# Lack of replication of thirteen single-nucleotide polymorphisms implicated in Parkinson's disease: a large-scale international study

*Alexis Elbaz, Lorene M Nelson, Haydeh Payami, John P A Ioannidis, Brian K Fiske, Grazia Annesi, Andrea Carmine Belin, Stewart A Factor, Carlo Ferrarese, Georgios M Hadjigeorgiou, Donald S Higgins, Hideshi Kawakami, Rejko Krüger, Karen S Marder, Richard P Mayeux, George D Mellick, John G Nutt, Beate Ritz, Ali Samii, Caroline M Tanner, Christine Van Broeckhoven, Stephen K Van Den Eeden, Karin Wirdefeldt, Cyrus P Zabetian, Marie Dehem, Jennifer S Montimurro, Audrey Southwick, Richard M Myers, Thomas A Trikalinos*

## Summary

**Background** **A genome-wide association study identified 13 single-nucleotide polymorphisms (SNPs) significantly associated with Parkinson's disease. Small-scale replication studies were largely non-confirmatory, but a meta-analysis that included data from the original study could not exclude all SNP associations, leaving relevance of several markers uncertain.**

**Methods** **Investigators from three Michael J Fox Foundation for Parkinson's Research-funded genetics consortia—comprising 14 teams—contributed DNA samples from 5526 patients with Parkinson's disease and 6682 controls, which were genotyped for the 13 SNPs. Most (88%) participants were of white, non-Hispanic descent. We assessed log-additive genetic effects using fixed and random effects models stratified by team and ethnic origin, and tested for heterogeneity across strata. A meta-analysis was undertaken that incorporated data from the original genome-wide study as well as subsequent replication studies.**

**Findings** **In fixed and random-effects models no associations with any of the 13 SNPs were identified (odds ratios 0·89 to 1·09). Heterogeneity between studies and between ethnic groups was low for all SNPs. Subgroup analyses by age at study entry, ethnic origin, sex, and family history did not show any consistent associations. In our meta-analysis, no SNP showed significant association (summary odds ratios 0·95 to 1.08); there was little heterogeneity except for SNP rs7520966.**

**Interpretation** **Our results do not lend support to the finding that the 13 SNPs reported in the original genome-wide association study are genetic susceptibility factors for Parkinson's disease.**

*Lancet Neurol* 2006; 5: 917–23

Published Online September 27, 2006 DOI:10.1016/S1474-4422(06)70579-8

See Reflection and Reaction page 896

All study investigators are listed at the end of the report

**INSERM, Unit 708, Paris, France** (A Elbaz MD); **Division of Epidemiology, Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA** (L M Nelson PhD); **Wadsworth Center, New York State Department of Health, Albany, NY, USA** (H Payami PhD, J S Montimurro BS); **Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece** (J P A Ioannidis MD, T A Trikalinos MD); **The Michael J Fox Foundation for Parkinson's Research, New York, NY, USA** (B K Fiske PhD); **Institute of Neurological Sciences, National Research Council, Mangone, Italy** (G Annesi PhD); **Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden** (A Carmine Belin PhD); **Parkinson's Disease and Movement Disorder Clinic, Albany Medical Center, Albany, NY, USA** (D S Higgins MD); **Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA** (S A Factor DO); **Department of Neuroscience-Section of Neurology, University of Milano-Bicocca Ospedale San Gerardo, Monza, Italy** (C Ferrarese MD); **University of Thessaly, School of Medicine, Larissa, Greece** (G M Hadjigeorgiou MD); **Department of Epidemiology, Research Institute for Radiation**

## Introduction

Genome-wide screening for genetic associations is a promising approach for identification of the genetic determinants of common complex diseases.[1] One of the first applications of this emerging approach has been in the genetics of Parkinson's disease. A high-resolution genome-wide analysis of 198 345 single-nucleotide polymorphisms (SNPs) identified 13 SNPs exhibiting significant association with Parkinson's disease in a two-tiered study of white Americans with Parkinson's disease and healthy related and unrelated controls.[2] After the publication of that study, several investigators tried to replicate one or more of these associations.[3–7] The results of these follow-up studies have been largely non-confirmatory, leading to much controversy.[8,9]

In view of the importance of understanding the contribution of genetics to Parkinson's disease and the desire to provide further clarity to this research area, The Michael J Fox Foundation for Parkinson's Research, which funded the original genome-wide study, coordinated its own independent large-scale multicentre international replication effort. This study consisted of 14 international centres that contributed a combined sample size of more than 12 000 individuals. This is the largest genetics study of its kind to date for Parkinson's disease and the largest replication effort of genome-wide-derived associations in any specialty.

## Methods

### Study population

Investigators from three existing Edmond J Safra Global Genetics Consortia funded by The Michael J Fox Foundation for Parkinson's Research were invited to participate (table 1).[10,11] Investigators involved in the original genome-wide study[2] were not invited in order to maintain independence between the two studies.

### Procedures

Genotyping of DNA samples was undertaken either on-site (seven teams at an investigator laboratory or core facility) or through commercial contract (seven teams at Genoscreen, Lille, France). Genotypes were ascertained for all 13 SNPs reported in the original genome-wide study[2] by use of several genotyping platforms following standard protocols: TaqMan SNP Genotyping Assays (Applied Biosystems, Foster City, CA, USA); LightCycler with HybProbes (Roche, Basel, Switzerland); MassARRAY Analyzer Compact (Sequenom, San Diego, CA, USA); and Pyrosequencing on a PSQ HS 96(A) system (Biotage AB, Uppsala, Sweden). A random selection of at least 5% of samples was regenotyped

Biology and Medicine, Hiroshima University, Japan (H Kawakami MD); Center of Neurology and Hertie-Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany (R Krüger MD); The Gertrude H Sergivesky Center, New York, NY, USA (K S Marder MD, R P Mayeux MD); Eskitis Institute for Cell and Molecular Therapies, Griffith University, Nathan, Australia (G D Mellick PhD); Department of Neurology, Oregon Health & Science University, Portland, OR, USA (J G Nutt MD); Department of Epidemiology, UCLA School of Public Health, Los Angeles, CA, USA (B Ritz MD); Parkinson Disease Research Education and Clinical Center, VA Puget Sound Health Care System, Seattle, WA, USA (A Samii MD); The Parkinson's Institute, Sunnyvale, CA, USA (C M Tanner MD); Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, Flanders Interuniversity Institute for Biotechnology, Antwerpen, Belgium (C Van Broeckhoven PhD); Division of Research, Kaiser Foundation Research Institute, Oakland, CA, USA (S K Van Den Eeden PhD); Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (K Wirdefeldt MD); Geriatric Research Education and Clinical Center, VA Puget Sound Health Care System, Seattle, WA, USA (C P Zabetian MD); Department of Neurology, University of Washington, Seattle, WA, USA (C P Zabetian); Genoscreen, Lille, France (M Dehem PhD); and Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA (A Southwick, R M Myers)

Correspondence to: Brian K Fiske, The Michael J Fox Foundation for Parkinson's Research, Grand Central Station, PO Box 4777, New York, NY 10163, USA
**bfiske@michaeljfox.org**

See Online for webtable 1

to determine precision; genotyping error rates were lower than 0·5% for all genotyping sites.

### Statistical analysis

In the analyses, participants were stratified according to the team that recruited them and to their ethnic origin (white non-Hispanic, Hispanic, Asian, African American, Native American, and other). We used an exact test to assess among controls in each stratum whether the genotype distributions for each of the 13 SNPs violated Hardy Weinberg equilibrium (HWE). This test was done only among women for the X-linked SNP9 (rs7878232). Deviation from HWE was deemed significant for $p<0·05$, but we accepted that given the extreme number of HWE tests done (n=403) some strata might exhibit significant HWE deviation simply by chance. We used the same allele coding as in the original genome-wide study;[2] the reference allele was the major frequency allele for all SNPs except for SNP3 (rs2313982), SNP8 (rs2245218), and SNP10 (rs1509269).

For quantitative syntheses we first did analyses adjusted for team and ethnic origin. Analyses were further adjusted for age at study entry and sex; for these analyses we excluded data from strata with fewer than 20 individuals. We assessed genetic effects with the assumption of log-additive (multiplicative) models in logistic regressions and we synthesised results across strata with both fixed and random-effects models. Fixed-effects models assume that odds ratios (ORs) are constant across all teams and ethnic subgroups and that observed differences are due to chance. Random-effects models allow that results might be genuinely different (heterogeneous) across teams and ethnic subgroups and they take into account between-study heterogeneity. In the presence of heterogeneity, random-effects syntheses are preferable.[12,13] We tested for between-study heterogeneity with the $\chi^2$-based Q statistic (formally deemed significant for $p<0·10$)[13] and quantified its extent with $I^2$, which ranges from 0% to 100% and represents the proportion of between-study variability ascribed to heterogeneity rather than to chance.[14] Values for $I^2$ of 0–24% suggest little heterogeneity, 25–49% reflect moderate heterogeneity, 50–74% reflect large heterogeneity, and more than 75% reflect very large heterogeneity. We assessed whether any summary results were nominally significant at $p<0·05$ and at $p<0·004$ (correcting for 13 polymorphisms).

Subgroup analyses were undertaken according to ethnic origin, age at study entry (cutoff at 25th percentile=60 years), sex, and presence or not of family history of Parkinson's disease in first-degree relatives. We assessed whether any differences between subgroups were nominally significant at $p<0·05$ and at $p<0·0013$ (correcting for 13 polymorphisms and three subgroup analyses per polymorphism).

Finally, we also did meta-analyses incorporating the data from tier two of the original genome-wide association study[2] as well as from preliminary replication efforts published until May 20, 2006.[3–7] We combined ORs and their variances (using consistent allele coding) with the ORs and variances from each team and subgroup included in our collaborative analysis using the inverse variance method. We used both fixed and random-effects models. Heterogeneity was assessed with the Q and $I^2$ statistics, as described above. All analyses were done with Intercooled Stata 8.2 (College Station, TX, USA).

### Role of the funding source

The Michael J Fox Foundation for Parkinson's Research played a part in the identification of study investigators; coordination of the study design; data collection, analysis, and interpretation; and writing of the report. Other funding sources played no role outside of study sponsorship. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## Results

The 14 teams contributed a total sample size of 5578 patients and 6765 controls. Of these, we excluded 46 participants because of missing information about sex, 19 because of missing information about ethnic origin, 52 twins from one study, and 18 male participants whose genotyping spuriously showed heterozygosity for SNP9 (located on the X chromosome), an indication of data error or Klinefelter syndrome. The remaining 5526 patients and 6682 controls were included in the analysis. Most (n=10767; 88%) were of white, non-Hispanic descent, whereas 896 (7%) were of Asian descent, 344 (3%) were of Hispanic descent, 141 (1%) were of African-American descent, 21 (0·2%) were Native Americans or Pacific Islanders, and 39 (0·3%) were of other descent. The proportion of men ranged between 41% and 62% across participating teams and ethnic groups. The mean age at diagnosis of Parkinson's disease ranged between 55·4 years and 67·5 years and the mean age at study entry ranged between 64·2 years and 74·6 years for patients and between 60·9 years and 73·8 years for controls. Among participants with Parkinson's disease, 547 (13%) had a documented family history of the disease, whereas 3846 reported no such family history, and for 1133 this information was unknown (table 1).

The distribution of genotypes was ascertained for each SNP and for each team and ethnic subgroup (webtable 1). Genotype distributions were consistent with HWE among controls in most of the strata for the thirteen SNPs. Genotypes were not in HWE for seven SNPs (1, 2, 3, 5, 6, 8, 9) in a total number of 13 strata: p values were greater than 0·01 in ten strata and lower than 0·01 in three strata (SNP3 in the Payami study, white non-Hispanic, SNP9 in the Kawakami study, Asian or Pacific Islander, and in the Hadjigeorgiou study, white non-Hispanic).

Summary ORs for all 13 SNPs in random-effects models adjusted for team and ethnic origin were very close to unity, ranging from 0·92 to 1·07 (table 2), indicating no significant association, even at an

uncorrected p value threshold of 0·05. The 95% CIs were also very tight, excluding ORs smaller than 0·84 or larger than 1·18. There was no heterogeneity for seven SNPs, little heterogeneity for five SNPs, and moderate heterogeneity for SNP13 (p=0·07). The fixed-effects calculation yielded similar—if not identical—results; nominal significance was seen only for SNP5 (p=0·049). Analyses restricted to participants of white, non-Hispanic descent also showed very similar estimates, with summary ORs between 0·93 and 1·09 in both fixed and random-effects calculations.

When adjustment for age at study entry and sex was taken into account, the results were the same (table 3). In random-effects calculations, all summary ORs were between 0·89 and 1·09 and their 95% CIs excluded ORs smaller than 0·81 or larger than 1·19. There was no significant heterogeneity for any SNP, and the largest $I^2$ was only 27%. Fixed-effects estimates were very similar. None of the SNPs had even nominal significance. Additionally, no significant associations were seen when analyses were restricted to white, non-Hispanic descent populations. Heterogeneity remained at low levels, except for SNP4 (p=0·02, $I^2$=51%). Exclusion of the strata in which controls were not in HWE did not modify these findings (data not shown).

Data were very limited for populations of non-white, non-Hispanic descent; however, the available results were consistent across ethnic subgroups (webtables 2–5). In analyses adjusted for team, p values for the comparison between populations of white, non-Hispanic descent and of Asian descent (the two most common groups) were less than 0·05, except for SNP11 (p=0·010). In analyses adjusted for age at study entry and sex, the corresponding p value was also significant for SNP7 (p=0·04). For none of the SNPs did the p value fall below the threshold of 0·004 (critical 95% alpha threshold accounting for 13 SNPs). For analyses by age at study entry, different cutoffs were used with similar results.

Nine of the 13 polymorphisms did not give any nominally significant signals in any subgroup analyses with random-effects modeling (figure 1). For some SNPs, nominal significance (p<0·05) was seen in specific subgroups: familial Parkinson's disease (OR 0·67, 95% CI 0·53–0·84; p=0·0007) and male population (0·90, 0·81–1·00; p=0·045) for SNP4; older (≥60 years) participants (1·09, 1·01–1·18; p=0·028) and familial Parkinson's disease (1·27, 1·06–1·53; p=0·009) for SNP8; non-familial Parkinson's disease (0·92, 0·84–1·00; p=0·049) and male population (0·88, 0·77–1·00; p=0·048) for SNP9; and familial Parkinson's disease (0·85, 0·74–0·99; p=0·030) for SNP13. Only the genetic effect of SNP4 in familial Parkinson's disease crossed the p value threshold after correction for 13 SNPs and four subgroup analyses (p<0·0013). However, given that random-effects models give more weight to smaller studies than to larger studies, this effect was probably mainly driven by strata with very small numbers. In most studies, the number of familial Parkinson's disease cases was small, whereas the Payami study contributed 44% of all familial Parkinson's

See Online for webtables 2–5

| Team* | Location | Cases | | | | | | | | Controls | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N | White† (%) | Male (%) | Mean age at study entry (years) | Mean age at diagnosis‡ (years) | Familial Parkinson's disease‡ (n) | Source | Diagnostic criteria | N | White (%) | Male (%) | Mean age at study entry (years) | Source |
| Nelson (1)* | USA | 597 | 80 | 61 | 65·8 | 64·1 | 61 | Population-based | CAPIT | 639 | 82 | 63 | 66·8 | Population-based |
| Ritz (1) | USA | 224 | 87 | 54 | 69·1 | 67·5 | 22 | Population-based | CAPIT | 198 | 85 | 56 | 67·7 | Population-based |
| Marder – Mayeux (1) | USA | 314 | 75 | 54 | 67·9 | 61·7 | 10 | Population-based/clinic | UKPDBB | 317 | 62 | 48 | 73·0 | Population-based |
| Payami (2) | USA | 1554 | 95 | 68 | 68·0 | 59·9 | 238 | Clinic | UKPDBB | 1934 | 94 | 37 | 66·8 | Spouses/community |
| Annesi (3) | Italy | 200 | 100 | 58 | 69·4 | 63·9 | 0 | Clinic | UKPDBB | 200 | 100 | 46 | 72·5 | Community |
| Carmine (3) | Sweden | 119 | 100 | 66 | 66·9 | 58·6 | 12 | Clinic | UKPDBB | 135 | 100 | 53 | 67·7 | Spouses/community |
| Elbaz (3) | France | 209 | 100 | 57 | 66·9 | 63·4 | 19 | Population-based | Bower | 501 | 100 | 59 | 66·9 | Population-based |
| Ferrarese (3) | Italy | 100 | 100 | 51 | 66·4 | 57·5 | 24 | Clinic | Gelb | 105 | 100 | 57 | 60·9 | Spouses/blood donors |
| Hadjigeorgiou (3) | Greece | 177 | 100 | 57 | 70·0 | 64·2 | 27 | Clinic | Bower | 155 | 100 | 46 | 67·6 | Community |
| Kawakami (3) | Japan | 304 | 0 | 48 | 68·8 | 62·9 | 12 | Clinic | Bower | 464 | 0 | 36 | 70·6 | Spouses/community |
| Krueger (3) | Germany | 834 | 100 | 43 | 64·2 | 55·4 | 0 | Clinic | UKPDBB | 1192 | 100 | 42 | 57·7 | Blood donors/MEMO study |
| Mellick (3) | Australia | 545 | 100 | 60 | 70·9 | 59·7 | 109 | Clinic | Calne | 446 | 100 | 41 | 69·2 | Spouses/community |
| Van Broeckhoven (3) | Belgium | 271 | 100 | 58 | 67·7 | 59·6 | 9 | Clinic | Pals/ Engelborghs§ | 278 | 100 | 56 | 65·2 | Spouses/community |
| Wirdefeldt (3) | Sweden | 78 | 100 | 50 | 74·6 | 64·1 | 4 | Population-based | Gelb | 118 | 100 | 46 | 73·8 | Community/unrelated twins |

*Teams participating in the same consortium are identified by the same number. †White, non-Hispanic descent. ‡Mean age at diagnosis and familial Parkinson's disease among first-degree relatives pertain to cases only. §For Pals/Engelborghs diagnostic criteria please see references 10 and 11.

***Table 1:*** **Descriptive and clinical characteristics of participants by team**

| | dbSNP number | Overall | | | | White, non-Hispanic participants | | | |
|---|---|---|---|---|---|---|---|---|---|
| SNP | | N (alleles) | RE OR (95% CI) | FE OR (95% CI) | $p_{Het}$ ($I^2$%) | N (alleles) | RE OR (95% CI) | FE OR (95% CI) | $p_{Het}$ ($I^2$%) |
| 1 | rs7702187 | 30 (24066) | 1·03 (0·94–1·13) | 1·00 (0·94–1·07) | 0·18 (19) | 13 (21210) | 1·01 (0·93–1·09) | 1·00 (0·93–1·08) | 0·40 (4) |
| 2 | rs10200894 | 28 (23666) | 1·05 (0·96–1·14) | 1·05 (0·96–1·14) | 0·66 (0) | 13 (20796) | 1·08 (0·98–1·18) | 1·08 (0·98–1·19) | 0·52 (0) |
| 3 | rs2313982 | 27 (24008) | 0·94 (0·85–1·04) | 0·94 (0·85–1·04) | 0·80 (0) | 13 (21138) | 0·96 (0·86–1·06) | 0·96 (0·86–1·06) | 0·70 (0) |
| 4 | rs17329669 | 31 (23852) | 0·92 (0·84–1·01) | 0·94 (0·87–1·01) | 0·28 (12) | 13 (20984) | 0·91 (0·80–1·02) | 0·94 (0·87–1·01) | 0·03* (47) |
| 5 | rs7723605 | 31 (23888) | 1·07 (0·98–1·18) | 1·08 (1·00–1·16)* | 0·24 (15) | 13 (21036) | 1·09 (1·01–1·18)* | 1·09 (1·01–1·19)* | 0·64 (0) |
| 6 | ss46548856 | 31 (24040) | 1·02 (0·94–1·11) | 1·02 (0·94–1·11) | 0·63 (0) | 13 (21176) | 1·04 (0·95–1·14) | 1·04 (0·95–1·14) | 0·61 (0) |
| 7 | rs16851009 | 29 (23158) | 1·05 (0·95–1·16) | 1·06 (0·97–1·15) | 0·30 (11) | 13 (20294) | 1·00 (0·89–1·13) | 1·01 (0·92–1·11) | 0·17 (27) |
| 8 | rs2245218 | 31 (23932) | 1·05 (0·98–1·12) | 1·05 (0·98–1·12) | 0·81 (0) | 13 (21066) | 1·05 (0·97–1·13) | 1·05 (0·97–1·13) | 0·45 (0) |
| 9 | rs7878232 | 25 (14714) | 0·93 (0·86–1·01) | 0·93 (0·86–1·01) | 0·85 (0) | 11 (12507) | 0·93 (0·85–1·01) | 0·93 (0·85–1·01) | 0·71 (0) |
| 10 | rs1509269 | 29 (23710) | 0·96 (0·87–1·05) | 0·95 (0·87–1·04) | 0·65 (0) | 13 (20848) | 0·97 (0·88–1·06) | 0·97 (0·88–1·06) | 0·42 (2) |
| 11 | rs11737074 | 29 (23956) | 0·98 (0·91–1·04) | 0·98 (0·91–1·04) | 0·60 (0) | 13 (21084) | 0·99 (0·92–1·06) | 0·99 (0·92–1·06) | 0·75 (0) |
| 12 | rs682705 | 31 (20788) | 1·00 (0·92–1·09) | 0·99 (0·93–1·05) | 0·12 (24) | 13 (17924) | 0·99 (0·92–1·06) | 0·99 (0·92–1·06) | 0·79 (0) |
| 13 | rs7520966 | 31 (23834) | 0·99 (0·91–1·08) | 0·98 (0·93–1·04) | 0·07* (29) | 13 (20964) | 0·98 (0·92–1·04) | 0·98 (0·92–1·04) | 0·71 (0) |

N=number of teams and ethnic subgroups contributing to the analysis. FE=fixed effects. RE=random effects. The reference allele was the major frequency allele for all SNPs, except for SNP3 (rs2313982), SNP8 (rs2245218), and SNP10 (rs1509269). *Nominal statistical significance without correction for multiple comparisons (p<0·05 for summary ORs and $p_{Het}$<0·10 for heterogeneity).

***Table 2:*** **Summary effects analyses adjusted for team and ethnic origin**

| | dbSNP number | Overall | | | | White, non-Hispanic only | | | |
|---|---|---|---|---|---|---|---|---|---|
| SNP | | N (alleles) | RE OR (95% CI) | FE OR (95% CI) | $p_{Het}$ ($I^2$%) | N (alleles) | RE OR (95% CI) | FE OR (95% CI) | $p_{Het}$ ($I^2$%) |
| 1 | rs7702187 | 24 (23554) | 1·01 (0·93–1·10) | 1·00 (0·93–1·07) | 0·30 (12) | 13 (20850) | 1·00 (0·92–1·08) | 1·00 (0·92–1·08) | 0·51 (0) |
| 2 | rs10200894 | 24 (23156) | 1·05 (0·96–1·15) | 1·05 (0·96–1·15) | 0·71 (0) | 13 (20440) | 1·08 (0·98–1·20) | 1·08 (0·98–1·20) | 0·62 (0) |
| 3 | rs2313982 | 22 (23366) | 0·94 (0·83–1·07) | 0·94 (0·85–1·04) | 0·29 (13) | 13 (20770) | 0·96 (0·85–1·09) | 0·96 (0·86–1·07) | 0·34 (11) |
| 4 | rs17329669 | 24 (23338) | 0·90 (0·81–1·01) | 0·94 (0·87–1·02) | 0·14 (24) | 13 (20622) | 0·89 (0·78–1·01) | 0·94 (0·87–1·02) | 0·02* (51) |
| 5 | rs7723605 | 24 (23366) | 1·08 (0·97–1·19) | 1·07 (0·99–1·16) | 0·12 (26) | 13 (20668) | 1·09 (0·99–1·19) | 1·08 (1·00–1·18) | 0·40 (5) |
| 6 | ss46548856 | 23 (23404) | 1·03 (0·95–1·13) | 1·03 (0·95–1·13) | 0·45 (1) | 13 (20808) | 1·06 (0·96–1·16) | 1·06 (0·96–1·16) | 0·63 (0) |
| 7 | rs16851009 | 24 (22654) | 1·05 (0·93–1·19) | 1·06 (0·97–1·16) | 0·11 (27) | 13 (19944) | 1·00 (0·89–1·13) | 1·01 (0·91–1·11) | 0·22 (22) |
| 8 | rs2245218 | 23 (23364) | 1·04 (0·97–1·12) | 1·04 (0·97–1·12) | 0·74 (0) | 13 (20698) | 1·05 (0·97–1·13) | 1·05 (0·97–1·13) | 0·70 (0) |
| 9 | rs7878232 | 24 (15046) | 0·93 (0·86–1·01) | 0·93 (0·86–1·01) | 0·75 (0) | 13 (12952) | 0·92 (0·85–1·00) | 0·92 (0·85–1·00) | 0·73 (0) |
| 10 | rs1509269 | 22 (23076) | 0·94 (0·82–1·07) | 0·95 (0·86–1·04) | 0·13 (26) | 13 (20488) | 0·96 (0·84–1·10) | 0·96 (0·88–1·06) | 0·11 (34) |
| 11 | rs11737074 | 24 (23432) | 0·95 (0·87–1·04) | 0·97 (0·90–1·04) | 0·25 (15) | 13 (20714) | 0·98 (0·92–1·06) | 0·98 (0·92–1·06) | 0·67 (0) |
| 12 | rs682705 | 24 (20362) | 0·99 (0·92–1·06) | 0·99 (0·93–1·05) | 0·41 (4) | 13 (17652) | 0·99 (0·92–1·06) | 0·99 (0·92–1·06) | 0·94 (0) |
| 13 | rs7520966 | 24 (23324) | 0·99 (0·92–1·06) | 0·99 (0·93–1·05) | 0·33 (10) | 13 (20608) | 0·98 (0·92–1·04) | 0·98 (0·92–1·04) | 0·92 (0) |

N=number of teams and ethnic subgroups contributing to the analysis. FE=fixed effects. RE=random effects. The reference allele was the major frequency allele for all SNPs, except for SNP3 (rs2313982), SNP8 (rs2245218), and SNP10 (rs1509269). *Nominal statistical significance without correction for multiple comparisons (p<0·05 for summary ORs and $p_{Het}$<0·10 for heterogeneity).

***Table 3:*** **Summary effects analyses adjusted for team, ethnic origin, age at study, and sex**

disease cases. In the Payami study the OR in familial Parkinson's disease for SNP4 was 1·02 (95% CI 0·78–1·33). When compared against their complementary subgroups, only three of these seven effects were nominally different at p<0·05: SNP4 in familial versus non-familial Parkinson's disease (p=0·008) and SNP8 in older versus younger age populations (p=0·010) and in familial versus non-familial Parkinson's disease (p=0·029). However, all of these three nominally significant subgroup effects were in the opposite direction compared with the effects reported in the original whole-genome association study;[2] thus, they do not indicate replication.

In meta-analyses including our data as well as the tier-two data from the original whole-genome association study[2] and several smaller replication studies,[3–7] none of the 13 SNPs showed nominally significant association even at p>0·064; all summary ORs were between 0·95 and 1·08 with 95% CIs excluding ORs smaller than 0·87 or larger than 1·18. Heterogeneity was formally significant only for SNP13 (p=0·03, $I^2$=34%). Apart from SNP13, there was generally limited heterogeneity ($I^2$<23%). Exclusion of the original tier-two data, which were also exploratory (almost 2000 SNPs were tested), did not affect our results. If anything, heterogeneity tended to diminish (SNP13 was only marginally significantly heterogeneous, p=0·09, $I^2$=26%). The 95% CIs of the tier-two data typically did not overlap at all with the overall meta-analysis (figure 2).

## Discussion

The present results do not lend support to the finding that the 13 SNPs reported in the original two-tier genome-wide association study[2] are genetic susceptibility loci for Parkinson's disease. Although effects were seen in a few subgroup analyses, these did not correspond in direction to the original report, suggesting that they were probably spurious.

Our collaborative study is the largest replication genetics study undertaken in Parkinson's disease to date. Although the study included populations with large ethnic heterogeneity, the results were consistent, with no major statistical between-study heterogeneity in any of the main analyses. Genotypes in a few strata were not in HWE, but this is not surprising given the large number of tests done. In most of these cases deviations were not very significant ($p>0·01$); exclusion from our analyses of the three strata where p values were below 0·01 did not change our findings.

One interpretation of this lack of replication is that the 13 SNPs highlighted in the original genome-wide study[2] were not true risk factors but were probably false-positive findings. Indeed, assuming that 1% (using $p<0·01$) of the SNPs assayed in the original report would be expected to show an association to Parkinson's disease by chance alone, most SNPs (1862 of 198 345; 0·9%) carried from tier one into tier two were probably false positives. Although slightly more SNPs (26 of 1793; 1·4%) than expected by chance showed significant association with Parkinson's disease in tier two, only around half of them (11 of 26) showed the same direction of effect in both tiers; the other 15 SNPs were extreme opposites that could be seen simply by chance.[15] Furthermore, none of the SNPs identified in either tier of the original report met statistical—albeit conservative—criteria after adjustment for multiple comparisons.

The study by Maraganore and colleagues[2] is the first of its type for Parkinson's disease and was a valiant effort given that technology and bioinformatics are still relatively immature for doing whole-genome association studies. However, based on the current replication results, and in hindsight, it is perhaps useful to discuss several points that could inform future efforts focused on the genetics of Parkinson's disease and other common diseases of multifactorial origin.

First, the effect sizes for genetic determinants of Parkinson's disease may be small; evidence from gene–disease associations identified to date suggests that, with some exceptions, most associations have ORs in the range of 1·1–1·6.[16] Because the power calculations of the original genome-wide study[2] were based on an OR of 2·0, the original study was underpowered to detect effect sizes in the lowest range. Tier one of the original genome-wide study[2] used a sibship-based association design to identify initial SNP associations, and was powered to detect ORs of 2·0 or higher. Although this design has the advantage of being most robust to population

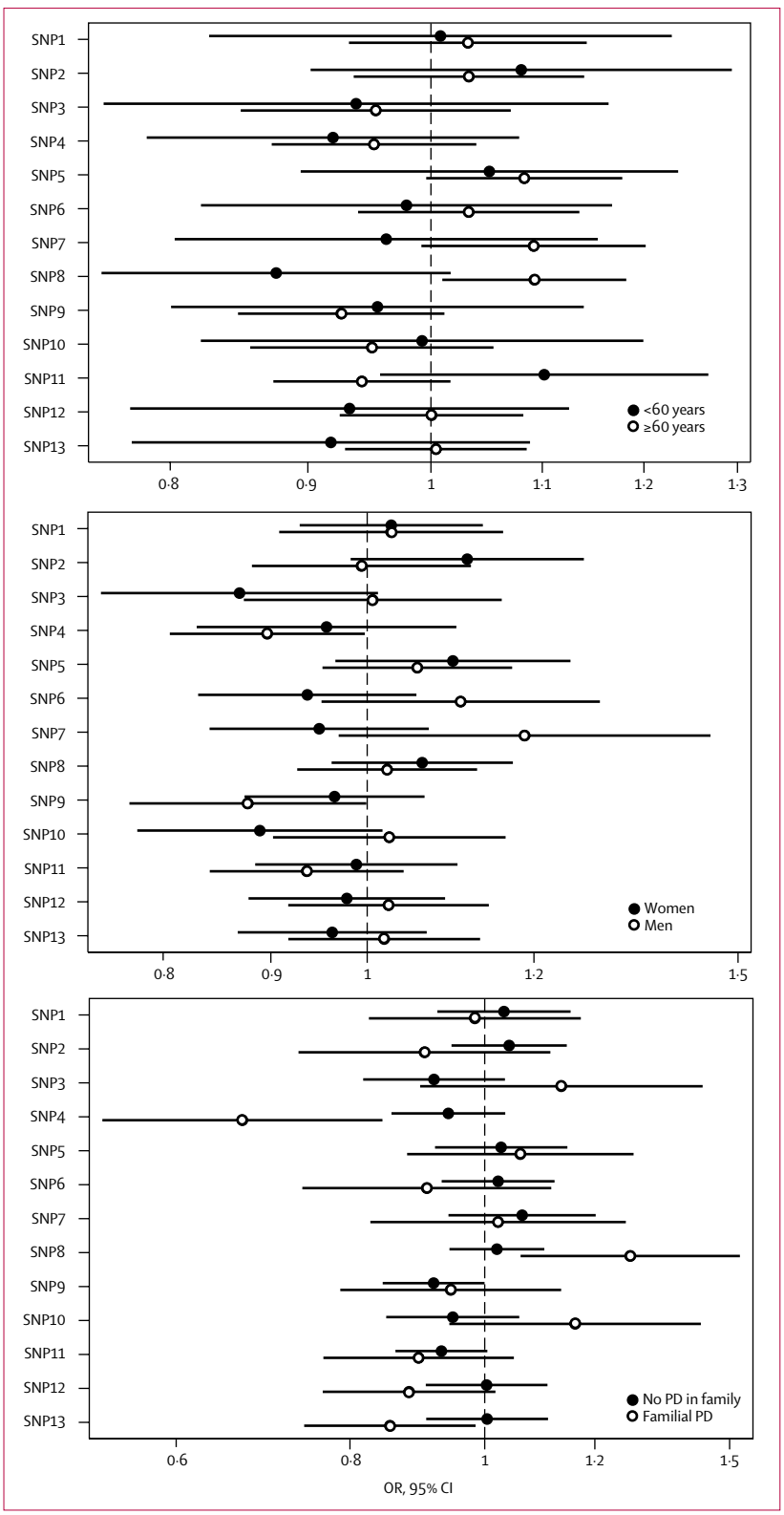


***Figure 1:*** **Subgroup analyses for (A) age at study entry, (B) sex, and (C) history of Parkinson's disease**
In each plot the two subgroups are shown for each of the 13 SNPs with point estimates of summary OR and 95% CIs according to random-effects calculations. Fixed-effects estimates are very similar (not shown).

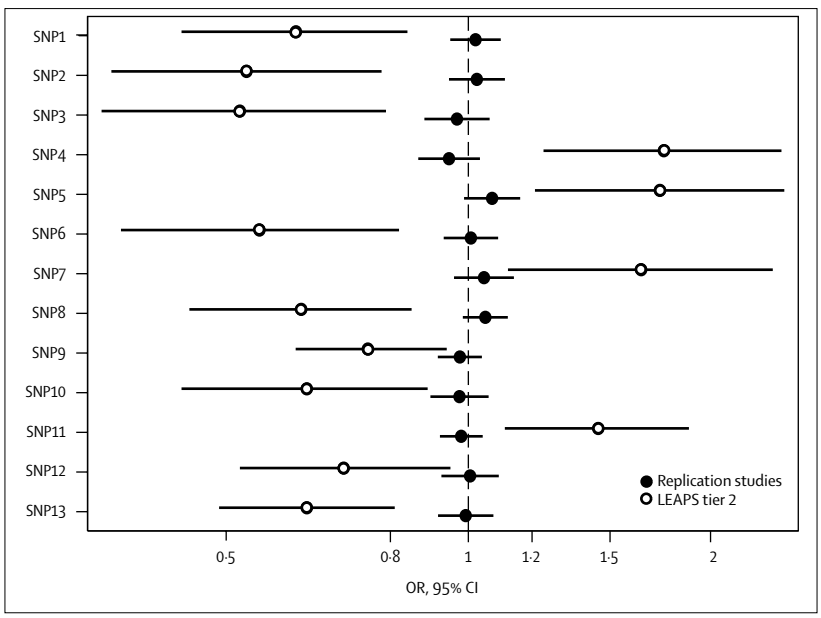


***Figure 2:*** **Tier 2 results from the whole genome-association versus meta-analysis of all replication data**
Point estimates and 95% CIs are shown (random-effects calculations for the meta-analysis). Results are shown for each of the 13 SNPs.

stratification, it can be underpowered relative to case-unrelated control studies.[17] With 332 matched case-unrelated control pairs, tier two was also underpowered to detect ORs in the very low range. Limited power may also result in false-negative findings. Some genuine determinants of Parkinson's disease could therefore still be among those screened in the original study. Along these lines, we applaud Fung and colleagues[18] for making their genome-wide SNP analysis dataset publicly available. In parallel, The Michael J Fox Foundation for Parkinson's Research has been planning a mechanism by which to make the original dataset of Maraganore and colleagues,[2] as well as the current replication data, available to researchers within the near future. Further scrutiny of these rich databases and integration with future efforts will be important.

Second, studies may require greater genome coverage and more informative markers to be able to detect true genetic determinants of Parkinson's disease. Coverage of the 250K SNP chip from Perlegen Sciences that was used by the original genome-wide study might be less than 60%, even when linkage disequilibrium of 80% is allowed. This might not have provided adequate coverage of the genome and informative markers could have been left out despite the relatively large number of SNPs. Even though current chips are often termed as whole-genome screens, in fact they offer only partial coverage of the estimated 10 million variants in the human genome.[19] This would be even more problematic if some genetic determinants are highly population-specific or if they express themselves through complex interactions between several markers or environmental factors. Our study only sought to replicate in a larger population the specific 13 SNPs highlighted in Maraganore and colleagues' study[2] as having significant association with Parkinson's disease because we felt such a replication would be a critical first step to understanding these results. Further study of the genetic regions surrounding these SNPs using greater marker density could still be useful.

Third, designing a whole-genome study that has adequate statistical power in the context of multiple hypothesis testing is challenging. Consensus is lacking for the most appropriate statistical methods to use. Some investigators favour an approach that uses an initial sample that is relatively small with the use of a liberal p value, followed by a second screen in an independent population that is of a larger size with a more stringent p value.[1] In the case of the original genome-wide study,[2] the second tier sample size was smaller (n=332 pairs) than the first tier sample size (n=443 pairs). Other approaches for selection of samples for each stage of a whole-genome association study of Parkinson's disease have also been proposed.[8] The relative merits of various multistage approaches need to be validated empirically based on whether their application is successful. Interestingly, the most clear success from genome-wide approaches to date (the identification of a genetic marker for age-related macular degeneration)[20] used an extremely small sample size. In this case, however, the effect size was unusually large. That being said, none of the other genetic determinants for complex diseases identified through genome-wide approaches has been subjected to the extent of replication that we have done, so conclusions from these studies need to be drawn cautiously.[21–24]

Finally, the main effects of genetic factors can be masked in the presence of underlying gene–environment interaction. If environmental factors are necessary for interacting with genetic variants to increase the risk of Parkinson's disease, studies that do not measure and account for this interaction might fail to identify important susceptibility genes.

Whole-genome association studies are still in their infancy and it is perhaps not surprising that the first attempt at a genome-wide association study of Parkinson's disease did not yield replicable results. Several recent advances could improve such studies in the future, including the identification of tag SNPs to more comprehensively cover the genome or the selection of a dense map of missense SNPs in coding and regulatory regions. A drop in the per-marker cost of genotyping platforms might bring this technology into practical reach of more investigators. Future approaches might also include doing whole-genome scans in founder populations or in early onset Parkinson's disease populations in which the genetic contributions to Parkinson's disease risk are strongest.[25] In any case, future whole-genome association studies of Parkinson's disease will necessitate larger numbers of participants and careful attention to patient selection for each stage of study.

### Contributors

AE, LN, HP, JI, and BF coordinated the study, contributed to study design, data analysis and interpretation, coordinated data collection, and assisted with writing of the manuscript. MD, JM, and RMM genotyped samples for multiple sites and provided additional overall technical assistance. TT assisted with overall data analysis and interpretation and manuscript preparation. All other authors assisted in coordinating sample preparation and data collection for each individual study site. All authors contributed to critical review of the manuscript and have seen and approved the final version.

### Conflicts of interest

BF is employed by The Michael J Fox Foundation for Parkinson's Research, which funded the study. MD is employed by Genoscreen, which has received funding from the Michael J Fox Foundation for Parkinson's Research. All other authors have no conflicts of interest.

### Acknowledgments

We thank the laboratory of Dr Matthew Farrer (Neurogenetic Laboratories, Udall Parkinson's Disease Research Center of Excellence, Mayo Clinic, Jacksonville, FL, USA) for providing details regarding TaqMan SNP Genotyping Assays for the 13 SNPs. Funding primarily to cover genotyping costs for the current study was provided to the three genetic consortia by The Michael J Fox Foundation for Parkinson's Research through a generous gift from the Edmond J Safra Philanthropic Foundation. Additional funding to individual investigators for other aspects of the original studies (eg, original sample collection and other support costs) was provided by: INSERM and the French Ministry of Environment (AE, PA, JCL, CT); NIH NS R01-31964 and Tobacco-Related Disease Research Fund Grant 8RT-0131 (LN); NIH NS R01-36960 and AG 08017 (HP); Swedish Research Council, Swedish Brain Foundation and the Hållsten Foundation, The Swedish Parkinson Foundation, Swedish Brain Power (ACB, DG, LO, OS, MW); German Ministry of Education and Research, Program NGFN2, No. 01GS0468 (RKr); National Parkinson Foundation (JN); US Department of Veterans Affairs Parkinson Disease Research Education and Clinical Center grant (ASa); the Special Research Fund of the University of Antwerp, the Fund for Scientific Research Flanders an EU contract LSHMCT-2003-503330 (CVB, PC, PDD, KN, PP, BP, JT); NIH Grant ES10758 (KW); NIH K08-NS044138; and a Veterans Affairs Merit Review Award (CZ).

### Study investigators

In addition to the main authors listed above, the individuals listed below also contributed to the study. *Australia*: Peter A Silburn (Institute for Cell and Molecular Therapies). *Belgium*: Karen Nuytemans (Flanders Interuniversity Institute for Biotechnology); Philippe Pals (Flanders Interuniversity Institute for Biotechnology); Barbara Pickut (Middelheim General Hospital); Patrick Cras (Institute Born-Bunge); Peter Paul De Deyn (Institute Born-Bunge); Jessie Theuns (Flanders Interuniversity Institute for Biotechnology). *France*: Christophe Tzourio (INSERM, Unit 708); Philippe Amouyel (INSERM, Unit 744 ); Jean-Charles Lambert (INSERM, Unit 744). *Germany*: Olaf Riess (University of Tübingen); Peter Bauer (University of Tübingen); Thomas Gasser (University of Tübingen); Daniela Berg (University of Tübingen). *Greece*: Georgia Xiromerisiou (University of Thessaly School of Medicine). *Italy*: Alessio Galbussera (University of Milano-Bicocca); Michela Zini (University of Milano-Bicocca and Parkinson Institute); Stefano Goldwurm (Parkinson Institute); Gianni Pezzoli (Parkinson Institute); Aldo Quattrone (Institute of Neurology University Magna Graecia); Ferdinanda Annesi (National Research Council); Patrizia Tarantino (National Research Council). *Japan*: Hiroyuki Morino (Hiroshima University); Hirofumi Maruyama (Hiroshima University); Yuisin Izumi (University of Tokushima Graduate School); Ryuji Kaji (University of Tokushima Graduate School). *Sweden*: Nancy L Pedersen (Karolinska Institutet); Lars Olson (Karolinska Institutet); Marie Westerlund (Karolinska Institutet); Dagmar Galter (Karolinska Institutet); Olof Sydow (Karolinska University Hospital). *United States*: Valerie McGuire (Stanford University School of Medicine); Barbara Topol (Stanford University School of Medicine); Angelika Wahner (UCLA School of Public Health); Yvette M Bordelon (UCLA Medical Center); Lorraine Clark (Columbia University); Ming-Xin Tang (Columbia University); Jeff Bronstein (UCLA School of Medicine); Shannon Brady (Stanford University School of Medicine); Amita Aggarwal (Stanford University School of Medicine); Xia Liu (Stanford University School of Medicine); Alida Griffith (Evergreen Hospital Medical Center); Berta C Leis (Evergreen Hospital Medical Center); John W Roberts (Virginia Mason Medical Center).

### References

1 Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005; **6:** 95–108.
2 Maraganore DM, De Andrade M, Lesnick TG, et al. High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet* 2005; **77:** 685–93.
3 Bialecka M, Kurzawski M, Klodowska-Duda G, Opala G, Tan EK, Drozdzik M. Polymorphism in semaphorin 5A (Sema5A) gene is not a marker of Parkinson's disease risk. *Neurosci Lett* 2006; **399:** 121–23.
4 Clarimon J, Scholz S, Fung HC, et al. Conflicting results regarding the semaphorin gene (SEMA5A) and the risk for Parkinson disease. *Am J Hum Genet* 2006; **78:** 1082–84.
5 Farrer MJ, Haugarvoll K, Ross OA, et al. Genomewide association, Parkinson disease, and PARK10. *Am J Hum Genet* 2006; **78:** 1084–88.
6 Goris A, Williams-Gray CH, Foltynie T, Compston DA, Barker RA, Sawcer SJ. No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. *Am J Hum Genet* 2006; **78:** 1088–90.
7 Li Y, Rowland C, Schrodi S, et al. A case-control association study of the 12 single-nucleotide polymorphisms implicated in Parkinson disease by a recent genome scan. *Am J Hum Genet* 2006; **78:** 1090–92.
8 Myers RH. Considerations for genomewide association studies in Parkinson's disease. *Am J Hum Genet* 2006; **78:** 1081–82.
9 Maraganore DM, Andrade M, Lesnick TG, Pant PV, Cox DR, Ballinger DG. Response from Maraganore et al. *Am J Hum Genet* 2006; **78:** 1092–94.
10 Pals P, Van Everbroeck B, Grubben B, et al. Case-control study of environmental risk factors for Parkinson's disease in Belgium. *Eur J Epidemiol* 2003; **18:** 1133–42.
11 Engelborghs S, Dermaut B, Goeman J. Prospective Belgian study of neurodegenerative and vascular dementai: APOE genotype effects. *J Neurol Neurosurg Psychiatry* 2003; **74:** 1148–51.
12 Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997; **127:** 820–26.
13 Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. *Eval Health Prof* 2002; **25:** 12–37.
14 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21:** 1539–58.
15 Ioannidis JP, Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol* 2005; **58:** 543–49.
16 Ioannidis JP, Trikalinos TA, Khoury MJ. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am J Epidemiol* 2006; published online Aug 7. DOI:10.1093/aje/kwj259.
17 Risch N, Teng J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. *Genome Res* 1998; **8:** 1273–88.
18 Fung H-C, Scholz S, Matarin M, et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol* 2006; published online Sept 27. DOI: 10.1016/S1474-4422(06)70578-6.
19 Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. *Nat Genet* 2006; **38:** 659–62.
20 Haines JL, Hauser MA, Schmidt S, et al. Complement factor H variant increases the risk of age-related macular degeneration. *Science* 2005; **308:** 419–21.
21 Arking DE, Pfeufer A, Post W, et al. A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat Genet* 2006; **38:** 644–51.
22 Herbert A, Gerry NP, McQueen MB, et al. A common genetic variant is associated with adult and childhood obesity. *Science* 2006; **312:** 279–83.
23 Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet* 2006; **38:** 652–58.
24 Smyth DJ, Cooper JD, Bailey R, et al. A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat Genet* 2006; **38:** 617–19.
25 Tanner CM, Ottman R, Goldman SM, et al. Parkinson disease in twins: an etiologic study. *JAMA* 1999; **281:** 341–46.