nature genetics

A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility

David Reich[1,2], Nick Patterson[2], Philip L De Jager[2–4], Gavin J McDonald[1,2], Alicja Waliszewska[2–4], Arti Tandon[1,2], Robin R Lincoln[5], Cari DeLoa[5], Scott A Fruhan[1–4], Philippe Cabre[6], Odile Bera[6], Gilbert Semana[7], M Ann Kelly[8], David A Francis[8], Kristin Ardlie[9], Omar Khan[10], Bruce A C Cree[5], Stephen L Hauser[5], Jorge R Oksenberg[5] & David A Hafler[2–4]

Multiple sclerosis is a common disease with proven heritability, but, despite large-scale attempts, no underlying risk genes have been identified. Traditional linkage scans have so far identified only one risk haplotype for multiple sclerosis (at HLA on chromosome 6), which explains only a fraction of the increased risk to siblings. Association scans such as admixture mapping have much more power, in principle, to find the weak factors that must explain most of the disease risk. We describe here the first high-powered admixture scan, focusing on 605 African American cases and 1,043 African American controls, and report a locus on chromosome 1 that is significantly associated with multiple sclerosis.

Admixture mapping is a new method for scanning the genome for gene variants that affect the risk for common, complex disease. The method has high statistical power to detect factors that differ markedly in frequency across human populations[1–6]. Although admixture mapping was proposed more than 50 years ago[1], the lack of a high-density map[7] of genetic markers or robust methods for analyzing the data[8–10] has prevented practical studies until recently[7,11]. Admixture mapping is inspired by the idea that by studying genetic markers whose frequencies differ between Europeans and Africans, one can classify the genome of an African American into sections that come from their African or European ancestors (the individuals in this study have, on average, 21% European and 79% African

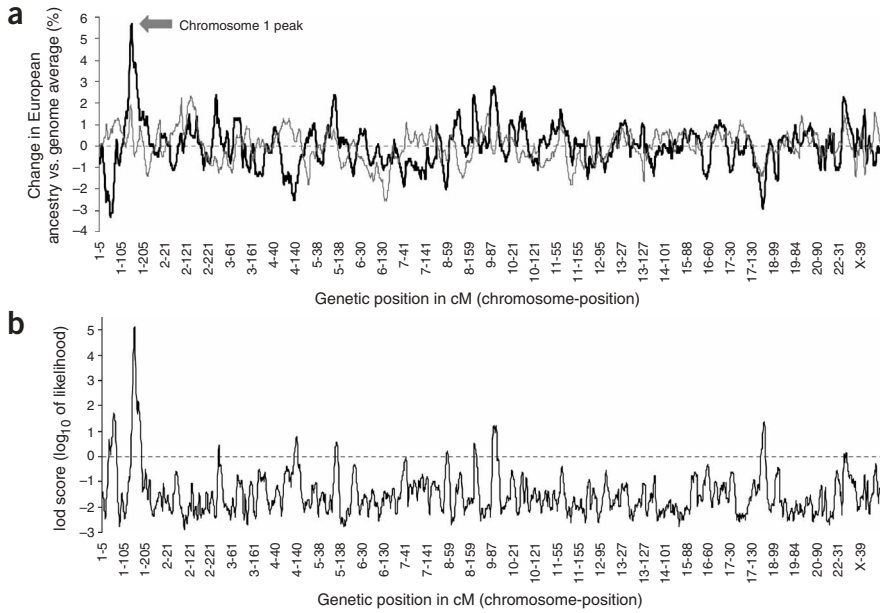

Figure 1 Results of the genome-wide admixture scan (run 3 in **Table 1**). (**a**) Estimate of European ancestry for 605 African Americans with multiple sclerosis compared with the genome-wide average for cases (black line) and controls (gray line). In cases, the strongest deviation from the genome-wide average is +5.9% on chromosome 1 near the centromere, with no strong rise or fall anywhere in the genome in 1,043 controls. (**b**) The chromosome 1 peak is also detected in a case-only scan for disease loci (lod = 5.2).

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [2]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [3]Department of Neurology, Harvard Medical School, Boston, Massachusetts, USA. [4]Laboratory of Molecular Immunology, Center for Neurologic Disease, Brigham and Women's Hospital, Boston, Massachusetts, USA. [5]Department of Neurology, University of California, San Francisco, California, USA. [6]Clinique Neurologique, L'Hôpital Quitman, Fort-de-France, France. [7]Laboratoire Universitaire d'Immunologie, University de Rennes 1 and Etablissement Français du Sang Bretagne, Rennes, France. [8]Department of Medicine, Division of Medical Sciences, University Hospital Birmingham, UK. [9]Genomics Collaborative, Division of SeraCare Life Sciences Inc, Cambridge, Massachusetts, USA. [10]Multiple Sclerosis Center, Department of Neurology, Wayne State University School of Medicine, Detroit, Michigan, USA. Correspondence should be addressed to D.R. (reich@receptor.med.harvard.edu).

Received 10 May; accepted 14 July; published online 25 September 2005; doi:10.1038/ng1646

Table 1 Summary of results from admixture scans

Run	Description of run	SNPs	Cases / controls	Burn-ins / follow-ons for MCMC	Genome score	Lod at chr. 1 peak	Case-control at chr. 1 peak
1	Initial African American samples, initial SNP set	1,082	484 / 1,043	100 / 200	2.2[a]	4.9	3.6
2	All African Americans, final SNP set	1,166	605 / 1,043	100 / 200	2.3[a]	5.2	3.3
3	All African Americans, final SNP set (10× more iterations)	1,166	605 / 1,043	1,000 / 2,000	2.3[a]	5.2	3.3
4	African American case-only analysis, final set of SNPs	1,166	605 / 0	100 / 200	2.3[a]	5.2	NA
5	Clinically definite African Americans only, final SNPs	1,166	572 / 1,043	100 / 200	1.2	4.0	2.2
6	All African Americans, final SNPs (drop every odd SNP)	679	605 / 1,043	100 / 200	2.7[a]	5.7	3.1
7	All African Americans, final SNPs (drop every even SNP)	697	605 / 1,043	100 / 200	1.4	3.8	2.7
8	All African Americans, final SNPs (SNPs spaced >1 Mb, >1 cM)	846	605 / 1,043	100 / 200	1.8	4.7	2.9
9[b]	Afro-Caribbean samples only, final SNP set	1,166	143 / 151	100 / 200	–0.1	–1.0	1.5
10[b]	African Americans and Afro-Caribbeans combined	1,166	748 / 1,194	100 / 200	1.5	4.3	3.2
11	Best-fit multiplicative model (1.44-fold multiplicative risk)	1,166	605 / 1,043	100 / 200	2.7[a]	5.7	3.3
12	Best-fit model (1.62 heterozygote, 1.78 homozygote risk)	1,166	605 / 1,043	100 / 200	3.4[a]	6.4	3.0

[a]Genome scores above 2.0 are formally significant; scores above 1.0 are suggestive. [b]Runs 9 and 10 are the only ones that include Afro-Caribbean data.
MCMC, Markov Chain Monte Carlo; NA, not applicable.

ancestry). The goal is to identify genomic regions where individuals with multiple sclerosis tend to have an unusually high proportion of ancestry from either Europeans or Africans (**Fig. 1**), indicative of the presence of a multiple sclerosis risk variant that differs in frequency between the ancestral populations.

Admixture mapping has two advantages over more commonly discussed strategies of whole-genome association. First, there has been an average of only six generations since African and European populations came into contact in North America; thus, there has been little recombination between chromosomes of African and European ancestry in the history of African American populations[7]. Chromosomal segments of one or the other ancestry are therefore typically tens of millions of base pairs long, and a genome scan tagging all these segments requires only 1,000–3,000 markers, rather than the 300,000–1,000,000 required for a whole-genome haplotype scan[12]. The second advantage is that admixture mapping has much more power to detect disease risk variants that are of very low or high frequency in Africans or Europeans but are highly differentiated in frequency across populations.

Only one published study has used admixture mapping in a scan for disease genes[11]. But the identification of two putative new loci underlying hypertension has not been confirmed by follow-up genotyping around the peaks of association or by formal significance testing. Mindful of the pitfalls associated with a new and complex method, we have reported rigorous criteria for declaring association in an admixture study[13], a high-density admixture map[7] that makes it practical to carry out a high-powered scan in African Americans and an analytical method that we tested exhaustively with simulations and real data[8].

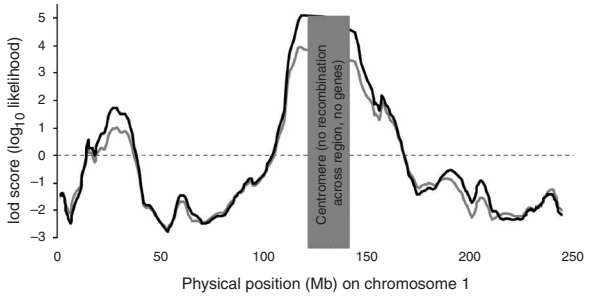

Figure 2 The strongest peak of association spans the centromere of chromosome 1. At this position, we observe a lod score of 5.2 (black line), which decreases to 4.0 (gray line) when we drop 33 probable multiple sclerosis samples whose physician-diagnosed multiple sclerosis could not be confirmed owing to incomplete medical records. The second strongest peak does not meet our criterion for significance or suggestiveness but is also on chromosome 1 and corresponds to a decrease in European ancestry (lod = 1.7).

Multiple sclerosis is an excellent candidate for admixture mapping because it is more prevalent in European Americans than African Americans[14,15]. We hypothesized that if there are genetic risk factors for multiple sclerosis that explain the epidemiology, they should be identifiable as regions with a high proportion of European ancestry in African Americans with multiple sclerosis compared with the average. We primarily tested for European ancestry conferring risk, but because variants more common in Africans might also confer risk, we also tested this hypothesis.

The initial scan of 484 cases and 1,043 controls (genotyped at 1,082 SNPs) detected a genome-wide significant association with multiple sclerosis risk around the chromosome 1 centromere (lod = 4.9; **Table 1**). The association grew stronger (lod = 5.2) with the addition of 121 new cases and 84 markers; at this position, there was a 5.9% rise in European ancestry compared with the genome-wide average (**Fig. 1**). We were concerned, however, that the lod score dropped to 4.0 when we eliminated 33 samples whose physician diagnosis of multiple sclerosis could not be confirmed owing to an incomplete medical record (**Table 1** and **Fig. 2**). To evaluate whether these samples were inhomogeneous with the rest of the cases, we repeatedly dropped subsets of 33 samples from the full set of 605 cases: 11 of 200 random replicates showed lod score reductions as large as in our data, which was odd but not unusual enough to reject these samples from analysis.

The second strongest association was also on chromosome 1 and corresponded to a decrease in European ancestry (lod = 1.7). Admixture association was not detected anywhere else in the genome (**Table 2**), including at the HLA locus (6p21), the one place where there are known genetic risk factors for multiple sclerosis[16]. Here, the admixture scan did not detect differential risk for multiple sclerosis comparing Africans and Europeans (95% credible interval = 0.87–1.24; **Supplementary Table 1** online). HLA genotyping in these samples[17] suggested why the locus was not detected: the overall effect of the multiple sclerosis risk haplotype HLA-DRB1*1501, which is

Table 2 Summary of main scan results by chromosome

Chromosome	Highest lod score	Highest case-control Z score	Lowest case-control Z score
1	5.2	3.3	−3.6
2	0.5	2.4	−2.1
3	−0.3	2.2	−1.9
4	0.8	1.8	−2.7
5	0.6	1.5	−1.8
6	−1.0	1.3	−1.2
7	0.0	0.2	−1.9
8	0.2	2.2	−1.7
9	1.2	2.5	−1.8
10	−0.5	1.7	−1.0
11	−0.4	1.4	−0.1
12	−0.6	0.9	−1.8
13	−1.0	2.2	−1.2
14	−0.8	0.5	−1.8
15	−0.6	0.9	−1.1
16	−0.3	0.4	−2.2
17	−0.6	0.0	−2.0
18	1.4	0.3	−1.6
19	−1.7	0.8	−0.7
20	−1.1	0.7	−1.3
21	−1.2	0.4	−0.9
22	0.2	0.9	−1.3
X	−0.3	1.9	−2.2

Results correspond to run 3 in **Table 1**.

more common in Europeans and which we estimate (on the basis of its over-representation compared with controls) to increase European ancestry by ~2%, on ancestry was canceled by HLA-DRB1*1503, which is more common in Africans and which we estimate to decrease European ancestry by ~1%. This highlights a limitation of admixture mapping: it cannot detect disease loci at which the total risk summed over all alleles in each population is similar in Africans and Europeans.

To assess the robustness of the peak, we re-examined the data from several perspectives. First, we formally assessed evidence for association by averaging the likelihood ratios (10 to the power of the lod score) at points spaced every centimorgan across the genome[8]. The logarithm of the average score was 2.3, indicating that the disease model was 10 to the power of 2.3 (~200) times more likely than the null model. This meets our published criterion[8] for association (lod > 2). Second, to ensure that the evidence for association was not dependent on particular markers, we divided the data into even- and odd-numbered SNPs. Each set independently showed evidence of association (**Table 1**). We also thinned the SNPs until all were separated by >1 Mb and >1 cM. The data continued to support association (**Table 1**). Third, to ensure that the detected association was not an artifact of mismatching of cases and controls, we carried out an analysis after removing all controls (605 cases, 0 controls). The score was significant genome-wide at 2.3 (**Table 1**). Fourth, we used computer simulations[8] to assess significance, creating 1,000 artificial data sets with no disease locus but otherwise the same structure (same number of markers and samples) as our real multiple sclerosis scan. Only one simulation generated a genome-wide score as high as in our real data ($P = 0.001$).

An alternative way to analyze admixture mapping data is to compare cases with controls. To obtain maximum power, we supplemented the 173 controls from the multiple sclerosis study with 870 controls from outside the collection. The new controls had more European ancestry overall than cases (23.5% versus 21.1%) reflecting their different geographic origin. A mismatch with respect to ancestry usually raises the possibility of false positives. That was not a problem here, because admixture mapping scans for a rise in European ancestry in cases (versus their genome-wide average) that is not seen in controls (compared with their average)[8,10,18]. The case-control Z statistic[8] at the chromosome 1 peak was 3.3, significant at $P < 0.0010$ (**Supplementary Table 2** online). We were concerned, however, by a 1.9% rise in European ancestry in controls within 4 Mb of the peak. This raised the possibility that the chromosome 1 peak might be an artifact observed in all African Americans (**Supplementary Table 1** online). We therefore increased the total number of controls to 2,035 by adding cases from a parallel prostate cancer study. The rise decreased to 1.4%; therefore, the chromosome 1 rise seemed to be specific to individuals with multiple sclerosis.

As a replication set, we studied 143 Afro-Caribbeans with multiple sclerosis from Martinique and the UK[19,20], recognizing at the outset that the genetic risk profile might be different in this cohort because of its different origin (**Supplementary Table 3** online). There was no evidence for replication, although the confidence intervals for increased risk due to European ancestry overlapped those in African Americans (**Fig. 3**). There are two possible explanations: first, the chromosome 1 locus does not have a role in Afro-Caribbean populations, perhaps owing to a difference between the Caribbean and North American environments, consistent with the well-known effect of latitude on risk for multiple sclerosis[21]; and second, the smaller size of the Afro-Caribbean cohort means that it cannot provide definitive replication.

To explore how the chromosome 1 locus might contribute to multiple sclerosis, we tested the likelihood of the data at the chromosome 1 peak under a range of multiplicative models of disease risk (**Table 3**). The best fit was a 1.44-fold increased risk owing to heterozygosity with respect to European ancestry (local lod = 5.7). The best fit for a nonmultiplicative model was a 1.62-fold increased risk resulting from one European allele and a 1.78-fold increased risk resulting from two European alleles (local lod = 6.4). The greater likelihood of the nonmultiplicative model suggested that the multiple sclerosis risk variants underlying the locus might have a dominant effect ($P < 0.07$ by a likelihood ratio test). We further tested whether the chromosome 1 ancestry association interacted epistatically with

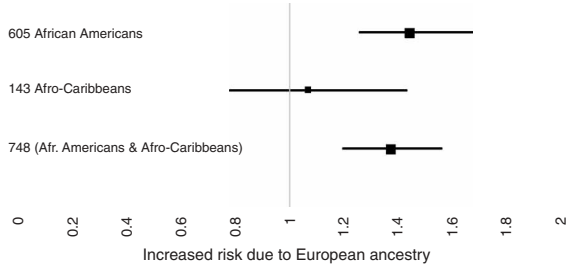

Figure 3 The 95% credible intervals for increased risk due to European ancestry at the chromosome 1 locus, for all African Americans ($n = 605$), all Afro-Caribbeans ($n = 143$) and both cohorts together ($n = 748$). Filled squares give the maximum likelihood estimates, whereas the credible intervals are the full range of multiplicative risk models for which the likelihood is within 0.83 of the maximum. There is no evidence for replication in the Afro-Caribbeans, although the confidence intervals for increased risk due to European ancestry overlap those in African Americans, and the data are consistent with the chromosome 1 locus modulating risk in both populations.

Table 3 Risk models evaluated in main admixture scan

Risk due to European ancestry	Genome-wide score with risk	Weight of model in our prior
0.4 = 60% lower per European allele	−5.8	6
0.5	−2.7	19
0.6	−0.6	39
0.7	0.4	46
0.8	0.4	25
1.3	2.3	91
1.4	2.7	100
1.5	2.7	88
1.6	2.3	70
1.7	1.6	54
1.8	0.6	40
1.9	−0.5	29
2.0	−1.9	20
2.1	−3.4	14
2.2	−5.0	10
2.3	−6.7	7
2.4	−8.5	5
2.5	−10.0	3
2.6	−11.1	2
2.7	−12.0	1
2.8	−12.9	1
2.9	−13.9	1
3.0 = 200% higher per European allele	−14.8	1
Total score	2.3	Weighted average of score across genome

the one other known genetic risk factor for multiple sclerosis, at the HLA locus. No interaction between the chromosome 1 peak and known risk alleles was found: individuals carrying the multiple sclerosis risk haplotypes HLA-DRB1*1501 or HLA-DRB1*1503 contribute about as much evidence per sample as noncarriers to the chromosome 1 peak.

The data also provide information about the location of the putative multiple sclerosis risk gene. The 95% credible interval (where the lod score is within 0.83 units of the maximum) extends from 114.9 Mb to 144.7 Mb in build 35 of the human genome assembly and contains 68 known genes. Previous linkage studies of multiple sclerosis never prioritized the chromosome 1 locus: in 1,002 affected relative pairs of European ancestry, the lod score was 0.08 (ref. 22). The fact that this locus was not identified might reflect the weak power of linkage scans to detect variants of modest effect. Alternatively, the multiple sclerosis risk variant might occur at a high frequency in Europeans but at much lower frequency in Africans. This would make it detectable in an admixture scan but undetectable in a linkage study in a population such as European Americans, in which nearly everyone would have the risk allele. Thus, admixture mapping might have power to find disease risk variants undetectable by other genome scan techniques.

These results validate admixture mapping as a promising way of finding risk genes for common and complex diseases. The immediate priority should be fine-mapping the chromosome 1 peak to clone the putative multiple sclerosis risk gene, as no polymorphism we have studied so far shows association to multiple sclerosis beyond the admixture association. Additional priority should be placed on collecting large samples of additional African Americans with multiple sclerosis, as this will provide power to confirm the chromosome 1 association, identify new risk loci and identify enough individuals with phenotypes such as rapid progression, opticospinal multiple sclerosis or poor response to immunosuppressive therapy to permit mapping of the relevant genes. This study should also lend momentum to the application of admixture mapping to other diseases in which epidemiological data suggest admixture mapping will be favorable[7], such as lupus, end-stage renal disease and type 2 diabetes.

METHODS

Multiple sclerosis samples. The samples from African Americans with multiple sclerosis used for the main analysis ($n = 605$) were collected primarily by the MS Genetics Group at the University of California San Francisco ($n = 572$) through a nationwide network of collaborating investigators[23] and through the assistance of the National Multiple Sclerosis Society and the Montel Williams Foundation. As described previously[23], these samples all met strict McDonald criteria[24] for a diagnosis of multiple sclerosis, except for 33 individuals who were considered to have multiple sclerosis by their clinicians but whose medical records were incomplete and therefore could not be evaluated for entry criteria. Additional samples from African Americans with multiple sclerosis came from Genomics Collaborative ($n = 21$) and from the Brigham and Women's Hospital ($n = 12$). The 143 Afro-Caribbean cases included 122 cases from Martinique and 21 from the UK[19,20]; all met criteria for a diagnosis of multiple sclerosis by the McDonald criteria (Martinique) or clinically definite multiple sclerosis as defined by Poser (UK)[24,25].

Control samples. The 1,043 control samples in this study were all from self-identified African Americans. They came from four sources: (i) spouses ($n = 106$) and unrelated friends ($n = 29$) of studied individuals with multiple sclerosis; (ii) pairs of untransmitted chromosomes from parents of individuals with multiple sclerosis ($n = 38$); (iii) samples previously genotyped in the context of building our admixture map[7] ($n = 109$); and (iv) African American cohort controls from Los Angeles ($n = 761$) from our parallel admixture mapping study of prostate cancer (B. Henderson and S. Ingles, personal communication). The 151 Afro-Caribbean controls included 98 from Martinique and 53 from the UK[19,20]. We also included 992 African American cases from the prostate cancer study (B. Henderson and S. Ingles, personal communication) as additional controls for one subanalysis.

Frequency estimates from ancestral populations. To obtain frequency estimates for each of the SNPs in Africans and Europeans, we used data previously published as part of building an admixture map[7], supplemented with new genotyping for SNPs specifically identified for this study. West African ancestral frequencies were estimated using samples from Ghana ($n = 33$), Cameroon ($n = 20$) and sometimes Nigeria ($n = 70$). European frequencies were obtained using samples from Baltimore ($n = 39$), Chicago ($n = 40$), Italy ($n = 41$), Poland ($n = 39$), Norway ($n = 48$) and sometimes Utah ($n = 87$) and Los Angeles ($n = 4$). The observed rates of reference and variant alleles in the African- and European-derived populations for each SNP are reported in **Supplementary Tables 1** and **4** online.

Whole-genome amplification of DNA. We subjected all Afro-Caribbean samples, all samples from the prostate cancer study, and the multiple sclerosis cases from Genomics Collaborative to whole-genome amplification[26] to produce DNA sufficient for these studies.

Elimination of poorly performing samples. We excluded DNA samples if they showed less than an 85% genotyping success rate (5 cases and 16 controls eliminated) or showed an excess or deficiency of heterozygous genotypes compared with the expectation from the individual's estimated proportion of European ancestry (5 cases and 2 controls eliminated; **Supplementary Methods** online).

SNP genotyping and quality control. We attempted assays for 1,831 SNPs using the Illumina[27] and Sequenom[28] genotyping platforms. For the Illumina genotyping, we prioritized the most informative SNPs from our published admixture map according to their predicted usefulness for determining

ancestry[7] and attempted assays for the top 1,536 SNPs. For Sequenom genotyping, we prioritized as follows: (i) 95 SNPs genome-wide that filled gaps in the map not successfully assayed by Illumina[7]; (ii) 16 SNPs to reconfirm Illumina genotypes around the chromosome 1 peak; and (iii) 200 SNPs to increase information about regions of interest, especially at the chromosome 1 peak. We eliminated 201 SNPs from analysis by considering only those that showed more than an 85% genotyping success rate in African American multiple sclerosis cases or that had reliable genotype clustering patterns as judged by an experienced research technician; this left a total of 1,630 SNPs useable for analysis. We eliminated 82 of these SNPs by considering only those in Hardy-Weinberg equilibrium ($P > 0.01$) in both the ancestral West African or European American populations, leaving 1,558 for analysis. Finally, we eliminated three more SNPs by requiring that the frequency in African American controls was appropriately intermediate between the frequencies in ancestral West Africans and European Americans (**Supplementary Methods** online). A total of 1,555 SNPs met all criteria for analysis and were used in our main study (**Supplementary Tables 1** and **4** online).

Genotyping accuracy and completeness. We obtained many genotypes in duplicate using the Illumina and Sequenom technologies (including a substantial amount of Sequenom genotyping that we did before the Illumina genotyping on many of the same SNPs, in the course of collecting preliminary data for this study). Of 203,311 overlapping genotypes, only 303 were different (0.15%), giving a maximum error rate attributable to either the Illumina or the Sequenom genotyping process. The overall genotyping completeness rate in African Americans and Afro-Caribbeans was 94.2% for the SNP-sample combinations studied.

Eliminating markers in linkage disequilibrium in ancestral populations. It was essential to identify a set of SNPs that were in linkage equilibrium with each other in the ancestral West African and European American populations, as analyzing SNPs that are in linkage disequilibrium (LD) in an admixture scan can result in false-positive associations to disease[8]. We implemented a "greedy" algorithm that picks a subset of successfully genotyped SNPs for analysis that are not in LD. First, it picks the SNP that has the highest calculated Shannon Information Content[7] for distinguishing African and European ancestry. Then it adds additional SNPs consecutively in order of their information content, considering only those that are not within a specified physical distance of any SNPs already chosen (we used a 200-kb exclusion zone, which was extended to 1 Mb for the conservative run 8; **Table 1**). Even for SNPs more than the minimum physical distance away from all other SNPs in the analysis, we explicitly tested[29] for LD with any other SNP within 20 cM. Only a slight hint of LD in the modern European or African populations was required to exclude pairs of markers if they were relatively close, but stronger evidence was required for more distantly spaced SNP pairs. The P value criteria decreased inversely with distance: <0.01, <0.001 and <0.0001 for pairs separated by 0.2 cM, 2 cM and 20 cM, respectively. Markers eliminated from the main analysis on the basis of these criteria are listed in **Supplementary Table 4** online.

Determination of physical and genetic positions of SNPs in the study. We mapped all SNPs to build 35 of the public genome reference sequence and discarded from analysis those markers that mapped to more than one location in the genome. To obtain genetic positions, we used the Rutgers integrated map[30], interpolating the genetic positions of the SNPs in this study based on their physical positions relative to those in the Rutgers map. Because the resolution of the Rutgers map breaks down below a scale of a few megabases[30], we modified these distances by using a smoothing algorithm to ensure that even if no recombination events had been observed between two markers in the scan, we could infer some minimal genetic distance between them proportional to their physical spacing (**Supplementary Methods** online).

Risk model used in the Markov Chain Monte Carlo data analysis. We used the ANCESTRYMAP software[8] as the central engine of the analysis. The lod score for association is defined as the ratio of the likelihood of the data under a disease model divided by the likelihood of the data under no disease model. The ANCESTRYMAP software uses Bayesian statistics and thus requires specification of a prior distribution on risk models before carrying out the analysis. Because multiple sclerosis occurs more often in Europeans than African Americans, we searched for disease risk loci under a prior distribution where European ancestry confers increased risk with 80% probability. We also considered the possibility of increased risk due to African ancestry (20% weight). The prior distribution on risks was given by a gamma distribution with a mean increase in risk of 1.6-fold, a standard deviation of 0.3 and an offset so that the minimum allowed risk was 1.2 (we did not expect to have power to detect risks with less than 1.2-fold given our sample size[8]). We implemented this prior distribution by testing a grid of 27 risk models from 0.4 to 3.0, weighting the lod scores obtained in each model according to the distribution (**Table 3**). Once we identified the chromosome 1 peak, we explored not only multiplicative risk models but also the possibility that the risk to homozygotes is different from the square of the risk in heterozygotes.

Computer simulations to evaluate statistical significance. We used our previously described software[8] to simulate data sets that had the same structure as our real data but no disease locus. For each of 1,000 replicates, we simulated (i) 605 cases and 1,043 controls with the same percentage of European ancestry and number of generations since mixture as estimated for our real samples (**Supplementary Table 5** online); (ii) 1,166 SNPs with the same genomic positions and ancestral frequencies as in the real markers (**Supplementary Table 1** online); and (iii) the same pattern of missing genotypes as in our real data. We used the ANCESTRYMAP admixture mapping software[8] to obtain a genome-wide score for association for the data produced by each simulation, and we counted the proportion of simulations that gave scores greater than the value of 2.3 seen in our data (1 of 1,000 simulations).

URLs. Our ANCESTRYMAP software is available at http://genepath.med.harvard.edu/~reich/. Build 35 of the public genome reference sequence is available from http://genome.ucsc.edu/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the individuals with multiple sclerosis, their families and friends and others for permission to use their DNA in these studies; M. Williams for his role in building the collection of samples from African Americans; B. Henderson and S. Ingles for permission to use data from the prostate cancer study; D. Goldstein for frequency information from the Italian and Norwegian samples; and D. Goldstein and E. Lander for providing comments throughout the project. This work was supported by grants to D.R. from the Wadsworth Foundation and the US National Institute of Neurological Disorders and Stroke; to D.A.H. from the National Multiple Sclerosis Society; to J.R.O. from the US National Institute of Neurological Disorders and Stroke and the National Multiple Sclerosis Society; and to S.L.H. from the Nancy David and Montel Williams Foundations. D.R. is the recipient of a Burroughs-Wellcome Career Development Award in the Biomedical Sciences; B.A.C.C. is a Sylvia Lawry fellow of the National Multiple Sclerosis Society; N.P. has a US National Institutes of Health career transition award; D.A.H. has a US National Institute of Neurological Disorders and Stroke Javits Investigator Award; and P.L.D.J. is the William C Fowler scholar for MS Research and is supported by US National Institute of Neurological Disorders and Stroke grant and the Clinical Investigator Training program: Harvard-MIT Health Sciences and Technology–Beth Israel Deaconess Medical Center, in collaboration with Pfizer, Inc.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Rife, D.C. Populations of hybrid origin as source material for the detection of linkage. *Am. J. Hum. Genet.* **6**, 26–33 (1954).
2. Chakraborty, R. & Weiss, K.M. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc. Natl. Acad. Sci. USA* **85**, 9119–9123 (1988).
3. Stephens, J.C., Briscoe, D. & O'Brien, S.J. Mapping by admixture linkage disequilibrium in human populations: Limits and guidelines. *Am. J. Hum. Genet.* **55**, 809–824 (1994).
4. Briscoe, D., Stephens, J.C. & O'Brien, S.J. Linkage disequilibrium in admixed populations: Applications in gene mapping. *J. Hered.* **85**, 59–63 (1994).

5. McKeigue, P.M. Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. *Am. J. Hum. Genet.* **60**, 188–196 (1997).
6. McKeigue, P.M. Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in populations, by conditioning on parental admixture. *Am. J. Hum. Genet.* **63**, 241–251 (1998).
7. Smith, M.W. *et al.* A high-density admixture map for disease gene discovery in African Americans. *Am. J. Hum. Genet.* **74**, 1001–1013 (2004).
8. Patterson, N. *et al.* Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* **74**, 979–1000 (2004).
9. Hoggart, C.J., Shriver, M.D., Kittles, R.A., Clayton, D.G. & McKeigue, P.M. Design and analysis of admixture mapping studies. *Am. J. Hum. Genet.* **74**, 965–978 (2004).
10. Montana, G. & Pritchard, J.K. Statistical tests for admixture mapping with case-control and cases-only data. *Am. J. Hum. Genet.* **75**, 771–789 (2004).
11. Zhu, X. *et al.* Admixture mapping for hypertension loci with genome-scan markers. *Nat. Genet.* **37**, 177–181 (2005).
12. Carlson, C.S. *et al.* Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat. Genet.* **33**, 518–521 (2003).
13. Reich, D. & Patterson, N. Will admixture mapping work to find disease genes? *Phil. Trans. R. Soc. Lond. B* **360**, 1605–1607 (2005).
14. Kurtzke, J.F., Beebe, G.W. & Norman, J.E. Jr. Epidemiology of multiple sclerosis in U.S. veterans: 1. Race, sex, and geographic distribution. *Neurology* **29**, 1228–1235 (1979).
15. Wallin, M.T., Page, W.F. & Kurtzke, J.F. Multiple sclerosis in US veterans of the Vietnam era and later military service: race, sex, and geography. *Ann. Neurol.* **55**, 65–71 (2004).
16. The Transatlantic Multiple Sclerosis Genetics Cooperative. A meta-analysis of genomic screens in multiple sclerosis. A meta-analysis of genomic screens in multiple sclerosis. *Mult. Scler.* **7**, 3–11 (2001).
17. Oksenberg, J.R. *et al.* Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *Am. J. Hum. Genet.* **74**, 160–167 (2004).
18. Hoggart, C.J. *et al.* Control of confounding of genetic associations in stratified populations. *Am. J. Hum. Genet.* **72**, 1492–1504 (2003).
19. Quelvennec, E. *et al.* Genetic and functional studies in multiple sclerosis patients from Martinique attest for a specific and direct role of the HLA-DR locus in the syndrome. *Tissue Antigens* **61**, 166–171 (2003).
20. Kelly, M.A. *et al.* An investigation of HLA-encoded genetic susceptibility to multiple sclerosis in subjects of Asian Indian and Afro-Caribbean ethnic origin. *Tissue Antigens* **45**, 197–202 (1995).
21. Lauer, K. The risk of multiple sclerosis in the U.S.A. in relation to sociogeographic features: a factor-analytic study. *J. Clin. Epidemiol.* **47**, 43–48 (1994).
22. International Multiple Sclerosis Genetics Consortium. A high-density screen for linkage in multiple sclerosis. *Am. J. Hum. Genet.* **77**, 454–467 (2005).
23. Cree, B.A. *et al.* Clinical characteristics of African Americans vs Caucasian Americans with multiple sclerosis. *Neurology* **63**, 2039–2045 (2004).
24. McDonald, W.I. *et al.* Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the Diagnosis of Multiple Sclerosis. *Ann. Neurol.* **50**, 121–127 (2001).
25. Poser, C.M. *et al.* New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann. Neurol.* **13**, 227–231 (1983).
26. Hosono, S. *et al.* Unbiased whole-genome amplification directly from clinical samples. *Genome Res.* **13**, 954–964 (2003).
27. Fan, J.B. *et al.* Highly parallel SNP genotyping. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 69–78 (2003).
28. Tang, K. *et al.* Chip-based genotyping by mass spectrometry. *Proc. Natl. Acad. Sci. USA* **96**, 10016–10020 (1999).
29. Excoffier, L. & Slatkin, M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* **12**, 921–927 (1995).
30. Kong, X. *et al.* A combined linkage-physical map of the human genome. *Am. J. Hum. Genet.* **75**, 1143–1148 (2004).