Human Molecular Genetics, 2010, Vol. 19, No. 15 **3080–3088**
doi:10.1093/hmg/ddq197
Advance Access published on May 12, 2010

Refining the association of MHC with multiple sclerosis in African Americans

Joseph P. McElroy[1,*], Bruce A. C. Cree[1], Stacy J. Caillier[1], Peter K. Gregersen[2], Joseph Herbert[3], Omar A. Khan[4], Jan Freudenberg[2], Annette Lee[2], S. Louis Bridges Jr[5], Stephen L. Hauser[1], Jorge R. Oksenberg[1] and Pierre-Antoine Gourraud[1]

[1]Department of Neurology, University of California, San Francisco, CA 94143, USA, [2]Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, North Shore University Hospital, Manhasset, NY 11030, USA, [3]Department of Neurology, New York University Hospital for Joint Diseases, New York, NY 10003, USA, [4]Department of Neurology, Wayne State Medical School, Detroit, MI 48201, USA and [5]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA

Received February 12, 2010; Revised April 11, 2010; Accepted May 6, 2010

Multiple sclerosis (MS) is a common demyelinating disease of the central nervous system mediated by autoimmune and neurodegenerative pathogenic mechanisms. Multiple genes account for its moderate heritability, but the only genetic region shown to have a large replicable effect on MS susceptibility is the major histocompatibility complex (MHC). Strong linkage disequilibrium (LD) across the MHC has made it difficult to fully characterize individual genetic contributions of this region to MS risk in previous studies. African Americans are at a lower risk for MS when compared with northern Europeans and Americans of European descent, but greater haplotypic diversity and distinct patterns of LD suggest that this population may be particularly informative for fine-mapping efforts. To examine the role of the MHC in African American MS, a case–control association study was performed with 499 African American MS patients and 750 African American controls that were genotyped for 6040 MHC region single nucleotide polymorphisms (SNPs). A replication data set consisting of 451 African American patients and 718 African American controls was genotyped for selected SNPs. Two MHC class II SNPs, rs2647040 and rs3135021, were significant in the replication cohort and partially tagged *DRB1*15* alleles. Surprisingly, in comparison to similar studies of individuals of European descent, the MHC seems to play a smaller role in MS susceptibility in African Americans, consistent with pervasive genetic heterogeneity across ancestral groups, and may explain the difference in MS susceptibility between African Americans and individuals of European descent.

INTRODUCTION

Multiple sclerosis (MS) is a common demyelinating disease of the central nervous system mediated by autoimmune and neurodegenerative pathogenic mechanisms (1). Evidence of measurable genetic influence on susceptibility includes increased familial risk (2–4) independent of environment (5–7) and a concordance rate of 30% between monozygous twins (8). In addition, genome-wide scans in populations of European descent have identified MS-associated loci that were replicated across different cohorts (9–13), such as *IL-2Ra* [OR = 1.25, $P = 2.96 \times 10^{-8}$ (13)] and *IL7R* [OR = 1.18, $P = 2.94 \times 10^{-7}$ (13)], but the only genetic region that has been shown to have a large replicable effect is the major histocompatibility complex (MHC) on chromosome 6p21.3 [OR = 1.99, $P = 8.94 \times 10^{-81}$ (13); see (14,15) for review]. The major signal maps to the MHC class II sub-region and is in linkage disequilibrium (LD) with the extended *HLA-DQB1*06:02*, *HLA-DQA1*01:02*, *HLA-DRB1*15:01*, *HLA-DRB5*01:01* haplotype. A recent study of 1472 single nucleotide polymorphisms (SNPs) genotyped in the MHC region of MS patients and controls of European descent confirmed putative independent association signals in the telomeric class I and class III regions in addition to the

*To whom correspondence should be addressed at: Department of Neurology, University of California, San Francisco, 513 Parnassus Avenue, S256, Box 0435, San Francisco, CA 94143, USA. Tel: +1 4154762508; Fax: +1 4154765229; Email: joseph.mcelroy@ucsf.edu

© *The Author 2010. Published by Oxford University Press. All rights reserved.*
For Permissions, please email: journals.permissions@oxfordjournals.org

well-established class II association (16). However, the strong LD in the region has hindered the ability to further refine the major locus of susceptibility or to identify other independent smaller effects in populations of European descent.

Although MS is rare in Black Africans, African Americans are susceptible (17). On average, MS appears to take a more aggressive course in African Americans than in Northern Europeans (18–20). In African Americans, the *HLA-DRB1* MS susceptibility association was shown to be independent of the neighboring *HLA-DQB1* (21) and *HLA-DRB5* (22) genes. Interestingly, African origin DNA at the MHC was associated with disease severity (19) and the *HLA-DRB5*null* allele was associated with MS progression (22). These studies show the usefulness of African American data sets for dissecting the effect of the MHC genes in disease susceptibility. To further characterize the African American MHC region and to clarify the role of the MHC genes in MS, 6040 MHC SNPs were genotyped and analyzed in an African American MS data set.

RESULTS

Population parameters

After all quality control methods were implemented, 496 patients and 735 controls remained for analysis in the discovery sample (Table 1). For the patients, 75% were female, average percent European ancestry was 20% (median = 17.8%), average age of onset was 33 years, average disease duration was 10.6 years and 53% had a relapsing remitting disease course. In this data set, the average percent European ancestry in controls was 18.5% (median = 16.5%), which is significantly different from the cases (Van der Waerden Normal Quantiles test $P = 0.002$). To adjust for this variance, the first principal component, which is very significantly correlated with percent European ancestry ($r^2 = 0.96$; Spearman's rho $P < 0.0001$), was used as a covariate throughout the discovery analysis.

Linkage disequilibrium

Figure 1 displays the LD decay for each of the four populations: the MS discovery cohort, the CEU HapMap population (European), the YRI HapMap population (African) and the ASW HapMap population (African American). The LD decay of the MS cohort overlapped with that of the ASW population, and compared to the CEU and YRI populations, the African American populations had the lowest extent of LD in the MHC region (see also Supplementary Material, Fig. S1). This confirmed that the African American population exhibits lower LD than the European populations, which could facilitate the generation of a more refined map of MS association signals in the MHC. Interestingly, the YRI population had LD decay intermediate to that of the CEU and African American populations.

Association analysis in the discovery sample

A total of 4942 post-QC MHC SNPs were analyzed in 496 African American MS patients and 735 African American control individuals for association with MS. Figure 2A displays the $-\log_{10}$(P-values) of the SNPs across the MHC region. The most significantly associated SNP occurred at rs2647040 in the class II region ($P = 3 \times 10^{-5}$). To select SNPs for the replication analysis, rs2647040 was fit as a covariate in the next round of analyses (Fig. 2B), for which rs2772372 was the most significant SNP ($P = 5 \times 10^{-4}$). Both SNPs were then fit into a third analysis for the selection of the next SNP. This process was continued until 10 SNPs were selected for replication (Table 2). The selected SNPs span the MHC region (3 of the 10 SNPs were in the MHC class I, 3 in class II, 1 in class III, 2 in class I extended and 1 in class II extended regions).

In addition, given the relatively low level of overall statistical significance observed in African Americans compared with individuals of European descent, equivalence intervals were computed. For 21 SNPs in five LD blocks ($r^2 > 0.8$), the discovery cohort failed to show that the carrier frequency difference between cases and controls was lower than 12%, suggesting a lack of evidence for non-association with the trait. One SNP from each block was then taken for the

Table 1. Characteristics for the discovery and replication populations

Parameter	Discovery		Replication	
	MS patients	MS controls	MS patients	MS controls
n	496	735	451	718
Average age of onset (years)	33.08	NA	32.23	NA
Average disease duration (years)	10.63	NA	10.46	NA
% European Ancestry	0.20335	0.18461[a]	NA	NA
% Relapsing remitting and secondary progressive	87	NA	89	NA

[a]Indicates significant difference within population ($P < 0.01$).

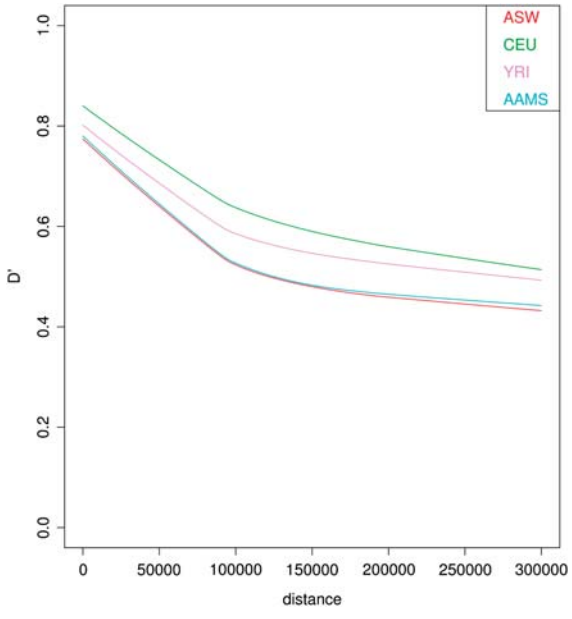

Figure 1. Linkage disequilibrium decay. LD (D') was calculated pair-wise between SNPs within 300 kb of each other using 2822 common polymorphic SNPs for 42 individuals from each population. HapMap populations: ASW, African American; CEU, European; YRI, Yoruban; AAMS, African American individuals in the current study.

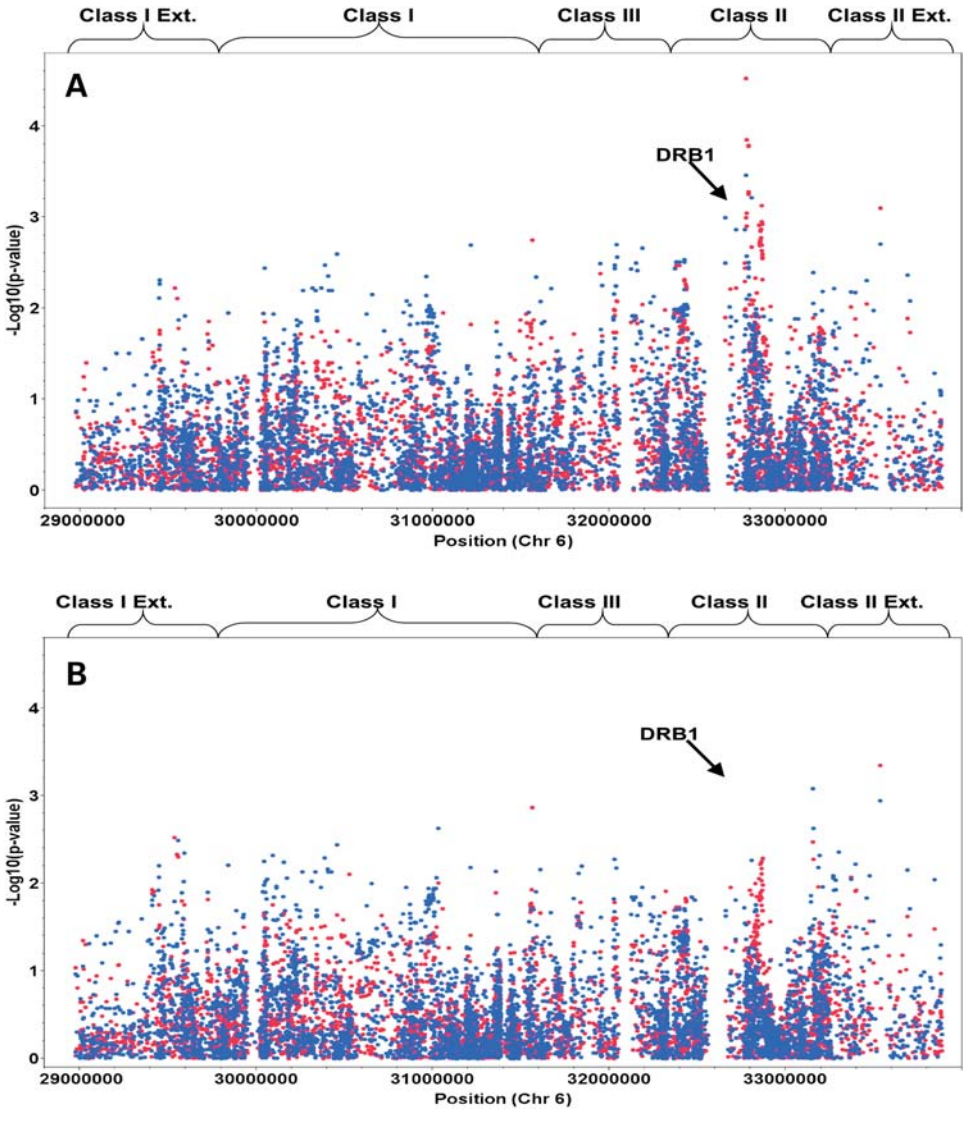

Figure 2. MHC-wide MS association analysis results in the discovery cohort. The first principal component, which was derived from principal component analysis of 4248 non-chromosome 6 SNPs and is highly correlated with percent European ancestry, was included as a covariate in the logistic regression models. Red and blue dots indicate the $-\log_{10}$ (P-value) for the genotypic and trend models, respectively. (**A**) No SNPs in the model. (**B**) Most significant SNP from first model in the model.

replication study. Three out of five of these SNPs were not included in previously identified set of 10 candidate SNPs (rs614549, rs12614 and rs2621419).

Replication

The 10 SNPs selected from difference analyses and the three selected from equivalency testing of the discovery population were genotyped in an independent cohort of 451 African American patients and 718 African American controls for replication (Table 2). Assays were unsuccessful for two SNPs rs7341211 and rs12614. The LD metrics (r^2 and D') between the 11 SNPs for which data were available is included in the Supplementary Material, Table S2. For replication SNPs chosen by difference testing, the largest r^2 (absolute value) was 0.24, although one equivalency testing selected SNP had a high r^2 with a difference testing SNP ($r^2 = 0.96$; rs2621419 and rs2261566, respectively). The only two SNPs significant [false discovery rate (FDR) (23) $P < 0.05$] in the replication cohort were rs2647040 and rs3135021, which are both class II region SNPs. Finally, the discovery and replication populations were combined to more accurately estimate effect sizes (Table 2). Although the P-values were smaller in the combined than the discovery analyses for 5 out of the 11 SNPs tested, the direction of the effects was similar in the discovery and replication analyses.

DRB1 allele tagging SNPs

Because HLA typing was only available for part of the data, adjustment and/or stratification for *HLA-DRB1*15:01* and/or *HLA-DRB1*15:03* was not possible. Nevertheless, on individuals typed for *HLA-DRB1*, rs2647040 and rs3135021 were tested for genotype correlation and LD with **15:01* and

Table 2. Analyses results for 13 SNPs chosen for replication based on results from difference testing and equivalency testing

Order into the models	SNP	Position	Gene	MHC class	Discovery Genotype *P*	Trend *P*	Replication Genotype FDR *P*	Trend FDR *P*	All Genotype *P*	Trend *P*	Trend odds ratio	Lower CL	Upper CL	Susceptible allele (alternate allele)
1	rs2647040	32,775,258	–	II	3.00E − 05	3.50E − 04	1.00E − 04[a]	1.60E − 05[a]	2.10E − 09	1.00E − 08	1.429	1.265	1.615	A(G)
2	rs2772372	33,535,328	–	II Extended	4.50E − 04	1.14E − 03	3.38E − 01	1.47E − 01	3.00E − 03	2.10E − 03	1.237	1.080	1.416	A(G)
3	rs3135021	33,153,536	*HLA-DPB1*	II	7.13E − 03	1.79E − 03	8.70E − 03[a]	1.87E − 02	1.00E − 04	2.00E − 04	0.785	0.691	0.891	G(A)
4	rs7341211	29,535,116	OR2H1	I Extended	3.32E − 03	1.47E − 02	NA[a]	NA	NA	NA	NA	NA	NA	NA
5	rs7753474	29,449,287	*OR12D3*	I Extended	4.61E − 03	1.58E − 03	1.62E − 01	6.16E − 02	3.20E − 03	8.00E − 04	0.796	0.696	0.910	G(A)
6	rs1029239	30,246,141	TRIM15	I	5.42E − 03	1.84E − 03	7.44E − 01	4.43E − 01	5.12E − 02	2.43E − 02	1.145	1.018	1.289	C(G)
7	rs707936	31,841,629	*C6ORF27*	III	9.27E − 03	2.56E − 03	1.32E − 01	4.42E − 02	7.50E − 03	3.20E − 03	0.803	0.694	0.929	G(A)
8	rs2261566	32,862,059	–	II	3.57E − 03	4.67E − 01	3.82E − 02	1.35E − 02	2.00E − 04	2.40E − 03	1.206	1.069	1.362	A(G)
9	rs9501035	31,020,393	–	I	5.91E − 03	1.40E − 03	2.16E − 01	2.72E − 01	2.48E − 02	1.19E − 02	0.803	0.676	0.953	G(A)
10	rs4711209	30,155,382	–	I	9.70E − 03	2.29E − 03	2.63E − 01	1.03E − 01	2.91E − 02	7.90E − 03	0.780	0.649	0.937	G(A)
Equivalency	rs614549	31,948,604	*SLC44A4*	III	4.14E − 03	3.21E − 03	4.10E − 01	2.48E − 01	5.00E − 03	5.00E − 03	0.844	0.751	0.950	A(G)
Equivalency	rs12614	32,022,158	*CFB*	III	1.83E − 02	1.35E − 02	NA	NA	NA	NA	NA	NA	NA	NA
Equivalency	rs2621419	32,848,984	–	II	1.22E − 03	3.05E − 02	1.04E − 01	5.28E − 02	5.00E − 04	5.10E − 03	1.193	1.054	1.350	A(G)

[a]Significant at false discovery rate = 0.05.

15:03 alleles. At the genotype level, rs2647040*A tags *HLA-DRB1*15* with a sensitivity (Se) 98.30% and a specificity (Sp) 46.76% (Table 3). At the haplotype level, the rs2647040*A–*HLA-DRB1*15* haplotype exhibits a high LD ($D' = 0.93$, $D' = 0.91$, $D' = 0.94$ for *HLA-DRB1*15*, **15:01* and **15:03*, respectively) because it is present on almost all the *HLA-DRB1*15* haplotypes (see haplotype frequencies in Table 4). The rs3135021*A SNP allele helps to refine the *HLA-DRB1*15:01* from the *HLA-DRB1*15:03* genotype by significantly better tagging ($P < 0.05$) the presence of *HLA-DRB1*15:03* Se = 62.56 versus Se = 43.04% for *HLA-DRB1*15:01*. The overall contribution of rs3135021 alone to **15:01/*15:03* tagging remained limited (Table 3). This is in line with the entrance of rs3135021 after rs2647040 in the sequential model. At the haplotype level, the coverage of *HLA-DRB1*15:01* and *HLA-DRB1*15:03* differs by the frequency of the SNP ($D' = 0.94$, $D' = 0.93$, $D' = 0.95$ for *HLA-DRB1*15*, **15:01* and **15:03*, respectively, Table 4). The haplotype frequencies presented in Table 4 suggest that several common SNPs interplay to capture the HLA polymorphism component when present in a model. In this case, rs3135021 subdivides the *HLA-DRB1*15:03-rs2647040*A* haplotype (13.3%) into two: **15:03-A-A* (7%) and **15:03-A-G* (6.3%). Similar observations are made by adding more MHC class II SNPs (data not shown).

DISCUSSION

The MHC class II region has been repeatedly shown to have a large association with MS susceptibility in European populations and their descendants (reviewed in 14), but the extent of LD in Europeans has been an obstacle in pinpointing the specific MHC genes associated. Utilizing populations with presumed lower LD may be a way to circumvent this problem. Therefore, an association study, reported herein, was performed using approximately 5000 extended MHC SNPs in 496 African American MS patients and 735 African American control individuals, with a follow-up analysis of selected SNPs in a replication data set of 451 African American MS patients and 718 African American MS controls.

As hypothesized, the extent of LD in the MHC was less in African Americans than in Europeans, showing that indeed the African American population is better suited to tease out the HLA effects on MS susceptibility than Europeans in regards to LD. Interestingly, while the African population had a lesser extent of LD than the European population, it had a greater extent of LD than the African American populations. This is likely a reflection of the recent influx of European haplotypes and alleles during admixing, thereby partially disrupting the ancestral African patterns of LD in the region.

Associations

SNPs in the MHC were tested for an association with MS susceptibility using difference and equivalency testing. While difference testing returns the probability of observing an association by chance, equivalency testing returns the probability of observing a lack of association by chance. The combination

Table 3. Tagging properties of rs2647040, and rs3135420 on *HLA-DRB1*15* alleles

HLA-DRB1	rs2647040 Sensitivity % (95% confidence interval)	Specificity % (95% confidence interval)	rs3135021 Sensitivity % (95% confidence interval)	Specificity % (95% confidence interval)
**15*	98.30% (96.79–99.10)	46.76% (43.67–49.88)	57.39% (53.13–61.54)	51.32% (48.20–54.42)
**1501*	98.10% (94.57–99.35)	34.46% (31.98–37.03)	43.04% (35.57–50.83)	47.28% (44.63–49.93)
**1503*	98.49% (96.75–99.31)	41.59% (38.74–44.51)	62.56% (57.71–67.18)	52.15% (49.22–55.06)

Table 4. Haplotype frequency of *HLA-DRB1*, rs2647040 and rs3135420

HLA-DRB1	rs2647040	rs3135021	Estimated frequency (%)
**non15:AB*	G	G	40.58
**non15:AB*	A	G	17.49
**non15:AB*	G	A	15.62
**non15:AB*	A	A	7.21
**15:03*	A	G	7.00
**15:03*	A	A	6.30
**15:03*	G	G	0.48
**15:01*	A	G	4.06
**15:01*	A	A	0.98
**15:01*	G	G	0.28

Haplotype frequency estimated by maximum likelihood implemented in the expectation maximization algorithm. **non15:AB* is any *HLA-DRB1* allele other than **15:01* or **15:03*.

of these two methods identifies the most likely SNPs to be associated, as well as SNPs whose association cannot be statistically ruled out even though they might not have the best *P*-values in the difference testing. Thirteen SNPs were selected for replication based on the difference and equivalency testing in the discovery cohort. Two of these SNPs, both class II SNPs, were significant (FDR $P < 0.05$) in the replication analyses. Ancestry informative SNP genotypes were not available for the replication population. It is possible that if an ancestry component was available to fit as a covariate for analysis of the replication population, more SNPs may have replicated. Both of the SNPs that did replicate show limited ability to tag *HLA-DRB1*15* alleles. Furthermore, we observed that in this admixed population multiple SNPs are needed to better capture the information carried by *HLA-DRB1* alleles, suggesting that low-frequency alleles associated with MS such as *HLA-DRB1*15:01* in African Americans might be difficult to identify using single SNPs. This contrasts to the observation that single SNPs are able to effectively tag the *HLA-DRB1*15:01* allele in European descended populations (24). Although only 64% ($n = 1526$) of the individuals in the data set had *HLA-DRB1* genotypes available, it is expected that the tagging results would be similar if *HLA-DRB1* genotypes were available for the entire population. Therefore, if all individuals had *HLA-DRB1* genotypes available, it would be interesting to fit allele count of *HLA-DRB1*15:01* into the models of analyses to determine if the partial tagging of *HLA-DRB1*15:01* by the two replicating markers is alone responsible for their significance.

Although the association of these two SNPs may be derived from their partial tagging of *HLA-DRB1*15* alleles, it is still of interest to discuss other loci close to the SNPs. The SNP rs2647040 is near another SNP associated with rheumatoid arthritis (25) (rs9275390; 1.9 kb away) and *HLA-DQB1* (32 kb away) is the nearest gene. In populations of European descent, *DRB1*15:01* is in high LD with *DQB1*06:02*, which hindered efforts to separate the effects of the two loci on MS susceptibility in those populations. However, Oksenberg *et al*. (21) utilized an African American MS data set to show that, at least in African Americans, *DRB1*15:01* had an association independent of *DQB1*06:02*, and that no significant association of *DQB1*06:02* independent of *DRB1*15:01* was identified. However, it is still possible that other alleles of *DQB1* are associated independently of *DRB1* with MS susceptibility in African Americans (26,27).

The SNP rs3135021 is located in an intron of *HLA-DPB1*. This locus was shown to have a strong association with MS susceptibility in a Chinese data set (28) in which there was a lack of association between *DRB1*15:01* and MS susceptibility. In Japanese data sets, *DRB1*15:01* was associated only with conventional MS, whereas *DPB1*05:01* was associated with opticospinal MS (29). Similarly, an opticospinal form of MS is more common in African Americans than in individuals of European descent and, unlike typical MS, is not associated with *HLA-DRB1*15* alleles (18,30). It is therefore possible that the association of rs3135021 with MS in African Americans may be through *HLA-DPB1* rather than *HLA-DRB1* alleles.

Comparing to studies in individuals of European descent

The MHC difference testing association analyses in the discovery cohort yielded no associations with MS susceptibility that were significant at an FDR $P = 0.1$. This is surprising considering the magnitude of the associations of the MHC in studies of MS susceptibility in European populations (10,13,16). The early success in identifying the relevance of the HLA reflects the strong effect of the MHC region, where nominally significant association was detected with just 32 cases (31) and a highly significant association was shown with fewer than 200 (32). The combined discovery and replication population in the current study consisted of 947 African American cases and 1453 African American controls, which is comparable in size to the case/control data set analyzed by Baranzini *et al*. (10) (978 patients and 883 controls of European descent). The most significant SNP in the MHC region in the Baranzini study was rs3129934, which had a *P*-value of 8.88×10^{-34} and an odds ratio of 2.657 (95%: 2.262–3.122) using only 1169 MHC SNPs. In the current study, which used almost five times as many MHC region SNPs, the most significant SNP had a *P*-value of 2.1×10^{-9} with an odds ratio of 1.429 (95%: 1.265–1.615). This suggests that the MHC influences MS susceptibility to a lesser extent

in African Americans than individuals of European descent, which may explain the lower prevalence of MS in African Americans. These results may be, in part, due to the difference in risk allele frequency between the two populations. As the frequency of a risk allele gets lower, the power to significantly detect differences decreases (Supplementary Material, Fig. S3). Alternatively, the difference in significance and effect sizes may be due to reduced LD in the region in African Americans resulting in lower total coverage. However, since almost five times as many SNPs were in the region in the current study, low LD is unlikely to have played a role in the modest MHC effect.

Conclusions

Results from analyses of discovery and replication populations together suggest that the MHC is less associated in African Americans than in individuals of European descent, indicating that part of the association observed in African Americans may come from European DNA which is diluted in the great genetic variation exhibited by the African Americans. Indeed, the large amount of genetic variation present in African Americans results in the population having a lesser extent of LD than either populations of European descent or African populations. Results from the *HLA-DRB1*15* tagging analyses for the two replicated SNPs indicate that the MHC association seen in this population is likely partially due to the effects of *HLA-DRB1*, which is the major susceptibility locus in individuals of European descent. However, given the differences between MS in African Americans and individuals of European descent in terms of disease severity and prevalence, it is possible that the genetic control of susceptibility might also be different between the two populations. This difference may be partially reflected in the weaker, yet significant, association of the MHC class II region with MS in African Americans identified in this study.

MATERIALS AND METHODS

Subjects

The discovery data set consisted of DNAs from 499 African American MS patients and 750 healthy controls genotyped for 6040 MHC region SNPs using a custom Infinium iSelect HD Custom Genotyping BeadChip (Illumina, Inc.). The MHC SNPs were selected based on previously described methods (33,34), having an Illumina design score of one or greater, in complete disequilibrium ($r^2 = 1$) with HapMap II SNPs and with a 5% or greater minor allele frequency. In addition, we included 'common untagged' SNPs from the Wellcome Trust MHC sequencing project (35) and selected ancestry informative SNPs in the MHC region. In addition, 4431 non-chromosome 6 SNPs were genotyped for population stratification correction on the same platform. These SNPs were selected based on ability to distinguish ancestry between major racial groups and subgroups (36,37). An independent data set consisting of 451 MS patients and 718 control individuals was available for the replication study of SNPs chosen from the results of the discovery analyses. Recruitment and demographic characteristics of all MS patients and part of the control individuals for both the discovery and replication data sets are reported elsewhere (19,21). The control individuals were supplemented with African American control individuals from the CLEAR Registry (discovery $n = 398$, replication $n = 404$) (38). All MS patients met the International Panel MS diagnostic criteria (39). African American ancestry was self reported. The UCSF institutional review board approved this study and all participants gave written informed consent. All quality control and analyses were completed using JMP Genomics 4.0 (SAS Institute Inc.) and R version 2.9 (40), unless otherwise noted. SNP names herein correspond to NCBI dbSNP build 130:human_9606 (41).

Quality control

Individuals were removed from the analysis if they had greater than 5% failed genotype calls. Markers had to meet several criteria to be included in the study: GenCall score (Illumina's genotype quality measure) >0.2, $<10\%$ missing genotypes, minor allele frequency $>2\%$ and Hardy–Weinberg equilibrium P-value > 0.001 in controls. In addition, markers were tested (chi-square) for differences in missing genotypes between cases and controls, and only markers with a P-value >0.001 were included in the study. After this standard quality control, 497 patients and 743 controls genotyped for 4942 SNPs across the MHC remained for the analysis. Additionally, to confirm the consistency of the control data, an association analysis was carried out in the discovery data set between the MS controls and the Clear Registry controls to identify MHC SNPs differing in allele frequency between the two groups. No SNPs were significantly different between the two groups [false discovery rate (23) $P > 0.6$].

Principal components analysis

For additional quality control, 4248 (post-standard quality control; Supplementary Material, Table S4) non-chromosome 6 SNPs were genotyped on all individuals to identify outliers, determine percent European ethnicity and correct for possible population stratification. These markers were used to generate the first 10 principal components for the population. Based on the scree plot (Supplementary Material, Fig. S5), only the first component, which accounts for 1.6% of the genetic variation captured by the SNPs, was of interest. For a subset ($n = 542$) of individuals for which the genome-wide percent European ancestry was previously calculated with ancestry informative markers (22), percent European was regressed onto the first component to verify that the component was capturing European ancestry. The resulting regression equation ($r^2 = 0.96$) was then used to calculate percent European ancestry for the full data set. Plotting of individuals onto their first component (Supplementary Material, Fig. S6) leads to the identification of nine outliers (one patient and eight controls) that were removed from the analysis because of unexpectedly high European ancestry, resulting in a total of 496 patients and 735 controls remaining for the analysis. The first component was retained for use as a covariate in the association analyses to correct for ancestry differences.

Linkage disequilibrium

For further characterization of the African American MHC region, genotypic data downloaded from the HapMap (42) website for the Yoruban (YRI), European (CEU) and African American (ASW) populations were used to compare the extent of LD in the MHC region between African American, African and European populations. Common polymorphic MHC SNPs ($n = 2822$) from the four populations (selected from 5044 SNPs in the MS population) were used for the analysis. In order to adjust for potential sample size effects, a random subset of unrelated individuals was selected from each of the MS, YRI and CEU populations to match the number of unrelated individuals in the ASW population ($n = 42$; the smallest of the four populations). LD (D') was calculated with Haploview (43) for each of the populations using a maximum marker distance of 300 kb. The 'plot' and 'lowess' options in the R statistical software (40) were used to generate the pair-wise LD by distance plot for the MS, ASW, YRI and CEU populations.

Genetic association analysis

Logistic regression was employed to identify associations between SNP genotypes and patient/control status. Both the 'genotypic', which considers genotype as a discrete variable, and 'trend', which assumes that heterozygotes are phenotypically intermediate the two homozygotes, models were used. The first principal component was included in the models as a covariate to correct for ancestry effects. The most significant SNP from the primary analysis was used as a covariate in subsequent analyses. The most significant SNP from the secondary analysis was then included in the model with the first SNP for another analysis. This process was carried out seven more times, with the final model containing 9 covariate SNPs, to select 10 SNPs for replication. Selecting the SNPs in this manner prevents the retention of SNPs that are redundantly associated with susceptibility through LD to a more significant SNP. Because a replication population of approximately the same size as the discovery population was available, multiple testing corrections were not considered for the analyses, and therefore the P-values given are uncorrected.

Refining the strength of association signal

Given the strength of the association between SNPs and MS observed in persons of European ancestry, statistical methods used in equivalence trials were employed in an attempt to demonstrate that, if existing, the MS association is lower in African Americans than individuals of European descent. We implemented a pooled variance method for binomial variables following Barker *et al*. (44). Considering a minimal carrier frequency difference between cases and controls for rs3135388 (the *HLA-DRB1*15:01* tagging SNP) in individuals of European descent of 12% [conservatively estimated from (45)], we defined an equivalence interval as defined by Schuirmann (46) at the nominal significance level of 5%. Equivalence testing was applied to determine the extent that our discovery analysis has the power to identify the absence of a major genetic association for each SNP tested. For example, a non-significant SNP in equivalency testing suggests that a larger sample size would be required to rule out that, if existing, the frequency difference between cases and controls is lower than the threshold. Technically, this is achieved by reversing the null and alternate hypotheses of classical difference testing. Independent SNPs ($r^2 < 0.8$) which fail to pass this equivalence tests were included in the replication analysis. Applied on both the discovery and the replication cohorts, such a method locally evaluates the power of our study to detect or exclude a strong association with MS in the African American MHC region.

Replication

As previously mentioned, an independent population consisting of 451 African American MS patients and 718 African American control individuals was available for the replication study of SNPs chosen from the results of the discovery analyses. SNP genotyping was completed in the replication data set using ABI custom TaqMan assays designed on File Builder 3.0 software and TaqMan predesigned SNP genotyping assays. TaqMan SNP genotyping assays are conducted in 384-well plates using TaqMan Universal PCR Master Mix on an ABI 7900HT Sequence Detection System using SDS 2.3 software. The statistical models used were the same as in the analysis of the discovery cohort excluding the principal component which could not be calculated for the replication cohort because genotypes for the non-chromosome 6 ancestry SNPs were not available. Each SNP was analyzed with no other SNPs in the model, and an FDR P-value of 0.1 within model was considered the threshold for successful replication.

DRB1 allele tagging

A subset ($n = 1526$) of the discovery and replication populations had known *HLA-DRB1*15:01* and *HLA-DRB1*15:03* statuses. To determine if any of the SNPs selected for replication were associated with MS through LD to these two *HLA-DRB1* alleles, an analysis was performed to identify associations between the SNPs and *HLA-DRB1*15:01* and *HLA-DRB1*15:03* status. Sensitivity (Se) and specificity (Sp) of the SNPs to predict *HLA-DRB1* alleles at the genotype level were computed. Confidence levels are given at 95%. In addition, haplotype frequencies were computed by maximum likelihood methods (47,48) in order to investigate long range LD patterns.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank the MS patients and controls that participated in this study. We thank R.R. Lincoln, R. Guerrero, H. Mousavi and A. Santaniello for sample and database management. We thank R. Gomez and E. Tailor for recruitment of participants. We thank Professor C. Cazaux, PhD for insightful discussion. We also thank the IMAGEN consortium and the CLEAR

Registry (see http://medicine.uab.edu/rheum/70918/) for access to samples and data.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by grants from the National Institute of Health U19AI067152 to S.L.H., RO1NS46297 to J.R.O., K23NS048869 to B.A.C.C. and National Multiple Sclerosis Society RG3060C8 to J.R.O. J.P.M. is a post-doctoral fellow supported by the National MS Society.

REFERENCES

1. Hauser, S.L. and Oksenberg, J.R. (2006) The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration. *Neuron*, **52**, 61–76.
2. Carton, H., Vlietinck, R., Debruyne, J., De Keyser, J., D'Hooghe, M.B., Loos, R., Medaer, R., Truyen, L., Yee, I.M. and Sadovnick, A.D. (1997) Risks of multiple sclerosis in relatives of patients in Flanders, Belgium. *J. Neurol. Neurosurg. Psychiatry*, **62**, 329–333.
3. Robertson, N.P., Fraser, M., Deans, J., Clayton, D., Walker, N. and Compston, D.A. (1996) Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. *Brain*, **119**, 449–455.
4. Sadovnick, A.D. (1993) Familial recurrence risks and inheritance of multiple sclerosis. *Curr. Opin. Neurol. Neurosurg.*, **6**, 189–194.
5. Ebers, G.C., Sadovnick, A.D. and Risch, N.J. (1995) A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. *Nature*, **377**, 150–151.
6. Ebers, G.C., Yee, I.M., Sadovnick, A.D. and Duquette, P. (2000) Conjugal multiple sclerosis: population-based prevalence and recurrence risks in offspring. Canadian Collaborative Study Group. *Ann. Neurol.*, **48**, 927–931.
7. Sadovnick, A.D., Ebers, G.C., Dyment, D.A. and Risch, N.J. (1996) Evidence for genetic basis of multiple sclerosis. The Canadian Collaborative Study Group. *Lancet*, **347**, 1728–1730.
8. Hawkes, C. and Macgregor, A. (2009) Twin studies and the heritability of MS: a conclusion. *Mult. Scler.*, **15**, 661–667.
9. Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene). (2009) Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat. Genet.*, **41**, 824–828.
10. Baranzini, S.E., Wang, J., Gibson, R.A., Galwey, N., Naegelin, Y., Barkhof, F., Radue, E.W., Lindberg, R.L., Uitdehaag, B.M., Johnson, M.R. *et al.* (2009) Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum. Mol. Genet.*, **18**, 767–778.
11. Comabella, M., Craig, D.W., Camina-Tato, M., Morcillo, C., Lopez, C., Navarro, A., Rio, J., Montalban, X. and Martin, R. (2008) Identification of a novel risk locus for multiple sclerosis at 13q31.3 by a pooled genome-wide scan of 500,000 single nucleotide polymorphisms. *PLoS ONE*, **3**, e3490.
12. De Jager, P.L., Jia, X., Wang, J., de Bakker, P.I., Ottoboni, L., Aggarwal, N.T., Piccio, L., Raychaudhuri, S., Tran, D., Aubin, C. *et al.* (2009) Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat. Genet.*, **41**, 776–782.
13. Hafler, D.A., Compston, A., Sawcer, S., Lander, E.S., Daly, M.J., De Jager, P.L., de Bakker, P.I., Gabriel, S.B., Mirel, D.B., Ivinson, A.J. *et al.* (2007) Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.*, **357**, 851–862.
14. McElroy, J.P. and Oksenberg, J.R. (2008) Multiple sclerosis genetics. *Curr. Top. Microbiol. Immunol.*, **318**, 45–72.
15. Oksenberg, J.R., Baranzini, S.E., Sawcer, S. and Hauser, S.L. (2008) The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. *Nat. Rev. Genet.*, **9**, 516–526.
16. Rioux, J.D., Goyette, P., Vyse, T.J., Hammarstrom, L., Fernando, M.M., Green, T., De Jager, P.L., Foisy, S., Wang, J., de Bakker, P.I. *et al.* (2009) Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc. Natl Acad. Sci. USA*, **106**, 18680–18685.
17. Marrie, R.A., Cutter, G., Tyry, T., Vollmer, T. and Campagnolo, D. (2006) Does multiple sclerosis-associated disability differ between races? *Neurology*, **66**, 1235–1240.
18. Cree, B.A., Khan, O., Bourdette, D., Goodin, D.S., Cohen, J.A., Marrie, R.A., Glidden, D., Weinstock-Guttman, B., Reich, D., Patterson, N. *et al.* (2004) Clinical characteristics of African Americans vs Caucasian Americans with multiple sclerosis. *Neurology*, **63**, 2039–2045.
19. Cree, B.A., Reich, D.E., Khan, O., De Jager, P.L., Nakashima, I., Takahashi, T., Bar-Or, A., Tong, C., Hauser, S.L. and Oksenberg, J.R. (2009) Modification of multiple sclerosis phenotypes by African ancestry at HLA. *Arch. Neurol.*, **66**, 226–233.
20. Weinstock-Guttman, B., Jacobs, L.D., Brownscheidle, C.M., Baier, M., Rea, D.F., Apatoff, B.R., Blitz, K.M., Coyle, P.K., Frontera, A.T., Goodman, A.D. *et al.* (2003) Multiple sclerosis characteristics in African American patients in the New York State Multiple Sclerosis Consortium. *Mult. Scler.*, **9**, 293–298.
21. Oksenberg, J.R., Barcellos, L.F., Cree, B.A., Baranzini, S.E., Bugawan, T.L., Khan, O., Lincoln, R.R., Swerdlin, A., Mignot, E., Lin, L. *et al.* (2004) Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *Am. J. Hum. Genet.*, **74**, 160–167.
22. Caillier, S.J., Briggs, F., Cree, B.A., Baranzini, S.E., Fernandez-Vina, M., Ramsay, P.P., Khan, O., Royal, W. 3rd, Hauser, S.L., Barcellos, L.F. *et al.* (2008) Uncoupling the roles of HLA-DRB1 and HLA-DRB5 genes in multiple sclerosis. *J. Immunol.*, **181**, 5473–5480.
23. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, **57**, 289–300.
24. de Bakker, P.I., McVean, G., Sabeti, P.C., Miretti, M.M., Green, T., Marchini, J., Ke, X., Monsuur, A.J., Whittaker, P., Delgado, M. *et al.* (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.*, **38**, 1166–1172.
25. Julia, A., Ballina, J., Canete, J.D., Balsa, A., Tornero-Molina, J., Naranjo, A., Alperi-Lopez, M., Erra, A., Pascual-Salcedo, D., Barcelo, P. *et al.* (2008) Genome-wide association study of rheumatoid arthritis in the Spanish population: KLF12 as a risk locus for rheumatoid arthritis susceptibility. *Arthritis Rheum.*, **58**, 2275–2286.
26. Kaushansky, N., Altmann, D.M., Ascough, S., David, C.S., Lassmann, H. and Ben-Nun, A. (2009) HLA-DQB1*0602 determines disease susceptibility in a new 'humanized' multiple sclerosis model in HLA-DR15 (DRB1*1501;DQB1*0602) transgenic mice. *J. Immunol.*, **183**, 3531–3541.
27. Mangalam, A., Luckey, D., Basal, E., Jackson, M., Smart, M., Rodriguez, M. and David, C. (2009) HLA-DQ8 (DQB1*0302)-restricted Th17 cells exacerbate experimental autoimmune encephalomyelitis in HLA-DR3-transgenic mice. *J. Immunol.*, **182**, 5131–5139.
28. Wu, X.M., Wang, C., Zhang, K.N., Lin, A.Y., Kira, J., Hu, G.Z., Qu, X.H., Xiong, Y.Q., Cao, W.F. and Gong, L.Y. (2009) Association of susceptibility to multiple sclerosis in Southern Han Chinese with HLA-DRB1, -DPB1 alleles and DRB1-DPB1 haplotypes: distinct from other populations. *Mult. Scler.*, **15**, 1422–1430.
29. Kira, J. (2003) Multiple sclerosis in the Japanese population. *Lancet Neurol.*, **2**, 117–127.
30. Cree, B. (2008) Neuromyelitis optica: diagnosis, pathogenesis, and treatment. *Curr. Neurol. Neurosci. Rep.*, **8**, 427–433.
31. Compston, D.A., Batchelor, J.R. and McDonald, W.I. (1976) B-lymphocyte alloantigens associated with multiple sclerosis. *Lancet*, **2**, 1261–1265.
32. Olerup, O. and Hillert, J. (1991) HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. *Tissue Antigens*, **38**, 1–15.
33. de Bakker, P.I., Burtt, N.P., Graham, R.R., Guiducci, C., Yelensky, R., Drake, J.A., Bersaglieri, T., Penney, K.L., Butler, J., Young, S. *et al.* (2006) Transferability of tag SNPs in genetic association studies in multiple populations. *Nat. Genet.*, **38**, 1298–1303.
34. Leslie, S., Donnelly, P. and McVean, G. (2008) A statistical method for predicting classical HLA alleles from SNP data. *Am. J. Hum. Genet.*, **82**, 48–56.
35. Stewart, C.A., Horton, R., Allcock, R.J., Ashurst, J.L., Atrazhev, A.M., Coggill, P., Dunham, I., Forbes, S., Halls, K., Howson, J.M. *et al.* (2004) Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res.*, **14**, 1176–1187.

36. Tian, C., Plenge, R.M., Ransom, M., Lee, A., Villoslada, P., Selmi, C., Klareskog, L., Pulver, A.E., Qi, L., Gregersen, P.K. *et al.* (2008) Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet.*, **4**, e4.
37. Tian, C., Gregersen, P.K. and Seldin, M.F. (2008) Accounting for ancestry: population substructure and genome-wide association studies. *Hum. Mol. Genet.*, **17**, R143–R150.
38. Bridges, S.L. Jr, Hughes, L.B., Mikuls, T.R., Howard, G., Tiwari, H.K., Alarcon, G.S., McNicholl, J.M. and Moreland, L.W. (2003) Early rheumatoid arthritis in African-Americans: the CLEAR Registry. *Clin. Exp. Rheumatol.*, **21**, S138–S145.
39. McDonald, W.I., Compston, A., Edan, G., Goodkin, D., Hartung, H.P., Lublin, F.D., McFarland, H.F., Paty, D.W., Polman, C.H., Reingold, S.C. *et al.* (2001) Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann. Neurol.*, **50**, 121–127.
40. R Development Core Team. (2009) R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria.
41. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
42. The International HapMap Consortium. (2003) The International HapMap Project. *Nature*, **426**, 789–796.
43. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
44. Barker, L., Rolka, H., Rolka, D. and Brown, C. (2001) Equivalence testing for binomial random variables. *Amer. Statistician*, **55**, 279–287.
45. Rubio, J.P., Stankovich, J., Field, J., Tubridy, N., Marriott, M., Chapman, C., Bahlo, M., Perera, D., Johnson, L.J., Tait, B.D. *et al.* (2008) Replication of KIAA0350, IL2RA, RPL5 and CD58 as multiple sclerosis susceptibility genes in Australians. *Genes. Immun.*, **9**, 624–630.
46. Schuirmann, D.J. (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biopharm.*, **15**, 657–680.
47. Gourraud, P.A., Gagne, K., Bignon, J.D., Cambon-Thomsen, A. and Middleton, D. (2007) Preliminary analysis of a KIR haplotype estimation algorithm: a simulation study. *Tissue Antigens*, **69** (Suppl. 1), 96–100.
48. Gourraud, P.A., Genin, E. and Cambon-Thomsen, A. (2004) Handling missing values in population data: consequences for maximum likelihood estimation of haplotype frequencies. *Eur. J. Hum. Genet.*, **12**, 805–812.