RESEARCH ARTICLE

Genetic susceptibility to multiple sclerosis in African Americans

Douglas S. Goodin1 *, Jorge R. Oksenberg1, Venceslas Douillard2, Pierre-Antoine Gourraud1,2, Nicolas Vince2

1 Department of Neurology, University of California, San Francisco, CA, United States of America, 2 Université de Nantes, CHU Nantes, Inserm, Centre de Recherche en Transplantation et Immunologie, UMR 1064, ITUN, Nantes, France

* douglas.goodin@ucsf.edu

Abstract

Objective
To explore the nature of genetic-susceptibility to multiple sclerosis (MS) in African-Americans.

Background
Recently, the number of genetic-associations with MS has exploded although the MS-associations of specific haplotypes within the major histocompatibility complex (MHC) have been known for decades. For example, the haplotypes HLA-DRB1*15:01~HLA-DQB1*06:02, and HLA-DRB1*03:01~HLA-DQB1*02:01 have odds ratios (ORs) for an MS-association orders of magnitude stronger than many of these newly-discovered associations. Nonetheless, all these haplotypes are part of much larger conserved extended haplotypes (CEHs), which span both the Class I and Class II MHC regions. African-Americans are at greater risk of developing MS compared to a native Africans but at lesser risk compared to Europeans. It is the purpose of this manuscript to explore the relationship between MS-susceptibility and the CEH make-up of our African-American cohort.

Design/methods
The African-American (AA) cohort consisted of 1,305 patients with MS and 1,155 controls, who self-identified as being African-American. For comparison, we used the 18,492 controls and 11,144 MS-cases from the predominantly European Wellcome Trust Case Control Consortium (WTCCC) and the 28,557 phased native Africans from the multinational “Be the Match” registry. The WTCCC and the African-Americans were phased at each of five HLA loci (HLA-A, HLA-C, HLA-B, HLA-DRB1 and HLA-DQB1) and the at 11 SNPs (10 of which were in non-coding regions) surrounding the Class II region of the DRB1 gene using previously-published probabilistic phasing algorithms.
Results
Of the 32 most frequent CEHs, 18 (56%) occurred either more frequently or exclusively in Africans whereas 9 (28%) occurred more frequently or exclusively in Europeans. The remaining 5 CEHs occurred in neither control group although, likely, these were African in origin. Eight of these CEHs carried the DRB1*15:03–DQB1*06:02–a36 haplotype and three carried the DRB1*15:01–DQB1*06:02–a1 haplotype. In African Americans, a single-copy of the European CEH (03:01_07:02_07:02_15:01_06:02_a1) was associated with considerable MS-risk (OR = 3.30; p = 0.0001)—similar to that observed in the WTCCC (OR = 3.25; p<10^{-168}). By contrast, the MS-risk for the European CEH (02:01_07:02_07:02_15:01_06:02_a1) was less (OR = 1.49; ns)—again, similar to the WTCCC (OR = 2.2; p<10^{-38}). Moreover, four African haplotypes were “protective” relative to a neutral reference, to three European CEHs, and also to the five other African CEHs.

Conclusions
The common CEHs in African Americans are divisible into those that are either African or European in origin, which are derived without modification from their source population. European CEHs, linked to MS-risk, in general, had similar impacts in African-Americans as they did in Europeans. By contrast, African CEHs had mixed MS-risks. For a few, the MS-risk exceeded that in a neutral-reference group whereas, for many others, these CEHs were “protective”—perhaps providing a partial rationale for the lower MS-risk in African-Americans compared to European-Americans.

Introduction
The pathogenesis of multiple sclerosis (MS) is complex and the susceptibility of an individual to developing this disease depends critically upon both environmental events and genetic factors [1–5]. Recently, a great deal of progress has been made with regard to our understanding of both aspects of MS pathogenesis. On the genetic side, for example, 233 loci, in diverse genomic regions, have now been identified as MS-associated by genome-wide association screens (GWAS), which use large arrays of single nucleotide polymorphisms (SNPs) scattered throughout the genome [4, 6–14]. Many of these implicated regions are located within or close to immune-related genes, which are involved in either the adaptive or innate arms of the human immune system [4]. Moreover, 32 of these independently MS-associated SNPs are located within the major histocompatibility complex (MHC) on the short arm of Chromosome 6 [4]. Despite this recent increase in the number and location of these MS-associations, however, certain human leukocyte antigens (HLA), located within the MHC, have long been known to be MS-associated [12, 15–22]. For example, among persons of European descent, the Class II HLA-DRB1*15 alleles have been known for decades to have a strong MS-association. In addition, the relationship of other HLA alleles (either to MS or to other diseases) has also been well-established [12, 15–23]. Typically, these studies have been focused on identifying the relationship of genetic susceptibility to specific alleles at specific HLA loci. For example, in individuals of European descent, the focus has been on the increased “risk” associated with carrying either HLA-DRB1*15:01 or HLA-DRB1*03:01 alleles and on the “protective” effect of carrying the HLA-A*02:01 allele [15–25]. For example, in the Wellcome Trust Case Control...
Consortium (WTCCC) dataset [13], the odds ratio (OR) of MS for individuals possessing one or more of these alleles is highly significant— for HLA-DRB1*15:01 (OR = 3.24; p < 10^{-300}); for HLA-DRB1*03:01 (OR = 1.27; p < 10^{-11}); and for HLA-A*02:01 (OR = 0.69; p < 10^{-54}).

Despite this focus on single alleles of specific genes, however, these HLA alleles don’t really exist in isolation. Indeed, it has been known for decades that multiple HLA alleles within both the Class I and II regions of the MHC influence, often interactively, the risk of developing MS [26]. For example, within the MHC, most HLA alleles are in tight linkage disequilibrium with each other and, overall, the HLA region consists of a relatively small collection of highly conserved extended haplotypes (CEHs), which stretch (at least) across the “classical” HLA genes (HLA-A, HLA-C, HLA-B, HLA-DRB1, and HLA-DQB1)—a distance spanning nearly 3 mb of DNA [27–31]. This haplotypic structure is found in all human populations, including Africans and persons of European descent [28]. Nevertheless, the CEH compositions, which account for this population structure, vary markedly between different regions [26–29]. Thus, in the predominantly European WTCCC, the most frequent 250 CEHs accounted for 57% of all CEHs present [29] and, in an African population [28], the most frequent 250 CEHs accounted for 31% of all CEHs present reflecting greater haplotypic diversity. Nevertheless, only 19 (4.0%) of these 500 “most-frequent” CEHs were shared as “most frequent” between the two populations. Thus, it seems that these CEHs are under a strong selection pressure, presumably based upon favorable biological properties of the complete haplotypes in certain environments [27–31].

In the HLA Class II region, this linkage disequilibrium is especially strong between (at least) between the HLA-DRB1 and HLA-DQB1 loci. For example, in the predominantly European data from the WTCCC, 97.5% of the HLA-DRB1*15:01 alleles (the most common DRB1 allele in Europeans; control frequency = 13.0%) are linked to the HLA-DQB1*06:02 allele. Similarly, 98.4% of the HLA-DRB1*03:01 alleles (control frequency = 11.8%) are linked to the HLA-DQB1*02:01 allele. The same is true in an African population [27]. Thus, in Africans, 98.3% of the HLA-DRB1*15:01 alleles (control frequency = 1.8%) are similarly linked to HLA-DQB1*06:02 allele and 99.5% of HLA-DRB1*03:01 alleles (control frequency = 7.3%) are also linked to the HLA-DQB1*02:01 allele [27]. Moreover, in Africans, 98.9% of the HLA-DRB1*15:03 alleles (the most common DRB1 allele in Africans; control frequency = 12.5%) are linked to the HLA-DQB1*06:02 allele. Similar tight linkages are found for most other DRB1–DQB1 combinations [29]. In addition, we have described a collection of SNP-haplotypes that are composed of unique combinations of the 11 SNPs (rs2395173; rs2395174; rs3129871; rs7192; rs3129890; rs9268832; rs532098; rs1753390; rs2187668; rs1063355; and rs9275141), and which span 0.25 mb of DNA surrounding the HLA-DRB1 locus [29]. Ten of these SNPs are within intergenic regions whereas rs1063355 is within exon 5 of the DQB1 gene. One such 11-sNP haplotype (a1), adds further specificity to the HLA-DRB1*15:01–HLA-DQB1*06:02 haplotype [29]. Thus, 99% of (a1) SNP-haplotypes carry the HLA-DRB1*15:01–HLA-DQB1*06:02 haplotype and, conversely, 99% of these HLA-haplotypes carry the (a1) SNP-haplotype [29]. This complete HLA Class II haplotype (DRB1*15:01–DQB1*06:02–a1) is referred to as the (H+) haplotype.

Regardless of such strong linkage disequilibrium in the Class II region, however, there are nuances to susceptibility that accrues because of the CEH structure. For example, in persons of European descent, the Class II HLA-DRB1*03:01–HLA-DQB1*02:01 haplotype comes in two forms. The first (present in 84% of the WTCCC controls) is coupled to the (a6) SNP-haplotype and the second (present in 15% of the WTCCC controls) is coupled to the (a2) SNP-haplotype [29]. Each form has a distinct relationship to susceptibility. For (a2) carriers, among non-(H +)-carrying individuals, a single copy is consistently associated with an increased MS-risk [29]. By contrast, for (a6) carriers, the risk associated with carrying a single copy varies from being associated with “risk” to being “protective” depending upon the Class I portion of the CEH.
being considered [29]. Similarly, all carriers of the (H+) haplotype have an increased MS-risk, although the degree of association varies depending upon the CEH involved [29]. By contrast, some HLA-DRB1*15:01~HLA-DQB1*06:02 haplotypes that don’t also carry the (a1) SNP-haplotype, seem not to be associated with any MS-risk [29]. And, finally, although the HLA-A’02:01 allele is “protective” when considered as a single allele, some of the CEHs on which this allele is present seem to have little impact on MS-risk whereas on other CEHs this allele seems to have a “protective” effect [29].

Given this strong linkage disequilibrium it is unclear what gene (or genes) within a “risk” haplotype is responsible for the increased susceptibility to MS that is observed. We have previously reported that, in an African American population, both the HLA-DRB1*15:01 and the HLA-DRB1*15:03 alleles (in the absence of the HLA-DQB1*06:02 allele) are associated with an increase in MS risk whereas the HLA-DQB1*06:02 allele (in the absence of the HLA-DRB1*15 alleles) is not [32]. A similar observation is noted in the WTCCC data where HLA-DRB1*15:01 in the absence of HLA-DQB1*06:02 is associated with MS (OR = 1.7; p = 0.0002) whereas HLA-DQB1*06:02 in the absence of HLA-DRB1*15:01 is not (OR = 1.2; ns). This asymmetry between loci the has been taken as evidence to suggest that MS susceptibility is related to something that lies telomeric to the DQB1 locus, possibly at the DRB1 locus itself [32]. Notably, however, the difference in OR between these two WTCCC observations is not significant (p = 0.11). In addition, another study utilized the fact that some African Americans lack the HLA-DRB5 gene (telomeric to DRB1) and demonstrated that MS-susceptibility was unchanged in individuals who were missing this gene [33]. This observation was interpreted as supporting the notion that MS susceptibility could be mapped to the DRB1 locus, although others have reported that DRB5, itself, may be related either to progression or susceptibility [34, 35]. Nevertheless, in this study, the authors also identified a single SNP (rs1035798), located in the region of the Class III AGER gene (telomeric to DRB5), which was independently associated with MS (in the absence of all carriers of DRB1*15 and DRB1*03 alleles were excluded from the analysis (OR = 1.85; p = 0.008). Similarly, the IMSGC reported 32 independent signals within the MHC [4].

It is unclear, however, given the haplotypic (CEH) structure of the MHC, whether these observations actually support any single gene (e.g., DRB1) as being responsible for the observed changes in MS-susceptibility. For example, using well-established MS epidemiologic parameters (e.g., the disease prevalence, the proportion of women among MS patients, the recurrence-risks for MS in siblings and twins of an MS proband, and the time-dependent changes in the sex-ratio) and based theoretical considerations, less than 7.3% of the general populations of North America and Europe have any chance, whatsoever, of getting MS [5]. Therefore, because 23% of controls in the WTCCC carry one or more copies of the DRB1*15:01~HLA-DBQ1*06:02~a1 or (H+) haplotype, this indicates that fewer than 32% (7.3/23) of these (H+)-carrying individuals have any chance at all of getting the disease—i.e., more than 68% of (H+)-carrying individuals have no chance of developing MS, regardless of their environmental experiences [5]. Moreover, as noted above, some carriers of the HLA-DRB1*15:01~HLA-DBQ1*06:02 haplotype, but not carriers of (a1), seem to have little, if any, MS-risk [29]. From these considerations, it seems clear that CEH composition within a population is a critical factor for MS pathogenies. It is the purpose of this manuscript, therefore, to explore the relationship between MS susceptibility and the CEH composition of our African American (AA) cohort.

Results

Among the 2,460 African American individuals in this study, there were 4,920 total CEHs present, of which 2,744 were unique and, of these, 679 had more than one representation in the
dataset. The 32 CEHs having at least 12 representations accounted for 16% of the total number of CEHs present (Tables 1 and 2) and, moreover, the 250 most frequently occurring CEHs accounted for 39% of the total. In addition, the likely source of these CEHs in the admixture (African or European) seemed, for the most part, clear because they remained unaltered in the AA cohort when compared either to their exclusive source population or to both populations. For example, of these 32 most frequent CEHs, 18 (56%) occurred either much more frequently or exclusively in an African compared to a European population (Fig 1) whereas 9 (28%) occurred much more frequently or exclusively in a European compared to a population (Fig 1) whereas 9 (28%) occurred much more frequently or exclusively in a European population. In all of these cases, the full haplotype was represented in the reference populations (Fig 1). The remaining 5 CEHs of Fig 1 were not found in either the African or the European control populations. Because (b5), (b16), and (b18) carry the predominantly (or exclusively) African HLA Class II motifs of DRB1*15:03~DQB1*06:02 or DRB1*13:01~DQB1*05:01, these CEHs seem likely to be of African origin (Fig 2). Because (b10) and (b14) carry the (apparently exclusive) European HLA Class II motifs of DRB1*07:01~DQB1*02:02 or DRB1*09:01~DQB1*02:02, these CEHs might seem likely to be of European origin (Fig 2).

Nevertheless, this is probably not the case. For example, it appears that the Class II haplotypes DRB1*07:01~DQB1*02:02 and DRB1*09:01~DQB1*02:02 in both the WTCCC and AA datasets are consistently identified as DRB1*07:01~DQB1*02:01 and DRB1*09:01~DQB1*02:01, respectively, in the data of Gragert and colleagues for both Africans and Europeans [28]. Thus, among the 2.4 million European Class II haplotypes in the “Be the Match” registry, there were no DRB1*07:01~DQB1*02:02 haplotypes whereas the DRB1*07:01~DQB1*02:01 haplotype accounted for 9.6% of all haplotypes present. By contrast, in the European WTCCC data here were no DRB1*07:01~DQB1*02:01 haplotypes whereas the DRB1*07:01~DQB1*02:02 haplotype accounted for 9.2% of all haplotypes present (Fig 2). Moreover, among the 28,557 Africans in the data of Gragert and colleagues [28], there were no DRB1*07:01~DQB1*02:02 haplotypes whereas the DRB1*07:01~DQB1*02:01 haplotype accounted for 9.6% of all haplotypes present. By contrast, in the AA dataset, there were no DRB1*07:01~DQB1*02:02 haplotypes whereas the DRB1*07:01~DQB1*02:02 haplotype accounted for 9.6% of all haplotypes present.

Similarly, among the Class II haplotypes in the “Be the Match” registry, among Europeans, there were no DRB1*09:01~DQB1*02:02 haplotypes whereas the DRB1*09:01~DQB1*02:01 haplotype accounted for 0.04% of all haplotypes present. By contrast, in the WTCCC data here were no DRB1*09:01~DQB1*02:01 haplotypes whereas the DRB1*09:01~DQB1*02:02 haplotype accounted for 0.01% of all haplotypes present (Fig 2). Moreover, in the “Be the Match” registry, among Africans, there were no DRB1*09:01~DQB1*02:02 haplotypes whereas the DRB1*09:01~DQB1*02:01 haplotype accounted for 2.7% of all haplotypes present. By contrast, in the AA dataset, there were no DRB1*09:01~DQB1*02:01 haplotypes whereas the DRB1*09:01~DQB1*02:02 haplotype accounted for 2.0% of all haplotypes present.

Although the same haplotype confusion seems to apply to other rare CEHs (i.e., not listed in Table 2), which carry the DQB1*02:02 allele in the AA dataset, these differences cannot simply be attributed to a general typing difference for the DQB1*02:01 and DQB1*02:02 alleles between these different sets of data. Thus, in each of these datasets, only the very common haplotype HLA-DRB1*03:01~DQB1*02:01 haplotype was represented. No dataset had any examples of a HLA-DRB1*03:01~DQB1*02:02 haplotype.

If, as suggested from above, these Class II haplotype pairs are, in fact, the same, then the (c7) CEH occurs in both groups although still more commonly in Europeans (p < 10^-8), the (b10) CEH still occurs in neither population, and the (b14) CEH occurs in only Africans (Fig 1). In addition, the DRB1*07:01~DQB1*02:01 haplotype is slightly, but not significantly, more common Africans (9.6%) than the DRB1*07:01~DQB1*02:02 haplotype is in Europeans.
(9.2%), whereas the \(DRB1^*09:01\)-\(DQB1^*02:02\) haplotype is significantly \((p < 10^{-12})\) more common in Africans compared to the \(DRB1^*09:01\)-\(DQB1^*02:02\) haplotype in Europeans (Fig 2).

Moreover, of the 364 AA individuals judged by admixture to be 99.999% African, 13 (3.6%) carried at least one of these 5 unknown CEHs and all of these full CEHs were carried by at least one of these "African" AA individuals. In addition, none of these 364 AA individuals carried any of the “European” CEHs listed in Fig 1. Conversely, of the 40 AA individuals judged by admixture to be 99.999% European, no one carried any of these 5 CEHs and, also, no one had any of the “African” CEHs listed in Fig 1. In addition, of the 1.24 million European individuals in the “Be the Match” registry [28], even considering the possible haplotype confusion (described above), 4 of these 5 CEHs were not carried by anyone and \((b14)\) was still significantly more common among Africans \((p < 10^{-12})\).

Taken together, this evidence suggests that each of these 5 CEHs are of African in origin and that, like the other frequent CEHs that we observed in this study, these CEHs have remained intact (unaltered) during the period of admixture. This breakdowns for CEH origin is also fully consistent with the average admixture (~73% African) that we observed in this cohort.

As noted above, the \(DRB1^*15:03\)-\(DQB1^*06:02\) haplotype is the most common haplotype among Africans. In our AA cohort, 87% of these HLA Class II haplotypes are linked to the \((a36)\) SNP-haplotype and 7.5% were linked to the \((a59)\) haplotype. Similarly, in the AA cohort, of all the \(DRB1^*15:01\)-\(DQB1^*06:02\) HLA Class II haplotypes present, 96% were linked to the \((a1)\) haplotype. In the WTCCC, 99% of the \(DRB1^*15:01\)-\(DQB1^*06:02\) haplotypes were linked to the \((a1)\) SNP-haplotype. However, in rare instances in the WTCCC, it was linked to other

### Table 1. CEHs in the AA population, which include either the \((H+)\) haplotype or the \(DRB1^*15:03\)-\(DQB1^*06:02\) Class II haplotype \(\dagger\).

| Name \(\dagger\) | HLA Haplotype | OR1 \(\dagger\) | Percentage | p-value \(\dagger\) |
|------------------|----------------|----------------|------------|-----------------|
| \(c_2\)          | 0302_07:02_07:02_15:01_06:02_a1 | 3.3 (1.7–6.7) | 0.6%       | 0.00005         |
| \(b_3\)          | 0302_04:01_44:03_15:03_06:02_a36 | 0.5 (0.2–1.1) | 0.3%       | 0.06            |
| \(c_3\)          | 0202_07:02_07:02_15:01_06:02_a1 | 1.5 (0.6–3.7) | 0.5%       | 0.33            |
| \(b_5\)          | 0602_07:18_58:01_15:03_06:02_a36 | 2.0 (0.8–5.9) | 0.3%       | 0.19            |
| \(b_6\)          | 0302_08:02_14:02_15:03_06:02_a36 | 1.9 (0.7–5.5) | 0.3%       | 0.19            |
| \(b_7\)          | 0602_04:01_53:01_15:03_06:02_a36 | 1.7 (0.6–5.2) | 0.3%       | 0.27            |
| \(b_9\)          | 0602_07:02_07:02_15:03_06:02_a36 | 2.0 (0.7–6.6) | 0.3%       | 0.17            |
| \(c_6\)          | 0202_07:02_07:02_15:01_06:02_a1 | 2.2 (0.7–8.1) | 0.4%       | 0.15            |
| \(b_{11}\)       | 0301_04:01_53:01_15:03_06:02_a36 | 0.7 (0.2–1.9) | 0.2%       | 0.47            |
| \(b_{12}\)       | 0301_17:01_42:01_15:03_06:02_a36 | 0.7 (0.2–2.2) | 0.4%       | 0.62            |
| \(b_{15}\)       | 0102_07:01_49:01_15:03_06:02_a59 | 1.3 (0.4–5.2) | 0.2%       | 0.78            |
| \(b_{18}\)       | 0301_02:10_15:03_03:06_02_a36 | 1.3 (0.4–5.2) | 0.2%       | 0.78            |
| \(b_{19}\)       | 0301_04:01_53:01_15:03_06:02_a36 | 1.3 (0.4–5.2) | 0.2%       | 0.78            |

\(\dagger\) CEHs either carrying the \(DRB1^*15:01\)-\(DQB1^*06:02\)-\(a1\) haplotype—i.e., the \((H+)\) haplotype—or carrying the \(DRB1^*15:03\)-\(DQB1^*06:02\) haplotype. The listed CEHs are all of those that have \(\geq 12\) representations in the AA cohort.

\(\dagger\) Arbitrary name for haplotype, sorted in descending order of frequency in the WTCCC [3, 29]—designated by \(c\)—and in the AA cohort for CEHs not found in the WTCCC—designated by \(b\).

* Odds ratio \((OR1)\) of disease for individuals having 1 copy of the listed CEH compared to a neutral reference group consisting of individuals having either no copies or no other copies of the \((H+)\) haplotype (see Methods). The 95% confidence interval \((CI)\) is in parenthesis. Percentage indicates the % of all CEHs in the AA Control population.

** The p-values for the \(OR1\) comparing cases to controls.

https://doi.org/10.1371/journal.pone.0254945.t001
Table 2. Other common CEHs in the AA population.

| HLA haplotype | A–C–B–DRB1–DQB1–SNP | ORI (CI) | Percentage | p-value |
|---------------|----------------------|----------|------------|---------|
| b1            | 3061_1:70:1_4201_0302_0402_a93 | 0.6 (0.4–1.0) | 2.2% | 0.06 |
| c1            | 0101_07:01_0701_0301_0201_a6 | 1.5 (0.9–2.5) | 1.3% | 0.12 |
| b2            | 6861_06:06_2:5802_1201_0501_a61 | 0.9 (0.4–1.8) | 1.0% | 0.75 |
| b4            | 3303_04:01_5301_0804_0301_a16 | 0.5 (0.2–1.1) | 0.9% | 0.05 |
| c6456         | 3601_04:01_5301_1101_0602_a79 | 0.5 (0.2–1.3) | 0.7% | 0.17 |
| c3707         | 6862_03:04_1510_0501_0201_a2 | 1.9 (0.7–5.5) | 0.3% | 0.19 |
| b8            | 3601_1701:4202_0804_0301_a49 | 0.8 (0.3–2.0) | 0.5% | 0.65 |
| c5            | 0201_05:01_4402_0401_0301_a3 | 0.6 (0.2–1.7) | 0.3% | 0.32 |
| b10           | 2301_02:10_1503_0701_0202_a3 | 1.2 (0.4–3.4) | 0.5% | 0.82 |
| c6651         | 7401_02:10_1503_1503_0609_a25 | 1.4 (0.4–4.8) | 0.3% | 0.61 |
| c6666         | 3303_14:02_1516_0102_0501_a23 | 0.3 (0.1–1.1) | 0.5% | 0.07 |
| c9472         | 6862_03:04_1510_0102_0501_a23 | 0.6 (0.2–2.0) | 0.4% | 0.44 |
| c14           | 0201_07:01_0801_0301_0201_a6 | 0.9 (0.2–3.5) | 0.3% | 1.00 |
| c4            | 0301_04:01_3501_0101_0501_a9 | 0.9 (0.2–3.5) | 0.3% | 1.00 |
| b13           | 6862_03:04_1510_0804_0301_a16 | 0.8 (0.2–2.8) | 0.3% | 0.78 |
| c7            | 2902_16:01_4403_0701_0202_a5 | 0.5 (0.1–1.7) | 0.2% | 0.25 |
| b14           | 3002_07:02_0702_0901_0202_a3 | 2.1 (0.6–9.4) | 0.4% | 0.27 |
| b16           | 3002_18:02_7503_1301_0501_a10 | 1.3 (0.4–5.2) | 0.2% | 0.78 |
| b17           | 6862_1701:4201_0302_0402_a93 | 0.9 (0.2–4.1) | 0.2% | 1.00 |

† CEHs not carrying either the (H+) haplotype (see Table 1) or the DRB1*15:03–DQB1*06:02 haplotype and also having ≥ 12 representations in the AA cohort.
†† Arbitrary name for haplotype, sorted in descending order of frequency in the WTCCC [13, 29]–designated by (c)–and in the AA cohort for CEHs not found in the WTCCC–designated by (b).
* Odds ratio (ORI) of disease for individuals having 1 copy of the listed CEH compared to a neutral reference group consisting of individuals having no copies of that particular CEH and also no copies of any (H+) carrying CEH (see Methods). The 95% confidence interval (CI) is in parenthesis. Percentage indicates the % of all CEHs in the AA control population.
** The p-values for the ORI comparing cases to controls.

https://doi.org/10.1371/journal.pone.0254945.t002

SNP-haplotypes [29]. The same was true for both of these HLA Class II motifs in the AA cohort. Many of these rare alternative linkages in both populations were shared across these two HLA haplotypes including (a1), (a34), (a36), (a43), and (a71). In the predominantly European WTCCC, (a59) was not linked to (H+) but rather to DRB1*01:02–DQB1*05:01 [29]. In addition, in the WTCCC controls [29], 99% of the DRB1*03:01–DQB1*02:01 Class II haplotypes are linked either to the (a6) SNP-haplotype (84%) or to the (a2) SNP-haplotype (15%).

By contrast, in the AA controls only 83% of these haplotypes are linked to one or the other of these two SNP-haplotypes and they are distributed quite differently– 53% linked to (a2) and 30% linked to (a6)–the opposite of the distribution in Europeans. The remaining 17% of these DRB1*03:01–DQB1*02:01 haplotypes are, thus, linked to other, non-(a2) and non-(a6), SNP-haplotypes. Again, of these most frequent CEHs with these linkages, (c3707)–linked to (a2)–is African in origin whereas (c1) and (c14)–linked to (a6)–are European (Fig 1).

The ORs for the most frequent CEHs in our AA cohort are presented in Tables 1 & 2. Only 11 of these CEHs–(b1), (c1), (c2), (b2), (b3), (b4), (c6456), (c3), (b7), (c3707), and (b6)–had more than 20 representations available in the AA cohort and, only 3 of these CEHs–(b1), (c1), and (c2)–had more than 40 representations available. Also, among these 11 most frequent CEHs, (b3), (b7), and (b6) carried the DRB1*15:03–DQB1*06:02–a36 haplotype and (c2) and (c3) carried the (H+) haplotype. These are the haplotypes considered in our primary analysis.
Fig 1. CEH frequency in African, European, and African American control populations. African control frequencies are from Gragert, et al. [26]. European control frequencies are from the WTCCC [13, 14], and AA control frequencies are from the AA dataset reported here. African CEHs are highlighted in orange whereas European CEHs are highlighted in yellow (see Text). Only CEHs with ≥12 representations in the AA cohort are listed. The "No SNP haplotypes" condition is for CEHs not including any associated SNP-haplotype (Tables 1 & 2). The "With SNP haplotypes" condition is for CEHs that include the associated SNP-haplotype as indicated in Tables 1 & 2. Name indicates the haplotype (Tables 1 & 2), sorted in descending order of frequency in the WTCCC [13, 29]–designated by (c)–and in the AA cohort for CEHs not found in the WTCCC–designated by (b). In the Table, (0) indicates true zero. Ratios are of CEH frequency in Europeans to that in Africans (No SNPs condition) and the frequency ratio in Europeans to that in African Americans (With SNPs condition). nd = not defined; ∞ = infinity. Significance of the difference in CEH frequency between Africans and Europeans are indicated as follows:

- \(10^{-4} \leq p < 0.05\)
- \(10^{-8} \leq p < 10^{-4}\)
- \(10^{-12} \leq p < 10^{-8}\)
- \(p < 10^{-12}\)

It appears that the Class II haplotypes DRB1*07:01_DQB1*02:02 and DRB1*09:01_DQB1*02:02 in the WTCCC data are consistently identified in both African and European populations as DRB1*07:01_DQB1*02:01 and DRB1*09:01_DQB1*02:01, respectively, in the data of Gragert and colleagues [28]–see Text. If these Class II haplotypes are, in fact, the same, there is still a significantly greater (c7) CEH frequency in Europeans compared to Africans (\(p<10^{-8}\)). However, in this case, the (b14) CEH occurs only in Africans, whereas the (b10) CEH still occurs in neither Africans nor Europeans.

https://doi.org/10.1371/journal.pone.0254945.g001
The only unequivocally significant association in the AA cohort (compared to neutral reference—see Methods), among individuals who didn’t carry any (H+) haplotypes, was for the possession of a single copy of the (c2) CEH (OR = 3.30; \( p < 0.0001 \)). There was only one individual in the AA cohort who possessed two copies of the (c2) CEH so the association for the

| Haplotype          | African Frequency | European Frequency | Ratio |
|--------------------|-------------------|--------------------|-------|
| 01:01_05:01 **     | 0.018             | 0.103              | 5.8   |
| 03:01_02:01 **     | 0.072             | 0.116              | 1.6   |
| 03:02_04:02 **     | 0.059             | 0                | 0     |
| 04:01_03:01 **     | 0.008             | 0.054              | 6.4   |
| 07:01_02:02 **     | 0                 | 0.092              | \( \infty \) |
| 08:04_03:01 **     | 0.054             | 0.0001             | 0.002 |
| 09:01_02:02 *      | 0                 | 0.0001             | \( \infty \) |
| 11:01_06:02 **     | 0.051             | 0.0001             | 0.003 |
| 12:01_05:01 **     | 0.030             | 0.0001             | 0.002 |
| 13:01_05:01 **     | 0.011             | 0.0002             | 0.014 |
| 13:02_06:09 **     | 0.035             | 0.009              | 0.24  |
| 15:01_06:02 **     | 0.018             | 0.126              | 7.1   |
| 15:03_06:02 **     | 0.124             | 0.0001             | 0     |

Fig 2. DRB1–DQB1 Class II haplotype frequencies in Africans and Europeans control populations. African control frequencies are taken from Gragert, et al. [28] whereas European control frequencies are taken from the WTCCC [13, 29]. Predominantly African haplotypes are highlighted in orange whereas predominantly European haplotypes are highlighted in yellow (see Text). Haplotype of uncertain origin, either because of a similar frequency in both groups (03:01_02:01) or because of ambiguities (see below) are highlighted in green. In the Table, (0) indicates true zero. Only Class II haplotypes for the CEHs with \( \geq 12 \) representations in the AA cohort (see Fig 3) are listed. Ratios are of the haplotype frequency in Europeans controls (from the WTCCC) to that in Africans. \( \infty \) = infinity. Significance of the difference in Class II haplotype frequency between Africans and Europeans are indicated as follows:

- ** \( 10^{-4} \leq p < 0.05 \)
- **\* \( 10^{-8} \leq p < 10^{-4} \)
- † \( 10^{-12} \leq p < 10^{-8} \)
- †† \( p < 10^{-12} \)

As noted in the legend of Fig 1, it appears that the Class II haplotypes DRB1*07:01_DQB1*02:02 and DRB1*09:01_DQB1*02:02 in the WTCCC data are consistently identified in both African and European populations as DRB1*07:01_DQB1*02:01 and DRB1*09:01_DQB1*02:01, respectively, in the data of Gragert and colleagues [28]—see Text. If these Class II haplotypes are, in fact, the same, there is no significant difference in the frequency of the Class II haplotype DRB1*07:01_DQB1*02:01 in Africans (9.6%) and the frequency of DRB1*07:01_DQB1*02:02 in Europeans (9.2%). By contrast, in this case, the Class II haplotype DRB1*09:01_DQB1*02:01 is significantly more common in Africans (2.7%) than the haplotype DRB1*07:01_DQB1*02:02 in Europeans (0.01%)—\( p < 10^{-12} \).

https://doi.org/10.1371/journal.pone.0254945.g002

(Fig 1). The only unequivocally significant association in the AA cohort (compared to neutral reference—see Methods), among individuals who didn’t carry any (H+) haplotypes, was for the possession of a single copy of the (c2) CEH (OR = 3.30; \( p < 0.0001 \)). There was only one individual in the AA cohort who possessed two copies of the (c2) CEH so the association for the
Table 3. Comparisons between different CEHs carrying the Class II motif of either DRB1*15:03–06:02–a36 or DRB1*15:03–06:02–a36 *.

|                | OR1  | SE  | OR1  | SE  |
|----------------|------|-----|------|-----|
| 1503–0602–a1   |      |     | 1.49 | 0.41|
| c3             | 3.30 | 0.32| 1.49 | 0.41|
| 1503–0602–a36  |      |     | 1.6 (0.5–4.9) | 0.7 (0.2–2.5) |
| b3             | 0.50 | 0.36| 6.5 (2.6–16.8) * | 3.0 (1.0–8.6) * |
| b6             | 1.88 | 0.46| 1.8 (0.6–5.3) | 0.8 (0.2–2.7) |
| b7             | 2.01 | 0.46| 1.6 (0.5–4.9) | 0.7 (0.2–2.5) |

* Comparisons of the odds ratio (OR1) for the different CEHs listed in the Table. The numbers in in parentheses represent the 95% confidence intervals, at the point of intersection, the OR1 in the column to that in the row. Only CEHs having more than 20 representations in the AA cohort are compared.
that have more than 15 representations available. Despite the lack of statistical power, however, it seems clear from Fig 3 that, in general, \( ORs \) for the \( (b1) \) \( (b3) \), \( (b4) \), and \( (c6456) \) CEHs are notably smaller than the \( ORs \) for the \( (c1) \) \( (c2) \) \( (c3) \), \( (b7) \), \( (c3707) \) \( (b6) \) \( (b5) \), \( (b8) \), \( (b9) \) \( (b10) \), \( (c6) \) \( (c11) \) \( (c5) \) CEHs. We previously undertook a more fine-grained analysis of such relationships as these in the predominantly European WTCCC data [29]. However, in that study, we were able to consider only CEHs that had at least 50 representations (and many with hundreds) in the dataset—a circumstance that gave us a statistical power, which was not possible in a cohort of this size.

**Discussion**

The present study provides considerable insight both to the haplotypic composition of an African American population and to the relationship that this composition has to MS-susceptibility. Indeed, of the 32 most frequent (independently phased) CEHs in our AA cohort, 27 (84%) of them were identical to CEHs (also independently phased) in African and/or European populations [28, 29]. Moreover, of these 18 were clearly of African origin, 9 were clearly of European origin. The remaining 5 CEHs—\( (b5) \) \( (b10) \) \( (b14) \), \( (b16) \), and \( (b18) \)—were probably also
African in origin (see Results)—a circumstance that fits well with the average admixture (73%) observed in our cohort. The high frequency of these particular CEHs in our AA cohort (Tables 1 and 2), therefore, seems likely to be due to their high frequency in certain sub-populations of Africa, which were not well-represented in the African controls of Gragert and colleagues [28]. As a result, all 32 of these most-frequent CEHs seem to have remained remarkably intact over the period of time (<600 years), during which the admixture of our AA cohort was taking place. This observation underscores the stability of this CEH composition over relatively short time-intervals. By contrast, the considerable variability of CEH composition between African, European, and other populations [27–31] indicates a the CEH composition of different populations must be remarkably divergent over much longer periods of time. Presumably, such divergence is due to specific environmental and/or biological pressures that vary with time, with geographic location, or with both [26–31].

Among individuals who either don’t carry any (H+) haplotypes, or don’t carry any other (H+) haplotypes, we noted that some CEHs seem to be “protective” (e.g., b1, b3, c5), whereas others seem to carry “risk” (e.g., c2, c3, b6, b7). However, this distinction is simply a matter of definition. As a purely hypothetical example, we can arbitrarily designate one of two different haplotypes in some genomic region as “A” haplotypes and the other as “B” haplotypes. In this circumstance, any “protective” effect in individuals carrying “A” haplotypes compared to a reference group of individuals carrying “B” haplotypes is equivalent to a “risk” effect in individuals carrying “B” haplotypes compared to a reference group of individuals carrying “A” haplotypes. Thus, any notion of “risk” or “protective” haplotypes depends completely upon risk ratio between each haplotype being considered and the reference group chosen [31, 36–39]. By contrast, when two ORs are directly compared to each other as an estimate of the relative risk ratio, any chosen reference group becomes irrelevant [31, 36–39]. This point is critical when assessing in MS-susceptibility because, as noted earlier, more than 92.7% of individuals have no risk of MS whatsoever [5] and, using this group as the reference, even the group of individuals who don’t carry the (H+) haplotype will have an infinite relative risk. Consequently, it is the relative risk ratios (Fig 1) that provide the most reliable information regarding susceptibility.

Understanding this and even using this small dataset, it is clear that CEH composition has an important impact on MS-susceptibility in an African American population, much as it does in the predominantly European WTCCC [29]. Thus, the strongest statistical association with MS in both populations was for the (c2) CEH, which carries an (H+) Class II haplotype in addition to its Class I haplotype (Table 1; Fig 1). Moreover, the degree of risk associated with this predominantly European CEH (when compared to a similar reference group) was the same in each population (OR ≈ 3.3). Despite this, the (H+) haplotype, overall, appeared to be associated with less “risk” in African Americans (p = 0.01). However, this observation may be an artifact of combining, into a single group, different CEHs, many of which are known to carry different risks and which have different relative frequencies in the two populations [29]. For example, in the WTCCC cohort [29], the odds of disease for the c3 CEH was (OR = 2.2; p<10^-38), which was significantly smaller (p<10^-9) than that observed for the c2 CEH (OR = 3.25; p<10^-168). Similarly, the AA cohort, the odds of disease for the c3 CEH (OR = 1.5; ns) was smaller than that for the c2 CEH (OR = 3.3; p<10^-4).

Also, the apparent risk difference between DRB1*15:03–DQB1*06:02–a36 and (H+) is likely explained, at least partly, in a similar manner. For example, the (b3) CEH is carries significantly less risk than (b7) and (b6) and, possibly, the (b11) and (b12) CEHs as well (Table 3; Fig 3). Therefore, combining all of these CEHs into a single group will lead to an intermediate assessment of risk (which it did). In fact, because the relationship between a specific CEH and MS...
depends upon the nature of the entire haplotype (Table 3), the relationship between the HLA Class II portion of a CEH and MS, will, necessarily, be heterogeneous [29].

Other investigators have also explored the differential MS susceptibility in Africans and Europeans. For example, in a cohort of African Americans, Chi and coworkers [40] reported that the MS-risk OR for HLA\(^{\ast}\)DRB1\(^{15:01}\) allele of European origin was three times that for the same allele of African origin. In addition, these authors found that there were differences between these alleles in the amino acid composition, especially in the region of exon 1, but also in the regions of exons 3 and 5 [40]. Because exon 2 codes for the extracellular loop of the DRB1 protein, which contains the antigen recognition site (ARS), there were no differences found in this exon between African and European versions of this protein [40]. The authors raised the possibility that these differences could have functional consequences for the DRB1 molecule, despite Europeans and Africans sharing the same ARS [40]. For example, potentially, alterations in the non-ARS regions of the protein might impact the transcription, the translation, or the expression of DRB1 gene even if these changes didn’t impact the binding and recognition of antigen by the mature protein. This is an intriguing possibility although it should be noted that, even among Europeans, there are differences in risk between different DRB1\(^{15:01}\) alleles. For example, in the WTCCC, individuals who carry the (H+) haplotype (OR = 3.0) have almost twice the MS-risk (p < 10\(^{-6}\)) compared to individuals who carry other DRB1\(^{15:01}\) containing haplotypes (OR = 1.6). Also, as discussed in the Introduction, the same allele resides on many different CEHs and often these CEHs have very different disease associations, even among persons of very similar ancestry [29]. And, finally, because so few (H+) carriers are even susceptible to (i.e., have any chance of) getting MS, it is unclear how any single variant of the DRB1\(^{15:01}\) allele could possibly be responsible for the relationship between DRB1\(^{15:01}\) and susceptibility to MS [5]. This is especially true for the circumstance in which 94% of European DRB1\(^{15:01}\) alleles are identical [40].

In summary, the haplotypic (CEH) structure of our AA cohort is quite similar to the structure of other world populations [28, 29]. The CEH composition of our AA cohort appears to be an admixture of common CEHs of either African or European origin, which seem not to have been modified during the period of admixture. Moreover, those CEHs, which are likely of European origin (Fig 1), and which are associated with MS-risk in the predominantly European WTCCC cohort [29]—i.e., (c1), (c2), (c3), (c5), and (c6)—generally seemed to have a similar impact in our AA cohort (Tables 1 & 2). Of the common African CEHs, which carried the DRB1\(^{15:03}\)-DQB1\(^{06:02}\)-a36 haplotype, many seemed to have an MS-risk, which exceeded that in a reference group of non-(H+)-carrying individuals. However, even with this haplotype, the actual risk (i.e., whether it was “protective” or carried “risk”) depended upon the specific CEH being considered (Table 2). By contrast, most other common CEHs of likely African origin (Tables 2& 3) seemed to be “protective” relative to this same reference group—a circumstance that might help to rationalize, at least partly, the lower risk of MS in African compared to European Americans. Nevertheless, even though the risk of MS may be less in African Americans, the disease may be more severe and the disability greater compared to European Americans [Cree 2004].

Methods

Ethics statement

This research has been approved by the University of California, San Francisco’s Institutional Review Board (IRB) and has been conducted according to the principles expressed in the Declaration of Helsinki.
Study participants

The study population consisted of 1,305 patients with MS and 1,155 controls, all of whom self-identified as being African American (AA). The diagnosis of MS in this cohort was made based upon internationally recognized criteria [41–43]. The UCSF Institutional Review Board approved the protocol and written informed consent was obtained from each study participant.

For comparison purposes, we used the data from the WTCCC. The patients enrolled in this multinational cohort study were predominantly of European ancestry [13]. This cohort consists of 18,492 controls and 11,144 cases with MS and has been described in detail previously [13, 29]. The WTCCC granted data access for this study.

Also, for comparison, we analyzed the 28,557 native Africans and 1.24 million Europeans from the multinational data-set of Gragert et al. [28]. This study calculated six-locus high resolution HLA-A~C~B~DRB3/4/5~DRB1~DQB1 haplotype frequencies using the “Be the Match” registry donors who volunteered to be typed by DNA methods at recruitment. Mixed resolution HLA typing data was inputted using a modified expectation–maximization (EM) algorithm in the form of genotype lists generated by interpretation of primary genomic typing data to the IMGT/HLA v3.4.0 allele list [28]. The full cohort consisted of 6.59 million subjects categorized at a broad level by race. In sum, 25.8% of the individuals were typed at the C locus, 5.2% typed at the DQB1 locus, and all individuals were typed for the A, B, & DRB1 loci. The purpose of this study was to improve match predictions regarding donor selection for hematopoietic stem cell transplantation.

Genotyping, and quality control

The genotyping methods and quality control for the AA cohort has been described in detail previously [44]. Briefly, DNA was extracted from whole blood and SNP genotyping was conducted using the MS Chip, which is a custom genotyping array of Illumina Infinium. This array includes content designed to contain ancestry informative markers and other genetic markers specific interest for multiple sclerosis. Genotyping was done by the Center for Genome Technology (part of the John P. Hussman Institute for Human Genomics; University of Miami) and genotype calling was made using GenomeStudio v2.0. The identities of the five HLA alleles in the MHC region (A, C, B, DRB1 and DQB1) were determined for each participant by imputation using the HIBAG method [45]. We built a custom reference panel using CAAPA data (dbGaP Study Accession: phs001123.v1.p1) to impute HLA alleles from African American ancestry as accurately as possible. We used best guess HLA alleles. The posterior probabilities cutoff was 0.5, as recommended by the original HIBAG authors [45]. The percentage of alleles with posterior probabilities (> 0.5) was: HLA-A: 98%; HLA-B: 82%; HLA-C: 95%; HLA-DRB1: 85%; HLA-DQB1: 98%.

The genotyping and quality control methods both for the WTCCC and for the study of Gragert et al. [28] have also been described in detail previously [13, 14, 16, 18, 19, 28].

Estimating admixture

The ancestry of individuals in our AA cohort was inferred using ADMIXTURE software [46]. On chromosome 6, we selected SNPs (n = 2504), which overlapped between the AA individuals and two subsets of 1000 Genomes project (CEU, n = 99; YRI, n = 108), and which were representative of the European and African populations [47, 48].
Data access

Due to limitations in the original signed consents and to IRB restrictions regarding patient confidentiality, we are unable to provide individual genotype data for our African American cohort. For further inquiries or information, individuals may contact the IRB Chair at UCSF (Victor I. Reus, MD) at https://irb.ucsf.edu/. Nevertheless, summary statistics for the MHC SNPs are available upon request from the authors of the original paper [32]. For access to the WTCCC data and the “Be the Match” registry data (which are not ours), the original authors should be contacted directly [13, 28]. Our group obtained the data as outlined above although, because the lead author and principal investigator of the original AA publication [32] was also our co-author (JRO), we had access to the individual AA genotype data.

Fig 4. A principal component analysis of the pairwise “identity by decent” distances between cases and controls demonstrated no difference between cases and controls.

https://doi.org/10.1371/journal.pone.0254945.g004
Statistical methods

Phasing. Both the phasing of alleles at each of five HLA loci (HLA-A, HLA-C, HLA-B, HLA-DRB1 and HLA-DQB1) and the phasing of the SNP-haplotypes surrounding the Class II region of the DRB1 gene were accomplished using previously-published probabilistic phasing algorithms [29, 49–51].

Haplotype frequencies and association testing. Disease association tests, as measured by ORs and confidence intervals (CIs) comparing cases to controls, were calculated for each of the CEHs. These ORs were determined relative to a so-called “neutral reference group”. For CEHs that did carry the (H+) motif, this reference group excluded all (H+) carriers. For CEHs that carried the (H+) motif, this reference group excluded all individuals who carries another copy of (H+). The AA data was considered in its entirety and not further stratified. The significance of the differences in ORs for disease association (comparing cases to controls) for any two haplotypes or genotypes was determined by z-scores calculated from the differences in the natural logarithm of the ORs such that:

\[
z = \frac{\ln(OR_1) - \ln(OR_2)}{\sqrt{(SE[\ln(OR_1)])^2 + (SE[\ln(OR_2)])^2}}
\]

Benjamini-Hochberg method was used to correct for multiple testing of possible MS-association for the different CEHs. To maximize the statistical power to detect differences between CEHs, our primary analysis was on only those African American CEHs, which carried either the (H+) haplotype or the related DRB1*15:03–DQB1*06:02–a36 haplotype, and which had more than 20 representations in our AA cohort. Other comparisons were included only to provide exploratory point-estimates. All ORs used for pair-wise comparisons within the MHC were estimated relative to a reference group that excluded individuals who either carried any (H+) haplotypes or carried any other (H+) haplotypes. Within the AA cohort, to assess population stratification, we performed a principal components (PC) analysis, which excluded MHC SNPs (Eigensoft) and used regression analysis to correct the observations in Tables 1 & 2 for the possible effects of either population stratification or admixture within the AA cohort. In this analysis we used the first 10 of these PC components which accounted for 71% of the variance. Neither of these adjustments significantly altered any of our observations. Also, a PC analysis of the pairwise “identity by decent” distances demonstrated no differences between cases and controls (Fig 4).

Author Contributions

Conceptualization: Douglas S. Goodin.

Data curation: Jorge R. Oksenberg, Nicolas Vince.

Formal analysis: Douglas S. Goodin, Venceslas Douillard.

Methodology: Douglas S. Goodin, Nicolas Vince.

Visualization: Douglas S. Goodin.

Writing – original draft: Douglas S. Goodin.

Writing – review & editing: Douglas S. Goodin, Jorge R. Oksenberg, Pierre-Antoine Gourraud, Nicolas Vince.
References

1. Gourraud PA, Harbo HF, Hauser SL, Baranzini SE. The genetics of multiple sclerosis: an up-to-date review. *Immunol Rev* 2012; 248:87–103. https://doi.org/10.1111/j.1600-065X.2012.01134.x PMID: 22725956

2. Hofker MH, Fu J, Wijmenga C. The genome revolution and its role in understanding complex diseases. *Biochim Biophys Acta* 2014; 1842:1889–1895. https://doi.org/10.1016/j.bbdis.2014.05.002 PMID: 24834846

3. Goodin DS. The Causal Cascade to Multiple Sclerosis: A model for MS pathogenesis. *PLoS One* 2009, 4(2):e4565. https://doi.org/10.1371/journal.pone.0004565 PMID: 19242548

4. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* 2019; 65: eaav7188. https://doi.org/10.1126/science.aav7188 PMID: 31604244

5. Gourraud DS, Khankhanian P, Gourraud PA, Vince N. The Nature of Genetic and Environmental Susceptibility to Multiple Sclerosis. *PLoS One* 2021; 16(3):e0246157. https://doi.org/10.1371/journal.pone.0246157 PMID: 33750973

6. GAMES, the Transatlantic Multiple Sclerosis Genetics Cooperative. A meta-analysis of whole genome linkage screens in multiple sclerosis. *J Neuroimmunol* 2003; 143:39–46. https://doi.org/10.1016/j.jneuroim.2003.06.009 PMID: 14575912

7. de Bakker PIW, Yelensky R, Pe’er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet* 2005; 37:1217–1223. https://doi.org/10.1038/ng1669 PMID: 16244653

8. Herrera BM, Cader MZ, Dyment DA, Bell JT, Deluca GC, Miller CJ, et al. Multiple sclerosis susceptibility and the X chromosome. *Mult Scler* 2007; 13:856–8. https://doi.org/10.1177/1350493X06007921 PMID: 17881398

9. Wellcome Trust Case Control Consortium & Australo-Anglo-American Spondylitis Consortium. Associations can of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nature Genet* 2007; 39:1329–1337. https://doi.org/10.1038/ng.2007.17 PMID: 17982073

10. Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, et al. Genome-wide association studies of multiple sclerosis. *Hum Mol Genet*. 2009; 18:767–778. https://doi.org/10.1093/hmg/ddn988 PMID: 19010793

11. De Jager PL, Jia X, Wang J, de Bakker PIW, Ottoboni L, Aggarwal NT, et al, and the International MS Genetics Consortium. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat Genet* 2009; 41:776–782. https://doi.org/10.1038/ng.401 PMID: 19525953

12. Sanna S, Pitzalis M, Zoledziowska M, Zara I, Sidore C, Murr R, et al. Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. *Nat Genet*. 2010; 42:495–497. https://doi.org/10.1038/ng.584 PMID: 20453840

13. International Multiple Sclerosis Genetics Consortium & Wellcome Trust Case Control Consortium. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011; 476:214–219. https://doi.org/10.1038/nature10251 PMID: 21833088

14. International Multiple Sclerosis Genetics Consortium. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis *Nat Genet* 2014; 46:1535–60.

15. Dyment DA, Herrera BM, Cader Z, Willer CJ, Lincoln MR, Sadovnick AD, et al. Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. *Hum Mol Genet* 2005; 14:2019–2026. https://doi.org/10.1093/hmg/ddv206 PMID: 15930013

16. Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, de Jager PL, et al. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med*. 2007; 357:851–862. https://doi.org/10.1056/NEJMoa073493 PMID: 17660530

17. Ramagopalan SV, Anderson C, Sadovnick AD, Ebers GC. Genomewide study of multiple sclerosis. *N Engl J Med*. 2007; 357:2199–2200. https://doi.org/10.1056/NEJMoa078363 PMID: 18032773

18. Link J, Kockum I, Lorentzen AR, Lie BA, Cellius EG, Westerlind H, et al. Importance of Human Leukocyte Antigen (HLA) Class I and II Alleles on the Risk of Multiple Sclerosis. *PLoS One* 2012; 7(5): e36779. https://doi.org/10.1371/journal.pone.0036779 PMID: 22586495

19. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Diijn CM, Nobel JA, et al. Fine- Mapping the Genetic Association of the Major Histocompatibility Complex in Multiple Sclerosis: HLA and Non-HLA Effects. *PLoS Genet* 2014; 9(11):e1003926.

20. Chao MJ, Barnardo MC, Lincoln MR, Ramagopalan SV, Herrera BM, Dyment DA, et al. HLA class I alleles tag HLA-DRB1*1501 haplotypes for differential risk in multiple sclerosis susceptibility. *Proc Natl Acad Sci USA* 2008; 105:13069–74. https://doi.org/10.1073/pnas.0801042105 PMID: 18765817
21. Lincoln MR, Ramagopalan SV, Chao MJ, Herrera BM, Deluca GC, Orton SM, et al. Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. Proc Natl Acad Sci USA 2009; 106:7542–7. https://doi.org/10.1073/pnas.0812664106 PMID: 19380721

22. Multiple Sclerosis Genetics Group. Linkage of the MHC to familial multiple sclerosis suggests genetic heterogeneity. Hum Molec Genet 1998; 7:1229–1234. https://doi.org/10.1093/hmg/7.8.1229 PMID: 9668163

23. Isobe N, Gourraud PA, Harbo HF, Caillier SJ, Santaniello A, Khankhanian P, et al. Genetic risk variants in African Americans with multiple sclerosis. Neurology 2013; 81:219–227 https://doi.org/10.1212/WNL.0b013e318289e2f1 PMID: 23771490

24. Brynedal B, Duvefelt K, Jonasdottig G, Roos IM, Akesson E, Palmgren J, et al. HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. PLoS One 2007; 2:e664. https://doi.org/10.1371/journal.pone.0000664 PMID: 17653284

25. International Multiple Sclerosis Genetics Consortium, Welcome Trust Case Control Consortium. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 2011; 476:214–219. https://doi.org/10.1038/nature10251 PMID: 2183088

26. Martin R, Sospedra M, Eiermann T, Olsson T. Multiple sclerosis: doubling down on MHC. Trends Genet 2021; https://doi.org/10.1016/j.tig.2021.04.012 PMID: 34006391

27. Stenzel A, Lu T, Koch WA, Hampe J, Guenther SM, De La Vega FM, et al. Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. Hum Genet. 2004; 114:377–385. https://doi.org/10.1007/s00439-003-1075-5 PMID: 14740295

28. Gragert L, Madbouly A, Freeman J, Maiers M. Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. Hum Immunol 2013; 74:1313–1320. https://doi.org/10.1016/j.humimm.2013.06.025 PMID: 23806270

29. Goodin DS, Khankhanian P, Gourraud PA, Vince N. Highly conserved extended haplotypes of the major histocompatibility complex and their relationship to multiple sclerosis susceptibility. PLoS One 2018; 13(2):e0190043. https://doi.org/10.1371/journal.pone.0190043 PMID: 29433832

30. Linjama T, Räther C, Ritari J, Peräsaari J, Eberhard HP, Korhonen M, et al. Extended HLA haplotypes and their impact on DPB1 matching of unrelated hematologic stem cell transplant donors. Biol Blood Marrow Transplant. 2019; 25:1956–1964. https://doi.org/10.1016/j.bbmt.2019.07.008 PMID: 31306777

31. Goodin DS, Khankhanian P, Gourraud PA, Vince N. Genetic susceptibility to multiple sclerosis: Interactions between conserved extended haplotypes of the MHC and other susceptibility regions. (submitted).

32. Oksenberg JR, Barcellos LF, Cree BA, Baranzini SE, Bugawan TL, Khan O, et al. Mapping Multiple Sclerosis Susceptibility to the HLA-DR Locus in African Americans. Am J Hum Genet. 2004; 74:160–167. https://doi.org/10.1086/380997 PMID: 14669136

33. Caillier SJ, Briggs F, Cree BA, Baranzini SE, Fernandez-Vila M, Ramsay PP, et al. Uncoupling the Roles of HLA-DRB1 and HLA-DRB5 Genes in Multiple Sclerosis J Immunol. 2008; 181:5473–5480. https://doi.org/10.4049/jimmunol.181.8.5473 PMID: 18832704

34. Quandt JA, Huh J, Baig M, Yao K, Ito N, Bryant M, et al. Myelin Basic Protein-Specific TCR/HLA-DRB5*01:01 Transgenic Mice Support the Etiologic Role of DRB5*01:01 in Multiple Sclerosis J Immunol. 2012; 189:2897–2908. https://doi.org/10.4049/jimmunol.1103087 PMID: 2288134

35. Etzensperger R, McMahon RM, Jones EY, Fugger L. Dissection of the multiple sclerosis associated DR2 haplotype. J Autoimmun. 2008; 31:201–207. https://doi.org/10.1016/j.jaut.2008.04.016 PMID: 18513924

36. Kedod RL, Gaylor DW. On the additive and multiplicative models of relative risk. Biometrical J 1989; 31:359–370.

37. Greenland S. Additive Risk versus Additive Relative Risk Models. Epidemiology 1993; 4:32–36. https://doi.org/10.1097/00001648-199301000-00007 PMID: 8420578

38. Rothman KJ, Greenland S. Modern Epidemiology. Lippincott, Williams & Wilkins, Philadelphia. PA, 1998

39. Viera AJ. Odds ratios and risk ratios: What's the difference and why does it matter?. South Med J. 2008; 101:730–734. https://doi.org/10.1097/SMJ.0b013e31817a7ee4 PMID: 18580722

40. Chi C, Shao X, Rhead B, Gonzales E, Smith JB, Xiang AH, et al. Admixture mapping reveals evidence of differential multiple sclerosis risk by genetic ancestry. PLoS Genet. 2019; 15(1):e1007808. https://doi.org/10.1371/journal.pgen.1007808 PMID: 30653506

41. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol 1993; 13:227–231. https://doi.org/10.1002/ana.410130302 PMID: 6847134
42. McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001; 50:121–127. https://doi.org/10.1002/ana.1032 PMID: 11456302

43. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Ann Neurol* 2005; 58:840–846. https://doi.org/10.1002/ana.20703 PMID: 16283615

44. Beecham AH, Amezcu L, Chinea A, Manrique CP, Rubi C, Isobe N, et al. The genetic diversity of multiple sclerosis risk among Hispanic and African American populations living in the United States. *Mult Scler J* 2020; 26:1329–1339. https://doi.org/10.1177/1352458519863764 PMID: 31368393

45. Zheng X, J Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, et al. (2014) HIBAG–HLA genotype imputation with attribute bagging. *Pharmacogenom J* 14:192–200. https://doi.org/10.1038/tjp.2013.18 PMID: 23712092

46. Fast model-based estimation of ancestry in unrelated individuals. Alexander DH, Novembre J, Lange K. *Genome Res* 2009; 19:1655–1654; https://doi.org/10.1101/gr.094052.109 PMID: 19648217

47. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81:559–575. https://doi.org/10.1086/519795 PMID: 17701901

48. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature* 2015; 526(7571):68–74. https://doi.org/10.1038/nature15393 PMID: 26432245

49. Gourraud PA, Lamiraux P, El-Kadhi N, Raffoux C, Cambon-Thomsen A. Inferred HLA haplotype information for donors from hematopoietic stem cells donor registries. *Hum Immunol* 2005; 66:563–70. https://doi.org/10.1016/j.humimm.2005.01.011 PMID: 15935894

50. Gourraud PA, Khankhanian P, Cereb N, Yang SY, Feolo M, Maiers M, et al. HLA diversity in the 1000 genomes dataset. *PLoS One* 2014; 9:e9782. https://doi.org/10.1371/journal.pone.0097282 PMID: 24988075

51. Khankhanian P, Gourraud PA, Lizee A, Goodin DS. Haplotype-based approach to known MS-associated regions increases the amount of explained risk. *J Med Genet*. 2015; 52:587–594. https://doi.org/10.1136/jmedgenet-2015-103071 PMID: 26185143