An ImmunoChip study of multiple sclerosis risk in African Americans

Noriko Isobe,1,2 Lohith Madireddy,1 Pouya Khankhanian,1 Takuya Matsushita,1,3 Stacy J. Caillier,1 Jayaji M. Moré,1 Pierre-Antoine Gourraud,1 Jacob L. McCauley,4 Ashley H. Beecham,4 International Multiple Sclerosis Genetics Consortium, Laura Piccio,5 Joseph Herbert,6 Omar Khan,7 Jeffrey Cohen,8 Lael Stone,8 Adam Santaniello,1 Bruce A. C. Cree,1 Suna Onengut-Gumuscu,9 Stephen S. Rich,9 Stephen L. Hauser,1 Stephen Sawcer10,* and Jorge R. Oksenberg1,*

*These authors contributed equally to this work.

The aims of this study were: (i) to determine to what degree multiple sclerosis-associated loci discovered in European populations also influence susceptibility in African Americans; (ii) to assess the extent to which the unique linkage disequilibrium patterns in African Americans can contribute to localizing the functionally relevant regions or genes; and (iii) to search for novel African American multiple sclerosis-associated loci. Using the ImmunoChip custom array we genotyped 803 African American cases with multiple sclerosis and 1516 African American control subjects at 130 135 autosomal single nucleotide polymorphisms. We conducted association analysis with rigorous adjustments for population stratification and admixture. Of the 110 non-major histocompatibility complex multiple sclerosis-associated variants identified in Europeans, 96 passed stringent quality control in our African American data set and of these, 470% (69) showed over-representation of the same allele amongst cases, including 21 with nominally significant evidence for association (one-tailed test \( P \leq 0.05 \)). At a further eight loci we found nominally significant association with an alternate correlated risk-tagging single nucleotide polymorphism from the same region. Outside the regions known to be associated in Europeans, we found seven potentially associated novel candidate multiple sclerosis variants \( (P < 10^{-4}) \), one of which (rs2702180) also showed nominally significant evidence for association (one-tailed test \( P = 0.034 \)) in an independent second cohort of 620 African American cases and 1565 control subjects. However, none of these novel associations reached genome-wide significance (combined \( P = 6.3 \times 10^{-5} \)). Our data demonstrate substantial overlap between African American and European multiple sclerosis variants, indicating common genetic contributions to multiple sclerosis risk.

1 Department of Neurology, School of Medicine, University of California, San Francisco, CA 94158, USA
2 Division of Neurology, Department of Internal Medicine, Saga University Faculty of Medicine, Saga, Saga 849-8501, Japan
3 Department of Neurological Therapeutics, Neurological Institute, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Fukuoka 812-8582, Japan
4 John P. Hussman Institute for Human Genomics and The Dr John T Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, FL 33136, USA
5 Department of Neurology, Washington University School of Medicine, St. Louis, MO 63108, USA
6 Department of Neurology, New York University School of Medicine, New York, NY 10016, USA
7 Multiple Sclerosis Centre and The Sastry Foundation Advanced Imaging Laboratory, Department of Neurology, Wayne State University School of Medicine, Detroit, MI 48201, USA
8 Mellen Centre for Multiple Sclerosis Treatment and Research, Cleveland Clinic, Cleveland, OH 44195, USA
9 Centre for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA
10 Department of Clinical Neurosciences, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0QQ, UK
Introduction

Multiple sclerosis is a chronic, inflammatory disease of the CNS and a common cause of neurological disability in young adults (Hauser and Goodin, 2012). Its modest heritability reflects complex polygenic effects and, most likely, gene-environment interactions [Simon et al., 2011; International Multiple Sclerosis Genetics Consortium (IMSGC), 2013a; Sawcer et al., 2014]. The results from genome-wide association studies (GWAS) have clarified important aspects of multiple sclerosis pathogenesis and provided strong empirical support for a model of inheritance driven primarily by allelic variants that are relatively common in the general population [Oksenberg and Baranzini, 2010; IMSGC and Wellcome Trust Case Control Consortium 2 (WTCCC2), 2011]. The strongest susceptibility signal genome-wide maps to HLA-DRB1 in the class II region of the major histocompatibility complex (MHC, 6p21.3) and explains up to 10.5% of the genetic variance underlying risk. The HLA association implies that mechanistically, multiple sclerosis clusters with other antigen-specific autoimmune diseases, a hypothesis supported by the observation that the non-MHC associated variants appear to locate predominantly in or near genes influencing the function of the adaptive immune system (IMSGC and WTCCC2, 2011). Interestingly, some of the non-MHC allelic variants associated with multiple sclerosis have also emerged in GWAS of other autoimmune diseases (IMSGC and WTCCC2, 2011; Cotsapas et al., 2011), suggesting that common underlying risk mechanisms might exist across multiple immune-related conditions. To better describe this overlap and refine the regions of interest in susceptibility loci, a mega-consortium was established to conduct cost-effective candidate loci association studies across multiple autoimmune diseases using a common, high-coverage single nucleotide polymorphisms (SNPs) array known as the ImmunoChip (Cortes and Brown, 2011). The chip was designed in 2010 and 207,728 variants were considered for inclusion, of which 196,524 passed manufacturing quality control (192,402 autosomal, 1,595 X-linked, 1,735 Y-linked, 791 pseudoautosomal and one mitochondrial). The multiple sclerosis input to the content came from two sources: an early analysis of a well-powered GWAS (IMSGC and WTCCC2, 2011) and a meta-analysis of previously published smaller GWAS (Patsopoulos et al., 2011).

Typing the ImmunoChip in a new independent data set identified 48 novel multiple sclerosis susceptibility variants with genome-wide significance (IMSGC, 2013b). These results considerably enhanced the roster of validated risk loci and demonstrated the discovery power of this array that has been similarly effective in other autoimmune diseases (Trynka et al., 2011; Cooper et al., 2012; Eyre et al., 2012; Jostins et al., 2012; Juran et al., 2012; Liu et al., 2012; Tsai et al., 2012; Hinks et al., 2013). However, the utility of this platform in non-Europeans remains to be addressed. In addition, consistent with their longer evolutionary history, populations of African origin are known to have, on average, characteristically smaller blocks of linkage disequilibrium compared to populations with European ancestry (Tishkoff and Kidd, 2004), implying that the study of populations of African origin could help to narrow the regions of interest and assist in identifying causative variants (Buyske et al., 2012; Gong et al., 2013).

Notwithstanding difficulties in surveillance, multiple sclerosis is almost non-existent in black Africans and early estimates suggested that the disease was significantly less prevalent in African Americans than in European Americans (relative risk of 0.64; Wallin et al., 2004). However, contemporary studies are challenging the long-held belief that African Americans are at a reduced risk for developing multiple sclerosis (Wallin et al., 2012; Langer-Gould et al., 2013). Furthermore, compared with whites, African Americans are more likely to have a more severe disease course, which at least in part appears to be genetically determined (Buchanan et al., 2004; Cree et al., 2004, 2009; Boster et al., 2009; Kimbrough et al., 2014). Here, we applied the ImmunoChip to a well-curated African American multiple sclerosis data set to investigate: (i) whether European multiple sclerosis-associated variants...
are also associated with the disease in African Americans; (ii) whether the smaller haplotype blocks, characteristic of African American genomes, can contribute to better mapping of the functionally relevant variants driving the association; and (iii) whether the array can identify novel multiple sclerosis-associated loci in African American patients.

Materials and methods

Patients and control subjects

The core screening data set consists of 842 de-identified DNA samples from African American cases and 498 African American controls. All multiple sclerosis subjects met established diagnostic criteria (McDonald et al., 2001; Polman et al., 2011). Ascertainment protocols and clinical and demographic characteristics have been summarized elsewhere (Cree et al., 2004; Oksenberg et al., 2004). All study participants are self-reported African Americans. Additionally, data from 1114 African American control subjects were provided by the International Consortium on the Genetics of Systemic Lupus Erythematosus (SLEGEN), totalling 842 cases and 1612 controls. The University of California at San Francisco Institutional Review Board approved this study.

SNP genotyping and quality control

SNP genotyping was conducted using the ImmunoChip, an Illumina Infinium HD custom array at the Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK for the cases and at the Centre for Public Health Genomics, University of Virginia, Charlottesville, VA, USA for the controls. Standard quality control measures were implemented using PLINK v1.07 (Purcell et al., 2007). SNPs with missing rates higher than 2%, Hardy-Weinberg proportion test \( P < 10^{-5} \) in controls and \( P < 10^{-8} \) in cases, and distinct missing proportion between cases and controls with \( P < 10^{-3} \) were excluded. For the further analysis, 130248 autosomal SNPs remained, including 96 of 110 SNPs known to be associated in Europeans (IMSGC, 2013b). Samples were excluded for missing genotyping rates exceeding 2%, extreme autosomal heterozygosity of \( > 3 \) standard deviations (SD), or excessive Identity By Descent (IBD) with \( \text{PI}_\text{HAT} > 0.20 \).

Population stratification

Principal component (PC) analyses were used to assess ancestry and control for the effects of population stratification. Principal component analysis was conducted using pruned autosomal non-MHC SNPs with minor allele frequency \( > 1\% \) and pairwise \( r^2 < 0.1 \) with a window size of 100 SNPs (25 408 SNPs). The scree plot indicated that PC1 explained the vast majority of the variance in the African American data set (Supplementary Fig. 1A). According to the plot, PC1 were used to remove two outlier samples with the values outside \( \pm 6 \) SD and all following association analyses were conducted using PC1 as a covariate. PC1 values were highly correlated with the previously reported percentage of African ancestry \( (r = 0.99, P < 2.2 \times 10^{-16}) \) (Reich et al., 2005; Isobe et al., 2013). When the PC1 components of individual samples were compared with each other, the control samples from SLEGEND had higher PC1 values compared to the UCSF cases and controls \( (P = 2.53 \times 10^{-27} \) and \( 1.77 \times 10^{-16} \), respectively), suggesting the proximity of SLEGEN controls to African ancestry (Supplementary Fig. 1B). Thus, to eliminate association signals derived from the different population admixture levels between the two control groups, association was analysed between the two with PC1 as a covariate, which identified 113 SNPs with \( P \)-values \( < 10^{-3} \) to be removed. Finally, 130 135 autosomal SNPs remained for the following analysis. Another principal component analysis was conducted including samples from the 1000 Genome Project (The 1000 Genomes Project Consortium, 2010) as a reference, with commonly available autosomal SNPs pruned with the same criteria as above (24 994 SNPs). From this principal component analysis, an additional 22 samples located far from the relevant reference populations were removed (Fig. 1). Ultimately, 803 African American multiple sclerosis cases and 1516 healthy African American control subjects remained for further analysis. Principal component analyses were performed using the R package SNPRelate (Zheng et al., 2012).

Association analysis

Following quality control analyses, association tests were conducted assuming the additive effect of the allele for the affection status with PC1 as a covariate to control for population stratification and admixture. First, the replication status of 96 European multiple sclerosis SNPs (IMSGC, 2013b) was evaluated using one-tailed tests. For each multiple sclerosis variant, the Cochrane Heterogeneity Q Test was also performed to test effect size differences between African Americans and Europeans. Additionally, SNPs with association \( P \)-values \( < 10^{-4} \) locating outside 2 Mb (1 Mb centromeric and 1 Mb telomeric) flanking the European multiple sclerosis-associated SNPs were nominated as candidates for novel multiple sclerosis-associated variants in African Americans. All association tests for genotyped SNPs were conducted using PLINK (v1.07) (Purcell et al., 2007). Power calculations were performed using Bioconductor’s GeneticsDesign package version 1.28.0 (Warrn et al., 2010).

Fine mapping with imputation

For multiple sclerosis SNPs, regardless of evidence of replication in African Americans, we assessed whether there was a more strongly associated risk-tagging SNP in the region flanking the multiple sclerosis SNP by testing for association amongst the genotyped SNPs from these regions, including those assessable by imputation. The regions of multiple sclerosis SNPs were defined as the range of chromosomal positions where SNPs in linkage disequilibrium around the multiple sclerosis SNPs with \( r^2 > 0.5 \) locate in the European populations of the 1000 Genome Project. Imputation was performed using IMPUTE2 (v2.3.0) (Howie et al., 2009) and the 1000 Genomes Phase 1 integrated haplotypes (released in September 2013) were used as a reference panel. In addition to the previously conducted quality control measures, those AT/GC SNPs with failed alignment were removed. Missing genotypes for the genotyped SNPs were not imputed. To increase the
Replication study of unreported multiple sclerosis variants

For the candidate multiple sclerosis loci previously unreported in Europeans, a replication study was conducted on an independent African American group consisting of 620 multiple sclerosis cases and 1565 controls by genotyping the top SNPs after imputation in the region of interest. SNP genotyping was completed in the replication data set using predesigned and custom TaqMan® SNP Genotyping Assays. TaqMan® SNP genotyping assays were conducted in 384-well plates on an ABI 7900HT Sequence Detection System using SDS 2.3 software. Association P-values were provided with one-tailed test. We also performed meta-analysis under a fixed-effects model with effect sizes and standards errors from the African American ImmunoChip (discovery) data set and the replication study. Here a SNP was considered to have replicated when the replication $P < 0.05$ (one-tailed test) and the combined $P$-value of meta-analysis is more significant than the discovery $P$-value. For the replication study, no adjustment of population admixture was conducted.

Results

Replication study of the multiple sclerosis-associated SNPs in non-MHC-associated regions

We screened 130,135 autosomal SNPs in 803 African American multiple sclerosis cases and 1516 African American control subjects. Figure 2 shows a Circos plot summarizing the results from this screen; as anticipated, the strongest association was observed in the MHC region on chromosome 6p21.3 ($P = 2.75 \times 10^{-8}$). In Europeans 110 SNPs from 103 discrete loci outside the MHC region have been established as risk variants in multiple sclerosis (IMSGC, 2013b); in our African American screen, results passing stringent quality control were available for 96 of these SNPs (including rs3190930 a proxy SNP for rs802734, Supplementary Table 1). Amongst these 96 we found that >70% (69/96) had the same allele over-represented in cases as in European multiple sclerosis cases, a highly significant excess of concordance (one-tailed binomial test $P = 1.07 \times 10^{-5}$). For 21 of these 69 the excess frequency in cases was nominally significant (one-tailed test $P < 0.05$) (Table 1); for all of these the effect sizes in African Americans were statistically indistinguishable from those observed in Europeans (heterogeneity test $P > 0.05$, Supplementary Table 1). Even including unreplicated multiple sclerosis SNP, the obtained effect sizes of multiple sclerosis variants in African Americans were generally correlated with those in Europeans (Supplementary Fig. 2). To estimate the level of concordance that might be expected if effects were the same in African Americans as in Europeans, we estimated for each of the 96 SNPs the power of a study with 803 cases and 1516 control subjects...
to identify nominally significant association (one-tailed test \( P < 0.05 \) or half the power to observe two-tailed test \( P < 0.1 \)), assuming effect sizes equivalent to those seen in the European screen (IMSGC, 2013b) and the risk allele frequencies observed in our African American control population (Supplementary Table 1). Across the 96 variants we found that the average power was 18.6\%, with values ranging from 5.5\% (at rs2028597) to 49.9\% (at rs6677309). Based on this average value we would anticipate seeing nominally significant association (one-tailed test \( P < 0.05 \)) at between 12 and 24 SNPs with the same risk allele as in Europeans. Our observation of 21 such SNPs is thus entirely consistent with these variants exerting equivalent effect in African Americans and Europeans.
Table 1 Replicated 21 SNPs in African Americans out of 96 non-MHC multiple sclerosis susceptibility variants of Europeans

<table>
<thead>
<tr>
<th>Chr</th>
<th>rsID</th>
<th>Position</th>
<th>Gene</th>
<th>Function</th>
<th>RA Europeans</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>African Americans</th>
<th>OR (95% CI)</th>
<th>het. P</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>rs6677309</td>
<td>117080166</td>
<td>CD58</td>
<td>Intronic</td>
<td>A</td>
<td>0.879</td>
<td>1.34</td>
<td>1.27–1.41</td>
<td>1.5 x 10^{-28}</td>
<td>0.547</td>
<td>0.471</td>
</tr>
<tr>
<td>16</td>
<td>rs12927355</td>
<td>11194771</td>
<td>CLEC16A</td>
<td>Intronic</td>
<td>G</td>
<td>0.678</td>
<td>1.21</td>
<td>1.17–1.26</td>
<td>8.2 x 10^{-27}</td>
<td>0.787</td>
<td>0.747</td>
</tr>
<tr>
<td>19</td>
<td>rs11554159</td>
<td>18285944</td>
<td>IFI30</td>
<td>Exonic</td>
<td>G</td>
<td>0.730</td>
<td>1.15</td>
<td>1.11–1.20</td>
<td>2.6 x 10^{-13}</td>
<td>0.795</td>
<td>0.755</td>
</tr>
<tr>
<td>5</td>
<td>rs71624119</td>
<td>55440730</td>
<td>ANKRDS5</td>
<td>Intronic</td>
<td>G</td>
<td>0.755</td>
<td>1.12</td>
<td>1.08–1.17</td>
<td>2.7 x 10^{-09}</td>
<td>0.943</td>
<td>0.935</td>
</tr>
<tr>
<td>7</td>
<td>rs9171161</td>
<td>28172778</td>
<td>CD58</td>
<td>Intronic</td>
<td>C</td>
<td>0.205</td>
<td>1.15</td>
<td>1.11–1.20</td>
<td>6.8 x 10^{-15}</td>
<td>0.728</td>
<td>0.700</td>
</tr>
</tbody>
</table>

Position is based on human genome 19 and dbSNP 137.

\( ^{a} \)When SNPs locate intergenic, the closest genes are shown in brackets.

\( ^{b} \)Results of Europeans are originated from IMSGC, 2013.

\( ^{c} \)SNPs with association \( P \)-values < 0.05 (one-tailed test) were shown.

\( ^{d} \)Cochrane Heterogeneity Q test.

AA = African Americans; Chr = chromosome; CI = confidence interval; cont. = controls; het. = heterogeneity; RA = risk allele; RAF = risk allele frequency.
Unsurprisingly, the two SNPs with the most significant association in the African Americans were those with the strongest effects in Europeans, rs6677309 (CD58) and rs12927355 (CLEC16A).

Among the 21 replicated variants, two exonic multiple sclerosis SNPs in IFI30 (rs11554159) and TYK2 (rs34536443), respectively, are predicted as probably damaging. For rs34536443 in TYK2, the protective allele drives T lymphocyte differentiation towards a Th2 phenotype (Couturier et al., 2011). This variant was also found to be associated with juvenile idiopathic arthritis, primary biliary cirrhosis, psoriasis and rheumatoid arthritis (Eyre et al., 2012; Liu et al., 2012; Tsoi et al., 2012; Hinks et al., 2013). Other replicated SNPs included rs1800693 in TNFRSF1A, which functionality mimics the effect of TNF blocking drugs (Gregory et al., 2012), and rs2104286 in IL2RA, which seems to increase the ratio of soluble/membrane IL2RA and inhibits IL2 signalling (Maier et al., 2009). Both rs1800693 (TNFRSF1A) and rs2104286 (IL2RA) accounted for > 50% of posterior probability of association in Europeans (IMSGC, 2013b).

In a previous study using an overlapping, albeit larger sample set of African Americans, we reported significant associations for 8 of 74 tested non-MHC multiple sclerosis risk SNPs with a two-tailed test \( P < 0.01 \) threshold, whereas (coincidentally) 21 variants exceeded the one-tailed test \( P < 0.05 \) threshold (Isobe et al., 2013). When compared to this study, all of these variants had the same direction of association despite the limited statistical power of both studies (Supplementary Table 2). However, associations in this study did not reach statistical significance in seven loci due most likely to the relatively low effect sizes of these variants. Additionally, lower minor allele frequency of the updated multiple sclerosis SNPs compared to the previous SNPs prevented replication for two loci (MMEL1 and IRF8). On the other hand, the TYK2 locus was replicated, this time with a different risk-tagging SNP from our previous study (rs8112449, not replicated; Isobe et al., 2013).

Given the possibility of allelic heterogeneity across ancestral groups, we reasoned that SNPs lying close to the European lead SNP have increased prior odds even if not in linkage disequilibrium. To look for such effects we searched in the African American data set for nominally associated SNPs within the intervals flanking each of the 110 European SNPs (with boundaries of the flanking intervals defined by the most distant SNP in linkage disequilibrium with the European lead SNP; \( r^2 > 0.5 \) in the 1000 Genome data set of Europeans). We considered first the 21 intervals containing European lead SNPs that showed nominal evidence of association, and found more significantly associated SNPs in 20 (data not shown). Among them only eight were in linkage disequilibrium (\( r^2 > 0.5 \)) with the European lead SNP. In the remaining 89 regions (= 110 – 21), after correction for independent testing at each locus, and setting a FDR of 0.05, we found eight regions that contained SNPs showing nominally significant evidence for association (Table 2, Figs 2 and 3, and Supplementary Fig. 3). One of the variants (rs1861842) in the PVT1/IRIIR208 locus shows modest linkage disequilibrium (\( r^2 = 0.409 \)) with the corresponding lead European SNP (rs759648) in the European population but rather little linkage disequilibrium with that SNP in African Americans (\( r^2 = 0.142 \)), suggesting that these two SNPs (rs759648 and rs1861842) tag the same signal in Europeans while only rs1861842 is correlated with the signal in African Americans, consistent with this SNP being a better tag for the functionally relevant variant (Fig. 3A).

Taking advantage of the unique linkage disequilibrium patterns in the African American genome enabled us to possibly narrow two additional disease-association regions, MMEL1 at 1p36 and ZFP36L1 at 14q22-q24. In the MMEL1 locus, the linkage disequilibrium block in African Americans (\( r^2 > 0.5 \)) flanking rs111375644 (lowest \( P \)-value in African Americans) is 2494816–2728455 bp and is 3.5 kb smaller than linkage disequilibrium block in Europeans flanking rs3748817 (lowest \( P \)-value in Europeans), excluding LOC115110 from the candidate disease-associated genes (Fig. 3B). Furthermore, the narrow linkage disequilibrium region (\( r^2 > 0.8 \)) around rs3748817 spreads across 237 kb in Europeans and includes five genes, whereas the size of the high linkage disequilibrium region in African Americans for rs111375644 was 16 kb and includes a single gene (TNFRSF1F). In the ZFP36L1 locus, the linkage disequilibrium region in African Americans around the most significantly associated SNP rs8011424, was 25.6 kb smaller (69265911–69310210 bp) than that in Europeans flanking the established multiple sclerosis SNP (rs2236262), highlighting the upstream region of ZFP36L1 (Supplementary Fig. 3D). However, as these variants are monomorphic or show no significant linkage disequilibrium with their respective European lead SNP even in the European population, they may represent additional risk alleles rather than successful fine mapping of European signals. Lastly, in the IRF8 locus, the optimal SNP in African Americans after imputation (rs13333054) coincides with the previously reported SNP in the 2011 GWAS (IMSGC and WTCCC2, 2011) rather than the ImmunoChip (IMSGC, 2013b) (Supplementary Fig. 3E).

In this African American cohort 4 of 96 European lead SNPs showed nominally significant evidence of association with the alternate allele to that seen in Europeans (Supplementary Table 1), raising the possibility that these variants might be exerting different, even opposite effects in this population. However, this number of seemingly opposite effects is consistent with that expected to result from random sampling variation overwhelming genuine but modest signals. In a study of this size and considering 96 variants, we would anticipate seeing up to five apparently reversed effects by chance alone. A similar low frequency of apparently reversed signals was seen in our previous African American study (Isobe et al., 2013) and also in
Table 2: Alternate SNPs in replicating known susceptibility regions

(A) European multiple sclerosis-associated SNPs | (B) Top SNP in African Americans

<table>
<thead>
<tr>
<th>rsID (LD region)</th>
<th>Gene</th>
<th>RA</th>
<th>AfAm</th>
<th>OR</th>
<th>P&lt;sup&gt;b&lt;/sup&gt;</th>
<th>rsID&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Gene</th>
<th>RA</th>
<th>OR (95%CI)</th>
<th>raw p</th>
<th>FDR P (No. Phy)&lt;sup&gt;d&lt;/sup&gt;</th>
<th>LD info (between A and B)&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3748817 (1: 2484–2721)</td>
<td>MMEL1</td>
<td>A</td>
<td>1.05</td>
<td>2.4 x 10&lt;sup&gt;-01&lt;/sup&gt;</td>
<td>rs11375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
</tr>
<tr>
<td>rs3748817 (2330–2721)</td>
<td>MMEL1</td>
<td>A</td>
<td>1.05</td>
<td>2.4 x 10&lt;sup&gt;-01&lt;/sup&gt;</td>
<td>rs11375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
</tr>
<tr>
<td>rs111375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
<td>_&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs111375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
<td>_&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3748817 (1: 2484–2721)</td>
<td>MMEL1</td>
<td>A</td>
<td>1.05</td>
<td>2.4 x 10&lt;sup&gt;-01&lt;/sup&gt;</td>
<td>rs11375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
</tr>
<tr>
<td>rs3748817 (2330–2721)</td>
<td>MMEL1</td>
<td>A</td>
<td>1.05</td>
<td>2.4 x 10&lt;sup&gt;-01&lt;/sup&gt;</td>
<td>rs11375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
</tr>
<tr>
<td>rs111375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
<td>_&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs111375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
<td>_&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3748817 (1: 2484–2721)</td>
<td>MMEL1</td>
<td>A</td>
<td>1.05</td>
<td>2.4 x 10&lt;sup&gt;-01&lt;/sup&gt;</td>
<td>rs11375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
</tr>
<tr>
<td>rs3748817 (2330–2721)</td>
<td>MMEL1</td>
<td>A</td>
<td>1.05</td>
<td>2.4 x 10&lt;sup&gt;-01&lt;/sup&gt;</td>
<td>rs11375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
</tr>
<tr>
<td>rs111375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
<td>_&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs111375644</td>
<td>TNFRSF14</td>
<td>G</td>
<td>0.065</td>
<td>0.039</td>
<td>1.92 (1.43–2.58)</td>
<td>1.5 x 10&lt;sup&gt;-05&lt;/sup&gt;</td>
<td>9.7 x 10&lt;sup&gt;-03&lt;/sup&gt; (664)</td>
<td>_&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Regions are defined as range of positions where SNPs in linkage disequilibrium (r^2 > 0.5) with the multiple sclerosis-associated SNPs in Europeans of the 1000 Genome Project locate. Chromosome number and genomic positions (human genome 19, kb) are shown.

*bOne-tailed test.

*cImputed SNPs are shown in italics.

*dFalse discovery rate P-values are shown with the number of SNP P-values in linkage disequilibrium regions in brackets.

*eLinkage disequilibrium information in Europeans originated from European data set of the 1000 Genome Project and that in African Americans are from our African American data set.

*fTop SNP in linkage disequilibrium region (B) is monomorphic in Europeans.

*AfAm = African Americans; Chr = chromosome; CI = confidence interval; dist = distance; EUR = Europeans; LD = linkage disequilibrium; MAF = minor allele frequency; NA = not available; OR = odds ratio; RA = risk allele in Europeans; SE = standard error.
Figure 3  Narrowing in the causative region using African American data set. Comparative association plots for the loci of (A) PVT1/MIR1208 and (B) MMEL1 of (i) African Americans after fine mapping with imputation; and (ii) the discovery data set of European ImmunoChip. SNPs with the top association $P$-values are shown in purple with SNP IDs. For the plots of African Americans, genotyped SNPs are shown in closed circles and imputed ones in closed triangles. Colours of the marks represent linkage disequilibrium ($r^2$) with the top SNP in each population. Chromosomal positions are based on human genome 19. Note differences in the scales of y-axis.
Exploring potential association outside the established multiple sclerosis-associated loci

Recognizing the limited power of the data set, we nevertheless explored the evidence for association seen at SNPs mapping outside the designated 110 multiple sclerosis loci and outside the MHC region. In this analysis we identified seven regions containing at least one SNP with \( P < 10^{-4} \) (Supplementary Table 3). Only one of these (rs11123495) showed nominally significant association in the European ImmunoChip data set (\( P = 1.98 \times 10^{-5} \)) but the direction of the association was opposite from that in African Americans (IMSGC, 2013b). When analysing these seven regions in an independent replication cohort (620 African American cases with multiple sclerosis and 1565 control subjects), we found evidence of association for only one variant (rs2702180 in SMG7) (one-tailed test \( P = 0.034 \), Table 3). However, in a combined analysis across both African American data sets, this SNP failed to reach genome-wide significance (\( P = 6.3 \times 10^{-5} \)).

Discussion

The recent completion of the ImmunoChip project raised to 110 the number of non-MHC multiple sclerosis risk DNA variants in Europeans (IMSGC, 2013b). In aggregate, the proportion of the genetic variance accounting for disease risk explained by these polymorphisms, including the MHC, is roughly 27% (IMSGC, 2013b). Our main goal was to assess the transferability of this updated multiple sclerosis genetic map to African Americans. The number of replicated variants (21 of 96) was within the range of expectation given the power of our study, suggesting that most, if not all, of the multiple sclerosis risk SNPs discovered in Europeans are also relevant in African Americans and possibly in other non-white populations as well. An excess of concordant direction for allelic effects of the European multiple sclerosis SNPs in African Americans is consistent with this generalization.

For several of the established loci even though we failed to see evidence for significant association with the European lead SNP, we did find evidence of association with independent flanking variants. Most of these new variants were uncorrelated with the European lead SNP in both

---

### Table 3

<table>
<thead>
<tr>
<th>Chr</th>
<th>rsID</th>
<th>Position</th>
<th>Gene</th>
<th>RA</th>
<th>RAf</th>
<th>Number of significant SNPs</th>
<th>Allelic Effects</th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>121</td>
<td>183571791 SMG7</td>
<td>T</td>
<td>0.24</td>
<td>0.70</td>
<td>1.33 (1.41–1.60)</td>
<td>1.31 (1.10–1.53)</td>
<td>( P = 0.04 )</td>
</tr>
<tr>
<td>2</td>
<td>211</td>
<td>183570833 (SMG7)</td>
<td>A</td>
<td>0.32</td>
<td>0.26</td>
<td>1.31 (1.21–1.42)</td>
<td>1.23 (1.09–1.37)</td>
<td>( P = 0.02 )</td>
</tr>
<tr>
<td>8</td>
<td>78</td>
<td>619270612 (LOC2000502)</td>
<td>T</td>
<td>0.30</td>
<td>0.21</td>
<td>1.31 (1.23–1.42)</td>
<td>1.23 (1.10–1.37)</td>
<td>( P = 0.01 )</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>101559080 LOC100287015</td>
<td>G</td>
<td>0.28</td>
<td>0.24</td>
<td>1.31 (1.23–1.42)</td>
<td>1.23 (1.10–1.37)</td>
<td>( P = 0.01 )</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>101559080 LOC2000502</td>
<td>T</td>
<td>0.30</td>
<td>0.21</td>
<td>1.31 (1.23–1.42)</td>
<td>1.23 (1.10–1.37)</td>
<td>( P = 0.01 )</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>101559080 LOC100287015</td>
<td>G</td>
<td>0.28</td>
<td>0.24</td>
<td>1.31 (1.23–1.42)</td>
<td>1.23 (1.10–1.37)</td>
<td>( P = 0.01 )</td>
</tr>
</tbody>
</table>

---
the African American and European populations, suggesting that these must be different effects although the possibility of artefacts cannot be excluded. However, for one variant (rs1861842 adjacent to MIR1208) we did see linkage disequilibrium in the European population but not in the African American population, indicating that the original disease risk signal has been successfully fine mapped beyond what was possible within the European population. Altogether, our data confirm that the potential of less extensive linkage disequilibrium structure present in African Americans to aid in fine mapping is only likely to be advantageous if the study has adequate power to demonstrate significant genome-wide association.

The signal we identified in the region of MMEL1, locates telomeric to MMEL1, between FAM213B and TNFRSF14, whereas the European signal locates to the 18th intron of MMEL1 itself. This increases the evidence supporting a role for these other genes. FAM213B is associated with biosynthesis of prostaglandin F2α and cyclooxygenase pathway whereas TNFRSF14, a member of the TNF receptor superfamily fits well within the activatory/inhibitory inflammation pathways mechanistically associated with multiple sclerosis (IMSGC and WTCCC2, 2011). Mutations in TNFRSF14 have also been associated with B cell lymphoma (Morin et al., 2011; Lohr et al., 2012), consistent with an increasing appreciation that disordered B cell function is intimately associated with multiple sclerosis pathogenesis (Hauser and Goodin, 2012). Interestingly, a recent network-based pathway analysis also ranked TNFRSF14 higher compared to MMEL1 as a multiple sclerosis susceptibility locus (IMSGC, 2013a). Similarly, for EVI5 on chromosome 1, which is a well-established risk locus with relatively large effect size (odds ratio = 1.20) (IMSGC, 2013b), the reported top risk-tagging SNP in Europeans locates at the 3’ untranslated region (UTR) of EVI5, while in African Americans, the associated SNP (rs115126543) identified in this study locates 11 kb upstream of the gene itself within transcription factor binding sites and a DNase I hypersensitive site (Bernstein et al., 2012), suggesting a role for transcriptional regulatory mechanisms mediating risk. However, it is notable that risk allele frequencies for both these variants are low and therefore so is power. As neither is identified with clear significance additional studies will be required to confirm the relevance of these observations.

The screen identified seven novel regions containing at least one SNP with suggestive evidence of association, of which only one (rs2702180 in SMG7 on chromosome 1) replicated in an independent data set using relatively lenient but predetermined replication criteria. SMG7 encodes a protein that is essential for nonsense-mediated mRNA decay, a process linked to autoimmunity (Bachmann et al., 2006). Interestingly, the risk-tagging SNP was reported to be associated with the expression level of SMG7 in brain tissues (Gibbs et al., 2010) and in lymphoblastoid cells (Stranger et al., 2007). In SLE, NCF2 adjacent to SMG7 is reported to be associated with the disease (Gateva et al., 2009; Cunningham Graham et al., 2011; Jacob et al., 2012) but a multi-ethnic study pointed out that SLE-associated variants located in NCF2 were significantly associated with the expression of SMG7 (Kim-Howard et al., 2014). Additional studies will be required to validate the association in an independent African American data set with larger sample size and to determine if the association with this locus can also be observed in European populations. The SMG7 region locates outside the highest peak in a genome-wide admixture scan, which may partially explain why the multiple sclerosis association with SMG7/NCF2 was found in African Americans but not in Europeans (Reich et al., 2005).

In conclusion, we show the extensive replication of European multiple sclerosis variants in African Americans, consistent with a shared genetic architecture for multiple sclerosis susceptibility across these different populations. However, as the ImmunoChip design was mainly based on reference European populations (Cortes and Brown, 2011), the utility of the array to genotype non-European populations, potentially lacking tag SNPs for some haplotypes and the full range of cross-ancestral genetic pleiotropy, remains unknown. Our results suggest that ImmunoChip-like platforms have substantial potential to fine-map regions of interest by taking advantage of different haplotypic structures, but the need for very large sample sizes and functional studies is still evident. Even with arrays capable of tagging variation in both populations, very large sample sizes would be necessary to exclude any modest effect of a European variant in an African American population and vice versa. Trans-ancestral studies are also likely to help in the discovery of new genes and pathways vital to disease susceptibility (Diabetes Genetics Replication and Meta-analysis Consortium et al., 2014). In addition, the clinical expression of multiple sclerosis, including its severity, is known to have a genetic basis, but to date no disease modifiers have been convincingly identified. The severe clinical course and treatment-resistance typical of multiple sclerosis in African Americans highlights an additional opportunity, i.e. to identify modifiers of disease severity and progression that could lead to much-needed therapeutic opportunities.

Supplementary material
Supplementary material is available at Brain online.

Acknowledgements
The authors thank the multiple sclerosis patients and healthy controls who participated in this study. The authors acknowledge the contributions of H. Mousavi and R. Guerrero (UCSF) for sample processing and management, and the genotyping teams at the Wellcome Trust Sanger Institute, UK, the Cambridge NIHR Biomedical
Research Centre, UK and the Centre for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. This manuscript is dedicated to the memory of Joseph Herbert, in recognition of his contributions and leadership in multiple sclerosis research, and committed dedication to people afflicted with the disease.

**Funding**

This study is supported by grants from the National Institute of Health (R01NS076492, R01NS046297 and R01NS049477) and the UK Multiple Sclerosis Society (898/08). Recruitment of study participants and sample acquisition was supported by National Multiple Sclerosis Society (RG2899-D11 and RC2 GM093080). N.I. was supported by Postdoctoral Fellowship for Research Abroad from Japan Society for the Promotion of Science (JSPS) and is currently a JSPS Research Fellow. L.P. is a Harry Weaver Neuroscience Scholar of the National Multiple Sclerosis Society (JF2144A2/1).

**References**


