Brain (2002), **125,** 1337–1347

A whole genome screen for linkage disequilibrium in multiple sclerosis confirms disease associations with regions previously linked to susceptibility

Stephen Sawcer,[1] Mel Maranian,[1] Efrosini Setakis,[2] Val Curwen,[3] Eva Akesson,[1] Anke Hensiek,[1] Francesca Coraddu,[1] Richard Roxburgh,[1] David Sawcer,[1] Julia Gray,[1] Jackie Deans,[1] Peter N. Goodfellow,[4,5] Neil Walker,[2] David Clayton[2] and Alastair Compston[1]

[1]*University of Cambridge Neurology Unit, Addenbrooke's Hospital,* [2]*MRC Biostatistics Unit, Institute of Public Health, University Forvie Site,* [3]*Human Genome Mapping Project Resource Centre, Hinxton and* [4]*University of Cambridge Department of Genetics, Cambridge, UK*

[5]*Present address: GlaxoSmithKline Pharmaceuticals, Gunnels Wood Road Stevenage, Hertfordshire SG1 2NY, UK*

Correspondence to: Alastair Compston, University of Cambridge Neurology Unit, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK
E-mail: alastair.compston@medschl.cam.ac.uk

Summary

Linkage analysis in multiplex families has provisionally identified several genomic regions where genes influencing susceptibility to multiple sclerosis are likely to be located. It is anticipated that association mapping will provide a higher degree of resolution, but this more powerful approach is limited by the substantial genotyping effort required. Here, we describe the first use of DNA pooling to screen the whole genome for association in multiple sclerosis based on a 0.5 cM map of microsatellite markers and using four DNA pools derived from cases (*n* = 216), controls (*n* = 219) and trio families (*n* = 745 affected individuals and their 1490 parents). The 10 markers showing the greatest evidence for association with multiple sclerosis that emerge from this analysis include three from the HLA region on chromosome 6p (D6S1615, D6S2444 and TNFa), providing a positive control for the method, four from regions previously identified by linkage analysis in UK multiplex families (two mapping to chromosome 17q GCT6E11 and D17S1535; one to chromosome 1p GGAA30B06; and one to 19q D19S585), and three from novel sites with respect to linkage analysis (D1S1590 at 1q; D2S2739 at 2p; and D4S416 at 4q). Our results thus provide further supporting evidence for the candidature of 6p, 17q, 19q and 1p as regions encoding susceptibilty genes for multiple sclerosis. The protocol used in this UK-based study is now being extended to 18 additional sites in Europe in order to search for susceptibility genes shared between populations of common ancestry, as well as those that exert ethnically more restricted effects.

Keywords: DNA pooling; multiple sclerosis; linkage disequilibrium

Abbreviations: LMS-HD5 = high density linkage mapping set; MMD = monocyte to macrophage differentiation; SNP = single nucleotide polymorphism

Introduction

Multiple sclerosis is a typical complex trait, globally affecting ~1 million people. Familial clustering is well established (Sadovnick *et al*., 1988; Robertson *et al*., 1996), with λ_s taking a value of ~10–15 (Sadovnick *et al*., 1988; Robertson *et al*., 1996; Carton *et al*., 1997). Concordance rates in twins (Ebers *et al*., 1986; Mumford *et al*., 1994), together with recurrence risks in adoptees (Ebers *et al*., 1995), half-siblings (Sadovnick *et al*., 1996) and conjugal pairs (Robertson *et al*., 1997; Ebers *et al*., 2000), indicate that familial clustering results, at least in part, from shared genetic factors. As in other complex diseases, results obtained from whole genome linkage screens suggest that effects attribut-

able to individual susceptibility genes are modest, such that studies based on association are likely to be considerably more powerful than those employing linkage (Risch and Merikangas, 1996).

Although tests for association have been extensively employed in the analysis of candidate genes, these are only now being considered in the context of systematic whole genome screening. Two conceptually different approaches, the 'direct' and 'indirect', can be considered (Collins *et al*., 1997). In a direct screen, all functionally relevant and common polymorphisms of every gene are tested, whereas in an indirect screen, only a uniformly spaced map of markers is considered. The indirect approach is more efficient but relies on linkage disequilibrium between the markers tested and the susceptibility genes being sought. The extent of linkage disequilibrium in the human population in general, and in relation to disease susceptibility alleles in particular, is critically important in determining the density of markers required to provide adequate power in an indirect screen. Crude theoretical modelling of human population history suggests that variants which are common in the population as a whole will generally be very old and therefore accompanied by rather little linkage disequilibrium (Kruglyak, 1999). In Europeans, however, empirical evidence shows that linkage disequilibrium is in fact much more extensive than predicted, with average distances ranging from 60 kb (Reich *et al*., 2001) to 385 kb (Collins *et al*., 2001) for single nucleotide polymorphism (SNP), and 0.5 Mb (Eaves *et al*., 2000) to 1 Mb (Kendler *et al*., 1999) for microsatellites. Empirical observations have also confirmed the expected stochastic nature of linkage disequilibrium, which is highly variable across the genome and therefore not uniformly accessible to indirect association screening (Dunning *et al*., 2000; Martin *et al*., 2000; Taillon-Miller *et al*., 2000). Others have pointed out that the number of markers required in an indirect screen is not only dependent upon the extent of linkage disequilibrium, but is further increased by confounding factors such as allele frequency mismatch between susceptibility genes and linked markers (Muller-Myhsok and Abel, 1997; Weiss and Terwilliger, 2000). Even in Europeans, where the extent and nature of linkage disequilibrium favour the detection of disease associations, it is clear that many thousands of markers will be required, and the task of genotyping these in several hundred individuals is likely to prove generally prohibitive unless highly efficient methods are employed. Barcellos *et al*. (1997) have proposed just such a method, employing DNA pooling of samples from cases, unrelated controls and nuclear families together with a panel of microsatellite markers that is sufficiently dense to ensure that, on average, no gene is >0.25 cM from its nearest marker. Several groups have shown that typing pooled DNA provides an accurate and reliable method for detecting differences in allele frequencies between cases and controls (Pacek *et al*., 1993; Barcellos *et al*., 1997; Daniels *et al*., 1998; Shaw *et al*., 1998; Breen *et al*., 1999; Fisher *et al*., 1999; Collins *et al*., 2000; Kirov *et al*., 2000). Here, we report the first indirect screen of the genome for association in a complex human disease based on typing a dense map of microsatellite markers in pooled DNA from cases, controls and family trios.

Material and methods

Patients and controls

The 961 unrelated patients considered in this study were recruited in the UK and are all Caucasian. They were classified using the Poser criteria (Poser *et al*., 1983) and diagnosed with either clinically definite (87%), laboratory supported definite (6%) or clinically probable (7%) multiple sclerosis, and also meet revised diagnostic procedures for establishing the diagnosis (McDonald *et al*., 2001). 225 were index cases from multiplex families studied previously in our linkage genome screen (Sawcer *et al*., 1996); 688 were index cases from independent trio families (an affected individual with both parents living); and 48 were local sporadic cases. DNA was available from both parents of the index case in 57 of the 225 multiplex families, and these were therefore used to boost the total number of available trio families to 745. The remaining 216 patients were used to form the cases pool, which thus includes the 48 local sporadic cases together with the 168 index probands from multiplex families where DNA was not available from both parents. Risch and Teng (1998) have shown that including patients from multiplex families increases the power of tests for association. The over-representation of patients with a family history of the disease in these pools will therefore have increased the power of our study. The basic demographic details of the patients are unremarkable: mean age was 41 years, mean expanded disability status score (EDSS) 4.5, mean disease duration 14 years, mean age at diagnosis 33 years, and female : male sex ratio 2.8 (709 female and 252 male). In order to ensure high power for common genetic factors, inclusion criteria were not restricted by disease course. In total, 58% of patients had a relapsing remitting course (RRMS), 31% had secondary progressive disease (SPMS) and 11% had primary progressive multiple sclerosis (PPMS) at the time of assessment. A total of 1709 individuals (219 unrelated blood donors together with the 1490 parents from the 745 trio families) were used as controls.

Ethical approval for the study was given by the Eastern Medical Research Ethics Committee (MREC).

DNA pooling

All individuals involved in this study gave written consent for genetic analysis. DNA was extracted from a sample of venous blood using standard methods. DNA concentration was measured (using a 1 : 20 dilution of the sample in a 1 ml cuvette) with an Ultraspec 3000 spectrophotometer (Pharmacia Biotech, Cambridge, UK). Based on these measurements, samples were diluted to convenient concentrations depending on the yield of DNA. The four separate

Table 1 *Markers typed on each chromosome*

Chromosome	Length (cM)	Number of markers	
		All	Successful*
1	263	609	579
2	255	441	406
3	214	394	364
4	203	289	256
5	194	299	267
6	183	370	334
7	171	312	280
8	155	236	214
9	145	244	229
10	144	268	237
11	144	343	330
12	143	310	300
13	114	137	127
14	109	205	199
15	106	201	191
16	98	167	161
17	92	246	227
18	85	102	95
19	67	216	199
20	72	148	126
21	50	60	57
22	56	110	99
X	160	293	252

*This column indicates the number of markers giving usable allele image profile data in the population screen, the trio family screen or both.

DNA pools were then constructed using 2 μg of DNA from each index case within trio families (n = 745), 2 μg of DNA from each parent in these families (n = 1490), and 20 μg of DNA from each of the remaining cases (n = 216) and controls (n = 219). These four pools are identified in the text as 'Index', 'Parents', 'Cases' and 'Controls', respectively. Each was extensively vortexed and mixed. Pools were then subjected to a full re-extraction process in order to 'clean up' the DNA and thereby reduce any variation in the extent of polymerase chain reaction (PCR) amplification from each individual. Each pool was diluted to 50 ng/μl after re-extraction.

Markers

We used the 811 microsatellite markers in Applied Biosystems (Foster City, USA) high density linkage mapping set (LMS-HD5) as a framework for this project. The additional 5189 markers were identified from the Co-operative for Human Linkage Centre (CHLC) database, the Genethon database (Gyapay *et al*., 1994; Dib *et al*., 1996) and the Genome database (GDB). The total number of markers studied on each chromosome (Table 1) was chosen to reflect the number of genes mapping to that chromosome according to the National Center for Biotechnology Information web site in December 1999. The density of markers is thus biased towards gene-rich and against gene-poor chromosomes (Antonarakis, 1994). Within each, we attempted to make uniform the distribution of markers according to their location scores as derived from the genetic location database (LDB) gmaps (Collins *et al*., 1996). This process necessarily restricted marker selection to those microsatellites included on the LDB. Where multiple markers shared the same or similar location scores, preference was given to tetra- and trinucleotides over dinucleotides, and to those with higher estimated heterozygosity. Despite these manoeuvres, there was persistent non-uniformity of the resulting maps, with 323 gaps of >2 cM and 11 gaps of >10 cM, most corresponding to pericentromeric regions, particularly on acrocentric chromosomes, and sections of the genome generally containing very few genes. Except for the LMS-HD5 markers, where Applied Biosystems has in many cases redesigned the primers, sequences for the markers included in our study were derived from the originating databases. In summary, 4723 (79%) of the markers were dinucleotides, 1019 (17%) tetranucleotides and 258 (4%) trinucleotides. The estimated average heterozygosity (according to data in the originating databases) of the 6000 markers was 70%. The forward primer was labelled with FAM (blue) in 2065 (34%) markers, HEX (green) in 1958 (33%) and NED (red) in 1976 (33%). One marker was labelled with TET (yellow). All but five markers included reverse primer tailing in order to reduce plus A artefact.

In order to explore the possibility that some apparently distinct markers might in fact be amplifying the same repeat, we used the primersearch program to screen the forward primer sequence (and its complement) for each marker against all the sequences from the following EMBL database subsections (as at December 1999): STS (n >90 000 sequences, >32 Mb), human EST (n >1.6 million sequences, >600 Mb) and human (n >86 000 sequences, >450 Mb). Those EMBL sequences matching the forward primer were then screened against the reverse primer sequence, or its complement, as appropriate. Exact sequence matches were found for 5924 markers. A further 18 were matched by relaxing the sequence comparison to allow a single mismatch in each primer. Most of the identified EMBL sequences only matched a unique primer pair. However, 118 sequences

Table 2 *Results for the 10 associated markers*

Marker	Screen*	Refining†
GGAA30B06	0.0261	0.0454
D1S1590	0.0008	0.0460
D2S2739	0.0227	0.0003
D4S416	0.0481	0.0354
D6S1615	0.0010	0.0002
D6S2444	0.0163	0.0392
TNFa	0.0273	0.0085
GCT6E11	0.0277	0.0048
D17S1535	0.0953	0.0058
D19S585	0.0921	0.0267

*The hypothetical P values from the screening phase. †The hypothetical P values from the refining phase.

matched two or more primer pairs. Closer analysis of these 118 sequences showed that, in most cases, the separate matching primer pairs were amplifying different repeats. In total, only 19 markers had to be dropped as they were found to be amplifying repeats already recognized by other primer pairs. These were replaced. In total, primer sequences for the 5189 markers completing our screening set represent >210 kb of sequence. Thus, given an average SNP rate of one per kilobase, we would expect 210 markers to have their primers based on alleles which are rare/uncommon in the UK. We thus anticipated that PCR amplification would probably fail in 210 cases, but in fact a total of 478 markers gave no usable signal, suggesting that failure was unrelated to null allele problems in 268 instances. The net success rate for PCR in the 5189 markers was thus 95%; success rate for the carefully redesigned LMS-HD5 markers was 100%, as expected. No primer optimization was performed and all markers were amplified using the same standard PCR protocol (as described below). Details of the markers used, including primer sequences, are available via the GAMES (genetic analysis of multiple sclerosis in Europeans) web site.

PCR

PCR was performed in 15 μl reactions using 9 μl of True Allele™ PCR Premix (Applied Biosystems), 5 pmol of labelled forward primer, 5 pmol of unlabelled reverse primer, 25 ng of template (pooled) DNA and 6 μl of MilliQ water. Thermal cycling was performed on an Applied Biosystems 9700 thermal cycler according to the following protocol: 12 min at 95°C (to activate the Taq Gold DNA polymerase); 10 cycles of 94°C for 15 s, 55°C for 15 s and 72°C for 30 s; 20 cycles of 89°C for 15 s, 55°C for 15 s and 72°C for 30 s; and a final step of 72°C for 20 min.

Capillary electrophoresis

The products from each PCR were electrophoresed twice on an Applied Biosystems 3700 Genetic Analyser (Applied Biosystems). Prior to electrophoresis, 1 μl of PCR product was denatured for 5 min at 95°C in combination with 9.55 μl of HiDi Formamide and 0.45 μl of GS400HD ROX labelled size standard (Applied Biosystems). Denatured samples were electrokinetically injected with 10 000 V applied for 10 s, while capillary electrophoresis used Performance Optimised Polymer 6 (Applied Biosystems), a cuvette temperature of 35°C, run temperature of 50°C, run voltage of 6000 V and run time of 4500 s. In the screening phase, each PCR product was electrophoresed in isolation, while in the refining phase, some multiplexing by colour was performed. Multiplexing by size was not used in either stage. The allele image profiles from each replicate were analysed using Applied Biosystems GENESCAN (version 3.5) and GENOTYPER (version 3.6) software. Peak height rather than area was quantified as this has been shown to be the more sensitive and specific measure (Breen *et al*., 1999).

In order to determine the linear range of the Applied Biosystems 3700, and to rescue data from allele image profiles where some peaks exceeded the limit of linearity, we diluted the PCR products of markers producing allele image profiles with peak heights >10 000 in one or more replicates, and these were re-electrophoresed. Comparison of allele image profiles from dilute and original replicates for each particular sample enabled expected signals to be calculated for the largest peaks and compared with those observed. This analysis showed that the fluorescent signal remained linear up to ~11 000. Beyond this level, the relationship showed discontinuity and the signal was unreliable. Applied Biosystems has shown that signal increases monotonically up to ~15 000, but variance in the detected signal becomes increasingly large. When the amount of DNA is even greater, the fluorescent signal may actually decrease. However, these saturated signals have a distorted and easily recognizable allele image profile appearance. Peaks of this type were ignored in our analysis of linearity and only those with a signal of <10 000 were considered in testing for association. Where peaks exceeded this limit in some replicates, the corresponding allele frequency was determined only from those replicate allele image profiles where the peak was 10 000.

Statistical analysis

Typing a microsatellite in pooled DNA generates an allele image profile (Daniels *et al*., 1998) consisting of a series of product fragments, with the strength of the signal from each reflecting the frequency of the corresponding allele. Various PCR artefacts confound this relationship, but these tend to affect each pooled sample similarly (see Discussion). In the population screen, each marker was amplified once by PCR and electrophoresed twice, so that two replicate allele image profiles were generated from each sample. Best-fit allele image profiles were calculated for both samples, normalized according to the total number of alleles in the respective pools and compared statistically using a χ^2 test. The same analysis was performed in the trio family screen, except that the control allele image profile was calculated by subtracting the normalized best-fit allele image profile for index cases from the normalized best-fit allele image profile for parents; the control allele image profile thus reflects the non-transmitted allele frequency distribution and the test for linkage disequilibrium is equivalent to the AFBAC (affected family base controls) test suggested by Thomson (1995). Alleles with a frequency of <5% were considered together.

In order to select markers for further investigation, we ranked the observed statistics according to their evidence for association. This task was complicated by the fact that the additional sources of variance introduced by pooling led to deviation of the distribution of these statistics away from the theoretical χ^2 form. In order to assign hypothetical P values, and thereby rank the results, we first grouped the markers according to the number of alleles considered (1 + the degrees

of freedom for the χ^2 test). In each group we then used probability plotting methods to compare the empirical distribution of test statistics with the theoretical χ^2 form that would have applied if we had carried out individual typing. In each case we saw over-distribution, as expected, but also observed deviation of the empirical distribution from the χ^2 shape. Plotting logarithms of the test statistics against logarithm of the expected χ^2 order statistic gave acceptably linear plots, suggesting that a power transformation of the tests, with an appropriate multiplication factor, could be used to transfer empirical distributions to the χ^2 form. We used this approach in each group of markers in order to calculate hypothetical *P* values (a multiplication factor of 2/3 was required for each group of markers in each screen) and thereby enable ranking of markers across all groups.

The deviation of the observed distribution from the theoretical χ^2 form was particularly pronounced in the trio family screen, especially for large values of the test statistics. Clearly, pooling added new categories of variance and, paradoxically, limited power in the trio family screen despite benefits usually associated with the use of a larger sample size (see Discussion). The need to take the difference between parent and index estimates in order to derive control allele frequencies also contributed to loss of power seen in the trio family screen; however, this approach is preferable to estimating control frequencies directly from the parents alone without subtraction, thereby halving any effect and leading to even greater loss of power.

Pilot study

In order to confirm the findings of Barcellos *et al*. (1997) and to validate the construction of our DNA pools, the 216 cases and 219 controls were individually typed for the microsatellite markers TNFa (a dinucleotide), D19S49 (a dinucleotide), D7S1791 (a trinucleotide) and D7S1821 (a tetranucleotide). Individual typing was performed using the semi-automated GENESCAN/GENOTYPER system, with PCR amplification using fluorescently labelled forward primers followed by slab gel electrophoresis on an Applied Biosystems 373A genetic analyser. Consistency of allele calling was ensured by running samples from an individual with known genotype in every twelfth lane. Allele frequencies established by this individual typing were then compared with those estimated by typing the pooled DNA samples. This required only two genotypings, compared with 435 per marker for the individual typing. After correcting for the effects of stutter bands and length-dependent amplification (Barcellos *et al*., 1997), estimated allele frequencies for the dinucleotide markers (TNFa and D19S49) showed no statistically significant difference from those observed by individual typing in either cases or controls. The same results were obtained for the trinucleotide D7S1791, where only correction for length-dependent amplification was required, and for the tetranucleotide D7S1821, where no correction was necessary. Analysis of the TNFa microsatellite provided a positive control for the pooling approach, since the 121 bp allele of this marker is known to be in linkage disequilibrium with DRB1*1501 (Kirk *et al*., 1997; Coraddu *et al*., 1998), the HLA class II allele associated with multiple sclerosis and the most likely candidate gene responsible for the increased susceptibility attributable to this region. Individual typing showed that frequency of the TNFa 121 bp allele was increased in cases (28%) compared with controls (20%; $P = 0.047$). Expected excess of the 121 bp allele was seen in all 15 replicate genotypings of the case and control DNA pools performed with the TNFa marker and, even without correction, the differences were statistically significant in 11 out of 15. This result illustrates that, even without correcting for artefacts, comparison of allele image profiles from separately pooled cases and controls is capable of identifying microsatellite markers in linkage disequilibrium with alleles increasing susceptibility to multiple sclerosis. It should be noted that although the association of multiple sclerosis with DRB1*1501 is well known, the effect of this allele is modest, and probably only increases the risk of disease by a factor of 2–4. Since the 121 bp allele is one of the longer alleles of TNFa, the power of pooled DNA typing to detect its association with multiple sclerosis is reduced by length-dependent amplification. Thus, typing this marker and analysing the result without correction represents a stringent test of the method.

Results

Whole genome screening phase

Following an adapted version of the method proposed by Barcellos *et al*. (1997) we established a set of 6000 microsatellite markers (Table 1) and performed independent whole genome linkage disequilibrium screens in two sets of pooled samples. The first used DNA separately pooled from 216 cases and 219 unrelated controls (the population screen), and the second employed DNA pooled separately from 745 additional cases and their 1490 parents (the trio family screen). Usable allele image profile data were obtained from 5396 markers in the population screen and 5330 markers in the trio family screen; 478 markers failed in both studies, and another 126 and 192 in the population and trio family screens, respectively.

Under the assumption that a single founder mutation will be responsible for the majority of susceptibility alleles at any given locus, it would be expected that genuinely associated markers would show the same allele image profile difference in each screen. We therefore performed a combined analysis of all data from both screens in order to identify the most promising markers. Comparison of the observed statistics in each screen alone with those expected from the normal distribution showed that the effective sample size for pools in the population screen was very nearly equivalent to their actual size, while the effective sample size in the trio family screen was significantly reduced and approximately equiva-

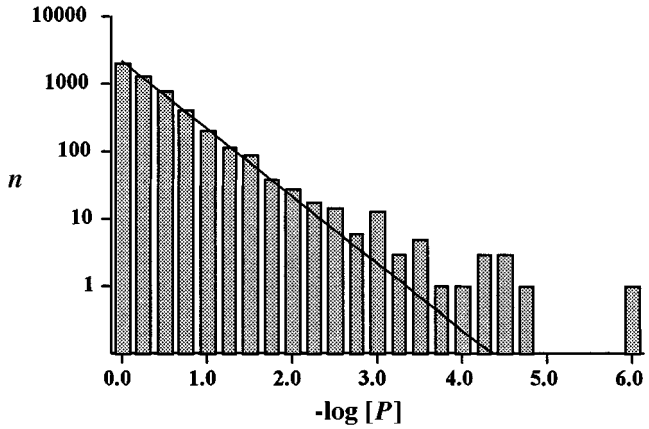

Fig. 1 Frequency histogram of results obtained in the whole genome screening phase (5204 markers). The line indicates the number of markers expected if the statistics were distributed as a χ^2.

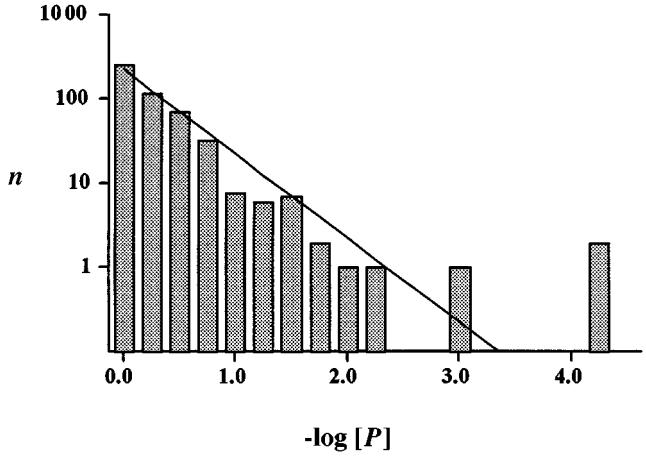

Fig. 2 Frequency histogram of the results obtained in the refining phase (520 markers). The line indicates the number of markers expected if the statistics were distributed as a χ^2. Note that the empirical *P* values assigned to the transformed statistics obtained in this phase were judged relative to the distribution obtained in the whole genome screening phase and are therefore highly conservative.

lent to that in the population screen. Therefore, in the combined analysis, the contribution from trio family samples was weighted to be equal to that from the population samples (ensuring that the trio family screen did not dominate the combined analysis inappropriately). Markers were then ranked according to their evidence for association and the 10% ($n = 520$) showing the strongest effect were considered for more detailed study. The distribution of results with respect to statistical significance judged empirically as described in Material and methods is shown in Fig. 1. Considering each screen separately and comparing the results rather than combining the data in a single analysis revealed 61 markers with *P* values <10% in both studies (slightly more than the number that would be expected by chance alone). All but seven of these are already included in the 520 identified in the combined analysis. None of these additional seven markers showed consistent associations across the screens (as expected), and were therefore not considered further.

Data relating to all of the markers studied is available via the GAMES web site.

Refining the list of associated markers

In order to refine this list of markers each pool was reconstructed *de novo*. The PCR and electrophoresis were then repeated for the 520 'best' markers (in both the newly constructed and original pools) in order to generate four new replicate allele image profiles from each repeat pool together with four new replicate allele image profiles from each of the original pools. A global analysis of these eight allele image profile replicates was then performed for each of the 520 best markers. Since this analysis combined data from independent samples and involved replication of pool construction, PCR and electrophoresis, it was expected that the importance of population allele frequency differences would be enriched. The 520 markers were then re-ranked on the basis of these new results and the refining process repeated for all markers in the top 5% of the refined ranking ($n = 26$). The final distribution of the best 520 markers in this refining analysis is shown in Fig. 2. As expected, the evidence for association decreased in the majority of these markers. In this analysis, the hypothetical *P* values assigned to each marker were judged with respect to the distribution as established in the whole genome screening phase. This is highly conservative, since this refining analysis is based on a greater number of replicate allele image profiles per sample and would therefore be expected to show less variance.

The refining process significantly reduced variance in observed results for these 520 markers by averaging out artefactual effects resulting from errors in pool construction, PCR and electrophoresis. This refinement could also have been achieved by individually typing the samples making up these pools, but this would have been considerably less efficient. Neither additional typing of individuals nor pooled samples can distinguish sampling variance and genuine population frequency differences resulting from true associations. In order to make this distinction, it will be necessary for the most promising markers to be typed in additional cohorts.

Markers associated with multiple sclerosis

Of the 10 markers showing the greatest evidence for association with multiple sclerosis (see Table 2), seven are located within regions of potential linkage identified in the UK linkage screen (GGAA30B06, D6S1615, D6S2444, TNFa, GCT6E11, D17S1535 and D19S585), whereas three represent novel associations at loci not previously implicated in susceptibility to multiple sclerosis (D1S1590, D2S2739 and D4S416).

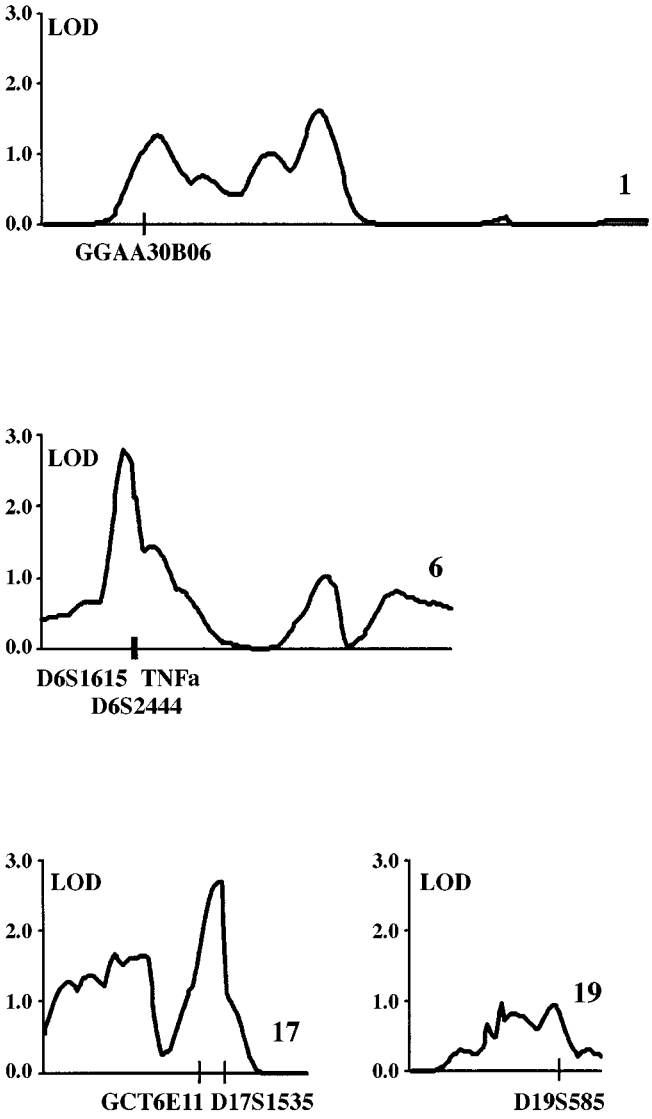

Fig. 3 MLS (maximum LOD score) profiles for chromosomes 1, 6, 17 and 19 obtained in our previously published linkage genome screen (Sawcer *et al*., 1996). The locations of the seven associated markers that lie within regions of potential linkage are indicated by vertical bars.

Three (D6S1615, D6S2444 and TNFa) of the 10 markers are located in the HLA region (6p21) and, in each case, the associated allele is known to be in linkage disequilibrium with the HLA class II allele DRB1*1501. For example, the TNFa marker was deliberately included in the 6000 marker screening set as a positive control because we have previously shown that the 121-bp allele from this marker is in linkage disequilibrium with DRB1*1501 (D′ = 52%; Coraddu *et al*., 1998). The identification of multiple markers from the HLA region confirms that the experiment has the power to identify microsatellites in linkage disequilibrium with known susceptibility alleles of modest effect.

The other seven markers show an equivalent degree of association to those located in the HLA region, i.e. to those in linkage disequilibrium with an established multiple sclerosis susceptibility gene. In addition, four of these markers are located in regions of interest identified in our previous linkage screen (Sawcer *et al*., 1996). Two markers are located under the peak of linkage on chromosome 17q (GCT6E11 and D17S1535); one lies under the peak on chromosome 1p (GGAA30B06); and another is under the peak on chromosome 19q (D19S585) (Fig. 3). The remaining three markers (D1S1590, D2S2739 and D4S416) map to 1p, 2p and 4q, respectively. These have not previously been identified as regions of interest on the basis of linkage genome screens in the UK or other populations.

In order to identify candidate genes mapping close to the associated markers we searched the Ensembl database for 250 kb in both directions from each marker. Six genes with known function were identified, one of which immediately suggested itself as a potential candidate for a multiple sclerosis susceptibility gene: the MMD (monocyte to macrophage differentiation associated) gene, mapping close to GCT6E11 on chromosome 17 (see Discussion). We also noted that the IFITM2 (interferon-induced transmembrane protein 2) gene mapped just beyond 250 kb from D4S416. In addition, nine predicted genes were also identified along with ~20 novel transcripts.

Discussion

We screened the whole genome for evidence of linkage disequilibrium in multiple sclerosis and identified seven novel associations. Our successful identification of microsatellite markers from the HLA region (6p21), known to be in linkage disequilibrium with susceptibility gene(s) of modest effect, validates the approach, in particular the efficiency of using pooled DNA samples. The high concordance between the novel results presented here and those from our previous linkage screen further supports the candidature of chromosomes 1p, 17q and 19q as regions encoding susceptibility genes for multiple sclerosis. All three of these potentially linked regions remained indicative of linkage after the UK linkage data were combined with those from the American (Haines *et al*., 1996) and Canadian (Ebers *et al*., 1996) screens in the recently published meta-analysis (Transatlantic Multiple Sclerosis Genetics Cooperative, 2001). The linkage peak on chromosome 17q is particularly well supported, with additional positive results found in this region in studies from Finland (Kuokkanen *et al*., 1997) and Denmark (Larsen *et al*., 2000), although these results are not found in all populations (Fontaine *et al*., 1999; Broadley *et al*., 2001; Coraddu *et al*., 2001). We anticipate that others will also now concentrate efforts at assessing the candidature of genes contained within these regions and achieve fine mapping using a combination of microsatellites and SNPs in other northern European Caucasian and cross-cultural studies.

In considering the concordance with previous UK linkage results, it is appropriate to point out that the index cases from our linkage study were re-used in this present study, although they represent <25% of the total number of affecteds considered. That said, linkage is an allele-independent test, and therefore the presence of linkage is unlikely to bias tests of association when only one affected member from each family is included, as here. Setting up a PCR requires a finite amount of genomic DNA and therefore places an upper limit on the number of individuals that can be included in a pool. We estimated that a pool consisting of 1470 parents, employing 25 ng of template DNA, was close to this limit. However, analysis of data from our trio family screen indicates that these pools behave as though comprised of significantly smaller numbers of individuals, with consequent reduction of power. Thus, although 745 trio families should in theory offer a resource of considerable statistical power, much of this potential was lost by constructing just two pools (index and all parents) rather than dividing the families into multiple smaller sub-groups and screening each separately. Essentially, it would appear that each PCR originating from these pools behaved as though it was constructed from a subset of the families actually included. Our data would suggest that 200–400 individuals represent a practical maximum for the construction of pools used in screens employing 25 ng of template DNA per PCR.

The major limitation of our study is density of the marker map employed. Even though a significant number of microsatellite markers was chosen, their distribution is not uniform and it remains possible that important susceptibility loci lie in regions not well covered by this set of markers, or in regions lacking significant linkage disequilibrium and hence outside the range of association mapping using the indirect approach. Although we successfully identified HLA susceptibility markers, a number of factors ensured that the power in this region was greater than for the screen as a whole. First, our map was more dense than average in this region; and secondly, linkage disequilibrium in this region is particularly extensive. Both factors would be expected to increase the power of an indirect screen for association. The failure to show association with other markers located within the HLA region illustrates the modest power of our screening phase and the stochastic nature of linkage disequilibrium. For example, we previously typed the marker D6S273 in multiplex families and know that the 138 bp allele is in modest linkage disequilibrium with the DRB1*1501 allele ($D' = 39\%$; Coraddu *et al*., 1998). Global analysis of the allele image profiles from this marker had a P value of 0.18 in the screening phase of the present study, which improved to 0.13 when additional replicates were tested. Although not significant, this marker clearly lies in the tail of the distribution and would be expected to emerge as significant in a more powerful study. In summary, our results from the HLA region clearly demonstrate that separate markers located close to a susceptibility gene exerting only a modest biological effect can perform very differently in association studies dependent on linkage disequilibrium. Microsatellites are significantly more informative than SNPs (Kruglyak, 1997), and are expected to be accompanied by more extensive linkage disequilibrium (Kendler *et al*., 1999). Although our map represents the most extensive that could be constructed when this experiment was designed (in 1999), and comparable whole genome screens cannot yet be performed using SNPs, it remains less dense than some investigators have recommended or considered necessary on theoretical grounds (Risch and Merikangas, 1996; Kruglyak, 1999).

Recent empirical data indicate that the pattern of linkage disequilibrium in Europeans is radically different both in extent and structure from that predicted by Kruglyak (1999). These data have shown that recombination is not uniformly distributed, as assumed, but rather is concentrated in hot spots separating regions of marked linkage disequilibrium (Daly *et al*., 2001; Jeffreys *et al*., 2001). These regions therefore have very limited haplotype diversity and can, in theory, easily be tested for association with appropriately informative markers (Johnson *et al*., 2001). Although modest linkage disequilibrium occurs between adjacent blocks, this is considerably less than within blocks. These blocks of extensive linkage disequilibrium with sharp boundaries vary in size, but are on average ~60 kb (Daly *et al*., 2001), implying that as few as 50 000 informative markers (one per block) could be sufficient reliably to screen the whole genome. It follows that a map employing >5000 microsatellites will provide one microsatellite in every tenth block, and therefore fully screen 10% of the genome and superficially screen a further 20% (the two adjacent blocks). In short, while SNP maps of sufficient density are awaited, a microsatellite map of 5000 markers is powered to find one in three susceptibility genes. If, as expected, several genes influence susceptibility to multiple sclerosis, a low density map such as ours cannot be expected to include markers in linkage disequilibrium with all such genes, other than by chance. Equally, it is unlikely that each of the relevant genes will have been missed. Our study significantly advances knowledge on the genetics of multiple sclerosis by bringing a first pass at whole genome association mapping alongside the existing linkage data and strengthening the status of four chromosomal regions. However, we recognize that is it is not definitive and denser maps need to be constructed when technology and reagents become available.

Many controversies have yet to be resolved concerning the preferred methodology for association studies. The choice of controls has been a source of considerable debate (Risch, 2000). Unrelated controls provide most power but are susceptible to the effects of population stratification and non-random mating patterns, whereas related controls (parents or siblings) protect against these confounding effects but are generally less powerful (Risch and Teng, 1998). We adopted the strategy promoted by Barcellos *et al*. (1997) and employed both approaches, combining the power of unrelated controls with the security of family-based samples. Various artefacts of the PCR, such as length-dependent amplification

(Demers *et al*., 1995), stutter bands (Weber and May, 1989) and plus A contamination (Smith *et al*., 1995), confound the interpretation of data from pooled samples. However, as these phenomena tend to affect cases and controls equally they do not introduce a systematic bias and have a limited negative effect on power (Daniels *et al*., 1998; Shaw *et al*., 1998; Breen *et al*., 1999; Fisher *et al*., 1999; Collins *et al*., 2000; Kirov *et al*., 2000). A number of mathematical corrections for these confounding factors have been suggested (Perlin *et al*., 1995; Kirov *et al*., 2000), but each requires typing individual samples and greatly reduces efficiency of the pooling approach with little dividend in terms of extra power.

Our results add to the growing body of data concerning the effectiveness of DNA pooling (Fisher *et al*., 1999) and whole genome screening for linkage disequilibrium (Lee *et al*., 2001; Uhl *et al*., 2001). Our screen confirms that pooling is an effective strategy that can be used to search for genes influencing susceptibility to complex disease. The protocol is being extended, through international collaboration, to other European and migrant European populations in a genetic analysis of multiple sclerosis in Europeans (GAMES). The results of these 18 additional whole genome linkage disequilibrium screens, each using samples sizes of ~200 cases, ~200 controls and ~200 family trios, will together provide a meta-analysis based on a gross sample of >7500 pooled cases and >11 000 parental or unrelated controls, the results from which are expected to be available during 2002. Combining data from across European populations will effectively increase the sample size by at least an order of magnitude, and make possible the detection of genes with effects too small to be identified in any one population. Since it is well recognized that the power to detect linkage disequilibrium is profoundly affected by the frequency of susceptibility alleles (Risch and Merikangas, 1996; Clayton, 1999), which is likely to vary across Europe, it is expected that some susceptibility factors will only be detectable in particular populations. Within the limitations discussed, whole genome linkage disequilibrium screening using the present protocol in the 18 populations contributing to GAMES will therefore have considerable power to identify regions containing genes conferring susceptibility to multiple sclerosis in all populations, irrespective of ethnicity, and those which exert their effect only in defined populations—the ubiquitous and domestic susceptibility genes, respectively.

Pending that result, the demonstration of associations with three markers encoded within the major histocompatibility region confirms the reliability of whole genome association mapping using microsatellites, and again shows that this region is associated with susceptibility in multiple sclerosis. Our study confirms the involvement of 17q, 19p and 1p, each of which has previously been implicated in susceptibility to multiple sclerosis using linkage analysis, and has revealed an interesting candidate gene worthy of further analysis (the MMD gene). Macrophage activation is a common feature in the pathogenesis of many immune diseases, including multiple sclerosis, and thus the identification of a gene, both by association and linkage, which is implicated in this process combines candidature based on genetic and mechanistic principles. The precise function of the MMD gene is unknown, but analysis of its nucleotide sequence reveals several potential transmembrane domains, indicating that it may act as a receptor or ion channel (Rehli *et al*., 1995). The importance of genes influencing the function and activation of macrophages is supported by the growing body of evidence, indicating that this activity may be responsible for many epi-phenomena that accompany autoimmune disease, such as increased expression of human endogenous retroviral (HERV) elements (Johnston *et al*., 2001). Although not exhaustive, chromosomes 1p, 6p, 17q and 19p can now be listed as the best supported regions of interest—linked and associated with multiple sclerosis—for the detailed analysis of positional candidates, and these positive findings also add weight to the candidature of the three other novel associated markers located at 1q, 2p and 4q.

Electronic database information

URLs for data in this article are as follows: GAMES web site, http://www-gene.cimr.cam.ac.uk/MSgenetics/GAMES (for details of markers used and additional experimental detail); Co-operative for Human Linkage Centre (CHLC) database, http://lpg.nci.nih.gov/CHLC (for original details of markers); Genethon database, http://www.genethon.fr (for original details of markers); Genome database (GDB), http://www.gdb.org (for original details of markers); National Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov (for details concerning the number of genes mapping to individual chromosomes); Genetic location database (LDB), http://cedar.genetics.soton.ac.uk (for details concerning the location of markers used); primersearch program, http://www.uk.embnet.org/Software/EMBOSS/Apps/primersearch.html; EMBL database, http://www3.ebi.ac.uk/Services/DBStats (for details concerning the sequences searched in the selection of primers); Ensembl database, http://www.ensembl.co.uk (for details concerning the genes mapping close to associated markers).

Acknowledgements

We thank B. Reid, R. Hatton, J. Cooper, M. Dronsfield (Applied Biosystems) and Gary Williams (HGMP) for help with technical aspects of the study. We are grateful to members of the Association of British Neurologists and the Multiple Sclerosis Society of Great Britain and Northern Ireland for notifying patients participating in this genetic analysis of multiple sclerosis. The work was supported by the Wellcome Trust (grant 057097).

References

Antonarakis S. Genome linkage scanning: systematic or intelligent? Nature Genet 1994; 8: 211–2.

Barcellos LF, Klitz W, Field LL, Tobias R, Bowcock AM, Wilson R, et al. Association mapping of disease loci, by use of a pooled DNA genomic screen. Am J Hum Genet 1997; 61: 734–47.

Breen G, Sham P, Li T, Shaw D, Collier DA, St Clair D. Accuracy and sensitivity of DNA pooling with microsatellite repeats using capillary electrophoresis. Mol Cell Probes 1999; 13: 359–65.

Broadley S, Sawcer S, D'Alfonso S, Hensiek A, Coraddu F, Gray J, et al. A genome screen for multiple sclerosis in Italian families. Genes Immun 2001; 2: 205–10.

Carton H, Vlietinck R, Debruyne J, De Keyser J, D'Hooghe MB, Loos R, et al. Risks of multiple sclerosis in relatives of patients in Flanders, Belgium. J Neurol Neurosurg Psychiatry 1997; 62: 329–33.

Clayton D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. Am J Hum Genet 1999; 65: 1170–7.

Collins A, Frezal J, Teague J, Morton NE. A metric map of humans: 23,500 loci in 850 bands. Proc Natl Acad Sci USA 1996; 93: 14771–5.

Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. Science 1997; 278: 1580–1.

Collins HE, Li H, Inda SE, Anderson J, Laiho K, Tuomilehto J, et al. A simple and accurate method for determination of microsatellite total allele content differences between DNA pools. Hum Genet 2000; 106: 218–26.

Collins A, Ennis S, Taillon-Miller P, Kwok PY, Morton NE. Allelic association with SNPs: metrics, populations, and the linkage disequilibrium map. Hum Mutat 2001; 17: 255–62.

Coraddu F, Sawcer S, Feakes R, Chataway J, Broadley S, Jones HB, et al. HLA typing in the United Kingdom multiple sclerosis genome screen. Neurogenetics 1998; 2: 24–33.

Coraddu F, Sawcer S, D'Alfonso S, Lai M, Hensiek A, Solla E, et al. A genome screen for multiple sclerosis in Sardinian multiplex families. Eur J Hum Genet 2001; 9: 621–6.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nature Genet 2001; 29: 229–32.

Daniels J, Holmans P, Williams N, Turic D, McGuffin P, Plomin R, et al. A simple method for analyzing microsatellite allele image patterns generated from DNA pools and its application to allelic association studies. Am J Hum Genet 1998; 62: 1189–97.

Demers DB, Curry ET, Egholm M, Sozer AC. Enhanced PCR amplification of VNTR locus D1S80 using peptide nucleic acid (PNA). Nucleic Acids Res 1995; 23: 3050–5.

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, et al. A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 1996; 380: 152–4.

Dunning AM, Durocher F, Healey CS, Teare MD, McBride SE, Carlomagno F, et al. The extent of linkage disequilibrium in four populations with distinct demographic histories. Am J Hum Genet 2000; 67: 1544–54.

Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, et al. The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nature Genet 2000; 25: 320–3.

Ebers GC, Bulman DE, Sadovnick AD, Paty DW, Warren S, Hader W, et al. A population-based study of multiple sclerosis in twins. New Engl J Med 1986; 315: 1638–42.

Ebers GC, Sadovnick AD, Risch NJ, and the Canadian Collaborative Study Group. A genetic basis for familial aggregation in multiple sclerosis. Nature 1995; 377: 150–1.

Ebers GC, Kukay K, Bulman DE, Sadovnick AD, Rice G, Anderson C, et al. A full genome search in multiple sclerosis. Nature Genet 1996; 13: 472–6.

Ebers GC, Yee IM, Sadovnick AD, Duquette P. Conjugal multiple sclerosis: population-based prevalence and recurrence risks in offspring. Canadian Collaborative Study Group. Ann Neurol 2000; 48: 927–31.

Fisher PJ, Turic D, Williams NM, McGuffin P, Asherson P, Ball D, et al. DNA pooling identifies QTLs on chromosome 4 for general cognitive ability in children. Hum Mol Genet 1999; 8: 915–22.

Fontaine B, Cournu I, Arnaud I, Babron MC, EichenbaumVoline S, Oksenberg JR, et al. Chromosome 17q22-q24 and multiple sclerosis genetic susceptibility. Genes Immun 1999; 1: 149–50.

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, et al. The 1993–94 Genethon human genetic linkage map. Nature Genet 1994; 7: 246–339.

Haines JL, Ter-Minassian M, Bazyk A, Gusella JF, Kim DJ, Terwedow H, et al. A complete genomic screen for multiple sclerosis underscores a role for the major histocompatibility complex. Nature Genet 1996; 13: 469–71.

Jeffreys AJ, Kauppi L, Neumann R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nature Genet 2001; 29: 217–22.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, et al. Haplotype tagging for the identification of common disease genes. Nature Genet 2001; 29: 233–7.

Johnston JB, Silva C, Holden J, Warren KG, Clark AW, Power C. Monocyte activation and differentiation augment human endogenous retrovirus expression: implications for inflammatory brain diseases. Ann Neurol 2001; 50: 434–42.

Kendler KS, MacLean CJ, Ma Y, O'Neill FA, Walsh D, Straub RE. Marker-to-marker linkage disequilibrium on chromosomes 5q, 6p, and 8p in Irish high-density schizophrenia pedigrees. Am J Med Genet 1999; 88: 29–33.

Kirk CW, Droogan AG, Hawkins SA, McMillan SA, Nevin NC, Graham CA. Tumour necrosis factor microsatellites show association with multiple sclerosis. J Neurol Sci 1997; 147: 21–5.

Kirov G, Williams N, Sham P, Craddock N, Owen MJ. Pooled genotyping of microsatellite markers in parent–offspring trios. Genome Res 2000; 10: 105–15.

Kruglyak L. The use of a genetic map of biallelic markers in linkage studies. Nature Genet 1997; 17: 21–4.

Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nature Genet 1999; 22: 139–44.

Kuokkanen S, Gschwend M, Rioux JD, Daly MJ, Terwilliger JD, Tienari PJ, et al. Genomewide scan of multiple sclerosis in Finnish multiplex families. Am J Hum Genet 1997; 61: 1379–87.

Larsen F, Oturai A, Ryder LP, Madsen HO, Hillert J, Fredrikson S, et al. Linkage analysis of a candidate region in Scandinavian sib pairs with multiple sclerosis reveals linkage to chromosome 17q. Genes Immun 2000; 1: 456–9.

Lee N, Daly MJ, Delmonte T, Lander ES, Xu F, Hudson TJ, et al. A genomewide linkage-disequilibrium scan localizes the Saguenay-Lac-Saint-Jean cytochrome oxidase deficiency to 2p16. Am J Hum Genet 2001; 68: 398–409.

Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, et al. SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. Am J Hum Genet 2000; 67: 383–94.

McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the Diagnosis of Multiple Sclerosis. Ann Neurol 2001; 50: 121–7.

Muller-Myhsok B, Abel L. Genetic analysis of complex diseases [letter]. Science 1997; 275: 1328–9.

Mumford CJ, Wood NW, Kellar-Wood H, Thorpe JW, Miller DH, Compston DA. The British Isles survey of multiple sclerosis in twins. Neurology 1994; 44: 11–5.

Pacek P, Sajantila A, Syvanen AC. Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples. PCR Methods Appl 1993; 2: 313–7.

Perlin MW, Lancia G, Ng SK. Toward fully automated genotyping: genotyping microsatellite markers by deconvolution. Am J Hum Genet 1995; 57: 1199–210.

Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol 1983; 13: 227–31.

Rehli M, Krause SW, Schwarzfischer L, Kreutz M, Andreesen R. Molecular cloning of a novel macrophage maturation-associated transcript encoding a protein with several potential transmembrane domains. Biochem Biophys Res Commun 1995; 217: 661–7.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, et al. Linkage disequilibrium in the human genome. Nature 2001; 411: 199–204.

Risch NJ. Searching for genetic determinants in the new millennium. [Review]. Nature 2000; 405: 847–56.

Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996; 273: 1516–7.

Risch N, Teng J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. [Review]. Genome Res 1998; 8: 1273–88.

Robertson NP, Fraser M, Deans J, Clayton D, Walker N, Compston DA. Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. Brain 1996; 119: 449–55.

Robertson NP, O'Riordan JI, Chataway J, Kingsley DP, Miller DH, Clayton D, et al. Offspring recurrence rates and clinical characteristics of conjugal multiple sclerosis. Lancet 1997; 349: 1587–90.

Sadovnick AD, Baird PA, Ward RH. Multiple sclerosis: updated risks for relatives. Am J Hum Genet 1988; 29: 533–41.

Sadovnick AD, Ebers GC, Dyment DA, Risch NJ. Evidence for genetic basis of multiple sclerosis. Lancet 1996; 347: 1728–31.

Sawcer S, Jones HB, Feakes R, Gray J, Smaldon N, Chataway J, et al. A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. Nature Genet 1996; 13: 464–8.

Shaw SH, Carrasquillo MM, Kashuk C, Puffenberger EG, Chakravarti A. Allele frequency distributions in pooled DNA samples: applications to mapping complex disease genes. Genome Res 1998; 8: 111–23.

Smith JR, Carpten JD, Brownstein MJ, Ghosh S, Magnuson VL, Gilbert DA, et al. Approach to genotyping errors caused by nontemplated nucleotide addition by Taq DNA polymerase. Genome Res 1995; 5: 312–7.

Taillon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen T, Cao A, et al. Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nature Genet 2000; 25: 324–8.

Thomson G. Mapping disease genes: family-based association studies. Am J Hum Genet 1995; 57: 487–98.

Transatlantic Multiple Sclerosis Genetics Cooperative. A meta-analysis of genomic screens in multiple sclerosis. Mult Scler 2001; 7: 3–11.

Uhl GR, Liu QR, Walther D, Hess J, Naiman D. Polysubstance abuse-vulnerability genes: genome scans for association, using 1,004 subjects and 1,494 single-nucleotide polymorphisms. Am J Hum Genet 2001; 69: 1290–300.

Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 1989; 44: 388–96.

Weiss KM, Terwilliger JD. How many diseases does it take to map a gene with SNPs? [Review]. Nature Genet 2000; 26: 151–7.

Received November 13, 2001. Revised January 15, 2002.
Accepted January 24, 2002