*, *ASAH1*, and *CTSD*

13   cause the rare LSDs, Salla disease, Farber Lipogranulomatosis, and Neuronal Ceroid

14   Lipofuscinosis (CLN10), respectively. Whereas Sialin (the protein product of *SLC17A5*) is a

15   lysosomal membrane transporter for sialic acid, Acid Ceramidase (*ASAH1*) participates in

16   ceramide metabolism, similar to Glucocerebrosidase and Sphingomyelinase (*SMPD1*). In

17   addition to promoting lysosomal stress, glucosylceramide, which accumulates in Gaucher

18   disease, has been suggested to directly promote the aggregation of alpha-synuclein (Mazzulli *et*

19   *al.*, 2011; Moors *et al.*, 2016). Interestingly, *CTSD* encodes a lysosomal aspartyl proteinase

20   which has been independently implicated in alpha-synuclein degradation (Cullen *et al.*, 2009;

21   McGlinchey and J. C. Lee, 2015). In sum, the LSD genes and variants implicated by our studies

22   are excellent candidates for further replication, including resequencing and/or genotyping in the

23   largest available Parkinson's disease case/control samples. Although we employed standard

1   quality control procedures for calling variants from WES and genotyping data, definitive

2   confirmation of specific variants will require additional studies.

3   There is a growing recognition of the importance of lysosomal biology in Parkinson's

4   disease pathogenesis (Moors *et al.*, 2016; Wong and Krainc, 2016). First, the lysosome is an

5   important route for alpha-synuclein degradation (Cuervo *et al.*, 2004; H. J. Lee, 2004; Vogiatzi

6   *et al.*, 2008). Genomic variants that elevate alpha-synuclein protein levels—such as rare locus

7   multiplication (Singleton *et al.*, 2003) or a common polymorphism that enhances promoter

8   activity (Soldner *et al.*, 2016)—also increase Parkinson's disease risk. Knockdown of selected

9   LSD genes, including *GBA* or *SCARB2*, in neuronal cells or in mouse models impairs alpha-

10   synuclein clearance (Cooper *et al.*, 2006; Rothaug *et al.*, 2014; Sardi *et al.*, 2011), whereas

11   increasing glucocerebrosidase activity has the opposite effect (Mazzulli, Zunke, Tsunemi, *et al.*,

12   2016; Migdalska-Richards *et al.*, 2016; Sardi *et al.*, 2011). Second, lysosomal autophagy plays a

13   critical role in mitochondrial quality control, and substantial evidence, including from genetics,

14   highlight mitochondrial dysfunction in Parkinson's disease (Haelterman *et al.*, 2014). Third,

15   there is accumulating evidence from numerous experimental models that alpha-synuclein

16   interferes with endoplasmic reticulum-to-Golgi vesicle trafficking, inducing reciprocal

17   disruptions in lysosomal biogenesis (Cooper *et al.*, 2006). Expression of alpha-synuclein

18   impeded trafficking of multiple hydrolases linked to LSDs, including GBA, within human

19   dopaminergic neurons (Mazzulli, Zunke, Isacson, *et al.*, 2016). In one recent study, subjects with

20   idiopathic Parkinson's disease, in which *GBA* carriers were excluded, were found to have modest

21   but significantly reduced glucocerebrosidase enzymatic activity based on peripheral blood testing

22   (Alcalay *et al.*, 2015). Fourth, besides *GBA* and the other genes implicated in our study,

23   mutations in *ATP13A2,* a rare cause of recessive juvenile-onset parkinsonism and dementia has

1 been independently implicated to cause the LSD Neuronal Ceroid Lipofuscinosis (Bras *et al.*,

2 2012). Lastly, many other common and rare Parkinson's disease risk alleles, including at

3 *RAB7L1*, *GAK, LRRK2,* and *VPS35* have strong functional links to vesicle trafficking, including

4 for lysosomal biogenesis and function. Together, these findings support a model in which partial

5 loss-of-function in genes regulating lysosomal activity, such as those that cause LSDs, may

6 increase vulnerability to alpha-synuclein-mediated mechanisms in Parkinson's disease.

7   While our analyses reveal a robust and replicable LSD variant burden in Parkinson's

8 disease cases, the overall magnitude of the difference between variant frequencies in cases and

9 controls appears modest (Figure 2). We speculate that this is probably an underestimate of the

10 true difference due to several assumptions. Specifically, only a subset of the 54 LSD genes and

11 760 nonsynonymous variants considered in our burden analyses are likely to be truly involved in

12 Parkinson's disease risk. Further, as noted above, while the CADD framework allowed us to

13 prioritize 596 variants as putative damaging alleles, larger Parkinson's disease exome datasets

14 with improved statistical power will be required to resolve the specific LSD genes and variants

15 that contribute to Parkinson's disease risk. Lastly, similar to *GBA* (Anheim *et al.* 2012; Rana *et*

16 *al.* 2013), we expect that many of the other LSD gene variants contributing to Parkinson's

17 disease risk may have individually modest and therefore incompletely penetrant effects, perhaps

18 modified by alleles at other loci (Cooper *et al.* 2013). In sum, the likely (i) incomplete

19 penetrance of many pathogenic variants along with (ii) contamination of our analyses by many

20 benign variants would be expected to inflate estimates for the LSD variant burden among

21 controls and attenuate the overall SKAT-O association.

22   Parkinson's disease heritability remains incompletely explained by the genes and variants

23 identified to date (Do *et al.*, 2011; Hamza and Paymi, 2010; Keller *et al.*, 2012; Verstraeten *et*

1 *al.*, 2015). Besides the likelihood of yet undiscovered loci, alternative explanations for familial

2 aggregation of disease include epigenetic changes due to shared environmental exposures or

3 even false positive diagnoses due to phenocopies (Mullin and Schapira 2015; Pihlstrom 2011). In

4 complex genetic disorders such as Parkinson's disease, the cumulative impact of common and

5 rare variants at multiple genomic loci, as well as non-additive interactions among alleles, likely

6 also play an important role (Cooper *et al.* 2013; Lupski *et al.* 2011). Polygenic modeling

7 approaches have previously demonstrated how common risk alleles can cumulatively impact

8 Parkinson's disease risk and age-of-onset (Escott-Price *et al.*, 2015; Nalls *et al.*, 2014). In

9 addition, a recently published analysis in the IPDGC WES and NeuroX cohorts identified

10 evidence for oligogenic interactions underlying Parkinson's disease risk, including alleles for

11 *GBA* and those for established Mendelian Parkinson's disease genes (Lubbe *et al.*, 2016). In the

12 IPDGC, WES reveals a substantial proportion of Parkinson's disease cases (21%) carrying two

13 or more likely damaging variants in LSD genes. Consistent with other reports (Clark *et al.*,

14 2015), our observation suggests the possibility that multiple LSD gene variants may interact in a

15 multi-hit, combinatorial manner to degrade lysosomal function, causing the accumulation of

16 alpha-synuclein and potentially other toxic substrates, and increasing susceptibility for

17 Parkinson's disease. Oligogenic interactions such as those proposed here may be an important

18 source for "missing heritability" in Parkinson's disease (Mullin and Schapira 2015; Pihlstrom

19 2011). Recent work has also implicated oligogenic inheritance in other neurologic disorders,

20 including amyotrophic lateral sclerosis (Cady *et al.*, 2015; Kenna *et al.*, 2013; van Blitterswijk *et*

21 *al.*, 2012) and idiopathic peripheral neuropathy (Gonzaga-Jauregui *et al.*, 2015), and further

22 reveals how pleiotropic genes causing early-onset, monogenic disorders may act in combination

23 to additionally trigger late-onset, complex genetic disorders (Cooper *et al.* 2013; Lupski *et al.*

2011). Future studies, including even-larger, case-control cohorts with WES and complementary experiments in Parkinson's disease cellular or animal models, are needed to further investigate whether a variant burden in LSD genes, perhaps in combination with other susceptibility loci, underlies oligogenic risk and contributes substantially to Parkinson's disease heritability.

Acknowledgements

Funding

Supplementary Material

Supplementary material, including 2 Figures and 6 Tables, is available at Brain online.

Appendix

IPDGC Consortium Members and Affiliations

Mike A Nalls (Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA), Vincent Plagnol (UCL Genetics Institute, London, UK), Dena G Hernandez (Laboratory of Neurogenetics, National Institute on Aging; and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Manu Sharma (Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tubingen and Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen Germany), Una-Marie Sheerin (Department of Molecular Neuroscience, UCL Institute of Neurology), Mohamad Saad (INSERM U563, CPTP, Toulouse, France; and Paul Sabatier University, Toulouse, France), Javier Simón-Sánchez (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany), Claudia Schulte (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research), Suzanne Lesage (INSERM, UMR_S975 [formerly UMR_S679], Paris, France;

1   Université Pierre et Marie Curie-Paris, Centre de Recherche de l'Institut du Cerveau et de la

2   Moelle épinière, Paris, France; and CNRS, Paris, France), Sigurlaug Sveinbjörnsdóttir

3   (Department of Neurology, Landspítali University Hospital, Reykjavík, Iceland; Department of

4   Neurology, MEHT Broomfield Hospital, Chelmsford, Essex, UK; and Queen Mary College,

5   University of London, London, UK), Sampath Arepalli (Laboratory of Neurogenetics, National

6   Institute on Aging), Roger Barker (Department of Neurology, Addenbrooke's Hospital,

7   University of Cambridge, Cambridge, UK), Yoav Ben-Shlomo (School of Social and

8   Community Medicine, University of Bristol), Henk W Berendse (Department of Neurology and

9   Alzheimer Center, VU University Medical Center), Daniela Berg (Department for

10   Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research and DZNE, German

11   Center for Neurodegenerative diseases), Kailash Bhatia (Department of Motor Neuroscience,

12   UCL Institute of Neurology), Rob M A de Bie (Department of Neurology, Academic Medical

13   Center, University of Amsterdam, Amsterdam, Netherlands), Alessandro Biffi (Center for

14   Human Genetic Research and Department of Neurology, Massachusetts General Hospital,

15   Boston, MA, USA; and Program in Medical and Population Genetics, Broad Institute,

16   Cambridge, MA, USA), Bas Bloem (Department of Neurology, Radboud University Nijmegen

17   Medical Centre, Nijmegen, Netherlands), Zoltan Bochdanovits (Department of Clinical Genetics,

18   Section of Medical Genomics, VU University Medical Centre), Michael Bonin (Department of

19   Medical Genetics, Institute of Human Genetics, University of Tübingen, Tübingen, Germany),

20   Jose M Bras (Department of Molecular Neuroscience, UCL Institute of Neurology), Kathrin

21   Brockmann (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain

22   Research and DZNE, German Center for Neurodegenerative diseases), Janet Brooks (Laboratory

23   of Neurogenetics, National Institute on Aging), David J Burn (Newcastle University Clinical

Ageing Research Unit, Campus for Ageing and Vitality, Newcastle upon Tyne, UK), Elisa

Majounie (Laboratory of Neurogenetics, National Institute on Aging), Gavin Charlesworth

(Department of Molecular Neuroscience, UCL Institute of Neurology), Codrin Lungu (National

Institutes of Health Parkinson Clinic, NINDS, National Institutes of Health), Honglei Chen

(Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes

of Health, NC, USA), Patrick F Chinnery (Neurology M4104, The Medical School, Framlington

Place, Newcastle upon Tyne, UK), Sean Chong (Laboratory of Neurogenetics, National Institute

on Aging), Carl E Clarke (School of Clinical and Experimental Medicine, University of

Birmingham, Birmingham, UK; and Department of Neurology, City Hospital, Sandwell and

West Birmingham Hospitals NHS Trust, Birmingham, UK), Mark R Cookson (Laboratory of

Neurogenetics, National Institute on Aging), J Mark Cooper (Department of Clinical

Neurosciences, UCL Institute of Neurology), Jean Christophe Corvol (INSERM, UMR_S975;

Université Pierre et Marie Curie-Paris; CNRS; and INSERM CIC-9503, Hôpital Pitié-

Salpêtrière, Paris, France), Carl Counsell (University of Aberdeen, Division of Applied Health

Sciences, Population Health Section, Aberdeen, UK), Philippe Damier (CHU Nantes, CIC0004,

Service de Neurologie, Nantes, France), Jean-François Dartigues (INSERM U897, Université

Victor Segalen, Bordeaux, France), Panos Deloukas (Wellcome Trust Sanger Institute,

Wellcome Trust Genome Campus, Cambridge, UK), Günther Deuschl (Klinik für Neurologie,

Universitätsklinikum Schleswig-Holstein, Campus Kiel, Christian-Albrechts-Universität Kiel,

Kiel, Germany), David T Dexter (Parkinson's Disease Research Group, Faculty of Medicine,

Imperial College London, London, UK), Karin D van Dijk (Department of Neurology and

Alzheimer Center, VU University Medical Center), Allissa Dillman (Laboratory of

Neurogenetics, National Institute on Aging), Frank Durif (Service de Neurologie, Hôpital

1 Gabriel Montpied, Clermont-Ferrand, France), Alexandra Dürr (INSERM, UMR_S975;

2 Université Pierre et Marie Curie-Paris; CNRS; and AP-HP, Pitié-Salpêtrière Hospital), Sarah

3 Edkins (Wellcome Trust Sanger Institute), Jonathan R Evans (Cambridge Centre for Brain

4 Repair, Cambridge, UK), Thomas Foltynie (UCL Institute of Neurology), Jing Dong

5 (Epidemiology Branch, National Institute of Environmental Health Sciences), Michelle Gardner

6 (Department of Molecular Neuroscience, UCL Institute of Neurology), J Raphael Gibbs

7 (Laboratory of Neurogenetics, National Institute on Aging; and Department of Molecular

8 Neuroscience, UCL Institute of Neurology), Alison Goate (Department of Psychiatry,

9 Department of Neurology, Washington University School of Medicine, MI, USA), Emma Gray

10 (Wellcome Trust Sanger Institute), Rita Guerreiro (Department of Molecular Neuroscience, UCL

11 Institute of Neurology), Clare Harris (University of Aberdeen), Jacobus J van Hilten

12 (Department of Neurology, Leiden University Medical Center, Leiden, Netherlands), Albert

13 Hofman (Department of Epidemiology, Erasmus University Medical Center, Rotterdam,

14 Netherlands), Albert Hollenbeck (AARP, Washington DC, USA), Janice Holton (Queen Square

15 Brain Bank for Neurological Disorders, UCL Institute of Neurology), Michele Hu (Department

16 of Clinical Neurology, John Radcliffe Hospital, Oxford, UK), Xuemei Huang (Departments of

17 Neurology, Radiology, Neurosurgery, Pharmacology, Kinesiology, and Bioengineering,

18 Pennsylvania State University– Milton S Hershey Medical Center, Hershey, PA, USA), Isabel

19 Wurster (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain

20 Research and German Center for Neurodegenerative diseases), Walter Mätzler (Department for

21 Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research and German Center for

22 Neurodegenerative diseases), Gavin Hudson (Neurology M4104, The Medical School,

23 Newcastle upon Tyne, UK), Sarah E Hunt (Wellcome Trust Sanger Institute), Johanna

30

1 Huttenlocher (deCODE genetics), Thomas Illig (Institute of Epidemiology, Helmholtz Zentrum

2 München, German Research Centre for Environmental Health, Neuherberg, Germany), Pálmi V

3 Jónsson (Department of Geriatrics, Landspítali University Hospital, Reykjavík, Iceland), Jean-

4 Charles Lambert (INSERM U744, Lille, France; and Institut Pasteur de Lille, Université de Lille

5 Nord, Lille, France), Cordelia Langford (Cambridge Centre for Brain Repair), Andrew Lees

6 (Queen Square Brain Bank for Neurological Disorders), Peter Lichtner (Institute of Human

7 Genetics, Helmholtz Zentrum München, German Research Centre for Environmental Health,

8 Neuherberg, Germany), Patricia Limousin (Institute of Neurology, Sobell Department, Unit of

9 Functional Neurosurgery, London, UK), Grisel Lopez (Section on Molecular Neurogenetics,

10 Medical Genetics Branch, NHGRI, National Institutes of Health), Delia Lorenz (Klinik für

11 Neurologie, Universitätsklinikum Schleswig-Holstein), Codrin Lungu (National Institutes of

12 Health Parkinson Clinic, NINDS, National Institutes of Health), Alisdair McNeill (Department

13 of Clinical Neurosciences, UCL Institute of Neurology), Catriona Moorby (School of Clinical

14 and Experimental Medicine, University of Birmingham), Matthew Moore (Laboratory of

15 Neurogenetics, National Institute on Aging), Huw R Morris (National Hospital for Neurology

16 and Neurosurgery, University College London, London, UK), Karen E Morrison (School of

17 Clinical and Experimental Medicine, University of Birmingham; and Neurosciences Department,

18 Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust,

19 Birmingham, UK), Valentina Escott-Price (MRC Centre for Neuropsychiatric Genetics and

20 Genomics, Cardiff University School of Medicine, Cardiff, UK), Ese Mudanohwo

21 (Neurogenetics Unit, UCL Institute of Neurology and National Hospital for Neurology and

22 Neurosurgery), Sean S O'Sullivan (Queen Square Brain Bank for Neurological Disorders), Justin

23 Pearson (MRC Centre for Neuropsychiatric Genetics and Genomics), Joel S Perlmutter

1 (Department of Neurology, Radiology, and Neurobiology at Washington University, St Louis),

2 Hjörvar Pétursson (deCODE genetics; and Department of Medical Genetics, Institute of Human

3 Genetics, University of Tübingen), Pierre Pollak (Service de Neurologie, CHU de Grenoble,

4 Grenoble, France), Bart Post (Department of Neurology, Radboud University Nijmegen Medical

5 Centre), Simon Potter (Wellcome Trust Sanger Institute), Bernard Ravina (Translational

6 Neurology, Biogen Idec, MA, USA), Tamas Revesz (Queen Square Brain Bank for Neurological

7 Disorders), Olaf Riess (Department of Medical Genetics, Institute of Human Genetics,

8 University of Tübingen), Fernando Rivadeneira (Departments of Epidemiology and Internal

9 Medicine, Erasmus University Medical Center), Patrizia Rizzu (Department of Clinical Genetics,

10 Section of Medical Genomics, VU University Medical Centre), Mina Ryten (Department of

11 Molecular Neuroscience, UCL Institute of Neurology), Stephen Sawcer (University of

12 Cambridge, Department of Clinical Neurosciences, Addenbrooke's hospital, Cambridge, UK),

13 Anthony Schapira (Department of Clinical Neurosciences, UCL Institute of Neurology), Hans

14 Scheffer (Department of Human Genetics, Radboud University Nijmegen Medical Centre,

15 Nijmegen, Netherlands), Karen Shaw (Queen Square Brain Bank for Neurological Disorders),

16 Ira Shoulson (Department of Neurology, University of Rochester, Rochester, NY, USA), Joshua

17 Shulman (Baylor College of Medicine, Houston, Texas), Ellen Sidransky (Section on Molecular

18 Neurogenetics, Medical Genetics Branch, NHGRI), Colin Smith (Department of Pathology,

19 University of Edinburgh, Edinburgh, UK), Chris C A Spencer (Wellcome Trust Centre for

20 Human Genetics, Oxford, UK), Hreinn Stefánsson (deCODE genetics), Francesco Bettella

21 (deCODE genetics), Joanna D Stockton (School of Clinical and Experimental Medicine), Amy

22 Strange (Wellcome Trust Centre for Human Genetics), Kevin Talbot (University of Oxford,

23 Department of Clinical Neurology, John Radcliffe Hospital, Oxford, UK), Carlie M Tanner

(Clinical Research Department, The Parkinson's Institute and Clinical Center, Sunnyvale, CA, USA), Avazeh Tashakkori-Ghanbaria (Wellcome Trust Sanger Institute), François Tison (Service de Neurologie, Hôpital Haut-Lévêque, Pessac, France), Daniah Trabzuni (Department of Molecular Neuroscience, UCL Institute of Neurology), Bryan J Traynor (Laboratory of Neurogenetics, National Institute on Aging), André G Uitterlinden (Departments of Epidemiology and Internal Medicine, Erasmus University Medical Center), Daan Velseboer (Department of Neurology, Academic Medical Center), Marie Vidailhet (INSERM, UMR_S975, Université Pierre et Marie Curie-Paris, CNRS, UMR 7225), Robert Walker (Department of Pathology, University of Edinburgh), Bart van de Warrenburg (Department of Neurology, Radboud University Nijmegen Medical Centre), Mirdhu Wickremaratchi (Department of Neurology, Cardiff University, Cardiff, UK), Nigel Williams (MRC Centre for Neuropsychiatric Genetics and Genomics), Caroline H Williams-Gray (Department of Neurology, Addenbrooke's Hospital), Sophie Winder-Rhodes (Department of Psychiatry and Medical Research Council and Wellcome Trust Behavioural and Clinical Neurosciences Institute, University of Cambridge), Kári Stefánsson (deCODE genetics), Maria Martinez (INSERM UMR 1043; and Paul Sabatier University), Nicholas W Wood (UCL Genetics Institute; and Department of Molecular Neuroscience, UCL Institute of Neurology), John Hardy (Department of Molecular Neuroscience, UCL Institute of Neurology), Peter Heutink (DZNE, German Center for Neurodegenerative Diseases and Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany), Alexis Brice (INSERM, UMR_S975, Université Pierre et Marie Curie-Paris, CNRS, UMR 7225, AP-HP, Pitié-Salpêtrière Hospital), Thomas Gasser (Department for Neurodegenerative Diseases, Hertie

Institute for Clinical Brain Research, and DZNE, German Center for Neurodegenerative

Diseases), Andrew B Singleton (Laboratory of Neurogenetics, National Institute on Aging).

References

The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 2012; 491: 56-65.

Aharon-Peretz J, Rosenbaum H, Gershoni-Baruch R. Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N Engl J Med 2004; 351: 1972–1977.

Alcalay RN, Levy OA, Waters CC, Fahn S, Ford B, Kuo S-H, et al. Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations. Brain 2015; 138: 2648–2658.

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Research 2015; 43: D789–98.

Amendola LM, Dorschner MO, Robertson PD, Salama JS, Hart R, Shirts BH, et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res. 2015; 25: 305–315.

Anheim M, Elbaz A, Lesage S, Durr A, Condroyer C, Viallet F, et al. Penetrance of Parkinson disease in glucocerebrosidase gene mutation carriers. Neurology 2012; 78: 417–420.

Boustany R-MN. Lysosomal storage diseases—the horizon expands. Nature Reviews Neurology 2013; 9: 583–598.

Bras J, Guerreiro R, Hardy J. SnapShot: Genetics of Parkinson's disease. Cell 2015; 160: 570–570.e1.

Bras J, Verloes A, Schneider SA, Mole SE, Guerreiro RJ. Mutation of the parkinsonism gene ATP13A2 causes neuronal ceroid-lipofuscinosis. Hum. Mol. Genet. 2012; 21: 2646–2650.

Brockmann K, Srulijes K, Pflederer S, Hauser A-K, Schulte C, Maetzler W, et al. GBA-associated Parkinson's disease: reduced survival and more rapid progression in a prospective longitudinal study. Mov Disord. 2015; 30: 407–411.

Cady J, Allred P, Bali T, Pestronk A, Goate A, Miller TM, et al. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. Ann. Neurol. 2015; 77: 100–113.

Clark LN, Chan R, Cheng R, Liu X, Park N, Parmalee N, et al. Gene-wise association of variants in four lysosomal storage disorder genes in neuropathologically confirmed Lewy body disease. PLoS ONE 2015; 10: e0125204.

1   Clark LN, Ross BM, Wang Y, Mejia-Santana H, Harris J, Louis ED, et al. Mutations in the
2   glucocerebrosidase gene are associated with early-onset Parkinson disease. Neurology 2007; 69:
3   1270–1277.

4   Cooper AA, Gitler AD, Cashikar A, Haynes CM, Hill KJ, Bhullar B, et al. Alpha-synuclein
5   blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. Science 2006; 313:
6   324–328.

7   Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., & Kehrer-Sawatzki, H.
8   Where genotype is not predictive of phenotype: towards an understanding of the molecular basis
9   of reduced penetrance in human inherited disease. Humangenetik 2013; 132: 1077–1130.

10 Cuervo AM, Stefanis L, Fredenburg R, Lansbury PT, Sulzer D. Impaired degradation of mutant
11 alpha-synuclein by chaperone-mediated autophagy. Science 2004; 305: 1292–1295.

12 Cullen V, Lindfors M, Ng J, Paetau A, Swinton E, Kolodziej P, et al. Cathepsin D expression
13 level affects alpha-synuclein processing, aggregation, and toxicity in vivo. Mol Brain 2009; 2: 5.

14 Davis MY, Johnson CO, Leverenz JB, Weintraub D, Trojanowski JQ, Chen-Plotkin A, et al.
15 Association of GBA Mutations and the E326K Polymorphism With Motor and Cognitive
16 Progression in Parkinson Disease. JAMA Neurol 2016; 73: 1217–1224.

17 de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. The Lancet Neurology 2006; 5:
18 525–535.

19 Deng H, Xiu X, Jankovic J. Genetic convergence of Parkinson's disease and lysosomal storage
20 disorders. Mol. Neurobiol. 2015; 51: 1554–1568.

21 Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, et al. Web-Based Genome-
22 Wide Association Study Identifies Two Novel Loci and a Substantial Genetic Component for
23 Parkinson's Disease. PLoS Genetics 2011; 7: e1002141.

24 Duran R, Mencacci NE, Angeli AV, Shoai M, Deas E, Houlden H, et al. The glucocerobrosidase
25 E326K variant predisposes to Parkinson's disease, but does not cause Gaucher's disease. Mov
26 Disord. 2012; 28: 232–236.

27 Escott-Price V, International Parkinson's Disease Genomics Consortium, Nalls MA, Morris HR,
28 Lubbe S, Brice A, et al. Polygenic risk of Parkinson disease is correlated with disease age at
29 onset. Ann. Neurol. 2015; 77: 582–591.

30 Filocamo M, Morrone A. Lysosomal storage disorders: molecular basis and laboratory testing.
31 Hum. Genomics 2011; 5: 156–169.

32 Foo J-N, Liany H, Bei J-X, Yu X-Q, Liu J, Au W-L, et al. Rare lysosomal enzyme gene SMPD1
33 variant (p.R591C) associates with Parkinson's disease. Neurobiol. Aging 2013; 34: 2890.e13–5.

34 Gan-Or Z, Orr-Urtreger A, Alcalay RN, Bressman S, Giladi N, Rouleau GA. The emerging role
35 of SMPD1 mutations in Parkinson's disease: Implications for future studies. Parkinsonism Relat.

Disord. 2015; 21: 1294–1295.

Gan-Or Z, Ozelius LJ, Bar-Shira A, Saunders-Pullman R, Mirelman A, Kornreich R, et al. The p.L302P mutation in the lysosomal enzyme gene SMPD1 is a risk factor for Parkinson disease. Neurology 2013; 80: 1606–1610.

Giri, A., Mok, K. Y., Jansen, I., Sharma, M, Tesson, C, Mangone, G, et al. Lack of evidence for a role of genetic variation in TMEM230 in the risk for Parkinson's disease in the Caucasian population. Neurobiol. Aging 2017; 50: 167.e11–167.e13.

Goker-Alpan O, Schiffmann R, LaMarca ME, Nussbaum RL, McInerney-Leo A, Sidransky E. Parkinsonism among Gaucher disease carriers. J. Med. Genet. 2004; 41: 937–940.

Gonzaga-Jauregui C, Harel T, Gambin T, Kousi M, Griffin LB, Francescatto L, et al. Exome Sequence Analysis Suggests that Genetic Burden Contributes to Phenotypic Variability and Complex Neuropathy. Cell Rep 2015; 12: 1169–1183.

Haelterman NA, Yoon WH, Sandoval H, Jaiswal M, Shulman JM, Bellen HJ. A mitocentric view of Parkinson's disease. Annu. Rev. Neurosci. 2014; 37: 137–159.

Hamza, T. H., & Payami, H. The heritability of risk and age at onset of Parkinson's disease after accounting for known genetic risk factors. Journal of Human Genetics 2010; 55: 241–243.

Hofman A, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2016 objectives and design update. European Journal of Epidemiology 2015; 30: 661–708.

Holm S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. Scand J Statistics 1979; 6: 65-70.

Jansen IE, Bras JM, Lesage S, Schulte C, Gibbs JR, Nalls MA, et al. CHCHD2 and Parkinson's disease. The Lancet Neurology 2015; 14: 678–679.

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; 18: 22.

Kalia LV, Lang AE. Parkinson's disease. The Lancet 2015; 386: 896–912.

Keller MF, Saad M, Bras J, Bettella F, Nicolaou N, Simon-Sanchez J, et al. Using genome-wide complex trait analysis to quantify 'missing heritability" in Parkinson"s disease. Hum. Mol. Genet. 2012; 21: 4996–5009.

Kenna KP, McLaughlin RL, Byrne S, Elamin M, Heverin M, Kenny EM, et al. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. J. Med. Genet. 2013; 50: 776–783.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for

estimating the relative pathogenicity of human genetic variants. Nature Genetics 2014; 46: 310–315.

Kluenemann HH, Nutt JG, Davis MY, Bird TD. Parkinsonism syndrome in heterozygotes for Niemann–Pick C1. Journal of the Neurological Sciences 2013; 335: 219–220.

Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Research 2016; 44: D862–8.

Lee HJ. Clearance of -Synuclein Oligomeric Intermediates via the Lysosomal Degradation Pathway. Journal of Neuroscience 2004; 24: 1888–1896.

Lee S, Fuchsberger C, Kim S, Scott L. An efficient resampling method for calibrating single and gene-based rare variant association analysis in case-control studies. Biostatistics 2016; 17: 1–15.

Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. Biostatistics 2012; 13: 762–775.

Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010; 26: 589–595.

Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide B-MM, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. PLoS Genetics 2012; 8: e1002548.

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: 25:5483-5489.

Lupski, J. R., Belmont, J. W., Boerwinkle, E., & Gibbs, R. A. Clan Genomics and the Complex Architecture of Human Disease. Cell 2011; 147: 32–43.

Marder K, Wang Y, Alcalay RN, Mejia-Santana H, Tang M-X, Lee A, et al. Age-specific penetrance of LRRK2 G2019S in the Michael J. Fox Ashkenazi Jewish LRRK2 Consortium. Neurology 2015; 85: 89–95.

Mazzulli JR, Xu Y-H, Sun Y, Knight AL, McLean PJ, Caldwell GA, et al. Gaucher disease glucocerebrosidase and α-synuclein form a bidirectional pathogenic loop in synucleinopathies. Cell 2011; 146: 37–52.

Mazzulli JR, Zunke F, Isacson O, Studer L, Krainc D. α-Synuclein-induced lysosomal dysfunction occurs through disruptions in protein trafficking in human midbrain synucleinopathy models. Proc. Natl. Acad. Sci. U.S.A. 2016; 113: 1931–1936.

Mazzulli JR, Zunke F, Tsunemi T, Toker NJ, Jeon S, Burbulla LF, et al. Activation of β-Glucocerebrosidase Reduces Pathological α-Synuclein and Restores Lysosomal Function in Parkinson's Patient Midbrain Neurons. J. Neurosci. 2016; 36: 7693–7706.

McGlinchey RP, Lee JC. Cysteine cathepsins are essential in lysosomal degradation of α-synuclein. Proc. Natl. Acad. Sci. U.S.A. 2015; 112: 9322–9327.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20: 1297–1303.

Migdalska-Richards A, Daly L, Bezard E, Schapira AHV. Ambroxol effects in glucocerebrosidase and α-synuclein transgenic mice. Ann. Neurol. 2016; 80: 766–775.

Moors T, Paciotti S, Chiasserini D, Calabresi P, Parnetti L, Beccari T, et al. Lysosomal Dysfunction and α-Synuclein Aggregation in Parkinson's Disease: Diagnostic Links. Mov Disord. 2016; 31: 791–801.

Mullin, S., & Schapira, A. The genetics of Parkinson's disease. British Medical Bulletin 2015; 114: 39–52.

Nalls MA, Bras J, Hernandez DG, Keller MF, Majounie E, Renton AE, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. Neurobiol. Aging 2015; 36: 1605.e7–12.

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. Nature Genetics 2014; 46: 989–993.

Pankratz N, Beecham GW, DeStefano AL, Dawson TM, Doheny KF, Factor SA, et al. Meta-analysis of Parkinson's Disease: Identification of a novel locus, RIT2. Ann. Neurol. 2012; 71: 370–384.

Parkinson Progression Marker Initiative. The Parkinson Progression Marker Initiative (PPMI). Prog. Neurobiol. 2011; 95: 629–635.

Perrett RM, Alexopoulou Z, Tofaris GK. The endosomal pathway in Parkinson's disease. Molecular and Cellular Neuroscience 2015; 66: 21–28.

Pihlstrøm, L., & Toft, M. Parkinson's disease: What remains of the "missing heritability"? Movement Disorders 2011; 26: 1971–1973.

Pringsheim T, Jette N, Frolkis A, Steeves TDL. The prevalence of Parkinson's disease: a systematic review and meta-analysis. Mov Disord. 2014; 29: 1583–1590.

Rana HQ, Balwani M, Bier L, Alcalay RN. Age-specific Parkinson disease risk in GBA mutation carriers: information for genetic counseling. Genet. Med. 2013; 15: 146–149.

Rothaug M, Zunke F, Mazzulli JR, Schweizer M, Altmeppen H, Lüllmann-Rauch R, et al. LIMP-2 expression is critical for β-glucocerebrosidase activity and α-synuclein clearance. Proc. Natl. Acad. Sci. U.S.A. 2014; 111: 15573–15578.

1 Sardi SP, Clarke J, Kinnecom C, Tamsett TJ, Li L, Stanek LM, et al. CNS expression of
2 glucocerebrosidase corrects alpha-synuclein pathology and memory in a mouse model of
3 Gaucher-related synucleinopathy. Proc. Natl. Acad. Sci. U.S.A. 2011; 108: 12101–12106.

4 Shachar T, Bianco CL, Recchia A, Wiessner C, Raas-Rothschild A, Futerman AH. Lysosomal
5 storage disorders and Parkinson's disease: Gaucher disease and beyond. Mov Disord. 2011; 26:
6 1593–1604.

7 Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, Barbosa ER, et al. Multicenter
8 Analysis of Glucocerebrosidase Mutations in Parkinson's Disease. N Engl J Med 2009; 361:
9 1651–1661.

10 Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. alpha-Synuclein
11 locus triplication causes Parkinson's disease. Science 2003; 302: 841–841.

12 Singleton AB, Farrer MJ, Bonifati V. The genetics of Parkinson's disease: progress and
13 therapeutic implications. Mov Disord. 2013; 28: 14–23.

14 Soldner F, Stelzer Y, Shivalila CS, Abraham BJ, Latourelle JC, Barrasa MI, et al. Parkinson-
15 associated risk variant in distal enhancer of α-synuclein modulates target gene expression. Nature
16 2016; 533: 95–99.

17 Tayebi N, Walker J, Stubblefield B, Orvisky E, LaMarca ME, Wong K, et al. Gaucher disease
18 with parkinsonian manifestations: does glucocerebrosidase deficiency contribute to a
19 vulnerability to parkinsonism? Mol. Genet. Metab. 2003; 79: 104–109.

20 Trinh J, Guella I, Farrer MJ. Disease penetrance of late-onset parkinsonism: a meta-analysis.
21 JAMA Neurol 2014; 71: 1535–1539.

22 van Blitterswijk M, van Es MA, Hennekam EAM, Dooijes D, van Rheenen W, Medic J, et al.
23 Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum. Mol. Genet. 2012; 21:
24 3776–3784.

25 van Rooij, Jhamai, M, Arp, PP, Nouwens, SCA, Verkerk, M, Hofman, A, et al. Population-
26 specific genetic variation in large sequencing datasets: why more data is still better. EJHG 2017;
27 Epub ahead of print. doi: 10.1038/ejhg.2017.110

28 Vekrellis K, Xilouri M, Emmanouilidou E, Rideout HJ, Stefanis L. Pathological roles of α-
29 synuclein in neurological disorders. The Lancet Neurology 2011; 10: 1015–1025.

30 Verstraeten A, Theuns J, Van Broeckhoven C. Progress in unraveling the genetic etiology of
31 Parkinson disease in a genomic era. Trends in Genetics 2015; 31: 140–149.

32 Vogiatzi T, Xilouri M, Vekrellis K, Stefanis L. Wild type alpha-synuclein is degraded by
33 chaperone-mediated autophagy and macroautophagy in neuronal cells. J. Biol. Chem. 2008; 283:
34 23542–23556.

35 Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-

throughput sequencing data. Nucleic Acids Research 2010; 38: e164–e164.

Winder-Rhodes SE, Garcia-Reitböck P, Ban M, Evans JR, Jacques TS, Kemppinen A, et al. Genetic and pathological links between Parkinson's disease and the lysosomal disorder Sanfilippo syndrome. Mov Disord. 2012; 27: 312–315.

Wong YC, Krainc D. Lysosomal trafficking defects link Parkinson's disease with Gaucher's disease. Mov Disord. 2016; 31: 1610–1618.

Wu R-M, Lin C-H, Lin H-I. The p.L302P mutation in the lysosomal enzyme gene SMPD1 is a risk factor for Parkinson disease. Neurology 2014; 82: 283–283.

Zech M, Nübling G, Castrop F, Jochim A, Schulte EC, Mollenhauer B, et al. Niemann-Pick C disease gene mutations and age-related neurodegenerative disorders. PLoS ONE 2013; 8: e82879.

Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, et al. Searching for missing heritability: Designing rare variant association studies. Proceedings of the National Academy of Sciences 2014; 111: E455–E464.

Figure Legend

Figure 1 Overall analytic strategy. (Left) Variant categories. Because the number, frequency, and effect sizes of Parkinson's disease risk variants remains incompletely defined, our analyses considered three nested categories based on increasing variant pathogenicity: (1) all non-synonymous variants (Nonsyn), (2) likely damaging variants based on combined annotation dependent depletion (CADD) score, and (3) loss-of-function (LoF) variants. Based on the known prevalence of Parkinson's disease and incomplete penetrance documented for many risk alleles, we also considered 2 frequency thresholds, including rare (MAF<1%) and somewhat more common (MAF<3%) variants. (Right) Analysis Flowchart. The Sequence Kernel Association Test-Optimal (SKAT-O) was initially performed for the complete LSD gene set in the IPDGC discovery cohort, considering each variant category separately. For those categories with a significant SKAT-O association in the full gene set, a secondary analysis was performed excluding all *GBA* variants in order to confirm the involvement of additional genes. This was repeated in each of the replication cohorts (PPMI and NeuroX). Lastly, in order to highlight those loci driving associations detected in the gene set, secondary analyses were performed using SKAT-O to evaluate variants in each LSD gene independently.

Figure 2 Distribution of LSD variants in the IPDGC cohort. The number of likely damaging LSD variants (MAF<3%, CADD C-score≥12.37) per individual is shown versus the proportional representation in the IPDGC discovery cohort. Cases (Red) and Controls (Blue) are plotted separately. Many individuals harbor multiple LSD alleles, and the distribution is right-skewed among Parkinson's disease cases. The analysis considers variants in all 54 LSD genes. Supplemental Figure 2 shows a similar plot restricted to the 5 top driver genes.

1 Tables

2 Table 1 LSD Genes and Variants in the IPDGC cohort

Disease	Gene	Variants[a]
Aspartylglucosaminuria	*AGA*	13 (10)
Metachromatic Leukodystrophy	*ARSA*	5 (5)
Maroteaux-Lamy disease	*ARSB*	11 (10)
Farber Lipogranulomatosis	*ASAH1*	20 (17)
Kufor-Rakeb syndrome	*ATP13A2*	24 (18)
Neuronal Ceroid Lipofuscinosis (CLN3)	*CLN3*	18 (17)
Neuronal Ceroid Lipofuscinosis (CLN5)	*CLN5*	-
Neuronal Ceroid Lipofuscinosis (CLN6)	*CLN6*	10 (7)
Neuronal Ceroid Lipofuscinosis (CLN8)	*CLN8*	9 (4)
Cystinosis	*CTNS*	13 (12)
Galactosialidosis	*CTSA*	14 (11)
Neuronal Ceroid Lipofuscinosis (CLN10)	*CTSD*	7 (4)
Neuronal Ceroid Lipofuscinosis (CLN13)	*CTSF*	11 (9)
Pycnodysostosis	*CTSK*	6 (5)
Neuronal Ceroid Lipofuscinosis (CLN4B)	*DNAJC5*	5 (5)
Fucosidosis	*FUCA1*	15 (12)
Pompe disease	*GAA*	15 (10)
Krabbe disease	*GALC*	36 (30)
Morquio A disease	*GALNS*	22 (14)
Gaucher disease	*GBA*	39 (32)
Fabry disease	*GLA*	9 (7)
GM1-Gangliosidosis/Morquio B	*GLB1*	8 (4)
GM2-Gangliosidosis	*GM2A*	1 (1)
I-Cell disease	*GNPTAB*	39 (31)
Sanfilippo D syndrome	*GNS*	20 (11)
Neuronal Ceroid Lipofuscinosis (CLN11)	*GRN*	19 (12)
Sly disease	*GUSB*	17 (10)
Tay-Sachs disease	*HEXA*	20 (18)
Sandhoff disease	*HEXB*	8 (6)
Sanfilippo C syndrome	*HGSNAT*	18 (15)
Mucopolysaccharidosis Type IX	*HYAL1*	13 (9)
Hunter syndrome	*IDS*	9 (8)
Hurler syndrome	*IDUA*	8 (4)
Neuronal Ceroid Lipofuscinosis (CLN14)	*KCTD7*	4 (3)
Danon disease	*LAMP2*	9 (7)
Wolman disease	*LIPA*	14 (10)
Alpha-Mannosidosis	*MAN2B1*	12 (11)
Beta-Mannosidosis	*MANBA*	18 (15)
Mucolipidosis Type IV	*MCOLN1*	19 (14)
Neuronal Ceroid Lipofuscinosis (CLN7)	*MFSD8*	18 (14)
Schindler Disease/Kanzaki disease	*NAGA*	9 (8)
Sanfilippo B syndrome	*NAGLU*	10 (9)
Sialidosis	*NEU1*	-
Niemann-Pick Disease Type C1	*NPC1*	43 (35)
Niemann-Pick Disease Type C2	*NPC2*	2 (2)
Neuronal Ceroid Lipofuscinosis (CLN1)	*PPT1*	9 (7)
Sphingolipid-activator deficiency	*PSAP*	22 (16)
Action myclonus-renal failure syndrome	*SCARB2*	10 (7)
Sanfilippo A syndrome	*SGSH*	10 (8)
Salla disease	*SLC17A5*	18 (17)
Niemann-Pick Disease Type A/B	*SMPD1*	25 (21)
GM3-Gangliosidosis	*ST3GAL5*	11 (11)
Multiple Sulfatase Deficiency	*SUMF1*	-
Neuronal Ceroid Lipofuscinosis (CLN2)	*TPP1*	15 (13)

3 [a]The number of variants (MAF < 3%) in each LSD gene is shown for the IPDGC
4 discovery cohort, including total number of nonsynonymous variants and likely
5 damaging variants based on CADD (in parentheses). Of the 54 LSD genes
6 considered, no exonic variants in *CLN5* or *NEU1* passed quality control filters (see
7 Methods), and no nonsynonymous variants were identified in *SUMF1*.
8 LSD=Lysosomal storage disorder; CADD=Combined Annotation Dependent
9 Depletion.

Table 2 Analyses of LSD Variant Burden in Parkinson's disease

Cohort	Cases (n)	Controls (n)	Variants[a]	(a) MAF < 1%		(b) MAF < 3%	
				n[b]	p_{LSD}[c]	n	p_{LSD} (p_{-GBA})[c]
Discovery							
IPDGC	1,167	1,685	(1) nonsyn	746 (709)	0.056	760 (721)	0.014 (0.026)
			(2) CADD	585 (555)	0.066	596 (564)	0.0055 (0.0198)
			(3) LoF	69 (65)	0.464	-[d]	-
Replication							
PPMI	436	169	(1) nonsyn	243 (237)	0.096	256 (248)	0.320
			(2) CADD	179 (174)	0.294	187 (180)	0.281
NeuroX	6,713	5,964	(1) nonsyn	452 (443)	0.068	467 (456)	0.0004 (0.002)
			(2) CADD	338 (331)	0.057	348 (339)	0.0003 (0.020)

[a]Variants were classified into nested categories (Figure 1A) based on two frequency thresholds, MAF < 1% (a) or 3% (b), and three functional filters, all nonsynonymous (1), CADD likely damaging (2), and LoF (3).

[b]n=total number of LSD variant (number of variants excluding *GBA*). In parentheses, the number of variants excluding those in *GBA* are shown.

[c]Empirical SKAT-O *p*-values are based on 10,000 permutations following randomization of case/control status, and adjusted for multiple comparisons using the Bonferroni-Holm method (see Methods). As shown in Figure 1, primary analyses consider the variant burden among 54 LSD genes (p_{LSD}). For significant SKAT-O results, secondary analyses were performed excluding all variants in *GBA* (**p_{-GBA}**). Unadjusted p-values are reported in Supplemental Table 5.

[d]No additional LoF variants met the relaxed frequency threshold (MAF < 3%).

LSD=lysosomal storage disorder; MAF=minor allele frequency; IPDGC=International Parkinson's Disease Genomics Consortium Discovery Cohort; PPMI= Parkinson's Progression Markers Initiative Replication Cohort; NeuroX = NeuroX exome array cohort; nonsyn=nonsynonymous variants; CADD=Combined Annotation Dependent Depletion; LoF= loss of function variants
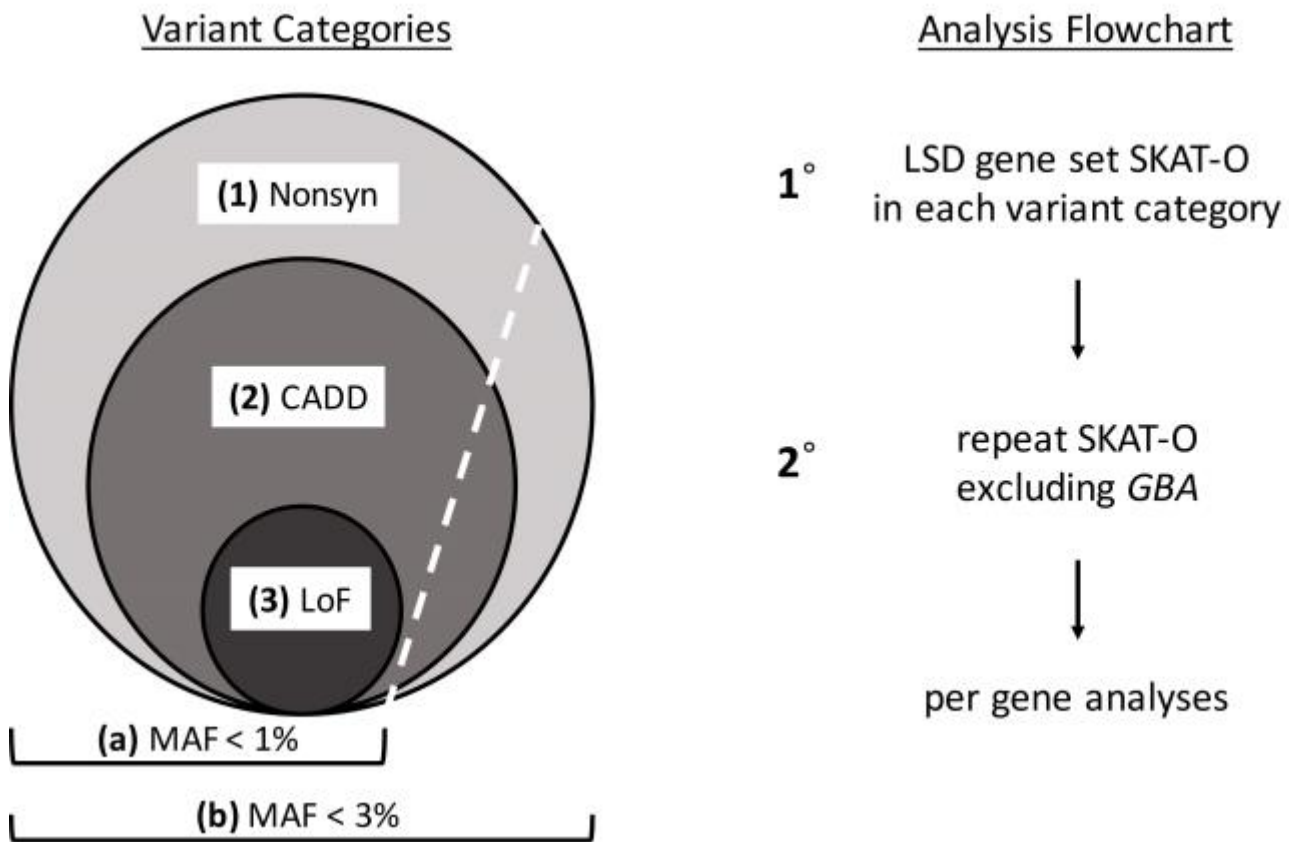
Figures

Figure 1 Overall Analytic Strategy.

Figure 2 Distribution of LSD variants in the IPDGC cohort.