npg

ORIGINAL ARTICLE

Fine mapping and resequencing of the PARK16 locus in Parkinson's disease

Lasse Pihlstrøm[1,2], Aina Rengmark[1], Kari Anne Bjørnarå[3], Nil Dizdar[4], Camilla Fardell[5], Lars Forsgren[6], Björn Holmberg[7], Jan Petter Larsen[8], Jan Linder[6], Hans Nissbrandt[5], Ole-Bjørn Tysnes[9], Espen Dietrichs[1,2] and Mathias Toft[1]

The PARK16 locus, spanning five genes on chromosome 1, was among the first genetic regions to show genome-wide association in Parkinson's disease (PD). Subsequent investigations have found variability in PARK16 top-hits and association patterns across populations, and the implicated genes and mechanisms are currently unclear. In the present study, we aimed to explore the contribution of PARK16 variability to PD risk in a Scandinavian population. We genotyped 17 single-nucleotide polymorphisms in a case-control sample set of 2570 individuals from Norway and Sweden to fine map the locus. Targeted resequencing of the full coding regions of *SLC45A3*, *NUCKS1, RAB7L1, SLC41A1* and *PM20D1* was performed in DNA pools from a subset of 387 patient samples. We find evidence for an association with PD for rs1775143 as well as a haplotype located around the 5′ region of *RAB7L1,* implicating variants which are not in high linkage disequilibrium with the strongest signal from a recent large meta-analysis in Caucasians. We also provide suggestive support for epistasis between *RAB7L1* and *LRRK2* as previously hypothesized by others. Comparing our results with previous work, allelic heterogeneity at PARK16 appears likely, and further studies are warranted to disentangle the complex patterns of association and pinpoint the functionally relevant variants.
Journal of Human Genetics (2015) **60**, 357–362; doi:10.1038/jhg.2015.34; published online 9 April 2015

INTRODUCTION

A major current challenge in the genetic study of complex disease is to identify the responsible genes and functional variants underlying the susceptibility loci detected in genome-wide association studies (GWAS). Parkinson's disease (PD) is a common neurodegenerative disease, for which 26 significant GWAS signals have been reported to date.[1–12] One of the first loci to be identified was designated as PARK16, spanning a 170-kb region on chromosome 1 entailing the genes *SLC45A3*, *NUCKS1, RAB7L1* (currently renamed *RAB29* in the HUGO nomenclature), *SLC41A1* and *PM20D1* (Figure 1).

The PARK16 locus has been robustly corroborated at genome-wide significance in both Asian and Caucasian populations, yet association data have shown considerable variability across single studies regarding allele frequencies and effects of individual single-nucleotide polymorphisms (SNPs). Originally, several PARK16 SNPs were reported as genome-wide significant in a Japanese GWAS.[1] Some of these SNPs, but not the Japanese top-hit, were reproduced at significance levels near the genome-wide threshold in a collaborating study from Europe and the US.[2] Subsequently, the locus failed to replicate in several studies of Caucasian subjects.[3,8,13] Larger sample sizes were needed before the unequivocal evidence confirmed the association signal also

outside of Asia, most recently demonstrated in the largest meta-analysis of PD GWAS to date.[12]

Reported PARK16 susceptibility variants include several SNPs that are not in high linkage disequilibrium (LD), and several authors have suggested that allelic heterogeneity is likely, with more than one underlying functional variant independently affecting the disease risk.[1,12] However, no systematic fine mapping study with stepwise conditional association analysis has been published to date to confirm this hypothesis. The implicated gene(s), the functionally relevant genetic variants, and the molecular mechanisms affecting disease risk are therefore currently unclear.

Expression data from both leukocytes and brain tissue have shown that PD risk SNPs are expression quantitative trait loci (eQTLs) for both *NUCKS1* and *RAB7L1*.[1,12] An association with DNA methylation across both these genes has also been reported.[12] A few studies have performed targeted resequencing of coding regions of PARK16,[14,15] and a possible functional mechanism has been proposed for a rare coding variant in *SLC41A1*.[16] Evidence has been presented supporting a common pathway for *RAB7L1* and *LRRK2*, a gene known to be implicated in both sporadic and dominant PD.[17] This study further suggested a functional mechanism affecting

[1]Department of Neurology, Oslo University Hospital, Oslo, Norway; [2]University of Oslo, Institute of Clinical Medicine, Oslo, Norway; [3]Department of Neurology, Drammen Hospital, Drammen, Norway; [4]Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden; [5]Department of Pharmacology, The Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden; [6]Department of Pharmacology and Clinical Neuroscience, Umeå University, Umeå, Sweden; [7]Department of Clinical Neuroscience and Rehabilitation, The Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden; [8]The Norwegian Centre for Movement Disorders, Stavanger University Hospital, Stavanger, Norway and [9]Department of Neurology, Haukeland University Hospital, Bergen, Norway
Correspondence: Dr L Pihlstrøm, Department of Neurology, Oslo University Hospital, PO Box 4950 Nydalen, N-0424 Oslo, Norway.
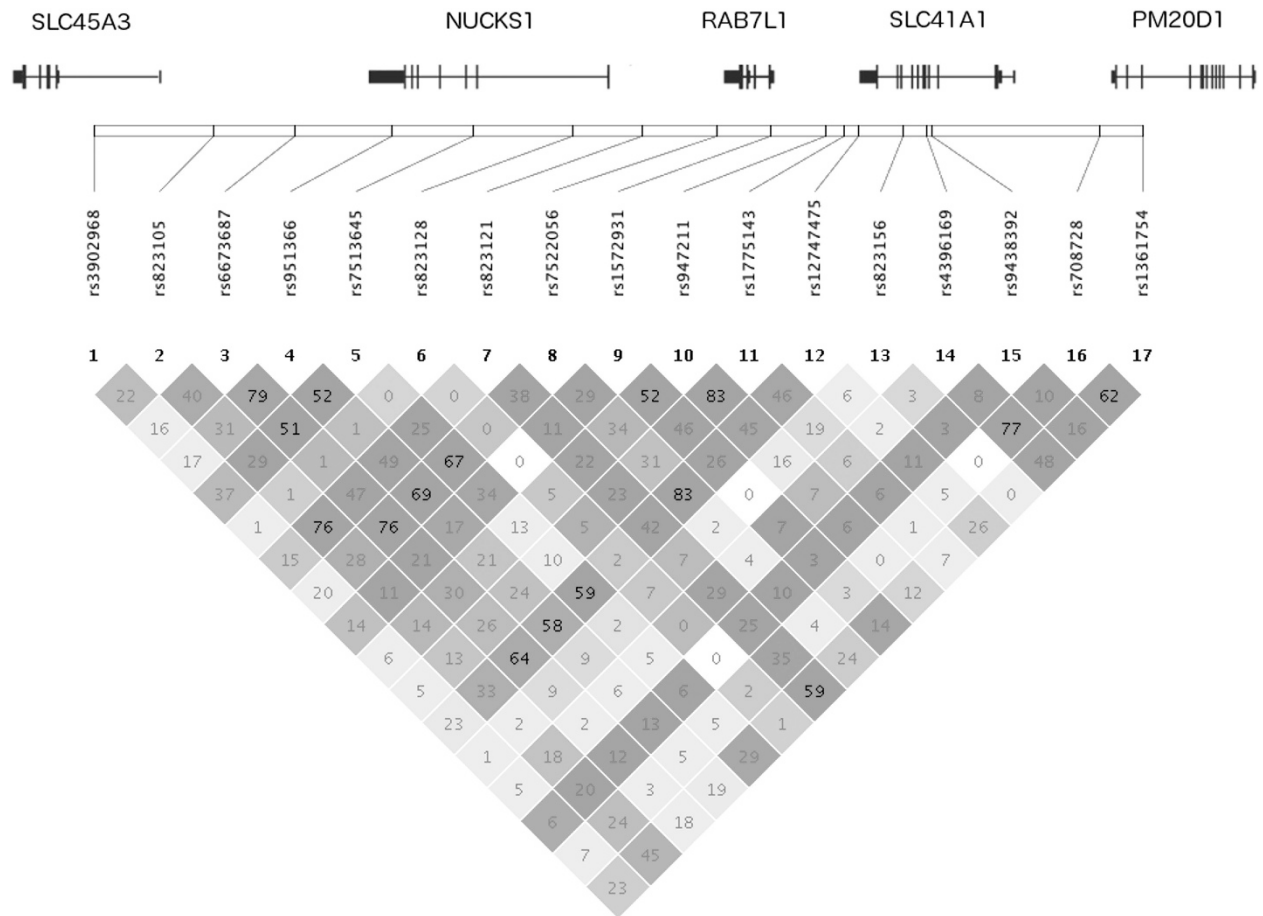E-mail: laspih@ous-hf.no

Figure 1 The PARK16 locus, genotyped markers and linkage disequilibrium plot. The top of the figure shows the five genes of the PARK16 locus with exons (thick) untranslated regions (intermediate) and introns (thin blue lines). The positions of the 17 single-nucleotide polymorphisms genotyped in the current study are indicated below. The bottom of the panel shows linkage disequilibrium as estimated in the software Haploview. The values in each square indicate pairwise r^2, darker red color corresponds to higher D'. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

RAB7L1 splicing isoforms, and epistatic effects with PD risk variants in *LRRK2*.

We previously published a replication study demonstrating supportive evidence for 11 PD GWAS loci in a relatively homogeneous sample set from Norway and Sweden.[18] Three PARK16 SNPs were included in the study, but we observed no significant association despite excellent estimated statistical power (99% power for $P < 0.05$ based on odds ratio (OR) from the original GWAS and minor allele frequency (MAF) from our study). Based on the emerging robust evidence for PARK16 also in Caucasian populations, we aimed to explore this region further in our Scandinavian sample set. In the present study, we genotyped 17 tag-SNPs spanning the PARK16 region in 1345 PD patients and 1225 controls free of neurodegenerative disease. In 387 PD samples we investigated the full coding regions of all five PARK16 genes by next generation targeted resequencing. We also analyzed potential epistatic effects between PARK16 and a common *LRRK2* GWAS signal. Finally, we reviewed our results in the context of previously published studies and publicly available gene expression data.

MATERIALS AND METHODS
Subjects
PD patients and controls were recruited from five collaborating centers in Norway and Sweden. Inclusion criteria and demographics are described in

detail in a previous publication.[18] The study was approved by the Regional Committee for Medical Research Ethics (Oslo, Norway). Sample and data collection at each study site was approved by local ethics committees. All participants gave written, informed consent.

Reanalysis of Scandinavian GWAS replication data
Our previous GWAS replication study included three PARK16 SNPs (rs947211, rs823128 and rs823156), which all failed to replicate.[18] We interpreted the lack of association as most probably related to the differences in LD structure across populations. However, we came to suspect some degree of heterogeneity in the PARK16 association signal between different sites included in our study, most apparent for rs947211. Based on this observation as well as the emerging hypothesis of a common pathway for *RAB7L1* and *LRRK2*, we decided to perform additional statistical analyses in our GWAS replication data as a starting point for further PARK16 investigations.

For all three PARK16 SNPs we performed the Cochran–Mantel–Haenszel test for $2 \times 2 \times K$ stratified tables to assess disease association while controlling for possible differences between study sites. In a logistic regression model we tested for SNPxSNP epistasis between PARK16 SNPs and the *LRRK2* variant rs1491942. Both analyses were performed with statistical tools implicated in the software package PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/).

SNP genotyping, association testing and haplotype analysis
Based on previous studies, we defined the PARK16 locus as corresponding to genomic region 1:205638377-205808377 (GHRC b37). We used the tag SNP function in the HapMap Genome Browser (release #27) to design a panel of

16 SNPs capturing all other HapMap variants above a MAF of 0.2 at an LD threshold of $r^2 > 0.8$. In addition, we included rs1572931, for which a causal role linked to splicing has been proposed in a previous study.[17] Genotyping was performed in the Scandinavian sample set of 1345 PD patients and 1225 healthy controls either by KASPar assays on a ViiA7 instrument (Life Technologies, Foster City, CA, USA) (rs7522056, rs1572931 and rs1775143) or by matrix-assisted laser desorption ionization time-of-flight mass spectrometry using the Sequenom MassARRAY (Sequenom, San Diego, CA, USA) system (all other SNPs). We included only samples that passed quality control in the GWAS replication study. 78 samples were unavailable for KASP genotyping. The overall genotype call rate was 0.98, with values above 0.95 for each individual SNP. We tested for Hardy–Weinberg equilibrium in controls, observing no significant departure ($P < 0.01$).

We assessed single-SNP associations with disease status by X^2-test and calculated ORs. LD patterns were visualized by plots generated in the software Haploview (www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview). By using a sliding window approach, we estimated the haplotype frequencies for all combinations of three adjacent SNPs across the data set and tested for association with PD. Minimal haplotype frequency to be included in analysis was the default 0.01. To further investigate the hypothesis of a genetic interaction with *LRRK2* we investigated SNPxSNP epistasis with the full genotyping panel and assessed how carrying two high-risk copies of the most associated haplotype affected the OR of rs1491942.

Statistical analyses were performed in PLINK or SPSS (Armonk, NY, USA). As our panel included 17 SNPs and we also performed further tests on three SNPs from our original GWAS replication study, we acknowledge that some form of correction for multiple testing would be required to claim strong significance. However, given the context of an established risk locus where we follow-up on specific findings reported by others, we decided to interpret *P*-values of nominal significance (< 0.05), combined with comparable size and direction of effects, as supportive of the relevant hypothesis. With respect to the sliding window haplotype association testing, we performed 100 000 permutations and estimated empirical *P*-values both on the level of each window as well as across all haplotypes combined.

Pooled sequencing of all PARK16 coding regions
To investigate if low-frequency coding variation might contribute to the PARK16 association signal, we performed targeted resequencing of the coding regions of *SLC45A3*, *NUCKS1*, *RAB7L1*, *SLC41A1* and *PM20D1*. The PARK16 genes were included as part of a larger capture panel, which was used to sequence 39 pools of DNA from 10 individuals each, all included from the Oslo sample series. Laboratory methods, bioinformatic analyses, validation and quality measures for this pooled sequencing experiment have been reported in detail in a previous publication.[19] In brief, equimolar amounts of DNA from 10 individuals were pooled together before target enrichment with a 200-kb HaloPlex kit (Currently Agilent Technologies, Santa Clara, CA, USA). Deep sequencing was performed on an Illumina HiSeq 2000 (Illumina, San Diego, CA, USA) at the Norwegian Sequencing Centre, Oslo. Reads were aligned to the reference genome with the Burrows-Wheeler aligner (bwa 0.5.9). We used the Genome Analysis Toolkit (GATK 2.5) (www.broadinstitute.org/gatk/) for bioinformatic processing of aligned sequence files and variant calling and

ANNOVAR (version 2012may25) (www.openbioinformatics.org/annovar/) for variant annotation.

Coding variants underlying GWAS signals would be expected to have intermediate frequency and effect size as compared to highly penetrant monogenic mutations on one side, and common noncoding susceptibility variants on the other. We applied a filtering algorithm to our results keeping only nonsynonymous variants with an estimated MAF in the pooled sequencing data between 0.005 and 0.1. Next, we compared the MAF in PD patients to variant frequency in the Exome Server Project, as well as an in-house database of 176 Norwegian exomes without relevance to PD. We calculated ORs to prioritize variants for validation and follow-up in a case-control data set. One variant showed a tendency to be overrepresented in PD with estimated OR > 2 against both databases, and was subsequently genotyped by KASPar assay in all Oslo samples.

RESULTS
SNP association, haplotype analysis and epistasis with *LRRK2*
As a starting point for our exploration of the PARK16 locus, we performed further analyses on relevant SNPs included in our previous PD GWAS replication study. In stratified association analysis, controlling for possible geographical differences within Scandinavia, we found weak suggestive evidence for rs947211 ($P = 0.048$, OR = 0.87). The two other SNPs showed insignificant *P*-values, yet ORs consistent with the original Caucasian GWAS[2] (rs832128; $P = 0.11$, OR = 0.77, rs823156; $P = 0.096$, OR = 0.88). Next, we assessed SNPxSNP interaction between these same three PARK16 SNPs and the *LRRK2* GWAS variant rs1491942 in a logistic regression model. Again, the results indicated a possible role for rs947211 ($P = 0.022$ for interaction), but not the others (rs832128; $P = 0.21$, rs823156; $P = 0.16$).

Together, these results were compatible with a possible role for PARK16 in Scandinavian PD patients despite our initial negative findings, and we extended our investigation to a broader genotyping panel. The results from single-SNP association analysis are shown in Supplementary Table 1. One SNP, rs1775143, showed evidence of association with PD ($P = 0.0095$, OR = 0.84), yet the *P*-value would not pass a significance threshold strictly corrected for 17 independent tests. Comparing association data for individual study sites, we found that the minor allele of rs1775143 was consistently underrepresented in PD patients across all five sites, yet with evidence of geographical variation regarding effect size and allele frequency (Supplementary Table 2). The minor allele of rs1775143 was more common in Sweden than in Norway and showed a larger effect size in association analysis (OR 0.76 vs 0.91). A comparison between the Norwegian and Swedish case-control sets is included in Supplementary Table 3.

We note that rs1775143 is located in the intergenic region between *RAB7L1* and *SLC41A1*, neighboring rs947211 in our panel. Plots visualizing the LD pattern across the PARK16 region in our Scandinavian data are shown in Figure 1 and Supplementary Figure 1. Although the picture is complex, with some markers showing high levels of LD over longer distances, rs1775143 correlates most strongly with its flanking SNPs upstream of *RAB7L1*. Consequently, we chose to investigate the association with PD further by estimating haplotypes from all combinations of three adjacent SNPs in a sliding window approach. Association results for all 15 haplotypes with empirical *P*-values are shown in Supplementary Table 4. We found that one three-SNP window showed association with PD (empirical $P = 0.046$ corrected for the multiple testing of all haplotypes across all windows). This haplotype spans the transcription start site for *RAB7L1* and includes the previously proposed functional SNP rs1572931, the original GWAS hit rs947211 and our most associated SNP rs1775143. Frequencies and ORs with confidence intervals for haplotypes at this window are shown in Table 1.

Table 1 Association analysis for the rs1572931- rs947211- rs1775143 haplotype

Haplotype	Frequency PD patients	Frequency controls	OR (95% CI)
CGT	0.76	0.72	1.18 (1.04–1.33)
TAC	0.13	0.14	0.90 (0.76–1.06)
CAC	0.090	0.095	0.94 (0.78–1.14)
CGC	0.015	0.028	0.51 (0.33–0.77)
CAT	0.0085	0.013	0.66 (0.38–1.16)

Abbreviations: CI, confidence interval; PD, Parkinson's disease; OR, odds ratio.
Haplotype frequencies were estimated by the expectation-maximization algorithm as implicated in the software PLINK. ORs were calculated for each individual haplotype compared against all other haplotypes combined.

In the study by MacLeod *et al.*,[17] proposing a common pathway for *RAB7L1* and *LRRK2*, increasing ORs for a common *LRRK2* variant was shown for individuals homozygous for a high-risk PARK16 allele. We attempted to replicate this pattern in our own data, calculating ORs for different genotypes of rs1491942 stratified by whether individuals carry two high-risk haplotypes or not. The results are shown in Table 2. While rs1491942 showed no association at baseline, we observed a tendency compatible with the interaction hypothesis, where the maximal risk combination across both loci is over-represented in PD patients. Homozygous high-risk at both PARK16 and *RAB7L1* was about twice as common in the PD group as compared to controls, but the confidence interval is wide (95%CI for OR = 1.18–3.86).

Results for logistic regression analysis of SNPxSNP epistasis between all genotyped PARK16 SNPs and *LRRK2* rs1491942 is provided in Supplementary Table 5. We observed no interactions passing a significance threshold adjusted for multiple testing. We note that five out of 17 SNPs reached P-values < 0.05 for interaction with rs1491942, yet these did not include rs1775143.

Targeted resequencing of PARK16 genes and follow-up of a low-frequency variant in *PM20D1*

Assessing coverage metrics for the targeted resequencing experiment, we set a threshold of 80x, corresponding to 4x per pooled allele on average, based on our previous experience.[19] Across all pools, the mean proportion of coding positions to reach this benchmark was 96%. At the level of individual genes, the 80x threshold was reached for more than 95% of positions for each gene except *NUCKS1*, for which the corresponding figure was 88%. Coverage statistics are summarized in Supplementary Table 6. The nonsynonymous variants

passing our frequency filters are listed in Table 3. Calculations of ORs as compared with frequency in the Exome Server Project database and an in-house database of Norwegian exomes were suggestive of a possible association with PD for one variant only, rs141605758 in *PM20D1*. We genotyped this SNP in all Oslo case-control samples. The presence of a heterozygous individual was validated for all positive pools in the sequencing experiment. However, we found no convincing difference between carrier frequencies in patients (6 heterozygotes out of 405) versus controls (5 heterozygotes out of 470).

A previous PARK16 sequencing study from the United Kingdom reported an insertion in *RAB7L1* (c.379-12insT) that associated with disease, as well as two novel coding SNPs seen only in patients.[14] Regardless of the frequency filter, we did not observe any of these variants in our data. All nonsynonymous variants detected in pooled sequencing are listed in Supplementary Table 7.

DISCUSSION
When the largest meta-analysis of PD GWAS performed to date was published recently, the authors highlighted PARK16 as a complicated locus, where both allelic heterogeneity and epistatic effects may have a role, warranting dedicated efforts to disentangle the nature of the association signal.[12] Following the original discovery, a number of replication studies have contributed further evidence for PARK16 association to a variable degree, often limited by sample size.[13,20–28] Although the locus has now been unequivocally corroborated in both Asian and Caucasian populations, large GWAS and meta-analyses have shown differences with respect to allele frequencies and effect sizes for various groups of SNPs. We have summarized previously published PARK16 association signals in Table 4, illustrating the variability across ethnic groups.

In our previous Scandinavian investigation of PD GWAS loci, we attempted to replicate three of these PARK16 SNPs with negative results. In the present study, expanding the analysis to a broader panel of tag-SNPs we find suggestive evidence for rs1775143 and a stronger association signal for a haplotype including the adjacent SNPs rs974211 and rs1572931. The original Asian top-hit rs974211 itself also showed a tendency to associate with disease when controlling for variability across study sites, and taken together we interpret these findings as supportive of the same signal that is represented by this SNP in previous studies, located upstream of *RAB7L1*. This variant reached genome-wide significance for Caucasians in a meta-analysis from the PDgene database[11] (Table 4).

As the largest meta-analyses of PD GWAS have reached genome-wide significance levels for PARK16 in Caucasians, the reported top-hit SNPs have been located in the 3′UTR of *RAB7L1* (rs708723)[9] and the intergenic region between this gene and *NUCKS1*

Table 2 Interaction with *LRRK2* rs1491942

	PARK16 haplotype rs1572931-rs947211-rs1775143	
	Low-risk haplotype carrier (N = 1064)	Homozygous for risk haplotype (N = 1437)
LRRK2 rs1491942 genotype		
CC	Reference	Reference
CG	0.49 (0.23–1.06)	0.99 (0.71–1.25)
GG	1.05 (0.80–1.39)	**2.14 (1.18–3.86)**

The table shows odds ratios with 95% confidence intervals for association between rs1491942 genotype and PD, when stratified by PARK16 haplotype carrier status. PARK16 haplotype groups were defined as homozygotes for the risk haplotype CGT versus carriers of at least one other low-risk haplotype. The effect of rs1491942 appears significant only against a PARK16 high-risk background, as highlighted in bold.

Table 3 Nonsynonymous PARK16 variants identified by targeted resequencing

dbSNP ID	Position	Gene	Amino acid change	Estimated frequency in pooled sequencing	Frequency in ESP database	Frequency in 176 Norwegian exomes	Estimated OR vs ESP database	Estimated OR vs 176 Norwegian exomes
rs137949511	1:205628616	*SLC45A3*	p.Val470Ile	0.8%	0.7%	0.0%	1.17	N.A.
rs41302139	1:205740668	*RAB7L1*	p.Gln104Glu	1.0%	1.1%	1.1%	0.96	0.91
rs114776144	1:205801756	*PM20D1*	p.Val419Ile	1.4%	0.9%	1.4%	1.54	1.00
rs141605758	1:205811865	*PM20D1*	p.Pro281Gln	0.9%	0.4%	0.3%	**2.27**	**3.20**
rs139066269	1:205812879	*PM20D1*	p.Met248Thr	0.5%	0.4%	0.3%	1.35	1.82

Abbreviations: ESP, Exome Server Project; OR, odds ratio; SNPs, single-nucleotide polymorphisms.
The table lists nonsynonymous PARK16 variants with an estimated MAF between 0.5 and 10% identified by targeted resequencing of pooled DNA from 387 Norwegian PD patients. Frequencies are compared to data from the ESP and an in-house database of 176 Norwegian exomes from individuals without known neurological disease. Estimated ORs for the variant selected for further genotyping are indicated in bold.
RefSeq accessions: *SLC45A3*, NP_149093.1; *RAB7L1*, NP_003920.1; *PM20D1*, NP_689704.4.

Table 4 Overview of PARK16 association signals

SNP	rs708723/rs823118[a]	rs823128	rs947211	rs823156/rs708730[a]
Minor allele frequency — Asians/Caucasians	0.39/0.46	0.18/0.01	0.38/0.22	0.26/0.16
LD in Asians/Caucasians — r^2 with rs708723	—	0.14/0.01	0.97/0.33	0.22/0.18
r^2 with rs823128	0.14/0.01	—	0.13/0.03	0.52/0.05
r^2 with rs947211	0.97/0.33	0.13/0.03	—	0.22/0.19
r^2 with rs823156	0.22/0.18	0.52/0.05	0.22/0.19	—
Association results in Asian populations — Satake et al.[1] GWAS	—	GWS	Top-hit, GWS	GWS
Lill et al.[11] Meta-analysis	—	—	—	GWS
Association results in Caucasians — Lill et al.[11] Meta-analysis; Nalls et al.[12] Meta-analysis of GWAS	GWS	—	GWS	—
Functional study — MacLeod et al.[17]	Interaction with LRRK2 (rs823114) RAB7L1 ($P=1.2\times10^{-8}$)	RAB7L1 ($P=4.0\times10^{-9}$)	Proposed functional mechanism (rs1572931) RAB7L1 ($P=5.9\times10^{-17}$) NUCKS1 ($P=7.8\times10^{-6}$)	—
eQTL data — Monocytes, Raj et al.[29]	PM20D1 ($P=2.2\times10^{-22}$) RAB7L1 ($P=2.3\times10^{-7}$)		PM20D1 ($P=7.9\times10^{-8}$) RAB7L1 ($P=2.5\times10^{-6}$)	
Brain, Ramasamy et al.[30]				PM20D1 ($P=6.2\times10^{-9}$) RAB7L1 ($P=5.0\times10^{-6}$) NUCKS1 ($P=3.9\times10^{-3}$)

Abbreviations: eQTL, expression quantitative trait locus; GWAS, genome-wide association study; GWS, genome-wide significance ($P<5\times10^{-8}$); LD, linkage disequilibrium; NS, nominal significance ($P<0.05$).
The table summarizes available data on PARK16 association signals that have been reported at genome-wide significance. We note that the original Japanese GWAS reported two further SNPs at P-values between 5×10^{-8} and 5×10^{-7} which are not included here. Allele frequencies and r^2 values were retrieved from the 1000genomes Pilot 1 data by using the online software tool SNAP. Only association results from the largest and most comprehensive GWAS and meta-analysis performed in each population are listed. When other SNPs in high LD with the association signal were reported, these are stated in parentheses. Monocyte data are from the supplement of reference 29. Brain eQTL data from the UK brain expression consortium were downloaded from www.braineac.org. The P-values reported in the table are the lowest for any probe in each of the five PARK16 genes when data across all investigated brain regions are combined. Only P-values<0.01 are listed.
aSNPs in high LD ($r^2>0.95$) are listed together as one signal.

(rs823118),[12] showing high LD with each other, but not with rs947211. We note that this signal may have been imperfectly tagged in our panel due to incomplete SNP data in HapMap ($r^2=0.61$ with rs823121). However, even if we genotyped rs823118, the reported OR of 1.12 implicates that the statistical power to detect an association in our study would have been weak. Variability in LD patterns across populations may affect how genotyped SNPs tag unknown functional variants in different studies. We still consider it most likely that the discrepancy between the present and previous reports also reflects allelic heterogeneity with multiple independent functional variants affecting PD risk.

In contrast to Caucasian populations, rs947211 and rs823118 show near perfect LD in Asians ($r^2=0.97$). However, two other signals in low LD with these SNPs have so far been reported at genome-wide significance in Asian studies (Table 4), supporting the hypothesis of allelic heterogeneity.[1,11] One of these SNPs, rs823128, is very rare in Caucasians (MAF 0.01) and may therefore be tagging a functional variant almost unique to Asians. Future large-scale efforts to characterize the PARK16 locus and identify independent signals should take care to consider the variability across both populations and larger ethnic groups within this region.

It is currently unclear whether one or several genes in the PARK16 locus contribute to PD pathogenesis. The recent GWAS meta-analysis reviewed a brain data set from a previous study and reported that rs823118 was an eQTL and also a methylation quantitative trait locus for both RAB7L1 and NUCKS1.[12] We further queried the different PARK16 top-SNPs in publicly available eQTL data sets from monocytes[29] and brain,[30] summarizing the results in Table 4. In conclusion, current evidence seems to link PARK16 susceptibility variants most strongly to the expression of RAB7L1, yet the presently available data are by no means conclusive regarding the implicated gene(s). The most associated single variant in the present study is also an eQTL for RAB7L1 (brain; $P=7.1\times10^{-6}$, monocytes; $P=2.9\times10^{-14}$).

From targeted deep sequencing of all PARK16 coding regions in 387 PD patients we found no evidence indicating that low-frequency nonsynonymous variants contribute to the association signal. The present study provides no direct data to support the speculation about possible functional mechanisms of noncoding variants. A functional role through an effect on mRNA splicing, with exon 2 skipping of RAB7L1, was previously proposed for rs1572931 by MacLeod et al. Although this SNP is part of our associated haplotype, we would expect a significant result for the SNP in itself if it were truly a causative variant. We also reviewed the evidence for eQTL properties of rs1572931 for an exon 2 specific probe in the brain eQTL data set, but found only moderate support for this hypothesis (probe ID: 2452676, $P=0.0096$, averaged across all brain regions).

The same study by MacLeod et al.[17] also found evidence from transcriptomics, as well as in vitro and in vivo functional experiments, for a common pathogenic pathway for RAB7L1 and LRRK2 involving retromer and lysosome functions. The authors furthermore proposed an epistatic effect, with nonlinear increase in PD susceptibility when individuals carry common high-risk alleles at both loci. An independent subsequent study attempting to replicate this finding reported no evidence of interaction between GWAS SNPs rs708723 and rs1491942.[27] In the present Scandinavian data set, we observed SNPxSNP interaction between rs947211 and rs1491942 that marginally passed a nominal significance threshold. We also found a similar pattern of increasing OR when high-risk variants are combined for both the LRRK2 SNP and the associated three-SNP haplotype in the 5′ region of RAB7L1. We recognize that the statistical significance of

these results is weak, and the findings should be interpreted with caution.

In conclusion, we have further explored the nature of the PARK16 association with PD in a Scandinavian sample set through fine mapping of SNPs and targeted resequencing. Our results provide supportive evidence for a signal located near the 5′ end of *RAB7L1*, which may be independent from the signal highlighted in the most recent meta-analysis of PD GWAS in Caucasian populations. Further studies are warranted to disentangle all disease-relevant functional variation at the PARK16 locus in PD, including possible epistatic effects and shared pathogenetic pathways between *RAB7L1* and *LRRK2*.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

WEB RESOURCES

The URLs for databases and software referenced in this article are as follows: 1000 Genomes, www.1000genomes.org. NHLBI Exome Sequencing Project (ESP), evs.gs.washington.edu/EVS/Burrows-Wheeler Aligner, bio-bwa.sourceforge.net. The Genome Analysis Toolkit, www.broadinstitute.org/gatk/. ANNOVAR, www.openbioinformatics.org/annovar/. SNAP, www.broadinstitute.org/mpg/snap/. Haploview, www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview. PLINK, http://pngu.mgh.harvard.edu/ ~ purcell/plink/.

1 Satake, W., Nakabayashi, Y., Mizuta, I., Hirota, Y., Ito, C. & Kubo, M. *et al.* Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat. Genet.* **41**, 1303–1307 (2009).
2 Simon-Sanchez, J., Schulte, C., Bras, J. M., Sharma, M., Gibbs, J. R. & Berg, D. *et al.* Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.* **41**, 1308–1312 (2009).
3 Hamza, T. H., Zabetian, C. P., Tenesa, A., Laederach, A., Montimurro, J. & Yearout, D. *et al.* Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat. Genet.* **42**, 781–785 (2010).
4 Edwards, T. L., Scott, W. K., Almonte, C., Burt, A., Powell, E. H. & Beecham, G. W. *et al.* Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson's disease. *Ann. Hum. Genet.* **74**, 97–109 (2010).
5 Spencer, C. C., Plagnol, V., Strange, A., Gardner, M., Paisan-Ruiz, C. & Band, G. *et al.* Dissection of the genetics of Parkinson's disease identifies an additional association 5′ of SNCA and multiple associated haplotypes at 17q21. *Hum. Mol. Genet.* **20**, 345–353 (2011).
6 Saad, M., Lesage, S., Saint-Pierre, A., Corvol, J. C., Zelenika, D. & Lambert, J. C. *et al.* Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. *Hum. Mol. Genet.* **20**, 615–627 (2011).
7 Do, C. B., Tung, J. Y., Dorfman, E., Kiefer, A. K., Drabant, E. M. & Francke, U. *et al.* Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet.* **7**, e1002141 (2011).
8 Nalls, M. A., Plagnol, V., Hernandez, D. G., Sharma, M., Sheerin, U. M. & Saad, M. *et al.* 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 (2011).
9 International Parkinson's Disease Genomics Consortium (IPDGC), Wellcome Truste Case Control Consortium 2 (WTCCC2). A two-stage meta-analysis identifies several new loci for Parkinson's disease. *PLoS Genet.* **7**, e1002142 (2011).
10 Pankratz, N., Beecham, G. W., DeStefano, A. L., Dawson, T. M., Doheny, K. F. & Factor, S. A. *et al.* Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann. Neurol.* **71**, 370–384 (2012).
11 Lill, C. M., Roehr, J. T., McQueen, M. B., Kavvoura, F. K., Bagade, S. & Schjeide, B. M. *et al.* Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: the PDGene database. *PLoS Genet.* **8**, e1002548 (2012).
12 Nalls, M. A., Pankratz, N., Lill, C. M., Do, C. B., Hernandez, D. G. & Saad, M. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).
13 Mata, I. F., Yearout, D., Alvarez, V., Coto, E., de Mena, L. & Ribacoba, R. *et al.* Replication of MAPT and SNCA, but not PARK16-18, as susceptibility genes for Parkinson's disease. *Mov. Disord.* **26**, 819–823 (2011).
14 Tucci, A., Nalls, M. A., Houlden, H., Revesz, T., Singleton, A. B. & Wood, N. W. *et al.* Genetic variability at the PARK16 locus. *Eur. J. Hum. Genet.* **18**, 1356–1359 (2010).
15 Yan, Y., Tian, J., Mo, X., Zhao, G., Yin, X. & Pu, J. *et al.* Genetic variants in the RAB7L1 and SLC41A1 genes of the PARK16 locus in Chinese Parkinson's disease patients. *Int. J. Neurosci.* **121**, 632–636 (2011).
16 Kolisek, M., Sponder, G., Mastrototaro, L., Smorodchenko, A., Launay, P. & Vormann, J. *et al.* Substitution p.A350V in Na(+)/Mg(2)(+) exchanger SLC41A1, potentially associated with Parkinson's disease, is a gain-of-function mutation. *PLoS One* **8**, e71096 (2013).
17 MacLeod, D. A., Rhinn, H., Kuwahara, T., Zolin, A., Di Paolo, G. & McCabe, B. D. *et al.* RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. *Neuron* **77**, 425–439 (2013).
18 Pihlstrom, L., Axelsson, G., Bjornara, K. A., Dizdar, N., Fardell, C. & Forsgren, L. *et al.* Supportive evidence for 11 loci from genome-wide association studies in Parkinson's disease. *Neurobiol. Aging* **34**, 1708e1707-1713 (2013).
19 Pihlstrom, L., Rengmark, A., Bjornara, K. A. & Toft, M. Effective variant detection by targeted deep sequencing of DNA pools: an example from Parkinson's disease. *Ann. Hum. Genet.* **78**, 243–252 (2014).
20 Tan, E. K., Kwok, H. H., Tan, L. C., Zhao, W. T., Prakash, K. M. & Au, W. L. *et al.* Analysis of GWAS-linked loci in Parkinson disease reaffirms PARK16 as a susceptibility locus. *Neurology* **75**, 508–512 (2010).
21 Vilarino-Guell, C., Ross, O. A., Aasly, J. O., White, L. R., Rajput, A. & Rajput, A. H. *et al.* An independent replication of PARK16 in Asian samples. *Neurology* **75**, 2248–2249 (2010).
22 Ramirez, A., Ziegler, A., Winkler, S., Kottwitz, J., Giesen, R. & Diaz-Grez, F. *et al.* Association of Parkinson disease to PARK16 in a Chilean sample. *Parkinsonism Relat. Disord.* **17**, 70–71 (2011).
23 Chang, X. L., Mao, X. Y., Li, H. H., Zhang, J. H., Li, N. N. & Burgunder, J. M. *et al.* Association of GWAS loci with PD in China. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **156B**, 334–339 (2011).
24 Yan, Y. P., Mo, X. Y., Tian, J., Zhao, G. H., Yin, X. Z. & Jin, F. Y. *et al.* An association between the PARK16 locus and Parkinson's disease in a cohort from eastern China. *Parkinsonism Relat. Disord.* **17**, 737–739 (2011).
25 Wang, C., Cai, Y., Zheng, Z., Tang, B. S., Xu, Y. & Wang, T. *et al.* Penetrance of LRRK2 G2385R and R1628P is modified by common PD-associated genetic variants. *Parkinsonism Relat. Disord.* **18**, 958–963 (2012).
26 Zhou, L. L., Zhang, X., Bao, Q. Q., Liu, R. P., Gong, M. Y. & Mao, G. Y. *et al.* Association analysis of PARK16-18 variants and Parkinson's disease in a Chinese population. *J. Clin. Neurosci.* **21**, 1029–1032 (2014).
27 Soto-Ortolaza, A. I., Heckman, M. G., Labbe, C., Serie, D. J., Puschmann, A. & Rayaprolu, S. *et al.* GWAS risk factors in Parkinson's disease: LRRK2 coding variation and genetic interaction with PARK16. *Am. J. Neurodegener. Dis.* **2**, 287–299 (2013).
28 Gan-Or, Z., Bar-Shira, A., Dahary, D., Mirelman, A., Kedmi, M. & Gurevich, T. *et al.* Association of sequence alterations in the putative promoter of RAB7L1 with a reduced parkinson disease risk. *Arch. Neurol.* **69**, 105–110 (2012).
29 Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M. N. & Replogle, J. M. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523 (2014).
30 Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C. & Walker, R. *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17**, 1418–1428 (2014).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)