Complete Genomic Screen in Parkinson Disease
Evidence for Multiple Genes

William K. Scott, PhD; Martha A. Nance, MD; Ray L. Watts, MD; Jean P. Hubble, MD; William C. Koller, MD; Kelly Lyons, PhD; Rajesh Pahwa, MD; Matthew B. Stern, MD; Amy Colcher, MD; Bradley C. Hiner, MD; Joseph Jankovic, MD; William G. Ondo, MD; Fred H. Allen, Jr, MD; Christopher G. Goetz, MD; Gary W. Small, MD; Donna Masterman, MD; Frank Mastaglia, MD; Nigel G. Laing, MD; Jeffrey M. Vance, PhD, MD; Margaret A. Pericak-Vance, PhD

PARKINSON DISEASE (PD) IS A NEURODEGENERATIVE DISEASE THAT AFFECTS MORE THAN A HALF-MILLION PEOPLE IN THE UNITED STATES. The economic, social, and emotional burden of PD will increase as the population ages. Controversy has surrounded the etiology of PD, with evidence suggesting that both genetic and environmental factors influence disease risk. Familial aggregation of PD has been observed for decades. Data from family studies, including a recent large study from Iceland, have shown that the sibling recurrence risk ratio ranges from 2 to 10, suggesting that a genetic component to PD exists. However, twin studies have produced conflicting results about the genetic contributions, suggesting that little if any genetic contribution exists in the development of PD.

Previous efforts to identify genetic risk factors for PD have focused primarily on rare, simple autosomal dominant and recessive forms of the disease. Mutations in the α-synuclein gene, located on chromosome 4q, have been shown to be rare causes of autosomal dominant, early-onset PD. Mutations in the parkin gene, located on chromosome 6q, have been reported in familial PD.

Objective To identify genetic risk factors for idiopathic PD.

Design, Setting, and Participants Genetic linkage study conducted 1995-2000 in which a complete genomic screen (n=344 markers) was performed in 174 families with multiple individuals diagnosed as having idiopathic PD, identified through probands in 13 clinic populations in the continental United States and Australia. A total of 870 family members were studied: 378 diagnosed as having PD, 379 unaffected by PD, and 113 with unclear status.

Main Outcome Measures Logarithm of odds (lod) scores generated from parametric and nonparametric genetic linkage analysis.

Results Two-point parametric maximum parametric lod score (MLOD) and multipoint nonparametric lod score (LOD) linkage analysis detected significant evidence for linkage to 5 distinct chromosomal regions: chromosome 6 in the parkin gene (MLOD=5.07; LOD=5.47) in families with at least 1 individual with PD onset at younger than 40 years, chromosomes 17q (MLOD=2.28; LOD=2.62), 8p (MLOD=2.01; LOD=2.22), and 9q (MLOD=2.39; LOD=1.50) overall and in families with late-onset PD, and chromosome 9q (MLOD=1.52; LOD=2.59) in families with both levodopa-responsive and levodopa-nonresponsive patients.

Conclusions Our data suggest that the parkin gene is important in early-onset PD and that multiple genetic factors may be important in the development of idiopathic late-onset PD.

JAMA. 2001;286:2239-2244 www.jama.com

See also pp 2245 and 2324.

©2001 American Medical Association. All rights reserved.
lies with rare autosomal recessive juvenile parkinsonism and autosomal recessive early-onset PD.\textsuperscript{8,9} Linkage of several large families with autosomal dominant PD to chromosome 2 has been reported,\textsuperscript{10} but a disease-causing gene remains to be identified. Collectively, these studies have demonstrated genetic effects only for rare single-gene variants of PD. To examine the broader issue of genetic effects in idiopathic PD, we performed a complete genomic screen for linkage analysis in 174 families with PD containing at least 1 affected relative pair.

**METHODS**

**Family Data**

In 1995-2000, we coordinated a 13-center effort selecting multiplex (≥2 individuals diagnosed as having PD) families for genetic studies of PD. History of PD was documented for each family by conducting a standard interview with the proband or a knowledgeable family informant. Diagnostic and exclusion criteria, based on previously published diagnostic criteria for PD,\textsuperscript{11-13} were adopted by all participating clinicians before beginning selection of families.

All participants were examined before enrollment in the study by a board-certified neurologist or a physician assistant trained in neurologic disease and supervised by a neurologist. Participants were classified as affected, unclear, or unaffected based on neurologic examination findings and clinical history. Affected individuals had at least 2 cardinal signs of PD (eg, rest tremor, bradykinesia, and rigidity) and no atypical clinical features or other causes of parkinsonism. Individuals with unclear status had only 1 sign of PD, a history of atypical clinical features, or both. Unaffected individuals had no signs of PD. Excluded from participation were individuals with a history of encephalitis, neuroleptic therapy within the year before diagnosis, evidence of normal pressure hydrocephalus, or a clinical course with unusual features suggestive of atypical or secondary parkinsonism.

Age at onset of PD was self-reported and defined as the age at which the affected individual could first recall noticing one of the primary signs of PD. Because a positive response to levodopa therapy was considered supportive of a diagnosis of idiopathic PD, physician and patient observations of whether symptoms of PD were significantly improved by levodopa therapy were used to classify individuals as responsive or nonresponsive to levodopa.\textsuperscript{12,13} Individuals for whom levodopa was of uncertain benefit or who never received levodopa therapy were classified as having unknown levodopa response. Within family variation in response to levodopa was considered a marker of potential phenotypic and thus genotypic heterogeneity.

To ensure diagnostic consistency across sites, clinical data for all participants were reviewed by a clinical adjudication board, which consisted of a board-certified neurologist with fellowship training in movement disorders (B.L.S.), a dually board-certified neurologist and medical geneticist (J.M.V.), and a certified physician assistant (J.M.S.). Forms with missing data or data inconsistent with the diagnosis assigned to the individual were referred back to the collaborating site for clarification. All participants gave written informed consent before venipuncture and data collection according to protocols approved by each center’s institutional review board.

**DNA Analysis**

Genomic DNA was extracted from whole blood using the Puregene system (Centra Systems, Minneapolis, Minn). Analysis was performed on 344 microsatellite markers with an average spacing of 10 cM (centimorgans). Genotyping was performed by the FAAST method.\textsuperscript{14} All samples in the laboratory were identified only by a sequential 6-digit identification number, which was given to the sample when it was received in the DNA Bank. No names or family relationships were provided to any laboratory technician.

Systematic genotyping errors were minimized using a system of quality control checks with duplicated samples. For each 96-well polymerase chain reaction plate, 2 standard samples from Centre d’Etude du Polymorphisme Humain (Paris, France) families were included and 6 samples were duplicated either on that plate or another plate in the screen. Laboratory technicians not involved in the determination of genotypes performed the placement of these duplicated quality control samples. Thus, the laboratory technicians who read the genotypes were blinded to the location of the matching partner for each quality control sample to avoid bias in interpretation of results. Statistical analysts used automated computer scripts to check each set of genotypes submitted by the technician for mismatches between the duplicated samples; mismatches are indicative of potential genotype reading errors, misloading of samples, and sample mix-ups. These mismatches were then sent as one of a large group of surrounding genotypes for rechecking. Thus, the technician had no knowledge of the actual genotype in question. As an additional quality control measure, potential pedigree errors were checked using the program RELPAIR,\textsuperscript{15} which infers likely relationships between pairs of relatives using identical by descent—sharing estimates from a set of microsatellite markers.

**Statistical Analysis**

Data analysis used a multianalytical approach consisting of both parametric lod score and nonparametric affected relative pair methods. Maximum parametric lod scores (MLODs) for each marker were calculated using the VITESSE and HOMOG program packages.\textsuperscript{16,17} The MLOD is the lod score maximized over the 2 genetic models tested, allowing for genetic heterogeneity. Dominant and recessive low-penetrance (affecteds-only) models were considered. Only individuals with a clear diagnosis of PD were considered affected in these analyses. Estimates of
prevalence of PD range from 0.3% in individuals 40 years or older to 2.5% in individuals 70 years or older. Based on these prevalence estimates and allowing for age-dependent or incomplete penetrance, disease allele frequencies of 0.001 for the dominant model and 0.20 for the recessive model were used. Marker allele frequencies were generated from more than 150 unrelated white individuals.

Multipoint nonparametric lod scores (LODs) were calculated using GENEHUNTER-PLUS. Sex-averaged inter-marker distances from the Marshfield Center for Medical Genetics, Marshfield, Wis, genetic linkage maps (http://www.marshfieldclinic.org/research/genetics/map_markers/maps/indexmapframes.html) were used in these analyses. In contrast to nonparametric linkage approaches that consider allele sharing in pairs of affected siblings, GENEHUNTER-PLUS considers allele sharing across pairs of affected relatives (or all affected relatives in a family) in moderately sized pedigrees. We selected this program because of the additional power contributed to the sample by the 75 affected relative pairs that would be excluded by an affected sibpair analysis. Because of computational constraints on pedigree size, 27 unaffected individuals from 12 families were omitted from this analysis.

Because of the potential genetic heterogeneity in this sample, a priori we stratified the data set into 3 subsets. The early-onset PD subset included 18 families with at least 1 member with early-onset PD (<40 years) (range, 12-66 years). The levodopa-nonresponsive subset included 9 families with late-onset PD that contained at least 1 affected individual who was determined to be nonresponsive to levodopa therapy. The late-onset idiopathic PD subset contained 147 families with late-onset PD.

Traditionally in linkage analysis (particularly of mendelian traits), a LOD of more than 3 (corresponding to 1000:1 odds in favor of linkage) is considered strong evidence for linkage. However, such a threshold is likely too stringent for initial efforts to find complex disease genes; using a lower threshold (such as a LOD of >1) in an initial genomic screen may help ensure that genes with modest effects are not missed. Recent genomic screens have used reduced thresholds for declaring results “interesting” for further study. In this study, only regions generating both MLODs and LODs of more than 1.5 were classified as having interesting results. Although this approach may increase the number of false-positive results that are subjected to more detailed examination, it decreases the more serious possibility of missing a true genetic effect.

RESULTS

All individuals potentially informative for linkage were considered for selection in each family, and all family members sampled at the time of the study were included in this analysis. The families contained an average of 2.3 affected individuals and an average of 1.5 affected relative pairs per family. Although most affected relative pairs were affected sibpairs (185/260), there were 75 other affected relative pairs (19 avuncular, 51 cousin, and 5 parent-child pairs) in the data set, indicating that families were often multigenerational in structure and that the study was not limited to affected sibpairs.

All families studied were white and included 870 individuals (an average of 5 per family). Of these individuals, 378 (43%) were diagnosed as having PD, 379 (44%) were unaffected, and 113 (13%) had unclear status. In affected individuals, the mean (SD) age at onset of PD was 59.9 (12.6) years (range, 12-90 years), and the mean (SD) age at examination was 69.9 (10.2) years (range, 33-90 years). Mean (SD) age at examination in unaffected individuals was 67.1 (12.9) years (range, 31-96 years), and mean (SD) age at examination in those with unclear PD status was 72.1 (11.6) years (range, 49-90 years). Mean age at onset in the families with early-onset PD was 39.7 years (range, 12-66 years), whereas mean age at onset in the families with late-onset PD was 62.7 years (range, 40-90 years). The 2 age-of-onset groups were similar with respect to average family size and structure.

Analysis of the clinical parameters of the collected data set did not differ significantly from the patient data collected for other studies in which the focus was on the enrollment of patients with PD for clinical trials. Genetic regions generating both MLODs and LODs of greater than 1.5 are listed in the Table. Markers on chromosomes 5q, 8p, and 17q generated interesting 2-point and multipoint lod scores (MLODs and LODs of more than 1.5) in the overall sample of 174 families (Figure 1). The strongest evidence for linkage in the overall data set was on chromosome 8p (MLOD=2.01 at D8S520; LOD=2.22). Other regions with interesting 2-point and multipoint results were 5q (MLOD=2.39

©2001 American Medical Association. All rights reserved.
at D5S816; LOD = 1.5) and 17q (MLOD = 1.92 at D17S921; LOD = 2.02).

Figure 2 presents the 2-point MLOD results for the data stratified by age at onset and levodopa response. In the subset of 18 families with early-onset PD, a significant MLOD was obtained at D6S305, located in intron 7 of the parkin gene (MLOD = 5.07). Multilocus analysis confirmed these findings, resulting in a LOD of 5.47. No other regions of the genome generated both MLODs and LODs > 1.5 in the early-onset subset.

The strongest linkage result obtained in the subset containing 147 families with late-onset idiopathic PD was on 17q (MLOD = 2.28 at D17S1293; LOD = 2.62). A second region of interest was located on 8p (MLOD = 1.92 at D8S520; LOD = 1.69).

The 9 levodopa-nonresponsive families produced novel results on chromosome 9q (MLOD = 1.52 at D9S301; LOD = 2.59). An additional region of interest was detected on 3q (MLOD = 1.62 at D3S2460; LOD = 1.54).

COMMENT

The results of this study, to our knowledge the largest complete genomic screen in idiopathic PD, suggest that genetic factors are involved in the etiology of both early- and late-onset PD. These results are in contrast to findings of the twin study by Tanner and colleagues.6 In that study of 161 twin pairs, the authors found similar concordance rates in identical (15.5%) and fraternal (11.1%) twin pairs and concluded that there was no role for genetics in late-onset PD. However, the authors acknowledged that there were limitations to the twin study design, including sample size and time to follow-up.25 Our results suggest that these concerns are warranted.

Analysis of 18 families with at least 1 early-onset PD case resulted in strong evidence for linkage to D6S305, which is located in intron 7 of the parkin gene. Parkin was originally described as the gene responsible for only autosomal recessive juvenile parkinsonism8 and later was implicated in other autosomal recessive, early-onset PD.9 The 18 families with early-onset PD were phenotypically similar to the overall data set, with the only notable difference being earlier age at onset of PD symptoms in at least 1 individual. The 18 families with early-onset PD did not have a recognizable mode of inheritance, contained a wide range of ages at onset (12-66 years), and were a mixture of affected sibling pairs and other affected relative pairs in multiple generations.

Through examination of the parkin gene in these 174 families and an additional 134 multiplex and singleton families, we discovered parkin mutations in 11 early-onset and 7 late-onset idiopathic PD families.26 These
linkage and mutation results indicate that parkin is an important genetic factor in PD and mutations are more prevalent than previously reported. Parkin mutations were not identified in 10 families with early-onset PD included in the genomic screen, indicating that additional loci underlying early-onset PD may exist.

In the 147 families with late-onset PD, the strongest overall evidence for linkage was on chromosome 17q at marker D17S1293, which is about 8 cM from the tau gene. The tau gene encodes a microtubule-associated protein that is expressed in the brain and forms paired helical filaments found in Alzheimer disease and other neurodegenerative disorders. Mutations in the tau gene cause frontotemporal dementia with parkinsonism (FTDP), and a haplotype of single nucleotide polymorphisms (SNPs) in the gene has been associated with progressive supranuclear palsy (PSP). The strict clinical criteria developed for use in this study ensured that these families have idiopathic PD and not atypical forms of parkinsonism. For example, to exclude potential cases of PSP from the sample, individuals with PD had to have asymmetrical motor symptoms at onset, no postural instability with falls early in the disease course, and no supranuclear down- or lateral-gaze palsy. Subjects with potential FTDP were excluded from the PD affected group by requiring the absence of dementia at onset and the presence of asymmetrical onset of motor symptoms. Cognitive status testing was not performed during the initial clinical examinations in these families, and therefore we cannot determine if linkage to this region is associated with development of dementia later in the disease. These data are being collected during follow-up evaluations of these families and will be examined in future studies. Therefore, the evidence for linkage of late-onset PD to chromosome 17q suggests a possible genetic link between FTDP, PSP, and idiopathic PD. We examined intragenic SNPs in tau for association with PD and found that a haplotype of 4 SNPs in the tau gene is significantly associated with increased risk of developing PD. Association of PD with a haplotype of tau and evidence for linkage to that region of chromosome 17q suggest that tau or a gene in linkage disequilibrium with tau is a genetic risk factor for PD.

Stratification by levodopa response identified additional regions of interest. In particular, multipoint LODs on chromosome 9q were stronger in these 9 families compared with the other families with late-onset idiopathic PD. Each family included both levodopa-responsive and nonresponsive members, indicating that variable response to levodopa within the family might be part of the phenotype associated with a susceptibility locus in this region. This region of 9q contains the taurina gene, deletions in which are responsible for causing idiopathic torsion dystonia, another movement disorder that is not responsive to levodopa therapy. These data suggest a potential etiologic connection between dystonia and non-levodopa-responsive parkinsonism.

Family-based genetic linkage studies are frequently used as initial attempts to identify susceptibility genes in complex diseases. This study design requires a selected sample of families with multiple related individuals diagnosed as having the disease. Only 10% to 20% of individuals with PD report a family history of the disease, and only a subset have a living relative with PD. Because of this ascertainment scheme, these multiplex families may not be representative of all cases of PD. Therefore, the results of this study may not apply to all PD cases and cannot be used to calculate the relative contribution of each gene to the overall risk of PD. However, we found no significant differences in the clinical presentation of patients in this data set and those seen in a general PD clinic population. Once susceptibility genes underlying each of these regions of linkage are identified, investigations of population-based random samples will be necessary to determine the contribution of each to the overall risk of idiopathic PD.

These results provide strong evidence that the parkin gene is influential in the development of early-onset PD, that several genes may influence the development of late-onset PD, and that age at onset and levodopa response pattern may be useful discriminators for genetic etiology. Like many complex traits, it is likely that PD is caused by an interaction of genetic and environmental risk factors, in which specific genetic templates are more susceptible to the influences of environmental exposures. Further studies to identify the molecular pathways affected by the responsible genes will provide valuable insight into this complex etiology and potential treatment for PD.
Critical revision of the manuscript for important intellectual content: W. K. Scott, Nance, Watts, Hubble, Koller, Lyons, Pahwa, Stern, Colcher, Hiner, Jankovic, Ondo, Allen, Goetz, Small, Masterman, Mastrogi glia, Laing, Stajich, Slatterbeck, Booze, Rampersaud, West, Middleton, Roses, Haines, B. L. Scott, Vance, Pericak-Vance.


REFERENCES