Am. J. Hum. Genet. 77:685–693, 2005

High-Resolution Whole-Genome Association Study of Parkinson Disease

Demetrius M. Maraganore,[1] Mariza de Andrade,[2] Timothy G. Lesnick,[2] Kari J. Strain,[2] Matthew J. Farrer,[3] Walter A. Rocca,[1,2] P. V. Krishna Pant,[4] Kelly A. Frazer,[4] David R. Cox,[4] and Dennis G. Ballinger[4]

Departments of [1]Neurology and [2]Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN; [3]Department of Neuroscience, Mayo Clinic College of Medicine, Jacksonville, FL; and [4]Perlegen Sciences, Mountain View, CA

We performed a two-tiered, whole-genome association study of Parkinson disease (PD). For tier 1, we individually genotyped 198,345 uniformly spaced and informative single-nucleotide polymorphisms (SNPs) in 443 sibling pairs discordant for PD. For tier 2a, we individually genotyped 1,793 PD-associated SNPs ($P < .01$ in tier 1) and 300 genomic control SNPs in 332 matched case–unrelated control pairs. We identified 11 SNPs that were associated with PD ($P < .01$) in both tier 1 and tier 2 samples and had the same direction of effect. For these SNPs, we combined data from the case–unaffected sibling pair (tier 1) and case–unrelated control pair (tier 2) samples and employed a liberalization of the sibling transmission/disequilibrium test to calculate odds ratios, 95% confidence intervals, and P values. A SNP within the semaphorin 5A gene (*SEMA5A*) had the lowest combined P value ($P = 7.62 \times 10^{-6}$). The protein encoded by this gene plays an important role in neurogenesis and in neuronal apoptosis, which is consistent with existing hypotheses regarding PD pathogenesis. A second SNP tagged the *PARK11* late-onset PD susceptibility locus ($P = 1.70 \times 10^{-5}$). In tier 2b, we also selected for genotyping additional SNPs that were borderline significant ($P < .05$) in tier 1 but that tested a priori biological and genetic hypotheses regarding susceptibility to PD ($n = 941$ SNPs). In analysis of the combined tier 1 and tier 2b data, the two SNPs with the lowest P values ($P = 9.07 \times 10^{-6}$; $P = 2.96 \times 10^{-5}$) tagged the *PARK10* late-onset PD susceptibility locus. Independent replication across populations will clarify the role of the genomic loci tagged by these SNPs in conferring PD susceptibility.

Introduction

Association-based genome scans provide localizing information that is much more precise (often extending over a few thousand base pairs) than the corresponding information from linkage-based studies (which often extend over many millions of base pairs). To date, there has been only one published high-resolution genome scan for association for any human disease (Klein et al. 2005). Here, we report the results of a high-resolution, whole-genome association study of Parkinson disease (PD [MIM 168600]). Our findings contribute to the creation of a genomic predisposition map for PD, and we illustrate a tiered genotyping approach that can be applied to the study of other complex diseases.

Material and Methods

We performed a high-resolution, whole-genome association study of PD, using a two-tiered genotyping approach (tier 1 and tiers 2a and 2b). All methods of the study were approved by the investigational review board of the Mayo Clinic.

Tier 1

For tier 1, we included sibling pairs that were discordant for PD. Cases were enrolled prospectively from the clinical practice of the Department of Neurology of the Mayo Clinic in Rochester, MN, from June 1996 through May 2004. They all resided within Minnesota or one of the surrounding four states (Wisconsin, Iowa, South Dakota, or North Dakota). All cases underwent a standardized clinical assessment performed by a neurologist subspecialized in movement disorders. Cases had at least two of four cardinal signs of parkinsonism (rest tremor, rigidity, bradykinesia, and/or postural instability) and no features atypical for PD (such as unexplained upper motor neuron signs or cerebellar signs). When nonmotor manifestations such as dysautonomia or dementia were present, they were mild and occurred late in the disease course. Subjects with secondary causes of parkinsonism (e.g., history of neuroleptic exposure, encephalitis, or multiple strokes) were excluded. All patients treated with a daily dosage total of ⩾1 g of levodopa (in combination with carbidopa) had a more than minimal improvement

Received June 3, 2005; accepted for publication July 28, 2005; electronically published September 9, 2005.

Address for correspondence and reprints: Dr. Dennis Ballinger, Perlegen Sciences, Inc., 2021 Stierlin Court, Mountain View, CA 22102. E-mail: dballinger@perlegen.com

© 2005 by The American Society of Human Genetics. All rights reserved.
0002-9297/2005/7705-0002$15.00

in parkinsonism symptoms and signs. We obtained a genealogical history from all cases, and, when permitted, we contacted available siblings for a telephone interview, to exclude parkinsonism via a validated screening instrument (Rocca et al. 1998). For cases, we obtained blood after a clinical assessment performed either at the Mayo Clinic or in the subjects' homes. For siblings who screened negative for PD (i.e., who had none of the following: prior diagnosis of PD, prior treatment with levodopa, or three or more of nine symptoms), we obtained blood via mail-in kits (we did no clinical assessment). All subjects provided written informed consent; their whole blood was obtained, via venipuncture, for DNA extraction (via the Puregene method [Gentra Sciences]) and storage. Cases were matched to a single participating sibling, without PD or parkinsonism, first by sex (when possible) and then by closest age.

For each subject (matched discordant sibling pairs), ~1 μg of DNA was shipped to Perlegen Sciences for laboratory study. Whole-genome amplification was performed as described elsewhere (Dean et al. 2002). For each subject, DNA was individually genotyped, for a set of 248,535 SNPs, with unique positions on National Center for Biotechnology Information (NCBI) build 34. These SNPs were selected to have relatively uniform spacing across the genome and to preferentially include haplotype-defining common SNPs from the Perlegen haplotype map (Patil et al. 2001). Details regarding the distribution of genomic gaps is provided in table 1. The genotyping platform employed high-density oligonucleotide, photolithographic microarrays (DNA chips), such that one hybridization yielded genotypes for 85,000 SNPs in a single individual.

We performed a liberalization of the sibling transmission/disequilibrium test (sTDT) (Schaid and Rowland 1998) to identify SNPs that had significant allele-frequency differences in cases versus unaffected siblings, adjusting the analyses for age and sex. For each SNP, we calculated odds ratios (ORs), 95% CIs, and P values (using a log-additive or "trend" model).

Tier 2

For tiers 2a and 2b, we individually genotyped case–unrelated control pairs. Cases were enrolled as for tier 1 but had no siblings available. Unrelated controls were identified via random digit dialing from the same five-state region as the cases and were screened negative for parkinsonism via the same validated telephone instrument as for siblings in tier 1. DNA was collected at the time of clinical assessment for cases and via mail-in kits for controls. Case–unrelated control pairs were matched for sex and age (±3 years). DNA was again shipped to Perlegen Sciences for whole-genome amplification and genotyping. The genotyping for tiers 2a and 2b was performed using customized oligonucleotide microarrays.

Table 1

Distribution of Gap Sizes between Adjacent Informative Tier 1 SNPs, within Chromosomal Contigs from NCBI Build 34

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

For tier 2a, a subset of SNPs that were associated with PD ($P < .01$ in tier 1) were genotyped. An additional predefined set of 311 SNPs was genotyped for genomic control (Hinds et al. 2004). Conditional logistic regression analyses were performed, to test for and model associations between the SNPs and PD. The analyses were adjusted for age and sex, as appropriate. We calculated ORs, 95% CIs, and P values (as for tier 1, using a log additive or "trend" model). We also checked for population structure, using the genomic-control method (Devlin and Roeder 1999; Bacanu et al. 2000; Devlin et al. 2004). This method treats the subjects as unmatched and does not include adjustments for covariates; however, it should still provide guidelines for assessing the degree of stratification in tier 2. For the SNPs associated with PD ($P < .01$) in both the tier 1 and 2a analyses, we pooled data for the case–unaffected sibling and case–unrelated control pairs and again analyzed the data, using a liberalization of the sTDT (Schaid and Rowland 1998). This allowed us to maximize statistical power for the analyses to obtain a more precise ranking of replicated SNPs by ascending P values and a more precise estimate of ORs and 95% CIs.

For tier 2b, we also selected for genotyping additional SNPs that in tier 1 were only borderline significant but that tested a priori biological or genetic hypotheses regarding genomic susceptibility to PD. These SNPs did not achieve a P value <.01 in the tier 1 sample overall. However, some of these SNPs (1) had P values <.001 in tier 1 strata defined by sex or by median age at study; (2) had P values <.05 and were positioned within 10 kb of the linkage-derived candidate genes α-synuclein (*SNCA* [MIM 163890]), parkin (MIM 602544), ubiquitin carboxy-terminal hydrolase L1 (*UCHL1* [MIM 191342]), microtubule-associated protein τ (*MAPT* [MIM 157140]), oncogene DJ1 (*DJ1* [MIM 602533]), and PTEN-induced kinase 1 (*PINK1* [MIM 608309]); (3) had P values <.05 and were positioned within additional loci linked to PD (i.e., *PARK3* [MIM 602404], *PARK8* [MIM 607060], *PARK9* [MIM 606693], *PARK10* [MIM 606852], *PARK11* [MIM 607688]); (4) had P values <.05 and were positioned within exons or within 10 kb 5′ of the transcript, genomewide; or (5) were redundant tags of European American linkage-disequilibrium bins (Hinds et al. 2005) highlighted by some

of the above criteria. We note that the tier 2b SNP selection predated the discovery of mutations in the leucine-rich repeat kinase 2 gene (*LRRK2* [MIM 609007]) as the cause of *PARK8*-linked parkinsonism (although SNPs within the gene were indeed selected and genotyped). We tested the association of these tier 2b SNPs with PD in the combined samples, using a liberalization of the sTDT (as above).

Results

Tier 1

For tier 1, we genotyped 443 case–unaffected sibling pairs. Table 2 summarizes demographic and clinical characteristics of the sample. For the 248,535 SNPs selected, the genotyping call rate was >80% for each of 220,143 SNPs. Of these SNPs, 205,031 (93%) were polymorphic within the study sample. The Hardy-Weinberg equilibrium (HWE) P value was >.001 in controls for 198,345 SNPs (97% of polymorphic SNPs, with an average gap between adjacent SNPs of 12,363 bp). Ultimately, for these subjects and informative SNPs, 172,420,019 genotype calls (98.1% of possible genotypes) were made. We regenotyped, in triplicate, 96 SNPs for each subject, with 99.8% concordance of genotypes. The concordance of genotypes called by the oligonucleotide array platform, as compared with genotypes called by other platforms employed as part of the multicenter HapMap project, was 99.5% (Hinds et al. 2005).

As a sensitivity analysis and to assess the genomic coverage of the informative SNPs, we determined our statistical power to detect unassayed, disease-associated variants with this SNP collection. Specifically, we previously genotyped these same SNPs in a different sample of European Americans who were also sequenced across selected genes by the SeattleSNPs Program for Genomic Applications (PGA) (SeattleSNPs) (Hinds et al. 2005). Our metric of coverage was the mean r^2 value for any common PGA SNP (minor-allele frequency >10%) with the most-correlated SNP in the same region from the 198,345 informative SNPs. The statistical power to detect an unassayed, disease-associated allele indirectly by use of a correlated allele of an assayed SNP is related to r^2. Specifically, the power to detect an association indirectly in N samples is equivalent to the power to detect it directly in Nr^2 samples (Pritchard and Przeworski 2001). Despite our assaying only 3.3% of all of the common SNPs in these intervals (and only 1% of the less common SNPs), the mean r^2 for unassayed SNPs was 0.57.

Table 2

Demographic and Clinical Characteristics (Tiers 1 and 2)

	Findings for Subjects			
	Tier 1		Tier 2	
Characteristics	PD Cases	Siblings	PD Cases	Controls
No. (%) of subjects:	443 (100)	443 (100)	332 (100)	332 (100)
Male	271 (61.2)	214 (48.3)	194 (58.4)	193 (58.1)
Female	172 (38.8)	229 (51.7)	138 (41.6)	139 (41.9)
Median age (range), in years, at PD onset	61 (31–94)	…	63 (36–88)	…
Median age (range), in years, at time of study	68 (33–96)	66 (29–90)	68 (42–90)	67 (42–91)
Percentage of subjects with family history of PD[a]	20.5	…	14.9	…
No. (%) of subjects with parental region of origin:				
Both parents of European origin:	381 (86.0)	363 (81.9)	269 (81.0)	272 (81.9)
Both parents northern European[b]	111 (29.1)	100 (27.5)	84 (31.2)	83 (30.5)
Both parents central European[c]	145 (38.1)	135 (37.2)	96 (35.7)	84 (30.9)
Both parents southern European[d]	3 (.8)	3 (.8)	2 (.7)	3 (1.1)
Both parents European, mixed region	122 (32.0)	125 (34.4)	87 (32.3)	102 (37.5)
Only one parent of European origin[e]	39 (8.8)	39 (8.8)	43 (13.0)	35 (10.5)
One parent declared "American"[f]	2 (.5)	1 (0.2)	1 (.3)	4 (1.2)
Both parents declared "American"	18 (4.1)	23 (5.2)	12 (3.6)	19 (5.7)
Both parents Asian	…	1 (.2)	1 (.3)	…
Origin unknown	3 (.7)	16 (3.6)	6 (1.8)	2 (.6)

[a] Family history was defined as having at least one affected first-degree relative; for tier 1, 90/439 cases had a family history of PD (information was missing for 4); for tier 2, 48/322 cases had a family history of PD (information was missing for 10).

[b] "Northern European" includes Scandinavian, Swedish, Norwegian, Finnish, Danish, Irish, or British origins.

[c] "Central European" includes French, Belgian, Dutch, Swiss, Luxemburgian, German, Austrian, Hungarian, Polish, Czechoslovakian, or Russian origins.

[d] "Southern European" includes Italian, Spanish, Portuguese, Greek, or Yugoslavian origins.

[e] Includes subjects for whom origin of one parent is unknown.

[f] These subjects were all European Americans and not Native Americans.

There were 1,862 SNPs that were associated with PD in tier 1 ($P < .01$), after exclusion of SNPs with fewer than nine discordances between sibling pairs (SNPs for which ORs could not be reliably estimated). We provide details for all informative tier 1 SNPs (including their dbSNP reference sequence identification numbers, gene names, case and control allele frequencies, ORs, 95% CIs, and *P* values) in tab-delimited text file 1 (online only); the columns are described in table 3.

Tier 2

For tier 2a, we genotyped 332 case–unrelated control pairs. Table 2 summarizes demographic and clinical characteristics for these samples. Genotypes and analyses were attempted for the 1,862 SNPs selected in tier 1 and for 311 genomic control SNPs. Of these, genotyping call rates of >80% and HWE *P* values >.001 (in controls) were achieved for 1,793 SNPs. We were able to call successfully 1,176,772 (99% of possible) genotypes for these SNPs.

To assess the impact of population structure on tier 2 associations, trend tests were performed on the genomic control SNPs. The mean trend test statistic over the genomic control SNPs, λ_m, was 0.96, which is less than the expected value of 1 for no population structure (Devlin et al. 2004). Therefore, there was no indication of substantial confounding due to population structure bias.

There were 26 SNPs with *P* values <.01 in both tier 1 and tier 2a. Of these SNPs, 11 had nominal *P* values <.01 and the same direction of the effect in tier 1 and tier 2a (termed "replicated SNPs"). A SNP (*rs7702187*) within the semaphorin 5A gene (*SEMA5A* [MIM 609297]) had the lowest *P* value (OR 1.74; 95% CI 1.36–2.24; $P = 7.62 \times 10^{-6}$). A second SNP (*rs10200894*) tagged the *PARK11* late-onset PD susceptibility locus (OR 1.84; 95% CI 1.38–2.45; $P = 1.70 \times 10^{-5}$). Table 4 lists the 11 replicated SNPs, including their dbSNP reference sequence identification numbers and, when applicable, the corresponding gene ontology annotation of nearby genes. The SNPs are ordered by ascending combined *P* values. We provide additional results for the tier 2a SNPs (including their dbSNP reference sequence identification numbers and gene names and, for the combined samples, the case and control allele frequencies, ORs, 95% CIs, and *P* values), in tab-delimited text file 2 (online only); the columns are described in table 3. The 11 replicated SNPs from tiers 1 and 2a (table 4) had 20 neighboring SNPs within a distance of 15 kb of the reported data. None of these 20 SNPs were significant enough to be selected for inclusion in tier 2a. However, 6 of the 20 SNPs had tier 1 *P* values <.1 ($P = .011$ on a binomial test), and 14 of these 20 had *P* values <.2 ($P = 1.8 \times 10^{-6}$). Thus, there is evidence of correlations among neighboring SNPs.

Table 3

Column Descriptions for Text Files 1 and 2

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

Table 5 summarizes the rationale for inclusion of the tier 2b SNPs, primarily on the basis of a priori biological and genetic hypotheses regarding susceptibility to PD. We considered 975 tier 2b SNPs. Of these, genotyping call rates of >80% and HWE *P* values >.001 in controls were achieved for 941 SNPs. We were able to call successfully 617,257 (99% of maximum possible) genotypes for these SNPs. Of these, the two SNPs (*rs682705* and *rs7520966*) with the lowest combined tier 1 and tier 2b *P* values ($P = 9.07 \times 10^{-6}$, OR 0.66, 95% CI 0.54–0.79 and $P = 2.96 \times 10^{-5}$ OR 0.67, 95% CI 0.55–0.81, respectively) tagged the *PARK10* late-onset PD susceptibility locus. The r^2 value between the two SNPs was 0.993 (highly correlated). We provide additional results for the tier 2b SNPs (including their dbSNP reference sequence identification numbers, gene names, case and control allele frequencies, ORs, 95% CIs, and *P* values) in tab-delimited text file 2 (online only).

Discussion

Twin studies and familial aggregation studies have suggested that the genetic contribution to PD is small and limited to younger-onset cases (Rocca et al. 2004; Wirdefeldt et al. 2004). However, parametric linkage studies have assigned 11 loci to PD, and causal mutations have been identified in six genes (*SNCA,* parkin, *UCHL1, DJ1, PINK1,* and *LRRK2*) (Polymeropoulos et al. 1997; Kitada et al. 1998; Leroy et al. 1998; Bonifati et al. 2003; Valente et al. 2004; Zimprich et al. 2004). Nonparametric linkage studies have also assigned the locus containing the *MAPT* gene to PD (Martin et al. 2001). Whereas mutations in these genes only rarely cause PD, common variations in some of these genes may confer susceptibility with a more sizable population-attributable risk (Farrer et al. 2001; West et al. 2002; Maraganore et al. 2004; Mamah et al. 2005; Mueller et al. 2005). However, the search for PD susceptibility genes has been largely limited to the candidate-genes approach.

Our study is unique because we aimed to discover common PD susceptibility gene variants via a whole-genome association design. Despite the low heritability of PD, we have nominated several new susceptibility genes for PD (table 4). Findings for the *SEMA5A* gene are particularly noteworthy and highlight the apoptosis pathway (Tatton et al. 2003).

The *SEMA5A* gene (sema domain, seven thrombospondin repeats [type 1 and type 1–like], transmembrane domain [TM], and short cytoplasmic domain, [semaphorin] 5A) has not been previously implicated in

Table 4

Genomic SNPs Associated with PD in Two Samples (Tiers 1 and 2)

dbSNP Number	Gene Name	NCBI Build 35.1	Brain[a]	Gene Ontology[b] Function	Process	Component	Summary[b,c]	OR[d] (95% CI)	P[d]
rs7702187	*SEMA5A*	5p15.2	Yes	Receptor activity	Cell adhesion; cell-cell signaling; neurogenesis	Integral to membrane	Axonal guidance (neural development); initiation of neuronal apoptosis	1.74 (1.36–2.24)	7.62×10^{-6}
rs10200894	...	2q36	...	...	...	...	*PARK11* locus	1.84 (1.38–2.45)	1.70×10^{-5}
rs2313982	...	4q31.1	...	...	...	...	...	2.01 (1.44–2.79)	1.79×10^{-5}
rs17329669	...	7p14	...	...	...	...	...	1.71 (1.33–2.21)	2.30×10^{-5}
rs7723605	...	5p15.3	...	...	...	...	...	1.78 (1.35–2.35)	3.30×10^{-5}
ss46548856[e]	...	...	...	...	...	...	...	1.88 (1.38–2.57)	3.65×10^{-5}
rs16851009	*GALNT3*	2q24	Yes	Manganese, sugar binding; transferase activity	Carbohydrate metabolism	Golgi apparatus; integral to membrane	Marker of differentiation and aggressiveness (several cancers)	1.84 (1.36–2.49)	4.17×10^{-5}
rs2245218	*PRDM2*	1p36.2	Yes	DNA, metal, zinc binding; transcription factor, regulator activity	Transcription regulation	Nucleus	Tumor suppression; neuronal differentiation; estrogen receptor binding; estrogen effector	1.67 (1.29–2.14)	4.61×10^{-5}
rs7878232	*PASD1*	Xq28	Yes	Signal transducer activity	Signal transduction	...	X-linked	1.38 (1.17–1.62)	6.87×10^{-5}
rs1509269	...	4q31.1	...	...	...	...	...	1.71 (1.30–2.26)	9.21×10^{-5}
rs11737074	...	4q27	...	...	...	...	...	1.50 (1.21–1.86)	1.55×10^{-4}

[a] Evidence of brain expression. Annotated from GeneCards.
[b] Summary of biological plausibility. Annotated from Entrez Gene.
[c] Annotated from OMIM.
[d] Derived for the combined tier 1 and tier 2 samples, with use of a liberalization of the sTDT (log additive model).
[e] This SNP is designated with a Perlegen Sciences internal SNP identifier. It has been submitted to dbSNP; the rs number and cytogenic location will be provided when available.

Table 5

Criteria for SNPs Selected for Genotyping in Tier 2

Category	Description	No. of SNPs
1	$P < .01$ in tier 1 overall analysis	1,862
2	$P < .001$ in tier 1 age- or sex-stratified analyses	188
3	$P < .05$ in tier 1 overall analysis, within 10 kb of the six linkage-derived candidate genes (*SNCA, PARK2, UCHL1, MAPT, DJ1,* and *PINK1*)	8
4	$P < .05$ in tier 1 overall analysis, within additional *PARK* linkage loci for which genes have not been cloned (*PARK3, PARK8, PARK9, PARK10,* and *PARK11*)	145
5	$P < .05$ in tier 1 overall analysis, within exons or within 10 kb 5′ of the transcript, genomewide	589
6	Genomic control SNPs	311
7	Other SNPs in linkage disequilibrium with those passing criteria 1 and 2	45
Total		3,148

PD. Although the effect size is small (OR 1.7; 95% CI 1.36–2.24), the disease-associated allele occurs with sufficient frequency to confer a sizable population-attributable risk (minor-allele frequency = 19.6% for unrelated controls). This gene maps to the deletion candidate interval for cri du chat syndrome. This chromosomal deletion syndrome is associated with severe abnormalities in brain development, with resultant mental retardation (Simmons et al. 1998). The encoded semaphorin protein has receptor activity and contributes to neurogenesis and apoptotic processes. It elicits multiple functional responses through its receptor plexin-B3 (Artigiani et al. 2004). Specifically, semaphorins may play a role in the development of the mesencephalic dopamine neuron system (Kawano et al. 2003). They are also positive mediators of dopamine-induced apoptosis; antibodies directed against semaphorins inhibit dopamine-induced neuronal death (Shirvan et al. 1999, 2000; Barzilai et al. 2000). The expression of semaphorin genes is abnormal in Alzheimer disease, which shares clinical, pathological, and etiological features with PD (Hirsch et al. 1999). It has also been suggested in a rat model of PD that the neuroprotective effects of vascular endothelial growth factor may be mediated via semaphorin inhibition (Yasuhara et al. 2004). To date, there have been no clinical studies of semaphorin inhibitors as neuroprotective therapy for PD. Our genetic findings converge with cell biology and animal experiments to support the research and development of therapies targeting semaphorins as neuroprotection for PD.

Several additional findings of the whole-genome scan are consistent with our current understanding of PD (table 4). One of the replicated SNPs mapped the *PARK11* locus, which showed significant linkage to PD in one study but not in another (Pankratz et al. 2003; Prestel et al. 2005). Another replicated SNP was within the cancer-related UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 3 gene (*GALNT3* [MIM 601756]). This may be consistent with epidemiological studies that suggest that patients with PD have a higher risk of cancers after the diagnosis of PD (Elbaz et al. 2005). Another replicated SNP was within the PR domain–containing protein 2 gene (*PRDM2* [MIM 601196]), which is also related to neuronal differentiation and cancer. That gene is also estrogen related. This may be consistent with epidemiological data suggesting that endogenous estrogen and estrogen treatment are protective against PD (Benedetti et al. 2001). Finally, a replicated SNP was within the chromosome X gene *PASD1* (PAS domain–containing 1). X inactivation of a major susceptibility gene could explain the sex difference in lifetime risk of PD (men have a 1.5-times-greater risk than women) (Elbaz et al. 2002; Carrel and Willard 2005).

Our tier 2b, hypothesis-based SNP analyses were noteworthy for highlighting a chromosome 1p32 locus previously linked to late-onset PD susceptibility. The *PARK10* locus was originally identified in Icelandic families via a genomewide scan with use of microsatellite markers (Hicks et al. 2002). This locus was subsequently shown to influence age at onset in a multicenter study of families with PD (Li et al. 2002). To date, a specific gene within the *PARK10* locus has not been identified. Our disease-associated SNPs (*rs682705* and *rs7520966*) are within a gene designated *LOC200008,* which encodes a hypothetical protein with inferred oxidoreductase activity and potential involvement in cholesterol biosynthesis and electron transport. The SNPs are also within 21 kb upstream of the mitochondrial ribosomal protein L37 gene (*MRPL37)*. Mitochondrial mechanisms of PD pathogenesis have been described (Dauer and Przedborski 2003), and a role for mitochondrial ribosomal protein genes in human disorders has been postulated (Kenmochi et al. 2001).

To our knowledge, our study is the first high-resolution, whole-genome association study published about PD. Recent review articles anticipated such studies (Hirschhorn and Daly 2005; Wang et al. 2005), but our present understanding of human genome variability and the available high-throughput genotyping technologies currently allow for whole-genome studies (Patil et al. 2001; Olivier et al. 2001; Hinds et al. 2005). Also, our

two-tiered study design had several strengths. First, we performed a family-based case-control study (tier 1). This limited false-positive results due to population stratification bias (Maraganore 2005). However, the relatedness of cases and siblings may have reduced statistical power to detect genetic associations (false-negative results). Second, we performed a case–unrelated control study (tiers 2a and 2b). The use of unrelated controls (rather than unaffected siblings) optimized statistical power for the replication study (Maraganore 2005). It is noteworthy that, for tier 2, there was minimal population structure, despite the inclusion of unrelated cases and controls. This is fortunate, since we therefore did not need to stratify the analyses for the tier 2 sample. It is also a strength of our study that all subjects were individually genotyped (no pooling of DNA samples was involved). This allowed analyses adjusted or stratified for age and sex, factors that have been shown elsewhere to influence heritability in PD (Rocca et al. 2004; Mueller et al. 2005). Individual genotyping will also allow us to perform studies of complex gene-gene and gene-environment interactions and genotype-phenotype correlations within the available data set. It is intriguing that 7 of the 11 replicated SNPs are intergenic. This may suggest that intergenic sequences are important to health, perhaps via gene regulatory effects (Martens et al. 2004).

It is likely that the number of SNPs genotyped—or the specific ones selected—as well as the available sample size resulted in our missing disease-associated loci across the human genome (Carlson et al. 2003, 2004; Wang et al. 2005). However, we observed an association of SNPs neighboring the 11 replicated SNPs ("clustering"), which suggests that we achieved an appreciable map density. Also, with the available samples, we detected significant ($P < .01$) ORs $\geqslant 1.34$ in tier 1, $\geqslant 1.33$ in tier 2a, and $\geqslant 1.21$ in the combined analysis. It is also likely that multiple statistical comparisons or other factors resulted in false-positive findings (Storey and Tibshirani 2003; Hirschhorn and Daly 2005). None of our results were significant with use of a strict Bonferroni correction for the number of tests (198,345 SNPs for tier 1 and the combined analyses, 1,466 SNPs for tier 2a, and 1,312 SNPs for tier 2b). However, this is a very conservative evaluation of significance; it assumes total independence of the tests, and properly correcting for the correlational structure among the SNPs is not straightforward. Eleven SNPs had P values <.01 and the same direction of the effect in both tiers 1 and 2a, and many of these SNPs have compelling potential biological relationships with PD or are in regions otherwise implicated in PD via linkage analysis ("convergence").

The results of this study draw a first genomic map of PD, which awaits replication across populations (including collaborative-pooled analyses within global consortia). Our findings may ultimately identify biomarkers that contribute to early disease detection and primary prevention strategies. Also, they may nominate new molecular targets for disease-modifying therapies (secondary prevention). Correlation studies of genotypes to prognostic outcomes may predict the effect of treatments targeting these genes and their proteins ("virtual clinical trials") and thus reduce research and developments costs and identify subgroups of patients most likely to benefit from treatment (personalized medicine).

Acknowledgments

This study was supported primarily by a Michael J. Fox Foundation Linked Efforts to Accelerate Parkinson's Solutions award. We also received support from National Institutes of Health grants ES10751, ES10751-S1, NS33978, and NS40256. We acknowledge the outstanding contributions of the many Perlegen Sciences and Mayo employees who provided technical and infrastructural support for this work. But mostly, we thank the patients and study subjects who made this work possible and who give it value.

Web Resources

The URLs for data presented herein are as follows:

Entrez Gene, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
GeneCards, http://bioinfo.weizmann.ac.il/cards/index.shtml
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for PD, *SNCA,* parkin, *UCHL1, MAPT, DJ1, PINK1, PARK3, PARK8, PARK9, PARK10, PARK11, LRRK2, SEMA5A, GALNT3,* and *PRDM2*)
SeattleSNPs, http://pga.gs.washington.edu/ (for PGA)

References

Artigiani S, Conrotto P, Fazzari P, Gilestro GF, Barberis D, Giordano S, Comoglio PM, Tamagnone L (2004) Plexin-B3 is a functional receptor for semaphorin 5A. EMBO Rep 5:710–714

Bacanu S-A, Devlin B, Roeder K (2000) The power of genomic control. Am J Hum Genet 66:1933–1944

Barzilai A, Zilkha-Falb R, Daily D, Stern N, Offen D, Ziv I, Melemed E, Shirvan A (2000) The molecular mechanism of dopamine-induced apoptosis: identification and characterization of genes that mediate dopamine toxicity. J Neural Transm Suppl 60:59–76

Benedetti MD, Maraganore DM, Bower JH, McDonnell SK, Peterson BJ, Ahlskog JE, Schaid DJ, Rocca WA (2001) Hysterectomy, menopause, and estrogen use preceding Parkinson's disease. Mov Disord 16:830–837

Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, Dekker MCJ, Squitieri F, Ibanez P, Joosse M, van Dongen JW, Vanacore N, van Swieten JC, Brice A, Meco G, van Duijn CM, Oostra BA, Heutink P (2003) Mutations in

the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 299:256–259

Carlson CS, Eberle MA, Rieder MJ, Qian Y, Kruglyak L, Nickerson DA (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 74:106–120

Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA (2003) Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat Genet 33:518–521

Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434:400–404

Dauer W, Przedborski S (2003) Parkinson's disease: mechanisms and models. Neuron 39:889–909

Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS (2002) Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci USA 99:5261–5266

Devlin B, Bacanu S, Roeder K (2004) Genomic control to the extreme. Nat Genet 36:1129–1130

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004

Elbaz A, Bower JH, Maraganore DM, McDonnell SK, Peterson BJ, Ahlskog JE, Schaid DJ, Rocca WA (2002) Risk tables for parkinsonism and Parkinson's disease. J Clin Epidemiol 55:25–31

Elbaz A, Peterson BJ, Bower JH, Yang P, Maraganore DM, McDonnell SK, Ahlskog JE, Rocca WA (2005) Risk of cancer after the diagnosis of Parkinson's disease: a historical cohort study. Mov Disord 20:719–725

Farrer M, Maraganore DM, Lockhart P, Singleton A, Lesnick TG, de Andrade M, West A, de Silva R, Hardy J, Hernandez D (2001) α-Synuclein gene haplotypes are associated with Parkinson's disease. Hum Mol Genet 10:1847–1851

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

Hinds DA, Stokowski RP, Patil N, Konvicka K, Kershenobich D, Cox DR, Ballinger DG (2004) Matching strategies for genetic association studies in structured populations. Am J Hum Genet 74:317–325

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR (2005) Whole-genome patterns of common DNA variation in three human populations. Science 307:1072–1079

Hirsch E, Hu LJ, Prignet A, Constatin B, Agid Y, Drabkin H, Roche J (1999) Distribution of semaphorin IV in adult human brain. Brain Res 823:67–79

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies of common diseases and complex traits. Nat Rev Genet 6:95–108

Kawano H, Horie M, Honma S, Kawamura K, Takeuchi K, Kimura S (2003) Aberrant trajectory of ascending dopaminergic pathway in mice lacking Nkx2.1. Exp Neurol 182:103–112

Kenmochi N, Suzuki T, Uechi T, Magoori M, Kuniba M, Higa S, Watanabe K, Tanaka T (2001) The human mitochondrial ribosomal protein genes: mapping of 54 genes to chromosomes and implications for human disorders. Genomics 77:65–70

Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimizu N (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392:605–608

Klein RJ, Zeiss C, Chew EY, Tsai J-Y, 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

Leroy E, Boyer R, Auburger G, Leube B, Ulm G, Mezey E, Harta G, Brownstein MJ, Jonnalagada S, Chernova T, Dehejia A, Lavedan C, Gasser T, Steinbach PJ, Wilkinson KD, Polymeropoulos MH (1998) The ubiquitin pathway in Parkinson's disease. Nature 395:451–452

Li Y-J, 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

Mamah CE, Lesnick TG, Lincoln SJ, Strain KJ, de Andrade M, Bower JH, Ahlskog JE, Rocca WA, Farrer MJ, Maraganore DM (2005) Interaction of α-synuclein and tau genotypes in Parkinson's disease. Ann Neurol 57:439–443

Maraganore DM (2005) Blood is thicker than water: the strengths of family-based case-control studies. Neurology 64:408–409

Maraganore DM, Lesnick TG, Elbaz A, Chartier-Harlin M-C, Gasser T, Kruger R, Hattori N, Mellick GD, Quattrone A, Satoh J-I, Toda T, Wang J, Ioannidis JPA, de Andrade M, Rocca WA, the UCHL1 Global Genetics Consortium (2004) UCHL1 is a Parkinson's disease susceptibility locus. Ann Neurol 55:512–521

Martens JA, Laprade L, Winston F (2004) Intergenic transcription is required to repress the *Sacharomyces cerevisiae* SER3 gene. Nature 429:571–574

Martin ER, Scott WK, Nance MA, Watts RL, Hubble JP, Koller WC, Lyons K, et al (2001) Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson's disease. JAMA 286:2245–2250

Mueller JC, Fuchs J, Hofer A, Zimprich A, Lichtner P, Illig T, Berg D, Wullner U, Meitinger T, Gasser T (2005) Multiple regions of α-synuclein are associated with Parkinson's disease. Ann Neurol 57:535–541

Olivier M, Bustos VI, Levy MR, Smick GA, Moreno I, Bushard JM, Almendras AA, Sheppard K, Zierten DL, Aggarwal A, Carlson CS, Foster BD, Vo N, Kelly L, Liu X, Cox DR (2001) Complex high-resolution linkage disequilibrium and haplotype patterns of single-nucleotide polymorphisms in 2.5 Mb of sequence on human chromosome 21. Genomics 78:64–72

Pankratz N, Nichols WC, Uniacke SK, Halter C, Rudolph A, Schults C, Conneally PM, Foroud T, the Parkinson Study Group (2003) Significant linkage of Parkinson disease to chromosome 2q36-37. Am J Hum Genet 72:1053–1057

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Majoribanks C, McDonough DP, Nguyen BTN, Norric MC, Sheehan JB, Shen N, Stern D,

Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SPA, Cox DR (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719–1723

Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehajia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenroos ES, Chandrasekharappa S, Athanassiadou A, Papapetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe LI, Nussbaum RL (1997) Mutation in the α-synuclein gene identified in families with Parkinson's disease. Science 276:2045–2047

Prestel J, Sharma M, Leitner P, Zimprich A, Vaughan JR, Durr A, Bonifati V, De Michele G, Hanagasi HA, Farrer M, Hofer A, Asmus F, Volpe G, Meco G, Brice A, Wood NW, Muller-Myhsok B, Gasser T, the European Consortium on Genetic Susceptibility in Parkinson's Disease (2005) PARK11 is not linked with Parkinson's disease in European families. Eur J Hum Genet 13:193–197

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1–14

Rocca WA, Maraganore DM, McDonnell SK, Schaid DJ (1998) Validation of a telephone questionnaire for Parkinson's disease. J Clin Epidemiol 51:517–523

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

Schaid DJ, Rowland C (1998) Use of parents, sibs, and unrelated controls for detection of associations between genetic markers and disease. Am J Hum Genet 63:1492–1506

Shirvan A, Shina R, Ziv I, Melamed E, Barzilai A (2000) Induction of neuronal apoptosis by semaphorin3A-derived peptide. Mol Brain Res 83:81–93

Shirvan A, Ziv I, Fleminger G, Shina R, He Z, Brudo I, Melamed E, Barzilai A (1999) Semaphorins as mediators of neuronal apoptosis. J Neurochem 73:961–971

Simmons AD, Puschel AW, McPherson JD, Overhauser J, Lovett M (1998) Molecular cloning and mapping of human semaphorin F from the cri-du-chat candidate interval. Biochem Biophys Res Commun 242:685–91

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci USA 100:9440–9445

Tatton WG, Chalmers-Redman R, Brown D, Tatton N (2003) Apoptosis in Parkinson's disease: signals for neuronal degeneration. Ann Neurol 53:S61–S72

Valente EM, Abou-Sleiman PM, Caputo V, Muqit MMK, Harvey K, Gispert S, Ali Z, Del Turco D, Bentivoglio AR, Healy DG, Albanese A, Nussbaum R, Gonzalez-Maldonado R, Deller T, Salvi S, Corelli P, Gilks WP, Latchman DS, Harvey RJ, Dallapiccola B, Auburger G, Wood NW (2004) Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science 304:1158–1160

Wang WYS, Barratt BJ, Clayton DG, Todd JA (2005) Genomewide association studies: theoretical and practical concerns. Nat Rev Genet 6:109–118

West AB, Maraganore D, Crook J, Lesnick T, Lockhart PJ, Wilkes KM, Kapatos G, Hardy JA, Farrer MJ (2002) Functional association of the parkin gene promoter with idiopathic Parkinson's disease. Hum Mol Genet 11:2787–2792

Wirdefeldt K, Gatz M, Schalling M, Pedersen NL (2004) No evidence for heritability of Parkinson disease in Swedish twins. Neurology 63:305–311

Yasuhara T, Shingo T, Kobayashi K, Takeuchi A, Yano A, Muraoka K, Matsui T, Miyoshi Y, Hamada H, Date I (2004) Neuroprotective effects of vascular endothelial growth factor (VEGF) upon dopaminergic neurons in a rat model of Parkinson's disease. Eur J Neurosci 19:1494–1504

Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, Kachergus J, Hulihan M, Uitti RJ, Calne DB, Stoessl AJ, Pfeiffer RF, Patenge N, Carbajal IC, Vieregge P, Asmus F, Muller-Myhsok B, Dickson DW, Meitinger T, Strom TM, Wszolek ZK, Gasser T (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44:601–607