Letters to the Editor

Am. J. Hum. Genet. 78:1081–1082, 2006

Considerations for Genomewide Association Studies in Parkinson Disease

To the Editor:

Although the magnitude of a genetic component of Parkinson disease (PD [MIM 168600]) remains to be determined, the disease has already shown remarkable genetic heterogeneity, with at least five monogenic forms identified, the most common of which is *LRRK2* (MIM 609007).[1] In this issue of *The American Journal of Human Genetics,* four investigative teams[2–5] report that they have sought to replicate the findings from a genomewide association (GWA) study of PD affection by Maraganore et al.[6] Taken together, these four studies appear to provide substantial evidence that none of the SNPs originally featured as potential PD loci are convincingly replicated and that all may be false positives. Furthermore, that the *LRRK2* gene was not identified may be considered a false-negative result. This conclusion is both disappointing and discouraging. The original study invested heavily in this venture, with 443 sibling pairs ($n = 886$) discordant for PD typed in tier 1 for 198,345 SNPs (172,420,019 genotype calls) and a tier 2 follow-up typing the strongest 1,892 SNPs in 332 matched case-control unrelated pairs (1,176,772 genotypes). Because this report is among the first GWA studies and because the effort appears to have failed to produce the desired objective, it is worth examining the implications for GWA studies in general and, specifically, the significance of this study for PD.

First, let's examine the original report. Tier 1 of the original study is founded upon sibling pairs discordant for PD recruited from the Mayo Clinic in Rochester, MN. The sample is composed of individuals substantially of northern and central European descent. Discordant sibling pairs were selected to limit false-positive results due to population stratification bias.[7] Population differences between case and control samples are recognized as the primary source of false-positive associations, and, clearly, every effort to minimize these effects is to be encouraged. However, in PD there is substantial evidence for reduced penetrance,[8] and the disease etiology is most likely a complex interaction of genetic and environmental factors.[9] Thus, the selection of randomly ascertained PD cases (often termed "sporadic") may include a substantial proportion of cases with little or no genetic basis for disease, and, even among familial cases, many unaffected siblings may carry PD risk alleles but remain unaffected for lack of critical environmental exposure, for essential modifying genes, or for follow-up to an advanced age. Case identification in tier 1 should focus on the selection of those most likely to carry the inherited form of the disease, whereas controls should be likely non–gene carriers drawn from the same population. Concerns for population stratification might best be addressed in tier 2 by the genotyping of families of tier 1 cases and by family-based association studies. SNPs showing association in these first phases can be typed in a second unrelated case-control sample as a tier 3, with case enrichment for familial disease when possible.

Fundamentally, scientific discovery relies first and foremost upon the independent replication of results. Investigators seeking to replicate the findings of association studies need to consider whether their sample provides an appropriate forum for the investigation. Because the overwhelming majority of SNPs in GWA studies will not be functionally related to the disease, one cannot reasonably expect that linkage-disequilibrium patterns will generalize across diverse ethnic groups. Thus, one may expect that there may not be replication for samples recruited from a restricted geographic region (e.g., Taiwan[2]). Whereas most of these replication samples are composed of Europeans (e.g., from Finland,[2] Norway and Ireland,[3] and the United Kingdom[4]), a few reveal minor-allele frequencies that vary from the original sample and that may deserve further study. Enrichment for familial PD would also be important, since none of these replication studies is described as familial PD.

Genomewide linkage studies have generally not been successful in finding genes responsible for common complex diseases, and whether GWA studies will prove to be more successful remains to be determined. There is at least one important positive precedent of the Maraganore et al.[6] study. Notably, all of their single-SNP association results (minor-allele frequencies and *P* values) are available in two online text files (available from http://www.journals.uchicago.edu/AJHG/journal/issues/

v77n5/42619/tableS2new.txt and http://www.journals.uchicago.edu/AJHG/journal/issues/v77n5/42619/tableS3new.txt) in the online-only version of the original article.[6] These results can be readily downloaded and searched for evidence of association with other interesting PD candidate genes. Maraganore and colleagues, with the Michael J. Fox Foundation, have the opportunity to establish a precedent for making the entire GWA study available online, since one may reasonably expect that true PD risk alleles may be found among the SNPs with lesser levels of statistical significance. The jury is still out on whether this GWA study holds important insights for PD.

RICHARD H. MYERS
Boston University School of Medicine
Boston

Web Resources

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/entrez/Omim/ (for PD and *LRRK2*)

References

1. Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, Kachergus J, Hulihan H, Uitti RJ, Calne DB, et al (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44:601–607
2. Clarimon J, Scholz S, Fung H-C, Hardy J, Eerola J, Hellström O, Chen C-M, Wu Y-R, Tienari PJ, Singleton A (2006) Conflicting results regarding the semaphorin gene (*SEMA5A*) and the risk for Parkinson disease. Am J Hum Genet 78:1082–1084 (in this issue)
3. Farrer MJ, Haugarvoll K, Ross OA, Stone JT, Milkovic NM, Cobb SA, Whittle AJ, Lincoln SJ, Hulihan MM, Heckman MG, White LR, Aasly JO, Gibson JM, Gosal D, Lynch T, Wszolek ZK, Uitti RJ, Toft M (2006) Genomewide association, Parkinson disease, and *PARK10*. Am J Hum Genet 78:1084–1088 (in this issue)
4. Goris A, Williams-Gray CH, Foltynie T, Compston DAS, Barker RA, Sawcer SJ (2006) No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. Am J Hum Genet 78:1088–1090 (in this issue)
5. Li Y, Rowland C, Schrodi S, Laird W, Tacey K, Ross D, Leong D, Catanese J, Sninsky J, Grupe A (2006) A case-control association study of the 12 single-nucleotide polymorphisms implicated in Parkinson disease by a recent genome scan. Am J Hum Genet 78:1090–1092 (in this issue)
6. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PVK, Frazer KA, Cox DR, Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 77:685–693
7. Maraganore DM (2005) Blood is thicker than water: the strengths of family-based case-control studies. Neurology 64:408–409
8. Maher NE, Currie LJ, Lazzarini AM, Wilk JB, Taylor CA, Saint-Hilaire MH, Feldman RG, Golbe LI, Wooten GF, Myers RH (2002) A segregation analysis of Parkinson disease revealing evidence for a major causative gene. Am J Med Genet 109:191–197
9. Dekker MC, Bonifati V, van Duijn CM (2003) Parkinson's disease: piecing together a genetic jigsaw. Brain 126:1722–1733

Address for correspondence and reprints: Dr. Richard H. Myers, Department of Neurology, E-304, Boston University School of Medicine, 715 Albany Street, Boston, MA 02118. E-mail: rmyers@bu.edu

0002-9297/2006/7806-0019$15.00

Am. J. Hum. Genet. 78:1082–1084, 2006

Conflicting Results Regarding the Semaphorin Gene (*SEMA5A*) and the Risk for Parkinson Disease

To the Editor:

The strongest variant (*rs7702187*) associated with Parkinson disease (PD [MIM 168600]) reported in the whole-genome association study by Maraganore et al.[1] was evaluated in two independent case-control series of patients from Finland and Taiwan, as were four other variants located within *SEMA5A* (MIM 609297). The Finnish series comprised 146 patients with sporadic PD (mean age 67.2 years, range 38–88 years; 41% women) and 135 neurologically normal, healthy control subjects (mean age 65.8 years, range 37–80 years; 64% women). All individuals were recruited from the neurological outpatient clinics of the Helsinki University Central Hospital and Seinäjoki Central Hospital. The Taiwanese series consisted of 303 patients with sporadic PD (mean age 61.9 years, range 24–91 years; 46.2% women) and 171 control individuals (mean age 60.1 years, range 31–86 years; 43.9% women). Patients were selected from the neurological clinic of Chang-Gung Memorial Hospital. Individuals with evidence of secondary parkinsonism or with atypical features such as early dementia, ophthalmoplegia, early autonomic failure, and pyramidal signs were not included in this study. All patients included in the study fulfilled PD diagnosis criteria.[2] All participants signed an informed consent form.

Taqman Assays-by-Design SNP Genotyping Assays (Applied Biosystems) were employed for allelic discrimination of all SNPs. Differences in allele and genotype distributions were analyzed using the χ^2 test, and two-tailed *P* values are presented. Haplotype frequency comparisons between cases and controls were performed with PHASE version 2.1 software.[3] One thousand permutations were performed for each comparison. The COCAPHASE module of the UNPHASED statistical package was used for linkage-disequilibrium (LD) analyses.[4] Power calculations were performed with PS version 2.1.30.[5]

Allele and genotype frequency information for each of the markers is shown in table 1. None of the markers showed any significant association with disease in the Finnish series. However, we were able to replicate the

Table 1

Genotype and Allele Frequency Distribution of the Polymorphisms Analyzed across *SEMA5A* on Chromosome 5

dbSNP Accession Number	Position	Genotype Frequency				Minor-Allele Frequency		OR	*P* (95% CI)[a]
		Control 11	Control 12	Case 11	Case 12	Control	Case		
Finnish series:									
rs3798097	9595529	.49	.35	.51	.35	.34	.32	.910	.610 (.63–1.31)
rs368226	9470056	.90	.10	.91	.09	.05	.05	.921	.838 (.42–2.02)
rs7702187	9385281	.69	.29	.74	.22	.16	.15	.901	.657 (.57–1.43)
rs1806151	9207659	.25	.50	.27	.53	.50	.47	1.160	.424 (.81–1.67)
rs786843	9093141	.68	.29	.66	.31	.18	.18	1.055	.814 (.67–1.65)
Taiwanese series:									
rs3798097	9595529	.81	.16	.88	.10	.11	.07	.586	**.025 (.37–.94)**
rs368226	9470056	.51	.37	.49	.42	.30	.30	.995	.976 (.74–1.33)
rs7702187	9385281	.61	.35	.49	.44	.22	.30	1.534	**.007 (1.12–2.10)**
rs1806151	9207659	.62	.34	.62	.35	.21	.22	.958	.805 (.68–1.35)
rs786843	9093141	.89	.10	.88	.12	.06	.06	1.057	.846 (.61–1.65)

[a] Values in bold denote statistical significance.

reported association with marker *rs7702187* in the Taiwanese cohort (odds ratio [OR] = 1.53, 95% CI 1.12–2.10, $P = .007$). Genotype analysis showed that individuals homozygous for the A allele had a significantly decreased risk of PD compared with those heterozygous or homozygous for the T allele (OR = 0.60, 95% CI 0.41–0.88, $P = .009$). A significant association was also found for the *rs3798097* marker, which is located in the 5′ UTR region of *SEMA5A* (OR for the C allele was 1.71, 95% CI 1.06–2.73, $P = .025$).

Both populations showed a complete lack of LD for any pairs of neighboring polymorphisms (all D' values were <0.5, independently of diagnostic group). Haplotype frequency comparisons did not reveal any significant differences between patients and controls in the Finnish series ($P = .901$) or between patients and controls in the Taiwanese series ($P = .091$) (table 2).

The present results point to differential risk effects of *SEMA5A* marker alleles across populations. In the Taiwanese population, we have found an associated risk in the same locus as the one reported elsewhere[1] but in an opposite direction. That is, the at-risk allele that we report was found to be protective in the sample from Minnesota described by Maraganore et al.[1] This could be due to the effect of LD between this polymorphism and another "true" risk variant within the gene. The lack of association shown in the Finnish population could be related to genetic heterogeneity, or, alternatively, the Finnish series might not be large enough to assess genes with modest effects (this sample has a 60% power to detect risks of 1.7, at $\alpha = 0.05$).

The replication of an association with *SEMA5A* in a Taiwanese population makes it a good candidate for further analyses in different populations.

Acknowledgments

We are indebted to the patients and their families for their generous participation in this work. This work was supported in part by the Helsinki University Central Hospital, the Finnish Cultural Foundation, the Finnish National Graduate School of Clinical Investigation, and the intramural program of the National Institute on Aging, National Institutes of Health, Bethesda.

Table 2

Haplotype Frequency Distribution in Finnish and Taiwanese Series

Haplotype	Frequency			
	Finnish[a]		Taiwanese[b]	
	Control	Cases	Control	Cases
CCACC	.252	.267	.391	.339
CCAGC	.212	.203	.069	.081
TCAGC	.104	.096	.010	.005
TCACC	.103	.109	.049	.024
CGACG	.010	.013	.164	.163
CCTCC	.038	.046	.105	.162
Other[c]	.281	.267	.097	.227

Note.—The order of SNPs is *rs3798097, rs368226, rs7702187, rs1806151,* and *rs786843.*

[a] Global significance for haplotype frequency differences: $P = .9$

[b] Global significance for haplotype frequency differences: $P = .091$

[c] Other haplotypes with frequencies <5%.

Jordi Clarimon,[1] Sonja Scholz,[1] Hon-Chung Fung,[1,2] John Hardy,[1] Johanna Eerola,[3,4] Olli Hellström,[5] Chiung-Mei Chen,[2] Yih-Ru Wu,[2] Pentti J. Tienari,[3,4] and Andrew Singleton[1]
[1]*Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda;* [2]*Department of Neurology, Chang Gung Memorial Hospital and College of Medicine, Chang Gung University, Taipei, Taiwan;* [3]*Department of Neurology, Helsinki University Central Hospital, and* [4]*Neuroscience Programme, University of Helsinki, Biomedicum-Helsinki, Helsinki; and* [5]*Department of Neurology, Seinäjoki Central Hospital, Seinäjoki, Finland*

Web Resources

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=snp (for *SEMA5A* markers)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/entrez/Omim/ (for PD and *SEMA5A*)

References

1. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PVK, Frazer KA, Cox DR, Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 77:685–693
2. Hughes AJ, Daniel SE, Kilford L, Lees AJ (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 55:181–184
3. Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162–1169
4. Dudbridge F (2003) Pedigree disequilibrium tests for multilocus haplotypes. Genet Epidemiol 25:115–121
5. Dupont WD, Plummer WD Jr (1998) Power and sample size calculations for studies involving linear regression. Control Clin Trials 19:589–601

Address for correspondence and reprints: Dr. Jordi Clarimon, Molecular Genetics Section, Laboratory of Neurogenetics, National Institutes of Health, Building 35, Room 1A1000, MSC 3707, 35 Lincoln Drive, Bethesda, MD 20892. E-mail: jclarimon@santpau.es

0002-9297/2006/7806-0020$15.00

Am. J. Hum. Genet. 78:1084–1088, 2006

Genomewide Association, Parkinson Disease, and *PARK10*

To the Editor:

Genomewide linkage analysis of rare familial forms of parkinsonism has identified mutations in seven genes, revealing a clinicopathologically and genetically heterogeneous syndrome.[1] Less progress has been made in the more typical late-onset form of Parkinson disease (PD [MIM 168600]), although the recently identified *LRRK2* (MIM 609007) G2019S substitution is estimated to account for ~1% of sporadic PD cases.[2] Common polymorphisms of familial genes may also influence susceptibility to idiopathic PD.[3,4] Of the 198,345 SNPs successfully genotyped in the recent genomewide association (GWA) study, 26 had notably different allele frequencies between patients and controls in both tiers ($P < .01$).[5] Fifteen of these SNPs had opposite directions of effect (disease risk or protection) in tiers 1 and 2. The remaining 11 SNPs were proposed as markers for new genes/chromosomal loci that influence susceptibility to PD. In addition, two SNPs in tier 2 (*rs682705* and *rs7520966*) were highlighted in the *PARK10* locus (MIM 606852), which nominated the gene *LOC200008* in disease susceptibility.

The *PARK10* locus on chromosome 1p32 was originally identified in a genomewide linkage analysis of 117 patients from 51 Icelandic families (maximum Z_{lr} = 4.8 at *D1S231,* with a LOD-1, 7.6-cM support interval from *D1S2874* to *D1S475*).[6] Iceland has a well-characterized genealogy that is powerful for family-based linkage studies. The ancestral founders of Iceland have Scandinavian patrilineal inheritance with a minor Celtic matrilineal component.[7] Assuming that the *PARK10* mutation predates the Icelandic settlement, we reasoned that the 1p32 susceptibility gene might be more readily found in patients with PD originating from Scandinavian or Celtic populations. In parallel to the study of Maraganore et al.,[5] we have been mapping the *PARK10* locus. Genotypes from 28 SNPs (including *rs682705* and *rs7520966)* within a 132-kb region of chromosome 1p32 located around the *LOC200008* gene have been analyzed in two well-characterized case-control series from Norway and Ireland. In addition, we attempted to replicate findings for the two *PARK10* SNPs in a U.S. series collected at the Mayo Clinic in Jacksonville, FL. We then employed all three case-control series to investigate the genotype/allele frequencies of the main 11 SNPs nominated to influence PD susceptibility.[5] Power was comparable to the original study (>80% at α = 0.05 for odds ratios [ORs] >2.0 and for disease-allele frequencies >0.035), and genotyping call rates were >95% for all markers (table 1).

In total, Norwegian samples included 676 subjects (cases and controls) with a mean age (±SD) of 70 ± 11 years, Irish samples included 372 subjects with a mean age (±SD) of 61 ± 13 years, and the U.S. samples included 522 subjects with a mean age (±SD) of 71 ± 10 years. All patients were examined and were observed longitudinally by a movement-disorders neurologist (J.O.A., J.M.G., D.G., T.L., Z.K.W., and R.J.U.), and they were

Table 1

Genotype/Allele Frequencies of the Main 11 SNPs Nominated to Influence PD Susceptibility

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

given a diagnosis of PD in accordance with published criteria.[8] Each patient was individually matched, on the basis of age (±4 years) and ethnicity, to an unrelated control without evidence of neurological disease. The ethical review boards at each institution involved approved the study, and all participants provided informed consent.

SNP genotyping was performed using TaqMan chemistry on an ABI7900 genetic analyzer; in cases where genotype data was available for only one subject of a matched pair, the other subject was retained in the analysis. For the controls in each population, χ^2 tests of Hardy-Weinberg equilibrium (HWE) were implemented using Haploview.[9] Optimal SNP coverage for association analysis of the *LOC200008* gene was determined empirically by the construction of linkage-disequilibrium (LD) maps in Norwegian and Irish samples, onto which haplotype blocks were assigned (fig. 1).[10,11] ORs for disease association, with corresponding 95% CIs, were subsequently calculated using logistic-regression models adjusted for age and sex. Overall ORs combining data from all three sites were additionally adjusted for site. Previous studies have nominated the *PARK10* locus as an age-at-onset modifier in PD[12]; thus, we also assessed the influence of 1p32 SNPs variability on this disease trait, using linear-regression models adjusted for sex.

There was no evidence of association with PD for any of the 28 genotyped 1p32 SNPs in our study (all SNP $P > .05$ after applying Bonferroni correction in both population samples). Haplotype frequencies between patients and controls were not significantly different for the haplotype blocks identified; nor was the age at onset in patients associated with any single marker or haplotype (all corrected $P > .05$). Of note, the ancestral recombination and haplotype blocks apparent within Norwegian and Irish samples were comparable for this interval at this marker resolution. The average number of SNPs per LD unit (LDU) was 6.8 (mean LDU between markers 0.15, range 0–0.63), indicating that the number of SNPs genotyped within and flanking *LOC200008* should be sufficient for examination of the region.[11] In addition, the two *PARK10* SNPs showed no significant association within the U.S. series ($P > .05$). None of the other 11 SNPs nominated by the GWA study had different allele frequencies or genotype distributions between affected subjects and matched controls (all SNP $P > .05$ in all populations independently or as a combined sample set) (table 2). There was no evidence of departure from HWE in controls ($P > .01$ in all population controls).

Our study indicates that genetic variability within the *LOC200008* gene is unlikely to explain the *PARK10* susceptibility locus for PD. Sadly, the lack of disease association and replication in an independent U.S. series of comparable power suggests that the original findings may be spurious. Failure to nominate *LOC200008* as the *PARK10* gene in our population samples provides empirical support for statistical caveats concerning GWA studies. Implicit in multiple testing is false discovery, even in well-designed studies, and there are several potential sources of bias.[13] Of note, neither *PARK10* SNP *rs682705* nor *rs7520966* fulfilled the main criterion for being genotyped in tier 2 ($P < .01$ in tier 1 overall analysis), but each was included with a less stringent association criterion ($P < .05$ in tier 1 overall analysis) because of its physical position within a *PARK* locus. Interestingly, the combined P value for *rs682705* ($P = 9.07 \times 10^{-6}$) is the second-lowest P value of the overall study, even though it did not fulfill the inclusion criteria. Individual-level data from the GWA study is not yet available, but, in our study, these two SNPs also appear to be in LD (pairwise $r^2 > 0.9$), as suggested by Maraganore et al.[5]; in addition, the minor-allele frequencies (MAFs) of the two SNPs are comparable across studies and populations. The former suggests less-than-optimal haplotype tagging in the initial study, whereas the latter argues against technical errors in genotyping, but neither provides sufficient explanation for the positive findings observed elsewhere.[5]

We found no evidence of direct association between the 11 SNPs nominated in the GWA study and disease in the three independent populations or in a combined sample group ($n = 1{,}570$) (table 2). However, for these loci, we did not employ a gene-based approach (nor did we fine-map each region as with *PARK10*), as advocated elsewhere[14]; we await the results of further replication studies. Of note, in the study by Maraganore et al.,[5] the *rs7702187* SNP within *SEMA5A* (MIM 609297) had the lowest combined P value ($P = 7.62 \times 10^{-6}$); however, a total of 53 SNPs were examined in this gene in tier 1. Only *rs7702187* was significant before correction ($P = .001$), which supports the possibly spurious nature of this and the other associations. The MAFs observed in our three populations and in that of the GWA study are comparable, which argues against population bias/heterogeneity (table 2).

The number of SNPs highlighted in each tier of the original study is consistent with chance—that is, 1% of SNPs use a significance level of $P < .01$. None of the P values obtained by Maraganore et al.[5] meets a Bonferroni correction for multiple testing, although this standard may be too conservative in GWA, since it fails to account for LD and incorrectly assumes that chromo-

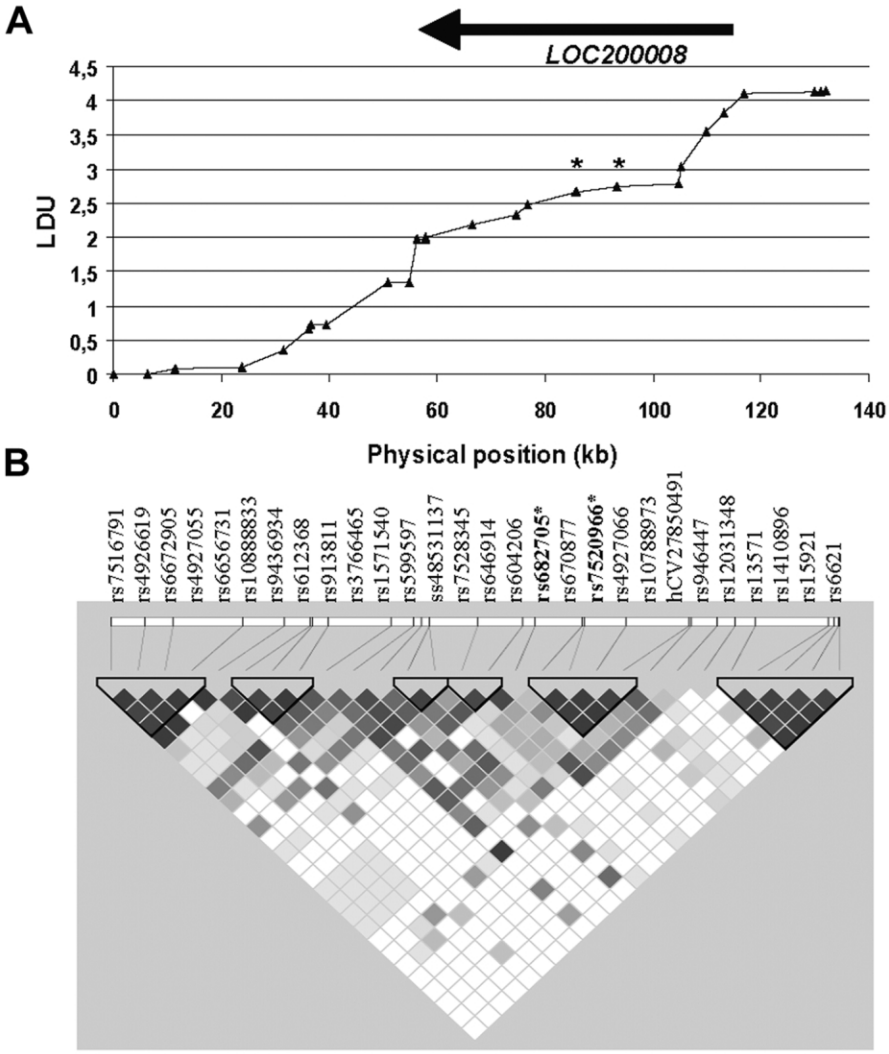

Figure 1 Metric LD map and haplotype block structure of the investigated region. *A,* LD map providing information about LD patterns in the investigated candidate region, through locations expressed in LDUs. LDUs have an inverse relationship with LD, with regions of extensive recombination having many LDUs. The physical position of the gene in the region *LOC200008* is marked with an arrow. All 28 SNPs genotyped are reported, although the symbols (▲) may be obscured for SNPs that lie in close physical proximity and high LD. SNPs *rs682705* and *rs7520966* are denoted by an asterisk (*). *B,* LD structure of the candidate region. *Black* and *dark gray cells,* strong LD; *gray cells,* intermediate; and *light gray* and *white cells,* evidence for historical recombination. The haplotype block structure of the region is defined according to Gabriel et al.[10] An asterisk denotes SNPs *rs682705* and *rs7520966.* The LD map and haplotype structure were constructed using genotypes from the Norwegian sample. Similar results were obtained for the Irish population.

somal markers are independent. A consensus on the most appropriate correction for multiple testing has yet to be reached. Now that genomewide data sets have been generated, there exists the possibility to use these to develop appropriate statistical methods to identify true positive results.[15]

In the interim, we recommend that enthusiasm for positive findings should be tempered by the strength of the evidence, the population-attributable risk, and the differences in SNP allele/genotype frequencies between cases and controls. If allele frequencies are significantly different, genomic controls might be used to assess population substructure. It is important that future studies employ multiple independent sample series, each with sufficient power to verify significant genetic associations, before publication.[16] However, lack of evidence for an association is not the same as evidence against one; thus, lack of replication should also be interpreted with caution.

Over the few next years, the number of GWA studies

Table 2

11 SNPs Nominated in GWA Study as Genetic Susceptibility Loci for PD

dbSNP Accession Number	Chromosome	Position (bp)	This Study: Control MAF Ireland	Control MAF Norway	Control MAF United States	Estimated OR (95% CI)[a]	Combined *P* (*n* = 1,570)	Maraganore et al.[5]: Control MAF Tier 1	Control MAF Tier 2	*P* Tier 1	*P* Tier 2	Estimated OR (95% CI)[b]	Combined *P* (*n* = 1,550)
rs7702187	5p15.2	9385281	.17	.18	.18	.88 (.74–1.06)	.18	.18	.20	.001	.002	1.74 (1.36–2.24)	7.62×10^{-6}
rs10200894	2q36	228642637	.13	.11	.09	.96 (.77–1.21)	.74	.12	.13	.009	.001	1.84 (1.38–2.45)	1.70×10^{-5}
rs2313982	4q31.1	139145665	.05	.11	.09	.93 (.73–1.18)	.54	.07	.06	.006	.002	2.01 (1.44–2.79)	1.79×10^{-5}
rs17329669	7p14	36625169	.13	.12	.14	1.01 (.82–1.24)	.92	.13	.11	.008	.001	1.71 (1.33–2.21)	2.30×10^{-5}
rs7723605	5p15.3	5407615	.13	.14	.13	.91 (.75–1.12)	.38	.12	.09	.010	.002	1.78 (1.35–2.35)	3.30×10^{-5}
ss46548856	10q21	58986929	.09	.08	.11	.93 (.73–1.19)	.58	.09	.11	.003	.002	1.88 (1.38–2.57)	3.65×10^{-5}
rs16851009	2q24	166456214	.11	.11	.12	.95 (.76–1.18)	.63	.09	.08	.002	.009	1.84 (1.36–2.49)	4.17×10^{-5}
rs2245218	1p36.2	13885132	.19	.17	.15	.95 (.79–1.14)	.57	.11	.13	.002	.002	1.67 (1.29–2.14)	4.61×10^{-5}
rs7878232	Xq28	150516943	.25	.23	.25	1.10 (.97–1.25)	.15	.29	.26	.003	.010	1.38 (1.17–1.62)	6.87×10^{-5}
rs1509269	4q31.1	139111329	.08	.13	.13	.94 (.76–1.17)	.58	.10	.09	.005	.008	1.71 (1.30–2.26)	9.21×10^{-5}
rs11737074	4q27	125438978	.21	.20	.21	1.05 (.89–1.25)	.55	.19	.19	.007	.005	1.50 (1.21–1.86)	1.55×10^{-4}

NOTE.—In this study, MAFs are not significantly different between the populations. No *P* values are corrected for multiple testing. SNPs are ordered by combined *P* value, per Maraganore et al.[5]

[a] The direction of effect of the estimated OR observed in this study for each SNP is shown (i.e., >1 risk and <1 protective).

[b] Estimated ORs in the study by Maraganore et al.[5] do not indicate the direction of effect relative to the MAF.

will increase, and it is important to learn from the experiences gained by the few studies performed to date. Although our negative findings suggest that the conclusions drawn from the study by Maraganore et al.[5] might be based on spurious associations, further analysis of individual-level raw data is now necessary. The recent identification of a complement factor H polymorphism in age-related macular degeneration in a GWA study and the identical findings by two other groups using other study designs demonstrates that this approach can be used successfully.[17–19] It may be that, because of the heterogeneous nature of PD, associations with a gestalt phenotype are masked by background variation in SNP informativeness, population strata, and insufficient power. It is, therefore, crucial that future associations are validated and that analysis is performed to resolve the underlying cause of association in the sample population. GWA studies may still provide direction for the genetic analysis of heterogenous complex traits, but, in the short term, they may exacerbate the problem of replication failure in association studies.

Acknowledgments

The authors thank the patients and families for their participation in this study. Mayo Clinic Jacksonville is an M. K. Udall Parkinson's Disease Research Center of Excellence (National Institute of Neurological Disorders and Stroke grant P01 NS40256), and the authors thank all collaborators at the Udall Center. This study was also supported by the National Institutes of Health grant R01 NS36960, the Research Council of Norway grant 153487/V50, and Reberg's Legacy. The Ireland research consortium was supported by a Programme for Research in Third-Level Institutions neurosciences award and by the Research and Development Office of the Health and Personal Social Services. O.A.R. and M.T. are partly funded by National Parkinson Foundation and Parkinson's Disease Foundation fellowships, respectively. We thank Minnie Schreiber for technical assistance.

MATTHEW J. FARRER,[1] KRISTOFFER HAUGARVOLL,[1,2] OWEN A. ROSS,[1] JEREMY T. STONE,[1] NICOLE M. MILKOVIC,[1] STEPHANIE A. COBB,[1] ANDREW J. WHITTLE,[1] SARAH J. LINCOLN,[1] MARY M. HULIHAN,[1] MICHAEL G. HECKMAN,[1] LINDA R. WHITE,[2] JAN O. AASLY,[3] J. MARK GIBSON,[4] DAVID GOSAL,[5,6] TIMOTHY LYNCH,[5,6] ZBIGNIEW K. WSZOLEK,[1] RYAN J. UITTI,[1] AND MATHIAS TOFT[1,2]

[1]*Departments of Neuroscience and Neurology, Mayo Clinic College of Medicine, Jacksonville, FL;* [2]*Department of Neuroscience, Norwegian University of Science and Technology, and* [3]*Department of Neurology, St. Olav's Hospital, Trondheim, Norway;* [4]*Department of Neurology, Royal Victoria Hospital, Belfast, Ireland;* [5]*Department of Neurology, Mater Misericordiae Hospital, and* [6]*The Conway Institute, University College, Dublin, Ireland*

Web Resources

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=snp

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for PD, *LRRK2, PARK10,* and *SEMA5A*)

References

1. Ross OA, Farrer MJ (2005) Pathophysiology, pleiotrophy and paradigm shifts: genetic lessons from Parkinson's disease. Biochem Soc Trans 33:586–590
2. Kachergus J, Mata IF, Hulihan M, Taylor JP, Lincoln S, Aasly J, Gibson JM, Ross OA, Lynch T, Wiley J, Payami H, Nutt J, Maraganore DM, Czyzewski K, Styczynska M, Wszolek ZK, Farrer

MJ, Toft M (2005) Identification of a novel *LRRK2* mutation linked to autosomal dominant parkinsonism: evidence of a common founder across European populations. Am J Hum Genet 76: 672–680
3. Skipper L, Wilkes K, Toft M, Baker M, Lincoln S, Hulihan M, Ross OA, Hutton M, Aasly J, Farrer M (2004) Linkage disequilibrium and association of *MAPT* H1 in Parkinson disease. Am J Hum Genet 75:669–677
4. Pals P, Lincoln S, Manning J, Heckman M, Skipper L, Hulihan M, Van den Broeck M, De Pooter T, Cras P, Crook J, Van Broeckhoven C, Farrer MJ (2004) α-Synuclein promoter confers susceptibility to Parkinson's disease. Ann Neurol 56:591–595
5. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PVK, Frazer KA, Cox DR, Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 77:685–693
6. Hicks AA, Petursson H, Jonsson T, Stefansson H, Johannsdottir HS, Sainz J, Frigge ML, Kong A, Gulcher JR, Stefansson K, Sveinbjornsdottir S (2002) A susceptibility gene for late-onset idiopathic Parkinson's disease. Ann Neurol 52:549–555
7. Helgason A, Sigurethardottir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, Gulcher JR, Ward R, Stefansson K (2000) Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. Am J Hum Genet 67:697–717
8. Gelb DJ, Oliver E, Gilman S (1999) Diagnostic criteria for Parkinson disease. Arch Neurol 56:33–39
9. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265
10. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229
11. Tapper WJ, Maniatis N, Morton NE, Collins A (2003) A metric linkage disequilibrium map of a human chromosome. Ann Hum Genet 67:487–494
12. Li YJ, Scott WK, Hedges DJ, Zhang F, Gaskell PC, Nance MA, Watts RL, et al (2002) Age at onset in two common neurodegenerative diseases is genetically controlled. Am J Hum Genet 70: 985–993
13. Wang WY, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6:109–118
14. Neale BM, Sham PC (2004) The future of association studies: gene-based analysis and replication. Am J Hum Genet 75:353–362
15. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108
16. Hattersley AT, McCarthy MI (2005) What makes a good genetic association study? Lancet 366:1315–1323
17. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh, J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385–389
18. Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, Spencer KL, Kwan SY, Noureddine M, Gilbert JR, Schnetz-Boutaud N, Agarwal A, Postel EA, Pericak-Vance MA (2005) Complement factor H variant increases the risk of age-related macular degeneration. Science 308:419–421
19. Edwards AO 3rd, Ritter R, Abel KJ, Manning A, Panhuysen C, Farrer LA (2005) Complement factor H polymorphism and age-related macular degeneration. Science 308:421–424

Address for correspondence and reprints: Dr. Owen A. Ross, Department of Neuroscience, Mayo Clinic College of Medicine, Birdsall Building, 4500 San Pablo Road, Jacksonville, FL 32224. E-mail: ross.owen@mayo.edu

0002-9297/2006/7806-0021$15.00

Am. J. Hum. Genet. 78:1088–1090, 2006

No Evidence for Association with Parkinson Disease for 13 Single-Nucleotide Polymorphisms Identified by Whole-Genome Association Screening

To the Editor:

The 13 SNPs identified by Maraganore et al.[1] as being potentially associated with Parkinson disease (PD [MIM 168600]) represent some of the first fruit produced by the whole-genome association screening era and are clearly worthy of follow-up. To further explore these exciting candidates, we typed each SNP in 538 patients with idiopathic PD and in 516 control individuals from the United Kingdom. Cases included 160 patients involved in a community-based epidemiological study of incident PD and 378 consecutive patients with prevalent PD attending our research clinic. All cases met United Kingdom Parkinson's Disease Society Brain Bank criteria for the diagnosis of PD. The mean age at disease onset was 63 years (range 25–91 years); 2% of patients had early-onset disease ($\leqslant$40 years), and 14% of patients reported a family history of one or more first-degree relatives with parkinsonian symptoms or tremor. The control group consisted of 146 spouses of patients with PD and 370 blood donors. All individuals were white, except for four patients and one spouse. All gave written informed consent and a blood sample from which DNA was extracted using standard methods. Genotyping was performed using Taqman Assay-on-Demand (*rs2245218*) and Assays-by-Design products on a 7900HT Sequence Detection System (Applied Biosystems). Only samples that typed successfully for at least one-third of markers were included in the analysis (520 cases and 499 controls). Genotyping success rates were all $\geqslant$97%, and no marker showed evidence of deviation from Hardy-Weinberg equilibrium. Two pairs of SNPs (*rs2313982* and *rs1509269*; *rs682705* and *rs7520966*) were found to be in strong linkage disequilibrium ($D' = 1.0$, $r^2 > 0.69$), which reduced the number of independent tests to 11. Allele frequencies in cases and controls were compared using the COCAPHASE program in the UNPHASED package.[2] Our study provides, on average, 85% power (range 68%–96%) to detect the case-control differences averaged over tier 1 and tier 2, as observed by Maraganore et al.[1]

Table 1

Thirteen SNPs Reported by Maraganore et al.,[1] Ranked in Accordance with Evidence for Association in a Meta-Analysis Combined with Data from This Study

dbSNP Accession Number	Gene	Chromosome	Position	Control MAF[a]	Case MAF	OR (95% CI)	*P*	
							This Study[b]	Meta-Analysis[c]
rs10200894	...	2q36	228642637	.09	.08	.91 (.67–1.24)	.53	.01
ss46548856[d]	...	10q21	58986929	.10	.09	.92 (.68–1.24)	.58	.02
rs7702187	*SEMA5A*	5p15	9385281	.16	.16	.97 (.76–1.23)	.81	.02
rs17329669	...	7p14	36625169	.13	.13	1.04 (.80–1.35)	.79	.06
rs7723605	...	5p15	5407615	.13	.14	1.07 (.83–1.39)	.59	.06
rs7878232	*PASD1*	Xq28	150516943	.23	.23	.99 (.78–1.26)	.95	.11
rs682705	*LOC200008*	1p32	54349438	.26	.28	1.08 (.89–1.31)	.44	.20
rs7520966	*LOC200008*	1p32	54357283	.26	.28	1.07 (.88–1.30)	.51	.22
rs2245218	*PRDM2*	1p36	13885132	.16	.14	.89 (.70–1.14)	.36	.28
rs2313982	...	4q31	139145665	.09	.08	.83 (.61–1.14)	.26	.33
rs1509269	...	4q31	139111329	.12	.12	.92 (.70–1.20)	.53	.41
rs11737074	...	4q27	125438978	.23	.21	.90 (.73–1.11)	.32	.86
rs16851009	*GALNT3*	2q24	166456214	.10	.09	.86 (.64–1.16)	.33	.94

[a] Minor-allele frequency.

[b] *P* value for comparison of case and control allele frequencies with the use of UNPHASED.[2]

[c] *P* value corresponding to Mantel-Haenszel test statistic for association, with data from this study and that from Maraganore et al.,[1] after correction for the number of independent tests.

[d] Perlegen Sciences internal SNP identifier, as used by Maraganore et al.[1]

In our data set, none of the 13 SNPs showed any evidence of association, all *P* values being >.25, even without correction for multiple testing (tables 1 and 2). Fewer than half of the SNPs (46%) showed allele frequency differences between cases and controls in the same direction as that reported by Maraganore et al.[1] The combination of our data with those from the original report, with the use of the Mantel-Haenszel test statistic (Statsdirect) and correction for the 11 independent tests performed, revealed that only three markers (*rs10200894, ss46548856,* and *rs7702187*) retain any evidence of significance at the 5% level in the total data (table 1). In summary, our study suggests that none of the 13 markers identified by Maraganore et al.[1] is associated with PD.

Under the null hypothesis that there are no genes influencing susceptibility to PD, a follow-up of 1.4% (2,734) of the 198,345 markers included in the screening stage, as performed by Maraganore et al.,[1] would be expected to identify 27–28 markers showing $P < .01$ in the replication stage, with half of these—that is, 13–14—showing an allele frequency difference in the same direction as that seen in the screening stage. The number of markers identified by Maraganore et al.[1] is, thus, in keeping with that expected under the null hypothesis. However, since such screens are not intended to identify all susceptibility genes and, indeed, would be considered successful if they identified even a single such locus, we would not expect to see a striking excess of markers above the predicted 13. In short, it could be anticipated that most of the 13 markers identified by Maraganore et al.[1] would be false positives. However, our failure to replicate results for any of the 13 markers identified by Maraganore et al.[1] suggests that their screen lacked power in one or more critical dimensions. Although typing 200,000 markers in 450 cases and controls is a substantial effort, it is clear that this will adequately interrogate only a part of the common variation in the genome. Increasing the density of markers and the number of samples studied would be the most effective way to increase the power of the study but, in practice, would be the most difficult. It must remain possible that a more generous threshold (such as $P < .1$) would have captured relevant loci currently lying high in the ranking of markers provided by the screening stage performed by Maraganore et al.[1] but falling outside their stringent threshold. On the downside, this approach would greatly increase the number of markers requiring follow-up, generating a list of nearly 1,000 instead of just 13 potentially associated loci.

Various strategies for multistage whole-genome association studies have been proposed,[3–6] and the importance of setting an appropriate threshold for following up first-stage results has been stressed. We feel that the present observations, regarding one of the first whole-genome association screens performed, strengthen the importance of these theoretical recommendations. To ensure that replication and follow-up phases are not

Table 2

Genotype Counts for 13 SNPs Studied

The table is available in its entirety in the online edition of *The American Journal of Human Genetics.*

overwhelmingly large, it is essential to ensure high power in the screening phase. If thresholds as stringent as $P < .01$ are to be used, the screening phase in future PD screens will need to be very much larger than that performed by Maraganore et al.[1]

Acknowledgments

This work was supported by the Medical Research Council and the Parkinson's Disease Society. A.G. is a postdoctoral fellow of the Research Foundation of Flanders (FWO Vlaanderen). C.H.W.G. is a Patrick Berthoud clinical research fellow and holds a Raymond and Beverley Sackler scholarship.

A. Goris,[1,3] C. H. Williams-Gray,[2] T. Foltynie,[2] D. A. S. Compston,[1] R. A. Barker,[2] and S. J. Sawcer[1]

[1]*Neurology unit, Department of Clinical Neurosciences, and* [2]*Cambridge Centre for Brain Repair, Department of Clinical Neurosciences, University of Cambridge, United Kingdom; and* [3]*Laboratory of Neuroimmunology, Section of Experimental Neurology, Katholieke Universiteit Leuven, Leuven, Belgium*

Web Resources

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=snp

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for PD)

References

1. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant KPV, Frazer KA, Cox DR, Dennis G, Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 77:685–693
2. Dudbridge F (2003) Pedigree disequilibrium tests for multilocus haplotypes. Genet Epidemiol 25:115–221
3. Wang WYS, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6:109–118
4. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108
5. Lowe CE, Cooper JD, Chapman JM, Barratt BJ, Twells RCJ, Green EA, Savage DA, Guja C, Ionescu-Tirgoviste C, Tuomilehto-Wolf E, Tuomilehto J, Todd JA, Clayton DG (2004) Cost-effective analysis of candidate genes using htSNPs: a staged approach. Genes Immun 5:301–305
6. van den Oord EJCG, Sullivan PF (2003) False discoveries and models for gene discovery. Trends Genet 19:537–542

Address for correspondence and reprints: Dr. An Goris, Neurology unit, Department of Clinical Neurosciences, Box 165, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, United Kingdom. E-mail: ag441@medschl.cam.ac.uk

0002-9297/2006/7806-0022$15.00

Am. J. Hum. Genet. 78:1090–1092, 2006

A Case-Control Association Study of the 12 Single-Nucleotide Polymorphisms Implicated in Parkinson Disease by a Recent Genome Scan

To the Editor:

To validate associations of SNPs that Maraganore et al.[1] reported as associated with Parkinson disease (PD [MIM 168600]), we constructed a case-control series from PD cases and matched population/convenience controls that are available through the National Institute of Neurological Disorders and Stroke (NINDS) Human Genetics Resources at the Coriell Institute. Cases met United Kingdom Brain Bank criteria for idiopathic PD,[2] and controls were neurologically normal. This series comprises 311 pairs of age- and sex-matched cases and controls. Cases had an age at disease onset ranging from 50 to 87 years (average [±SD] 63.8 ± 8.9 years) and were sampled at the age of 52–92 years (average [±SD] 70.1 ± 8.5 years). Controls were also sampled at the age of 52–92 years (average [±SD] 70.2 ± 8.5 years). All cases and controls are white, and each group includes 165 females (53.1%) and 146 males (46.9%), respectively. Cases in this series do not carry the Gly2019Ser mutation in *LRRK2* [MIM 609007], which may occur in idiopathic PD,[3] and several tests did not reveal evidence of significant population stratification for 78 individually genotyped null markers (data not shown). We individually genotyped the 11 SNPs that were reported significant and one of the two SNPs that map to the *PARK10* [MIM 606852] locus (the two reported-significant SNPs are highly correlated: $r^2 = 0.99$), using allele-specific real-time PCR in our PD case-control sample set. Cases and controls were run on the same plate in a blinded fashion. Our genotyping method has an overall accuracy of >99%.[4] As an additional indication of genotyping quality, we calculated deviation from Hardy-Weinberg equilibrium (HWE) in cases and controls. One marker had an HWE exact P value of <.05 (.017 for *rs2245218* in cases), but further examination of our genotype data did not reveal questionable calls. Therefore, these data were included in our analysis. All SNPs were tested for allelic association with PD with the use of χ^2 statistics to calculate two-sided P values (table 1). Power calculations were done for a sample size of 311 pairs for each SNP, with the use of a one-sided

Table 1

Allelic Tests of SNPs Associated with Late-Onset PD

dbSNP Accession Number[a]	Gene	Chromosome	Position (Mbp)	Case[b]					Control[b]					Allelic Test		
				11	12	22	Sum	MAF[c]	11	12	22	Sum	MAF	OR (95% CI)	*P*[d]	Power (%)
rs7702187	*SEMA5A*	5	9.4	9	86	215	310	.168	8	83	217	308	.161	1.05 (.78–1.42)	.74	97
rs10200894		2	228.6	1	48	262	311	.080	3	66	238	307	.117	.66 (.45–.96)	.03	93
rs2313982		4	139.1	2	47	258	307	.083	4	45	256	305	.087	.95 (.64–1.42)	.81	96
rs17329669		7	36.6	10	73	224	307	.151	3	60	246	309	.107	1.49 (1.07–2.09)	.02	94
rs7723605		5	5.4	6	86	218	310	.158	5	70	232	307	.130	1.25 (.91–1.72)	.17	95
ss46548856		10	59.0	2	52	256	310	.090	4	61	241	306	.113	.78 (.54–1.13)	.19	92
rs16851009		2	166.5	6	53	251	310	.105	4	45	259	308	.086	1.24 (.85–1.82)	.26	94
rs2245218	*PRDM2*	1	13.9	12	64	234	310	.142	6	83	215	304	.156	.89 (.65–1.22)	.48	94
rs7878232:																
Male and female		X	150.5	…	…	…	307	.230	…	…	…	301	.244	.92 (.71–1.20)	.55	66
Female		…	…	13	47	101	161	.227	9	61	89	159	.248	.89 (.62–1.28)	.52	…
Male		…	…	…	…	…	146	.233	…	…	…	142	.239	.96 (.56–1.66)	.90	…
rs1509269		4	139.1	5	63	243	311	.117	7	65	235	307	.129	.90 (.64–1.26)	.55	92
rs11737074		4	125.4	16	116	178	310	.239	13	102	192	307	.208	1.19 (.91–1.56)	.20	90
rs7520966	*LOC200008*	1	54.4	16	117	175	308	.242	19	129	160	308	.271	.86 (.66–1.11)	.24	97

[a] The top 11 markers are presented in the same order as in table 4 in Maraganore et al.[1]
[b] Counts of genotype 11, 12, and 22.
[c] Minor-allele frequency.
[d] Two-sided *P* value for all strata and for female and male substrata in *rs7878232*.

allelic χ^2-hypothesis test at a significance level of 0.05 and with the assumption that the control-allele frequencies of the unrelated controls and odds ratios (ORs) in table 4 in Maraganore et al.[1] are true population parameters. Power calculation for *rs7520966* was based on the tier 2 OR given in the text of Maraganore et al.,[1] since it did not appear in their table 4.

Two markers, *rs10200894* and *rs17329669,* were replicated in our sample set at $P < .1$ ($P = .03$ and $P = .02$, respectively) with the same risk alleles as in Maraganore et al.,[1] although with slightly lower ORs. *rs10200894* is an intergenic variant located on chromosome 2 near a linkage peak previously identified in late-onset PD,[5] and *rs17329669* is in an intergenic region on chromosome 7. Further investigations in these regions, including further genetic mapping and the identification of potential causative variants, are thus warranted. Indeed, several SNPs in the vicinity of *rs10200894* and *rs17329669* reached significance in the Maraganore et al.[1] discovery sample set ($P < .05$) but were not followed up because they did not reach their significance threshold of $P < .01$. *ELMO1* [MIM 606420], a gene whose product is predicted to be involved in apoptosis and cell migration, resides in a region that, according to the HapMap, is in high linkage disequilibrium with *rs17329669*. The more abundant splice variant of *ELMO1* appears to be exclusively expressed in brain[6] and, thus, constitutes an excellent biological candidate gene for PD. All other markers were not significant in our sample set at the 0.1 level, including the marker reported most significant in *SEMA5A* [MIM 609297] and the marker in *LOC200008,* which maps to the *PARK10* locus that appears to affect both disease risk and age of onset.[7–9] Our failure to replicate the majority of the associated markers may be due to false-positive results in the initial study or to locus heterogeneity. Although the power in our validation sample set is ≥90% for 11 of the 12 tested SNPs, this may be an overestimation due to an OR inflation ("jackpot effect") in the original study. In addition, our sample set included only late-onset cases, commonly defined by age at onset >50 years, whereas the study by Maraganore et al. included both early- and late-onset cases.[1] Thus, it is possible that nonreplicated markers are associated with early-onset PD but make a lesser contribution to the more common, late-onset form of the disease. Additional studies are required to further assess the association of these markers with PD.

Acknowledgments

We thank the contributors to and the organizers of NINDS Human Genetics Resources, particularly Dr. Katrina Gwinn Hardy and Jeanne Beck, for making the clinical samples available to the Parkinson disease research community, and we thank the families and individuals for their participation. We also thank our colleagues at Celera Diagnostics, particularly Alla Smolgovsky and David Wolfson, for providing expert technical support.

Yonghong Li, Charles Rowland, Steven Schrodi, Walter Laird, Kristina Tacey, David Ross, Diane Leong, Joseph Catanese, John Sninsky, and Andrew Grupe

Celera Diagnostics
Alameda, CA

Web Resources

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=snp

International HapMap Project, http://www.hapmap.org/

NINDS Human Genetics Resources at the Coriell Institute, http://locus.umdnj.edu/ninds

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for PD, *LRRK2, PARK10, ELMO1,* and *SEMA5A*)

References

1. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PVK, Frazer KA, Cox DR, Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 77:685–693
2. Hughes AJ, Daniel SE, Kilford L, Lees AJ (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 55:181–184
3. Gilks WP, Abou-Sleiman PM, Gandhi S, Jain S, Singleton A, Lees AJ, Shaw K, Bhatia KP, Bonifati V, Quinn NP, Lynch J, Healy DG, Holton JL, Revesz T, Wood NW (2005) A common LRRK2 mutation in idiopathic Parkinson's disease. Lancet 365:415–416
4. Li Y, Nowotny P, Holmans P, Smemo S, Kauwe JS, Hinrichs AL, Tacey K, et al (2004) Association of late-onset Alzheimer's disease with genetic variation in multiple members of the GAPD gene family. Proc Natl Acad Sci USA 101:15688–15693
5. Pankratz N, Nichols WC, Uniacke SK, Halter C, Rudolph A, Shults C, Conneally PM, Foroud T (2003) Significant linkage of Parkinson disease to chromosome 2q36-37. Am J Hum Genet 72:1053–1057
6. Gumienny TL, Brugnera E, Tosello-Trampont AC, Kinchen JM, Haney LB, Nishiwaki K, Walk SF, Nemergut ME, Macara IG, Francis R, Schedl T, Qin Y, Van Aelst L, Hengartner MO, Ravichandran KS (2001) CED-12/ELMO, a novel member of the CrkII/Dock180/Rac pathway, is required for phagocytosis and cell migration. Cell 107:27–41
7. Hicks AA, Petursson H, Jonsson T, Stefansson H, Johannsdottir HS, Sainz J, Frigge ML, Kong A, Gulcher JR, Stefansson K, Sveinbjornsdottir S (2002) A susceptibility gene for late-onset idiopathic Parkinson's disease. Ann Neurol 52:549–555
8. Li YJ, Scott WK, Hedges DJ, Zhang F, Gaskell PC, Nance MA, Watts RL, et al (2002) Age at onset in two common neurodegenerative diseases is genetically controlled. Am J Hum Genet 70:985–993
9. Oliveira SA, Li YJ, Noureddine MA, Zuchner S, Qin X, Pericak-Vance MA, Vance JM (2005) Identification of risk and age-at-onset genes on chromosome 1p in Parkinson disease. Am J Hum Genet 77:252–264

Address for correspondence and reprints to: Dr. Yonghong Li, Celera Diagnostics, 1401 Harbor Bay Parkway, Alameda, CA 94502. E-mail: yonghong.li@celeradiagnostics.com

0002-9297/2006/7806-0023$15.00

Am. J. Hum. Genet. 78:1092–1094, 2006

Response from Maraganore et al.

To the Editor:

In this issue, four independent research teams present new genetic association data for 13 SNPs previously reported by us to be potentially associated with Parkinson disease (PD [MIM 168600]).[1] Two groups[2,3] report statistically significant association between one or more of these SNPs and PD, whereas two groups[4,5] find no statistically significant association between PD and any of the SNPs investigated. In an accompanying letter,[6] Dr. Richard H. Myers provides his qualitative assessment of the implications of these new results.

We have performed a Mantel-Haenszel analysis, using 10 of the 13 SNPs not displaying linkage disequilibrium (LD) with each other—combining the data of Li et al.,[3] Farrer et al.,[4] and Goris et al.[5]—to provide an overall quantitative assessment of the new results. The odds ratios (ORs) are reported for the SNP alleles that increase the risk of PD[1] (table 1). The X-linked SNP *rs7878232* was not included in this analysis, since subgroup-level data for males and females were not reported by all groups. The results of Clarimon et al.[2] were also not included, given the significant difference in SNP allele frequency observed between the European and Taiwanese control samples. This analysis reveals that none of the 10 SNPs shows statistically significant association with PD (i.e., $P < .05$). As pointed out in many of the accompanying letters, this failure to replicate may be due, in part, to differences in sample ascertainment and demographics.

A Mantel-Haenszel analysis combining these new results with those from tier 2 of Maraganore et al. reveal five SNPs with $P < .05$ and smaller effect sizes than were originally reported[1] (table 1). Although we are aware that these low *P* values may, at least in part, be explained by multiple testing, additional data are required to determine if these SNPs truly confer PD susceptibility or if they represent false-positive associations. Despite the small ORs, the point estimates of attributable risk for PD in the total data is still quite large for two of these SNPs (*rs10200894* population-attributable risk 0.27, 95% CI 0.04–0.77; *rs7520966* population-attributable risk 0.21, 95% CI 0.1–0.39). If these are true associations, they may have substantial practical impact on PD.

We do not agree with Dr. Myers[6] that our failure to identify an association between the *LRRK2* gene and PD in our original study is evidence of a false-negative result. Farrer et al. have reported elsewhere that only a very small number of the individuals with PD studied in our original whole-genome scan have a mutation in the *LRRK2* (MIM 609007) gene.[7]

We also do not consider the positive association findings between SNP *rs7702187* and PD in a Taiwanese population by Clarimon et al.[2] to be a replication of our original study results, since the SNP allele associated with PD susceptibility is not the same in the two studies. However, further work to follow up these results in the Taiwanese population seems warranted.

It is gratifying that our hypotheses have been tested

Table 1

Meta-Analysis of Genetic Association for 10 SNPs

dbSNP Accession Number	Alleles (High-Risk Allele)	Maraganore et al.[1] Tiers 1 and 2					Meta-Analysis (Replication Studies)					Meta-Analysis (Replication Studies and Maraganore et al.[1] Tier 2)				
		Cases	Controls	Allele Frequency (Controls)	OR (95% CI)	*P*	Cases	Controls	Allele Frequency (Controls)	OR (95% CI)	*P*	Cases	Controls	Allele Frequency (Controls)	OR (95% CI)	*P*
rs10200894	C/G (C)	772	772	.88	1.84 (1.38–2.45)	1.70×10^{-5}	1,566	1,546	.89	1.14 (.96–1.35)	.125	1,926	1,955	.89	1.25 (1.07–1.45)	.004
rs11737074	G/A (A)	764	764	.19	1.50 (1.21–1.86)	1.55×10^{-4}	1,563	1,542	.21	1.02 (.9–1.15)	.770	1,925	1,952	.21	1.09 (.97–1.21)	.142
rs16851009	C/T (T)	741	741	.08	1.84 (1.36–2.49)	4.17×10^{-5}	1,539	1,544	.1	.98 (.83–1.16)	.853	1,899	1,953	.1	1.08 (.93–1.26)	.312
rs17329669	A/G (G)	768	768	.12	1.71 (1.33–2.21)	2.30×10^{-5}	1,554	1,525	.12	1.13 (.97–1.32)	.102	1,914	1,933	.12	1.22 (1.06–1.39)	.004
rs2245218	A/G (G)	770	770	.12	1.67 (1.29–2.14)	4.61×10^{-5}	1,571	1,563	.16	.94 (.82–1.08)	.369	1,933	1,971	.16	1.02 (.9–1.16)	.752
rs2313982	C/T (T)	740	740	.07	2.01 (1.44–2.79)	1.79×10^{-5}	1,562	1,554	.09	.88 (.73–1.04)	.138	1,924	1,964	.09	1.01 (.86–1.18)	.935
rs7520966	C/T (C)	769	769	.7	.67 (.55–.81)	2.96×10^{-5}	1,563	1,550	.72	1.07 (.96–1.2)	.242	1,923	1,956	.72	1.15 (1.04–1.27)	.007
rs7702187	T/A (T)	761	761	.81	1.74 (1.36–2.24)	7.62×10^{-6}	1,541	1,541	.83	1.07 (.93–1.22)	.334	1,900	1,950	.82	1.14 (1.01–1.29)	.030
rs7723605	T/C (C)	773	773	.11	1.78 (1.35–2.35)	3.30×10^{-5}	1,567	1,571	.13	1.03 (.89–1.19)	.684	1,927	1,981	.13	1.12 (.98–1.28)	.105
ss46548856	G/C (G)	765	765	.9	1.88 (1.38–2.57)	3.65×10^{-5}	1,551	1,528	.9	1.12 (.94–1.33)	.196	1,913	1,933	.9	1.21 (1.03–1.42)	.016

rapidly by many groups. The Michael J. Fox Foundation, which funded our original research, also has a large-scale replication study under way. Given the low heritability estimates for PD,[8] our initial study may have been underpowered for the detection of significant genetic associations, in part, because of the large number of genetic markers tested. Therefore, it may be prudent not to limit replication of our study to the 13 SNPs that we initially highlighted but to also consider additional SNPs and genes that had suggestive findings (as in the text files published in the online-only version of our original article).[1]

DEMETRIUS M. MARAGANORE,[1] MARIZA DE ANDRADE,[2] TIMOTHY G. LESNICK,[2] P. V. KRISHNA PANT,[3] DAVID R. COX,[3] AND DENNIS G. BALLINGER[3]

Departments of [1]Neurology and [2]Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN; and [3]Perlegen Sciences, Mountain View, CA

Web Resources

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=snp

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for PD and *LRRK2*)

References

1. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PVK, Frazer KA, Cox DR, Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 77:685–693
2. Clarimon J, Scholz S, Fung H-C, Hardy J, Eerola J, Hellström O, Chen C-M, Wu Y-R, Tienari PJ, Singleton A (2006) Conflicting results regarding the semaphorin gene (*SEMA5A*) and the risk for Parkinson disease. Am J Hum Genet 78:1082–1084 (in this issue)
3. Li Y, Rowland C, Schrodi S, Laird W, Tacey K, Ross D, Leong D, Catanese J, Sninsky J, Grupe A (2006) A case-control association study of the 12 single-nucleotide polymorphisms implicated in Parkinson disease by a recent genome scan. Am J Hum Genet 78: 1090–1092 (in this issue)
4. Farrer MJ, Haugarvoll K, Ross OA, Stone JT, Milkovic NM, Cobb SA, Whittle AJ, Lincoln SJ, Hulihan MM, Heckman MG, White LR, Aasly JO, Gibson JM, Gosal D, Lynch T, Wszolek ZK, Uitti RJ, Toft M (2006) Genomewide association, Parkinson disease, and *PARK10*. Am J Hum Genet 78:1084–1088 (in this issue)
5. Goris A, Williams-Gray CH, Foltynie T, Compston DAS, Barker RA, Sawcer SJ (2006) No evidence for association with Parkinson disease for 13 single-nucleotide polymorphisms identified by whole-genome association screening. Am J Hum Genet 78:1088–1090 (in this issue)
6. Myers RH (2006) Considerations for genomewide association studies in Parkinson disease. Am J Hum Genet 78:1081–1082 (in this issue)
7. Farrer M, Stone J, Mata IF, Lincoln S, Kachergus J, Hulihan M, Strain KJ, Maraganore MD (2005) LRRK2 mutations in Parkinson disease. Neurology 65:738–740
8. Rocca WA, McDonnell SK, Strain KJ, Bower JH, Ahlskog JE, Elbaz A, Schaid DJ, Maraganore DM (2004) Familial aggregation of Parkinson's disease: the Mayo Clinic Family Study. Ann Neurol 56: 495–502

Address for correspondence and reprints: Dr. Demetrius Maraganore, Department of Neurology, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905. E-mail: dmaraganore@mayo.edu

0002-9297/2006/7806-0024$15.00

Am. J. Hum. Genet. 78:1094–1095, 2006

A Note on Permutation Tests in Multistage Association Scans

To the Editor:

There is currently a great deal of interest in performing whole-genome scans for association between genetic markers—mainly SNPs—and biological or clinical end points.[1] Often, the most cost-effective strategy for these studies is a staged design in which a subset of the full sample is genotyped for all SNPs, and only those SNPs that show a trend of association are genotyped in the remainder of the sample.[2]

For calculating the significance of a genome scan, permutation tests have been suggested to adjust for multiple testing while preserving the correlation structure among linked markers.[3] In the staged design, however, permutation may result in a marker being selected for the second stage that had not been selected in the original analysis. Such a marker will not have been genotyped in the full sample, and data will not be available to complete the analysis of the permuted data. Recently, Lin[4] proposed a Monte Carlo method for assessing significance in two-stage association scans. The method is sound but is limited to analysis based on efficient score functions and does not use permutation. Other investigators have reported methods to address this problem.[5]

I wish to draw attention to a property of genome scans that permits a simple permutation procedure for staged designs, which is that the sample sizes are large enough for the null distributions to be asymptotically stable. Although this observation is trivial, its utility might have escaped some readers, because of the origins of permutation testing in small-sample inference. It means that any large subset of the data can be used to simulate the null distribution. In particular, we can simulate a staged design with just the first-stage subjects, by using a subset of the first stage as the simulated first stage, selecting markers on the basis of that subset, and using the remainder of the first stage as the simulated second stage. This ensures that full genotype data are always available

and will generate approximately the same null distribution as exists for the full sample.

More precisely, consider a two-stage scan of a set of markers, M, in a set of subjects, S. In the first stage, all markers in M are genotyped in a subset of subjects, $S_1 \subset S$. An algorithm, $A(M; S_1)$, selects a subset of markers, M_1, on the basis of the data for S_1, which are then genotyped in the remaining subjects $S_2 = S \setminus S_1$. Next, perform a permutation test by using just the first-stage subjects as follows. Choose a simulated first-stage subsample, $S_1^* \subset S_1$, and a second-stage subsample, $S_2^* = S_1 \setminus S_1^*$. After each permutation, select markers $M_1^* = A(M; S_1^*)$. Compute statistics for markers M_1^* in subjects S_1, and compare them with the statistics of the original data for markers M_1 in subjects S. Assume that (i) there exists an asymptotic joint null distribution of test statistics on M and (ii) subjects are exchangeable between S_1 and S_2. Then, for sufficiently large $|S_1^*|$, $|S_2^*|$, and $|S_2|$, the permutation test will sample from the same null distribution (up to an arbitrary accuracy) as holds for the two-stage analysis of the full sample S.

For illustration and to confirm that the sample sizes proposed for genomewide scans are sufficiently large, a simulation was performed using 1,000 cases and 1,000 controls, which is a smaller sample than current estimates for well-powered scans.[6] Chromosomes were drawn from the phased CEU (CEPH subjects from Utah) data of chromosome 1, released in phase 1 of the International HapMap Project.[7] Parental chromosomes were drawn independently and grouped in pairs, and gametes were constructed using the supplied recombination maps, under the assumption of the Kosambi function with no interference between adjacent SNPs. Chromosomes of children were assigned from the constructed gametes according to Mendelian transmission and random union of gametes and were randomly assigned to the case or control group. In each replicate, 50% of subjects were used in the first stage, with the 10% most-significant markers considered in the second stage.[2] The significance of individual SNPs was calculated by the trend test,[8] and empirical distributions of the maximum trend statistic were generated from 1,000 replicates.

It is sufficient to show that the two-stage analysis of the first 500 cases and controls yields the same distribution as the analysis of all 1,000. The distributions were compared by the two-sample Kolmogorov-Smirnov test and also by the Kuiper test, which is more sensitive in the tail. No significant difference was found, implying that the null distribution is indeed stable at this sample size.

The main assumption of this approach is that subjects are exchangeable between stages, meaning that the null distribution is independent of the allocation of subjects to stages. This is true when the sample population is homogeneous but not when there are systematic differences between subpopulations. In particular, different patterns of linkage disequilibrium will invalidate this approach, as will population stratification in which differences in both allele frequency and trait distribution create a relationship between the null distribution and the specific subjects analyzed. When the sample consists of known proportions of different populations, the approach can be used if the proportions in the original data are preserved in the permutation test. Also, the large-sample assumption implies that only common variation is included; this is true for Hapmap SNPs, but, if rare variation is included, the permutation test will be less accurate. Nevertheless, for most well-designed scans of common variation, this approach is a practical and easily implemented solution for permutation testing in staged designs.

Acknowledgments

F.D. is supported by European Union contract LSHM-CT-2004-503485. Thanks to Doug Levinson and Peter Holmans for discussions.

FRANK DUDBRIDGE
Medical Research Council Biostatistics Unit
Cambridge
United Kingdom

References

1. Thomas DC, Haile RW, Duggan D (2005) Recent developments in genomewide association scans: a workshop summary and review. Am J Hum Genet 77:337–345
2. Sagatopan JM, Venkatraman ES, Begg CB (2004) Two-stage designs for gene-disease association studies with sample size constraints. Biometrics 60:589–597
3. Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971
4. Lin DY (2006) Evaluating statistical significance in two-stage genomewide association studies. Am J Hum Genet 78:505–509
5. Lewinger JP, Thomas DC (2005) Controlling the family-wise error rate in multistage genome-wide association studies [abstract]. Genet Epidemiol 29:262
6. Wang WY, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6:109–118
7. International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299–1320
8. Sasieni P (1997) From genotypes to genes: doubling the sample size. Biometrics 53:1253–1261

Address for correspondence and reprints: Dr. Frank Dudbridge, MRC Biostatistics Unit, Robinson Way, Cambridge CB2 2SR, United Kingdom. E-mail: frank.dudbridge@mrc-bsu.cam.ac.uk

0002-9297/2006/7806-0025$15.00

Am. J. Hum. Genet. 78:1096–1095, 2006

Reply to Dudbridge

To the Editor:

The standard permutation approach cannot be applied to two-stage association studies, because a marker that was not originally selected for the second stage of the study may be selected after permutation. To get around this difficulty, Frank Dudbridge proposes[1] (in this issue) to simulate a two-stage design by using only the first-stage subjects. This is a very clever idea and seems to be in a spirit similar to my Monte Carlo method,[2] in that both methods use only the data from the first-stage subjects to estimate the correlations of the test statistics. I believe that Dudbridge's permutation method (implicitly) requires that the same design (in terms of the proportion of subjects used in the first stage) be adopted in the permutation process as in the original study; otherwise, the joint distribution between the two stages obtained by permutation will not properly reflect the true joint distribution.

I wish to respond briefly to Dudbridge's comment that my Monte Carlo method "is limited to analysis based on efficient score functions and does not use permutation."[1] As mentioned in my report,[2] all test statistics can be represented by efficient score functions. Thus, the use of efficient score functions in generating the null distribution of the test statistics does not, in any way, limit the scope of application. As discussed in an earlier article,[3] the Monte Carlo approach has important advantages over the permutation approach. First, the permutation approach requires repeated calculations of the test statistics for each permuted data set, which can be prohibitively time consuming if the calculation of each test statistic is nontrivial, as will be the case if proper statistical methods are employed to test haplotype-disease associations,[4] whereas the Monte Carlo approach involves simulation of normal random variables only and is thus very efficient. Second, the permutation method can be used only to test the global null hypothesis that the variable being permuted is independent of all other variables and cannot be used to test, for example, gene-environment interactions, whereas the Monte Carlo approach can be used to test any kind of hypothesis.

D. Y. Lin

Department of Biostatistics
University of North Carolina
Chapel Hill

References

1. Dudbridge F (2006) A note on permutation tests in multistage association scans. Am J Hum Genet 78:1094–1095 (in this issue)
2. Lin DY (2006) Evaluating statistical significance in two-stage genomewide association studies. Am J Hum Genet 78:505–509
3. Lin DY (2005) An efficient Monte Carlo approach to assessing statistical significance in genomic studies. Bioinformatics 21:781–787
4. Lin DY, Zeng D, Millikan R (2005) Maximum likelihood estimation of haplotype effects and haplotype-environment interactions in association studies. Genet Epidemiol 29:299–312

Address for correspondence and reprints: Dr. Danyu Lin, Department of Biostatistics, University of North Carolina, McGavran-Greenberg Hall, CB #7420, Chapel Hill, NC 27599-7420. E-mail: lin@bios.unc.edu

0002-9297/2006/7806-0026$15.00