

ASSOCIATION STUDIES ARTICLE

Identification of genetic modifiers of age-at-onset for familial Parkinson's disease

Erin M. Hill-Burns¹, Owen A. Ross², William T. Wissemann¹, Alexandra I. Soto-Ortolaza², Sepideh Zareparsis³, Joanna Siuda⁴, Timothy Lynch⁵, Zbigniew K. Wszolek⁶, Peter A. Silburn⁷, George D. Mellick⁷, Beate Ritz⁸, Clemens R. Scherzer⁹, Cyrus P. Zabetian¹⁰, Stewart A. Factor¹¹, Patrick J. Breheny¹² and Haydeh Payami^{1,13,*}

¹Department of Neurology, University of Alabama at Birmingham, AL, USA, ²Department of Neuroscience, Mayo Clinic Jacksonville, FL, USA, ³Department of Molecular and Medical Genetics, Oregon Health & Sciences University, Portland, OR, USA, ⁴Department of Neurology, Medical University of Silesia, Katowice, Poland, ⁵Dublin Neurological Institute at the Mater Misericordiae University Hospital, Conway Institute of Biomolecular & Biomedical Research, University College Dublin, Ireland, ⁶Department of Neurology, Mayo Clinic Jacksonville, FL, USA, ⁷Eskitis Institute for Drug Discovery, Griffith University, Queensland, Australia, ⁸Department of Epidemiology, Fielding School of Public Health and Neurology, Geffen School of Medicine at UCLA, Los Angeles, CA, USA, ⁹The Neurogenomics Laboratory, Harvard Medical School and Brigham & Women's Hospital, Cambridge, MA, USA, ¹⁰VA Puget Sound Health Care System and Department of Neurology, University of Washington, Seattle, WA, USA, ¹¹Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA, ¹²Department of Biostatistics, University of Iowa, Iowa City, IA, USA and ¹³Center for Genomic Medicine, HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

*To whom correspondence should be addressed at: MCLM 401, 1918 University Blvd, Birmingham AL 35233-0005, USA. Tel: +205-994-1909; Fax: +205-996-0116; Email: haydehpayami@uabmc.edu

Abstract

Parkinson's disease (PD) is the most common cause of neurodegenerative movement disorder and the second most common cause of dementia. Genes are thought to have a stronger effect on age-at-onset of PD than on risk, yet there has been a phenomenal success in identifying risk loci but not age-at-onset modifiers. We conducted a genome-wide study for age-at-onset. We analysed familial and non-familial PD separately, per prior evidence for strong genetic effect on age-at-onset in familial PD. GWAS was conducted in 431 unrelated PD individuals with at least one affected relative (familial PD) and 1544 non-familial PD from the NeuroGenetics Research Consortium (NGRC); an additional 737 familial PD and 2363 non-familial PD were used for replication. In familial PD, two signals were detected and replicated robustly: one mapped to *LHFPL2* on 5q14.1 ($P_{\text{NGRC}} = 3\text{E-}8$, $P_{\text{Replication}} = 2\text{E-}5$, $P_{\text{NGRC+Replication}} = 1\text{E-}11$), the second mapped to *TPM1* on 15q22.2 ($P_{\text{NGRC}} = 8\text{E-}9$, $P_{\text{Replication}} = 2\text{E-}4$, $P_{\text{NGRC+Replication}} = 9\text{E-}11$). The variants that were associated with accelerated onset had low frequencies (<0.02). The *LHFPL2*

Received: February 24, 2016. Revised: June 15, 2016. Accepted: June 23, 2016

© The Author 2016. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

variant was associated with earlier onset by 12.33 [95% CI: 6.2; 18.45] years in NGRC, 8.03 [2.95; 13.11] years in replication, and 9.79 [5.88; 13.70] years in the combined data. The TPM1 variant was associated with earlier onset by 15.30 [8.10; 22.49] years in NGRC, 9.29 [1.79; 16.79] years in replication, and 12.42 [7.23; 17.61] years in the combined data. Neither LHFPL2 nor TPM1 was associated with age-at-onset in non-familial PD. LHFPL2 (function unknown) is overexpressed in brain tumours. TPM1 encodes a highly conserved protein that regulates muscle contraction, and is a tumour-suppressor gene.

Introduction

Genetics plays a significant role in PD [MIM*168600], both in determining risk (if one will develop PD: cause) as well as age-at-onset (when a disease might manifest: modifier) (1). Several rare causative genes (2–11) and 28 common risk alleles (12–16) have been confirmed for PD. The known genes and risk factors account for ~5% of the heritability (17), hence much of the genetic component of PD is still missing.

Age-at-onset of PD varies by approximately 80 years (Fig. 1). The factors that contribute to the variation in age-at-onset are unknown, although genes are thought to be important. Heritability of PD has been estimated as 98% (SE = 0.25) for age-at-onset and 60% (SE = 0.10) for risk (1). Data from the most recent PD meta genome-wide association study (GWAS) have provided significant evidence for a polygenic component to age-at-onset (18), although no specific genes were identified. Three independent complex segregation analyses have reported a significantly better fit for a genetic model than for an environmental model for PD, and found the genetic effect on age-at-onset to be significantly greater than the genetic effect on risk (19–21). In one study, the best-fit model was rare alleles with large effects on age-at-onset in familial PD (19). Another study estimated an average decrease in age-at-onset of approximately 18 years for each copy of the putative allele (21). Thus, taken collectively, the clues from complex segregation analyses were “rare variant”, “large impact on age-at-onset”, and “positive family history”.

The loci that affect risk have little effect on age-at-onset. The International PD Genetic Consortium (6,249 PD cases) (18) and studies from Denmark (1,526 cases) (22) and from Norway and Sweden (1,340 cases) (23) independently reported that the risk alleles identified to date account for <1% of the variation in age-at-onset. Thus, 99% of the 80-year variation in age-at-onset of PD remains unexplained.

Here, we report evidence for the existence of variants with low allele frequencies and large effects on age-at-onset of familial PD, which we identified via GWAS and replicated independently. We analyzed familial and non-familial PD separately because complex segregation analyses had suggested a strong genetic effect on age-at-onset of familial PD specifically (19).

About one-fourth of persons with PD report a positive family history (Table 1), but their families rarely show a Mendelian inheritance pattern and most are not caused by known PD mutations (3–11). The vast majority of familial PD remains idiopathic, and like non-familial PD, is thought to involve complex interactions between the genome and environmental exposures (24–27). It is usually assumed that the same genes operate in familial and non-familial PD; in fact, GWAS for risk have successfully uncovered numerous susceptibility loci without separating the subtypes (12–16,26–28). However, familial and non-familial PD might differ in the relative burden of genetic and non-genetic modifiers (13,29,30). If certain variants are involved predominantly in one subtype (e.g. in familial PD as segregation analysis has suggested for age-at-onset modifiers), their signal may become diluted and undetectable if familial and non-familial PD are mixed. A positive family history does not necessarily imply a genetic aetiology because non-genetic disease can also cluster in families due to a common exposure. Similarly, genetic disease may present as non-familial due to incomplete penetrance (e.g. LRRK2 mutations (29)). Moreover, a familial case may be classified as non-familial given the difficulty in recall and knowledge of family members. Despite these uncertainties, stratifying by presence/absence of family history proved to be key to identifying two genes that each affect age-at-onset by a decade.

Results

Genome-wide genotyping was conducted using Illumina HumanOmni1-Quad_v1-0_B BeadChips on 3986 subjects from NGRC (13), including 435 familial PD (one person per family), 1565 non-familial PD and 1986 controls (PD subjects were used for analysis of age-at-onset, and controls were used for ancillary tests). Subjects were unrelated (subjects with cryptic relatedness $PI_HAT > 0.15$ were excluded). Over 800,000 genotyped SNPs passed quality control (13). We used imputation and expanded the coverage to 7.2 million SNPs (30). Statistical testing for GWAS was conducted using Cox regression survival analysis, treating age-at-onset as a quantitative trait. Linear

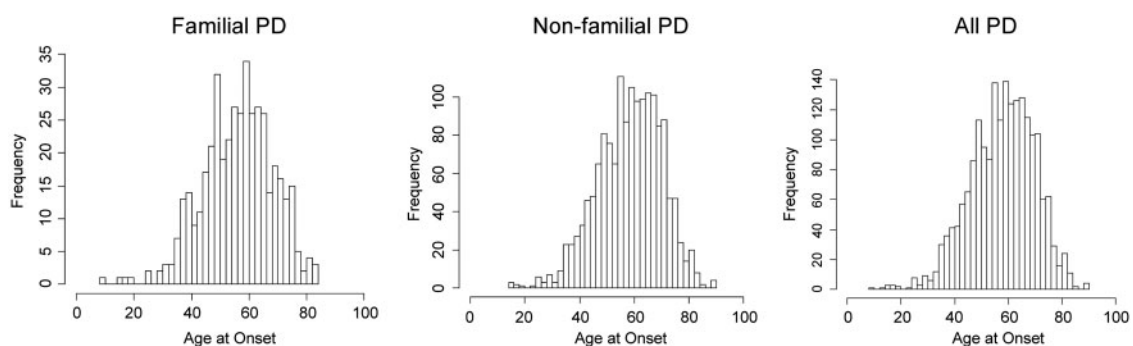


Figure 1. Variation in age-at-onset of PD Age-at-onset distribution in NGRC subjects shows nearly 80 years of variation in both familial and non-familial PD. The tails (age at onset ≤ 20 or ≥ 89 years) were excluded from analyses.

regression was also performed which yielded similar but less significant results than Cox. Cox regression is particularly suited for the analysis of time-to-event data, such as age-at-onset, where subjects are treated as unaffected from birth until the age when they develop symptoms (event) (31–34). Using an additive genetic model, genotypes were compared for age-specific incidence of PD symptoms using Cox regression, and hazard ratios (HR) were calculated with their associated *P*-values. The resulting Manhattan plots and quantile-quantile (QQ) plots are shown in Figure 2. Genomic inflation factors were close to one ($\lambda_{\text{familial}}=0.989$, $\lambda_{\text{non-familial}}=0.996$, $\lambda_{\text{all-PD}}=1.007$) indicating the *P*-values were not inflated. Genome-wide significant signals ($P < 5E-8$) were seen only in familial PD. Complete genome-wide results, including HR and *P*-values for 7.2 million SNPs for familial, non-familial and all PD, are provided in the Supplementary Tables.

Familial PD

Four loci reached $P < 5E-8$ in familial PD (Fig. 2A, Table 2). They were on chromosome 5q14.1 (rs344650: minor allele frequency (MAF)=0.016; HR=4.77, $P=3E-8$), chromosome 8q23.3 (rs74335301: MAF=0.014; HR=4.46, $P=3E-8$), chromosome 14q21.3 (rs192855008: MAF=0.012; HR=7.12, $P=4E-9$), and chromosome 15q22.2 (rs116860970: MAF=0.013; HR=6.52, $P=8E-9$). Genome-wide results for familial PD are provided in Supplementary Material, Table S1.

The signal on 5q14.1 included a variant that was directly genotyped on the GWAS array. The other three peaks were imputed. Since the fidelity of imputation for rare variants is unknown (35), we genotyped a subset of samples for the three imputed peaks (see Methods for details). Concordance between genotyped and imputed results was 98% for 15q22.2, 99% for 8q23.3 and 100% for 14q21.3. Replication samples were all genotyped. Adjusting for the first two principal components improved the association signals (chromosome 5q14.1 rs344650 $P=2E-8$; chromosome 8q23.3 rs74335301 $P=3E-8$, chromosome 14q21.3 rs192855008 $P=3E-9$, chromosome 15q22.2 rs116860970 $P=8E-9$).

The loci that achieved $P < 5E-8$ in discovery were carried to replication and were genotyped in 3100 additional PD samples (737 unrelated familial PD and 2363 non-familial PD; Table 1).

Potential for overlap across discovery and replication datasets was tested by comparing 74 SNP genotypes and all available phenotype data; no evidence of overlap was found. To correct for sparse numbers of minor-allele carriers in individual replication datasets, we applied Firth's Penalized correction for Cox regression (36,37). The signal from 5q14.1 and 15q22.2 replicated robustly in familial PD; i.e., the associations in the familial subset of replication were significant and the combination of NGRC and replication produced a more significant signal than the NGRC data alone (Table 2, Figs 3 and 4). The replication signals for 8q23.3 and 14q21.3 were borderline significant and when combined with NGRC, the signals were less significant than NGRC alone (Table 2). The discovery signal for 8q23.3 included only one SNP (down to $P=1E-6$), which adds to the uncertainty about the original finding at this peak.

The signal from 5q14.1 mapped to the *LHFPL2* (Lipoma HMGIC Fusion Partner-Like 2) gene. *LHFPL2* rs344650_G vs. A (5q14.1) yielded HR=4.77 ($P=3E-8$) in familial PD in GWAS, HR=2.68 ($P=2E-5$) in familial PD in replication, and HR=3.40 ($P=1E-11$) in a meta-analysis of familial PD in GWAS and replication (Fig. 3A). Presence of the rs344650_G allele was associated with 12 years earlier onset in NGRC ($\beta=-12.33$ [-18.45; -6.21]), 8 years in replication ($\beta=-8.03$ [-13.11; -2.95]), and 9.79 years in combined data ($\beta=-9.79$ [-13.70; -5.88]) (Fig. 3B). The Kaplan Meier plots show an accelerated age-at-onset distribution for rs344650_GA vs. AA genotype ($P_{\text{NGRC}}=2E-9$ (Fig. 5A), $P_{\text{Replication}}=6E-3$ (Fig. 5B)). rs344650_G was not associated with risk in familial PD (OR=1.04, $P=0.91$). rs344650_G was not associated with age in controls ($P=0.57$) or in patients ($P=0.42$ adjusted for age-at-onset). The Moving Average Plot (MAP) (38) of rs344650_G was consistent with the pattern expected for an age-at-onset modifier and distinct from the patterns for a risk allele like *SNCA* rs356220 which is associated with PD ubiquitously (13) (Fig. 6A) or like *PARK2* deletions/duplications which are risk factors for early-onset PD (7,39) (Fig. 6B). Note that the overall frequency of rs344650_G was the same in cases and controls (MAF_{familial_PD}=0.016±.004; MAF_{non-familial_PD}=0.014±.002; MAF_{all_PD}=0.014±.002; MAF_{controls}=0.014±.002); the distinguishing feature, as depicted in the *LHFPL2* MAPs in NGRC (Fig. 6C) and replication (Fig. 6D), was the enrichment of rs344650_G in cases with earlier onsets and gradual depletion of the allele with increasing ages-at-onset.

Table 1. Datasets and subject characteristics

Dataset	Familial PD				Non-familial PD				All			
	N	M/F	Age	Onset age	N	M/F	Age	Onset age	N	M/F	Age	Onset age
NGRC												
PD	431	280/151	66.2±10.4	56.9±11.7	1554	1057/497	67.5±10.6	58.9±11.4	1985	1337/648	67.2±10.6	58.5±11.5
Control									1986	769/1217	70.3±14.1	
REPLICATION												
AUST	293	170/123	69.8±10.3	57.5±11.2	842	532/310	71.9±10.1	60.4±11.2	1135	702/433	71.3±10.2	59.6±11.3
HBS*	99	67/32	63.8±9.2	58.7±9.5	350	227/123	66.7±10.0	62.6±10.6	449	294/155	66.1±9.9	61.7±10.4
MCJI	12	5/7	61.8±9.6	55.1±11.0	229	134/95	59.3±10.0	51.3±10.6	241	139/102	59.4±10.0	51.5±10.7
MCJE	142	90/52	69.5±9.7	63.0±10.7	182	113/69	69.4±10.7	63.8±12.0	324	203/121	69.5±10.3	63.5±11.4
MCJP	39	22/17	62.9±8.6	55.3±10.1	272	172/100	67.2±10.3	59.1±11.1	311	194/117	66.7±10.2	58.6±11.0
MCJU	112	74/38	66.5±12.6	59.9±12.8	217	139/78	70.7±10.7	64.8±12.2	329	213/116	69.2±11.5	63.2±12.6
UCLA*	40	21/19	70.8±9.9	68.8±9.7	271	156/115	71.4±10.5	69.3±10.6	311	177/134	71.3±10.4	69.2±10.5
Total	737	449/288	68.0±10.6	59.6±11.4	2363	1473/890	69.0±10.9	61.4±12.0	3100	1922/1178	68.8±10.8	60.9±11.9

NGRC and replication datasets were tested for potential overlap; no evidence was found for overlap. Subjects with age-at-onset at the extreme tails of the distribution (≤ 20 years, and ≥ 89 years) were excluded from analysis. Control subjects were used to test and rule out association of SNPs with age and with disease risk. M/F = N male/N female. Age = Age-at-enrollment ± standard deviation. Onset age = age-at-onset of first motor symptom of PD (*age-at-diagnosis) ± standard deviation.

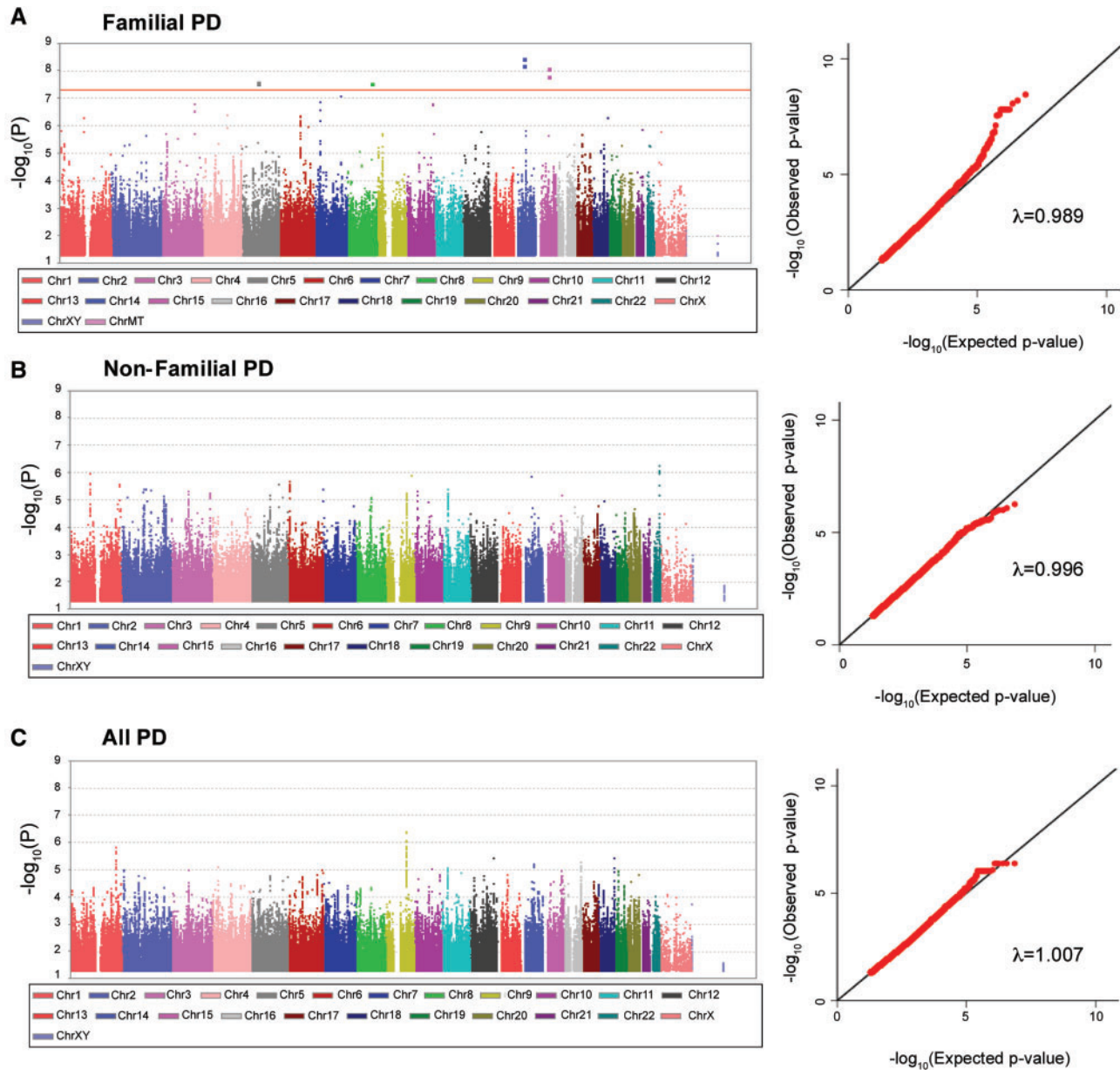


Figure 2. GWAS. Left panel: Manhattan Plots. Using Cox regression, four signals achieved $P < 5E-8$ in familial PD (A). No signals were detected in non-familial PD (B) or in all PD (C). SNPs with $P \geq 0.05$ are not plotted. Right panel: QQ plots. The observed P-values were consistent with the expected distributions and did not appear to be inflated ($\lambda_{\text{familial}}=0.989$, $\lambda_{\text{non-familial}}=0.996$, $\lambda_{\text{all-PD}}=1.007$).

The signal from 15q22.2 mapped to the *TPM1* (tropomyosin) gene. *TPM1* rs117267308_A vs. T (15q22.2) yielded HR = 6.47 in familial PD in GWAS ($P = 2E-8$), HR = 3.20 ($P = 2E-4$) in familial PD in replication, and HR = 4.55 ($P = 9E-11$) in a meta-analysis of familial PD in GWAS and replication (Fig. 4A). The presence of the rs117267308_A allele was associated with 15 years earlier onset in NGRC ($\beta = -15.30$ [-22.49; -8.10]), 9 years in replication ($\beta = -9.29$ [-16.79; -1.79]), and 12 years in combined data ($\beta = -12.42$ [-17.61; -7.23]) (Fig. 4B). Age-at-onset distribution curves generated by the Kaplan Meier method showed significant separation between rs117267308_AT and rs117267308_TT genotypes in familial PD ($P_{\text{NGRC}} = 2E-10$ (Fig. 5C), $P_{\text{Replication}} = 7E-3$ (Fig. 5D)). rs117267308 was not associated with risk of familial PD (OR = 1.18, $P = 0.67$). rs117267308 was not associated with age in controls ($P = 0.78$) or in patients ($P = 0.57$ adjusted for age-at-onset). The MAPs of *TPM1*

were consistent with the signature pattern for an age-at-onset modifier (Fig. 6E and F).

There was no significant difference in association with age-at-onset between sexes for *LHFPL2* or *TPM1*. In familial PD, carriers of rare alleles were heterozygous. One *LHFPL2* rs344650_GG rare homozygote was observed in non-familial PD.

Non-familial PD

No signal reached $P < 5E-8$ in non-familial PD (Fig. 2B). Genome-wide results for non-familial PD are provided in Supplementary Material, Table S2. The strongest signal in non-familial PD was at $P = 6E-7$ (Table 3). Note that the sample size for non-familial PD was three times larger than the sample size for familial PD,

Table 2. Signals that achieved the significance threshold in GWAS for associations with age-at-onset of familial PD

CHR	Gene	SNP	MAF	Discovery (NGRC)			Replication			Discovery + Replication								
				GWAS Test		P	Effect on AAO		P	HR	P	Effect on AAO		P				
				HR	P		Beta	95% CI				Linear regression	Beta		95% CI			
5	LHFPL2	rs10035651	0.016	4.76	3E-8	-12.31	-18.42; -6.19	8E-5	-	-	-	-	-	-	-	-	-	
5	LHFPL2	rs344650	0.016	4.77	3E-8	-12.33	-18.45; -6.21	8E-5	2.68	2E-5	-8.03	-13.11; -2.95	1E-3	-	-	-	-	-
5	LHFPL2	rs344657	0.016	4.77	3E-8	-12.33	-18.45; -6.21	8E-5	-	-	-	-	-	-	-	-	-	-
8	TRPS1	rs74335301	0.014	4.46	3E-8	-11.76	-17.98; -5.54	2E-4	1.39	0.07	0.44	-4.34; 5.23	0.86*	2.20	3E-6	-4.09	-7.89; -0.30	0.03
14	KLHDC1	rs79503702	0.012	6.95	7E-9	-14.81	-22.00; -7.62	5E-5	1.89	0.04	-2.03	-10.46; 6.41	0.32	3.82	5E-8	-9.43	-14.90; -3.96	7E-4
14	KLHDC1_ARF6	rs192855008	0.012	7.12	4E-9	-15.01	-22.22; -7.79	5E-5	-	-	-	-	-	-	-	-	-	-
15	TPM1	rs117267308	0.012	6.47	2E-8	-15.30	-22.49; -8.10	3E-5	3.20	2E-4	-9.29	-16.79; -1.79	8E-3	4.55	9E-11	-12.42	-17.61; -7.23	3E-6
15	TPM1	rs141049631	0.012	6.47	2E-8	-15.30	-22.49; -8.10	3E-5	-	-	-	-	-	-	-	-	-	-
15	TPM1	rs116860970	0.013	6.52	8E-9	-15.13	-22.17; -8.09	3E-5	-	-	-	-	-	-	-	-	-	-
15	TPM1	rs77362326	0.012	6.47	2E-8	-15.30	-22.49; -8.10	3E-5	-	-	-	-	-	-	-	-	-	-
15	TPM1	rs201411148	0.012	6.47	2E-8	-15.30	-22.49; -8.10	3E-5	-	-	-	-	-	-	-	-	-	-
15	TPM1	rs142383316	0.012	6.47	2E-8	-15.30	-22.49; -8.10	3E-5	-	-	-	-	-	-	-	-	-	-
15	TPM1	rs117484764	0.012	6.46	2E-8	-15.29	-22.49; -8.10	3E-5	-	-	-	-	-	-	-	-	-	-

SNPs that achieved $P < 5E-8$ in GWAS in familial PD are shown. They are in four LD blocks. One SNP per block was genotyped in additional samples of familial PD for replication. GWAS was conducted using Cox regression. Replication testing was conducted using Cox regression, and datasets were combined using Meta analysis. For data sets with 6 or fewer observations, Firth Penalization correction for Cox was applied. Similarly, tests for combined Discovery and Replication were conducted using Cox regression and Meta analysis. The effect on age-at-onset was calculated using linear regression. No SNPs achieved $P < 5E-8$ in non-familial PD or in all PD. For the list of signals that achieved $P < 1E-6$ see Table 3, and for genome-wide results see Supplementary Material, Tables S1–S3. CHR = chromosome, MAF = minor allele frequency, HR = age-specific Hazard Ratio calculated using Cox regression with its associated test P-value, Beta = years difference in age-at-onset per each allele (additive model) with its 95% confidence interval. – indicates not tested. Replication P values are one-sided, both for Cox and linear regression, *except for the linear regression result for rs74335301 because it was in opposite direction compared to discovery. P values for Discovery + Replication are all two-sided.

thus the weaker signals in non-familial PD cannot be attributed to lower power.

LHFPL2 and TPM1 gave no evidence for association with age-at-onset or risk in non-familial PD. LHFPL2 rs344650 was not associated with age-at-onset in non-familial PD in GWAS (Cox $P = 0.79$, $\beta = 1.87$ years) or in replication (Cox with Firth correction $P = 0.73$, $\beta = 0.90$ years). Similarly, TPM1 rs117267308 was not associated with age-at-onset in non-familial PD in GWAS (Cox $P = 1.00$, $\beta = 0.02$ years) and had only a weak trend in replication (Cox with Firth correction $P = 0.02$, $\beta = -1.80$ years), which may be due to misclassification of some familial cases as non-familial due to the difficulty in recall and knowledge of family members. When NGRC and replication were combined, neither LHFPL2 (Cox with Firth correction $P = 0.91$, $\beta = 1.25$ years) nor TPM1 (Cox with Firth correction $P = 0.06$, $\beta = -1.14$ years) was associated with age-at-onset in non-familial PD. Neither LHFPL2 (OR = 0.94, $P = 0.77$) nor TPM1 (OR = 1.18, $P = 0.53$) was associated with risk in non-familial PD. The Kaplan Meier curves best illustrate the contrast between the marked difference in genotype-specific age-at-onset distributions in familial PD (Fig 5A–D) and the lack of a difference in non-familial PD (Fig 5E–H).

All PD

No signal reached $P < 5E-8$ in all PD (Fig. 2C). Only one locus reached $P < 1E-6$ in all PD (Table 3): it was from the LPPR1 gene on chromosome 9q31.1, had similar effect sizes in familial (HR = 1.7, $\beta = -4.45$) and non-familial PD (HR = 1.7, $\beta = -5.11$), and achieved $P = 4E-7$ in the combined data. In most cases, however, loci that showed a strong signal in familial PD ($P < 1E-6$) did not have a signal in non-familial PD, and vice versa, hence the effects were diluted when all PD were combined. Genome-wide results for all PD are provided in Supplementary Material, Table S3.

Discussion

The present findings provide evidence for the existence of uncommon variants with large effects on the age-at-onset of PD. Although 28 susceptibility alleles have so far been identified for PD via GWAS, much of the heritability is still unaccounted for. As a result, modifiers of age-at-onset and rare variants are now receiving increasing attention. It was recently shown that all known PD risk loci identified via GWAS account for <1% of the 80-year variation in age-at-onset (18,22,23). The loci observed in the present study would not have been detected in prior PD GWAS because they affect age-at-onset and not risk, and because the signals are undetectable unless familial and non-familial PD are separated. The present study provides proof of concept that some of the missing heritability is in age-at-onset modifiers and uncommon variants. It demonstrates that the genetic architecture of familial and non-familial PD is only partially overlapping (modifiers that operate predominantly in one and not the other subtype produce diluted undetectable signals when all PD are combined). Our study also corroborates the results of the complex segregation analyses that predicted the existence of rare genetic variants with large effects on age-at-onset of familial PD (1,19–21).

The most significant finding was the detection and replication of two signals on chromosomes 5q14.1 and 15q22.2. Each locus achieved genome-wide significance in familial PD and had no signal in non-familial PD. The minor alleles had low frequencies (0.016 and 0.012) but each locus shifted onset age by

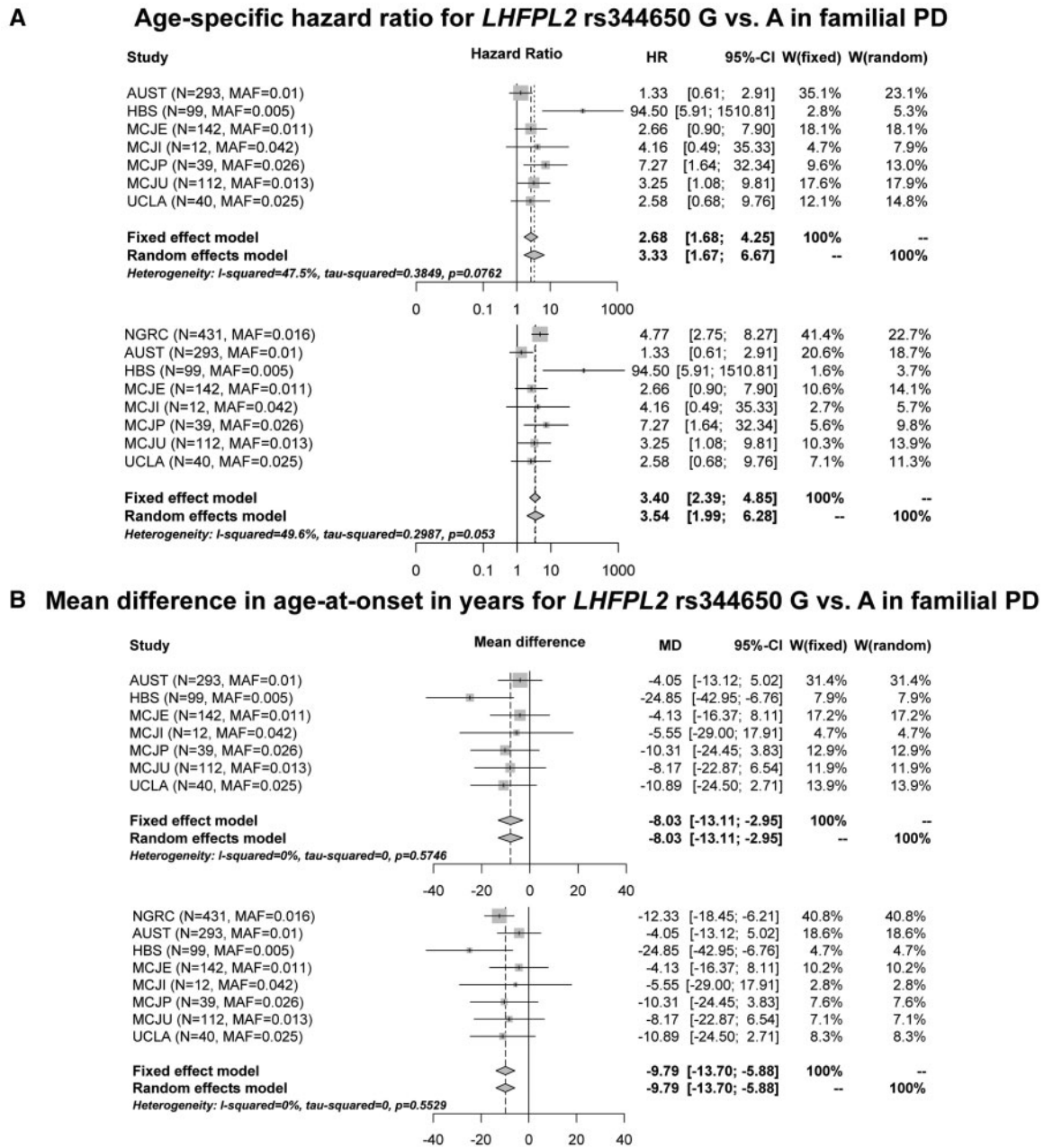


Figure 3. Replication results for rs344650 in *LHFPL2* in familial PD. In the replication datasets, excluding NGRC dataset (GWAS), the rs344650_G allele was associated with more than two-fold higher age-specific hazard ratio (HR) and approximately 8 years earlier onset than rs344650_A allele. (A). HR were generated using Cox regression, with Firth's Penalized correction for datasets with 6 or fewer observations. The forest plot depicts the HR with SE for each dataset individually, and combined using Fixed and Random Effects meta-analysis. (B) Mean differences in age-at-onset were calculated using linear regression. Additive models were used (estimates are per allele). Each panel shows the replication datasets only on top, followed by NGRC plus replication datasets. W: weight of each dataset in meta-analysis under fixed or random effects model.

10–12 years. The loci accounted for 3.5% (5q14.1) and 3.9% (15q22.2) of variation in age-at-onset.

The 5q14.1 signal maps to *LHFPL2* [MIM*609718], a member of the lipoma HMGIC fusion partner (LHFP) gene family. The function of *LHFPL2* is unknown. Interestingly, *LHFPL2* is expressed in all normal tissues and cell lines except brain and leukocytes (40); however, while healthy brain tissue has no detectable *LHFPL2* transcript, *LHFPL2* protein is abundant in malignant brain tissue (41). The 15q22.2 signal maps to the tropomyosin 1 gene (*TPM1* [MIM*191010]). *TPM1* encodes a highly conserved actin-binding protein that plays a central role in

calcium-dependent regulation of muscle contraction. *TPM1* is a tumour suppressor gene (42).

Cancer and Parkinson's disease are often likened to the two sides of a coin. Epidemiological studies have shown that the risk of developing PD is inversely associated with the risk of developing cancer (except skin cancer) (43). The pathways that lead to neuronal apoptosis, such as mitogen-activated protein kinase (MAPK) signalling, can also lead to their uncontrolled growth (44). There is also evidence from genetics for overlap, best exemplified by *PARK2*, which is both a tumour suppressor gene (45,46) and the most common cause of early-onset PD

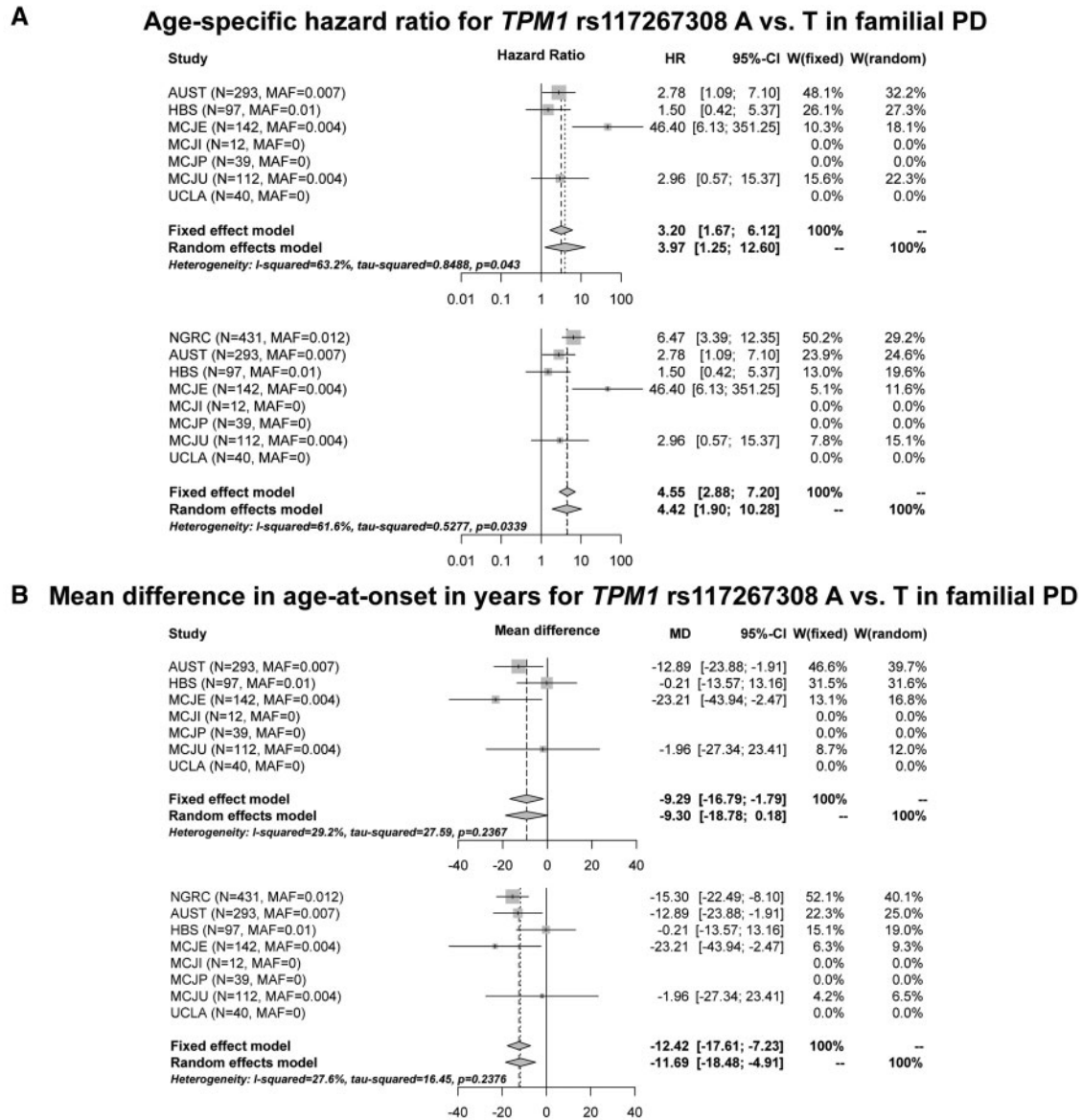


Figure 4. Replication results for rs117267308 in *TPM1* in familial PD. In the replication datasets, excluding NGRC dataset (GWAS), the rs117267308_A allele was associated with more than three-fold higher age-specific hazard ratio (HR) and approximately 9 years earlier onset than rs117267308_T allele. (A) HR were generated using Cox regression, with Firth's Penalized correction for datasets with 6 or fewer observations. The forest plot depicts the HR with SE for each dataset individually, and combined using Fixed and Random Effects meta-analysis. (B) Mean differences in age-at-onset were calculated using linear regression. Additive models were used (estimates are per allele). Each panel shows the replication datasets only on top, followed by NGRC plus replication datasets. W: weight of each dataset in meta-analysis under fixed or random effects model.

(7,47). *LHFPL2* and *TPM1* may also be genetic links between cancer and PD.

Many of the markers that associated with onset of familial PD map to sequences that are identified by the Roadmap Epigenomics Project (<http://genomebrowser.wustl.edu>) and ENCODE (48) as being active regulatory elements in the brain (Figs 7 and 8). The variants were not found in eQTL or mQTL databases Genevar (49), eqtl (<http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>), SCAN (50), or BRAINEAC (51), likely due to their low frequencies, thus we could not test their association with the expression or methylation of *LHFPL2*, *TPM1*, or adjacent genes.

We did not attempt to replicate signals that had $P > 5E-8$. It is noteworthy, however, that a block of variants mapping to

9q31.1 produced similar signals in familial (HR = 1.7, $\beta = -4.45$) and non-familial PD (HR = 1.7, $\beta = -5.11$), and when combined, the signal reached $P = 4E-7$. Low analytic power could have kept the 9q31.1 signal from reaching the significance threshold. The 9q31.1 signal maps to the neuronal plasticity gene *LPPR1* which is highly expressed in the brain and is involved in glutamate-receptor mediated neuronal excitation (52), one of the mechanisms that is believed to cause neuronal death in PD (53).

Our study was a GWAS, which was designed to detect common variants; in fact variants with $MAF < 0.01$ were excluded before analysis. If the age-at-onset modifiers for PD are uncommon alleles, as our results would suggest, our findings could be the tip of the iceberg. A related limitation was our sample size: the discovery dataset was barely powered to detect uncommon

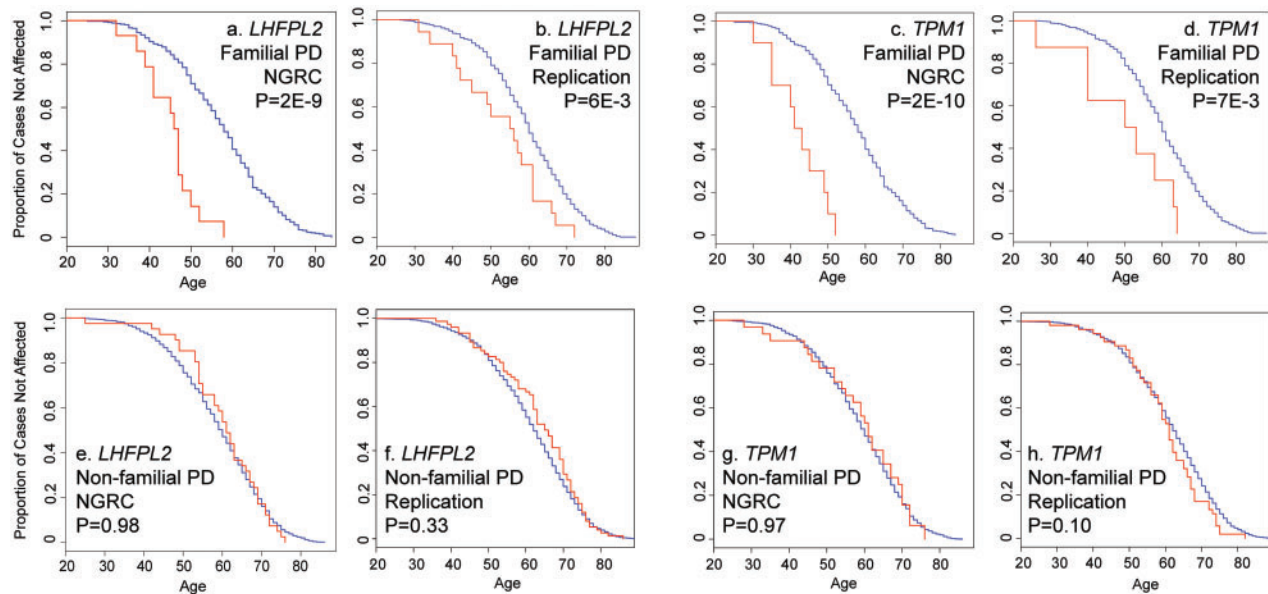


Figure 5. Kaplan-Meier plots of age-at-onset for *LHFPL2* (rs344650) and *TPM1* (rs117267308). Familial PD (A–D): *LHFPL2* genotype and *TPM1* genotype show markedly significant effects on age-at-onset of familial PD. Non-familial PD (E–H): Age at onset distributions did not vary by *LHFPL2* genotype or by *TPM1* genotype in non-familial PD. Kaplan-Meier survival curves are plots of age-specific cumulative probability of survival without disease. Here, survival is defined as not yet being affected with PD, the event is onset of PD, and the time of event is age-at-onset. Patients are divided by genotype (presence vs. absence of the minor allele), and cumulative disease-free survival is plotted for each group. Red: individuals with the minor allele. Blue: individuals without the minor allele.

variants. Given these limitations, that two loci reached genome-wide significance in discovery and replicated robustly is remarkable. Our study revealed several signals for variants that achieved $P < 1E-6$, which is promising enough to warrant studies that are specifically designed to detect and validate uncommon and rare variants.

Materials and Methods

Human subjects and data collection

Subjects: Institutional Review Boards and Human Subject Committees at participating institutions approved the study. Subject characteristics are shown in Table 1. For the discovery phase (GWAS) we used the subjects from NGRC (13). Uniform methods were used for diagnosis, subject selection, data collection, DNA preparation, genotyping, imputation, and analysis. Subjects included 2,000 individuals with the diagnosis of PD (54) whom we used to study age-at-onset, and 1,986 control subjects whom we used to rule out confounding due to associations with age. NGRC patients were on average 8 years past diagnosis, thus excluding early misdiagnoses which occur at a rate of 25% (55). Controls were free of neurodegenerative disease by self-report; a subset of older controls were examined and confirmed by neurologists to be unaffected (13). All patients and controls were American of European origin and unrelated to each other ($PI_HAT \leq 0.15$) (13). For replication, seven datasets were used, made available by investigators at Griffith University Australia (AUST) (56), Harvard Biomarker Study (HBS) (57), University of California, Los Angeles (UCLA) (58), and Mayo Clinic Jacksonville (MCJ) (56) which included four cohorts of Irish (MCJI), Polish (MCJP), and Caucasian of European descent with mixed (MCJE) or unknown (MCJU) European countries of origin. In total, replication included DNA, age-at-onset or age-at-diagnosis, family history data, sex, and age-at-enrolment on a total of 3100 persons with PD (Table 1). All subjects were Caucasian. No overlaps: We

compared all subjects across all datasets (NGRC and replications) for 74 SNP genotypes, sex, family history and age-at-onset/age-at-diagnosis. Eight pairs of individuals matched on all items. We reached out to the investigators for each dataset, obtained additional information on the 8 pairs, and were able to clear all of them as unique individuals. Additionally, we were able to confirm that there were no first-degree relatives among the carriers of *LHFPL2* or *TPM1* rare alleles across datasets.

Age-at-onset & Family history: NGRC subjects used for GWAS were recruited from neurology clinics sequentially and irrespective of age-at-onset or family history. Age-at-onset was defined as the age when the subject noticed the first motor symptom of PD. Age-at-onset was obtained at three independent occasions, several years apart: at the time of diagnosis by the movement disorder specialist as noted in medical records, at enrolment in our genetic study (59,60), and at enrolment in our environmental study (61). The three sources were compared, and inconsistencies that were >2 years were either resolved or the subject was designated as having unknown age-at-onset ($n = 1$). The outliers (onset ≤ 20 years or ≥ 89 years) were excluded from analysis ($n = 14$). Family history was obtained using a standardized self-administered questionnaire (59). Patients who reported a first or second-degree relative with PD were classified as familial PD; all others were classified as non-familial PD. Only one person per family was used. GWAS consisted of 1985 persons with PD, with known age-at-onset; 431 were familial PD and 1554 were non-familial PD. Datasets used for replication were each collected with a different study design and ascertainment method necessitating tests of heterogeneity and the use of meta-analysis. Each group had classified their samples as familial or non-familial. AUST, MCJE, MCJI, MCJP, and MCJU had collected age-at-onset. HBS and UCLA had collected age-at-diagnosis instead of age-at-onset, but age-at-diagnosis and age-at-onset are highly correlated (tested in NGRC $r^2 = 0.93$, $P < 1E-16$). For HBS and UCLA we used age-at-diagnosis instead of age-at-onset. Each dataset had either age-at-onset or age-at-diagnosis, but not a mix of both. In

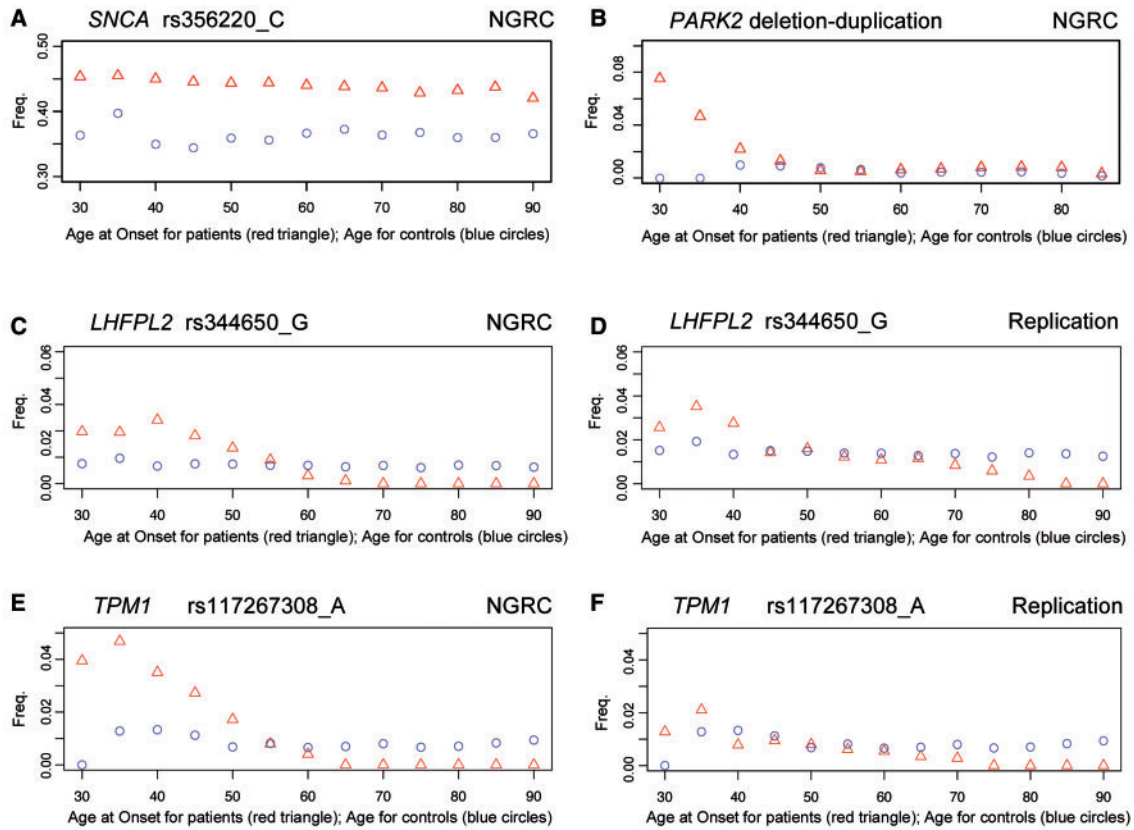


Figure 6. Moving average plots (MAP). Minor allele frequencies are plotted in a moving-average window across the age spectrum in NGRC controls (blue) and as a function of age-at-onset in patients (red). For the description of the MAP method see (38). Data are shown for the LHFPL2 rs344650_G allele and the TPM1 rs117267308_A allele, as well as for two well-established PD loci for the purpose of demonstration: SNCA rs356220, which is associated with risk in all PD (A), and PARK2 deletion/duplication, which is associated with risk of early-onset PD. (B) The MAP of SNCA rs356220 demonstrates the expected pattern for a variant that is associated with increased risk ubiquitously: allele frequency is higher in patients and parallels the control frequency, always staying higher, with no variation with age or age-at-onset. The plot for PARK2 is the signature pattern for variants that are associated with the risk of early-onset disease: allele frequency in patients is the highest in early-onset cases and decreases with increasing age-at-onset until it reaches the control frequency when it stops declining and remains superimposed on controls. LHFPL2 rs344650 has the signature pattern for an age-at-onset modifier in familial PD (C,D): accelerated onset in rs344650_G carriers causes the allele frequency to be highest in early-onset cases, decrease with increasing ages-at-onset, cross the control frequency and continue to drop below the control frequency – yet overall, rs344650_G frequency in all patients is the same as in controls. TPM1 rs117267308 exhibited a similar pattern consistent with an age-at onset modifier in familial PD (E,F).

Table 3. Signals that achieved $P < 1E-6$ in GWAS

CHR	BP	Gene	SNP	INFO	Familial PD			Non-familial PD			All PD		
					MAF	HR	P	MAF	HR	P	MAF	HR	P
Signals that Reached $P < 1E-6$ in Familial PD													
1	118901768	SPAG17	rs78024109	0.92	0.017	4.06	5E-7	0.018	1.10	0.49	0.018	1.28	0.04
3	161715295	OTOL1	rs12494760	0.92	0.019	3.84	2E-7	0.019	1.05	0.73	0.019	1.24	0.08
4	120114558	MYOZ2_USP53	rs116379732	0.95	0.013	5.14	4E-7	0.013	1.07	0.67	0.013	1.27	0.09
5	77860608	LHFPL2	rs344650	0.99	0.016	4.77	3E-8	0.014	0.96	0.79	0.014	1.16	0.25
6	105863402	PREP	rs6930232	0.98	0.158	1.65	4E-7	0.164	1.07	0.19	0.163	1.14	2E-3
7	30936024	AQP1	rs12112389	0.96	0.049	2.38	1E-7	0.052	0.96	0.64	0.051	1.09	0.24
7	129160558	SMKR1	rs62490863	0.90	0.013	5.84	8E-8	0.014	0.99	0.93	0.014	1.16	0.30
8	116638637	TRPS1	rs74335301	0.93	0.014	4.46	3E-8	0.016	1.04	0.77	0.015	1.23	0.12
10	129028001	DOCK1	rs149188358	0.97	0.011	6.02	2E-7	0.005	1.81	0.02	0.007	2.39	1E-5
14	50358528	KLHC1_ARF6	rs192855008	0.95	0.012	7.12	4E-9	0.011	0.77	0.16	0.011	0.97	0.86
15	63351500	TPM1	rs116860970	0.95	0.013	6.52	8E-9	0.011	0.88	0.48	0.012	1.11	0.51
18	73807596	LOC339298	rs11660883	0.93	0.012	5.19	5E-7	0.015	1.02	0.90	0.015	1.18	0.24
Signals that Reached $P < 1E-6$ in Non-Familial PD													
22	45356065	PHF21B	rs116305353	0.99	0.022	0.74	0.20	0.022	1.86	6E-7	0.022	1.43	1E-3
Signals that Reached $P < 1E-6$ in All PD													
9	103982633	LPPR1	rs62576890	0.92	0.025	1.72	0.02	0.024	1.73	7E-6	0.024	1.73	4E-7

Signals that reached $P < 1E-6$ in either of the three groups (familial, non-familial, all PD) are shown with the corresponding results for that signal in the other groups. Only one SNP is shown for each peak. CHR = chromosome, BP = base pair position of the top SNP (genome build 37), INFO = info score for imputed SNPs, MAF = minor allele frequency, HR = age-specific Hazard Ratio.

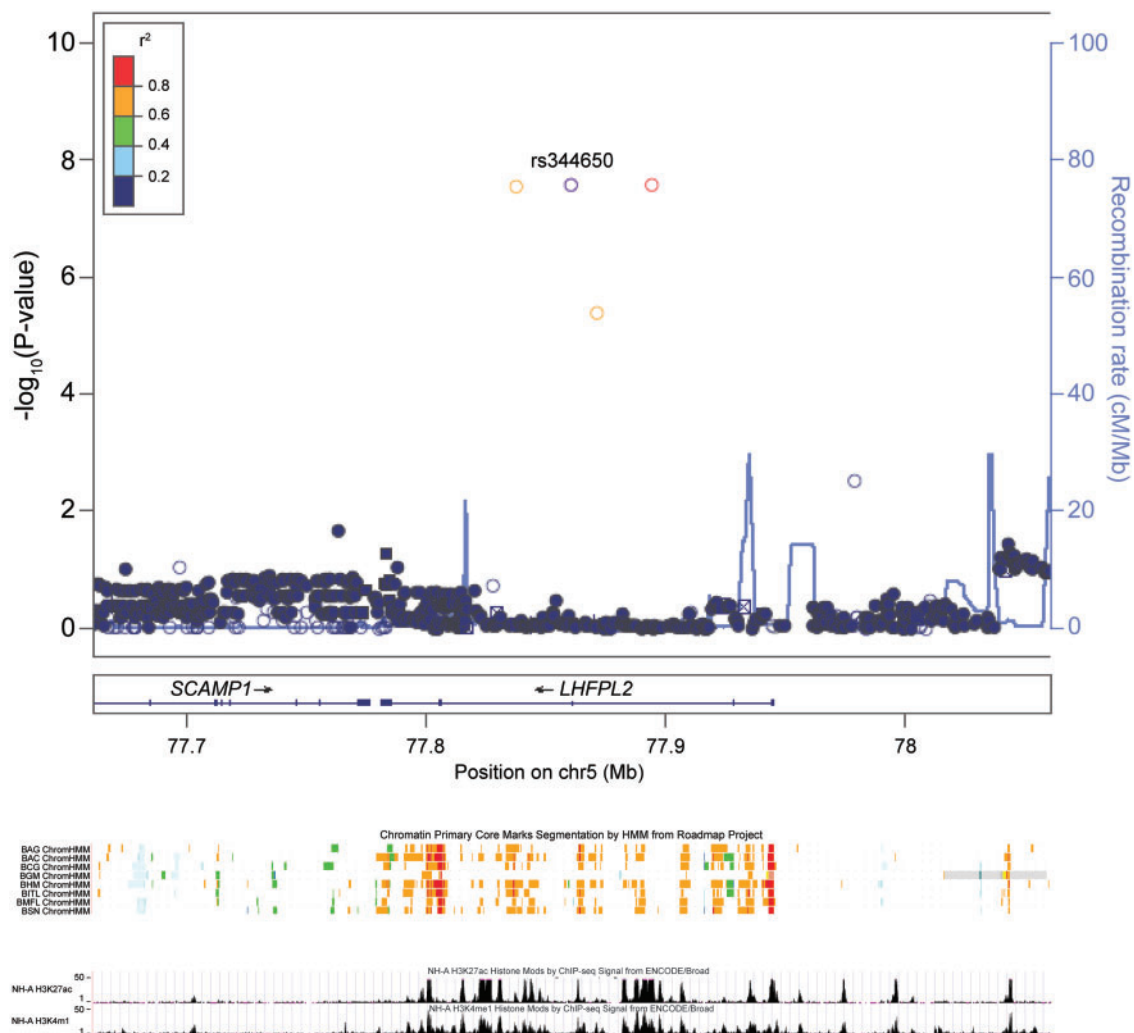


Figure 7. Alignment of *LHFPL2* variants with regulatory markers. Shown is a 400 kb segment of DNA surrounding the variants that associate with age-at-onset of PD in the *LHFPL2* region ($rs344650 \pm 200$ kb; chr5: 77,660,608–78,060,608, genome build 37). The box on top was generated using LocusZoom and shows the SNPs with their associated *P*-values (left Y-axis) and their positions on the chromosome (X-axis). *rs344650* is shown in purple. LD (r^2) was calculated in relation to *rs344650*. The colors denote the strength of LD. The top four SNPs shown in purple, red, and orange are all in the same intron. The next section is from the Roadmap Epigenomics Project and shows regulatory marks (orange=enhancers and red=transcription start sites) predicted by ChromHMM, with each line representing a different brain tissue that was analyzed (BAG=brain angular gyrus; BAC=brain anterior caudate; BCG=brain cingulate gyrus; BGM=brain germinal matrix; BHM=brain hippocampus middle; BITL=brain inferior temporal lobe; BMFL=brain mid frontal lobe; BSN=brain substantia nigra). The bottom panel is from ENCODE and shows histone acetylation and methylation marks (black) in brain cells (NH-A cell line).

total, replication included 3100 persons with PD with known age-at-onset or age-at-diagnosis; 737 were familial PD and 2363 were non-familial PD.

Genotyping and imputation

NGRC subjects were genotyped using Illumina HumanOmni1-Quad_v1-0_B BeadChips (Illumina, San Diego, CA, USA) and the Illumina Infinium II assay protocol (13). Technical genotyping quality-control criteria have been described in detail (13). The array genotyping call rate was 99.92% and reproducibility rate was $\geq 99.99\%$. Subjects who were inadvertently enrolled twice, or had cryptic relatedness ($PI-HAT > 0.15$) were excluded. SNPs were excluded if $MAF < 0.01$, call-rate $< 99\%$, $HWE P < 1E-6$, MAF difference in males vs. females > 0.15 , or missing rate in PD vs. control $P < 1E-5$. 811,597 SNPs passed quality-control measures (genotype and phenotype data for NGRC are available on dbGaP; <http://www.ncbi.nlm.nih.gov/gap>, accession number phs000196.v2.p1).

Principal component analysis (PCA) was conducted with HelixTree (<http://www.goldenhelix.com>) using a pruned subset of 104,064 SNPs, as described previously (13). No association was detected between PC 1-4 and age-at-onset in all PD (P -values for PC 1-4 = 0.09, 0.15, 0.81, 0.99), in familial PD ($P = 0.21, 0.57, 0.73, 0.66$), or in non-familial PD ($P = 0.21, 0.19, 0.80, 0.95$). Thus GWAS was carried out without adjustment for PC. However, we did reexamine the significant findings by including PC1 and PC2 in the model, and found the results to be similar and slightly more significant when corrected for PCs. Imputation was conducted using the IMPUTEv2.2.2 software (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html) (62) and the 1000 Genomes Phase I integrated variant set release v3. Imputed SNPs with info score < 0.9 or $MAF < 0.01$ were excluded. 6.4 million imputed SNPs passed quality control. In sum, GWAS included 7.2 million SNPs (0.8 million genotyped and 6.4 million imputed). Three of the four signals that reached $P < 5E-8$ were imputed. We genotyped a subset of the samples because the variants had low frequencies and the

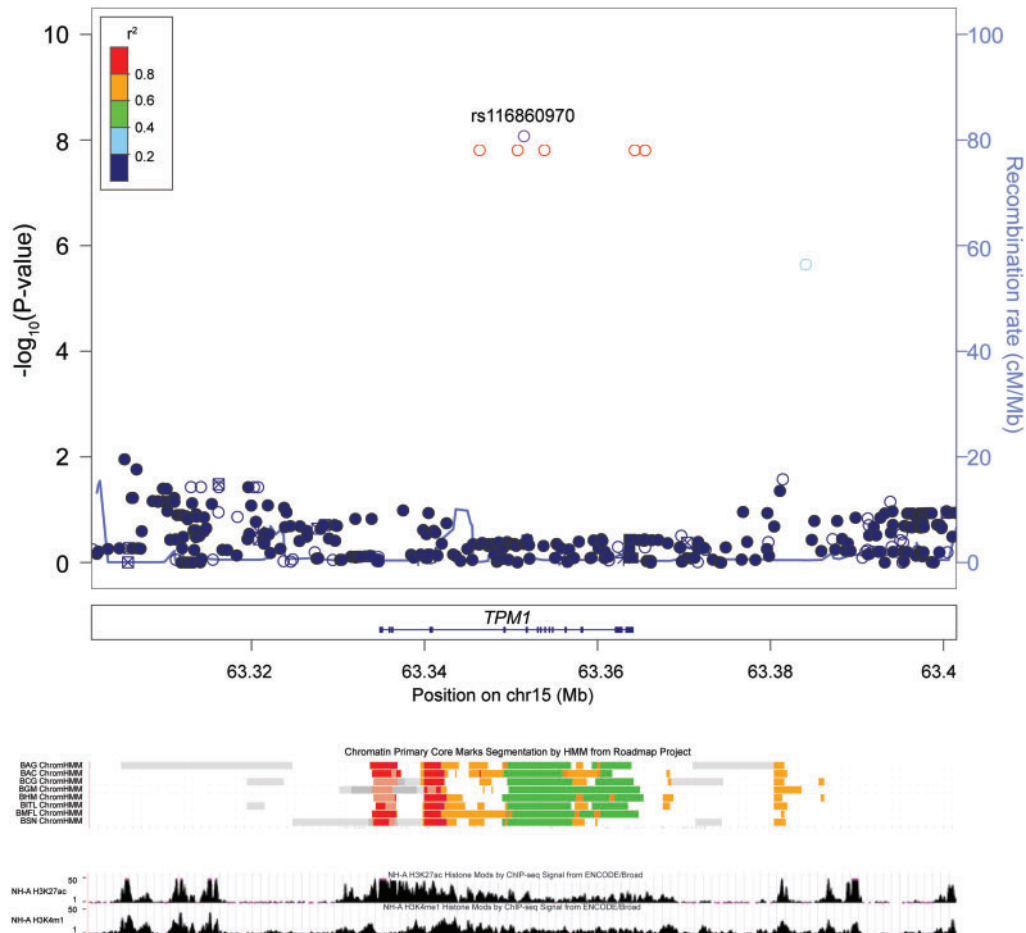


Figure 8. Alignment of *TPM1* variants with regulatory markers. Shown is a 100 kb segment of DNA surrounding the variants that associate with age-at-onset of PD in the *TPM1* region ($rs116860970 \pm 50\text{kb}$; chr15: 63,301,500–63,401,500, genome build 37). The box on top was generated using LocusZoom and shows the SNPs with their associated P-values (left Y-axis) and their positions on the chromosome (X-axis). $rs116860970$ is shown in purple. LD (r^2) was calculated in relation to $rs116860970$. The colors denote the strength of LD. The top SNPs shown in purple and red span from Intron 3 to 3' of *TPM1*. The next section is from the Roadmap Epigenomics Project and shows regulatory marks (orange=enhancers, red=transcription start sites, and green=transcribed regions) predicted by ChromHMM, with each line representing a different brain tissue that was analyzed (BAG=brain angular gyrus; BAC=brain anterior caudate; BCG=brain cingulate gyrus; BGM=brain germinal matrix; BHM=brain hippocampus middle; BITL=brain inferior temporal lobe; BMFL=brain mid frontal lobe; BSN=brain substantia nigra). The bottom panel is from ENCODE and shows histone acetylation and methylation marks (black) in brain cells (NH-A cell line).

quality of imputation for uncommon variants is unclear. For *TPM1*: 29 heterozygotes and 53 common homozygotes (no rare homozygotes were observed) as predicted by imputation were genotyped. Genotyping results were 98% concordant with imputed genotypes. For *TRPS1*: 1 rare homozygote, 28 heterozygotes, and 53 common homozygotes as predicted by imputation were genotyped. Genotyping results were 99% concordant with imputed genotypes. For *KLHDC1*: 29 heterozygotes and 53 common homozygotes (no rare homozygotes were observed) as predicted by imputation were genotyped. Genotyped results were 100% concordant with imputed genotypes. Replication samples were all directly genotyped using genomic DNA on Sequenom iPLEX (Sequenom, San Diego, CA, USA) and TaqMan assays (Life Technologies, Grand Island, NY, USA). None were imputed. Primers are available on request.

Statistical analyses

Discovery: GWAS was conducted using the Cox regression survival analysis, where age-at-onset was treated as a quantitative trait, and an additive genetic model was used for SNP genotypes:

[Survival(Age-at-onset, PD status) \sim SNP]. Using the Cox method, dosages (from 0 to 2 copies) of the minor allele of each SNP were compared, age-for-age, for the hazard of developing PD. Survival was measured as disease-free lifespan, from birth to age-at-onset. A hazard ratio (HR) and P-value was calculated for each SNP under the additive model. Significance was set at $P = 5E-8$. The “survival” package in R software (63) was used for Cox regression (<http://www.r-project.org/>). Manhattan plots were generated using Haploview v 4.2 (64). QQ plots were generated using R. Genomic inflation factors (λ) were calculated using the “GenABEL” package version 1.8-0 in R. Effect size on age-at-onset was estimated as the difference in mean age-at-onset (β) using linear regression: [Age-at-onset \sim SNP]. Linear regression was performed in ProbABEL v. 0.1-9d software (<http://www.genabel.org/packages/ProbABEL/>) (65). **Replication testing:** SNPs that generated $P < 5E-8$ in discovery were genotyped in all replication samples (familial and non-familial). Replication samples were stratified by family history for statistical testing. For each SNP, we tested the following hypotheses in replication; (a) SNP is associated with age-at-onset in familial PD, with the minor allele being associated with earlier onset, and (b) SNP is not associated with age-at-onset in non-familial PD. Each SNP was

tested in each of the replication datasets individually, using Cox regression in R, followed by meta-analyses of replication datasets using the “meta” package version 3.2-1 in R. For datasets that had 6 or fewer observations, Firth’s Penalized estimation was used to improve precision of Cox estimates (36,37). Datasets with zero observations (lacking rare allele) were not included in the Cox or linear regression, but were included in Kaplan Meier analysis. The effect size on age-at-onset was calculated for each dataset separately using linear regression in R, and then for all datasets combined using “meta” package in R. Meta-analysis forest plots were generated using the “meta” package in R. Moving Average Plots (MAP) of allele frequencies were generated using the algorithm described previously (38) and implemented in the “freqMAP” package in R. Kaplan Meier Survival plots were generated, and log-rank tests were performed using “survival” package in R. **Power:** The study was designed as a GWAS for common variants. Discovery of uncommon variants was a surprise. Post-hoc power calculation for GWAS suggested we had only ~1% power to detect variants with frequencies and effect sizes that we actually detected. The replication datasets had >80% power to detect the signals from the discovery at $P=0.05$ assuming no heterogeneity across datasets. PS program was used for power calculation (<http://biostat.mcu.vanderbilt.edu/wiki/Main/PowerSampleSize>).

Functional annotation

We used LocusZoom Version 1.1 (<http://locuszoom.sph.umich.edu/locuszoom/>) (66) to visualize the location and LD of the top association peaks. We examined Epigenomics Roadmap (via <http://genomebrowser.wustl.edu>) and ENCODE (via <http://genome.ucsc.edu/index.html>) (48) annotations of putative regulatory elements in the regions of our associated signals. We searched eQTL and mQTL databases Genevar (<https://www.sanger.ac.uk/resources/software/genevar/>) (49), eqtl (<http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>), SCAN (<http://www.scandb.org/newinterface/about.html>) (50) and BRAINEAC (<http://www.braineac.org>) (51) for eQTL or mQTL association results for the associated variants, but the variants were not found in any of the databases, likely due to their low frequencies.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We thank the persons with PD and volunteers who participated in this study. We thank Ryan J. Donahue for assistance with data management and double-checking.

Conflict of Interest statement. None declared.

Funding

This work was supported by a grant from the National Institute of Neurological Disorders And Stroke [grant number R01NS036960]. Additional support was provided by National Institutes of Health [grant number P30AG08017]; a Merit Review Award from the Department of Veterans Affairs [grant number 1I01BX000531]; Office of Research & Development, Clinical Sciences Research & Development Service, Department of Veteran Affairs; and the Close to the Cure Foundation. Genome-wide array genotyping was conducted by the Center for Inherited Disease Research, which is funded by the National Institutes of

Health [grant number HHSN268200782096C]. Studies providing samples and data for replication were supported by National Institutes of Health [grant numbers U01NS082157, P50AG005134, R01ES010544, U54ES012078, R01NS078086 and P50NS72187]; a gift from Carl Edward Bolch, Jr. and Susan Bass Bolch; the American Parkinson’s Disease Association; the Stowarzyszenie na Rzecz Rozwoju Neurologii Wieku Podeszlego grant; the Harvard NeuroDiscovery Center (HNDC); the Parkinson’s Disease Biomarkers Program (PDBP); the U.S. Department of Defense; and the M.E.M.O. Hoffman Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Funding to pay the Open Access publication charges for this article was provided by the University of Alabama at Birmingham.

References

- Hamza, T.H. and Payami, H. (2010) The heritability of risk and age at onset of Parkinson’s disease after accounting for known genetic risk factors. *J. Hum. Genet.*, **55**, 241–243.
- Bras, J., Guerreiro, R. and Hardy, J. (2015) SnapShot: Genetics of Parkinson’s disease. *Cell*, **160**, 570–570 e571.
- Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., et al. (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson’s disease. *Science*, **276**, 2045–2047.
- Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R., et al. (2003) alpha-Synuclein locus triplication causes Parkinson’s disease. *Science*, **302**, 841.
- Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R.J., Calne, D.B., et al. (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron*, **44**, 601–607.
- Paisan-Ruiz, C., Jain, S., Evans, E.W., Gilks, W.P., Simon, J., van der Brug, M., de Munain, A.L., Aparicio, S., Gil, A.M., Khan, N., et al. (2004) Cloning of the gene containing mutations that cause PARK8-linked Parkinson’s disease. *Neuron*, **44**, 595–600.
- Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y. and Shimizu, N. (1998) Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. *Nature*, **392**, 605–608.
- Valente, E.M., Abou-Sleiman, P.M., Caputo, V., Muqit, M.M., Harvey, K., Gispert, S., Ali, Z., Del Turco, D., Bentivoglio, A.R., Healy, D.G., et al. (2004) Hereditary early-onset Parkinson’s disease caused by mutations in PINK1. *Science*, **304**, 1158–1160.
- Bonifati, V., Rizzu, P., van Baren, M.J., Schaap, O., Breedveld, G.J., Krieger, E., Dekker, M.C., Squitieri, F., Ibanez, P., Joosse, M., et al. (2003) Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science*, **299**, 256–259.
- Ramirez, A., Heimbach, A., Grundemann, J., Stiller, B., Hampshire, D., Cid, L.P., Goebel, I., Mubaidin, A.F., Wriekat, A.L., Roeper, J., et al. (2006) Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. *Nat. Genet.*, **38**, 1184–1191.
- Vilarino-Guell, C., Wider, C., Ross, O.A., Dachsel, J.C., Kachergus, J.M., Lincoln, S.J., Soto-Ortolaza, A.I., Cobb, S.A., Wilhoite, G.J., Bacon, J.A., et al. (2011) VPS35 mutations in Parkinson disease. *Am. J. Hum. Genet.*, **89**, 162–167.

12. Satake, W., Nakabayashi, Y., Mizuta, I., Hirota, Y., Ito, C., Kubo, M., Kawaguchi, T., Tsunoda, T., Watanabe, M., Takeda, A., et al. (2009) Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat. Genet.*, **41**, 1303–1307.
13. Hamza, T.H., Zabetian, C.P., Tenesa, A., Laederach, A., Montimurro, J., Yearout, D., Kay, D.M., Doheny, K.F., Paschall, J., Pugh, E., et al. (2010) Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat. Genet.*, **42**, 781–785.
14. 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., et al. (2011) Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet.*, **7**, e1002141.
15. 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., et al. (2012) Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann. Neurol.*, **71**, 370–384.
16. Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., et al. (2014) Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.*, **46**, 989–993.
17. Keller, M.F., Saad, M., Bras, J., Bettella, F., Nicolaou, N., Simon-Sanchez, J., Mittag, F., Buchel, F., Sharma, M., Gibbs, J.R., et al. (2012) Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum. Mol. Genet.*, **21**, 4996–5009.
18. Nalls, M.A., Escott-Price, V., Williams, N.M., Lubbe, S., Keller, M.F., Morris, H.R. and Singleton, A.B. and International Parkinson's Disease Genomics Consortium (2015) Genetic risk and age in Parkinson's disease: Continuum not stratum. *Mov. Disord.*, **30**, 850–854.
19. Zarepari, S., Taylor, T.D., Harris, E.L. and Payami, H. (1998) Segregation analysis of Parkinson disease. *Am. J. Med. Genet.*, **80**, 410–417.
20. Maher, N.E., Currie, L.J., Lazzarini, A.M., Wilk, J.B., Taylor, C.A., Saint-Hilaire, M.H., Feldman, R.G., Golbe, L.I., Wooten, G.F. and Myers, R.H. (2002) Segregation analysis of Parkinson disease revealing evidence for a major causative gene. *Am. J. Med. Genet.*, **109**, 191–197.
21. McDonnell, S.K., Schaid, D.J., Elbaz, A., Strain, K.J., Bower, J.H., Ahlskog, J.E., Maraganore, D.M. and Rocca, W.A. (2006) Complex segregation analysis of Parkinson's disease: The Mayo Clinic Family Study. *Ann. Neurol.*, **59**, 788–795.
22. Lill, C.M., Hansen, J., Olsen, J.H., Binder, H., Ritz, B. and Bertram, L. (2015) Impact of Parkinson's disease risk loci on age at onset. *Mov. Disord.*, **30**, 847–850.
23. Pihlstrom, L. and Toft, M. (2015) Cumulative genetic risk and age at onset in Parkinson's disease. *Mov. Disord.*, **30**, 1712–1713.
24. McCulloch, C.C., Kay, D.M., Factor, S.A., Samii, A., Nutt, J.G., Higgins, D.S., Griffith, A., Roberts, J.W., Leis, B.C., Montimurro, J.S., et al. (2008) Exploring gene-environment interactions in Parkinson's disease. *Hum. Genet.*, **123**, 257–265.
25. Ritz, B.R., Manthripragada, A.D., Costello, S., Lincoln, S.J., Farrer, M.J., Cockburn, M. and Bronstein, J. (2009) Dopamine transporter genetic variants and pesticides in Parkinson's disease. *Environ. Health Perspect.*, **117**, 964–969.
26. Hamza, T.H., Chen, H., Hill-Burns, E.M., Rhodes, S.L., Montimurro, J., Kay, D.M., Tenesa, A., Kusel, V.I., Sheehan, P., Eaaswarkhanth, M., et al. (2011) Genome-Wide Gene-Environment Study Identifies Glutamate Receptor Gene GRIN2A as a Parkinson's Disease Modifier Gene via Interaction with Coffee. *PLoS Genet.*, **7**, e1002237.
27. Hill-Burns, E.M., Singh, N., Ganguly, P., Hamza, T.H., Montimurro, J., Kay, D.M., Yearout, D., Sheehan, P., Frodey, K., McLearn, J.A., et al. (2013) A genetic basis for the variable effect of smoking/nicotine on Parkinson's disease. *Pharmacogenomics J.*, **13**, 530–537.
28. 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., et al. (2009) Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.*, **41**, 1308–1312.
29. Healy, D.G., Falchi, M., O'Sullivan, S.S., Bonifati, V., Durr, A., Bressman, S., Brice, A., Aasly, J., Zabetian, C.P., Goldwurm, S., et al. (2008) Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol.*, **7**, 583–590.
30. Hill-Burns, E.M., Wissemann, W.T., Hamza, T.H., Factor, S.A., Zabetian, C.P. and Payami, H. (2014) Identification of a novel Parkinson's disease locus via stratified genome-wide association study. *BMC Genomics*, **15**, 118.
31. Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L. and Pericak-Vance, M.A. (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*, **261**, 921–923.
32. Payami, H., Kaye, J., Heston, L.L., Bird, T.D. and Schellenberg, G.D. (1993) Apolipoprotein E genotypes and Alzheimer's disease. *Lancet*, **342**, 738.
33. Scott, I.C., Seegobin, S.D., Steer, S., Tan, R., Forabosco, P., Hinks, A., Eyre, S., Morgan, A.W., Wilson, A.G., Hocking, L.J., et al. (2013) Predicting the risk of rheumatoid arthritis and its age of onset through modelling genetic risk variants with smoking. *PLoS Genet.*, **9**, e1003808.
34. Forno, E., Lasky-Su, J., Himes, B., Howrylak, J., Ramsey, C., Brehm, J., Klanderma, B., Ziniti, J., Melen, E., Pershagen, G., et al. (2012) Genome-wide association study of the age of onset of childhood asthma. *J. Allergy Clin. Immunol.*, **130**, 83–90. e84.
35. Hoffmann, T.J. and Witte, J.S. (2015) Strategies for Imputing and Analyzing Rare Variants in Association Studies. *Trends Genet.*, **31**, 556–563.
36. Heinze, G. and Dunkler, D. (2008) Avoiding infinite estimates of time-dependent effects in small-sample survival studies. *Stat. Med.*, **27**, 6455–6469.
37. Lin, I.F., Chang, W.P. and Liao, Y.N. (2013) Shrinkage methods enhanced the accuracy of parameter estimation using Cox models with small number of events. *J. Clin. Epidemiol.*, **66**, 743–751.
38. Payami, H., Kay, D.M., Zabetian, C.P., Schellenberg, G.D., Factor, S.A. and McCulloch, C.C. (2009) Visualizing disease associations: graphic analysis of frequency distributions as a function of age using moving average plots (MAP) with application to Alzheimer's and Parkinson's disease. *Genet. Epidemiol.*, **34**, 92–99.
39. Kay, D.M., Moran, D., Moses, L., Poorkaj, P., Zabetian, C.P., Nutt, J., Factor, S.A., Yu, C.E., Montimurro, J.S., Keefe, R.G., et al. (2007) Heterozygous parkin point mutations are as common in control subjects as in Parkinson's patients. *Ann. Neurol.*, **61**, 47–54.
40. Nagase, T., Seki, N., Ishikawa, K., Ohira, M., Kawarabayasi, Y., Ohara, O., Tanaka, A., Kotani, H., Miyajima, N. and

- Nomura, N. (1996) Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain. *DNA Res*, **3**, 321–329. 341–354.
41. Schaab, C., Geiger, T., Stoehr, G., Cox, J. and Mann, M. (2012) Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Mol. Cell Proteomics*, **11**, M111 014068.
 42. Raval, G.N., Bharadwaj, S., Levine, E.A., Willingham, M.C., Geary, R.L., Kute, T. and Prasad, G.L. (2003) Loss of expression of tropomyosin-1, a novel class II tumor suppressor that induces anoikis, in primary breast tumors. *Oncogene*, **22**, 6194–6203.
 43. Bajaj, A., Driver, J.A. and Schernhammer, E.S. (2010) Parkinson's disease and cancer risk: a systematic review and meta-analysis. *Cancer Causes Control*, **21**, 697–707.
 44. Kim, E.K. and Choi, E.J. (2010) Pathological roles of MAPK signaling pathways in human diseases. *Biochim. Biophys. Acta*, **1802**, 396–405.
 45. Cesari, R., Martin, E.S., Calin, G.A., Pentimalli, F., Bichi, R., McAdams, H., Trapasso, F., Drusco, A., Shimizu, M., Masciullo, V., et al. (2003) Parkin, a gene implicated in autosomal recessive juvenile parkinsonism, is a candidate tumor suppressor gene on chromosome 6q25-q27. *Proc. Natl. Acad. Sci.*, **100**, 5956–5961.
 46. Veeriah, S., Taylor, B.S., Meng, S., Fang, F., Yilmaz, E., Vivanco, I., Janakiraman, M., Schultz, N., Hanrahan, A.J., Pao, W., et al. (2010) Somatic mutations of the Parkinson's disease-associated gene PARK2 in glioblastoma and other human malignancies. *Nat. Genet.*, **42**, 77–82.
 47. Kay, D.M., Stevens, C.F., Hamza, T.H., Montimurro, J.S., Zabetian, C.P., Factor, S.A., Samii, A., Griffith, A., Roberts, J.W., Molho, E.S., et al. (2010) A comprehensive analysis of deletions, duplications, and copy number variations in PARK2. *Neurology*, **75**, 1189–1194.
 48. Ram, O., Goren, A., Amit, I., Shoshitaishvili, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M., et al. (2011) Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell*, **147**, 1628–1639.
 49. Yang, T.P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P. and Dermitzakis, E.T. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics*, **26**, 2474–2476.
 50. Gamazon, E.R., Zhang, W., Konkashbaev, A., Duan, S., Kistner, E.O., Nicolae, D.L., Dolan, M.E. and Cox, N.J. (2010) SCAN: SNP and copy number annotation. *Bioinformatics*, **26**, 259–262.
 51. Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T., Consortium, U.K.B.E., North American Brain Expression, C., Coin, L. et al. (2014) Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.*, **17**, 1418–1428.
 52. Savaskan, N.E., Brauer, A.U. and Nitsch, R. (2004) Molecular cloning and expression regulation of PRG-3, a new member of the plasticity-related gene family. *Eur. J. Neurosci.*, **19**, 212–220.
 53. Greenamyre, J.T. and Porter, R.H. (1994) Anatomy and physiology of glutamate in the CNS. *Neurology*, **44**, S7–13.
 54. Gibb, W. and Lees, A. (1988) The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry.*, **51**, 745–752.
 55. Hughes, A.J., Daniel, S.E., Ben-Shlomo, Y. and Lees, A.J. (2002) The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain*, **125**, 861–870.
 56. 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., et al. (2011) Association of LRRK2 exonic variants with susceptibility to Parkinson's disease: a case-control study. *Lancet Neurol*, **10**, 898–908.
 57. Ding, H., Dhima, K., Lockhart, K.C., Locascio, J.J., Hoising, A.N., Duong, K., Trisini-Lipsanopoulos, A., Hayes, M.T., Sohr, U.S., Wills, A.M., et al. (2013) Unrecognized vitamin D3 deficiency is common in Parkinson disease: Harvard Biomarker Study. *Neurology*, **81**, 1531–1537.
 58. Costello, S., Cockburn, M., Bronstein, J., Zhang, X. and Ritz, B. (2009) Parkinson's disease and residential exposure to maneb and paraquat from agricultural applications in the central valley of California. *Am. J. Epidemiol.*, **169**, 919–926.
 59. Payami, H., Larsen, K., Bernard, S. and Nutt, J. (1994) Increased risk of Parkinson's disease in parents and siblings of patients. *Ann. Neurol.*, **36**, 659–661.
 60. Kay, D.M., Zabetian, C.P., Factor, S.A., Nutt, J.G., Samii, A., Griffith, A., Bird, T.D., Kramer, P., Higgins, D.S. and Payami, H. (2006) Parkinson's disease and LRRK2: frequency of a common mutation in U.S. movement disorder clinics. *Mov. Disord.*, **21**, 519–523.
 61. Powers, K., Kay, D., Factor, S., Zabetian, C., Higgins, D., Samii, A., Nutt, J., Griffith, A., Leis, B., Roberts, J., et al. (2008) Combined effects of smoking, coffee and NSAIDs on Parkinson's disease risk. *Mov. Disord.*, **23**, 88–95.
 62. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
 63. Therneau, T. and Grambsch, P. (2000) *Modeling Survival Data: Extending the Cox Model*. Springer, New York.
 64. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
 65. Aulchenko, Y.S., Struchalin, M.V. and van Duijn, C.M. (2010) ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*, **11**, 134.
 66. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.