Identification of Two Independent Risk Factors for Lupus within the MHC in United Kingdom Families

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Published Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>doi:10.1371/journal.pgen.0030192</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Citable link</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:4454176">http://nrs.harvard.edu/urn-3:HUL.InstRepos:4454176</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Terms of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Identification of Two Independent Risk Factors for Lupus within the MHC in United Kingdom Families

Michelle M. A. Fernando1, Christine R. Stevens2, Pardis C. Sabeti2, Emily C. Walsh2, Alasdair J. M. McWhinnie3, Anila Shah3, Todd Green2, John D. Rioux2,4*, Timothy J. Vyse1*

1 Section of Molecular Genetics and Rheumatology, Imperial College London, London, United Kingdom, 2 The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, 3 Histocompatibility Laboratories and Research Institute, The Anthony Nolan Trust, London, United Kingdom, 4 Université de Montréal, Montréal Heart Institute, Montréal, Québec, Canada

The association of the major histocompatibility complex (MHC) with SLE is well established yet the causal variants arising from this region remain to be identified, particularly due to inadequate study design and the strong linkage disequilibrium demonstrated by genes across this locus. The majority of studies thus far have identified strong association with classical class II alleles, in particular HLA-DRB1*0301 and HLA-DRB1*1501. Additional associations have been reported with class III alleles; specifically, complement C4 null alleles and a tumor necrosis factor promoter SNP (TNF-308G/A). However, the relative effects of these classes II and class III variants have not been determined. We have thus used a family-based approach to map association signals across the MHC class II and class III regions in a cohort of 314 complete United Kingdom Caucasian SLE trios by typing tagging SNPs together with classical typing of the HLA-DRB1 locus. Using TDT and conditional regression analyses, we have demonstrated the presence of two distinct and independent association signals in SLE: HLA-DRB1*0301 (nominal p = 4.9 × 10^{-6}, permuted p < 0.0001, OR = 2.3) and the T allele of SNP rs419788 (nominal p = 4.3 × 10^{-6}, permuted p < 0.0001, OR = 2.0) in intron 6 of the class III region gene SKIV2L. Assessment of genotypic risk demonstrates a likely dominant model of inheritance for HLA-DRB1*0301, while rs419788-T confers susceptibility in an additive manner. Furthermore, by comparing transmitted and untransmitted parental chromosomes, we have delimited our class II signal to a 180 kb region encompassing the alleles HLA-DRB1*0301-HLA-DQA1*0501-HLA-DQB1*0201 alone. Our class III signal importantly excludes independent association at the TNF promoter polymorphism, TNF-308G/A, in our SLE cohort and provides a potentially novel locus for future genetic and functional studies.

Introduction

Since the early 1970s, the human major histocompatibility complex (MHC) has been shown to be associated with a number of autoimmune, inflammatory, and infectious diseases, and it continues to be the focus of intense research [1]. The recently defined extended MHC (xMHC) encompasses 7.6 Mb of genome on 6p21.3 and is divided into five subregions from telomere to centromere: extended class I, classical class I, classical class III, classical class II, and extended class II. In addition, the MHC contains two hypervariable regions, the RCCX module in class III (spanning complement C4) and the HLA-DRB genes in class II, that both exhibit copy number polymorphism. Examination of the sequence across the extended MHC reveals the presence of 421 genes, and over 252 (60%) are thought to be expressed [2]. Around 40% of genes expressed within the classical MHC encode proteins with putative immunomodulatory function [3]. The classical class I and class II loci encode the human leucocyte antigen (HLA) proteins involved in antigen presentation to T cells, initiating the adaptive immune response. The class III region contains the greatest density of genes in the genome (58 expressed genes), which are often found in functionally related clusters [2].

A major obstacle in the identification of disease-specific causal variants within the MHC has been the strong linkage disequilibrium (LD) exhibited by certain alleles in this region, resulting in the existence of long-range, conserved, extended haplotypes [4], also known as ancestral haplotypes [5], sometimes spanning more than 2 Mb [6]. Thus, for many MHC-associated diseases, it has only been possible to delimit association signals to a particular extended haplotype or segment of one.

Copyright: © 2007 Fernando et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: CEPH, Utah residents with ancestry from northern and western Europe; CI, confidence interval; EHH, extended haplotype homozygosity; geno-PDT, genotype-pedigree disequilibrium test; HLA, human leucocyte antigen; HWE, Hardy–Weinberg equilibrium; LD, linkage disequilibrium; MHC, major histocompatibility complex; OR, odds ratio; REHH, relative extended haplotype homozygosity; SLE, systemic lupus erythematosus; T, transmitted; UT, untransmitted; Yoruba, Yoruba in Ibadan, Nigeria;

* To whom correspondence should be addressed. E-mail: John.Rioux@inflammgen.org (JDR); t.vyse@imperial.ac.uk (TJV)
Author Summary

Systemic lupus erythematosus (SLE/lupus) is a complex autoimmune disease in which the body’s immune system attacks its own tissues, causing inflammation in a variety of different organs such as the skin, joints, and kidneys. The cause of lupus is not known, but genes play a significant role in the predisposition to disease. The major histocompatibility complex (MHC) on Chromosome 6 contains at least 100 different genes that affect the immune system, including the genes with the strongest effect on lupus susceptibility. Despite the importance of the MHC in SLE, the identity of the actual genes in the MHC region that cause SLE has remained elusive. In the present study, we used the latest set of genetic markers present at the MHC in lupus families to identify the actual genes that affect the disease. To our knowledge, we have shown for the first time that two separate groups of genes are involved in SLE. One group of genes alters how the immune system may inappropriately target its own tissues in the disease. How the second set of genes predisposes to SLE is the subject of ongoing study.

Systemic lupus erythematosus (SLE/lupus) is a chronic, multi-system autoimmune disease affecting young women ten times more commonly than men. The worldwide prevalence of SLE is estimated at between 12 and 124 cases per 100,000 individuals [7]. SLE is characterized by the presence of pathogenic autoantibodies to nuclear and cell-surface antigens that show affinity maturation. The consequent immune complexes deposit in tissues, causing inflammation and damage. It is well established that there is a complex genetic component to lupus etiology, with hormonal and environmental influences also contributing to disease susceptibility [8,9].

The MHC has been the most consistently confirmed genetic risk factor for SLE, and multiple different genes within the region have been significantly implicated with disease susceptibility. For example, hereditary and acquired deficiencies of the early classical complement component C4, located within the MHC class III locus, leads to a lupus-like syndrome. A role for another class III gene, tumour necrosis factor alpha (TNF), in SLE was suggested following the observation that the lupus-prone New Zealand F1 mouse hybrid exhibits constitutively low TNF expression [10]. Recently, the development of antinuclear antibodies in patients treated with TNF antagonists has also stimulated interest in the possible role of TNF in SLE [11–13]. Murine and human candidate gene studies, together with genome-wide linkage screens, provide further support that multiple genetic loci, including the mouse MHC complex H2 and the human MHC locus, contribute to disease susceptibility [14–17].

It should be noted that the human MHC was first associated with SLE in 1971, when studies demonstrated that lupus probands were enriched for the class I alleles HLA-A8 (now known as HLA-B8) and HLA-W15 (now known as HLA-B15) when compared with healthy controls [18,19]. Further case control association studies were small, performed in ethnically diverse populations, and tested only a small number of the classical HLA and complement C4 alleles. The most consistent findings reported to date are associations with the class II alleles HLA-DR2 (DRB1*1501) and HLA-DR3 (DRB1*0301) and their respective haplotypes in Caucasian populations. The complement C4A null allele (C4A*Q0) has shown inconsistent association with lupus in a number of studies—a situation that may reflect genetic heterogeneity in disease susceptibility [20–23]. In addition, a recent study has demonstrated that low C4A copy number is a risk factor for lupus in a European American cohort [24]. However, the C4A null allele lies on the lupus-associated DR3 “autoimmune” extended haplotype (AH8.1), which exhibits extremely strong LD [6]. It therefore remains to be definitively established whether this locus constitutes a distinct susceptibility allele to that of the class II association or is merely in LD with it. Similarly, certain TNF promoter polymorphisms, including the much-studied SNP TNF-308G/A, have shown association with SLE; but again, many of these variants are carried on the highly conserved 8.1 ancestral haplotype, thus restricting interpretation of these data.

In 2002, a family-based study employing microsatellites as surrogate markers for HLA-DRB1 haplotypes in Caucasian lupus families demonstrated association with DR3, DR2, and DR8 (DRB1*0801)-containing haplotypes. In that study, Graham and colleagues reported that, taking advantage of recombinant chromosomes, the disease risk region could be limited to a 1 Mb region encompassing classical class II and class III [25].

We have performed a medium resolution association mapping study of the MHC in lupus families, utilizing a combination of SNPs and four-digit typing at the HLA-DRB1 locus in order to anchor haplotypes. Sixty-eight SNPs were successfully genotyped across a 2.4 Mb region of the MHC, from the class I locus KIAA1949 to the class II gene HLA-DPB2, in 314 UK Caucasian SLE trios. We used these data to perform a family-based association study in an attempt to distinguish the relative effects of the class II and class III regions of the MHC in lupus susceptibility. In addition, we employed the long-range haplotype test to search for the presence of high-frequency, extended haplotypes indicative of recent positive selection [26]. We have also used family-based and case-control strategies to examine genotypic risk at HLA-DRB1 and rs419788.

Results

Association Testing of HLA-DRB1 and MHC Region SNPs

In order to define the causal variation within the MHC region, we typed 314 complete SLE trios for the HLA-DRB1 gene as well as for 86 SNPs across a 2.4 Mb region encompassing the HLA class I locus HLA-B to HLA-DPB2. High-quality genotype data was obtained for HLA-DRB1 and 68 MHC SNPs (see Table S1 for quality control data). Association testing of the HLA-DRB1 gene revealed a significant association with HLA-DRB1*0301 (nominal p = 4.9 × 10⁻⁸, permuted p < 0.0001, T:U = 129:55) in our lupus cohort (Table 1). There was also a trend for under transmission of the HLA-DRB1*0701 allele (nominal p = 0.0013, T:U 42:77); however, this association was no longer significant after correction for multiple testing as determined by 10,000 permutations of the dataset (permuted p = 0.09). Furthermore, we did not find evidence of association with HLA-DRB1*1501 (nominal p = 1.0, T:U 70:70) or HLA-DRB1*0801 (nominal p = 1.0, T:U = 11:11) in our cohort (see Table S2 for complete HLA-DRB1 association data); alleles previously suggested by microsatellite typing of a US lupus cohort [25].
Association testing of the MHC region SNPs also identified significant evidence of association to SLE (Table 1 for associated markers and Table S3 for all MHC SNPs). The SNP with the most significant association, rs419788 (nominal \( p = 4.3 \times 10^{-8} \), permuted \( p < 0.0001 \)) was of similar strength to that of the HLA-DRB1\*0301 allele, with odds ratios (ORs) and 95\% confidence intervals (CIs) of 2.0 (1.6–2.6) and 2.3 (1.7–3.2), respectively. This SNP is located within intron 6 of the \( HLA-DRB1 \) promoter, which is superkiller viral-like and is located approximately 500 kb telomeric to the \( HLA-DRB1 \) gene. Of the other 12 SNPs that were significantly associated with SLE (nominal \( p = 4.0 \times 10^{-4} \) to \( 2.5 \times 10^{-2} \); permuted \( p = 0.03 \) to <0.0001), one was located in the class I region between \( HLA-B \) and MICA, seven were located in the class III region, and four were situated in the class II region (Table 1; Figure S1). Specifically, the seven associated class III SNPs were located in or close to the following genes: the TNF promoter, \( BAT3, SLC44A4, EHM12, TNXB, GPSM3, \) and \( NOTCH4 \). One of the four class II associated SNPs was close to \( HLA-DRB1 \), two were between \( HLA-DRB1 \) and \( HLA-DQA1 \) and one was in intron 1 of \( HLA-DQA1 \). The correlation between all 68 SNPs and \( HLA-DRB1 \) in our UK SLE cohort is illustrated in Figure 1. The markers showing significant association are highlighted.

### Conditional Analyses Identify Two Independent Association Signals in the MHC

In order to establish whether the two most associated signals identified in this association-mapping experiment are likely to represent a single causal allele or independent risk factors, we first examined the association data conditioned upon the presence of the \( HLA-DRB1\*0301 \) allele. We found that four of the 13 associated SNPs showed evidence of signals independent of \( HLA-DRB1\*0301 \) in our dataset, the strongest of which was rs419788 (Table 1). We therefore conditioned the three remaining SNPs (rs2523589, rs1052486, and rs605203) on rs419788 to assess whether these signals are truly independent of each other or show association due to LD with rs419788. In addition, we included \( HLA-DRB1 \) in stepwise conditional regression analyses performed on the SNPs showing association independent of \( HLA-DRB1 \) (unpublished data). These analyses demonstrated that the observed association signals at rs2523589, rs1052486, and rs605203 were predominantly dependent upon the association at rs419788, and suggested that there are two major independent association signals in the MHC in UK SLE: \( HLA-DRB1 \) and rs419788. The independence of the association signals at \( HLA-DRB1 \) and rs419788 is further supported by the observation that there is only modest LD between these two (\( r^2 = 0.24 \)). There was no association with any other \( HLA-DRB1 \) allele and the four SNPs independent of \( HLA-DRB1\*0301 \) (TRANSMIT, unpublished data).

The association of the tumour necrosis factor gene promoter SNP TNF-308G/A with SLE is lost after conditioning for \( HLA-DRB1\*0301 \) in our cohort. If we perform the reverse analysis and condition \( HLA-DRB1\*0301 \) on the presence of the TNF promoter SNP, we find that the association remains, confirming that our TNF association is secondary to that of \( HLA-DRB1\*0301 \).

### Genotypic Risk for Class II and Class III Association Signals

Having established independent association at the allelic level with \( HLA-DRB1\*0301 \) and rs419788-T in our UK SLE cohort, we wanted to further determine the genotypic risk conferred by these variants and hence gain insight into their underlying mode of inheritance in lupus. We used case-control and family-based analyses to assess genotypic risk at \( HLA-DRB1 \), while the family-based test alone was used for

---

**Table 1. Single Marker Association and Conditional Regression Analysis in 314 UK SLE Trios**

<table>
<thead>
<tr>
<th>Marker #</th>
<th>Marker</th>
<th>Location</th>
<th>MAF</th>
<th>Minor Allele</th>
<th>Associated Allele</th>
<th>Single Marker Association</th>
<th>Conditional Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>rs2523589</td>
<td>Intergenic HLA-B and MICA</td>
<td>0.46</td>
<td>T</td>
<td>G</td>
<td>T:U Nominal p Value</td>
<td>DRB1*0301 p Value</td>
</tr>
<tr>
<td>12</td>
<td>rs1800629</td>
<td>TNF promoter</td>
<td>0.25</td>
<td>T</td>
<td>T</td>
<td>139:85 3.0 \times 10^{-4}</td>
<td>0.200</td>
</tr>
<tr>
<td>14</td>
<td>rs1052486</td>
<td>BAT3 exon 14</td>
<td>0.41</td>
<td>G</td>
<td>A</td>
<td>159:92 2.4 \times 10^{-5}</td>
<td>0.40</td>
</tr>
<tr>
<td>21</td>
<td>rs605203</td>
<td>5' SLC44A4</td>
<td>0.39</td>
<td>C</td>
<td>C</td>
<td>163:95 2.3 \times 10^{-5}</td>
<td>0.70</td>
</tr>
<tr>
<td>22</td>
<td>rs589428</td>
<td>EHM12 intron 27</td>
<td>0.39</td>
<td>T</td>
<td>T</td>
<td>172:107 1.0 \times 10^{-4}</td>
<td>0.60</td>
</tr>
<tr>
<td>23</td>
<td>rs419788</td>
<td>SKV2L intron 6</td>
<td>0.37</td>
<td>T</td>
<td>T</td>
<td>180:90 4.3 \times 10^{-8}</td>
<td>0.001</td>
</tr>
<tr>
<td>25</td>
<td>rs433061</td>
<td>TNXB intron 31</td>
<td>0.18</td>
<td>A</td>
<td>A</td>
<td>110:47 5.0 \times 10^{-7}</td>
<td>0.200</td>
</tr>
<tr>
<td>27</td>
<td>rs204989</td>
<td>GPSM3 intron 3</td>
<td>0.26</td>
<td>A</td>
<td>A</td>
<td>140:78 2.7 \times 10^{-4}</td>
<td>0.100</td>
</tr>
<tr>
<td>29</td>
<td>rs3131926</td>
<td>NOTCH4 intron 18</td>
<td>0.21</td>
<td>T</td>
<td>T</td>
<td>150:73 2.5 \times 10^{-7}</td>
<td>0.600</td>
</tr>
<tr>
<td>31</td>
<td>rs2227139</td>
<td>Intergenic HLA-DRB1 and HLA-DRB9</td>
<td>0.46</td>
<td>G</td>
<td>G</td>
<td>145:91 4.0 \times 10^{-4}</td>
<td>0.500</td>
</tr>
<tr>
<td>48</td>
<td>HLA-DRB1*0301</td>
<td>Intergenic HLA-DRB1 and HLA-DRB1</td>
<td>0.21</td>
<td>A</td>
<td>A</td>
<td>129:55 4.9 \times 10^{-8}</td>
<td>0.0008</td>
</tr>
<tr>
<td>49</td>
<td>rs3129763</td>
<td>Intergenic HLA-DRB1 and HLA-DQA1</td>
<td>0.28</td>
<td>T</td>
<td>T</td>
<td>145:77 5.0 \times 10^{-6}</td>
<td>0.300</td>
</tr>
<tr>
<td>50</td>
<td>rs2040406</td>
<td>Intergenic HLA-DRB1 and HLA-DQA1</td>
<td>0.23</td>
<td>G</td>
<td>G</td>
<td>112:59 5.0 \times 10^{-5}</td>
<td>0.050</td>
</tr>
<tr>
<td>51</td>
<td>rs2187668</td>
<td>HLA-DQA1 intron 1</td>
<td>0.21</td>
<td>T</td>
<td>T</td>
<td>132:65 1.8 \times 10^{-6}</td>
<td>0.002</td>
</tr>
</tbody>
</table>

*10,000 permutations in Haploview.

a HLA-DRB1*0301 tag SNP.

MAF, minor allele frequency; N/A, not applicable.
doi:10.1371/journal.pgen.0030192.t001
rs419788. Common family-based tests of LD, such as those used in this study (Genehunter), measure transmission distortion based on allele counts rather than genotype counts; the former has been shown to be more powerful under additive models, while the latter has greater power under recessive or dominant genetic models [27]. The genotype-pedigree disequilibrium test (geno-PDT) determines LD between a locus genotype and disease by comparing genotypes that are transmitted from parent to proband with those that are not [27]. We used the geno-PDT to assess genotypic risk for our class II and class III association signals: HLA-DRB1 and the SNP rs419788. In the case-control analysis for HLA-DRB1, ORs with 95% CI were calculated and Fisher’s exact test employed to assess statistically significant differences between HLA-DRB1 genotypes in lupus probands and healthy controls. For HLA-DRB1, the alleles were coded as follows: HLA-DRB1*0301, HLA-DRB1*1501, HLA-DRB1*X where X represents all HLA-DRB1 alleles other than HLA-DRB1*0301, and HLA-DRB1*1501. We included HLA-DRB1*1501 in the analysis, even though we find no allelic association in our cohort, because previous studies have shown a greater risk for lupus in individuals who are compound heterozygotes for HLA-DRB1*0301- and HLA-DRB1*1501-containing haplotypes [25,28]. Overall the results are consistent with a dominant effect from HLA-DRB1*0301 (Table 2) and a dose-dependent (additive) effect from rs419788-T (Table 3). Specifically, both case-control and geno-PDT demonstrate that there is no dose-dependent increase in disease risk for HLA-DRB1*0301. Rather, it appears that the presence of a single copy of HLA-DRB1*0301 alone is sufficient to increase susceptibility to disease. Moreover the 0301/X genotypes constitute the greatest risk in our cohort rather than the 0301/1501 heterozygotes. Genotypes containing HLA-DRB1*1501 in the absence of HLA-DRB1*0301 revealed no significant association in our cohort.

All three rs419788 genotypes demonstrated significant association in our lupus families (Table 3). The common CC

Figure 1. LD Plot of All 69 MHC Markers Typed in 314 UK SLE Trios

The figure illustrates the correlation between all 69 MHC markers successfully typed across 2.4 Mb of genome encompassing MHC class II and class III in 314 UK SLE trios. The markers are identified by rs numbers for the 68 SNPs while HLA-DRB1 indicates all HLA-DRB1 alleles coded as a biallelic marker: HLA-DRB1*0301/non-HLA-DRB1*0301. The numbers shown below the marker names correspond to the marker numbering in Tables 1, S1, and S3 for ease of interpretation. The grey ticks indicate SNP location to scale. The triangular units designate haplotype blocks. The open rectangles represent the 14 markers that show significant association in the cohort after permutation testing; the green rectangle highlights our class III signal, SNP rs419788; and the red rectangle designates our class II signal, HLA-DRB1. The 53 markers used for the EHH analyses and haplotype bifurcation plots are designated by open squares. The degree of correlation between pairs of markers is indicated by the correlation coefficient r’ (where r’ = 0 = no correlation, white; 0 < r’ < 1, shades of grey; r’ = 1 = complete correlation, black).

doi:10.1371/journal.pgen.0030192.g001

rs419788. Common family-based tests of LD, such as those used in this study (Genehunter), measure transmission distortion based on allele counts rather than genotype counts; the former has been shown to be more powerful under additive models, while the latter has greater power under recessive or dominant genetic models [27]. The genotype-pedigree disequilibrium test (geno-PDT) determines LD between a locus genotype and disease by comparing genotypes that are transmitted from parent to proband with those that are not [27]. We used the geno-PDT to assess genotypic risk for our class II and class III association signals: HLA-DRB1 and the SNP rs419788. In the case-control analysis for HLA-DRB1, ORs with 95% CI were calculated and Fisher’s exact test employed to assess statistically significant differences between HLA-DRB1 genotypes in lupus probands and healthy controls. For HLA-DRB1, the alleles were coded as follows: HLA-DRB1*0301, HLA-DRB1*1501, HLA-DRB1*X where X represents all HLA-DRB1 alleles other than HLA-DRB1*0301, and HLA-DRB1*1501. We included HLA-DRB1*1501 in the analysis, even though we find no allelic association in our cohort, because previous studies have shown a greater risk for lupus in individuals who are compound heterozygotes for HLA-DRB1*0301- and HLA-DRB1*1501-containing haplotypes [25,28]. Overall the results are consistent with a dominant effect from HLA-DRB1*0301 (Table 2) and a dose-dependent (additive) effect from rs419788-T (Table 3). Specifically, both case-control and geno-PDT demonstrate that there is no dose-dependent increase in disease risk for HLA-DRB1*0301. Rather, it appears that the presence of a single copy of HLA-DRB1*0301 alone is sufficient to increase susceptibility to disease. Moreover the 0301/X genotypes constitute the greatest risk in our cohort rather than the 0301/1501 heterozygotes. Genotypes containing HLA-DRB1*1501 in the absence of HLA-DRB1*0301 revealed no significant association in our cohort.

All three rs419788 genotypes demonstrated significant association in our lupus families (Table 3). The common CC
Table 2. Case-Control and Family-Based Genotypic Risk for HLA-DRB1 in UK SLE Trios

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Case-Control Analysisa</th>
<th>Genotype-PDTb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR 95% CI Fisher's p Value</td>
<td>T UT T:UT Z p Value</td>
</tr>
<tr>
<td>1501/1501</td>
<td>1.3 0.6–2.9 0.5</td>
<td>7 5 1.4 0.6 0.6</td>
</tr>
<tr>
<td>1501/X</td>
<td>1.2 0.9–1.7 0.2</td>
<td>42 56 0.8 −1.6 0.1</td>
</tr>
<tr>
<td>1501/X+</td>
<td>1.5 1.1–2.0 0.009</td>
<td>65 69 0.9 −0.4 0.7</td>
</tr>
<tr>
<td>0301</td>
<td>1.9 1.2–3.0 0.009</td>
<td>23 13 1.8 1.8 0.08</td>
</tr>
<tr>
<td>0301/0301</td>
<td>2.6 1.5–4.5 0.001</td>
<td>17 11 1.5 1.1 0.26</td>
</tr>
<tr>
<td>0301/X+</td>
<td>2.9 2.2–3.8 &lt;0.0001</td>
<td>91 39 2.3 5.0 7.1 × 10⁻¹</td>
</tr>
<tr>
<td>0301/1501</td>
<td>3.0 2.3–3.9 &lt;0.0001</td>
<td>114 52 2.2 5.5 3.3 × 10⁻¹</td>
</tr>
<tr>
<td>X/X</td>
<td>0.3 0.2–0.4 &lt;0.0001</td>
<td>87 143 0.6 −4.7 2.6 × 10⁻³</td>
</tr>
</tbody>
</table>

aData from 296 unrelated UK SLE cases and 1,667 unrelated healthy male controls.

bData from 267 unrelated complete UK SLE trios corresponding to a total of 534 transmitted and untransmitted genotypes. 1501 = HLA-DRB1*1501, 1502, 1504, and 1506 (the majority of alleles are 1501) in controls and 1501 = HLA-DRB1*1501 and 1502 in UK SLE, 0301 = HLA-DRB1*0301, X is any DRB1 allele other than 0301 or 1501.

doi:10.1371/journal.pgen.0030192.t002

genotype was significantly under transmitted, while the rare T allele displayed dose-dependent over transmission to lupus probands.

Characterization of HLA-DRB1*0301 Risk Haplotype

Next, we wanted to further delimit the MHC class II association signal that we have detected at HLA-DRB1. We used phased parental genotype data to compare the allelic composition of HLA-DRB1*0301-bearing haplotypes that were transmitted (T) to affected probands to those that were not transmitted (or untransmitted, UT) with the aim of identifying differences that could delineate the lupus susceptibility interval(s) arising from this haplotype (summarized in Figures 2A, 2B, and S2). We observed a striking difference between transmitted and untransmitted chromosomes within the class II region: nearly all transmitted HLA-DRB1*0301 haplotypes (99%) are identical across a 180 kb region defined by eight SNPs, whereas the corresponding region within untransmitted HLA-DRB1*0301 haplotypes exhibits significant recombination. These data strongly suggest the existence of a risk haplotype that, interestingly, contains only three expressed genes: HLA-DRB1, HLA-DQA1, and HLA-DQB1. Furthermore, we can confidently define the allelic composition of this risk haplotype, as these three genes are in strong LD and occur in one common haplotype in Caucasians: HLA-DRB1*0301-HLA-DQA1*0501-HLA-DQB1*0201. Thus, we hypothesize that the specific combination of all three alleles is required to confer disease risk in lupus or that disease susceptibility lies with either HLA-DRB1*0301 or the HLA-DQ alleles. We do not have sufficient numbers of recombinant chromosomes in this risk region to further delimit this signal: 2/176 (1.1%) transmitted HLA-DRB1*0301 haplotypes are recombinant at HLA-DQA1-HLA-DQB1; 3/178 (1.7%) transmitted haplotypes identical across HLA-DQA1-HLA-DQB1 do not possess HLA-DRB1*0301.

The composite relative extended haplotype homozygosity (REHH) versus frequency plot for UK SLE; Utah residents with ancestry from northern and western Europe (CEPH); and Yoruba in Ibadan, Nigeria (Yoruba) populations is shown in Figure 3A. We can only comment on evidence for positive selection in CEPH individuals, as we have used this population alone to assess background variation on Chromosome 6. The SLE and Yoruba cohorts are shown for comparative purposes. We find no evidence of positive selection for HLA-DRB1*0301 in the CEPH population. However, this allele is enriched in our lupus cohort (21% of parental chromosomes) and displays greater extended homozgyosity when compared with HLA-DRB1*0301-bearing haplotypes in CEPH and Yoruba. Hence, the HLA-DRB1*0301 allele in lupus is observed as an outlier on the plot when compared to background variation in CEPH. These data support our previous observations (outlined above) of the highly conserved nature of HLA-DRB1*0301 haplotypes in lupus. In addition, the haplotype bifurcation plots centered on HLA-DRB1*0301 for UK SLE, CEPH, and Yoruba populations in Figure 3B illustrate preservation of the common HLA-DRB1*0301 haplotype in CEPH and UK SLE, while that seen in the Yoruba is significantly different. The class II regions of all three populations are essentially identical across our chosen SNPs; the main differences lie in class III. The difference in African populations in the class III region is one possible explanation for the lack of evidence for an association between HLA-DRB1*0301 and SLE in African or African American populations. However, HLA-DRB1*0301 has a lower frequency (~7%–10%) in African populations compared with Europeans (~13%), and the number of HLA association studies conducted in African populations is very limited.

Table 3. Genotype-Pedigree Disequilibrium Test for SNP rs419788 in 281 UK SLE Trios

<table>
<thead>
<tr>
<th>Genotype</th>
<th>T</th>
<th>UT</th>
<th>T:UT</th>
<th>Z Score</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C/C</td>
<td>83</td>
<td>142</td>
<td>0.6</td>
<td>−5.2</td>
<td>2.5 × 10⁻⁷</td>
</tr>
<tr>
<td>C/T</td>
<td>138</td>
<td>110</td>
<td>1.3</td>
<td>2.4</td>
<td>0.016</td>
</tr>
<tr>
<td>T/T</td>
<td>60</td>
<td>29</td>
<td>2.1</td>
<td>3.5</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Data from 281 complete unrelated UK SLE trios corresponding to a total of 562 transmitted and untransmitted genotypes. doi:10.1371/journal.pgen.0030192.t003

The MHC and SLE
Figure 2. Structure of Transmitted and Untransmitted HLA-DRB1*0301 and rs419788-T Allele Haplotypes

Haplotype bifurcation plots were constructed using 120 randomly selected parental chromosomes from each of the four datasets for comparative purposes except (B), where there were only 90 chromosomes in the entire dataset: (A) T HLA-DRB1*0301 haplotypes, (B) UT HLA-DRB1*0301 haplotypes, (C) T rs419788-T allele haplotypes, and (D) UT rs419788-T allele haplotypes. The allelic composition of the most common haplotype in each subset is shown: the core allele is represented as a dark blue double bar indicating haplotypes to the right and to the left of the core; otherwise, the common haplotype is depicted by dark grey bars. In parts (A) and (B), the rs419788-T allele in class III that shows association independent of HLA-DRB1*0301 in our cohort is indicated in green, while in parts (C) and (D), the allele HLA-DRB1*0301 is shown in green. The key difference between HLA-DRB1*0301 T and UT haplotypes lies within the class II region of the MHC. All HLA-DRB1*0301 T haplotypes are identical across a 180 kb region defined by eight SNPs (light blue), whereas the corresponding region within UT HLA-DRB1*0301 haplotypes exhibits significant recombination. This conserved class II interval encompasses only three expressed genes: HLA-DRB1, HLA-DQA1, and HLA-DQB1. Given the strong LD exhibited by HLA-DRB1*0301 haplotypes, the allelic composition of this risk region is known to be HLA-DRB1*0301-HLA-DQA1*0501-HLA-DQB1*0201. Both T and UT rs419788-T allele haplotypes show similar structure overall. The rs419788-T allele is clearly present on HLA-DRB1*0301 and non-HLA-DRB1*0301-containing haplotypes, lending credence to our observation that rs419788-T or another variant in LD with it constitutes an association signal independent of HLA-DRB1*0301 in our UK SLE cohort. (A) 120 from a total of 176 parental chromosomes, (B) 90 from a total of 90 parental chromosomes, (C) 120 from a total of 284 parental chromosomes, and (D) 120 from a total of 182 parental chromosomes.

doi:10.1371/journal.pgen.0030192.g002
(A) REHH vs. Hapmap Allele Frequency for Chromosome 6 in UK SLE, CEPH and Yoruban populations

(B) Comparison of DRB1*0301 haplotypes in (i) UK SLE, (ii) CEPH and (iii) Yoruban populations
passes roughly 270 kb containing class III genes from SLC44A4 to AGER, including the RCCX module. Next, we analyzed the haplotype block structure of this region in CEPH families using SNP data dumped from the International HapMap Project (http://www.hapmap.org/). The greater density of SNP typing available in the HapMap CEPH population compared to our current UK SLE map allowed us to potentially refine our signal by exploring correlations between our associated SNP and those surrounding it. Analysis of these data (Figure 4) suggests the presence of short-range LD around our associated variant, rs419788, in CEPH families, encompassing approximately 40 kb of the genome which includes the five genes: complement factor B (CFB), RD RNA binding protein (RDBP), SKIV2L, dom-3

Figure 3. REHH versus Frequency Analysis and Comparison of HLA-DRB1*0301 Haplotype Bifurcation Plots in UK SLE, CEPH, and Yoruba Populations
(A) Composite REHH versus frequency analysis comparing CEPH with UK SLE and Yoruba populations generated in SWEEP. REHH is shown on the vertical axis and haplotype frequency on the horizontal axis. Background CEPH genotype data for Chromosome 6 is indicated by grey data points; the 53 SNPs common to the UK SLE, CEPH, and Yoruba datasets are shown as red, blue, and turquoise data points, respectively. The 95th percentile for background variation in CEPH is indicated. The position of the HLA-DRB1*0301 allele is shown in all three cohorts (a, b, and c). The alleles rs2187668-T (HLA-DRB1*0301 tag SNP in UK SLE) (d) and HLA-DRB1*0301 in UK SLE are the only core markers observed above the 95th percentile. The associated class III SNP, rs419788-T, is also indicated (e). We can only examine evidence for positive selection in CEPH, as these are the data we have used to assess background variation. There is no evidence for positive selection of the HLA-DRB1*0301 allele in CEPH. This allele is enriched in our lupus cohort (21% of parental chromosomes), and displays greater extended homozygosity when compared with HLA-DRB1*0301-bearing haplotypes in CEPH and Yoruba.
(B) Comparison of HLA-DRB1*0301 haplotype bifurcation plots for (i) UK SLE, (ii) CEPH, and (iii) Yoruba populations. We show preservation of the common HLA-DRB1*0301 haplotype in CEPH and UK SLE, while that seen in the Yoruba is significantly different (differences indicated in yellow, core allele shown in dark blue). The class II regions of all three populations are essentially identical across our chosen SNPs; the main differences lie in class III. 120 (of 1,256) randomly selected transmitted and untransmitted parental UK SLE chromosomes used for haplotype bifurcation plot for comparison with 120 total CEPH and 120 total Yoruba chromosomes.
doi:10.1371/journal.pgen.0030192.g003

Figure 4. LD Plot of HapMap CEPH Data Centered on the Associated Class III SNP rs419788
One hundred and twenty kilobases of the MHC class III region centered on rs419788 (boxed) in HapMap CEPH families is illustrated. The grey ticks indicate SNP location with gene location shown above and rs numbers below. The triangular units designate haplotype blocks. The degree of correlation (LD) between pairs of markers is indicated by the correlation coefficient $r^2$ (where $r^2 = 0$ = no correlation, white; $0 < r^2 < 1$, shades of grey; $r^2 = 1$ = complete correlation, black). The dashed arrows depict LD between rs419788 and surrounding SNPs. Significant LD ($r^2 > 0.8$) is observed with rs419788 and three further SNPs within SKIV2L, two SNPs in STK19, and one SNP in CFB.
doi:10.1371/journal.pgen.0030192.g004
homolg Z (C. elegans) (DOM3Z), and serine/threonine kinase 19 (STK19), and does not include the complement C4 locus. Furthermore, assessment of marker association in our lupus dataset demonstrates that after conditioning for HLA-

Subphenotype Analysis

In order to gain further insight into disease pathogenesis, we examined common lupus subphenotypes. Such subsets are more homogeneous than lupus per se and thus maybe enriched for specific predisposing variants. In addition, one might expect a close association between MHC class II alleles and autoantibody subsets in lupus if these are indeed causal variants, given their role in antigen presentation and subsequent humoral immunity. We therefore tested our two main MHC association signals, HLA-DRB1*0301 and rs419788, for association with renal disease and autoantibody subsets in our lupus cohort.

We found that HLA-DRB1*0301 was associated with the presence of anti-Ro and anti-La antibodies in our UK SLE cohort, with the latter showing the greatest evidence of association (anti-La nominal \( p < 0.001 \) compared with anti-Ro nominal \( p < 0.025 \)). We found no association of HLA-DRB1*0301 with renal disease or any other autoantibody subsets in our dataset (see Table S4 for detailed results).

Genotypes of the SNP rs419788 were not associated with any of the tested lupus subphenotypes after controlling for the effect of HLA-DRB1*0301 (unpublished data).

Discussion

We present the first family-based SNP association study of the MHC in SLE. We have genotyped 69 markers (HLA-DRB1 and 68 SNPs) across 2.4 Mb of the MHC, encompassing class III and class II, in a cohort of 314 UK Caucasian SLE trios. Transmission disequilibrium testing of these data has shown predominant association with the alleles HLA-DRB1*0301 and rs419788-T, together with 12 other MHC SNPs. Moreover, using conditional analyses, we have shown that the two primary signals of association at the MHC are independent of each other. Specifically, one signal arises from HLA-

Table 4. Comparison of HLA-DRB1 Allele/Haplotype Frequencies in UK SLE and US (Minnesota) SLE Cohorts

<table>
<thead>
<tr>
<th>HLA-DRB1 Allele</th>
<th>UK SLE(^a) Allele Frequency (%)</th>
<th>UK(^b) Control Allele Frequency (%)</th>
<th>MNSLE Inferred HLA-DRB1-DQB1 Haplotype Frequency (%)</th>
<th>MNSLE(^c) Haplotype Frequency (%)</th>
<th>MN(^a) Control Haplotype Frequency (%)</th>
<th>MN(^b) Control Allele Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0301</td>
<td>27.7</td>
<td>12.7</td>
<td>17.3</td>
<td>9.8</td>
<td>13.7</td>
<td>12.4</td>
</tr>
<tr>
<td>0801</td>
<td>2.2</td>
<td>2.0</td>
<td>3.2</td>
<td>2.2</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>1501</td>
<td>14.4</td>
<td>14.7</td>
<td>21.8</td>
<td>14.2</td>
<td>14.2</td>
<td>14.2</td>
</tr>
</tbody>
</table>

\(^a\)Source: current publication.

\(^b\)Source: http://www.allelefrequencies.net.

\(^c\)Source: [25].

MN, Minnesota; n, number of chromosomes.
cohorts. Differences in disease severity and subphenotype frequency between the two populations could account for the observed discrepancy. From the limited data available we found that the presence of renal disease appears to be similar in both cohorts (UK SLE 36% compared with MN SLE 40%), while the gender ratios are significantly different (female: male UK SLE 11:1 compared with MN SLE 57:1, Chi square p value < 0.001). We were unable to compare other lupus subphenotypes. Furthermore, closer inspection of these data reveals that microsatellite-inference of HLA-DRB1 alleles in the MN SLE dataset may underestimate the frequency of HLA-DRB1*0301 and overestimate that of HLA-DRB1*1501, thus diminishing the effect of the former and enhancing that of the latter. It is also possible that the MN SLE cohort shows greater racial heterogeneity in comparison to our UK SLE cohort, despite both being characterized as Caucasian.

Previous studies have demonstrated increased risk for lupus in individuals carrying particular combinations of microsatellite-inferred HLA-DRB1-HLA-DQB1 haplotypes [25,28]. The highest risk genotype was found to be the compound heterozygote HLA-DRB1*0301-DQB1*0201/HLA-DRB1*1501-DQB1*0602, while HLA-DRB1*0301-DQB1*0201-containing genotypes demonstrated a dose-dependent effect in increasing lupus susceptibility [25,28]. In the present study, we have examined genotypic risk at the classically typed HLA-DRB1 locus and in contrast to the aforementioned data of Graham et al. [25,28] we have shown a likely dominant effect of the associated allele, HLA-DRB1*0301. The case-control and family-based analyses for HLA-DRB1 also show the greater power of the former to detect significant association (Table 2). Specifically, all genotypes containing HLA-DRB1*0301 show increased transmission to lupus probands; however, homozygotes show no greater risk compared with heterozygotes, as would be expected under additive or multiplicative models. Thus, a dominant model of inheritance, requiring the presence of a single copy of the disease-predisposing variant alone, likely underlies the susceptibility conferred by HLA-DRB1*0301 in UK SLE. Such a model would fit an antigen presentation hypothesis where susceptible individuals carrying an HLA-DRB1*0301 allele are able to present auto-antigens to CD4+ lymphocytes, thus stimulating an autoimmune response. The differences between our UK SLE and the previously published US SLE data may reflect disease, ethnic, and haplotypic heterogeneity.

Interestingly, analysis of genotypic risk at the associated class III marker, rs419788, suggests an additive (dose-dependent) pattern of inheritance for the rare T allele, where one copy confers a low risk of disease and two copies result in greater susceptibility. The different inheritance patterns for our class II and class III association signals provide further evidence for their independency.

A variety of HLA-DR and HLA-DQ alleles have been associated with autoantibody subsets in ethnically diverse populations of lupus. The strongest associations have been demonstrated between anti-Ro and anti-La antibodies and HLA-DRB3 and HLA-DQ2 (HLA-DQB1*0201), which are in strong LD [41–45] in case-control studies. Here, we confirm the association of HLA-DRB1*0301 with anti-Ro and anti-La antibody production in our family-based cohort.

Examination of LD structure around our second independent association, rs419788-T in class III, coupled with the results of our conditional analysis, suggests that this signal could also be delimited to a relatively narrow genomic interval of about 40 kb given further SNP mapping in our cohort. This region includes the genes CFB, RDBP, SKIV2L, DOM3Z, and STK19, but does not include complement C4. Thus, complement C4 null alleles, which have been implicated in lupus pathogenesis, may not be responsible for our class III signal. We conclude, therefore, that our family-based mapping study has potentially revealed a hitherto unknown lupus susceptibility interval in the class III region of the MHC. However, we cannot conclusively exclude association at complement C4IRCCX without direct determination of C4 polymorphism/copy number in our cohort.

With respect to the genes implicated in our study, CFB is a vital component of the alternate complement pathway and disregulation may clearly affect the inflammatory response [46]. RD and Skiv2L are proteins potentially involved in RNA processing. The RD protein forms part of a negative elongation factor (NELF) complex that represses RNA polymerase II transcript elongation, while Skiv2L is a DEAD box protein with possible function as an RNA helicase. The function of Dom3z is currently unknown, although the homologous yeast protein binds nuclear exoribonuclease. Moreover, its ubiquitous expression suggests a housekeeping role. STK19 is a protein kinase of unknown function with primary nuclear localization [47]. Interestingly, RDBP and SKIV2L are found to be highly expressed in T lymphocytes, B lymphocytes, and dendritic cells (SymAtlas, http://symatlas.gnf.org/SymAtlas/).

A number of studies have demonstrated conflicting evidence for and against association with various TNF locus polymorphisms in SLE [48]. A recent meta-analysis of the TNF-308GA promoter polymorphism in SLE [48] revealed evidence of association for the minor allele (A) in European populations; however, this study did not account for LD with class II alleles. On conditioning our dataset for HLA-DRB1*0301, we find that the TNF promoter signal is lost, suggesting that this association is not independent and is due to LD with HLA-DRB1*0301 (or another variant in LD with HLA-DRB1*0301).

In summary, we have found association with two distinct and independent variants within the class II (HLA-DRB1*0301) and class III (SKIV2L) regions of the MHC in UK SLE trios. We can delimit our class II signal in lupus to three genetic variants (HLA-DRB1*0301-HLA-DQA1*0501-HLA-DQB1*0201) that may confer disease risk in combination or as separate signals. Our class III signal importantly excludes independent association at the TNF promoter polymorphism TNF-308GA and potentially provides a novel locus for further study.

Materials and Methods

Study cohorts. SLE families. The cohort comprises 314 complete SLE trios (that is, mother, father, and affected lupus proband) collected as previously described [49]. All study participants are European Caucasian on the basis of grandparental origin. All 314 lupus probands (288 female, 26 male) fulfill the revised American College of Rheumatology (ACR) criteria for SLE [50], 36% of whom have a diagnosis of lupus nephritis. Written consent was obtained from all study participants and ethical approval for this study was obtained from the Multi-Centre Research Ethics Committee (MREC 2 June 1998).

Healthy controls. The control population for the HLA-DRB1 genotypic risk case-control analysis constitutes 1,667 healthy males of Northern European origin. The individuals are potential hematopoietic...
poietic stem cell donors and were typed to four digits for HLA-DRB1 at the Anthony Nolan Trust, UK for this purpose. The level of resolution used for the typing of HLA-DRB1*15 alleles in these healthy controls resulted in the ambiguous allele string HLA-DRB1*1501/1502/1504/1506. However, it is likely that the great majority of these typed alleles were HLA-DRB1*1501. There is no gender bias in HLA-DRB1 allele frequencies, so although we have used a male cohort, the frequencies would be expected to be the same in a similar female cohort (Steven Marsh, personal communication).

SNP genotyping. Eighty-six SNPs were chosen for genotyping in our including six each. Specifically, we selected 49 MHC class II and class III haplotypes using SNPs from a preliminary MHC SNP map [51] that had previously shown robust genotyping efficacy. In addition, we typed 36 MHC class II tag SNPs from a subsequent high-resolution MHC study [52]. We also included the TNF-308G/A promoter SNP, together with nine further SNPs in the region of HLA-B and MICA obtained from the database. 4883 SNPs were typed using PLINK projects/SNPs). The latter SNPs had not been well characterized. All variants were typed in the entire cohort (n = 942). The SNPs span approximately 2.4 Mb of the MHC from the class I gene, KIAA1949 to the class II pseudogene, HLA-DPB2 and thus encompass MHC class III and class II. SNP genotyping was performed at the Broad Institute of MIT and Harvard and at Imperial College London by matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry using the Sequenom MassARRAY platform as previously described. SNPs that failed Sequenom genotyping were retyped using KBiosciences (http://www.kbioscience.co.uk), a competitive allele-specific PCR technique. HLA-DRB1 genotyping. HLA-DRB1 typing was performed at the Anthony Nolan Trust, UK. All samples (n = 942 UK SLE trios and n = 1,067 healthy controls) were genotyped using L-Type SSO (sequence-specific oligonucleotide) typing technology according to the manufacturer’s written recommendations (http://www.onelambda.com). Briefly, a locus-specific biotinylated PCR amplicon is produced, denatured, and rehybridized to complementary oligonucleotide probes conjugated to fluorescein isothiocyanate (FITC) or Cy5 dye-labeled. The biotinylated PCR product can be detected using R-phycocerythrin-conjugated streptavidin. A flow analyzer identifies the fluorescent intensity of phycocerythrin on each bead. The assignment of HLA type is based on the reaction pattern compared to patterns associated with known sequences.

High resolution testing was performed using the Dynal AllSet’ SSP (sequence-specific primers) DRB1 assay according to the manufacturer’s protocol (Invitrogen) for SLE family only. The presence or absence of PCR amplification was detected in a gel electrophoresis step using visualization by ethidium bromide incorporation. Genotypes were determined using SSPIT software.

Samples that could not be resolved to four digits using SSO and PCR-SSP were analyzed by DNA sequencing of exon 2 of HLA-DRB1. Primers, reagents, and protocols were provided by The Anthony Nolan Trust, UK (primer sequences are available on request). Specific HLA-DRB1 alleles were assigned by comparing the resulting sequence with reference sequence from the IMGT/HLA Database [54].

Data analysis. Mendelian inconsistencies were removed using PedCheck [55]. Families in which more than eight markers demonstrated Mendel errors were removed from further analysis. Markers with less than 90% genotyping efficiency and markers where more than eight families showed Mendel errors were also excluded from analysis. Two markers located within the SLE associated class II region (rs2239802 in intron 4 of HLA-DRB and rs6457594 in the region between HLA-DRB9 and HLA-DRB5) show deviation from Hardy-Weinberg equilibrium (HWE), which may reflect an undetected SLE association or systematic genotyping error. HWE were assessed in parental samples in our cohort. There is currently no uniform opinion in the community regarding the inclusion or exclusion of SNPs that show deviation from HWE, hence we elected to include these markers in the final analysis.

Sixty-eight out of the total 86 SNPs passed our quality-control measures (see Table S1 for details). In summary, one SNP was monomorphic in our dataset, four SNPs yielded low genotyping efficiency, and 13 SNPs were excluded for unacceptable Mendel error rate. The mean call rate for all markers post-quality control was 94% (range 83% to 99%).

Family-based association testing was performed using Genehunter TDT (version 2.1) [56] and TRANSMIT (version 2.5) [57]. Haplotypes were constructed and permutation testing performed using Haploview v3.82 [58]. The significance level of all analyses was based on permutation testing (10,000 permutations). The data are reported both as nominal and permuted p values. ORs with 95% CI for family-based analyses were calculated in PLINK (http://pngu.mgh.harvard.edu/purcell/plink/) [59]. Conditional regression analyses were undertaken using WHAP [60].

The geno-PDT was performed using PDT version 5.1 with default settings [27]. The HLA-DRB1 alleles were coded into three groups for the geno-PDT and the case-control analysis: HLA-DRB1*0301, HLA-DRB1*1501 and HLA-DRB1 alleles other than HLA-DRB1*0301 or HLA-DRB1*1501. The HLA-DRB1*1501 code in the healthy controls represents the allele string HLA-DRB1*1501/1502/1504/1506, as described previously. The HLA-DRB1*1501 code in the lupus probands represents the alleles HLA-DRB1*1501/1504/1506 out of 942 (9301 and *1502 alleles) and HLA-DRB1*1502 (*1501 and *1502 alleles), as *1504 and *1506 were not present in this population. Fisher’s exact test was used to assess significance of association in the case-control analysis.

Subphenotype analysis. We looked for association of the HLA-DRB1*0301 allele with autoantibody subsets and renal disease in our cohort using the Chi-square test. We compared cases with and without the subphenotype of interest with DRB1*0301 homozygosity, heterozygosity, combined homozygosity and heterozygosity, and non-DRB1*0301 status. We performed the same analyses for homozygous and heterozygous genotypes of the associated SNP rs419788. The autoantibody subsets compared were anti-Dsg, anti-C1q, Igs, and IgM anti-cardiolipin antibodies (ACLg and ACLM), anti-Ro, anti-La, anti-RNP, anti-Sm, and anti-dsDNA.

Delineation of associated MHC haplotypes and evidence for positive selection. We looked for positively selected alleles in our study to provide the long-term selection signature. We first identified extended haplotype homozygosity (EHH), previously described by Sabeti et al. [26]. Essentially, such an analysis allows assessment of positive selection by mining datasets for high frequency extended haplotypes in comparison to the other core haplotypes at a locus.

EHH is defined as the probability that two randomly chosen chromosomes carrying the core haplotype of interest will be identical by descent (homozygosity at all SNPs) for the entire interval from the core to a distance r. The REHH is the ratio of the EHH on the tested core haplotype compared with the combined EHH of all the other core haplotypes at the region excluding the tested core; as such, REHH accounts for local variation in recombination rate while EHH does not [26].

The program emphase was employed to assign the phase of parental genotypic data and reconstruct missing information. Emphase is a simple phaser similar to the phaser of Excoffier and Slatkin [61]. It is very fast, especially on large datasets, and sufficiently accurate for most genetic applications. EHH analysis was performed on the phased parental data using the software program SWEEP (http://www.broad.mit.edu/mpg/sweep/index.html).

Haploype bisation plots. We represent the breakdown of LD on core haplotypes using haplotype bisation diagrams generated in the program TRE [62] (also explained in [52]).

REHH versus frequency plots. Fifty-three SNPs (identified in Figure 1) are common to our dataset and the CEPH and Yoruba HapMap [65] populations. These three datasets, together with CEPH SNP data for its Chromosome 6 (http://ftp.1000genomes.ebi.ac.uk) and the CEPH allele frequency data for Chromosome 6, were used to generate separate REHH versus frequency plots in SWEEP. The plots from the four cohorts were combined for visual, not statistical, comparison. Evidence for positive selection was quantitatively assessed in CEPH individuals, as this population alone was used to assess background variation on Chromosome 6. The UK SLE and Yoruba cohort data are shown for comparison. The 95th percentile based on total CEPH Chromosome 6 SNP data is shown.

Supporting Information

Figure S1. OR Plot for Associated MHC Markers in 314 UK SLE Trios OR with 95% CI and marker name in genomic order (class I to class II, left to right) are indicated on the vertical and horizontal axes respectively. The greatest ORs are seen in the class III and class II regions.

Supporting Information

Figure S2. Structure of All Transmitted and Untransmitted HLA-DRB1*0301 and rs419788-T Allele Haplotypes Haploype bisation plots constructed for all (A) transmitted (T) HLA-DRB1*0301 haplotypes, (B) untransmitted (UT) HLA-DRB1*0301 haplotypes, (C) transmitted HLA-DRB1*0301 alleles, and (D) untransmitted rs419788-T allele haplotypes. The allelic composition of the most common haplotype in each subset is shown: the core allele is represented as a dark blue double bar indicating
haplotypes to the right and to the left of the core; otherwise, the common haplotype is depicted by dark grey bars. In parts (A) and (B), the rs149788-T allele in class III, which shows association independent of HLA-DRB1*0301 in our cohort, is indicated in green, while in parts (C) and (D), the allele HLA-DRB1*0301 is shown in dark grey. The key difference between HLA-DRB1*0301 T and UT haplotypes lies within the class II region of the MHC. All HLA-DRB1*0301 T haplotypes are identical across a 180 kb region defined by eight SNPs (light blue), whereas the corresponding region within UT HLA-DRB1*0301 haplotypes exhibits significant recombination. This conserved class II interval encompasses only three expressed genes: HLA-DRB1, HLA-DQA1, and HLA-DQB1. Given the strong LD exhibited by HLA-DRB1*0301 haplotypes, the allelic composition of this risk region is known to be HLA-DRB1*0301-HLA-DQA1*0501-HLA-DQB1*0201. Both T and UT rs149788-T allele haplotypes show similar structure overall. The rs149788-T allele is clearly present on HLA-DRB1*0301 and non-HLA-DRB1*0301-containing haplotypes, lending credence to our observation that rs149788-T or another variant in LD with it constitutes an association signal independent of HLA-DRB1*0301 in our UK SLE cohort.

(A) total of 176 parental chromosomes, (B) total of 90 parental chromosomes, (C) total of 284 parental chromosomes, and (D) total of 182 parental chromosomes.

Table S1. Quality Control Information for Four-Digit HLA-DRB1 Typing and All 86 SNPs Chosen for Genotyping in a Cohort of 314 UK SLE Trios

Table S2. OR and TDT Data for Four-Digit HLA-DRB1 Typing in a Cohort of 314 UK SLE Trios

Table S3. Complete Single Marker (SNP) and HLA-DRB1*0301 Association Data in a Cohort of 314 UK SLE Trios

Table S4. Association of HLA-DRB1*0301 Genotypes with SLE Subphenotypes in 314 UK SLE Probands

Acknowledgments

We would like to thank Ann-Margaret Little and Steven Marsh from the Anthony Nolan Trust for their assistance, expertise, and provision of control HLA-DRB1*0301 genotyping data. We are grateful to Hazel Maldonado Torres for collating the control HLA-DRB1 data. Our thanks and appreciation is extended to all participants and family members for participation in this study, as well as to Andrew Wong for sample collection and maintenance of our study database. We are grateful to the many clinicians and nursing staff who also contributed to sample collection. We are indebted to all our collaborators for providing study participants.

Author contributions. JDR and TJV conceived the study design. ECW and MMAF performed SNP genotyping. AS performed HLA-DRB1 typing and contributed to the Materials and Methods section of the manuscript. AJMM provided advice, protocols, and reagents to MMAF for HLA-DRB1 sequencing. CRS and PCS performed, analyzed and interpreted EHH analysis and haplotype bifurcation plots. MMAF, ECW, TG, and JDR contributed to data analysis. MMAF, TJV, and JDR wrote the manuscript with contributions from PCS, CRS, ECW, and AS.

Funding. This study was funded through a Wellcome Trust Senior Fellowship for TJV, an Arthritis Research Campaign Clinical Research Fellowship for MMAF, and a Cancer Research Institute Fellowship for ECW. The Broad Institute Center for Genotyping and Analysis is supported by grant U54 RR020278–01 from the National Center for Research Resources, JDR is funded by grants from the National Institutes of Allergy and Infectious Diseases (AI065867; A1067152) and from the National Institute of Diabetes and Digestive and Kidney Diseases (DK064869; DK002432).

Competing interests. The authors have declared that no competing interests exist.

References


Accession Numbers

The Online Mendelian Inheritance in Man (OMIM, http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim) accession numbers for the genes described in this study are as follows: AGER, 609214; BAT3, 142590; C4A, 120810; C4B, 120829; CFB, 138470; DOM3Z, 605996; DOM3Z, 605996; EHMT2, 604599; HLA-A, 142850; HLA-DRB2, 120920; HLA-DQA1, 146880; HLA-DQB1, 604305; HLA-DRA, 142860; HLA-DRB1, 142850; HLA-DRB1, 191635; HLA-DRB1, 142860; HLA-DRB1, 142850; KIAA1949, 610990; MICA, 600169; NOTCH4, 164951; RDBP, 134049; RDBP, 154409; SKIV2L, 60478; SLC4A4A, 606107; STK19, 604977; TNF, 191160; and TNXB, 600985.

Competing interests. The authors have declared that no competing interests exist.
number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE); low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. Am J Hum Genet 80: 1037–1054.


