Establishing the role of rare coding variants in known Parkinson's disease risk loci

Iris E. Jansen a,b, J. Raphael Gibbs c, Mike A. Nalls c,d, T. Ryan Price e, Steven Lubbe f, Jeroen van Rooij g, André G. Uitterlinden g,h,i, Robert Kraaij g,h,i, Nigel M. Williams j, Alexis Brice k,l, John Hardy m, Nicholas W. Wood n, Huw R. Morris n, Thomas Gasser o, Andrew B. Singleton c, Peter Heutink b,o, Manu Sharma o,*,1, for International Parkinson's Disease Genomics Consortium1

a Department of Clinical Genetics, VU University Medical Center, Amsterdam, the Netherlands
b Genome Biology of Neurodegenerative Diseases, German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany
c Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA
d Data Tecnica International, Glen Echo, MD, USA
e University California Irvine, Irvine, CA, USA
f Ken and Ruth Davee Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
g Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands
h Netherlands Consortium for Healthy Ageing (NCHA), Rotterdam, the Netherlands
i Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands
j MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales, UK
k Inserm U1127, CNRS UMR7225, Sorbonne Universités, UPMC Univ Paris 06, UMR_S1127, Institut du Cerveau et de la Moelle épinière, Paris, France
l Assistance Publique Hôpitaux de Paris, Hôpital de la Salpêtrière, Département de Génétique et Cytogénétique, Paris, France m Reta Lila Weston Institute, University College London, London, UK
n Department of Clinical Neuroscience, UCL Institute of Neurology, London, UK
o Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tübingen, Tübingen, Germany

Abstract

Many common genetic factors have been identified to contribute to Parkinson's disease (PD) suscepti-bility, improving our understanding of the related underlying biological mechanisms. The involvement of rarer variants in these loci has been poorly studied. Using International Parkinson's Disease Genomics Consortium data sets, we performed a comprehensive study to determine the impact of rare variants in 23 previously published genome-wide association studies (GWAS) loci in PD. We applied Prix fixe to select the putative causal genes underneath the GWAS peaks, which was based on underlying functional similarities. The Sequence Kernel Association Test was used to analyze the joint effect of rare, common, or both types of variants on PD susceptibility. All genes were tested simultaneously as a gene set and each gene individually. We observed a moderate association of common variants, confirming the involvement of the known PD risk loci within our genetic data sets. Focusing on rare variants, we identified additional association signals for LRRK2, STBD1, and SPATA19. Our study suggests an involvement of rare variants within several putatively causal genes underneath previously identified PD GWAS peaks.

1. Introduction

Genetic factors play an important role in Parkinson's disease (PD) pathogenesis. In addition to the discovery of rare variants using family based linkage studies, resulting in the identification of, for example, SNCA, LRRK2, parkin, DJ-1, PINK1, and VPS35, numerous genome-wide association studies (GWAS) have shown that com-mon genetic variants increase PD risk (Bras et al., 2015). The most recent and largest PD association study (Nalls et al., 2014) identified over 20 common risk variants, confirming many previously associated risk factors.

Nevertheless, heritability estimates indicate that additional genetic risk factors remain to be discovered since a relatively large fraction of PD heritability cannot be explained by known PD risk loci or Mendelian genes (Do et al., 2011; Keller et al., 2012; Pihlstrom and Toft, 2011). GWAS approaches are primarily designed to identify common risk variants by the usage of geno-typing arrays. However, emerging evidence suggests that rare var-iants may explain part of the missing heritability (Manolio et al., 2009; Zuk et al., 2014). Rare variants in protein-coding regions are more likely to affect the function of a gene than common var-iants, which tag the causal variants via linkage disequilibrium (LD) and are often located in noncoding regions of the genome (Nelson et al., 2012; Tennessen et al., 2012). Therefore, rare variants might be of more importance to complex diseases than predicted by the Common DiseaseeCommon Variant hypothesis (Botstein and Risch, 2003; Lander, 1996; Pritchard and Cox, 2002; Sharma et al., 2014). In contrast to GWAS, exome sequencing studies aim at systemati-cally analyzing coding regions of the genome to identify causal variants in complex diseases (Kiezun et al., 2012). Exome studies have been proven to be effective for studying familial diseases (Bamshad et al., 2011), but an increasing number of applications for population-based studies have been developed (Cirulli et al., 2015; Purcell et al., 2014).

In PD, multiple genes have been shown to harbor both common and rare variants, which contribute to disease pathogenesis. SNCA and LRRK2 both contain PD riskeassociated rare variants with Mendelian effects as common variants that increase PD risk in sporadic patients (Edwards et al., 2010; Nalls et al., 2011, 2014; Paisan-Ruiz et al., 2004; Polymeropoulos et al., 1997; Simon-Sanchez et al., 2009; Zimprich et al., 2004). GBA, for which an as-sociation was first seen in families with Gaucher's disease and parkinsonism (Goker-Alpan et al., 2004), is furthermore shown to play a role in PD by both rare coding variants and common risk variants (Do et al., 2011; Nalls et al., 2014; Pankratz et al., 2012). Thus, we hypothesize that rare coding variants in the known risk loci for sporadic PD are involved in the genetic etiology of PD. The combined effect of rare variants within recently identified PD risk loci will likely explain an additional portion of PD heritability. We aim to assess this hypothesis by determining the genetic burden of rare variants in the PD risk loci using 2 exome cohorts of the In-ternational Parkinson's Disease Genomics Consortium (IPDGC).

2. Methods

2.1. Subjects

All PD cases included in this study have given written informed consent. Relevant local ethical committees for medical research approved involvement in genetic studies. The PD patients were diagnosed using the UK Brain Bank criteria (Hughes et al., 1992).

2.2. Whole exome sequencing data set. The whole exome sequencing (WES) data set includes 1167 PD cases and 1685 controls (post QC) of European ancestry. The PD patients

have a tendency toward a young age of onset with an average of 41.2 years (SD ¼ 10.9). 1201 controls originate from the Rotterdam Study, version 1 (RSX1), as we merged the IPGDC WES data with the RSX1 WES data (Hofman et al., 2015). The samples were sequenced in different batches with 2 exome capture kits: EZ Exome Library v2.0 (Roche/Nimblegen) and Truseq Exome Enrich-ment Kit targeting 44.1 Mb and 62 Mb, respectively (Supplementary Table 1). To account for putative technical differences between the different capture kits, we only considered variants that were tar-geted by both capture protocols and included preQC individual sample missingness (as a reference to sequencing coverage) as covariates during all genetic analyses.

On average, 94.4% of the exome was covered for at least 10x. The 100-bp paired-end reads were sequenced on a HiSeq2000 and aligned to the human reference genome (build hg19) using Barrow Wheeler Aligner (Li and Durbin, 2009). Genome Analysis Toolkit (McKenna et al., 2010) called single nucleotide variants and small insertions/deletions (indels) for each sample, resulting in individual gVCF files. Genotypes of all IPDGC and RSX1 exome samples were then jointly called and recalibrated, allowing to merge the distinct WES data sets in a correct manner. Standard Genome Analysis Toolkit filter steps were applied, together with a minimum geno-type quality Phred score of 20 and depth of 8, to only select high-quality variants. Only biallelic calls were considered that were located in regions targeted by both capture kits. Supplementary Table 2 reports the exons that have been excluded due to insuffi-cient coverage within one of the exome capture protocols.

2.3. NeuroX data set

The NeuroX data set encompasses 6801 PD cases and 5970 controls (post QC) of European ancestry. Overlapping samples with the WES data set were excluded. The average age of onset of the PD patients is 63.0 years (SD ¼ 12.4). The Exome NeuroX array (Nalls et al., 2015) was used consisting of w240,000 exonic variants standard to the Illumina HumanExome array v1.1 and w25,000 variants focused on neurologic and neurodegenerative diseases.

2.4. Quality procedures

For individual QC in both the WES and the NeuroX data sets, samples were removed when showing gender ambiguity, dubious heterozygosity/genotype calls, evidence of relatedness, or being a population outlier. The latter 2 were calculated with LD-pruned common variants. Variant QC procedures were slightly different for the 2 different data sets. For the WES data set, variants passed QC when having a minimum call rate >85% and being in Har-dyeWeinberg equilibrium (HWE p-values >1e-8 based on con-trols). For the NeuroX data set, variants were excluded for subsequent analyses with a minimum call rate <95%, an HWE p-value <1e-6, or with significant differences in missingness rate between cases and controls.

2.5. Causal gene selection within PD risk loci

Based on the most recent and largest GWAS (Lill et al., 2012; Nalls et al., 2014), we selected 26 loci containing at least 1 top single nucleotide polymorphism (SNP) nominated in meta-analysis with p < 5.00e-08 (as reported by pdgene.org). The published SNPs associated with PD are not the causal variants but rather tag the unknown causal variants with which they are in LD. As the causal variant (and therefore also the related gene) has not been deter-mined for most of the PD risk loci, we explored the involvement of rare variants in PD susceptibility by using the Prix fixe strategy, which selects 1 gene per locus based on functional similarities of genes within the LD-blocks from the different loci.

The functional similarity is defined as the degree of shared biological function and is determined by overlapping biological features such as protein domains, transcription factor binding sites, gene expression, phylogenetic profiles, and protein-protein in-teractions. Based on these features, cofunction networks are generated which connect genes that are likely to share the same underlying molecular pathway. Genes that are strongly connected to other candidate genes obtain a higher Prix fixe score and therefore prioritized as causal gene. As this approach is based on genome-wide data sets and is not performed with disease-related biological assumptions, the Prix fixe strategy aims to prioritize

Table 1
Selected set of genes

Polymorphism	Location (hg19)	*P*-value	Seeding SNP	Prix fixe gene
rs71628662	chr1:155359992	6.86×10^{-28}	NA	
rs823118	chr1:205723572	1.96×10^{-16}	rs823114	*RAB7L1*
rs10797576	chr1:232664611	1.76×10^{-10}	rs2182431	*SIPA1L2*
rs6430538	chr2:135539967	3.35×10^{-19}	rs6430538	*ACMSD*
rs1955337	chr2:169129145	1.67×10^{-20}	rs2390669	*STK39*
rs12637471	chr3:182762437	5.38×10^{-22}	rs12637471	*LAMP3*
rs11724635	chr4:15737101	4.26×10^{-17}	rs11724635	*FBXL5*
rs6812193	chr4:77198986	1.85×10^{-11}	rs6812193	*STBD1*
rs356182	chr4:90626111	1.85×10^{-82}	rs356219	*SNCA*
rs34311866	chr4:951947	6.0×10^{-41}	rs748483	*MFSD7*
rs9275326	chr6:32666660	5.81×10^{-13}	rs9275311	*HLA-DR*
rs199347	chr7:23293746	5.62×10^{-14}	rs199347	*GPNMB*
rs591323	chr8:16697091	3.17×10^{-8}	NA	
rs117896735	chr10:121536327	1.21×10^{-11}	rs10886515	*RGS10*
rs329648	chr11:133765367	8.05×10^{-12}	rs329648	*SPATA19*
rs3793947	chr11:83544472	2.59×10^{-08}	rs1400313	*DLG2*
rs11060180	chr12:123303586	3.08×10^{-11}	rs11060180	*HIP1R*
rs76904798	chr12:40614434	4.86×10^{-14}	rs2708435	*LRRK2*
rs7155501	chr14:55347827	1.25×10^{-10}	rs2878174	*LGALS3*
rs1555399	chr14:67984370	5.70×10^{-16}	rs7155830	*ARG2*
rs2414739	chr15:61994134	3.59×10^{-12}	NA	
rs14235	chr16:31121793	3.63×10^{-12}	rs14235	*PRSS8*
rs17649553	chr17:43994648	6.11×10^{-49}	rs17649553	*MAPT*
rs12456492	chr18:40673380	2.15×10^{-11}	rs12456492	*RIT2*
rs62120679	chr19:2363319	2.52×10^{-09}	rs2074546	*PLEKHJ1*
rs55785911	chr20:3153503	3.30×10^{-10}	rs2295545	*AVP*

P-value, meta p-value as reported on pdegene.org; seeding SNP, input SNP for Prix fixe software; prix fixe gene, genes selected based on underlying functional simi-larities, which is determined by overlapping biological features such as protein domains, transcription factor binding sites, gene expression, phylogenietic profiles and protein-protein interactions. genes without the usual text mining bias caused by literature-based knowledge (Edwards et al., 2011).
The most significantly associated SNPs from the recent meta-analysis by Nalls et al. (2014) were used as seeding SNPs to define the LD region per PD locus. If an SNP was not applicable to be used as seeding SNP (not present in either the current dbSNP 137 or HapMap public resources), the next strongest associated SNP or an SNP in high LD (r2 >0.8) within the same locus was used as a seed. We were unable to define a legitimate seeding

SNP for 3 loci (rs71628662, rs591323, rs2414739). LD-regions were based on the CEU phase III population with a minimal R2 of 0.5. The final Prix fixe gene set consists of 23 genes for downstream analyses (Table 1).

2.6. Variant selection

To enrich both genetic data sets for deleterious variants, we selected multiple subsets of variants, differing in the method and stringency to select pathogenic variants. Based on variant annota-tion with annotate variation (Wang et al., 2010), 3 distinct subsets of variants were created, including: (1) all exonic variants (disruptive, splicing, [non]synonymous, and [non]frameshift indels); (2) amino acidechanging variants (same as previous except for synonymous); and (3) amino acidechanging (AAchanging) variants that are pre-dicted to be deleterious. The latter subset includes variants that are predicted to be pathogenic (Combined Annotation Dependent Depletion [CADD] score 12.37 (Amendola et al., 2015)) by Com-bined Annotation Dependent Depletion, v1 (Kircher et al., 2014). Fig. 1 displays a workflow of the classification of the different variant subsets. The exonic subset was exclusively tested for the gene set analysis to determine the involvement of the known PD risk loci in the WES and NeuroX data set.

On average, 94.4% of the exome was covered for at least 10x. The 100-bp paired-end reads were sequenced on a HiSeq2000 and aligned to the human reference genome (build hg19) using Barrow Wheeler Aligner (Li and Durbin, 2009). Genome Analysis Toolkit (McKenna et al., 2010) called single nucleotide variants and small insertions/deletions (indels) for each sample, resulting in individual gVCF files. Genotypes of all IPDGC and RSX1 exome samples were then jointly called and recalibrated, allowing to merge the distinct WES data sets in a correct manner. Standard Genome Analysis Toolkit filter steps were applied, together with a minimum geno-type quality Phred score of 20 and depth of 8, to only select high-quality variants. Only biallelic calls were considered that were located in regions targeted by both capture kits. Supplementary Table 2 reports the exons that have been excluded due to insuffi-cient coverage within one of the exome capture protocols.

2.7. Variant aggregation analysis

The Sequence Kernel Association Test (SKAT) (Ionita-Laza et al., 2013; Wu et al., 2011) was used to perform burden analyses. The MAF threshold, separating the rare and common variants, was based on the total sample size using the formula (T ¼ 1/(O(2n)) suggested by SKAT (Ionita-Laza et al., 2013), therefore resulting in the MAF thresholds of 0.013 and 0.006 for the WES data set and NeuroX data set, respectively. We performed polygenetic burden analyses for exclusively rare variants, exclusively common variants and both types of variants together. The common variants were pruned (PLINK (Purcell et al., 2007) indep settings 50 5 1.5) aiming to only consider independent variants in our genetic analyses. For the gene sets, we performed a 2-sided SKAT test allowing variants within a gene-set to have different directions and magnitudes of effects, which is in concordance with both damaging and protective effect estimates observed for the 23 published PD loci. To test individual genes, we performed a 1-sided burden test, as we hy-pothesized that variants in individual genes are likely to have the same direction of effect. We also performed a 2-sided SKAT analysis per gene in case we were interested which genes are driven an observed rare variant association in the total gene set.

To correct for confounding factors (e.g., population stratification and technical artifacts), we included 20 multidimensional scaling components, gender, individual missingness rate

preQC (as a reference to the individual WES coverage) and gender for the WES data set. As the NeuroX data set is more homogeneous, we corrected for the first 4 MDS components and gender. Empirical p-values were calculated for sig-nificant sample results ($p < 0.05$). For the gene set analysis, the original sample p-value of the gene set of interest was compared to p values of 1000 randomly drawn gene sets of the same size. For the individual gene associations, empirical p-values were calculated using the resampling method implemented by SKAT, by 10,000 permutations of the affection status. Empirical p-values are calculated by (n1þ1)/(nþ1), where n1 ¼ the number of resampling p-values smaller than the original sample p-value and n ¼ the number of resampling.

2.8. Power calculations

We estimated the power of our study design to detect rare variant associations. Supplementary Table 3 displays the parame-ters that were chosen for the calculations. For both data sets, the PD prevalence was set to 0.0057 (Pringsheim et al., 2014). As approx-imately half of the loci in PDgene.org have an odds ratio below 1, the percentage protective effect was set to 50%. A thousand simulations (a ¼ 0.025) were performed on a haplotype matrix of SKAT, mimicking linkage disequilibrium structure of European ancestry, comprising 10,000 haplotypes over 200 kb regions.

3. Results

3.1. WES and rare variants

First, we analyzed the WES data set as it represents all exonic variants, of which the study design has 65% power to detect a rare variant association signal considering individual genes. Testing the aggregated effect of grouped variants within a gene set has the potential to increase power. Supplementary Table 4 shows the re-sults of the gene set analyses in the WES data set. Common exonic variants are moderately associated to PD. The nominal p-value is significant, but the empirical p-value exceeds 0.05. Although we anticipated a significant association of common variants, we attri-bute the moderate association to a relatively low sample size (compared to the original GWAS) and the selection of genes (by Prix fixe) with variants in moderate LD with the original highest SNP.

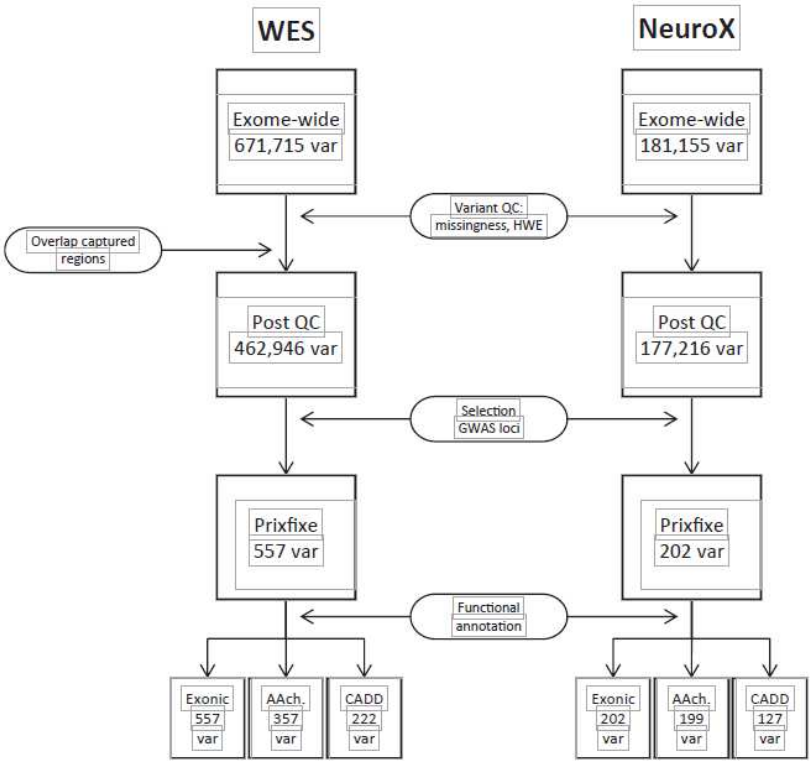

Fig. 1. Flowchart of variant subset classification. The variants undergo multiple analyses procedures, including quality control, selection of variants within PD loci, and functional annotation. Each genetic data set (WES and NeuroX) is tested for 3 different variant categories, differing in causal gene selection approach and functionality of variant.

The gene set association is absent when focusing on the common amino acid changing and CADD variants, which is probably due to a decrease in power as the number of variants drops.

No rare variant or common and rare variant associations were observed for the gene set in either of the functional variant cate-gories (nominal p ? 0.223; Supplementary Table 4). An alternative approach to study the putative rare variant associations is to test each gene individually within the gene set. Table 2 displays the 3 strongest associated genes per functional category. Using the AAchanging variants category, we observed a significant association for STBD1 (p ¼ 0.046).

3.2. NeuroX and rare variants

The NeuroX data set contains previously identified exonic vari-ants, of which a large proportion is rare (Nalls et al., 2015). In contrast to the WES data sets, our NeuroX cohort has enough power (estimated at 96%), due to the larger sample size (6804 cases 5970 controls), to detect a rare variant association signal. Similarly, to the WES data set, a moderate common variant association is detected (nominal p ¼ 0.031). In contrast to the WES data set, we do observe significant associations of the gene set with PD, even when only considering rare variants (AAchanging ¼ 0.007; CADD ¼ 0.002; Supplementary Table 5a).

To discover whether specific genes drive this observed rare variant association as observed in our cohort, the variants were grouped per gene and 2-sided tested for their association to PD.

LRRK2 is the gene driving the association observed in the total gene sets (Supplementary Table 6). Focusing on the CADD subset, this association (nominal p ¼ 5.17 ? 10?13) is considerably stron-ger than the second most significant SPATA19 (nominal p ¼ 0.050). The NeuroX array custom content is primarily driven by neuro-degenerative diseases; therefore,

NeuroX chip biases toward capture of in-depth genetic variability within genes, which are known to cause disease pathogenesis (Tennessen et al., 2012). Likewise, NeuroX harbors many variants of the known PD genes. For example, NeuroX contains 32 harmful (predicted by CADD) LRRK2 variants, whereas only 2 harmful variants are present for SPATA19.The variants in LRRK2 are overrepresented and biasing the results of the total gene sets. We, therefore, performed the same gene set analyses on the NeuroX data set excluding the variants of LRRK2 (Supplementary Table 5b), resulting in the absence of a rare variant association in the NeuroX data set (nominal p 0.28). This suggests that the previously observed association of rare variants within the total gene set to PD was solely driven by LRRK2.

The 2-sided SKAT analysis per gene aimed at the discovery of genes driving the rare variant association in the total gene set. Next, we were interested to explore the genetic burden of rare variants for each gene individually when assuming all rare variants to have the same direction of effect (1-sided BURDEN test). Table 3 shows again that LRRK2 is the strongest associated gene. Furthermore, SPATA19 (p ¼ 0.017) is significantly associated when specifically considering rare CADD variants.

3.3. Directionality of effect

We further explored the significant individual association signals (empirical $p < 0.05$) for LRRK2 and STBD1, and SPATA19.By focusing on the variant level, we aimed to comprehend the di-rection of effect estimates. LRRK2 showed a significant burden of 32 rare damaging variants in the NeuroX data set. Single-marker association analysis of LRRK2 variants revealed that the observed association (p ¼ 3.17 10 13) is attributed to the p.G2019S (rs34637584), the most common cause of monogenetic forms of PD. Interestingly, this particular variant was present in 78 cases (MAF ¼ 0.006). Performing the rare variant aggregation test on 31 pathogenic LRRK2 variants, excluding p.G2019S, resulted in no association (p ¼ 0.98) to PD, and thus suggesting that the observed rare variant association in LRRK2 was solely driven by the p.G2019S variant. As this variant is only present in 7 cases in the WES data set (MAF ¼ 0.003) with a single-marker p-value of 0.002 (LRRK2 mutations generally observed in late-onset PD), it explains the discrepancy of results for LRRK2 locus as observed in the WES data set, whereas it showed a strong association in the NeuroX data set.

In addition to the rare variant association test in LRRK2,we explored the presence of the previously published common LRRK2 haplotype with a protective effect of 3 exonic variants (p.N551K-p.R1398H-p.K1423K) (Ross et al., 2011). p.K1423K is not included in the NeuroX genotyping array but is in high linkage disequilibrium (r2 ¼ 1.00) with p.R1398H. We therefore tested the p.N551K-p.R1398H (G-A) haplotype and confirmed the protective effect (OR ¼ 0.89, p ¼ 0.027) of this haplotype for the PD cases, showing a minor haplotype frequency of 6.2% in cases and 6.9% in controls. All 3 variants were detected in the WES data set, allowing to test the full haplotype (G-A-A). Although the haplotype association was not significant in the WES data set (OR ¼ 0.81, p ¼ 0.223), the trend of effect is similar with a minor haplotype frequency of 7.0% in cases and 7.5% in controls. The smaller sample size of the WES data set is a plausible reason for not obtaining a significant association.

Next, the WES-based STBD1 and NeuroX-based SPATA19 were investigated for their variant frequencies. Single-marker associa-tion analysis showed no significant results for the 8 variants within STBD1. It therefore appears that the observed rare variant associa-tion is not caused by 1 exclusive variant but is rather the effect of multiple rare variants. Seven of the 8 variants are control-specificas they are only present in 10 control individuals. In contrast,

only 1 variant is present in a single case. The direction of effect of the variants that are generating the STBD1 gene association is therefore implied to be protective. The significant gene-based association for SPATA19 is relatively strong considering that it is driven by only 2 CADD variants that are present in 7 cases and 0 controls. The absence of SPATA19 CADD variants in controls suggests that the association signal is damaging.

Table 2
Gene-based rare variant association results for WES data set

Variant type	Gene	*p*-value (emp)	n variants	MAF cases (%)	MAF controls (%)
AAchanging	*STBD1*	**0.018 (0.046)**	8	0.05	0.32
	HIP1R	0.082	20	0.61	0.53
	STK39	0.126	4	0.20	0.00
CADD	*STBD1*	0.105	5	0.05	0.16
	SPATA19	0.122	4	0.19	0.06
	GPNMB	0.141	18	0.85	0.92

p-value, theoretical *p*-value; (emp.), emperical *p*-value calculated by comparison to 10,000 permutations of affection status; AAchanging, amino acid changing variants; CADD, variants predicted pathogenic; WES, whole exome sequencing.
p-values in bold are statistically significant.

Table 3
Gene-based rare variant association results for neuroX data set

Variant type	Gene	*p*-value (emp)	*n* variants	MAF cases (%)	MAF controls (%)
AAchanging	*LRRK2*	**0.0004 (0.0005)**	48	1.70	1.13
	RIT2	0.051	2	0.00	0.03
	PRSS8	0.098	1	0.04	0.01
CADD	*LRRK2*	**0.0003 (0.0005)**	32	1.38	0.86
	SPATA19	**0.014 (0.017)**	2	0.05	0.00
	RIT2	0.051	2	0.00	0.03

p-value, theoretical *p*-value; (emp.), emperical *p*-value calculated by comparison to 10,000 permutations of affection status; AAchanging, amino acid changing variants; CADD, variants predicted pathogenic.
p-values in bold are statistically significant.

4. Discussion

To establish the influence of rare variants in sporadic PD risk loci, we explored 2 independent PD data sets (WES and NeuroX) enriched for coding rare variants. We used the Prix fixe strategy to select the most likely causal genes underlying the PD loci peaks, which is based on overlapping biological functional similarities. We tested both the effect of rare variants in the gene sets at once, as each gene individually. Aggregating variants simultaneously across a set of genes has the potential to increase power to detect an as-sociation signal, given that the selected genes are enriched for a group of genes that are genuinely involved in the disease pathogenesis.

The average age of onset within the cases of the WES data set (w41 years) is 20 years younger than in the meta-analysis of the most recent PD GWAS (w61 years) where the PD risk loci were based on. As some rare genetic risk factors (DJ-1, parkin

PINK1)(Bras et al., 2015) are specific for young onset PD (YOPD), we acknowl-edge the putative existence of YOPD-specific common genetic risk factors within the WES data set. However, risk factors related to late-onset sporadic PD might also play a role in YOPD. PD risk loci, such as SNCA and GBA (Klein and Westenberger, 2012; Nalls et al., 2014) overlap between late and young onset. We therefore expect that our WES data set is an adequate data set to study the rare exonic variants in PD risk loci. Furthermore, YOPD is often geneti-cally explained through rare variants (Bras et al., 2015). The YOPD patient group in the WES data set could therefore be enriched for cases which are genetically influenced by rare variants, possibly increasing the likelihood of detecting rare variant association.

Using gene set approach in the WES data set, we did not detect a burden of rare variants when comparing PD subjects to controls. However, it is undetermined whether the absence of a rare variant association is genuine or due to insufficient power. A genuine rare variant association might furthermore be impeded by the gene set composition. By using Prix fixe, we increase the likelihood of selecting the truly involved PD genes underneath the known PD risk loci, yet unrelated genes might still be included, possibly diluting an association signal. In contrast, with the gene-based as-sociation test for the genes selected with the Prix fixe strategy, we observed a rare variant association for STBD1, implying that rare variants in this gene could decrease the risk to develop PD. STBD1 has its function in lysosomal-mediated autophagy to specifically guide glycogen to lysosomes for sequestration and degeneration (Jiang et al., 2011). It, therefore, seems that variants in STBD1 could have beneficial effects for the removal of glycogen. The lysosomal-mediated autophagy has been implied to be involved in PD through the association of multiple genes, such as LRRK2, ATP13A2, and GBA (Trinh and Farrer, 2013). However, the involvement of STBD1 in PD pathogenesis has to be carefully considered, as we currently did not have an adequate independent data set to replicate the associationthat was generally based on singletons. The NeuroX genotyping array typically includes variants that have been observed in previ-ous data sets, minimizing the probability to detect similar single-tons with an extremely low minor allele frequency. Only 3 of the 8 STBD1 variants of the WES data set were present within the NeuroX data set reducing the power to detect the single gene association. Hence, further genetic validation studies are warranted to establish the role of STBD1 in PD. Once a legitimate replication is realized, functional assays on lysosomal-mediated autophagy should further decipher the contribution of STBD1, preferably in relation to well-established PD genes.

We detected a strong association of rare variants within the gene sets for the NeuroX data set. However, subsequent analyses showed that these associations were dominated by LRRK2 variants. Asso-ciation analysis on variant level revealed that the LRRK2 gene signal was driven by the known p.G2019S variant. This observation highlights the importance of cataloguing the individual rare vari-ants to fully resolve the impact of rare variants in disease suscep-tibility for PD. As shown for the LRRK2 association and even the total gene set association, it is driven by only 1 variant, which also could have been detected with the performance of a simple single-marker association test. Besides the pathogenic association signal of rare variant p.G2019S, we observed a significant protective effect of a previously published common haplotype (Ross et al., 2011). This observation supports the

theory that other variants with opposite effects could interact and potentially influence the penetrance of pathogenic LRRK2 variants, such as p.G2019S. Besides LRRK2,we furthermore detected a NeuroX-based burden of rare CADD vari-ants for SPATA19 that increases PD risk (p ¼ 0.017). This association signal is relatively strong considering that it is driven by only 2 CADD variants that are present in 7 cases and 0 controls. As SPATA19 is involved in spermatogenesis (Nourashrafeddin et al., 2014) and the GTEx portal displays specific high expression for the testis, it diminishes the likelihood that defects of this gene would contribute to neurodegeneration. Further genetic and functional studies are warranted to decipher a role of this gene in PD.

In contrast to selecting the physically closest gene to the strongest SNP within each PD locus, we followed a comprehensive strategy to define true causal gene, which is based on biological similarities. As we expect that only 1 gene per locus is the true causal gene, we did not define a gene set including all the genes underneath the GWAS loci assuming that the overrepresentation of noncausal genes would dilute a putative association signal. We acknowledge that the ultimate strategy to test the effect of rare variants in the PD loci would be to sequence all genes in a large cohort and test the effect of rare variants in each gene individually. Furthermore, sequencing rather than genotyping will define novel rare variants and contribute to cataloguing the influence of rare variants underneath the PD risk loci. Acknowledging these caveats, our study suggests for the first time that, apart from LRRK2, SNCA, and GBA, other common PD risk loci might harbor rare variants that contribute to PD risk.

Disclosure statement

Mike A. Nalls' participation is supported by a consulting contract between Data Tecnica International and the National Institute on Aging, NIH, Bethesda, MD, USA. As a possible conflict of interest, Dr Nalls also consults for Illumina Inc, the Michael J. Fox Foundation, and University of California Healthcare among others.

Acknowledgements

This work was supported in part by the Prinses Beatrix Spier-fonds (IEJ and PH). MS is supported by the Michael J Fox Foundation, USA. TG and PH were supported by the Federal Ministry of Educa-tion and Research (BMBF) under the grant numbers 031A430A (TG) and 031A430D (PH) (e:Med Module II).

This work was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services; project ZO1 AG000949.

The generation and management of the exome sequencing data for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. The Exome Sequencing data set was funded by the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) sponsored Netherlands Consortium for Healthy Aging (NCHA; project nr. 050-060e810), by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by a Complementation Project of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL; www.bbmri.nl; project number CP2010-41).

The au-thors also thank Mr. Pascal Arp, Ms. Mila Jhamai, BSc, and Mr. Marijn Verkerkj for their help in creating the RS-Exome Sequencing database.
The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Cul-ture, and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study, and the participating general practitioners and pharmacists.

This study is also supported by the Courage-PD is an EU Joint ProgrammedNeurodegenerative Disease Research (JPND) project (MS, TG). The project is supported through the following funding organizations under the aegis of JPNDewww.jpnd.eu: the Medical Research Council, United Kingdom, the French National Research Agency, the German Bundesministerium für Bildung und For-schung, the Italian Ministry of Health/Ministry of Education, Uni-versities and Research, the Israeli Ministry of Health, the Luxembourgian National Research Fund, the Netherlands Organi-sation for Health Research and Development, the Research Council of Norway, the Portuguese Foundation for Science and Technology, and the Spanish National Institute of Health Carlos III.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.neurobiolaging. 2017.07.009.

References

Amendola, L.M., Dorschner, M.O., Robertson, P.D., Salama, J.S., Hart, R., Shirts, B.H., Murray, M.L., Tokita, M.J., Gallego, C.J., Kim, D.S., Bennett, J.T., Crosslin, D.R., Ranchalis, J., Jones, K.L., Rosenthal, E.A., Jarvik, E.R., Itsara, A., Turner, E.H., Herman, D.S., Schleit, J., Burt, A., Jamal, S.M., Abrudan, J.L., Johnson, A.D., Conlin, L.K., Dulik, M.C., Santani, A., Metterville, D.R., Kelly, M., Foreman, A.K., Lee, K., Taylor, K.D., Guo, X., Crooks, K., Kiedrowski, L.A., Raffel, L.J., Gordon, O., Machini, K., Desnick, R.J., Biesecker, L.G., Lubitz, S.A., Mulchandani, S., Cooper, G.M., Joffe, S., Richards, C.S., Yang, Y., Rotter, J.I., Rich, S.S., O'Donnell, C.J., Berg, J.S., Spinner, N.B., Evans, J.P., Fullerton, S.M., Leppig, K.A., Bennett, R.L., Bird, T., Sybert, V.P., Grady, W.M., Tabor, H.K., Kim, J.H., Bamshad, M.J., Wilfond, B., Motulsky, A.G., Scott, C.R., Pritchard, C.C., Walsh, T.D., Burke, W., Raskind, W.H., Byers, P., Hisama, F.M., Rehm, H., Nickerson, D.A., Jarvik, G.P., 2015. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res. 25, 305e315. Shendure, J., 2011. Exome sequencing as a tool for Mendelian disease gene discovery. Nat. Rev. Genet. 12, 745e755. Botstein, D., Risch, N., 2003. Discovering genotypes underlying human phenotypes:
past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 33 (Suppl), 228-237.
Bras, J., Guerreiro, R., Hardy, J., 2015. SnapShot: genetics of Parkinson's disease. Cell 160, 570-570.e1.
Cirulli, E.T., Lasseigne, B.N., Petrovski, S., Sapp, P.C., Dion, P.A., Leblond, C.S., outhouis, J., Lu, Y.F., Wang, Q., Krueger, B.J., Ren, Z., Keebler, J., Han, Y., Levy, S.E., Boone, B.E., Wimbish, J.R., Waite, L.L., Jones, A.L., Carulli, J.P., Day-Williams, A.G., Staropoli, J.F., Xin, W.W., Chesi, A., Raphael, A.R., McKenna-Yasek, D., Cady, J.,

Vianney de Jong, J.M., Kenna, K.P., Smith, B.N., Topp, S., Miller, J., Gkazi, A., Al-Chalabi, A., van den Berg, L.H., Veldink, J., Silani, V., Ticozzi, N., Shaw, C.E., Baloh, R.H., Appel, S., Simpson, E., Lagier-Tourenne, C., Pulst, S.M., Gibson, S., Trojanowski, J.Q., Elman, L., McCluskey, L., Grossman, M., Shneider, N.A., Chung, W.K., Ravits, J.M., Glass, J.D., Sims, K.B., Van Deerlin, V.M., Maniatis, T., Hayes, S.D., Ordureau, A., Swarup, S., Landers, J., Baas, F., Allen, A.S., Bedlack, R.S., Harper, J.W., Gitler, A.D., Rouleau, G.A., Brown, R., Harms, M.B., Cooper, G.M., Harris, T., Myers, R.M., Goldstein, D.B., 2015. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science (New York, NY) 347, 1436-1441.
Do, C.B., Tung, J.Y., Dorfman, E., Kiefer, A.K., Drabant, E.M., Francke, U., Mountain, J.L., Goldman, S.M., Tanner, C.M., Langston, J.W., Wojcicki, A., Eriksson, N., 2011. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. PLoS Genet. 7, e1002141.
Edwards, A.M., Isserlin, R., Bader, G.D., Frye, S.V., Willson, T.M., Yu, F.H., 2011. Too many roads not taken. Nature 470, 163-165.
Edwards, T.L., Scott, W.K., Almonte, C., Burt, A., Powell, E.H., Beecham, G.W., Wang, L., Zuchner, S., Konidari, I., Wang, G., Singer, C., Nahab, F., Scott, B., Stajich, J.M., Pericak-Vance, M., Haines, J., Vance, J.M., Martin, E.R., 2010. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. Ann. Hum. Genet. 74, 97e109. Goker-Alpan, O., Schiffmann, R., LaMarca, M.E., Nussbaum, R.L., McInerney-Leo, A., Sidransky, E., 2004. Parkinsonism among Gaucher disease carriers. J. Med. Genet. 41, 937-940.
Hofman, A., Brusselle, G.G., Darwish Murad, S., van Duijn, C.M., Franco, O.H., Goedegebure, A., Ikram, M.A., Klaver, C.C., Nijsten, T.E., Peeters, R.P., Stricker, B.H., Tiemeier, H.W., Uitterlinden, A.G., Vernooij, M.W., 2015. The Rotterdam Study: 2016 objectives and design update. Eur. J. Epidemiol. 30, 661e708. Hughes, A.J., Daniel, S.E., Kilford, L., Lees, A.J., 1992. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases.
J. Neurol. Neurosurg. Psychiatry 55, 181-184.
Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J.D., Lin, X., 2013. Sequence kernel association tests for the combined effect of rare and common variants. Am. J. Hum. Genet. 92, 841-853.
Jiang, S., Wells, C.D., Roach, P.J., 2011. Starch-binding domain-containing protein 1 (Stbd1) and glycogen metabolism: identification of the Atg8 family interacting motif (AIM) in Stbd1 required for interaction with GABARAPL1. Biochem. Bio-phys. Res. Commun. 413, 420-425.
Keller, M.F., Saad, M., Bras, J., Bettella, F., Nicolaou, N., Simon-Sanchez, J., Mittag, F., Sklar, P., Sullivan, P.F., Moran, J.L., Hultman, C.M., Lichtenstein, P., Magnusson, P., Lehner, T., Shugart, Y.Y., Price, A.L., de Bakker, P.I., Purcell, S.M., Sunyaev, S.R., 2012. Exome sequencing and the genetic basis of complex traits. Nat. Genet. 44, 623-630.
Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 46, 310-315.
Klein, C., Westenberger, A., 2012. Genetics of Parkinson's disease. Cold Spring Harbor Perspect. Med. 2, a008888.
Lander, E.S., 1996. The new genomics: global views of biology. Science (New York,

NY) 274, 536-539.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) 25, 1754e1760.
Lill, C.M., Roehr, J.T., McQueen, M.B., Kavvoura, F.K., Bagade, S., Schjeide, B.M., Schjeide, L.M., Meissner, E., Zauft, U., Allen, N.C., Liu, T., Schilling, M., Anderson, K.J., Beecham, G., Berg, D., Biernacka, J.M., Brice, A., DeStefano, A.L., Do, C.B., Eriksson, N., Factor, S.A., Farrer, M.J., Foroud, T., Gasser, T., Hamza, T., Hardy, J.A., Heutink, P., Hill-Burns, E.M., Klein, C., Latourelle, J.C., Maraganore, D.M., Martin, E.R., Martinez, M., Myers, R.H., Nalls, M.A., Pankratz, N., Payami, H., Satake, W., Scott, W.K., Sharma, M., Singleton, A.B., Stefansson, K., Toda, T., Tung, J.Y., Vance, J., Wood, N.W., Zabetian, C.P., Young, P., Tanzi, R.E., Khoury, M.J., Zipp, F., Lehrach, H., Ioannidis, J.P., Bertram, L., 2012. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: the PDGene database. PLoS Genet. 8, e1002548.
Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F., McCarroll, S.A., Visscher, P.M., 2009. Finding the missing heritability of complex diseases. Nature 461, 747e753.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297-1303.
Nalls, M.A., Bras, J., Hernandez, D.G., Keller, M.F., Majounie, E., Renton, A.E., Saad, M., ansen, I., Guerreiro, R., Lubbe, S., Plagnol, V., Gibbs, J.R., Schulte, C., Pankratz, N., Sutherland, M., Bertram, L., Lill, C.M., DeStefano, A.L., Faroud, T., Eriksson, N., Tung, J.Y., Edsall, C., Nichols, N., Brooks, J., Arepalli, S., Pliner, H., Letson, C., Heutink, P., Martinez, M., Gasser, T., Traynor, B.J.,Wood, N., Hardy, J., Singleton, A.B., 2015. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. Neurobiol. Aging 36, 1605.e7-1605.e12.
DeStefano, A.L., Kara, E., Bras, J., Sharma, M., Schulte, C., Keller, M.F., Arepalli, S., Letson, C., Edsall, C., Stefansson, H., Liu, X., Pliner, H., Lee, J.H., Cheng, R., Ikram, M.A., Ioannidis, J.P., Hadjigeorgiou, G.M., Bis, J.C., Martinez, M., Perlmutter, J.S., Goate, A., Marder, K., Fiske, B., Sutherland, M., Xiromerisiou, G., Myers, R.H., Clark, L.N., Stefansson, K., Hardy, J.A., Heutink, P., Chen, H., Wood, N.W., Houlden, H., Payami, H., Brice, A., Scott, W.K., Gasser, T., Bertram, L., Eriksson, N., Foroud, T., Singleton, A.B., 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's dis-ease. Nat. Genet. 46, 989e993.
Nalls, M.A., Plagnol, V., Hernandez, D.G., Sharma, M., Sheerin, U.M., Saad, M., Simon-Sanchez, J., Schulte, C., Lesage, S., Sveinbjornsdottir, S., Stefansson, K., Martinez, M., Hardy, J., Heutink, P., Brice, A., Gasser, T., Singleton, A.B., Wood, N.W., 2011. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet 377, 641-649.
Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.A., Fraser, D., Warren, L., Aponte, J., Zawistowski, M., Liu, X.,

Zhang, H., Zhang, Y., Li, J., Li, Y., Li, L., Woollard, P., Topp, S., Hall, M.D., Nangle, K., Wang, J., Abecasis, G., Cardon, L.R., Zollner, S., Whittaker, J.C., Chissoe, S.L., Novembre, J., Mooser, V., 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science (New York, NY) 337, 100-104.
Nourashrafeddin, S., Ebrahimzadeh-Vesal, R., Modarressi, M.H., Zekri, A., Nouri, M.,2014. Identification of Spata-19 new variant with expression beyond meiotic phase of mouse testis development. Rep. Biochem. Mol. Biol. 2, 89e93.
Paisan-Ruiz, C., Jain, S., Evans, E.W., Gilks, W.P., Simon, J., van der Brug, M., Lopez de Munain, A., Aparicio, S., Gil, A.M., Khan, N., Johnson, J., Martinez, J.R., Nicholl, D., Carrera, I.M., Pena, A.S., de Silva, R., Lees, A., Marti-Masso, J.F., Perez-Tur, J., Wood, N.W., Singleton, A.B., 2004. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron 44, 595-600.
Pankratz, N., Beecham, G.W., DeStefano, A.L., Dawson, T.M., Doheny, K.F., Factor, S.A., Hamza, T.H., Hung, A.Y., Hyman, B.T., Ivinson, A.J., Krainc, D., Latourelle, J.C., Clark, L.N., Marder, K., Martin, E.R., Mayeux, R., Ross, O.A., Scherzer, C.R., Simon, D.K., Tanner, C., Vance, J.M., Wszolek, Z.K., Zabetian, C.P., Myers, R.H., Payami, H., Scott, W.K., Foroud, T., 2012. Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. Ann. Neurol. 71, 370-384.
Pihlstrom, L., Toft, M., 2011. Parkinson's disease: what remains of the "missing heritability"? Mov. Disord. 26, 1971-1973.
Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., Stenroos, E.S., Chandrasekharappa, S., Athanassiadou, A., Papapetropoulos, T., Johnson, W.G., Lazzarini, A.M., Duvoisin, R.C., Di Iorio, G., Golbe, L.I., Nussbaum, R.L., 1997. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science (New York, NY) 276, 2045-2047.
Pringsheim, T., Jette, N., Frolkis, A., Steeves, T.D., 2014. The prevalence of Parkinson's disease: a systematic review and meta-analysis. Mov. Disord. 29, 1583e1590. Pritchard, J.K., Cox, N.J., 2002. The allelic architecture of human disease genes: common disease-common variant...or not? Hum. Mol. Genet. 11, 2417e2423.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559-575.
Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kahler, A., Duncan, L., Stahl, E., Genovese, G., Fernandez, E., Collins, M.O., Komiyama, N.H., Choudhary, J.S., Magnusson, P.K., Banks, E., Shakir, K., Garimella, K., Fennell, T., DePristo, M., Grant, S.G., Haggarty, S.J., Gabriel, S., Scolnick, E.M., Lander, E.S., Hultman, C.M., Sullivan, P.F., McCarroll, S.A., Sklar, P., 2014. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 506, 185-190.
Ross, O.A., Soto-Ortolaza, A.I., Heckman, M.G., Aasly, J.O., Abahuni, N., Annesi, G., Bacon, J.A., Bardien, S., Bozi, M., Brice, A., Brighina, L., Van Broeckhoven, C., Carr, J., Chartier-Harlin, M.C., Dardiotis, E., Dickson, D.W., Diehl, N.N., Elbaz, A., Ferrarese, C., Ferraris, A., Fiske, B., Gibson, J.M., Gibson, R., Hadjigeorgiou, G.M., Hattori, N., Ioannidis, J.P., Jasinska-Myga, B., Jeon, B.S., Kim, Y.J., Klein, C., Kruger, R., Kyratzi, E., Lesage, S., Lin, C.H., Lynch, T., Maraganore, D.M., Mellick, G.D., Mutez, E., Nilsson, C.,

Opala, G., Park, S.S., Puschmann, A., Quattrone, A., Sharma, M., Silburn, P.A., Sohn, Y.H., Stefanis, L., Tadic, V., Theuns, J., Tomiyama, H., Uitti, R.J., Valente, E.M., van de Loo, S., Vassilatis, D.K., Vilarino-Guell, C., White, L.R., Wirdefeldt, K., Wszolek, Z.K., Wu, R.M., Farrer, M.J., 2011. Association of LRRK2 exonic variants with susceptibility to Parkinson's disease: a case-control study. Lancet Neurol. 10, 898-908.
Sharma, M., Kruger, R., Gasser, T., 2014. From genome-wide association studies to next-generation sequencing: lessons from the past and planning for the future. JAMA Neurol. 71, 5-6.
Simon-Sanchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G., Kruger, R., Federoff, M., Klein, C., Goate, A., Perlmutter, J., Bonin, M., Nalls, M.A., Illig, T., Gieger, C., Houlden, H., Steffens, M., Okun, M.S., Racette, B.A., Cookson, M.R., Foote, K.D., Fernandez, H.H., Traynor, B.J., Schreiber, S., Arepalli, S., Zonozi, R., Gwinn, K., van der Brug, M., Lopez, G., Chanock, S.J., Schatzkin, A., Park, Y., Hollenbeck, A., Gao, J., Huang, X., Wood, N.W., Lorenz, D., Deuschl, G., Chen, H., Riess, O., Hardy, J.A., Singleton, A.B., Gasser, T., 2009. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat. Genet. 41, 1308-1312.
Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., Kang, H.M., Jordan, D., Leal, S.M., Gabriel, S., Rieder, M.J., Abecasis, G., Altshuler, D., Nickerson, D.A., Boerwinkle, E., Sunyaev, S., Bustamante, C.D., Bamshad, M.J., Akey, J.M., 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science (New York, NY) 337, 64-69.
Trinh, J., Farrer, M., 2013. Advances in the genetics of Parkinson disease. Nat. Rev. Neurol. 9, 445-454.
Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164.
Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., Lin, X., 2011. Rare-variant associationesting for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. 89, 82-93.
Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R.J., Calne, D.B., Stoessl, A.J., Pfeiffer, R.F., Patenge, N., Carbajal, I.C., Vieregge, P., Asmus, F., Muller-Myhsok, B., Dickson, D.W., Meitinger, T., Strom, T.M., Wszolek, Z.K., Gasser, T., 2004. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44, 601-607.
Zuk, O., Schaffner, S.F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M.J., Neale, B.M., Sunyaev, S.R., Lander, E.S., 2014. Searching for missing heritability: designing rare variant association studies. Proc. Natl. Acad. Sci. U. S. A. 111, E455-E464.