

# Genome-Wide Association Study Confirms SNPs in *SNCA* and the *MAPT* Region as Common Risk Factors for Parkinson Disease

Todd L. Edwards<sup>1,2</sup>, William K. Scott<sup>1</sup>, Cherylyn Almonte<sup>1</sup>, Amber Burt<sup>1</sup>, Eric H. Powell<sup>1</sup>, Gary W. Beecham<sup>1</sup>, Liyong Wang<sup>1</sup>, Stephan Züchner<sup>1</sup>, Ioanna Konidari<sup>1</sup>, Gaofeng Wang<sup>1</sup>, Carlos Singer<sup>4</sup>, Fatta Nahab<sup>4</sup>, Burton Scott<sup>5</sup>, Jeffrey M. Stajich<sup>5</sup>, Margaret Pericak-Vance<sup>1</sup>, Jonathan Haines<sup>3</sup>, Jeffery M. Vance<sup>1</sup> and Eden R. Martin<sup>1,\*</sup>

<sup>1</sup>John P. Hussman Institute for Human Genomics and the Dr. John T. Macdonald Foundation Department of Human Genetics, Miller School of Medicine, University of Miami, FL

<sup>2</sup>Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Vanderbilt University, TN

<sup>3</sup>Center for Human Genetics Research, Vanderbilt University Medical Center, Vanderbilt University, TN

<sup>4</sup>Department of Neurology, Miller School of Medicine, University of Miami, FL

<sup>5</sup>Department of Medicine, Duke University Medical Center, Duke University, NC

## Summary

Parkinson disease (PD) is a chronic neurodegenerative disorder with a cumulative prevalence of greater than one per thousand. To date three independent genome-wide association studies (GWAS) have investigated the genetic susceptibility to PD. These studies implicated several genes as PD risk loci with strong, but not genome-wide significant, associations.

In this study, we combined data from two previously published GWAS of Caucasian subjects with our GWAS of 604 cases and 619 controls for a joint analysis with a combined sample size of 1752 cases and 1745 controls. SNPs in *SNCA* (rs2736990, p-value =  $6.7 \times 10^{-8}$ ; genome-wide adjusted p = 0.0109, odds ratio (OR) = 1.29 [95% CI: 1.17–1.42] G vs. A allele, population attributable risk percent (PAR%) = 12%) and the *MAPT* region (rs11012, p-value =  $5.6 \times 10^{-8}$ ; genome-wide adjusted p = 0.0079, OR = 0.70 [95% CI: 0.62–0.79] T vs. C allele, PAR% = 8%) were genome-wide significant. No other SNPs were genome-wide significant in this analysis. This study confirms that *SNCA* and the *MAPT* region are major genes whose common variants are influencing risk of PD.

Keywords: Parkinson disease, association study, alpha-synuclein, microtubule associated protein tau

## Introduction

Parkinson disease [PD (OMIM 168600)] is a chronic neurodegenerative disorder with a cumulative prevalence of greater than one per thousand (Kuopio et al., 1999) with at least 1.5 million cases in the United States and 6 million worldwide (Thomas & Beal, 2007). Some genetic contributions to PD are well recognised. Mutations with high penetrance were initially identified in PD due to their easily detectable

effects in relatively rare, early-onset, Mendelian forms of PD (Polymeropoulos et al., 1997; Kitada et al., 1998; Bonifati et al., 2003; Paisan-Ruiz et al., 2004; Zimprich et al., 2004). These known mutations explain less than 10% of PD cases (Lesage & Brice, 2009). Over the last several years significant effort has been focused on investigating the contributions of common variants to PD risk and age-at-onset. Candidate gene approaches to identify genetic associations with PD have been used to follow up family-based genome-wide linkage studies to identify genomic regions containing risk loci (reviewed by Lesage & Brice, 2009). This focused approach restricts the number of association tests performed, but is limited to identifying loci detectable by linkage analysis (Risch & Merikangas,

\*Corresponding author: Eden R. Martin, John P. Hussman Institute for Human Genomics, 1501 NW 10<sup>th</sup> Avenue, BRB 305 (M-860), Miami FL 33136. Tel: 305-243-2372; Fax: 305-243-1968; E-mail: Emartin1@med.miami.edu

1996). More recently three genome-wide association studies (GWAS) have been conducted in PD, albeit with results that did not reach genome-wide significance (Maraganore et al., 2005; Fung et al., 2006a; Pankratz et al., 2009). While genes that achieve genome-wide significance in a complex disease GWAS are important, the presence of genetic heterogeneity, unobserved environmental interactions and sub-phenotypes with distinct genetic etiologies can all reduce the apparent contribution of important genes to below the threshold for genome-wide significance, requiring that large samples be assembled for adequate power.

Associations with PD have been replicated in the candidate gene and GWAS contexts, including those described early in PD association studies, such as alpha-synuclein (*SNCA*, Entrez Gene ID (EGI):6622) (Kruger et al., 1999; Farrer et al., 2001; Mueller et al., 2005; Maraganore et al., 2006; McCulloch et al., 2008; Myhre et al., 2008; Sutherland et al., 2009) and the microtubule-associated protein tau (*MAPT*, EGI:4137) inversion region on chromosome 17 (Martin et al., 2001; Scott et al., 2001; Zappia et al., 2003; Healy et al., 2004; Kwok et al., 2004; Levecque et al., 2004; Skipper et al., 2004; Mamah et al., 2005; Fidani et al., 2006; Fung et al., 2006b; Goris et al., 2007; Vandrovcova et al., 2007; Winkler et al., 2007; Zabetian et al., 2007), as well as ubiquitin-specific protease 24 (*USP24*, EGI:23358) (Oliveira et al., 2005; Li et al., 2006; Haugarvoll et al., 2009), ELAV-like 4 (*ELAVL4*, EGI:1996) (Noureddine et al., 2005; Haugarvoll et al., 2007; DeStefano et al., 2008), monoamine oxidase B (*MAOB*, EGI:4129) (Kurth et al., 1993), Apolipoprotein E (*APOE*, EGI:348) (Rubinsztein et al., 1994), and the mitochondrial haplogroups (Kosel et al., 1998; Ross et al., 2003; van der Walt et al., 2003; Autere et al., 2004; Ghezzi et al., 2005; Huerta et al., 2005; Pyle et al., 2005; Gaweda-Walerych et al., 2008). The consistency of results, particularly for *SNCA* and *MAPT*, suggest that the failure to reach genome-wide significance in previous studies is due to the relatively small GWAS datasets.

We have conducted a GWAS of PD at the Hussman Institute for Human Genomics at the University of Miami (HIHG) in a sample of 604 unrelated cases and 619 unrelated controls using 491,376 autosomal and sex chromosome SNPs. To increase the power of the analysis we included data from two of the three previous PD GWAS: The National Institute of Neurological Disorders and Stroke (NINDS) (Fung et al., 2006a), and the joint dataset from the Progeni/GenePD studies that was genotyped at the Center for Inherited Disease Research (CIDR) (Pankratz et al., 2009), and excluding the Mayo clinic GWAS which used discordant sibling pairs as subjects for analysis (Maraganore et al., 2005). This provided a joint analysis dataset with a combined sample size of 1752 cases and 1745 controls with genotypes at 495,715 SNPs after imputation and sample and SNP quality control procedures.

We demonstrate in our Caucasian-based population that the *SNCA* and *MAPT* regions are the strongest genetic contributors to PD risk, reaching genome-wide significance and establishing these factors without controversy. In addition, several genes replicated in all three datasets, but with less stringent significance. Although they did not achieve genome-wide significance in the joint analysis, the consistency of their effects makes them strong candidates and may provide additional insight into the pathological mechanisms of PD.

## Materials and Methods

### Samples

Samples in the HIHG GWAS include individuals with PD collected by one of 13 ascertainment centers in the PD Genetics Collaboration (Scott et al., 2001) or by the Morris K. Udall Parkinson Disease Center of Excellence (J.M. Vance, PI) ascertainment core. These participants were recruited by participating movement disorder and neurology clinics, referrals, and advertisements. Unaffected spouse and friend controls were recruited when available and willing to participate. All participants provided written informed consent, in accord with protocols established by institutional review boards at each centre.

All individuals with PD were examined by a board-certified neurologist. A neurological exam and standard clinical evaluation was performed on all participants with PD. Affected individuals exhibited at least two cardinal symptoms of PD, e.g. bradykinesia, resting tremor, and rigidity and no other causes of parkinsonism or atypical clinical features. Unaffected individuals had no symptoms of PD upon physical examination and self-reported symptom questionnaire (Rocca et al., 1998). Individuals were excluded if there was a history of encephalitis, neuroleptic therapy within one year before diagnosis, evidence of normal pressure hydrocephalus, or a clinical course with unusual features suggesting atypical or secondary Parkinsonism. Additionally, a blood sample, family history, medical history, and standard cognitive test (Blessed Orientation Memory Concentration (BOMC) test (Katzman et al., 1983) or Modified Mini Mental Status exam (3MS) (Folstein et al., 1975)) were obtained for each individual. To ensure diagnostic consistency across sites, clinical data for all participants were reviewed by a panel consisting of a board-certified neurologist with fellowship training in movement disorders, a board certified neurologist and medical geneticist, and a certified physician assistant.

### Genotyping

Genotypes for 635 PD cases and 255 PD controls were generated using the Illumina Infinium 610-quad BeadChip (Illumina, San Diego, CA, USA) and the Illumina Infinium II assay protocol (Gunderson et al., 2005). Additionally, we included 223 cognitively-normal controls with no PD symptoms by self-reported symptom questionnaire (Rocca et al., 1998) from

a previous GWAS (Beecham et al., 2009) of late-onset Alzheimer disease (LOAD) genotyped using the Illumina HumanHap 550 BeadChip, and another 164 cognitively-normal controls from a second LOAD study with no self-reported PD symptoms by questionnaire (Rocca et al., 1998) genotyped using the 1M-Duo Infinium HD BeadChip. Genotypes were determined using Illumina BeadStudio Genotyping Module version 3.2.33; samples with 99% genotyping efficiency were used to redefine genotype clusters, per the manufacturer's recommendation. Concordance of genotype calls for two CEPH samples with six replicates each was 99.98%.

### Quality Control

For the HIHG data, samples with genotyping efficiency of greater than 98% were included in subsequent quality control (QC) and statistical analysis steps. One case and seven control samples were removed for low efficiency. Population stratification was assessed using Structure (Pritchard et al., 2000; Falush et al., 2003) and Eigenstrat (Price et al., 2006). For Structure analysis, 5000 independent autosomal SNPs with minor allele frequency (MAF) > 0.25 were chosen using PLINK (Purcell et al., 2007) with an  $r^2$  threshold of 0.2 using 10,000 iterations of burn-in and 15,000 iterations of estimation. These analyses indicated that no stratification was present in our sample (Fig. S1a). For Eigenstrat analysis, 30,000 independent autosomal SNPs with MAF > 0.25 were used to generate plots of principal component loadings for samples and to remove outliers (6 cases, 7 controls) using the top ten principal components over five iterations with a threshold of six standard deviations (Fig. S1b).

Further quality control steps for the HIHG samples included checks for duplicated and related samples using mean identity by state (22 cases were removed as duplicate samples). Two cases and nine controls were identified for excess or insufficient heterozygosity at sex chromosomes and were removed. Finally SNPs with MAF < 0.01 in the combined case-control dataset were removed (21,976 SNPs), SNPs with significant deviation from Hardy-Weinberg equilibrium (HWE) in controls were removed (78 SNPs with  $p < 1.00 \times 10^{-7}$ ), and differential missing rate between cases and controls (29 SNPs with  $p < 1 \times 10^{-5}$ ) were removed.

Since controls from the LOAD GWAS dataset were genotyped on different Illumina chips, we examined all controls for homogeneity of genotype frequencies. This filter is used to detect where genotyping error rates differ across studies, and is a critical QC step when multiple genotyping experiments are combined into a single data set for analysis. A 4 degree of freedom (df) Fisher's exact test with 10,000 permutations was used at all SNPs to test frequency differences for genotypes generated from the three genotyping chips in the study. SNPs were removed if the  $p$ -value for this test was less than 0.001 (546 SNPs). After all QC procedures, 604 cases and 619 controls with 491,376 SNPs were available for association analysis in the HIHG GWAS dataset.

The sex and age distributions for these samples are described in Table 1.

Data were downloaded from dbGAP (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap>) for the CIDR (Pankratz et al., 2009) and NINDS (Fung et al., 2006a) PD GWAS. Imputation of SNP genotypes from these and the HIHG data was performed independently using the software package Impute (Marchini et al., 2007). Samples were removed from these GWAS for genotyping efficiency < 0.98 (CIDR: 9 cases, 4 controls removed; NINDS 11 cases, 9 controls removed). SNP data for these GWAS were removed for genotyping efficiency < 0.98 (CIDR: 7886 SNPs removed; NINDS: 16,886 SNPs removed), MAF < 0.01 (CIDR: 7,676 SNPs removed; NINDS: 4596 SNPs removed), HWE  $p < 10^{-7}$  (CIDR: 790 SNPs removed; NINDS: 67 SNPs removed), and differential missingness by disease status  $p < 10^{-5}$  (CIDR: 29 SNPs removed). Eigenstrat and Structure were used to ensure that the merged dataset did not contain stratified samples (Figs. S2a, b). The Fisher's exact test for homogeneity described above was run in controls for all 5 genotyping platforms using an 8 df test, removing SNPs for  $p < 0.001$ . Imputation was performed on each dataset independently after QC filters had been applied. The remaining high-quality samples and genotypes from each study were used to impute SNPs based on the HapMap reference panel. Individual genotypes with probability of 90% or greater were included. Imputed SNPs with greater than 5% missing genotypes were excluded from analysis. Upon merging the imputed GWAS files in PLINK, HWE, MAF, SNP genotyping efficiency, and tests for differential missing rate by phenotype filters were applied as described above. The final merged dataset for analysis contained 1752 cases and 1745 controls with genotypes at 495,715 genotyped and imputed SNPs.

Association analysis of the genotype data was conducted with PLINK (Purcell et al., 2007). Cochran-Armitage (Armitage, 1955) trend tests were calculated at each SNP to assess allelic association. Additional analyses evaluating dominant, recessive and genotypic exposures were performed using logistic regression in PLINK. To avoid over-correcting for multiple comparisons by using the conservative Bonferroni correction, we used PLINK to generate empirically adjusted  $p$ -values based on 10,000 permutations to adjust for multiple tests (Purcell et al., 2007). Additional logistic regression analysis fitting covariates for age of onset/age at exam, sex or history of smoking was conducted to assess confounding by these variables at all SNPs. If there was an indication that an individual had reported any history of smoking, they were scored as a smoker, a non-smoker if they reported not smoking, and missing otherwise.

The population proportional attributable risk (PAR%) was calculated using the formula for retrospective studies,  $[(\text{SNP allele frequency}) \times (\text{OR}-1)] / [1 + (\text{SNP allele frequency}) \times (\text{OR}-1)]$ , using the SNP allele frequency in controls as the exposure frequency, and the odds ratio (OR) in place of the relative risk (Woodward, 2005).

Assessment of associated SNPs and the presence of chromosome 17q21.31 alleles in the H1-H2 haplotype clades in *MAPT* was accomplished using rs1981997 as a haplotype tag SNP,

**Table 1** Demographic properties of the HIHG, CIDR and NINDS samples.

	HIHG		CIDR		NINDS	
	All HIHG	Cases	Controls	All CIDR	Cases	Controls
Male (freq)	595(0.49)	381(0.63)	214(0.34)	875(0.50)	358(0.41)	517(0.60)
Female (freq)	631(0.51)	224(0.37)	407(0.66)	869(0.50)	523(0.59)	346(0.40)
AAO(SD)	—	56.39(12.95)	—	—	62.34(10.32)	—
AAE(SD)	66.97(11.31)	64.56(12.18)	69.32(9.86)	70.99(9.32)	71.21(8.46)	54.83(13.08)
AAO—Age at onset of PD						
AAE—Age at examination						
SD—Standard deviation						
freq—frequency						
				All NINDS	Cases	Controls
				339(0.54)	162(0.61)	127(0.48)
				289(0.46)	104(0.39)	136(0.52)
				—	65.33(7.86)	—
				71.73(7.82)	71.99(6.53)	69.69(8.77)

because the major (G) and minor (A) alleles of this SNP are fixed in the H1 and H2 clades respectively (Stefansson et al., 2005). A 2-locus haplotype association analysis was performed using PLINK with SNPs in the *MAPT* region to determine which alleles were on the H1 haplotype, which has previously been associated with PD Golbe et al., 2001; Maraganore et al., 2001; Martin et al., 2001; Farrer et al., 2002; Skipper et al., 2004; Zabetian et al., 2007; Refenes et al., 2009).

In addition to the statistical analysis of the joint sample, as a final check against cryptic bias arising due to genotyping error rate heterogeneity across studies, we evaluated association with meta-analysis techniques using METAL (Abecasis & Willer, 2007).

## Results

In the HIHG sample no SNPs were statistically significantly associated with PD at the genome-wide level using permutation tests with a multiple testing corrected threshold of  $p < 0.05$  (Table 2). *SNCA* showed the strongest association in the HIHG study (*SNCA* intron, rs356220  $p = 2.7 \times 10^{-6}$ , OR = 1.48; 95% CI [1.25–1.74]). Other top associations included SNPs in the chromosome 1p22 gene chloride channel accessory 4 (*CLCA4*, EGI:22802, intron, rs1543467  $p = 3.0 \times 10^{-6}$ , OR = 1.51; 95% CI [1.27–1.80]), the chromosome 3q26 gene neuroligin 1 (*NLGN1*, EGI:22871, intron, rs976683  $p = 1 \times 10^{-5}$ , OR = 1.49; 95% CI [1.25–1.78]) and the chromosome 20p11 gene solute carrier family 24, member 3 (*SLC24A3*, EGI:57419, intronic SNPs, rs1406968  $p = 1.2 \times 10^{-5}$ , OR = 1.47; 95% CI [1.24–1.75]; rs4813368  $p = 1.2 \times 10^{-5}$ , OR = 1.48; 95% CI [1.24–1.76];  $r^2$  for rs1406968:rs4813368 = 0.94). Neither sex nor age-at-onset (age-at-exam in controls) were confounders for the top associations in the HIHG sample.

Two SNPs were statistically significant after empirical adjustment for multiple comparisons (permutation  $p < 0.05$ ) in the joint analysis of the HIHG, CIDR and NINDS samples (Table 3, Table S1); notably these SNPs were genotyped, not imputed, in all three datasets. These SNPs are in the genes pleckstrin homology domain-containing protein, family M, member 1 (*PLEKHM1*, EGI:9842, promoter, rs11012  $p = 5.6 \times 10^{-8}$ , OR = 0.71; 95% CI [0.62–0.79]), and alpha-synuclein (*SNCA* intron, rs2736990  $p = 6.7 \times 10^{-8}$ , OR = 1.30; 95% CI [1.18–1.43]). In addition to rs2736990 in *SNCA*, there were two SNPs with  $9 \times 10^{-5} > p\text{-value} > 1 \times 10^{-5}$  (rs11931074, rs356220), two SNPs with  $9 \times 10^{-4} > p\text{-value} > 1 \times 10^{-4}$  (rs365188, rs1866995), two with  $9 \times 10^{-3} > p\text{-value} > 1 \times 10^{-3}$  (rs2583985, rs3775439) as well as three more SNPs with  $p < 0.05$  (rs3857059, rs12502363, rs3822095), out of 15 total *SNCA* SNPs. The SNP rs356220, which was the most

SNP rs#	Trend p	HWE p	Chromosome	Gene
rs356220	$2.67 \times 10^{-6}$	0.930	4q22.1	<i>SNCA</i>
rs1543467	$2.97 \times 10^{-6}$	0.117	lp22.3	<i>CLCA4</i>
rs12063142	$5.02 \times 10^{-6}$	0.468	lp36.13	–
rs9513249	$5.92 \times 10^{-6}$	0.011	13q32.1	–
rs12870589	$9.45 \times 10^{-6}$	0.002	13q32.2	–
rs976683	$1.04 \times 10^{-5}$	0.529	3q26.31	<i>NLGN1</i>
rs1816879	$1.16 \times 10^{-5}$	0.784	15q22.1	–
rs1406968	$1.16 \times 10^{-5}$	0.705	20p11.23	<i>SLC24A3</i>
rs4813368	$1.22 \times 10^{-5}$	0.701	20p11.23	<i>SLC24A3</i>
rs7646773	$1.30 \times 10^{-5}$	0.865	3pll.2	–
rs1992695	$1.34 \times 10^{-5}$	0.422	4q28.3	–
rs7322222	$1.35 \times 10^{-5}$	0.061	13q32.1	–
rs10851073	$1.44 \times 10^{-5}$	0.001	13q32.2	–
rs4238458	$1.52 \times 10^{-5}$	0.407	15q24.1	–
rs11625012	$1.61 \times 10^{-5}$	0.507	14q22.1	–
rs7698161	$1.76 \times 10^{-5}$	0.738	4q28.3	–
rs7496513	$1.85 \times 10^{-5}$	0.402	15q24.1	–
rs135066	$1.91 \times 10^{-5}$	1.000	22ql3.2	<i>MPPED1</i>
rs13032621	$1.95 \times 10^{-5}$	1.000	2q37.2	–
rs9457743	$2.03 \times 10^{-5}$	0.931	6q25.3	–
rs1159278	$2.21 \times 10^{-5}$	0.025	13q32.1	<i>RAP2A</i>
rs12142266	$2.31 \times 10^{-5}$	0.406	lp22.3	–
rs358079	$2.99 \times 10^{-5}$	0.794	3pl4.3	<i>CACNA12D3</i>
rs13411180	$3.09 \times 10^{-5}$	0.832	2q31.2	<i>ZNF533</i>
rs13009601	$3.13 \times 10^{-5}$	0.848	2q31.2	<i>ZNF533</i>
rs6930229	$3.14 \times 10^{-5}$	1.000	6q25.3	–
rs11515274	$3.39 \times 10^{-5}$	0.850	2q31.2	<i>ZNF533</i>
rs13157	$3.52 \times 10^{-5}$	1.000	lp36.11	<i>RUNX3</i>
rs929708	$3.70 \times 10^{-5}$	0.378	3p25.3	<i>ATP2B2</i>
rs10882088	$3.89 \times 10^{-5}$	0.612	10q23.33	<i>KIF11</i>
rs11833635	$4.09 \times 10^{-5}$	0.743	12ql5	<i>PTPRR</i>
rs6798732	$4.51 \times 10^{-5}$	0.736	3pll.2	–
rs4777585	$4.57 \times 10^{-5}$	0.496	15q24.1	<i>NEO1</i>
rs12938031	$4.61 \times 10^{-5}$	1.000	17q21.31	<i>CRHR1</i>
rs6598020	$4.63 \times 10^{-5}$	0.270	llpl5.5	<i>TMEM16J</i>
rs4927602	$4.79 \times 10^{-5}$	0.936	2p25.3	<i>SNTG2</i>

**Table 2** Trend test of association with PD results from the HIHG GWAS sample and exact Hardy-Weinberg test results for all SNPs associated with PD at  $p < 5 \times 10^{-5}$  or less.

associated SNP from the HIHG study, had a p-value of  $9.7 \times 10^{-5}$  in the joint sample. The LD in *SNCA* is weaker than that at chromosome 17q21.31, due to the large inversion in the 17q region which inhibits recombination. The 3' terminus of *PLEKHM1* is 400 kilobases from the 3' terminus of *MAPT* on chromosome 17, and rs11012 is in linkage disequilibrium (LD; mean  $r^2 = 0.75$ ; Fig. S3) with SNPs in *MAPT*. Strong association ( $p < 1 \times 10^{-5}$ ) with PD spans the entire region around *MAPT*, including SNPs in intramembrane protease 5 (*IMP5* exonic SNPs, rs12373139  $p = 2.9 \times 10^{-6}$ , OR = 0.75; 95% CI [0.67–0.85]; rs12185268  $p = 3.6 \times 10^{-6}$ , OR = 0.76; 95% CI [0.67–0.85]), *MAPT* intron (rs8070723  $p = 4.5 \times 10^{-6}$ , OR = 0.76; 95% CI [0.67–0.85]) and chromosome 17 open reading frame 69 (C17orf69 exon:Y132C, rs393152  $p = 3.5 \times 10^{-6}$ , OR = 0.76; 95% CI

[0.67–0.85]). The smallest-to-largest ranking of the p-values for association with these SNPs and PD varied greatly within the three datasets (rs11012 ranked 290 in HIHG, 11 in CIDR and 78,027 in NINDS; rs2736990 ranked 38 in HIHG, 440 in CIDR and 41,827 in NINDS) illustrating that real associations might be deep within the ranked associations in underpowered GWAS studies, and increased sample size will help true association rise to the top. Neither sex nor age-at-onset (age-at-exam in controls) were confounders for the top associations in the joint analysis. Additionally, no SNP was significant after correction for multiple tests in the dominant, recessive or genotypic analyses.

Haplotype analysis with associated chromosome 17q21.31 SNPs and the H1-H2 haplotype clade tag SNP rs1981997 revealed high  $D'$  and  $r^2$  values for all SNPs in the region, and

**Table 3.** Results for SNPs with  $p < 1 \times 10^{-6}$  from the joint analysis with imputation of the HIHG, CIDR and NINDS GWAS data.

SNP	Chromosome	Gene	Type	$P_H$	$P_C$	$P_N$	$P_3$	Empirical p
rs11012	17q21.31	<i>PLEKHM1</i>	GGG	$6.58 \times 10^{-4}$	$4.40 \times 10^{-5}$	0.188	$5.65 \times 10^{-8}$	0.011
rs2736990	4q22.1	<i>SNCA</i>	GGG	$5.08 \times 10^{-5}$	$8.67 \times 10^{-4}$	0.102	$6.74 \times 10^{-8}$	0.014
rs12063142	1p36.13	<i>TAS1R2</i>	GGG	$6.41 \times 10^{-6}$	0.032	0.033	$4.83 \times 10^{-7}$	0.159
rs4837628	9q33.1	<i>DBC1</i>	GGG	$4.83 \times 10^{-3}$	$5.66 \times 10^{-5}$	0.984	$1.03 \times 10^{-6}$	0.281
rs11248060	4p16.3	<i>DGKQ</i>	GGG	0.244	$2.05 \times 10^{-5}$	0.005	$2.26 \times 10^{-6}$	0.509
rs1981997	17q21.31	<i>MAPT</i>	GGG	$4.96 \times 10^{-3}$	$1.37 \times 10^{-4}$	0.733	$2.45 \times 10^{-6}$	0.539
rs12373139	17q21.31	<i>IMP5</i>	GGI	$4.96 \times 10^{-3}$	$1.92 \times 10^{-4}$	0.349	$2.85 \times 10^{-6}$	0.602
rs10464059	5q35.3	–	III	$1.86 \times 10^{-4}$	$4.78 \times 10^{-4}$	0.939	$2.91 \times 10^{-6}$	0.610
rs13355682	5q35.3	–	III	$1.86 \times 10^{-4}$	$4.78 \times 10^{-4}$	0.888	$3.16 \times 10^{-6}$	0.640
rs393152	17q21.31	<i>C17orf69</i>	GGG	$5.86 \times 10^{-3}$	$2.58 \times 10^{-4}$	0.281	$3.46 \times 10^{-6}$	0.674
rs12185268	17q21.31	<i>IMP5</i>	GGG	$4.96 \times 10^{-3}$	$2.52 \times 10^{-4}$	0.349	$3.62 \times 10^{-6}$	0.689
rs8070723	17q21.31	<i>MAPT</i>	GGG	$5.86 \times 10^{-3}$	$2.30 \times 10^{-4}$	0.392	$4.49 \times 10^{-6}$	0.723
rs611199	9q33.1	<i>DBC1</i>	GGG	$5.43 \times 10^{-3}$	$9.61 \times 10^{-4}$	0.212	$7.65 \times 10^{-6}$	0.908
rs17080196	5q35.3	<i>GFPT2</i>	III	$8.35 \times 10^{-4}$	$1.52 \times 10^{-3}$	0.633	$7.73 \times 10^{-6}$	0.911
rs1635291	17q21.31	–	GGG	$7.27 \times 10^{-3}$	$1.89 \times 10^{-4}$	0.517	$7.76 \times 10^{-6}$	0.912
rs7703402	5q35.3	<i>GFPT2</i>	III	$1.57 \times 10^{-4}$	$1.06 \times 10^{-3}$	0.192	$8.41 \times 10^{-6}$	0.936
rs6864729	5q35.3	<i>GFPT2</i>	III	$8.35 \times 10^{-4}$	$1.52 \times 10^{-3}$	0.696	$8.87 \times 10^{-6}$	0.945
rs2303012	5q35.3	<i>GFPT2</i>	III	$9.09 \times 10^{-4}$	$1.52 \times 10^{-3}$	0.696	$9.49 \times 10^{-6}$	0.951
rs974002	14q13.1	<i>NPAS3</i>	GGG	$6.29 \times 10^{-3}$	$7.37 \times 10^{-3}$	0.017	$9.85 \times 10^{-6}$	0.954

$P_H$  – p-values from the HIHG GWAS

$P_C$  – p-values from the CIDR GWAS

$P_N$  – p-values from the NINDS GWAS

$P_3$  – p-values from the joint analysis of the three GWAS

G – Genotyped SNP identifier for each sample in order HIHG, CIDR, NINDS

I – Imputed SNP identifier for each sample in order HIHG, CIDR, NINDS

alleles at those SNPs with positive effect sizes were in strong LD with the G allele of rs1981997, indicating their presence in the H1 haplotype clade (Table S3, Fig. S3).

Additional strong association signals in the joint data were observed in SNPs in the chromosome 9q33 gene deleted in bladder cancer 1 (*DBC1*, EGI:1620, intron, rs4837628  $p = 1.07 \times 10^{-6}$ , OR = 0.79; 95% CI [0.72–0.87]), the chromosome 14q13 gene neuronal PAS domain protein 3 (*NPAS3*, EGI:64067, intron, rs974002  $p = 9.9 \times 10^{-6}$ , OR = 1.32; 95% CI [1.17–1.49]), and in imputed SNPs in the chromosome 5q35 gene glucose fructose-6-phosphate transaminase 2 (*GFPT2*, EGI:9945, intronic SNPs, rs17080196  $p = 7.7 \times 10^{-6}$ , OR = 0.77; 95% CI 0.69–0.87; rs7703402 (promoter)  $p = 8.4 \times 10^{-6}$ , OR = 0.76; 95% CI [0.66–0.85]; rs6864729  $p = 8.9 \times 10^{-6}$ , OR = 0.77; 95% CI [0.69–0.87]; rs2303012  $9.5 \times 10^{-6}$ , OR = 0.77; 95% CI [0.69–0.87]; Table 3, Table S1). However, none of these associations survive multiple testing adjustments.

Weaker associations in the joint data that replicated at an uncorrected  $p \leq 0.05$  with consistent effect directions in each of the three GWAS were observed at several loci (Table 4, Table S2). These associations were observed at rs12063142 in the chromosome 1p36.13 gene taste receptor 1, member 2 (*TAS1R2*, intron, rs12063142

$p = 4.83 \times 10^{-7}$ , OR = 0.76; 95% CI [0.69–0.85]), rs974002 in *NPAS3*, the chromosome 8q21 gene matrix metalloproteinase 16 (*MMP16*, EGI:4325, intron, rs3851539  $p = 3.8 \times 10^{-5}$ , OR = 1.21; 95% CI [1.11–1.34]), the chromosome 12q24 gene kinase suppressor of ras 2 (*KSR2*, EGI:283455, intron, rs7960736  $p = 6.3 \times 10^{-5}$ , OR = 1.22; 95% CI [1.11–1.34]), the chromosome 14q24 SNP rs11159221 ( $p = 1 \times 10^{-4}$ , OR = 0.79; 95% CI [0.69–0.89]), the chromosome 2q31 gene WAS/WASL-interacting protein family, member1 (*WASPIP*, EGI:7456, intron, rs1991601  $p = 3 \times 10^{-4}$ , OR = 1.23; 95% CI [1.09–1.38]), the chromosome 11q32 gene family with sequence similarity 55, family A (*FAM55A*, EGI:120400, intron, rs1080074  $p = 4 \times 10^{-4}$ , OR = 1.29; 95% CI [1.12–1.48]), and the chromosome 15q22 gene RAR-related orphan receptor A (*RORA*, EGI:6095, intron, rs1863270  $p = 5 \times 10^{-4}$ , OR = 1.19; 95% CI [1.08–1.33]).

Results from each GWAS and the joint analysis for SNPs in the top 10 genes from the PDgene database (www.pdgene.org) are summarised in Table 5. The PDgene database is a free online meta-analysis summary of PD genetic association studies. No associations with  $p < 2.00 \times 10^{-3}$  were observed in genes other than *SNCA* and *MAPT*. The strongest association among all SNPs in these genes

**Table 4.** Association results that replicate in the HIHG, CIDR and NINDS studies at the  $p < 0.05$  level with effects in the same direction. All SNPs were genotyped.

SNP	Chromosome	Gene	$P_H$	$P_C$	$P_N$	$P_3$
rs12063142	1p36.13	<i>TAS1R2</i>	$6.41 \times 10^{-6}$	0.032	0.033	$4.83 \times 10^{-7}$
rs974002	14q13.1	<i>NPAS3</i>	$6.32 \times 10^{-3}$	$7.39 \times 10^{-3}$	0.017	$9.85 \times 10^{-6}$
rs3851539	8q21.3	<i>MMP16</i>	$7.86 \times 10^{-3}$	0.015	0.034	$3.81 \times 10^{-5}$
rs7960736	12q24.23	<i>KSR2</i>	0.026	0.022	$6.94 \times 10^{-3}$	$6.26 \times 10^{-5}$
rs11159221	14q24.3	–	0.024	0.045	$5.21 \times 10^{-3}$	$1.32 \times 10^{-4}$
rs1991601	2q31.1	<i>WIPF1</i>	0.019	0.043	0.043	$3.26 \times 10^{-4}$
rs1080074	11q23.2	<i>FAM55A</i>	0.038	0.036	0.035	$3.77 \times 10^{-4}$
rs1863270	15q22.2	<i>RORA</i>	0.047	0.044	0.032	$5.33 \times 10^{-4}$

$P_H$  – p-values from the HIHG GWAS

$P_C$  – p-values from the CIDR GWAS

$P_N$  – p-values from the NINDS GWAS

$P_3$  – p-values from the joint analysis of the three GWAS

**Table 5** Summary of results for the top 10 genes in order from PDgene (June 2009) for each GWAS dataset for the SNP in each gene with the smallest joint analysis p-value.

Gene	SNP	$P_{MH}$	$P_C$	$P_N$	$P_3$	#SNPs in gene	Gene Size (bp)
<i>GBA</i>	rs9628662	0.694	0.165	0.581	0.153	1	10,245
<i>LRRK2</i>	rs1907632	0.423	0.255	0.023	0.032	42	144,273
<i>SNCA</i>	rs2736990	$5.08 \times 10^{-5}$	$8.67 \times 10^{-4}$	0.102	$6.70 \times 10^{-8}$	15	112,876
<i>PINK1</i>	rs2296223	0.546	0.031	0.247	0.212	2	18,056
<i>MAPT</i>	rs1981997	$4.96 \times 10^{-3}$	$1.37 \times 10^{-4}$	0.733	$2.50 \times 10^{-6}$	15	134,002
<i>USP24</i>	rs12065953	$8.96 \times 10^{-3}$	0.195	0.129	$7.81 \times 10^{-3}$	2	149,006
<i>CYP2D6</i>	rs769258	0.572	0.848	0.033	0.574	1	4,382
<i>MAOB</i>	rs209766	0.031	0.172	0.056	0.003	31	115,864
<i>ELAVL4</i>	rs17105974	$3.39 \times 10^{-3}$	0.059	0.142	0.042	10	153,854
<i>APOE</i>	rs439401	0.568	0.411	0.829	0.41	2	3,611

$P_{MH}$  – p-values from the HIHG GWAS

$P_C$  – p-values from the CIDR GWAS

$P_N$  – p-values from the NINDS GWAS

$P_3$  – p-values from the joint analysis of the three GWAS

was in monoamine oxidase B (*MAOB* intron, rs209766  $p = 0.0026$ , OR = 0.86; 95% CI [0.76–0.96]). Weaker associations were also detected in leucine-rich repeat kinase 2 (*LRRK2*, EGI:120892, intron, rs1907632  $p = 0.0321$ , OR = 1.15; 95% CI [1.01–1.30]), *USP24* (intron, rs12065953  $p = 0.0078$ , OR = 1.46; 95% CI [1.10–1.94]), and *ELAVL4* (intron, rs17105974  $p = 0.0418$ , OR = 0.82; 95% CI [0.65–0.99]).

Results from the meta-analysis of the results from the three GWAS were consistent with those observed in the joint analysis (Data not shown. Pearson's correlation coefficient for p-values from the two analyses = 0.95). The SNPs rs11012 (meta-analysis p-value =  $7.33 \times 10^{-8}$ ) in the chromosome 17q21.31 region and rs2736990 (meta-analysis p-value =  $7.88 \times 10^{-8}$ ) in *SNCA* were the most strongly associated SNPs in the meta-analysis.

## Discussion

This study unambiguously identifies *SNCA* and *MAPT* as risk factors for idiopathic PD. The power to declare association at genome-wide levels of significance in any of these GWAS is low for the effects observed in this study; however, in aggregate these datasets provide conclusive evidence for major risk genes and highlight other genes of interest. Our study illustrates the utility of freely available online data resources and large collaborative studies that are better powered to detect associations with modest effects in GWAS data. It is only with joint analysis of three independent GWAS datasets that SNPs in the *SNCA* and the *MAPT* region reached genome-wide significance. The exact mechanisms by which variation in *SNCA* and *MAPT* influence risk for PD are unknown (Devine & Lewis, 2008). Joint analysis with trend tests of allelic association also suggests other candidates that show

strong, but not genome-wide evidence of association: *DBC1*, *NPAS3*, and *GFPT2*.

The *SNCA* gene product  $\alpha$ -synuclein (AS) is an abundant brain protein which is localised in axon terminals where it may mediate synaptic processes (Cabin et al., 2002). PD is considered a synucleinopathy in which aggregations of precipitated filamentous AS (which form Lewy bodies) are a common finding in PD brains at autopsy (Lewy, 1912; Tretakoff, 1919; Okazaki et al., 1961; Kosaka et al., 1976; Spillantini et al., 1997). Mutations in *SNCA* have been shown to cosegregate with PD with an autosomal dominant mode of inheritance (Polymeropoulos et al., 1997; Kruger et al., 1998; Singleton et al., 2003; Chartier-Harlin et al., 2004; Ibanez et al., 2004; Zarranz et al., 2004). Other variation in *SNCA* has been reported to increase the risk of PD (Kruger et al., 1999).

The *MAPT* protein (tau) is critical for the assembly and stabilisation of the microtubule network, which is essential for axonal transport in neurons (Garcia & Cleveland, 2001). Neurological disorders where aggregations of tau are observed in brain tissue are known as tauopathies. The most common tauopathy is AD, where hyperphosphorylated tau accumulates in intraneuronal neurofibrillary tangles (NFTs) (Goedert et al., 1989). Additionally, progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) are neurodegenerative diseases characterised by tau deposition in neurons and glia (Litvan et al., 1996; Dickson et al., 2002) and parkinsonism. A further set of disorders referred to as frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17) is also characterised by the abnormal deposition of tau (Ingram & Spillantini, 2002). Haplotypes in the chromosome 17q21.31 region have been associated with progressive supranuclear palsy (PSP), corticobasal degeneration (CBD) (Pittman et al., 2005), and PD (Golbe et al., 2001; Maraganore et al., 2001; Martin et al., 2001; Farrer et al., 2002; Healy et al., 2004; Skipper et al., 2004; Zabetian et al., 2007; Refenes et al., 2009). In the current study we report haplotypic associations of alleles on the H1 haplotype clade which is consistent with previous investigations of PD, PSP, and CBD. The H1 clade has also been shown to have higher expression levels of tau than H2 (Kwok et al., 2004), which may provide some insight into PD pathology as previous studies have shown that tau mediates and promotes the polymerisation of AS (Giasson et al., 2003).

The gene Diacylglycerol kinase, theta (*DGKQ*) was previously observed to be associated with PD in the Pankratz et al. study (Pankratz et al., 2008). This gene is thought to be active in the phosphatidylinositol signaling pathway and is expressed in the brain.

Interestingly, associations between SNPs in *NPAS3* and smoking cessation have been observed in two of three GWAS datasets investigating genetic differences between the ability

of individuals to quit smoking (Uhl et al., 2008). Given the known relationship between smoking and PD (Dorn, 1959; Allam et al., 2004), this gene may merit further investigation in PD studies. We saw no evidence of confounding by or interaction with history of smoking at *NPAS3* SNPs (data not shown).

The HIHG data and the CIDR data were analysed for interactions with and confounding by history of smoking. Although no SNPs were significant after correction for multiple tests (data not shown), these and other exposures may explain or modify genetic susceptibilities to PD. Previous research in PD families has observed interactions with smoking and the genes nitric oxide synthase 2A (*NOS2A*) (Hancock et al., 2006), glutathione S-transferase omega 1 (*GSTO1*) (Wahner et al., 2007) and *SNCA* (McCulloch et al., 2008). This study only included two *NOS2A* SNPs, neither of which were included in the previous study, and which do not provide sufficient coverage of the gene to evaluate the previous findings of interaction. Similarly, the report of interaction with smoking with the Rep1 variant in *SNCA* was not replicable with the SNPs in this study. The interaction with history of smoking and *GSTO1* at rs4925 was not replicated.

*DBC1* is a gene that has been observed to be deleted in some bladder cancer cell lines (Habuchi et al., 1998). The *DBC1* gene product is detectable in several tissues, including brain, spinal cord, cortex and cerebellum (<http://www.genecards.org/>). The *DBC1* protein inhibits cell proliferation by negative regulation of the G1/S transition (Nishiyama et al., 2001). Additionally, this protein mediates non-apoptotic cell death (Wright et al., 2004), and is a regulator of components of the plasminogen pathway (Louhelainen et al., 2006). The relationship between PD and *DBC1* is not obvious based on the known biology of this gene; however, the cell-death phenotypes for this gene in sensitive brain regions might play a role in PD pathophysiology.

It should also be noted that the *GFPT2* results were obtained by imputation in the three GWAS samples used here. These genotypes are likely to be assigned with higher error rates than assayed SNPs. *GFPT2* is the rate-limiting enzyme for the entry of glucose into the hexosamine biosynthesis pathway and is thus a biologically important gene (Oki et al., 1999). Energy production in the central nervous system is a mixture of glycolysis and electron transport in the mitochondrion. As mitochondria are clearly affected in PD (Beal, 2007; Schapira, 2008), enzymes such as *GFPT2* may become increasingly important in energy production and thereby cell survival. Additionally, variants in this gene have been associated with type II diabetes mellitus (T2DM) and diabetic nephropathy in Caucasian and African American samples (Zhang et al., 2004); T2DM has been suggested as a risk factor for PD (Hu et al., 2007; Driver et al., 2008).



Despite the intriguing results of our study, there are some limitations. The error rate of imputed genotypes is likely to be higher than that of assayed genotypes, although this effect should be mild for imputed genotypes with high posterior probabilities (Marchini et al., 2007). It is possible that PD is composed of several sub-phenotypes, each with a distinct genetic etiology. This analysis would have optimal power to detect only associations that underlie the common features of the trait definitions for all three GWAS. The power of the joint analysis is also not optimal for the magnitude of effects seen in this study ( $\geq 80\%$  power for the combined sample at  $\alpha = 1 \times 10^{-7}$  and MAF 0.25 for odds ratios  $<0.59$  or  $>1.59$ ;  $\geq 80\%$  power for the HIHG sample at  $\alpha = 1 \times 10^{-7}$  and MAF 0.25 for odds ratios  $<0.39$  or  $>2.14$ ); however, with the addition of larger PD studies more genes might be conclusively associated. These results support the meta-analysis of Pankratz et al. of the CIDR GWAS results and the NINDS GWAS results to the extent that the *MAPT* region is prominently represented in the top associations (Pankratz et al., 2009); however, the overall similarity among top hits is sparse. A final limitation is the lack of detailed covariate data across datasets for exposures such as smoking, pesticide exposure and coffee drinking for model adjustment and gene  $\times$  environment interaction analysis. This is an important point, as the value of these datasets would increase greatly with the inclusion of well-documented environmental and sub-phenotype data, such as diabetes status, depression and dementia. Future studies of PD should include important environmental factors in the data collection protocol.

In conclusion, *SNCA* and *MAPT* are significantly associated at a genome-wide level with idiopathic PD. Several other biologically plausible genes are associated with PD but do not meet the genome-wide significance threshold. PD is a complex phenotype with substantial variation in age of onset and clinical course. It is possible that more genes are associated with subsets of PD cases or are associated with more modest effects than those detected here. These results suggest that observing additional PD-gene associations will require the pooling of datasets to achieve very large samples of thousands of cases and controls from which to obtain statistically significant p-values for loci of more modest effects after correction for multiple tests. Collaborative projects with coordinated ascertainment are necessary to advance PD genetic epidemiology.

## Conflict of Interest

None.

## Acknowledgements

We are grateful to the patients and control subjects who participated in this study. We thank the members of the PD Genetics

Collaboration: Martha A. Nance, Ray L. Watts, Jean P. Hubble, William C. Koller, Kelly Lyons, Rajesh Pahwa, Matthew B. Stern, Amy Colcher, Bradley C. Hiner, Joseph Jankovic, William G. Ondo, Fred H. Allen, Jr., Christopher G. Goetz, Gary W. Small, Donna Masterman, Frank Mastaglia, and Jonathan L. Haines who contributed subjects to this study. Some of the samples used in this study were collected while the Udall PDRCE was at Duke University. This work was supported by National Institutes of Health grant NS39764. We also thank the investigators from the Pankratz et al. (2009) (accession number: phs000126.v1.p1), and the Fung et al. (2006a) (accession number: phs000089.v1.p1) studies for making their data available on dbGAP.

## References

- Abecasis, G. & Willer, C. (2007) METAL, Meta Analysis Helper. Ref Type: Unpublished Work.
- Allam, M. F., Campbell, M. J., Hofman, A., Del Castillo, A. S. & Fernandez-Crehuet, N. R. (2004) Smoking and Parkinson's disease: systematic review of prospective studies. *Mov Disord* **19**(6), 614–621.
- Armitage, P. (1955) Tests for Linear Trends in Proportions and Frequencies. *Biometrics* **11**, 375–386.
- Autere, J., Moilanen, J. S., Finnila, S., Soininen, H., Mannermaa, A., Hartikainen, P., Hallikainen, M. & Majamaa, K. (2004) Mitochondrial DNA polymorphisms as risk factors for Parkinson's disease and Parkinson's disease dementia. *Hum Genet* **115**, 29–35.
- Beal, M. F. (2007) Mitochondria and neurodegeneration. *Novartis Found Symp* **287**, 183–192.
- Beecham, G. W., Martin, E. R., Li, Y. J., Slifer, M. A., Gilbert, J. R., Haines, J. L. & Pericak-Vance, M. A. (2009) Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease. *Am J Hum Genet* **84**, 35–43.
- Bonifati, V., Rizzo, P., van Baren, M. J., Schaap, O., Breedveld, G. J., Krieger, E., Dekker, M. C., Squitieri, F., Ibanez, P., Joosse, M., van Dongen, J. W., Vanacore, N., van Swieten, J. C., Brice, A., Meco, G., van Duijn, C. M., Oostra, B. A. & Heutink, P. (2003) Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science* **299**, 256–259.
- Cabin, D. E., Shimazu, K., Murphy, D., Cole, N. B., Gottschalk, W., McIlwain, K. L., Orrison, B., Chen, A., Ellis, C. E., Paylor, R., Lu, B. & Nussbaum, R. L. (2002) Synaptic vesicle depletion correlates with attenuated synaptic responses to prolonged repetitive stimulation in mice lacking alpha-synuclein. *J Neurosci* **22**, 8797–8807.
- Chartier-Harlin, M. C., Kachergus, J., Roumier, C., Mouroux, V., Douay, X., Lincoln, S., Levecque, C., Larvor, L., Andrieux, J., Hulihan, M., Waucquier, N., Defebvre, L., Amouyel, P., Farrer, M. & Destee, A. (2004) Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167–1169.
- DeStefano, A. L., Latourelle, J., Lew, M. F., Suchowersky, O., Klein, C., Golbe, L. I., Mark, M. H., Growdon, J. H., Wooten, G. F., Watts, R., Guttman, M., Racette, B. A., Perlmutter, J. S., Marlor, L., Shill, H. A., Singer, C., Goldwurm, S., Pezzoli, G., Saint-Hilaire, M. H., Hendricks, A. E., Gower, A., Williamson, S., Nagle, M. W., Wilk, J. B., Massood, T., Huskey, K. W., Baker, K. B., Itin, I., Litvan, I., Nicholson, G., Corbett, A., Nance, M., Drasby, E., Isaacson, S., Burn, D. J., Chinnery, P. F., Pramstaller, P. P., Al-Hinti, J., Moller, A. T., Ostergaard, K., Sherman, S. J.,

- Roxburgh, R., Snow, B., Slevin, J. T., Cambi, F., Gusella, J. F. & Myers, R. H. (2008) Replication of association between ELAVL4 and Parkinson disease: the GenePD study. *Hum Genet* **124**, 95–99.
- Devine, M. J. & Lewis, P. A. (2008) Emerging pathways in genetic Parkinson's disease: tangles, Lewy bodies and LRRK2. *FEBS J* **275**, 5748–5757.
- Dickson, D. W., Bergeron, C., Chin, S. S., Duyckaerts, C., Horoupian, D., Ikeda, K., Jellinger, K., Lantos, P. L., Lippa, C. F., Mirra, S. S., Tabaton, M., Vonsattel, J. P., Wakabayashi, K. & Litvan, I. (2002) Office of Rare Diseases neuropathologic criteria for corticobasal degeneration. *J Neuropathol Exp Neurol* **61**, 935–946.
- Dorn, H. F. (1959) Tobacco consumption and mortality from cancer and other diseases. *Public Health Rep* **74** 581–593.
- Driver, J. A., Smith, A., Buring, J. E., Gaziano, J. M., Kurth, T. & Logroscino, G. (2008) Prospective cohort study of type 2 diabetes and the risk of Parkinson's disease. *Diabetes Care* **31**, 2003–2005.
- Falush, D., Stephens, M. & Pritchard, J. K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587.
- Farrer, M., Maraganore, D. M., Lockhart, P., Singleton, A., Lesnick, T. G., de Andrade, M., West, A., de Silva, R., Hardy, J. & Hernandez, D. (2001) alpha-Synuclein gene haplotypes are associated with Parkinson's disease. *Hum Mol Genet* **10**, 1847–1851.
- Farrer, M., Skipper, L., Berg, M., Bisceglia, G., Hanson, M., Hardy, J., Adam, A., Gwinn-Hardy, K. & Aasly, J. (2002) The tau H1 haplotype is associated with Parkinson's disease in the Norwegian population. *Neurosci Lett* **322**, 83–86.
- Fidani, L., Kalinderi, K., Bostantjopoulou, S., Clarimon, J., Goulas, A., Katsarou, Z., Hardy, J. & Kotsis, A. (2006) Association of the Tau haplotype with Parkinson's disease in the Greek population. *Mov Disord* **21**, 1036–1039.
- Folstein, M. F., Folstein, S. E. & McHugh, P. R. (1975) Mini-Mental State—Practical Method for Grading Cognitive State of Patients for Clinician. *Journal of Psychiatric Research* **12**, 189–198.
- Fung, H. C., Scholz, S., Matarin, M., Simon-Sanchez, J., Hernandez, D., Britton, A., Gibbs, J. R., Langefeld, C., Stiebert, M. L., Schymick, J., Okun, M. S., Mandel, R. J., Fernandez, H. H., Foote, K. D., Rodriguez, R. L., Peckham, E., de Vrieze, F. W., Gwinn-Hardy, K., Hardy, J. A. & Singleton, A. (2006a) Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol* **5**, 911–916.
- Fung, H. C., Xiomerisiou, G., Gibbs, J. R., Wu, Y. R., Eerola, J., Gourbali, V., Hellstrom, O., Chen, C. M., Duckworth, J., Papadimitriou, A., Tienari, P. J., Hadjigeorgiou, G. M., Hardy, J. & Singleton, A. B. (2006b) Association of tau haplotype-tagging polymorphisms with Parkinson's disease in diverse ethnic Parkinson's disease cohorts. *Neurodegener Dis* **3**, 327–333.
- Garcia, M. L. & Cleveland, D. W. (2001) Going new places using an old MAP: tau, microtubules and human neurodegenerative disease. *Curr Opin Cell Biol* **13**, 41–48.
- Gaweda-Walerych, K., Maruszak, A., Safranow, K., Bialecka, M., Klodowska-Duda, G., Czyzewski, K., Slawek, J., Rudzinska, M., Styczynska, M., Opala, G., Drozdziak, M., Canter, J. A., Barcikowska, M. & Zekanowski, C. (2008) Mitochondrial DNA haplogroups and subhaplogroups are associated with Parkinson's disease risk in a Polish PD cohort. *J Neural Transm* **115**, 1521–1526.
- Ghezzi, D., Marelli, C., Achilli, A., Goldwurm, S., Pezzoli, G., Barone, P., Pellecchia, M. T., Stanzione, P., Brusa, L., Bentivoglio, A. R., Bonuccelli, U., Petrozzi, L., Abbruzzese, G., Marchese, R., Cortelli, P., Grimaldi, D., Martinelli, P., Ferrarese, C., Garavaglia, B., Sangiorgi, S., Carelli, V., Torroni, A., Albanese, A. & Zeviani, M. (2005) Mitochondrial DNA haplogroup K is associated with a lower risk of Parkinson's disease in Italians. *Eur J Hum Genet* **13**, 748–752.
- Giasson, B. I., Forman, M. S., Higuchi, M., Golbe, L. I., Graves, C. L., Kotzbauer, P. T., Trojanowski, J. Q. & Lee, V. M. (2003) Initiation and synergistic fibrillization of tau and alpha-synuclein. *Science* **300**, 636–640.
- Goedert, M., Spillantini, M. G., Potier, M. C., Ulrich, J. & Crowther, R. A. (1989) Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain. *EMBO J* **8**, 393–399.
- Golbe, L. I., Lazzarini, A. M., Szychala, J. R., Johnson, W. G., Stenroos, E. S., Mark, M. H. & Sage, J. I. (2001) The tau A0 allele in Parkinson's disease. *Mov Disord* **16**, 442–447.
- Goris, A., Williams-Gray, C. H., Clark, G. R., Foltynie, T., Lewis, S. J., Brown, J., Ban, M., Spillantini, M. G., Compston, A., Burn, D. J., Chinnery, P. F., Barker, R. A. & Sawcer, S. J. (2007) Tau and alpha-synuclein in susceptibility to, and dementia in, Parkinson's disease. *Ann Neurol* **62**, 145–153.
- Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G. & Chee, M. S. (2005) A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* **37**, 549–554.
- Habuchi, T., Luscombe, M., Elder, P. A. & Knowles, M. A. (1998) Structure and methylation-based silencing of a gene (DBCCR1) within a candidate bladder cancer tumor suppressor region at 9q32–q33. *Genomics* **48**, 277–288.
- Hancock, D. B., Martin, E. R., Fujiwara, K., Stacy, M. A., Scott, B. L., Stajich, J. M., Jewett, R., Li, Y. J., Hauser, M. A., Vance, J. M. & Scott, W. K. (2006) NOS2A and the modulating effect of cigarette smoking in Parkinson's disease. *Ann Neurol* **60**, 366–373.
- Haugarvoll, K., Toft, M., Ross, O. A., Stone, J. T., Heckman, M. G., White, L. R., Lynch, T., Gibson, J. M., Wszolek, Z. K., Uitti, R. J., Aasly, J. O. & Farrer, M. J. (2007) ELAVL4, PARK10, and the Celts. *Mov Disord* **22**, 585–587.
- Haugarvoll, K., Toft, M., Skipper, L., Heckman, M. G., Crook, J. E., Soto, A., Ross, O. A., Hulihan, M. M., Kachergus, J. M., Sando, S. B., White, L. R., Lynch, T., Gibson, J. M., Uitti, R. J., Wszolek, Z. K., Aasly, J. O. & Farrer, M. J. (2009) Fine-mapping and candidate gene investigation within the PARK10 locus. *Eur J Hum Genet* **17**, 336–343.
- Healy, D. G., Abou-Sleiman, P. M., Lees, A. J., Casas, J. P., Quinn, N., Bhatia, K., Hingorani, A. D. & Wood, N. W. (2004) Tau gene and Parkinson's disease: a case-control study and meta-analysis. *J Neurol Neurosurg Psychiatry* **75**, 962–965.
- Hu, G., Jousilahti, P., Bidel, S., Antikainen, R. & Tuomilehto, J. (2007) Type 2 diabetes and the risk of Parkinson's disease. *Diabetes Care* **30**, 842–847.
- Huerta, C., Castro, M. G., Coto, E., Blazquez, M., Ribacoba, R., Guisasola, L. M., Salvador, C., Martinez, C., Lahoz, C. H. & Alvarez, V. (2005) Mitochondrial DNA polymorphisms and risk of Parkinson's disease in Spanish population. *J Neurol Sci* **236**, 49–54.
- Ibanez, P., Bonnet, A. M., Debarges, B., Lohmann, E., Tison, F., Pollak, P., Agid, Y., Durr, A. & Brice, A. (2004) Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet* **364**, 1169–1171.
- Ingram, E. M. & Spillantini, M. G. (2002) Tau gene mutations: dissecting the pathogenesis of FTDP-17. *Trends Mol Med* **8**, 555–562.

- Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R. & Schimmel, H. (1983) Validation of A Short Orientation-Memory-Concentration Test of Cognitive Impairment. *American Journal of Psychiatry* **140**, 734–739.
- Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y. & Shimizu, N. (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605–608.
- Kosaka, K., Oyanagi, S., Matsushita, M. & Hori A. (1976) Presenile dementia with Alzheimer-, Pick- and Lewy-body changes. *Acta Neuropathol* **36**, 221–233.
- Kosel, S., Grasbon-Frodl, E. M., Mautsch, U., Egensperger, R., von Eitzen, U., Frishman, D., Hofmann, S., Gerbitz, K. D., Mehraein, P. & Graeber, M. B. (1998) Novel mutations of mitochondrial complex I in pathologically proven Parkinson disease. *Neurogenetics* **1**, 197–204.
- Kruger, R., Kuhn, W., Muller, T., Woitalla, D., Graeber, M., Kosel, S., Przuntek, H., Epplen, J. T., Schols, L. & Riess, O. (1998) Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nat Genet* **18**, 106–108.
- Kruger, R., Vieira-Saecker, A. M., Kuhn, W., Berg, D., Muller, T., Kuhn, N., Fuchs, G. A., Storch, A., Hungs, M., Woitalla, D., Przuntek, H., Epplen, J. T., Schols, L. & Riess, O. (1999) Increased susceptibility to sporadic Parkinson's disease by a certain combined alpha-synuclein/apolipoprotein E genotype. *Ann Neurol* **45**, 611–617.
- Kuopio, A. M., Marttila, R. J., Helenius, H. & Rinne, U. K. (1999) Changing epidemiology of Parkinson's disease in southwestern Finland. *Neurology* **52**, 302–308.
- Kurth, J. H., Kurth, M. C., Poduslo, S. E. & Schwankhaus, J. D. (1993) Association of a monoamine oxidase B allele with Parkinson's disease. *Ann Neurol* **33**, 368–372.
- Kwok, J. B., Teber, E. T., Loy, C., Hallupp, M., Nicholson, G., Mellick, G. D., Buchanan, D. D., Silburn, P. A. & Schofield, P. R. (2004) Tau haplotypes regulate transcription and are associated with Parkinson's disease. *Ann Neurol* **55**, 329–334.
- Lesage, S. & Brice, A. (2009) Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum Mol Genet* **18**(R1):R48–R59.
- Levecque, C., Elbaz, A., Clavel, J., Vidal, J. S., Amouyel, P., Alperovitch, A., Tzourio, C. & Chartier-Harlin, M. C. (2004) Association of polymorphisms in the Tau and Saitohin genes with Parkinson's disease. *J Neurol Neurosurg Psychiatry* **75**, 478–480.
- Lewy, F. H. (1912) *Handbuch der Neurologie Band III*. Berlin: Springer.
- Li, Y., Schrodi, S., Rowland, C., Tacey, K., Catanese, J. & Grupe, A. (2006) Genetic evidence for ubiquitin-specific proteases USP24 and USP40 as candidate genes for late-onset Parkinson disease. *Hum Mutat* **27**, 1017–1023.
- Litvan, I., Hauw, J. J., Bartko, J. J., Lantos, P. L., Daniel, S. E., Horoupian, D. S., McKee, A., Dickson, D., Bancher, C., Tabaton, M., Jellinger, K. & Anderson, D. W. (1996) Validity and reliability of the preliminary NINDS neuropathologic criteria for progressive supranuclear palsy and related disorders. *J Neuropathol Exp Neurol* **55**, 97–105.
- Louhelainen, J. P., Hurst, C. D., Pitt, E., Nishiyama, H., Pickett, H. A. & Knowles, M. A. (2006) DBC1 re-expression alters the expression of multiple components of the plasminogen pathway. *Oncogene* **25**, 2409–2419.
- Mamah, C. E., Lesnick, T. G., Lincoln, S. J., Strain, K. J., de Andrade, M., Bower, J. H., Ahlskog, J. E., Rocca, W. A., Farrer, M. J. & Maraganore, D. M. (2005) Interaction of alpha-synuclein and tau genotypes in Parkinson's disease. *Ann Neurol* **57**, 439–443.
- Maraganore, D. M., de Andrade, M., Elbaz, A., Farrer, M. J., Ioannidis, J. P., Kruger, R., Rocca, W. A., Schneider, N. K., Lesnick, T. G., Lincoln, S. J., Hulihan, M. M., Aasly, J. O., Ashizawa, T., Chartier-Harlin, M. C., Checkoway, H., Ferrarese, C., Hadjigeorgiou, G., Hattori, N., Kawakami, H., Lambert, J. C., Lynch, T., Mellick, G. D., Papapetropoulos, S., Parsian, A., Quattrone, A., Riess, O., Tan, E. K. & Van Broeckhoven, C. (2006) Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease. *JAMA* **296**, 661–670.
- Maraganore, D. M., de Andrade, M., Lesnick, T. G., Strain, K. J., Farrer, M. J., Rocca, W. A., Pant, P. V., Frazer, K. A., Cox, D. R. & Ballinger, D. G. (2005) High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet* **77**, 685–693.
- Maraganore, D. M., Hernandez, D. G., Singleton, A. B., Farrer, M. J., McDonnell, S. K., Hutton, M. L., Hardy, J. A. & Rocca, W. A. (2001) Case-Control study of the extended tau gene haplotype in Parkinson's disease. *Ann Neurol* **50**, 658–661.
- Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906–913.
- Martin, E. R., Scott, W. K., Nance, M. A., Watts, R. L., Hubble, J. P., Koller, W. C., Lyons, K., Pahwa, R., Stern, M. B., Colcher, A., Hiner, B. C., Jankovic, J., Ondo, W. G., Allen, F. H. Jr., Goetz, C. G., Small, G. W., Masterman, D., Mastaglia, F., Laing, N. G., Stajich, J. M., Ribble, R. C., Booze, M. W., Rogala, A., Hauser, M. A., Zhang, F., Gibson, R. A., Middleton, L. T., Roses, A. D., Haines, J. L., Scott, B. L., Pericak-Vance, M. A. & Vance, J. M. (2001) Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease. *JAMA* **286**, 2245–2250.
- 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., Zabetian, C. P. & Payami, H. (2008) Exploring gene-environment interactions in Parkinson's disease. *Hum Genet* **123**, 257–265.
- Mueller, J. C., Fuchs, J., Hofer, A., Zimprich, A., Lichtner, P., Illig, T., Berg, D., Wullner, U., Meitinger, T. & Gasser, T. (2005) Multiple regions of alpha-synuclein are associated with Parkinson's disease. *Ann Neurol* **57**, 535–541.
- Myhre, R., Toft, M., Kachergus, J., Hulihan, M. M., Aasly, J. O., Klungland, H. & Farrer, M. J. (2008) Multiple alpha-synuclein gene polymorphisms are associated with Parkinson's disease in a Norwegian population. *Acta Neurol Scand* **118**, 320–327.
- Nishiyama, H., Gill, J. H., Pitt, E., Kennedy, W. & Knowles, M. A. (2001) Negative regulation of G(1)/S transition by the candidate bladder tumour suppressor gene DBCCR1. *Oncogene* **20**, 2956–2964.
- Noureddine, M. A., Qin, X. J., Oliveira, S. A., Skelly, T. J., Van Der Walt, J., Hauser, M. A., Pericak-Vance, M. A., Vance, J. M. & Li, Y. J. (2005) Association between the neuron-specific RNA-binding protein ELAVL4 and Parkinson disease. *Hum Genet* **117**, 27–33.
- Okazaki, H., Lipkin, L. E. & Aronson, S. M. (1961) Diffuse intracytoplasmic ganglionic inclusions (Lewy type) associated with progressive dementia and quadriplegia in flexion. *J Neuropathol Exp Neurol* **20**, 237–244.
- Oki, T., Yamazaki, K., Kuromitsu, J., Okada, M. & Tanaka, I. (1999) cDNA cloning and mapping of a novel subtype of glutamine:fructose-6-phosphate amidotransferase (GFAT2) in human and mouse. *Genomics* **57**, 227–234.

- Oliveira, S. A., Li, Y. J., Noureddine, M. A., Zuchner, S., Qin, X., Pericak-Vance, M. A. & Vance, J. M. (2005) Identification of risk and age-at-onset genes on chromosome 1p in Parkinson disease. *Am J Hum Genet* **77**, 252–264.
- Paisan-Ruiz, C., Jain, S., Evans, E. W., Gilks, W. P., Simon, J., Van Der Brug, M., Lopez de Munain, A., Aparicio, S., Gil A. M., Khan, N., Johnson, J., Martinez, J. R., Nicholl, D., Carrera, I. M., Pena, A. S., de Silva, R., Lees, A., Marti-Masso, J. F., Perez-Tur, J., Wood, N. W. & Singleton, A. B. (2004) Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* **44**, 595–600.
- Pankratz, N., Wilk, J. B., Latourelle, J. C., DeStefano, A. L., Halter, C., Pugh, E. W., Doheny, K. F., Gusella, J. F., Nichols, W. C., Foroud, T. & Myers, R. H. (2009) Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum Genet* **124** 593–605.
- Pittman, A. M., Myers, A. J., Abou-Sleiman, P., Fung, H. C., Kaleem, M., Marlowe, L., Duckworth, J., Leung, D., Williams, D., Kilford, L., Thomas, N., Morris, C. M., Dickson, D., Wood, N. W., Hardy, J., Lees, A. J. & de Silva, R. (2005) Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. *J Med Genet* **42**, 837–846.
- Polymeropoulos, M. H., Lavedan, C., Leroy, E., Ide, S. E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., Stenroos, E. S., Chandrasekharappa, S., Athanassiadou, A., Papapetropoulos, T., Johnson, W. G., Lazzarini, A. M., Duvoisin, R. C., Di, I. G., Golbe, L. I. & Nussbaum, R. L. (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045–2047.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909.
- Pritchard, J. K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J. & Sham, P. C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575.
- Pyle, A., Foltynie, T., Tiangyou, W., Lambert, C., Keers, S. M., Allcock, L. M., Davison, J., Lewis, S. J., Perry, R. H., Barker, R., Burn, D. J. & Chinnery, P. F. (2005) Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. *Ann Neurol* **57**, 564–567.
- Refenes, N., Bolbrinker, J., Tagaris, G., Orlacchio, A., Drakoulis, N. & Kreutz, R. (2009) Role of the H1 haplotype of microtubule-associated protein tau (MAPT) gene in Greek patients with Parkinson's disease. *BMC Neurol* **9**, 26.
- Risch, N. & Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517.
- Rocca, W. A., Maraganore, D. M., McDonnell, S. K. & Schaid, D. J. (1998) Validation of a telephone questionnaire for Parkinson's disease. *Journal of Clinical Epidemiology* **51**, 517–523.
- Ross, O. A., McCormack, R., Maxwell, L. D., Duguid, R. A., Quinn, D. J., Barnett, Y. A., Rea, I. M., El-Agnaf, O. M., Gibson, J. M., Wallace, A., Middleton, D. & Curran, M. D. (2003) mt4216C variant in linkage with the mtDNA TJ cluster may confer a susceptibility to mitochondrial dysfunction resulting in an increased risk of Parkinson's disease in the Irish. *Exp Gerontol* **38**, 397–405.
- Rubinsztein, D. C., Hanlon, C. S., Irving, R. M., Goodburn, S., Evans, D. G., Kellar-Wood, H., Xuereb, J. H., Bandmann, O. & Harding, A. E. (1994) Apo E genotypes in multiple sclerosis, Parkinson's disease, schwannomas and late-onset Alzheimer's disease. *Mol. Cell Probes* **8**, 519–525.
- Schapira, A. H. (2008) Mitochondria in the aetiology and pathogenesis of Parkinson's disease. *Lancet Neurol* **7**, 97–109.
- Scott, W. K., Nance, M. A., Watts, R. L., Hubble, J. P., Koller, W. C., Lyons, K., Pahwa, R., Stern, M. B., Colcher, A., Hiner, B. C., Jankovic, J., Ondo, W. G., Allen, F. H. Jr., Goetz, C. G., Small, G. W., Masterman, D., Mastaglia, F., Laing, N. G., Stajich, J. M., Slotterbeck, B., Booze, M. W., Ribble, R. C., Rampersaud, E., West, S. G., Gibson, R. A., Middleton, L. T., Roses, A. D., Haines, J. L., Scott, B. L., Vance, J. M. & Pericak-Vance, M. A. (2001) Complete genomic screen in Parkinson disease: evidence for multiple genes. *JAMA* **286**, 2239–2244.
- Singleton, A. B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R., Lincoln, S., Crawley, A., Hanson, M., Maraganore, D., Adler, C., Cookson, M. R., Muenter, M., Baptista, M., Miller, D., Blancato, J., Hardy, J. & Gwinn-Hardy, K. (2003) alpha-Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841.
- Skipper, L., Wilkes, K., Toft, M., Baker, M., Lincoln, S., Hulihan, M., Ross, O. A., Hutton, M., Aasly, J. & Farrer, M. (2004) Linkage disequilibrium and association of MAPT H1 in Parkinson disease. *Am J Hum Genet* **75**, 669–677.
- Spillantini, M. G., Schmidt, M. L., Lee, V. M., Trojanowski, J. Q., Jakes, R. & Goedert, M. (1997) Alpha-synuclein in Lewy bodies. *Nature* **388**, 839–840.
- Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., Baker, A., Jonasdottir, A., Ingason, A., Gudnadottir, V. G., Desnica, N., Hicks, A., Gylfason, A., Gudbjartsson, D. F., Jonsdottir, G. M., Sainz, J., Agnarsson, K., Birgisdottir, B., Ghosh, S., Olafsdottir, A., Cazier, J. B., Kristjansson, K., Frigge, M. L., Thorgeirsson, T. E., Gulcher, J. R., Kong, A. & Stefansson, K. (2005) A common inversion under selection in Europeans. *Nat Genet* **37**, 129–137.
- Sutherland, G. T., Halliday, G. M., Silburn, P. A., Mastaglia, F. L., Rowe, D. B., Boyle, R. S., O'Sullivan, J. D., Ly, T., Wilton, S. D. & Mellick, G. D. (2009) Do polymorphisms in the familial Parkinsonism genes contribute to risk for sporadic Parkinson's disease? *Mov Disord* **24**, 833–838.
- Thomas, B. & Beal, M. F. (2007) Parkinson's disease. *Hum Mol Genet* **16**(Spec No. 2), R183–R194.
- Tretiakoff, M. C. (1919) Thesis. University of Paris.
- Uhl, G. R., Liu, Q. R., Drgon, T., Johnson, C., Walther, D., Rose, J. E., David, S. P., Niaura, R. & Lerman, C. (2008) Molecular genetics of successful smoking cessation: convergent genome-wide association study results. *Arch Gen Psychiatry* **65**, 683–693.
- Van Der Walt, J. M., Nicodemus, K. K., Martin, E. R., Scott, W. K., Nance, M. A., Watts, R. L., Hubble, J. P., Haines, J. L., Koller, W. C., Lyons, K., Pahwa, R., Stern, M. B., Colcher, A., Hiner, B. C., Jankovic, J., Ondo, W. G., Allen, F. H. Jr., Goetz, C. G., Small, G. W., Mastaglia, F., Stajich, J. M., McLaurin, A. C., Middleton, L. T., Scott, B. L., Schmechel, D. E., Pericak-Vance, M. A. & Vance, J. M. (2003) Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. *Am J Hum Genet* **72**, 804–811.
- Vandrovicova, J., Pittman, A. M., Malzer, E., Abou-Sleiman, P. M., Lees, A. J., Wood, N. W. & de Silva, R. (2007) Association of

- MAPT haplotype-tagging SNPs with sporadic Parkinson's disease. *Neurobiol Aging* **30**, 1477–1482.
- Wahner, A. D., Glatt, C. E., Bronstein, J. M. & Ritz, B. (2007) Glutathione S-transferase mu, omega, pi, and theta class variants and smoking in Parkinson's disease. *Neurosci Lett* **413**, 274–278.
- Winkler, S., Konig, I. R., Lohmann-Hedrich, K., Vieregge, P., Kostic, V. & Klein, C. (2007) Role of ethnicity on the association of MAPT H1 haplotypes and subhaplotypes in Parkinson's disease. *Eur J Hum Genet* **15**, 1163–1168.
- Woodward, M. (2005) *Epidemiology Study Design and Data Analysis*. 2nd edition. Boca Raton: Chapman & Hall/CRC.
- Wright, K. O., Messing, E. M. & Reeder, J. E. (2004) DBCCR1 mediates death in cultured bladder tumor cells. *Oncogene* **23**, 82–90.
- Zabetian, C. P., Hutter, C. M., Factor, S. A., Nutt, J. G., Higgins, D. S., Griffith, A., Roberts, J. W., Leis, B. C., Kay, D. M., Yearout, D., Montimurro, J. S., Edwards, K. L., Samii, A. & Payami, H. (2007) Association analysis of MAPT H1 haplotype and subhaplotypes in Parkinson's disease. *Ann Neurol* **62**, 137–144.
- Zappia, M., Annesi, G., Nicoletti, G., Serra, P., Arabia, G., Pugliese, P., Messina, D., Caracciolo, M., Romeo, N., Annesi, F., Pasqua, A. A., Spadafora, P., Civitelli, D., Romeo, N., Epifanio, A., Morgante, L. & Quattrone A. (2003) Association of tau gene polymorphism with Parkinson's disease. *Neurol Sci* **24**, 223–224.
- Zarranz, J. J., Alegre, J., Gomez-Esteban, J. C., Lezcano, E., Ros, R., Ampuero, I., Vidal, L., Hoenicka, J., Rodriguez, O., Ateas, B., Llorens, V., Gomez, T. E., del Ser, T., Munoz D. G. & de Yebenes J. G. (2004) The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann Neurol* **55**, 164–173.
- Zhang, H., Jia, Y., Cooper, J. J., Hale, T., Zhang, Z. & Elbein, S. C. (2004) Common variants in glutamine:fructose-6-phosphate amidotransferase 2 (GFPT2) gene are associated with type 2 diabetes, diabetic nephropathy, and increased GFPT2 mRNA levels. *J Clin Endocrinol Metab* **89**, 748–755.
- Zimprich, A., Biskup, S., Leitner, P., Lichtner, P., Farrer, M., Lincoln, S., Kachergus, J., Hulihan, M., Uitti, R. J., Calne, D. B., Stoessl, A. J., Pfeiffer, R. F., Patenge, N., Carbajal, I. C., Vieregge, P., Asmus, F., Muller-Myhsok, B., Dickson, D. W., Meitinger, T., Strom, T. M., Wszolek, Z. K. & Gasser, T. (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* **44**, 601–607.

## Supporting Information

Additional supporting information may be found in the online version of this article:

**Figure S1a** Structure plot for the joint data for  $K = 2$  with 10,000 iterations of burn-in and 15,000 iterations of estimation on 5,000 independent SNPs with minimum minor allele

frequency 0.2. Illumina 610 samples (yellow), Illumina 550 samples (blue), Illumina 1M samples (red), HapMap CEU samples (green).

**Figure S1b** Plot of the 1st 2 principal components (X-axis = PC1, Y-axis = PC2) for a dataset containing the HIHG PD (orange circle), the CEPH HapMap samples (CEU: red square), the Yoruban HapMap samples (YRI: green triangle), the Japanese/Chinese HapMap samples (JPT+CHB: purple circle), the AD GWAS 1 controls (light blue circle), and AD GWAS 2 controls (black circle). Outlier removal was disabled for this plot.

**Figure S2a** The first principal component (X-axis) vs. the second principal component (Y-axis) plotted with Eigenstrat without outlier removal for the three GWAS samples used in this study (HIHG, NINDS, CIDR) along with the three primary HapMap samples (CEPH, JPT/CHB, YRI).

**Figure S2b** Structure triangle plot for  $K = 3$  with 10,000 iterations of burn-in and 15,000 iterations of estimation on 5,000 independent SNPs with minimum minor allele frequency 0.2 using CIDR (red), HIHG (blue) and NINDS (green) GWAS genotypes.

**Figure S3** Strong linkage disequilibrium in the 17q31.21 (MAPT) region. Here  $r^2$  is plotted (below) with the HapMap information track available in Haploview v4.1 for SNPs from the joint analysis with  $p < 1E-5$ .

**Table S1** Odds ratios and 95% confidence intervals using the major allele as the referent exposure for SNPs with trend test  $p < 1E-6$ .

**Table S2** Odds ratios and 95% confidence intervals using the major allele as the referent exposure for SNPs with trend test  $p < 0.05$  in all three studies.

**Table S3** Haplotype analysis for SNPs associated with PD in the chromosome 17q21.31 region with H1-H2 tagging SNP rs1981997.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received: 1 September 2009

Accepted: 3 November 2009