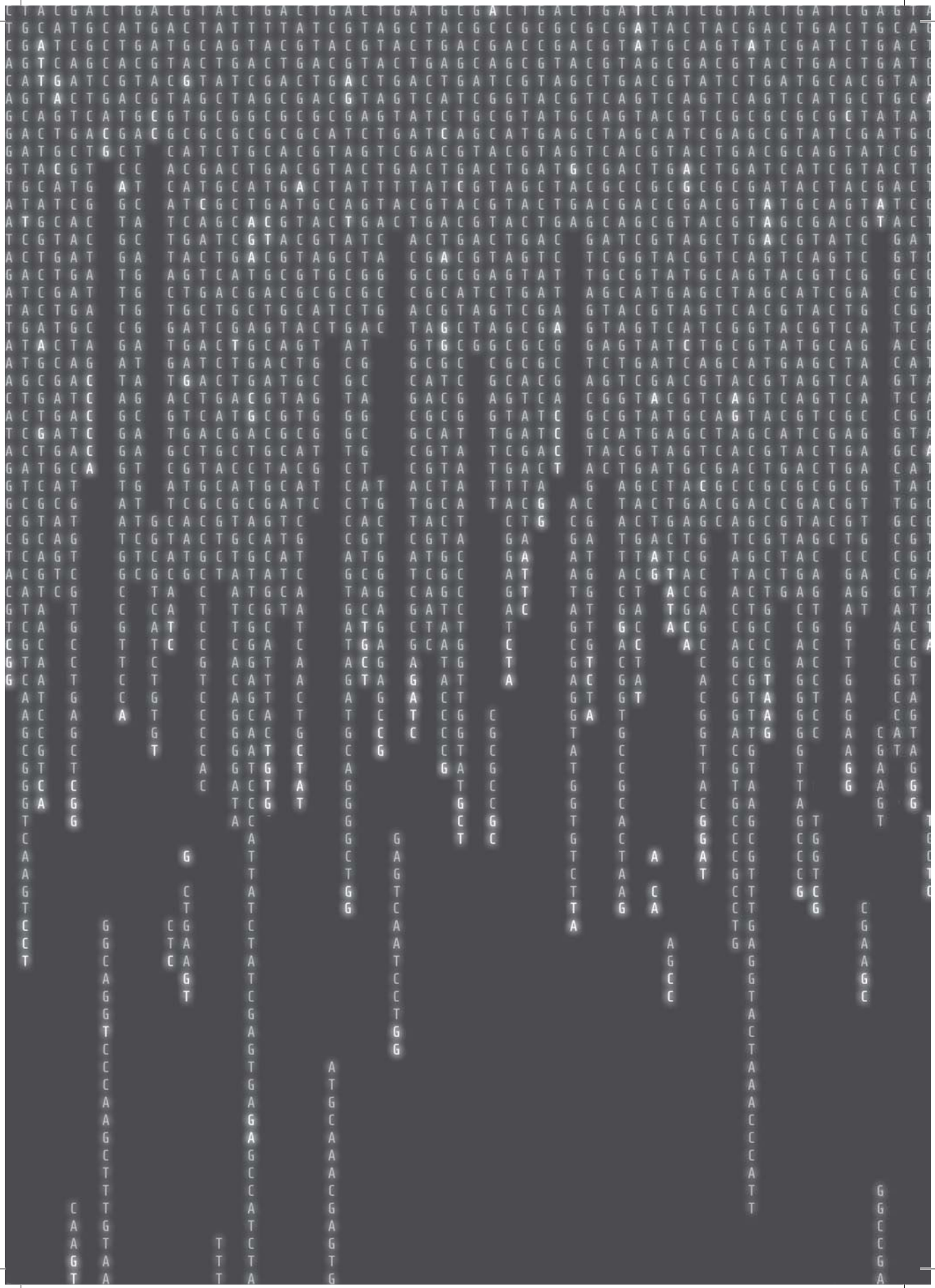
Cracking the code-ing sequence for Parkinson’s disease

Jansen, I.E.

2017

document version
Publisher's PDF, also known as Version of record



CHAPTER 4

EXCESSIVE BURDEN OF LYSOSOMAL STORAGE DISORDER GENE VARIANTS IN PARKINSON'S DISEASE

This chapter has been submitted as: Jansen IE*, Robak L*, van Rooij J, Uitterlinden AG, Kraaij R, Jankovic J, International Parkinson's Disease Genetics Consortium, Heutink P, Shulman J. Excessive Burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain*; 2017.

* These authors contributed equally to this study

ABSTRACT

Objective: Mutations in the glucocerebrosidase gene (*GBA*), which cause Gaucher disease, are also potent risk factors for Parkinson's disease (PD). We examined whether a genetic burden of variants in other lysosomal storage disorder (LSD) genes is more broadly associated with PD susceptibility.

Methods: The sequence kernel association test (SKAT-O) was used to interrogate variant burden within 54 LSD genes, leveraging whole exome sequencing (WES) data from 1,167 PD cases and 1,685 control subjects recruited from across the United States and Europe. For replication, we interrogated two independent datasets, including WES from an additional 436 cases and 169 controls, and exome-wide genotyping from 6,713 cases and 5,964 controls. Secondary analyses were also performed to highlight the specific LSD genes driving the aggregate association signal.

Results: In the discovery cohort, we demonstrate a significant burden of rare, likely damaging LSD gene variants in association with PD risk. The association signal was robust to the exclusion of *GBA*, and similar results were obtained in our replication cohorts. Secondary analyses confirm associations at the *GBA* and *SMPD1* loci, and newly implicate *CTSD*, *SLC17A5*, and *HGSNAT* as candidate PD susceptibility genes. The majority of PD cases (56%) have at least one putative damaging variant in an LSD gene and 22% carry multiple alleles.

Interpretation: Our results highlight several promising new susceptibility loci and reinforce the importance of lysosomal mechanisms in PD pathogenesis. We suggest that multiple genetic hits may act in combination to degrade lysosomal function, enhancing PD susceptibility.

INTRODUCTION

Parkinson's disease (PD) is a common neurodegenerative disorder with evidence for a substantial genetic etiology.[1] Studies in families as well as large population-based cohorts have implicated more than 30 genes;[2-5] however, the risk alleles identified to date explain only a fraction of PD heritability estimates,[6-8] suggesting the involvement of additional loci. Beyond discovering the responsible genes, a major challenge remains to understand the mechanisms by which these factors alter disease onset and/or progression, including whether they act independently or function within coherent biologic pathways.

Substantial evidence highlights the importance of lysosomal mechanisms in PD susceptibility and pathogenesis.[1,9,10] Prior to its discovery as a PD risk locus, the *glucocerebrosidase* gene, *GBA*, was known to cause Gaucher disease, an autosomal recessive lysosomal storage disorder (LSD). Increased risk for PD in heterozygous carriers of *GBA* loss-of-function alleles was first recognized in families of individuals with Gaucher disease.[11,12] A follow-up study in a large, case-control sample confirmed that heterozygous *GBA* variants confer at least a five-fold increased risk of PD.[13] *GBA* variants may also modify PD clinical manifestations, causing earlier age-of-onset and a higher risk of cognitive impairment. LSDs—of which there are more than 50—are Mendelian-inherited, metabolic disorders collectively caused by dysfunction in lysosomal biogenesis or function, and similarly characterized by the abnormal accumulation of non-degraded metabolites in the lysosome.[14-16] The strong genetic evidence linking Gaucher disease and PD risk leads to the intriguing hypothesis that more broadly, LSDs and PD may share a common genetic mechanism. Other LSD genes have therefore become attractive candidate PD risk genes.[17-19] Several studies have consistently supported a role for *SMPD1*,[20-23] which causes Niemann-Pick type C disease. Initial reports evaluating other LSD genes, including *NPC1*, *NPC2*, *MCOLN1, NAGLU* and *ARSB*, have either shown conflicting results or await further replication.[24-28] LSDs are individually quite rare in populations of European ancestry, as are the known genetic variants established to cause these disorders.[14-16] 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 (GWAS) in large PD case-control cohorts have independently implicated more common risk alleles at another LSD gene, *SCARB2*,[7,29] 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 α-synuclein[30-32] which aggregates to form Lewy body pathology in PD. Reciprocally, α-synuclein disrupts neuronal vesicle trafficking and lysosomal function.[33,34]

In this study, we leverage the largest PD whole exome sequencing (WES) dataset currently available to systematically examine the overlap between genes responsible for

LSDs and PD risk. Our results reveal an aggregate burden for genetic variants among 54 genes established to cause LSDs and suggest that many genes besides *GBA* contribute to PD risk.

METHODS

Subjects

The International Parkinson's Disease Genomics Consortium (IPDGC) WES discovery dataset used for this study consists of 2,852 samples of Northern and Western European ancestry, including 1,167 PD cases and 1,685 controls not known to have PD.[28,35-37] Subjects were recruited from academic medical centers across the United States and Europe. PD cases were diagnosed with PD at a mean age of 41.2 years (SD=10.9); 40.4% of which have a family history of PD. Control subjects were on average 63.8 years of age (SD=17.1). 1,201 control exomes originated from the Rotterdam Study exome dataset version 1 (RSX1). The Rotterdam Study is a prospective population-based cohort study based in Rotterdam, the Netherlands. WES was performed on DNA from participants from the RSX1 subcohort, enrolled in 1990, with an average age at baseline of 68.6 (SD=8.6, 54.4% female).[38] All IPDGC and RSX1 subjects gave written informed consent for participation in genetic research, which was approved by relevant oversight committees/institutional review boards. PD patients harboring known pathogenic variants in PD genes were excluded from analysis. Following quality control filters, the Parkinson's Progression Markers Initiative (PPMI) replication dataset[39] includes 436 cases and 169 controls of Northwest European descent. The PD cases were diagnosed at an average age of 59.8 years (SD=10.0), and 27.1% were known to have a family history of PD. PPMI controls were an average of 61.8 years of age (SD=10.1) at the time of evaluation. Data used in the preparation of this article were obtained from the PPMI database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. Samples analyzed for both the IPDGC and PPMI cohorts were derived from whole blood. The NeuroX dataset has been previously described in detail,[28,35] including 6,713 individuals with PD and 5,964 controls. NeuroX cases were diagnosed at an average age of 61.6 (SD=12.4) and controls were evaluated at an average age of 64.1 (SD=14.3).

Sequencing/genotyping and quality control

WES for the IPDGC and RSX1 cohorts was performed using the Roche Nimblegen SeqCap v2 or Illumina exome capture kits to prepare sample libraries, followed by paired-end sequencing with Illumina HiSeq2000. The generation of the PPMI WES dataset are described elsewhere (www.ppmi-info.org). Although the datasets originate from different consortia, the same algorithms were used for read processing. The Burrows-Wheeler Aligner (BWA)-MEM[40] was used for alignment of sequencing reads to the human reference genome

(hg19). Using Picard tools (http://broadinstitute.github.io/picard), Binary Alignment/Map (BAM) files were generated in a sorted and indexed manner. Alignments were Base-Quality score recalibrated and indels realigned using the Genome Analysis Toolkit (GATK)[41] v3.3-0, after which single nucleotide variants and small insertions/deletions were called with the HaplotypeCaller (GATK) to one gVCF file per individual. The IPDGC and RSX1 WES datasets (hereafter referred to as simply the IPDGC discovery dataset) were merged by joint variant calling of the individual gVCF files. Variants that were not assigned with the standard GATK quality annotation 'PASS' were excluded for subsequent analyses. 94.4% and 98.0% of the IPDGC and PPMI exomes, respectively, achieved a minimum of 10x coverage.

For individual quality control (QC), samples were excluded for ambiguous gender, deviating heterozygosity/genotype calls, low genotype calls, cryptic relatedness following identity-by-descent analyses, or poor clustering on multi-dimensional scaling (MDS) component analysis indicating population outliers. Analyses of relatedness and MDS were based on linkage disequilibrium-pruned common variants. Genotype and variant QC was accomplished by removal of low-quality genotypes (Phred-scaled genotype quality score < 20, depth < 8) and variants with low call rates or departure from Hardy-Weinberg equilibrium (HWE). Furthermore, for the IPDGC discovery dataset, variants were only considered when located within the overlapping targeted regions of the applied library preparation capture kits. Post-QC procedures, a total of 462,946 and 192,421 variants were called for the IPDGC and PPMI datasets, respectively.

NeuroX consists of 242,901 exonic variants from the Illumina Infinium HumanExome BeadChip and 24,706 custom variants related to neurologic disease.[35,42] For individual QC, as above, samples were excluded for gender ambiguity, dubious heterozygosity/genotype calls, evidence of relatedness, or poor MDS clustering. For variant QC, variants were excluded for subsequent analyses with low call rates, departure from HWE, or with significant differences in missingness rate between cases and controls. Post-QC procedures, we called 177,028 exonic variants from the NeuroX LSD gene-set.

Variant selection

Our analyses initially considered 54 LSDs (Table 1), based on widely accepted clinical, pathologic, and metabolic criteria.[14-16] All variants within the LSD gene-set were extracted from the three datasets. For the IPDGC WES dataset, no variants in the genes *CLN5* and *NEU1* passed the pre-specified maximum missingness criteria of 15%, yielding 1,175 total exonic variants for consideration in these analyses. Variants were categorized in nested groups including (1) nonsynonymous (n=786 variants in 51 genes), (2) likely damaging (n=609 variants in 51 genes), or (3) loss-of-function (n=69 variants in 27 genes) (see Table 1 and Supplementary Table 1). Loss-of-function variants included stop gain/loss, frameshift, and splicing mutations falling within 2 base pairs of exon-intron junctions. Predictions of variant pathogenicity were obtained from ANNOVAR,[43] based on the Combined

Table 1 - part 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*	15 (13)
Galactosialidosis	*CTSA*	14 (11)
Neuronal Ceroid Lipofuscinosis (CLN10)	*CTSD*	7 (4)
Neuronal Ceroid Lipofuscinosis (CLN13)	*CTSF*	12 (10)
Pycnodysostosis	*CTSK*	6 (5)
Neuronal Ceroid Lipofuscinosis (CLN4B)	*DNAJC5*	5 (5)
Fucosidosis	*FUCA1*	16 (13)
Pompe disease	*GAA*	15 (10)
Krabbe disease	*GALC*	37 (31)
Morquio A disease	*GALNS*	22 (14)
Gaucher disease	*GBA*	42 (33)
Fabry disease	*GLA*	9 (7)
GM1-Gangliosidosis/Morquio B	*GLB1*	8 (4)
GM2-Gangliosidosis	*GM2A*	1 (1)
I-Cell disease	*GNPTAB*	40 (32)
Sanfilippo D syndrome	*GNS*	20 (11)
Neuronal Ceroid Lipofuscinosis (CLN11)	*GRN*	20 (13)
Sly disease	*GUSB*	18 (11)
Tay-Sachs disease	*HEXA*	20 (18)
Sandhoff disease	*HEXB*	10 (7)
Sanfilippo C syndrome	*HGSNAT*	22 (15)
Mucopolysaccharidosis Type IX	*HYAL1*	14 (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*	15 (10)
Alpha-Mannosidosis	*MAN2B1*	12 (11)

Table 1 - part 2. LSD genes and variants in the IPDGC cohort.

Disease	Gene	Variants[a]
Beta-Mannosidosis	*MANBA*	18 (15)
Mucolipidosis Type IV	*MCOLN1*	19 (14)
Neuronal Ceroid Lipofuscinosis (CLN7)	*MFSD8*	19 (15)
Schindler Disease/Kanzaki disease	*NAGA*	9 (8)
Sanfilippo B syndrome	*NAGLU*	10 (9)
Sialidosis	*NEU1*	-
Niemann-Pick Disease Type C1	*NPC1*	44 (35)
Niemann-Pick Disease Type C2	*NPC2*	2 (2)
Neuronal Ceroid Lipofuscinosis (CLN1)	*PPT1*	9 (7)
Sphingolipid-activator deficiency	*PSAP*	22 (16)
Action mycolonus-renal failure syndrome	*SCARB2*	10 (7)
Sanfilippo A syndrome	*SGSH*	12 (9)
Salla disease	*SLC17A5*	18 (17)
Niemann-Pick Disease Type A/B	*SMPD1*	27 (23)
GM3-Gangliosidosis	*ST3GAL5*	11 (11)
Multiple Sulfatase Deficiency	*SUMF1*	-
Neuronal Ceroid Lipofuscinosis (CLN2)	*TPP1*	16 (13)

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

Annotation Dependent Depletion (CADD) algorithm (v1.3, http://cadd.gs.washington.edu).[44] In accordance with prior work,[45] we selected a CADD C-score≥12.37, representing the most damaging 2% of all possible nucleotide changes in the genome.

For the PPMI cohort, no variants were called in *DNAJC5*, resulting in a dataset of 515 total exonic variants, of which 256 variants from 49 genes were nonsynonymous and 187 variants in 47 genes met the CADD criteria for putative damaging changes (Supplementary Table 1). For the NeuroX cohort, all genes in the 54-gene-set were represented, resulting in 467 nonsynonymous variants, of which 348 were classified as likely damaging (Supplementary Table 1). Within these categories, variants were based on two minor allele frequency (MAF) thresholds: (a) <1% and (b) <3%. The latter, more relaxed frequency threshold is based on the population prevalence[46,47] and known incomplete penetrance of PD risk alleles.[48-50]

Statistical analysis

The sequence kernel association test – optimal (SKAT-O)[51,52] was performed, using the R-package SKAT v1.0.9 to determine the difference in the aggregate burden of rare LSD gene variants between PD cases and controls. Covariates were included to adjust analyses for gender and WES coverage (pre-QC missingness). Twenty MDS components were also included to account for other possible confounding factors, such as latent population stratification. In order to establish statistical significance, all SKAT-O results with an unadjusted $p < 0.05$, were subject to permutation testing, implemented within SKAT-O. An adjusted *p*-value was derived from an empirical distribution of null results based on 10,000 trials in which case/control assignment was randomized. Following this permutation test, an adjusted *p*-value (p_{adj}) < 0.05 was considered significant. Independently, we also performed permutation testing based on 1,000 random gene-sets similar in size to the LSD set. All *p*-values reported in the text are adjusted *p*-values; Table 2 reports both unadjusted and adjusted *p*-values. SKAT-O analysis was initially performed for the complete LSD gene-set, considering each class of variants defined based on frequency and functional characteristics. If the result was significant ($p_{adj} < 0.05$), the analysis was repeated excluding all *GBA* variants in order to confirm the involvement of additional genes. Lastly, secondary analyses were performed using SKAT-O to evaluate variants in each LSD gene independently.

To estimate statistical power, we performed 1,000 SKAT simulations of causal subregions within the discovery or replication datasets. We assumed a PD prevalence of 0.0041 and 0.0017 for the IPDGC and PPMI datasets, respectively, based on their distinct ages of onset.[46] 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 PD will have a damaging effect, all causal variants were assumed to have a positive coefficient (risk rather than protective alleles).

RESULTS

Variants were extracted from 54 LSDs, based on widely accepted clinical, pathologic, and metabolic criteria (Table 1).[14-16,53] To test our hypothesis that an aggregate burden of variants in the LSD gene-set contributes to PD risk, we implemented SKAT-O, which aggregates genetic information across defined genomic regions to test for associations.[51,52] Importantly, SKAT-O is robust to a wide frequency spectrum, including rare and more

common alleles, and to variants with different magnitudes and directions of effect. The genetic architecture of PD susceptibility, including the number, frequency, and effect sizes of responsible variants, remains incompletely defined. We therefore performed complementary analyses considering 3 nested categories of variants based on increasing potential pathogenicity: (1) all non-synonymous variants, (2) likely damaging variants, and (3) loss-of-function variants. For category 2, we leveraged the CADD framework, which integrates predictions from numerous bioinformatic algorithms into a single "C-score" and ranks all possible nucleotide changes in the genome based on potential to disrupt gene/protein function.[44] We selected a stringent C-score threshold predicting the top ~2% most damaging genomic variants; this subset is highly enriched for known pathogenic alleles.[45] Although most LSD gene variants that cause disease are individually rare, we further reasoned that more frequent variants might also contribute to PD given the increased prevalence, along with the incomplete penetrance documented for many established risk alleles, including *GBA* variants.[13,49,54] We therefore sampled variants at two frequency thresholds: (a) MAF < 1% for rare variants, and (b) MAF < 3% to include somewhat more common variants. In sum, our analytic strategy considers multiple frequency and functional characteristics, affording optimal sensitivity for detection of associations between LSD gene variants and PD susceptibility.

The results of the SKAT-O analyses for LSD gene-set variants within the IPDGC discovery cohort are presented in Table 2. Following permutation (see Methods), significant associations were detected for the LSD gene-set considering either all non-synonymous variants (category 1b, p_{adj} = 0.010) or likely damaging variants (category 2a, p_{adj} = 0.038 and 2b, p_{adj} = 0.003). No association was observed when considering only loss-of-function alleles (category 3), possibly due to the relative paucity of such variants limiting statistical power (Supplementary Table 1). The observed associations between LSD gene variant burden and PD risk also remained significant (p_{adj} < 0.05 for categories 1b and 2b) in an independent permutation test considering the likelihood of obtaining similar results based on randomly selected gene-sets. We next repeated each analysis with significant results, but excluding all *GBA* variants. As expected, the strength of the associations was attenuated; however, the association between likely damaging variants (MAF < 3%) and PD was robust to the exclusion of *GBA* and remained significant (category 2b, p_{adj}=0.026). Further, the association of LSD gene variants in category 1b remained suggestive. Our results indicate that the association between variant burden and PD risk in the IPDGC discovery cohort is mediated, at least in part, by the effects of LSD genes other than *GBA,* an established PD susceptibility locus.

To replicate our findings, we leveraged two independent cohorts, including an additional WES dataset from PPMI (436 PD cases and 169 controls)[39] and the NeuroX exome-wide genotyping dataset from IPDGC (6,713 PD cases and 5,964 controls)[28,35]. We again implemented SKAT-O to detect a potential variant burden in PD cases versus controls (Table 2).

Table 2. Analyses of LSD Variant Burden in PD.

				(a) MAF < 1%			(b) MAF < 3%		
					all LSD	w/o GBA[d]		all LSD	w/o GBA
Cohort	**# Cases**	**# Controls**	**Variants[a]**	**n[b]**	***p* (p_{adj})[c]**	***p* (p_{adj})**	**n**	***p* (p_{adj})**	***p* (p_{adj})**
Discovery									
IPDGC	1,167	1,685	(1) nonsyn	772 (732)	0.025 (0.051)	-	786 (744)	0.005 (0.010)	0.040 (0.068)
			(2) CADD	598 (567)	0.025 (0.038)	0.257	609 (576)	0.002 (0.003)	0.017 (0.026)
			(3) LoF	69 (65)	0.409	-	-[e]	-	-
Replication									
PPMI	436	169	(1) nonsyn	243 (237)	0.016 (0.024)	0.017 (0.025)	256 (248)	0.118	-
			(2) CADD	179 (174)	0.074	-	187 (180)	0.221	-
NeuroX	6,713	5,964	(1) nonsyn	452 (443)	0.031 (0.034)	0.024 (0.027)	467 (456)	$1.1x10^{-10}$ (0.0001)	0.001 (0.001)
			(2) CADD	338 (331)	0.057	-	348 (339)	$1.5x10^{-11}$ (0.0001)	0.018 (0.020)

[a]Variants were classified using 2 frequency criteria, MAF < 1% (a) or 3% (b), and 3 functional categories, all nonsynonymous (1), CADD likely damaging (2), and LoF (3). [b]n=total number of LSD variant. In parentheses, the number of variants excluding those in *GBA* are shown. [c]Unadjusted SKAT-O *p*-value (*p*) is shown, reflecting variant burden in LSD genes among PD cases versus controls. For analytic results with an unadjusted $p<0.05$, an empirical, adjusted p-value (p_{adj}) was determined based on 10,000 permutations following randomization of case/control status. [d]If the analysis of the total LSD gene-set was significant ($p_{adj} < 0.05$), an additional SKAT-O analysis was performed excluding all variants in GBA. [e]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.

In PPMI, we discovered consistent evidence for an excessive LSD variant burden in PD, and this signal was independent of *GBA*. However, the association was detected exclusively among rare alleles (MAF < 1%) and only when considering all non-synonymous variants (category 1a). It is possible that SKAT-O is sensitive to cohort differences between PPMI and the IPDGC, including both sample size and pertinent demographic features (e.g. age of onset and family history; see Methods). In the substantially larger NeuroX dataset, burden associations were detected among 3 out of 4 variant classes (categories 1a, 1b, and 2b), despite the less comprehensive genotyping coverage compared to WES. A major driver for the robust LSD gene-set association in NeuroX (categories 1b and 2b, p_{adj}=0.0001) appears to be the more common GBA^{E326K} variant ($Freq_{Cases}$=0.021, $Freq_{Controls}$=0.011), which has been reported to be associated with PD risk in several large studies.[55,56] Importantly, consistent with our findings in the IPDGC discovery cohort, the LSD gene-set burden association for all variant categories remained significant in NeuroX following exclusion of *GBA*. Thus, based on analyses in three independent PD case-control datasets, we demonstrate a burden of variants in LSD genes associated with PD risk, and this signal is at least partially independent of *GBA*.

To determine which additional LSD genes/variants may be responsible for the observed association with PD risk, we performed secondary analyses using SKAT-O to assess for potential contribution of variants within each gene considered independently. For these analyses, we returned to the IPDGC discovery dataset, and again focused on likely damaging variants, which showed the strongest association signal in our primary analysis (category 2b). In these gene-based analyses, besides the expected result for *GBA* (p_{adj} = 0.0001) and confirmation of *SMPD1* (p_{adj} = 0.029), we discover evidence of novel aggregate associations for variants in *CTSD* (p_{adj} = 0.002)*, SLC17A5* (p_{adj} = 0.005)*,* and *HGSNAT* (p_{adj} = 0.046). The specific variants implicated within each of these genes are included in the Supplemental Data (Supplementary Table 2), along with all other putative damaging variants considered in our full LSD gene-set analysis. While our datasets are underpowered to definitively assess the contributions of a particular rare variant in any single gene (see Discussion), these results identify the most likely specific loci driving the aggregate LSD gene-set association signal detected in the IPDGC discovery sample.

Lastly, we examined the distribution of putative damaging LSD gene variants (MAF < 3%, category 2b) within the IPDGC WES cohort (Figure 1). Consistent with our finding of an excessive variant burden in PD, the distribution of variants appeared modestly right-skewed in cases. The average variant burden among IPDGC cases was 0.9 alleles per individual, which was slightly higher than that seen in controls (0.8 alleles per individual). Given their commonality, the majority of IPDGC cases (56%) have at least 1 putative damaging variant in an LSD gene, and 22% carry multiple alleles. As discussed further below, this is consistent with a hypothetical model in which multiple LSD gene variants may interact to influence PD risk.

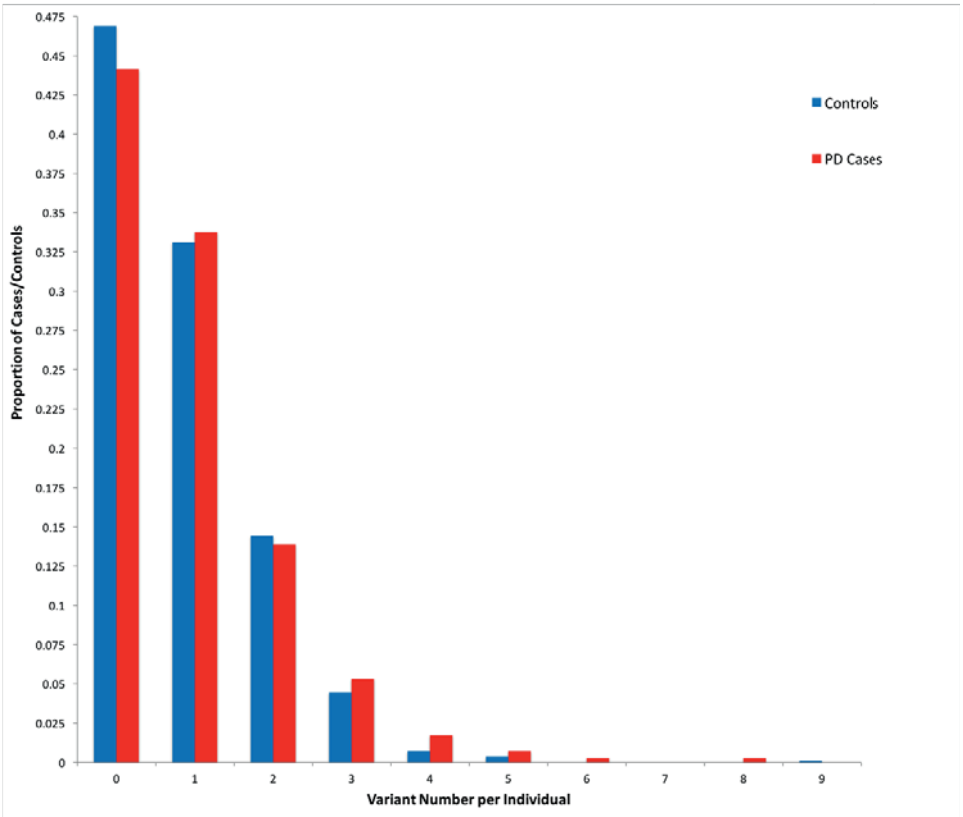


Figure 1. Distribution of LSD variants in the IPDGC cohort. The number of likely damaging LSD variants (MAF<3%, CADD C-score≥12.37) per individual is shown versus the fraction of Cases or Controls in the IPDGC discovery cohort. Many individuals harbor multiple LSD alleles, and the distribution is right-skewed among PD cases.

DISCUSSION

This study reveals an important connection between the genetic factors broadly responsible for LSDs, which are predominantly pediatric Mendelian disorders, and PD, an adult-onset neurodegenerative disorder with complex genetic etiology. Specifically, among 54 genes that cause LSDs, we find evidence for a burden of damaging alleles in association with PD risk. This association persisted after excluding *GBA*, consistent with a contribution from additional LSD genes. More than half of PD cases in our cohort harbor one or more putative damaging variants among the LSD genes. Thus, our results implicate several promising new PD susceptibility loci and reinforce the importance of lysosomal mechanisms in PD pathogenesis.

The strengths of this study include a large PD case/control discovery cohort as well as two independent datasets for replication of our findings. Since our understanding of the characteristics of causal alleles—including in both PD and LSDs—is incomplete, our initial analyses systematically considered multiple variant classes binned into categories based on frequency and putative functional impact. Interestingly, consideration of likely damaging alleles based on bioinformatic predictions, including more common LSD variants (MAF < 3%), appeared to offer optimal sensitivity for detection of a significant aggregate variant association. Critically, the implementation of burden association tests for joint consideration of LSD genes significantly improves statistical power over single gene and variant tests.[57] In populations of European ancestry, loss-of-function alleles, including those established to cause LSDs, are individually rare (Table 1), and based on post-hoc simulations (see Methods), we estimate poor power for discovery of rare PD risk alleles at isolated loci. For example, assuming a rare variant model (MAF = 0.035%, as for *GBA* loss-of-function alleles in our sample) and even assuming full penetrance, the IPDGC discovery cohort has only 30% power to discover an association for a single gene. However, a similar simulation considering the full set of 54 LSD genes was fully powered (100%). Our consideration of higher frequency variants further enhances power for both discovery and replication, especially when coupled with filtering based on potential pathogenicity. For example, allowing for more common variants (MAF < 3%) and assuming 10% of such alleles are causal, we estimate that the smaller PPMI cohort achieves 95% power for replication of a gene-set association, whereas negligible power (1%) is available for interrogation of a single gene candidate. We anticipate that larger WES datasets will significantly improve power, including for per gene analyses. In the IPDGC and PPMI cohorts, WES offers comprehensive characterization of LSD gene variants. By contrast, since the NeuroX data is restricted to those variants included on the genotyping array, it is possible that many potential pathogenic variants would be missed. Nevertheless, a total of 348 putative damaging variants were detected, including alleles for all LSD genes (Supplementary Table 1). The substantially larger NeuroX sample size makes this cohort more broadly representative of the population, and our findings likely therefore more generalizable. Since our cohort was composed of individuals with European ancestry, it will also be important to examine other ethnic populations in the future, especially those potentially enriched for LSD-causing variants due to genetic bottlenecks.

We also performed analyses in the IPDGC cohort to pinpoint the specific drivers from the LSD gene-set responsible for increasing PD risk. Our results (i) recapitulate the established association with *GBA*, (ii) strengthen the emerging evidence in support of *SMPD1,* and (iii) newly implicate *SLC17A5, HGSNAT* and *CTSD* as candidate PD susceptibility genes*.* Recessive mutations in *SMPD1* cause Niemann-Pick type A/B disease and this locus has been independently implicated in PD risk based on several published studies.[20-23] Similar to *GBA*, *SMPD1* (encoding sphingomyelinase) participates in ceramide metabolism.

While our analysis identified 23 candidate *SMPD1* risk alleles (Supplementary Table 2), they appear distinct from those reported in other studies of PD. Notably, we did find a non-synonymous variant, p.P332L, predicted to be damaging (CADD=32), and another substitution at this same amino acid position, p.P332R, was previously implicated in a Chinese PD cohort.[22] Among the novel candidate genes, *SLC17A5, HGSNAT*, and *CTSD*, most of the implicated variants are rare (MAF<1%). Only one of these variants (*rs16883930* in *SLC17A5*) is present in the 1000 Genomes reference,[58] having been previously examined in large GWAS, and this was non-associated with PD risk (p>0.05) based on available data.[59] Mutations in *SLC17A5*, *HGSNAT*, and *CTSD* cause the rare LSDs, Salla disease, Sanfilippo C syndrome (Mucopolysaccharidosis IIIC), and Neuronal Ceroid Lipofuscinosis (CLN10), respectively, which are characterized by the accumulation of sialic acid, heparan sulfate, or lipofuscin pigment. In addition to promoting lysosomal stress, glucosylceramide, which accumulates in Gaucher disease, has been suggested to directly promote the aggregation of α-synuclein.[34] Interestingly, *CTSD* encodes a lysosomal aspartyl proteinase which has been independently implicated in α-synuclein degradation.[60,61] Besides *HGSNAT* (MPSIIIC), other genes linked to mucopolysaccharidoses, including *NAGLU* (Sanfilippo B syndrome / MPSIIIB)[27] and *ARSB* (MPS6)[28] have been previously implicated as PD risk loci, and α-synuclein Lewy body pathology has also been documented in brain autopsies from selected Sanfilippo B cases.[62] In sum, the LSD genes and variants implicated by our studies are excellent candidates for further replication, including resequencing and/or genotyping in the largest available PD case/control samples. Although we employed rigorous quality control procedures for calling variants from WES and genotyping data, definitive confirmation of specific variants will require additional studies.

There is a growing recognition of the importance of lysosomal biology in PD pathogenesis.[10] First, the lysosome is an important route for α-synuclein degradation.[30-32] Genomic variants that elevate α-synuclein protein levels—such as rare locus multiplication[63] or a common polymorphism that enhances promoter activity[64]—also increase PD risk. Knockdown of selected LSD genes, including *GBA* or *SCARB2*, in neuronal cells or in mouse models impairs α-synuclein clearance,[34,65,66] whereas increasing glucocerebrosidase activity has the opposite effect.[67] Second, lysosomal autophagy plays a critical role in mitochondrial quality control, and substantial evidence, including from genetics, highlight mitochondrial dysfunction in PD.[68] Third, there is accumulating evidence from numerous experimental models that α-synuclein interferes with endoplasmic reticulum-to-Golgi vesicle trafficking, inducing reciprocal disruptions in lysosomal biogenesis.[33] Expression of α-synuclein impeded trafficking of multiple hydrolases linked to LSDs, including GBA, within human dopaminergic neurons.[69] In one recent study, subjects with idiopathic PD, in which *GBA* carriers were excluded, were found to have modest but significantly reduced glucocerebrosidase enzymatic activity based on peripheral blood testing.[70] Fourth, besides *GBA* and the other genes implicated in our study, mutations in *ATP13A2,* a rare cause of

recessive juvenile-onset parkinsonism and dementia has been independently implicated to cause the LSD Neuronal Ceroid Lipofuscinosis.[71] Lastly, many other common and rare PD risk alleles, including at *RAB7L1*, *GAK, LRRK2,* and *VPS35* have strong functional links to vesicle trafficking, including for lysosomal biogenesis and function. Together, these findings support a model in which partial loss-of-function in genes regulating lysosomal activity, such as those that cause LSDs, may increase vulnerability to α-synuclein-mediated mechanisms in PD.

Complex genetic disorders such as PD likely result from the cumulative impact and interaction of both common and rare allelic variants at multiple genomic loci.[2,3] Polygenic modeling approaches have previously demonstrated how common risk alleles can cumulatively impact PD risk and age-of-onset.[29,72] In addition, a recently published analysis in the IPDGC WES and NeuroX cohorts identified evidence for oligogenic interactions underlying PD risk, including alleles for *GBA* and those for established Mendelian PD genes.[73] In the IPDGC, WES reveals a substantial proportion of PD cases (22%) carrying 2 or more likely damaging variants in LSD genes. This observation suggests the possibility that multiple variants may interact in a multi-hit, combinatorial manner to degrade lysosomal function, causing the accumulation of α-synuclein and potentially other toxic substrates, and thereby increasing susceptibility for PD. Recent work has also implicated oligogenic inheritance in other neurologic disorders including, amyotrophic lateral sclerosis[74-76] and idiopathic peripheral neuropathy,[77] and further reveals how genes causing early-onset, monogenic disorders may act in combination to additionally trigger late-onset, complex genetic disorders. Future studies, including even-larger, case-control cohorts with WES and complementary experiments in PD 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 in PD.

ACKNOWLEDGEMENTS

We thank Drs. Andrew Singleton, John Hardy, V. Reid Sutton, Mike Nalls, Manu Sharma, and Zongqi Xia for critical review and helpful discussions related to the manuscript. We also thank Anamika Giri for assistance with data processing. IJ was funded by the Prinses Beatrix Spierfonds.

REFERENCES

1. Kalia LV, Lang AE. Parkinson's disease. *Lancet* 2015; 386(9996): 896-912.
2. Shulman JM, De Jager PL, Feany MB. Parkinson's disease: genetics and pathogenesis. *Annual review of pathology* 2011; 6: 193-222.
3. Singleton AB, Farrer MJ, Bonifati V. The genetics of Parkinson's disease: progress and therapeutic implications. *Movement disorders : official journal of the Movement Disorder Society* 2013; 28(1): 14-23.
4. Bras J, Guerreiro R, Hardy J. SnapShot: Genetics of Parkinson's disease. *Cell* 2015; 160(3): 570-.e1.
5. Verstraeten A, Theuns J, Van Broeckhoven C. Progress in unraveling the genetic etiology of Parkinson disease in a genomic era. *Trends in genetics : TIG* 2015; 31(3): 140-9.
6. Hamza TH, Payami H. The heritability of risk and age at onset of Parkinson's disease after accounting for known genetic risk factors. *Journal of human genetics* 2010; 55(4): 241-3.
7. Do CB, Tung JY, Dorfman E, et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS genetics* 2011; 7(6): e1002141.
8. Keller MF, Saad M, Bras J, et al. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Human molecular genetics* 2012; 21(22): 4996-5009.
9. Vekrellis K, Xilouri M, Emmanouilidou E, Rideout HJ, Stefanis L. Pathological roles of alpha-synuclein in neurological disorders. *Lancet neurology* 2011; 10(11): 1015-25.
10. Wong YC, Krainc D. Lysosomal trafficking defects link Parkinson's disease with Gaucher's disease. *Movement disorders : official journal of the Movement Disorder Society* 2016; 31(11): 1610-8.
11. Tayebi N, Callahan M, Madike V, et al. Gaucher disease and parkinsonism: a phenotypic and genotypic characterization. *Molecular genetics and metabolism* 2001; 73(4): 313-21.
12. Goker-Alpan O, Schiffmann R, LaMarca ME, Nussbaum RL, McInerney-Leo A, Sidransky E. Parkinsonism among Gaucher disease carriers. *Journal of medical genetics* 2004; 41(12): 937-40.
13. Sidransky E, Nalls MA, Aasly JO, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *The New England journal of medicine* 2009; 361(17): 1651-61.
14. Boustany RM. Lysosomal storage diseases--the horizon expands. *Nature reviews Neurology* 2013; 9(10): 583-98.
15. Filocamo M, Morrone A. Lysosomal storage disorders: molecular basis and laboratory testing. *Human genomics* 2011; 5(3): 156-69.
16. Warrel DA, Cox, T.M., Firth, J.D., Benz, C. Oxford Textbook of Medicine. Oxford: Oxford University Press; 2003.
17. Shachar T, Lo Bianco C, Recchia A, Wiessner C, Raas-Rothschild A, Futerman AH. Lysosomal storage disorders and Parkinson's disease: Gaucher disease and beyond. *Movement disorders : official journal of the Movement Disorder Society* 2011; 26(9): 1593-604.
18. Deng H, Xiu X, Jankovic J. Genetic convergence of Parkinson's disease and lysosomal storage disorders. *Molecular neurobiology* 2015; 51(3): 1554-68.
19. Perrett RM, Alexopoulou Z, Tofaris GK. The endosomal pathway in Parkinson's disease. *Molecular and cellular neurosciences*

2015; 66(Pt A): 21-8.
20. Gan-Or Z, Ozelius LJ, Bar-Shira A, et al. The p.L302P mutation in the lysosomal enzyme gene SMPD1 is a risk factor for Parkinson disease. *Neurology* 2013; 80(17): 1606-10.
21. Wu RM, Lin CH, Lin HI. The p.L302P mutation in the lysosomal enzyme gene SMPD1 is a risk factor for Parkinson disease. *Neurology* 2014; 82(3): 283.
22. Foo JN, Liany H, Bei JX, et al. Rare lysosomal enzyme gene SMPD1 variant (p.R591C) associates with Parkinson's disease. *Neurobiology of aging* 2013; 34(12): 2890. e13-5.
23. Gan-Or Z, Orr-Urtreger A, Alcalay RN, Bressman S, Giladi N, Rouleau GA. The emerging role of SMPD1 mutations in Parkinson's disease: Implications for future studies. *Parkinsonism & related disorders* 2015; 21(10): 1294-5.
24. Kluenemann HH, Nutt JG, Davis MY, Bird TD. Parkinsonism syndrome in heterozygotes for Niemann-Pick C1. *Journal of the neurological sciences* 2013; 335(1-2): 219-20.
25. Zech M, Nubling G, Castrop F, et al. Niemann-Pick C disease gene mutations and age-related neurodegenerative disorders. *PloS one* 2013; 8(12): e82879.
26. Clark LN, Chan R, Cheng R, et al. Gene-wise association of variants in four lysosomal storage disorder genes in neuropathologically confirmed Lewy body disease. *PloS one* 2015; 10(5): e0125204.
27. Winder-Rhodes SE, Garcia-Reitbock P, Ban M, et al. Genetic and pathological links between Parkinson's disease and the lysosomal disorder Sanfilippo syndrome. *Movement disorders : official journal of the Movement Disorder Society* 2012; 27(2): 312-5.
28. Jansen IE, Ye H, Heetveld S, et al. Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing. *Genome biology* 2017; 18(1): 22.
29. Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics* 2014; 46(9): 989-93.
30. Cuervo AM, Stefanis L, Fredenburg R, Lansbury PT, Sulzer D. Impaired degradation of mutant alpha-synuclein by chaperone-mediated autophagy. *Science (New York, NY)* 2004; 305(5688): 1292-5.
31. Vogiatzi T, Xilouri M, Vekrellis K, Stefanis L. Wild type alpha-synuclein is degraded by chaperone-mediated autophagy and macroautophagy in neuronal cells. *The Journal of biological chemistry* 2008; 283(35): 23542-56.
32. Lee HJ, Khoshaghideh F, Patel S, Lee SJ. Clearance of alpha-synuclein oligomeric intermediates via the lysosomal degradation pathway. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2004; 24(8): 1888-96.
33. Cooper AA, Gitler AD, Cashikar A, et al. Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. *Science (New York, NY)* 2006; 313(5785): 324-8.
34. Mazzulli JR, Xu YH, Sun Y, et al. Gaucher disease glucocerebrosidase and alpha-synuclein form a bidirectional pathogenic loop in synucleinopathies. *Cell* 2011; 146(1): 37-52.
35. Simon-Sanchez J, Heutink P, Gasser T. Variation in PARK10 is not associated with risk and age at onset of Parkinson's disease in large clinical cohorts. *Neurobiology of aging* 2015; 36(10): 2907.e13-7.
36. Jansen IE, Bras JM, Lesage S, et al. CHCHD2 and Parkinson's disease. *Lancet neurology*

2015; 14(7): 678-9.
37. Nichols WC, Kissell DK, Pankratz N, et al. Variation in GIGYF2 is not associated with Parkinson disease. *Neurology* 2009; 72(22): 1886-92.
38. Hofman A, Brusselle GG, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *European journal of epidemiology* 2015; 30(8): 661-708.
39. Parkinson Progression Marker Initiative The Parkinson Progression Marker Initiative (PPMI). *Progress in neurobiology* 2011; 95(4): 629-35.
40. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 2009; 25(14): 1754-60.
41. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 2010; 20(9): 1297-303.
42. Nalls MA, Bras J, Hernandez DG, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. *Neurobiology of aging* 2014.
43. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* 2010; 38(16): e164.
44. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM. A general framework for estimating the relative pathogenicity of human genetic variants. 2014; 46(3): 310-5.
45. Amendola LM, Dorschner MO, Robertson PD, et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome research* 2015; 25(3): 305-15.
46. Pringsheim T, Jette N, Frolkis A, Steeves TD. The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Movement disorders : official journal of the Movement Disorder Society* 2014; 29(13): 1583-90.
47. de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. *Lancet neurology* 2006; 5(6): 525-35.
48. Rana HQ, Balwani M, Bier L, Alcalay RN. Age-specific Parkinson disease risk in GBA mutation carriers: information for genetic counseling. *Genetics in medicine : official journal of the American College of Medical Genetics* 2013; 15(2): 146-9.
49. Trinh J, Guella I, Farrer MJ. Disease penetrance of late-onset parkinsonism: a meta-analysis. *JAMA neurology* 2014; 71(12): 1535-9.
50. Marder K, Wang Y, Alcalay RN, et al. Age-specific penetrance of LRRK2 G2019S in the Michael J. Fox Ashkenazi Jewish LRRK2 Consortium. *Neurology* 2015; 85(1): 89-95.
51. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics (Oxford, England)* 2012; 13(4): 762-75.
52. 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 (Oxford, England)* 2016; 17(1): 1-15.
53. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic acids research* 2015; 43(Database issue): D789-98.
54. Anheim M, Elbaz A, Lesage S, et al. Penetrance of Parkinson disease in glucocerebrosidase gene mutation carriers. *Neurology* 2012; 78(6): 417-20.

55. Pankratz N, Beecham GW, DeStefano AL, et al. Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Annals of neurology* 2012; 71(3): 370-84.
56. Duran R, Mencacci NE, Angeli AV, et al. The glucocerobrosidase E326K variant predisposes to Parkinson's disease, but does not cause Gaucher's disease. *Movement disorders : official journal of the Movement Disorder Society* 2013; 28(2): 232-6.
57. Zuk O, Schaffner SF, Samocha K, et al. Searching for missing heritability: designing rare variant association studies. *Proceedings of the National Academy of Sciences of the United States of America* 2014; 111(4): E455-64.
58. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature* 2015; 526(7571): 68-74.
59. Lill CM, Roehr JT, McQueen MB, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS genetics* 2012; 8(3): e1002548.
60. Cullen V, Lindfors M, Ng J, et al. Cathepsin D expression level affects alpha-synuclein processing, aggregation, and toxicity in vivo. *Molecular brain* 2009; 2: 5.
61. McGlinchey RP, Lee JC. Cysteine cathepsins are essential in lysosomal degradation of alpha-synuclein. *Proceedings of the National Academy of Sciences of the United States of America* 2015; 112(30): 9322-7.
62. Hamano K, Hayashi M, Shioda K, Fukatsu R, Mizutani S. Mechanisms of neurodegeneration in mucopolysaccharidoses II and IIIB: analysis of human brain tissue. *Acta neuropathologica* 2008; 115(5): 547-59.
63. Singleton AB, Farrer M, Johnson J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science (New York, NY)* 2003; 302(5646): 841.
64. Soldner F, Stelzer Y, Shivalila CS, et al. Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* 2016; 533(7601): 95-9.
65. Sardi SP, Clarke J, Kinnecom C, et al. CNS expression of glucocerebrosidase corrects alpha-synuclein pathology and memory in a mouse model of Gaucher-related synucleinopathy. *Proceedings of the National Academy of Sciences of the United States of America* 2011; 108(29): 12101-6.
66. Rothaug M, Zunke F, Mazzulli JR, et al. LIMP-2 expression is critical for beta-glucocerebrosidase activity and alpha-synuclein clearance. *Proceedings of the National Academy of Sciences of the United States of America* 2014; 111(43): 15573-8.
67. Migdalska-Richards A, Daly L, Bezard E, Schapira AH. Ambroxol effects in glucocerebrosidase and alpha-synuclein transgenic mice. *Annals of neurology* 2016; 80(5): 766-75.
68. Haelterman NA, Yoon WH, Sandoval H, Jaiswal M, Shulman JM, Bellen HJ. A mitocentric view of Parkinson's disease. *Annual review of neuroscience* 2014; 37: 137-59.
69. Mazzulli JR, Zunke F, Isacson O, Studer L, Krainc D. alpha-Synuclein-induced lysosomal dysfunction occurs through disruptions in protein trafficking in human midbrain synucleinopathy models. *Proceedings of the National Academy of Sciences of the United States of America* 2016; 113(7): 1931-6.
70. Alcalay RN, Levy OA, Waters CC, et al. Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations.

Brain : a journal of neurology 2015; 138(Pt 9): 2648-58.
71. Bras J, Verloes A, Schneider SA, Mole SE, Guerreiro RJ. Mutation of the parkinsonism gene ATP13A2 causes neuronal ceroid-lipofuscinosis. *Human molecular genetics* 2012; 21(12): 2646-50.
72. Escott-Price V, Nalls MA, Morris HR, et al. Polygenic risk of Parkinson disease is correlated with disease age at onset. *Annals of neurology* 2015; 77(4): 582-91.
73. Lubbe SJ, Escott-Price V, Gibbs JR, et al. Additional rare variant analysis in Parkinson's disease cases with and without known pathogenic mutations: evidence for oligogenic inheritance. *Human molecular genetics* 2016.
74. van Blitterswijk M, van Es MA, Hennekam EA, et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Human molecular genetics* 2012; 21(17): 3776-84.
75. Kenna KP, McLaughlin RL, Byrne S, et al. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. *Journal of medical genetics* 2013; 50(11): 776-83.
76. Cady J, Allred P, Bali T, et al. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. *Annals of neurology* 2015; 77(1): 100-13.
77. Gonzaga-Jauregui C, Harel T, Gambin T, et al. Exome Sequence Analysis Suggests that Genetic Burden Contributes to Phenotypic Variability and Complex Neuropathy. *Cell reports* 2015; 12(7): 1169-83.

SUPPLEMENTAL DATA

Table 1 - part 1. LSD genes and variants in the discovery and replication cohorts.

		IPDGC			PPMI		NeuroX	
Disease	**Gene**	**Nonsyn**	**CADD**	**LoF**	**Nonsyn**	**CADD**	**Nonsyn**	**CADD**
Aspartylglucosaminuria	*AGA*	13	10	2	4	3	7	6
Metachromatic Leukodystrophy	*ARSA*	5	5	1	3	2	5	5
Maroteaux-Lamy disease	*ARSB*	11	10	0	8	6	11	9
Farber Lipogranulomatosis	*ASAH1*	20	17	2	10	10	10	8
Kufor-Rakeb syndrome	*ATP13A2*	24	18	1	13	7	24	16
Neuronal Ceroid Lipofuscinosis (CLN3)	*CLN3*	18	17	3	4	4	5	4
Neuronal Ceroid Lipofuscinosis (CLN5)	*CLN5*	0	0	0	0	0	6	5
Neuronal Ceroid Lipofuscinosis (CLN6)	*CLN6*	10	7	1	6	5	10	9
Neuronal Ceroid Lipofuscinosis (CLN8)	*CLN8*	9	4	0	0	0	10	5
Cystinosis	*CTNS*	15	13	0	5	3	8	5
Galactosialidosis	*CTSA*	14	11	0	3	1	4	2
Neuronal Ceroid Lipofuscinosis (CLN10)	*CTSD*	7	4	0	2	1	5	3
Neuronal Ceroid Lipofuscinosis (CLN13)	*CTSF*	12	10	0	4	3	3	3
Pycnodysostosis	*CTSK*	6	5	0	3	2	3	3
Neuronal Ceroid Lipofuscinosis (CLN4B)	*DNAJC5*	5	5	1	0	0	1	1
Fucosidosis	*FUCA1*	16	13	0	7	5	5	4
Pompe disease	*GAA*	15	10	0	14	10	11	8
Krabbe disease	*GALC*	37	31	8	4	3	8	6
Morquio A disease	*GALNS*	22	14	0	8	2	10	6
Gaucher disease	*GBA*	42	33	4	8	7	11	9

Table 1 - part 2. LSD genes and variants in the discovery and replication cohorts.

		IPDGC			PPMI		NeuroX	
Disease	**Gene**	**Nonsyn**	**CADD**	**LoF**	**Nonsyn**	**CADD**	**Nonsyn**	**CADD**
Fabry disease	*GLA*	9	7	0	1	1	6	5
GM1-Gangliosidosis/Morquio B	*GLB1*	8	4	0	9	7	15	11
GM2-Gangliosidosis	*GM2A*	1	1	0	0	0	1	1
I-Cell disease	*GNPTAB*	40	32	13	10	8	15	11
Sanfilippo D syndrome	*GNS*	20	11	3	1	1	2	2
Neuronal Ceroid Lipofuscinosis (CLN11)	*GRN*	20	13	0	5	4	14	8
Sly disease	*GUSB*	18	11	0	5	3	3	2
Tay-Sachs disease	*HEXA*	20	18	1	4	3	12	11
Sandhoff disease	*HEXB*	10	7	2	2	1	10	7
Sanfilippo C syndrome	*HGSNAT*	22	15	1	7	6	7	4
Mucopolysaccharidosis Type IX	*HYAL1*	14	9	0	3	0	8	3
Hunter syndrome	*IDS*	9	8	2	0	0	8	5
Hurler syndrome	*IDUA*	8	4	0	2	1	12	11
Neuronal Ceroid Lipofuscinosis (CLN14)	*KCTD7*	4	3	0	1	1	3	2
Danon disease	*LAMP2*	9	7	1	1	1	4	3
Wolman disease	*LIPA*	15	10	2	3	2	3	2
Alpha-Mannosidosis	*MAN2B1*	12	11	2	14	10	22	17
Beta-Mannosidosis	*MANBA*	18	15	4	6	5	10	7
Mucolipidosis Type IV	*MCOLN1*	19	14	0	6	5	8	6
Neuronal Ceroid Lipofuscinosis (CLN7)	*MFSD8*	19	15	2	4	2	14	11
Schindler Disease/Kanzaki disease	*NAGA*	9	8	0	7	7	6	6

Sanfilippo B syndrome	*NAGLU*	10	9	3	2	1	5	3
Sialidosis	*NEU1*	0	0	0	3	2	5	5
Niemann-Pick Disease Type C1	*NPC1*	44	35	5	12	10	26	18
Niemann-Pick Disease Type C2	*NPC2*	2	2	0	3	2	6	5
Neuronal Ceroid Lipofuscinosis (CLN1)	*PPT1*	9	7	1	2	1	5	5
Sphingolipid-activator deficiency	*PSAP*	22	16	0	3	3	13	11
Action mycolonus-renal failure syndrome	*SCARB2*	10	7	0	3	0	10	7
Sanfilippo A syndrome	*SGSH*	12	9	0	4	3	10	7
Salla disease	*SLC17A5*	18	17	1	5	5	8	8
Niemann-Pick disease Type A/B	*SMPD1*	27	23	1	7	6	17	11
GM3-Gangliosidosis	*ST3GAL5*	11	11	1	2	2	3	2
Multiple Sulfatase Deficiency	*SUMF1*	0	0	0	2	2	3	2
Neuronal Ceroid Lipofuscinosis (CLN2)	*TPP1*	16	13	1	11	8	16	12

The number of variants (MAF < 3%) in each LSD gene is shown for the IPDGC discovery cohort and replication cohorts (PPMI and NeuroX), including total number of nonsynonymous variants (nonsyn) and likely damaging variants (CADD). For the IPDGC discovery dataset, the number of loss-of-function (LoF) variants is also indicated.

Table 2. LSD variants in IPDGC dataset.

Due to the size of this table, it is only available upon request (ie.jansen@vumc.nl)