ORIGINAL ARTICLE

Genome-wide scan linkage analysis for Parkinson's disease: the European genetic study of Parkinson's disease

M Martinez, A Brice, J R Vaughan, A Zimprich, M M B Breteler, G Meco, A Filla, M J Farrer, C Bétard, J Hardy, G De Michele, V Bonifati, B Oostra, T Gasser, N W Wood, A Dürr, the French Parkinson's Disease Genetics Study Group, the European Consortium on Genetic Susceptibility in Parkinson's Disease

....

See end of article for authors' affiliations

....

Correspondence to:
Dr Maria Martinez, Unité de Recherche INSERM EMI00-06, Tour Evry 2, 523 Place des Terrasses de l'Agora, Evry cedex 91068, France; maria@evry.inserm.fr

....

Objective: To undertake a full genome-wide screen for Parkinson's disease susceptibility loci.

Methods: A genome-wide linkage study was undertaken in 227 affected sibling pairs from 199 pedigrees with Parkinson's disease. The pedigree sample consisted of 188 pedigrees from five European countries, and 11 from the USA. Individuals were genotyped for 391 microsatellite markers at ~10 cM intervals throughout the genome. Multipoint model-free affected sibling pair linkage analyses were carried out using the MLS (maximum LOD score) test.

Results: There were six chromosomal regions with maximum MLS peaks of 1 or greater (pointwise $p < 0.018$). Four of these chromosomal regions appear to be newly identified regions, and the highest MLS values were obtained on chromosomes 11q (MLS = 1.60, at 91 cM, D11S4175) and 7p (MLS = 1.51, at 5 cM, D7S531). The remaining two MLS peaks, on 2p11–q12 and 5q23, are consistent with excess sharing in regions reported by other studies. The highest MLS peak was observed on chromosome 2p11–q12 (MLS = 2.04, between markers D2S2216 and D2S160), within a relatively short distance (~17 cM) from the PARK3 region. Although a stronger support of linkage to this region was observed in the late age of onset subgroup of families, these differences were not significant. The peak on 5q23 (MLS = 1.05, at 130 cM, D5S471) coincides with the region identified by three other genome scans. All peak locations fell within a 10 cM distance.

Conclusions: These stratified linkage analyses suggest linkage heterogeneity within the sample across the 2p11–q12 and 5q23 regions, with these two regions contributing independently to Parkinson's disease susceptibility.

P arkinson's disease is the second most common neurodegenerative disorder after Alzheimer's disease, affecting 2% of the population above 65 years of age. It is characterised by resting tremor, rigidity, slowness of movement, and postural instability. The main pathological hallmark of this disorder is a pronounced loss of dopamine producing neurones within the substantia nigra. In most cases, Parkinson's disease is a sporadic disorder of unknown aetiology. Familial cases do, however, exist, as do several rare Parkinson's disease families displaying a Mendelian mode of transmission, either dominant or recessive. The identification of rare monogenic variants has contributed substantially to our understanding of the underlying molecular mechanisms.

To date, eight loci have been implicated in monogenically inherited forms of Parkinson's disease, but only four of the causative genes have been identified. Four loci explain recessive forms with early onset: PARK2 (parkin gene) on chromosome 6q25[1 2]; and PARK9,[3] PARK6 (PINK1 gene),[4 5] and PARK7 (DJ-1 gene)[6 7] on chromosome 1p36 and 1p36–p35, respectively. Two genes have been identified in Mendelian forms with dominant transmission: PARK1/PARK4 (α synuclein gene), and PARK5 (ubiquitin carboxy-terminal hydrolase L1 gene), located on chromosome 4q21 and 4p14 respectively.[8–10] The remaining loci have been mapped on chromosomes 1p32 (PARK10), 2p13 (PARK3), and 12p11 (PARK8),[11–13] mostly under a presumed dominant model. However, the genetic aetiology of the common forms of familial Parkinson's disease remains mostly unknown. One locus (PARK11) on chromosome 2q36 has recently been

implicated,[14] and PARK3 has been suggested to affect age of onset.[15]

To obtain further evidence for genetic susceptibility loci, five genome linkage screens have been undertaken in four large independent cohorts of sibling pairs and small nuclear families. All but one dataset originated from the USA. Although these studies failed to provide definite evidence for linkage to any locus, they suggested several positive regions.[11 16–19] Most of the regions with the best evidence for linkage were not consistent across studies. However, some other regions with less evidence for linkage appear to have been reported by at least two study groups. For instance, non-parametric LOD scores of >1 have been reported on chromosome 5q in three datasets.[11 17 18]

Because of the large number of tests in genome-wide approaches, positive findings require confirmation in other independent panels of families originating from different geographical areas and using similar methods for ascertainment, disease definition, and statistics. We therefore undertook a full genome-wide screen for Parkinson's disease susceptibility loci using model-free affected sibling pair methods in our European and American collection of 199 multiplex pedigrees, ascertained through a definite Parkinson's disease case.

....

Abbreviations: ASP, affected sibling pair; IBD, identical by descent; LOD, logarithm of odds; MLS, maximum LOD score; UPDRS, unified Parkinson's disease rating score

METHODS
Families

Multiplex families were ascertained through an index case with a definite clinical diagnosis of Parkinson's disease originating from five European countries (France, United Kingdom, Netherlands, Germany, and Italy), and from the USA. The same clinical criteria were used for inclusion of the European and American families. All but two of the American families were of European descent; the two remaining being Hispanic from Chile. Index cases were collected from specialised movement disorders clinics.

Diagnostic criteria were established as follows:

- *Clinically definite Parkinson's disease*: the presence of at least three of four cardinal signs (bradykinesia, rigidity, rest tremor, asymmetry of signs of onset); at least 30% improvement with levodopa treatment; and absence of exclusion criteria (supranuclear gaze palsy, Babinski sign, cerebellar signs, dyspraxia, prominent and early urinary symptoms, mini-mental state examination (MMSE) score of less than 24/30 within two years of onset);
- *Probable Parkinson's disease*: three of the four cardinal signs, or two of the cardinal signs and at least 30% improvement with levodopa treatment, and absence of exclusion criteria;
- *Possible Parkinson's disease*: doubtful diagnosis, with two of four cardinal signs and unclear response to levodopa treatment.

In all countries the individuals with Parkinson's disease were examined and interviewed by a movement disorder specialist in a hospital or at their home, using a standardised protocol. All index cases were videotaped. Most families with transmission of the disease compatible with autosomal recessive inheritance and age of onset before 45 years in at least one patient in the family were tested for parkin mutations. Forty families carrying mutations in the parkin gene were excluded before the genome scan study. For comparison of frequencies and means among patients from different countries we used Pearson's χ^2 tests and analysis of variance (ANOVA) running on SPSS software. The ''on'' and ''off'' UPDRS (unified Parkinson's disease rating score)[20] and Hoehn and Yahr[21] scores were defined as follows: ''on'' means L-dopa treatment; ''off'' means after having interrupted L-dopa treatment from the evening before evaluation.

Molecular analyses

Genotyping was carried out at the Centre National de Génotypage. The linkage marker set MD 10 (Applied Biosystems, Foster City, California, USA) formed the core marker set for the genome-wide screen. These 391 microsatellite markers, labelled with fluorescent dyes (FAM[TM], HEX[TM], NED[TM]), are distributed at an average marker density of 10 cM (roughly every 10 million bases in the genome) and have an average heterozygosity of 75%. The internal size standard is fluorescently labelled with a fourth dye (ET-ROX[TM] 400, Amersham Biosciences, Amersham, Buckinghamshire, UK). CNG has developed a protocol allowing the co-amplification of up to six of these markers in a single reaction to be robust using dual 384-well GeneAmp® PCR 9700 cyclers (Applied Biosystems) and an automated procedure for polymerase chain reaction (PCR) and purification set-up. All PCR reaction mixes were prepared with a 96-tip head Automation Partnership BasePlate liquid handling robot. PCR were carried out with 4 µl of DNA (diluted to 5 ng/µl) + 6 µl of PCR mix. The PCR fragments obtained were pooled and purified before separation on automatic sequencers. All the steps to perform the pooling and G50 purification were done using a 96-tip head

Automation Partnership BasePlate robot. Two microlitres of the purified product were transferred to a 96-well plate and mixed with 3 µl of MegaBace loading cocktail (for one reaction: 2.75 µl H_2O+0.25 µl ET-ROX[TM] 400). The purified dye labelled fragments were separated according to size on Amersham Biosciences MegaBACE 1000 96-capillary sequencers. After injection (45 seconds at 3 kV) the samples were run for 65 minutes at 10 kV, using data collection software (Instrument Control Manager, version 2.1).

Genotype interpretation and quality control

After collection, raw data from runs were transferred to an independent team for blinded analysis. First, automatic genotyping was undertaken, based on a series of software processes implemented in Genetic Profiler software (version 1.1) applied to the raw MegaBACE data: trace processing, fragment sizing, allele calling, and assigning genotype quality scores. Briefly, DNA fragment sizes are estimated by identifying the peak intensities for each sample using internal DNA standards of known fragment sizes (ET-ROX[TM] 400). Fragments of similar estimated size for the same markers in the 96 samples of a single run are clustered, providing an allele assignment. A quality score is assigned to each genotype, based on a measure of the deviation of the observation from the median value for the allele, weighted by the standard deviation of the distribution over all observations. Before statistical analysis, rigorous genotype quality assurance was carried out to ensure accurate binning of alleles. Consistency of the data with Mendelian inheritance and lack of recombination between loci was evaluated using PEDCHECK[22] and other purpose written software. The genetic position of the markers was determined from the sex-average map of the Marshfield Research Foundation. We estimated marker allele frequencies from the data using the computer program ILINK from VITESSE.[23]

Statistical linkage methods

All linkage analyses were carried out using a narrow definition of disease—that is, subjects with definite or probable Parkinson's disease were coded as affected, whereas all remaining subjects were considered as having an unknown disease phenotype. Multipoint model-free affected sibling pair (ASP) linkage analyses were undertaken using the MLS LOD score test,[24] as implemented in the SIBS qualitative trait mapping option of the GENEHUNTER v2.1 program,[25] which provides likelihood based test statistics for linkage. At any chromosomal location, the likelihood L of the observed marker information among affected sibling pairs is maximised as a function of the proportion of pairs sharing two, one, or no alleles identical by descent (IBD), and is compared—through a likelihood ratio test—with the likelihood of the marker data under the null hypothesis of no linkage:

$$T = 2 \ln[L(z_2, z_1, z_0)/L(\tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4})]$$

The likelihood is maximised under the possible triangle constraints among the z parameters ($2z_0 \leqslant z_1$ and $z_1 \leqslant 1/2$), and the resulting distribution of T is a mixture of χ^2 results with one and two degrees of freedom.[26] The statistic can also be reported as a LOD score: $MLS = T/2 \, en(10)$. For X chromosome marker data, the likelihood ratio test is a function of one parameter only:

$$T = 2 \ln L(z_1, 1-z_1)/L(\tfrac{1}{2}, \tfrac{1}{2}))$$

The resulting distribution of T is a χ^2 distribution with one degree of freedom. As creating all possible ASPs from families with more than two (r>2) affected siblings may underestimate the p values in the far tail of the statistic distribution,[27–29] we computed multipoint MLS scores using all possible pairs (r(r−1)/2) and weighted the information provided by each pair by 2/r.[30] Because the X chromosome

Table 1 Clinical characteristics of subjects with definite and probable Parkinson's disease in 199 European and American families

	Country						
	DE	FR	IT	NL	UK	US	Total
Pedigrees (n)	30	72	28	26	32	11	199
Parkinson's disease relatives (n)	70	181	63	58	77	22	471
Definite/probable/possible cases (n)	67/2/1	158/9/14	55/6/2	55/3/0	76/0/1	23/7/2	424/27/20
Non-Parkinson's disease siblings examined (n)	39	77	24	46	63	0	249
Age at examination, number of cases	62	162	56	56	67	20	424
Mean (SD) (years)	70.4 (8.5)	67.5 (10.8)	69.4 (8.4)	67.7 (9.2)	68.1 (8.5)	71.9 (11.0)	68.5 (9.7)
Range (years)	45 to 87	32 to 86	50 to 86	50 to 89	48 to 86	46 to 87	32 to 89
Age of onset, number of cases	61	160	56	56	67	18	416
Mean (SD) (years)	56.7 (10.3)	57.5 (12.2)	59.7 (9.1)	59.3 (9.5)	57.7 (10.3)	60.1(11.8)	58.4 (11.0)
Range (years)	34 to 77	24 to 80	38 to 84	40 to 84	31 to 80	38 to 80	24 to 84
UPDRS +off, number of cases	10	49	12	£NA	18	2	91
Mean (SD) score	48.6 (16.2)	31.8 (18.5)	46.2 (24.4)		64.3 (18.4)	55 (75)	**44.2**** (23.3)
UPDRS ++on, number of cases	28	119	44	56	44	7	298
Mean (SD) score	27.1 (12.6)	23.3 (14.6)	34.0 (16.1)	24.4 (6.9)	43.1 (21.6)	35.4 (21.6)	**28.7**** (16.5)
Hoehn &Yahr* off, number of cases	NA	49	3	NA	NA	NA	52
Mean (SD) score		2.8 (1.3)	2.5 (0.5)				2.7 (1.2)
Hoehn &Yahr* on, number of cases	NA	126	54	56	NA	NA	236
Mean (SD) score		2.4 (0.99)	2.98 (0.96)	2.5 (0.8)		NA	2.5 (0.97)
Parkinsonian triad							
Tremor at rest (n)	43/61	125/161	44/57	50/56	62/67	11/19	**80%*** (335/421)
Rigidity (n)	52/61	153/162	57/57	55/56	60/67	15/18	**93%*** (392/421)
Bradykinesia (n)	58/61	161/162	57/57	52/56	67/67	17/17	**98%*** (412/420)
Dose of L-dopa, number of cases	58	154	52	51	67	16	398
Mean (SD) dose (mg)	507 (263)	476 (231)	636 (358)	406 (196)	771 (518)	382 (470)	**538**** (347)

*Hoehn and Yahr score for severity of handicap (Hoehn and Yahr, 1967): +off means after having interrupted L-dopa treatment from the evening before evaluation; ++on means with L-dopa treatment.
*p<0.01, **p<0.001 for between country comparisons.
DE, Germany; FR, France; IT, Italy; NA, not assigned; NL, Netherlands; UK, United Kingdom; UPDRS, united Parkinson's disease rating scale, motor subscore (Fahn and Elton, 1987).

version of SIBS/GENEHUNTER program is not functional, multipoint MLS values for X marker data were computed using the LOD score from the affected-only sharing method,[31] as implemented within ALLEGRO.[32]

For positive linkage regions (MLS >1), potential genetic heterogeneity in our Parkinson's disease data was investigated by means of the predivided χ^2 homogeneity test.[33] Linkage homogeneity of two distinct samples can be tested by computing twice the difference between the likelihood of the whole sample and the summed likelihoods of the subsamples (D). Under the null hypothesis of linkage homogeneity, D is asymptotically distributed as χ^2 with n−1 degrees of freedom, where n is the number of estimated parameters. Because of the triangle constraints in the MLS computations, D is a mixture of two χ^2 distributions with two and three degrees of freedom, respectively. Significance values of the predivided χ^2 test were therefore assessed through simulations. From the total sample of families, each family was assigned by random drawing to one of the two subsamples. Multipoint MLS were computed in each simulated subsample and the corresponding value of D recorded. The procedure was repeated 500 times. The significance of the test was obtained by the number of times D exceeded the observed value over the total number of replicates.

Using Nyholt's calculations,[34] MLS thresholds of 1, 1.5, 2, and 3 are associated, in our study, to pointwise nominal p values of 0.026, 0.0075, 0.0022, and 0.0002, respectively. For our chromosome X analyses, the corresponding pointwise nominal p values are 0.016, 0.0043, 0.0012, and 0.0001, respectively, as the likelihood is a function of only one parameter. The theoretical genome-wide threshold values required for claiming ''suggestive'' and ''significant'' linkage in our study are MLS = 2.19 (2.45) (nominal p = $7.4 10^{-4}$) and MLS = 3.63 (3.93) (nominal p = $2.2 10^{-5}$), respectively, for autosomal (X) linkage analyses.[34][35] These thresholds are derived assuming an infinitely dense map of markers and

independent sibling pairs with parental data. They also vary according to the allele sharing method.

Our genome scan study is based on a sparse marker map, parental marker data are often missing, and some families have multiple affected siblings. Thus to clarify the interpretation of our linkage results, we used simulated data to obtain empirical genome-wide significance levels. One thousand replicates of the genome were generated under the null hypothesis of no locus influencing the disease. In all our simulations the family structures, number of subjects genotyped, marker allele frequencies, and marker locations were maintained as in the observed data. Each replicate simulated genome was subject to SIBS/GENEHUNTER multipoint linkage analysis and the number of multipoint peaks at or above a given MLS threshold was stored. Distinct peaks in the distributions of multipoint MLS scores were defined as being at least 30 cM apart. The genome-wide significance was obtained by deriving the expected number of given MLS per genome screen. Empirical pointwise MLS thresholds at a given map position were derived as the proportion of replicates with at least one MLS above a given value.

The power of the present genome scan study was estimated by a similar approach. Replicate datasets were simulated under the assumption of tight linkage (recombination fraction = 0) between a marker (with five equally frequent alleles) and a disease locus. Different genetic models were considered by varying the relative sibling risk value, from 2 to 4. These figures were based on the reported estimates found in common forms of familial Parkinson's disease. The power of the study to detect a particular MLS score was estimated as the proportion of replicates with at least that MLS value.

RESULTS

In all, 213 European or American pedigrees with a definite case of Parkinson's disease were ascertained. Thirteen of
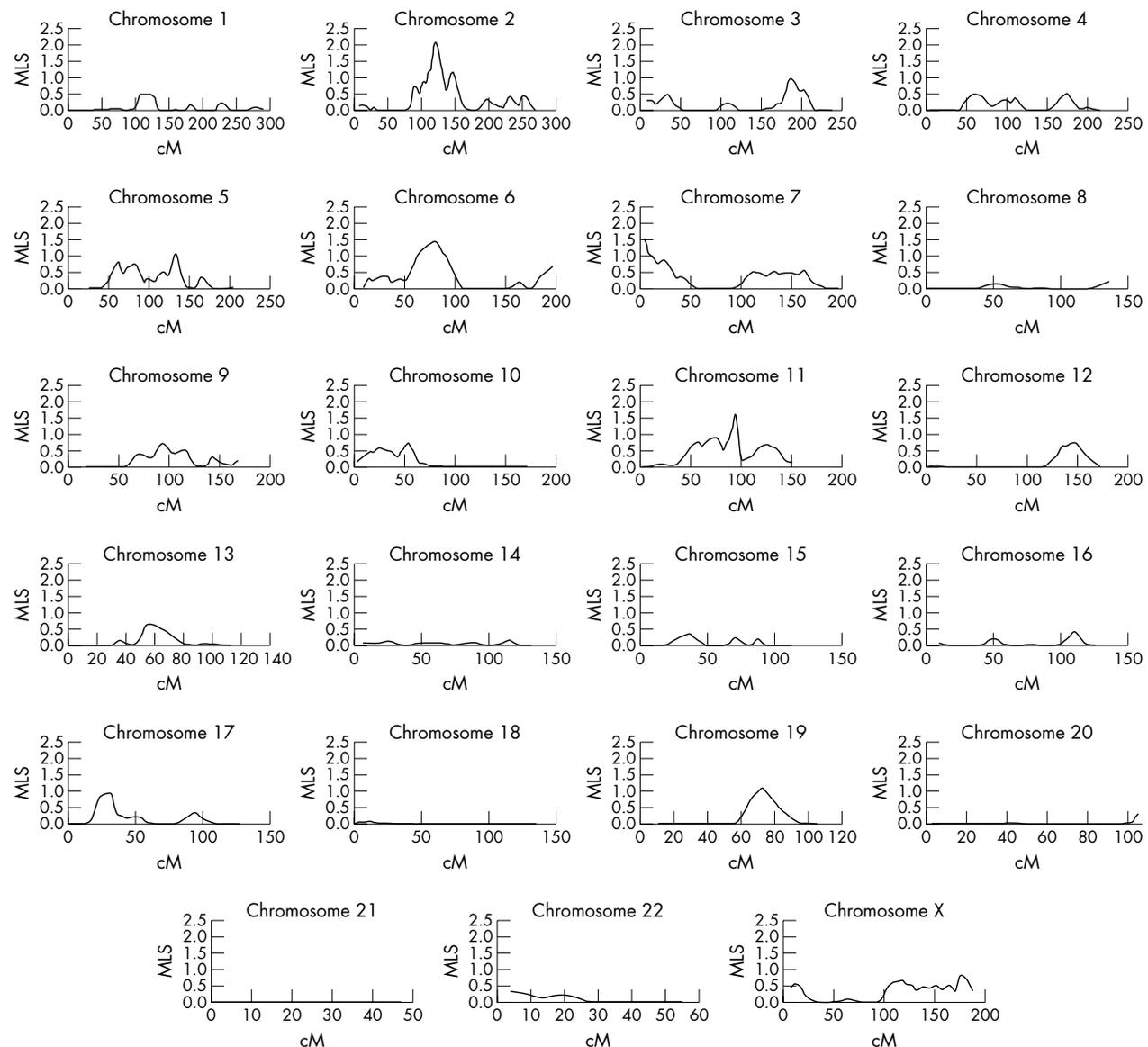
Figure 1 Genome-wide scan linkage results: multipoint LOD score (MLS) values at each point location for each chromosome.

these pedigrees were excluded because no secondary definite or probable Parkinson's disease case with DNA was available. In all but one of the remaining 200 multiplex pedigrees, the index case had at least one sibling with definite or probable Parkinson's disease and with DNA available.

The detailed clinical characteristics of all patients examined from the 199 pedigrees are given in table 1 according to the country of recruitment. The total number of cases was 471, with 424 definite phenotypes (90%), 27 probable (6%), and 20 possible (4%). Age of onset was known for 416 patients with either definite or probable Parkinson's disease. The mean (SD) age of onset was 58.4 (11.0) years (range 24 to 84) and of examination, 68.5 (9.7) years (range 32 to 89, n = 424), without significant differences among the countries. The UPDRS[20] ''on'' (p<0.001) and ''off'' (p<0.001), as well as the Hoehn and Yahr score[21] ''on'' (p<0.01), determined for a subgroup of patients, were significantly different among countries. This showed that the severity of the disease varied among countries, despite the similar age at examination and age of onset. This difference probably

reflects the small numbers of patients with Parkinson's disease in whom UPDRS and Hoehn and Yahr scores were determined. The frequency of tremor at rest, ranging from 53% in the USA to 93% in the UK (p<0.001), and that of rigidity ranging from 75% in the USA to 100% in Italy (p<0.01), were significantly different among countries. This reflects the variability of the predominant cardinal sign— tremor or rigidity—and was not used to differentiate subgroups of patients for the analysis.

The genome scan was carried out in the 199 multiplex pedigrees including at least two affected siblings or half-siblings genotyped (with definite or probable Parkinson's disease). The total number of subjects was 1159, including 735 with DNA available (422 affected and 313 unaffected). The multiplex pedigrees contained an average of 5.8 subjects and 2.3 affected individuals. The total number of ASPs informative for linkage was 222 (208 independent) and 227 (213 independent) for autosomal and X linkage analyses, respectively. The other affected relative pairs were four and 19 avuncular and cousins, respectively. Of the 422 affected

Table 2 Genome scan affected sibling pair linkage results: maximum multipoint LOD score (MLS) values >1

Chromosome and marker region	Position (cM)*	MLS (% IBD sharing)†
Chromosome 2		
D2S2216	111	1.24 (56%)
Peak	117	2.04 (58%)
D2S160	123	1.77 (57%)
Chromosome 5		
D5S471	130	1.05 (57%)
Chromosome 6		
D6S257	80	1.37 (56%)
Peak	85	1.41 (55%)
D6S460	90	1.14 (52%)
Chromosome 7		
D7S531	5	1.51 (57%)
Chromosome 11		
D11S4175	91	1.60 (56%)
Chromosome 19		
D19S902	73	1.05 (56%)

*Sex average position derived from Marshfield Research Foundation.
†Estimated value of IBD sharing data (IBD = $z_2 + \frac{1}{2}z_1$).
IBD, identical by descent.

cases with DNA, age of onset was known for 387. Mean age of onset in families ranged from 29 to 80 years (average mean age of onset in families = 58.3 years).

With a pointwise MLS threshold of >1, the power of our study ranged from 59% to 96% when the sibling relative risk varied from 2 to 4. However, with more conservative significance thresholds (that is, MLS = 3), our power to detect effect sizes of sibling relative risk 4 and 2 was only 59% and 10%, respectively. Simulation analyses under the null hypothesis of no linkage showed that multipoint MLS thresholds of 1, 1.5, and 2 were associated to pointwise p values of 0.018, 0.004, and 0.0015, respectively. Further, simulation studies of our genome scan suggested that we would expect to obtain, on average, one multipoint MLS of 1.62 per genome scan, while an MLS of 2.78 would be expected to occur once in every 20 genome scans in the absence of linkage. These therefore correspond to "suggestive" and "significant" thresholds for genome-wide significance. They are lower than the recommended thresholds obtained with an infinitely dense marker map and in a sample of 100 independent sibling pairs.[35] The empirical thresholds are specific to our study, given the family structures, subjects with DNA, marker allele frequencies, and marker positions as observed in our data.

Genome scan MLS results are shown in fig 1. We identified six chromosomal regions with an MLS >1 (table 2)—on chromosomes 2 (MLS = 2.04, position = 117 cM, between markers D2S2216 and D2S160); 11 (MLS = 1.60, position = 91 cM at D11S4175); 7 (MLS = 1.51, position = 5 cM at marker D7S531); 6 (MLS = 1.41, position = 85 cM between markers D6S257 and D6S460); 5 (MLS = 1.05, position = 130 cM at D5S424); and 19 (MLS = 1.05, position = 73 cM at D19S202).

We further carried out stratified linkage analyses in these six chromosomal regions to search for genetic heterogeneity within our Parkinson's disease family sample. Our first goal was to determine whether the evidence favouring linkage to these regions depends on the family mean age of onset. We used 60 years—the median of the observed mean age of onset distribution—as the cut off age to define two subgroups of families. Age of onset was unknown in 11 American families. Of the remaining pedigrees, 95 ("early") have a mean age of

onset less than 60 years. The "late" subgroup included the families with mean age of onset of 60 years or more (n = 93). Excess of allele sharing was observed in both subgroups and none of the χ^2 homogeneity tests reached significance (results not shown). The most pronounced differences were observed on chromosome 7, where excess allele sharing was higher in the "late" subgroup (IBD = 61%, MLS = 1.70) than in the "early" one (IBD = 53%, MLS = 0.15), but these differences were not statistically significant (p = 0.22).

We then examined the potential role of a Parkinson's disease susceptibility locus on chromosome 2, specifically its contribution to the excess allele sharing observed in the other five linkage regions. Families were stratified by their estimated allele sharing rate (IBD_Chr2) in the 2p11–q12 region delimited by D2S2216 and D2S160. To identify the most positively linked families, we used 75% as the cut off value of IBD_Chr2. This subgroup was denoted "Chr2_linked" (n = 84 families). The remaining families (n = 115) were combined into the "Chr2_unlinked" subgroup. Mean age of onset in the two subgroups was comparable, at 59.47 (37 to 80 years) and 57.04 (29 to 75 years), respectively. Table 3 shows the linkage heterogeneity results, with the MLS values for the two subgroups. As is evident from table 3, estimates of excess allele sharing on chromosomes 7, 6, 11, and 19 were not significantly different in the subgroups, the significance being proportional to the sample size. Evidence for linkage reached significance between the subgroups on chromosome 5 (p = 0.046). Evidence for linkage was greater in the "Chr2_unlinked" subgroup (MLS = 2.19, IBD = 61% at 78 cM) than in the "Chr2_linked" subgroup (MLS = 0.81, IBD = 58% at 133cM).

DISCUSSION

Our genome-wide scan has identified six chromosomal regions positively linked to Parkinson's disease, with LOD score (MLS) values of more than 1. Two positive findings, on chromosomes 2p11–q12 and 5q23, show evidence consistent with excess sharing in regions reported by other genome scans of Parkinson's disease. Based on our simulation genome-wide thresholds, we obtained one region with "suggestive" evidence for linkage (that is, MLS ⩾1.62). The expected genome-wide occurrence expectation, corresponding to our strongest MLS of 2.04, is 0.34. This means that one would expect to see an MLS of 2.04 once in every 2.9 genome scans. The finding on chromosome 11 (MLS = 1.60) fell just short of genome-wide significance for suggestive linkage; the estimated genome-wide occurrence expectation of an MLS = 1.60 is 1.03.

The highest MLS value was observed on chromosome 2p11–q12 at 117 cM between D2S2216 and D2S160. The same region was reported as the most significant in a sample of 103 multiplex families by the GenePD study[15] when analysing age of onset in Parkinson's disease (multipoint LOD score = 2.08, position = 99 cM at D2S1777). Our peak is observed at a more centromeric location but remains within a relatively short distance (17 cM) to the PARK3 region.[12 36] For complex disorders, simulation studies have shown how far from the peak the true susceptibility gene may be located.[37] Thus the scatter in terms of position of linkage peaks from our own and other groups in the PARK3 region is probably a reflection of the high variability of linkage peaks in sibling pair studies.

Interestingly, the GenePD study further reported significant association for age of onset with a more telomeric marker D2S1394 (~90 cM) and at 10 cM of their linkage peak. The results suggested that later age of onset was associated with one allele of D2S1394. This allele was also present in the haplotype shared by the patients in two large pedigrees showing linkage to PARK3.[36] More recently,

Table 3 Linkage homogeneity testing by excess allele sharing at 2p11–q12

Chromosome	"Chr2_linked" (n = 84)	"Chr2_unlinked" (n = 115)	χ^{2}* (empirical p value)
Chromosome 5			
MLS (%IBD)	0.81 (58)	2.19 (61)	8.98 (0.046)
Location (cM) (marker)	133 (D5S471)	78 (D5S647)	
Chromosome 6			
MLS (%IBD)	0.69 (57)	0.86 (57)	0.64 (>0.50)
Location (cM) (marker)	92 (D6S460)	76 (D6S257)	
Chromosome 7			
MLS (%IBD)	0.72 (58)	1.01 (57)	1.01 (>0.50)
Location (cM) (marker)	133 (D7S486)	5 (D7S531)	
Chromosome 11			
MLS (%IBD†)	2.08 (61)	1.07 (59)	7.14 (0.13)
Location (cM) (marker)	77 (D11S937)	127 (D11S4151)	
Chromosome 19			
MLS (%IBD)	0.46 (56)	0.60 (56)	0.05 (>0.50)
Location (cM) (marker)	73 (D19S902)	73 (D19S902)	

*χ^{2} statistic for homogeneity of linkage among the two subgroups. Likelihood for the combined sample in table 2.
†Proportion of alleles shared IBD by affected siblings.
IBD, identical by descent.

significant associations with other polymorphisms have been identified in the GenePD data, but to a younger age of onset.[38] The fine mapping study of 20 single nucleotide polymorphisms (SNPs) of genes located in the PARK3 region revealed an association with SNP rs1876487, located in the predicted promoter region of the sepiapterine reductase gene (SPR). The more rare allele (T) was significantly associated with a younger age of onset. If both true association and linkage exist, then higher estimates of IBD rates are expected in the subset of affected relatives from a specific tail of the age of onset distribution. Evidence for linkage should vary with family mean age of onset. The mean age of onset of Parkinson's disease in our family sample (58 years) is very similar to those reported in other genome scan studies and the GenePD study (60 years) as well as in the families linked to PARK3 (59 years).[12] Our stratified linkage analyses by family mean age of onset showed, at 2p11–q12, both a higher estimated sharing rate (62% *v* 55%) and MLS score (1.61 *v* 0.39) value in the "late" than in the "early" subgroup of families. Figure 2 plots the corresponding multipoint MLS values around the PARK3 locus and shows that subgroup specific linkage peak locations are ~60 cM distant: MLS = 1.72 at 85 cM near D2S2368 in "late" subgroup; MLS = 1.11 at 142 cM near D2S112 in "early" subgroup. These results tend to favour the hypothesis of a putative PARK3 gene affecting Parkinson's disease with late age of onset. However, these differences were not statistically significant and thus do not confirm a late (or young) age of onset specific genetic effect. More studies are needed to confirm or refute this hypothesis in other independent
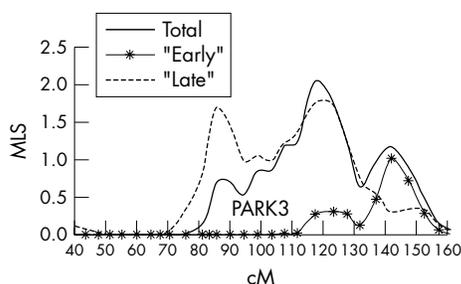

Figure 2 Multipoint MLS scores in the total and subgroup of families with "early" and "late" mean age of onset around PARK3 locus.

samples of Parkinson's disease cases, especially in samples of unrelated cases. So far, the reported associations to age of onset have been identified in a single dataset, showing positive linkage to age of onset variability. Appropriate statistical inference of association testing in the presence of linkage using related cases might be challenging especially when parental marker data are missing and when few unaffected siblings with DNA are available, as it is the case in most Parkinson's disease families.[39]

Three other genome scan studies for Parkinson's disease have reported linkage to the same chromosome 5q23.[11 17 18] Peak locations were all reported within a 10 cM distance. The Duke study[17] and the deCODE study in Iceland[11] reported peaks at D5S816 (139 cM, multipoint LOD score = 1.50) and at D5S666 (135 cM, multipoint LOD score = 1.6) respectively, which are ~9 and 5 cM telomeric to our peak at D5S471 (130 cM). The Parkinson Study Group[18] reported a peak (multipoint LOD score = 0.5 to 1.6) near our same marker D5S47.

Overall, except for PARK3, no evidence for linkage was observed in our family sample to any of the reported PARK loci (fig 1). Our peak on chromosome 6p11 does not overlap the PARK2 region (the parkin gene is at 6q34, ~100 cM to D6S257).

Our genome screen suggests four new regions—on chromosomes 11q, 7p, 6p, and 19q—with MLS >1. The strongest results were obtained on chromosomes 11q and 7p. None of these regions has been reported in other genome-wide linkage studies of Parkinson's disease.[11 16–19] We note, however, that the Duke genome scan study for Parkinson's disease age of onset[40] reported a multipoint LOD score of 1.88 at D6S1017, which is less than 30 cM apart from our peak. The chromosome 19 peak maps at the position of apolipoprotein E, and several studies have reported an association between Parkinson's disease and apolipoprotein E2.[41] This may therefore underlie this linkage peak.

From our linkage homogeneity testing, we have shown that the reported linkages in the six identified regions are not significantly explained by a specific mean age of onset of Parkinson's disease. Stratified analyses by evidence for linkage on 2p11–q12 suggested that, in our Parkinson's disease sample, 2p11–q12 and 5q23 linkage regions may independently contribute to Parkinson's disease susceptibility. Interestingly, the 5q23 region encompasses the location of the synphilin-1 gene, a protein shown to interact in vivo

with α synuclein and parkin, and a component of Lewy bodies, the pathological hallmark of Parkinson's disease.[42]

In conclusion, our genome-wide scan study has implicated six chromosomal regions in Parkinson's disease susceptibility. None of the identified regions satisfies genome-wide statistical criteria (p<0.05) for ''significant'' linkage. Nevertheless, these results merit further investigation, especially the two regions—2p11–q12 and 5q23—where excess allele sharing has now been consistently reported in several independent Parkinson's disease family datasets. Two challenging issues for genome scan studies of complex disorders are the well known multitest problem with the likely high rate of false positive findings but also the lack of replication of a specific linkage result.[43] None of the published genome scan studies found linkage to these regions at the genome-wide significance of p = 0.05, although the studies were based on a similar and substantial number of affected sibling pairs (~200). Nonetheless, there is agreement that replication studies are essential to distinguish true from false positive linkage findings. Thus consistent replications of specific linkage areas—even at relatively modest significance levels—are, in our opinion, of major interest. In our view, there is now sufficiently compelling evidence to undertake detailed mapping studies of these two regions. It is also clear that much larger datasets are probably required to find statistical evidence for significant linkage. To this end, we are developing the follow up study of the 5q23 area, with linkage analysis of additional microsatellite markers as well as association testing of polymorphisms in positional candidate genes, in collaboration with the Parkinson Study Group. The analyses from this global genetic consortium will be done in the largest Parkinson's disease family sample to date (more than 700 affected sibling pairs), and other groups are invited to participate. All genotypic data that will be generated by this consortium will be made available in a public setting to the scientific community.

ACKNOWLEGEMENTS

. . . .

Authors' affiliations

M Martinez, INSERM EMI00-06, Evry, France

A Brice, A Dürr, INSERM U289 and Département de Génétique, Cytogénétique et Embryologie, Hôpital La Pitié-Salpêtrière, Paris, France

J R Vaughan, Department of Neurology Charing Cross Hospital, London, UK

A Zimprich, T Gasser, Department of Neurodegenerative Diseases, Hertie-Institute for Clinical Brain Research, University of Tübingen, Germany

M M B Breteler, Department of Epidemiology and Biostatistics, Erasmus University, Rotterdam, Netherlands

G Meco, V Bonifati, Department of Neurological Sciences, La Sapienza University, Rome, Italy

A Filla, G De Michele, Department of Neurological Sciences, Federico II University, Naples, Italy

M J Farrer, Neurogenetics Laboratory, Mayo Clinic Jacksonville, Florida, USA

C Bétard, Centre National de Génotypage, Evry, France

J Hardy, Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, Maryland, USA

B Oostra, Department of Clinical Genetics, Erasmus University, Rotterdam, Netherlands

N W Wood, Department of Molecular Neuroscience, Institute of Neurology, University College London, London, UK

Conflicts of interest: none declared

The French Parkinson's Disease Genetics Study Group: Y Agid, A-M Bonnet, M Borg, A Brice, E Broussolle, P Damier, A Destée, A Dürr, F Durif, J Feingold, G Fénelon, E Lohmann, M Martinez, C Penet, P Pollak, O Rascol, F Tison, C Tranchant, M Vérin, F Viallet, M Vidailhet, and J-M Warter

The European Consortium on Genetic Susceptibility in Parkinson's Disease: N W Wood and D Nicholl (UK); A Brice, A Dürr, M Martinez and Y Agid (France); T Gasser and B Müller-Myhsok (Germany); M Breteler, S Harhangi and B Oostra (Netherlands); V Bonifati, E Fabrizio, N Vanacore, G Meco, G De Michele, G Volpe and A Filla (Italy).

REFERENCES

1 Matsumine H, Saito M, Shimoda-Matsubayashi S, Tanaka H, Ishikawa A, Nakagawa-Hattori Y, Yokochi M, Kobayashi T, Igarashi S, Takano H, Sanpei K, Koike R, Mori H, Kondo T, Mizutani Y, Schaffer AA, Yamamura Y, Nakamura S, Kuzuhara S, Tsuji S, Mizuno Y. Localization of a gene for an autosomal recessive form of juvenile Parkinsonism to chromosome 6q25.2–27. Am J Hum Genet 1997;60:588–96.

2 Abbas N, Lucking CB, Ricard S, Durr A, Bonifati V, De Michele G, Bouley S, Vaughan JR, Gasser T, Marconi R, Broussolle E, Brefel-Courbon C, Harhangi BS, Oostra BA, Fabrizio E, Bohme GA, Pradier L, Wood NW, Filla A, Meco G, Denefle P, Agid Y, Brice A. A wide variety of mutations in the parkin gene are responsible for autosomal recessive parkinsonism in Europe. French Parkinson's Disease Genetics Study Group and the European Consortium on Genetic Susceptibility in Parkinson's Disease. Hum Mol Genet 1999;8:567–74.

3 Hampshire DJ, Roberts E, Crow Y, Bond J, Mubaidin A, Wriekat AL, Al-Din A, Woods CG. Kufor-Rakeb syndrome, pallido-pyramidal degeneration with supranuclear upgaze paresis and dementia, maps to 1p36. J Med Genet 2001;38:680–2.

4 Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, Gispert S, Ali Z, Del Turco D, Bentivoglio AR, Healy DG, Albanese A, Nussbaum R, Gonzalez-Maldonado R, Deller T, Salvi S, Cortelli P, Gilks WP, Latchman DS, Harvey RJ, Dallapiccola B, Auburger G, Wood NW. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science 2004;304:1158–60.

5 Valente EM, Bentivoglio AR, Dixon PH, Ferraris A, Ialongo T, Frontali M, Albanese A, Wood NW. Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35–p36. Am J Hum Genet 2001;68:895–900.

6 van Duijn CM, Dekker MC, Bonifati V, Galjaard RJ, Houwing-Duistermaat JJ, Snijders PJ, Testers L, Breedveld GJ, Horstink M, Sandkuijl LA, van Swieten JC, Oostra BA, Heutink P. Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36. Am J Hum Genet 2001;69:629–34.

7 Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, Dekker MC, 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. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 2003;299:256–9.

8 Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia 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. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 1997;276:2045–7.

9 Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R, Lincoln S, Crawley A, Hanson M, Maraganore D, Adler C, Cookson MR, Muenter M, Baptista M, Miller D, Blancato J, Hardy J, Gwinn-Hardy K. alpha-Synuclein locus triplication causes Parkinson's disease. Science 2003;302:841.

10 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. The ubiquitin pathway in Parkinson's disease. Nature 1998;395:451–2.

11 Hicks AA, Petursson H, Jonsson T, Stefansson H, Johannsdottir HS, Sainz J, Frigge ML, Kong A, Gulcher JR, Stefansson K, Sveinbjornsdottir S. A susceptibility gene for late-onset idiopathic Parkinson's disease. Ann Neurol 2002;52:549–55.

12 Gasser T, Muller-Myhsok B, Wszolek ZK, Oehlmann R, Calne DB, Bonifati V, Bereznai B, Fabrizio E, Vieregge P, Horstmann RD. A susceptibility locus for Parkinson's disease maps to chromosome 2p13. Nat Genet 1998;18:262–5.

13 Funayama M, Hasegawa K, Kowa H, Saito M, Tsuji S, Obata F. A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2–q13.1. Ann Neurol 2002;51:296–301.

14 Pankratz N, Nichols WC, Uniacke SK, Halter C, Rudolph A, Shults C, Conneally PM, Foroud T. Significant linkage of Parkinson disease to chromosome 2q36–37. Am J Hum Genet 2003;72:1053–7.

15 DeStefano AL, Lew MF, Golbe LI, Mark MH, Lazzarini AM, Guttman M, Montgomery E, Waters CH, Singer C, Watts RL, Currie LJ, Wooten GF, Maher NE, Wilk JB, Sullivan KM, Slater KM, Saint-Hilaire MH, Feldman RG, Suchowersky O, Lafontaine AL, Labelle N, Growdon JH, Vieregge P, Pramstaller PP, Klein C, Hubble JP, Reider CR, Stacy M, MacDonald ME,

Gusella JF, Myers RH. PARK3 influences age at onset in Parkinson disease: a genome scan in the GenePD study. *Am J Hum Genet* 2002;**70**:1089–95.

16 **DeStefano AL**, Golbe LI, Mark MH, Lazzarini AM, Maher NE, Saint-Hilaire M, Feldman RG, Guttman M, Watts RL, Suchowersky O, Lafontaine AL, Labelle N, Lew MF, Waters CH, Growdon JH, Singer C, Currie LJ, Wooten GF, Vieregge P, Pramstaller PP, Klein C, Hubble JP, Stacy M, Montgomery E, MacDonald ME, Gusella JF, Myers RH. Genome-wide scan for Parkinson's disease: the GenePD Study. *Neurology* 2001;**57**:1124–6.

17 **Scott WK**, Nance MA, Watts RL, Hubble JP, Koller WC, Lyons K, Pahwa R, Stern MB, Colcher A, Hiner BC, Jankovic J, Ondo WG, Allen FH, Goetz CG, Small GW, Masterman D, Mastaglia F, Laing NG, Stajich JM, Slotterbeck B, Booze MW, Ribble RC, Rampersaud E, West SG, Gibson RA, Middleton LT, Roses AD, Haines JL, Scott BL, Vance JM, Pericak-Vance MA. Complete genomic screen in Parkinson disease: evidence for multiple genes. *JAMA* 2001;**286**:2239–44.

18 **Pankratz N**, Nichols WC, Uniacke SK, Halter C, Rudolph A, Shults C, Conneally PM, Foroud T. Genome screen to identify susceptibility genes for Parkinson disease in a sample without parkin mutations. *Am J Hum Genet* 2002;**71**:124–35.

19 **Pankratz N**, Nichols WC, Uniacke SK, Halter C, Murrell J, Rudolph A, Shults CW, Conneally PM, Foroud T. Genome-wide linkage analysis and evidence of gene-by-gene interactions in a sample of 362 multiplex Parkinson disease families. *Hum Mol Genet* 2003;**12**:2599–608.

20 **Fahn SER**, UPDRS program members. Unified Parkinson's Disease Rating Scale. In: Fahn SMC, Goldstein M, Calne DB, eds. *Recent developments in Parkinson's disease.* Florham Park: Macmillan Healthcare Information, 1987:153–63.

21 **Hoehn MM**, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology* 1967;**17**:427–42.

22 **O'Connell JR**, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998;**63**:259–66.

23 **O'Connell JR**, Weeks DE. The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. *Nat Genet* 1995;**11**:402–8.

24 **Risch N**. Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 1990;**46**:222–8.

25 **Kruglyak L**, Lander ES. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 1995;**57**:439–54.

26 **Holmans P**. Asymptotic properties of affected-sib-pair linkage analysis. *Am J Hum Genet* 1993;**52**:362–74.

27 **Tores FBE**, Phillippi A, Vaysseix G, Demenais F, Martinez M. Affected sib pair strategies for late-onset diseases: type I error and marker allelic frequencies [abstract]. *Genet Epidemiol* 1996;**13**:A302.

28 **Meunier FPA**, Martinez M, Demenais F. Affected sib-pair tests for linkage: type I errors with dependent sib-pairs. *Genet Epidemiol* 1997;**14**:1107–11.

29 **Davis S**, Weeks DE. Comparison of nonparametric statistics for detection of linkage in nuclear families: single-marker evaluation. *Am J Hum Genet* 1997;**61**:1431–44.

30 **Suarez BK**, Hodge SE. A simple method to detect linkage for rare recessive diseases: an application to juvenile diabetes. *Clin Genet* 1979;**15**:126–36.

31 **Kong A**, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997;**61**:1179–88.

32 **Gudbjartsson DF**, Jonasson K, Frigge ML, Kong A. Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 2000;**25**:12–13.

33 **Morton NE**. The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am J Hum Genet* 1956;**8**:80–96.

34 **Nyholt DR**. All LODs are not created equal. *Am J Hum Genet* 2000;**67**:282–8.

35 **Lander E**, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995;**11**:241–7.

36 **West AB**, Zimprich A, Lockhart PJ, Farrer M, Singleton A, Holtom B, Lincoln S, Hofer A, Hill L, Muller-Myhsok B, Wszolek ZK, Hardy J, Gasser T. Refinement of the PARK3 locus on chromosome 2p13 and the analysis of 14 candidate genes. *Eur J Hum Genet* 2001;**9**:659–66.

37 **Hauser ERBM**. Confirmation of linkage results in affected sib-pair linkage analysis for complex traits [abstract]. *Am J Hum Genet* 1997;**61**:A278.

38 **Karamohamed S**, DeStefano AL, Wilk JB, Shoemaker CM, Golbe LI, Mark MH, Lazzarini AM, Suchowersky O, Labelle N, Guttman M, Currie LJ, Wooten GF, Stacy M, Saint-Hilaire M, Feldman RG, Sullivan KM, Xu G, Watts R, Growdon J, Lew M, Waters C, Vieregge P, Pramstaller PP, Klein C, Racette BA, Perlmutter JS, Parsian A, Singer C, Montgomery E, Baker K, Gusella JF, Fink SJ, Myers RH, Herbert A. A haplotype at the PARK3 locus influences onset age for Parkinson's disease: the GenePD study. *Neurology* 2003;**61**:1557–61.

39 **Martin ER**, Bass MP, Hauser ER, Kaplan NL. Accounting for linkage in family-based tests of association with missing parental genotypes. *Am J Hum Genet* 2003;**73**:1016–26.

40 **Li YJ**, Scott WK, Hedges DJ, Zhang F, Gaskell PC, Nance MA, Watts RL, Hubble JP, Koller WC, Pahwa R, Stern MB, Hiner BC, Jankovic J, Allen FA, Goetz CG, Mastaglia F, Stajich JM, Gibson RA, Middleton LT, Saunders AM, Scott BL, Small GW, Nicodemus KK, Reed AD, Schmechel DE, Welsh-Bohmer KA, Conneally PM, Roses AD, Gilbert JR, Vance JM, Haines JL, Pericak-Vance MA. Age at onset in two common neurodegenerative diseases is genetically controlled. *Am J Hum Genet* 2002;**70**:985–93.

41 **Maraganore DM**, Farrer MJ, Hardy JA, McDonnell SK, Schaid DJ, Rocca WA. Case-control study of debrisoquine 4-hydroxylase, N-acetyltransferase 2, and apolipoprotein E gene polymorphisms in Parkinson's disease. *Mov Disord* 2000;**15**:714–19.

42 **Engelender S**, Kaminsky Z, Guo X, Sharp AH, Amaravi RK, Kleiderlein JJ, Margolis RL, Troncoso JC, Lanahan AA, Worley PF, Dawson VL, Dawson TM, Ross CA. Synphilin-1 associates with alpha-synuclein and promotes the formation of cytosolic inclusions. *Nat Genet* 1999;**22**:110–14.

43 **Suarez BKHC**, Van Eerdewegh P. Problems of replicating linkage claims in psychiatry. In: Gershon ESCC, eds. *Genetic approaches to mental disorders.* Washington DC: American Psychiatric Press, 1994.