EXTENDED REPORT

Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at MSH5, HLA-DPB1 and HLA-G

Michelle M A Fernando,¹ Jan Freudenberg,² Annette Lee,² David Lester Morris,¹ Lora Boteva,¹ Benjamin Rhodes,¹ Maria Francisca Gonzalez-Escrivano,³ Miguel Angel Lopez-Nevot,⁴ Sandra V Navarra,⁵ Peter K Gregersen,² Javier Martin,⁶ IMAGEN,* Timothy J Vyse¹

ABSTRACT

Objectives Systemic lupus erythematosus (SLE) is a chronic multisystem genetically complex autoimmune disease characterised by the production of autoantibodies to nuclear and cellular antigens, tissue inflammation and organ damage. Genome-wide association studies have shown that variants within the major histocompatibility complex (MHC) region on chromosome 6 confer the greatest genetic risk for SLE in European and Chinese populations. However, the causal variants remain elusive due to tight linkage disequilibrium across disease-associated MHC haplotypes, the highly polymorphic nature of many MHC genes and the heterogeneity of the SLE phenotype.

Methods A high-density case-control single nucleotide polymorphism (SNP) study of the MHC region was undertaken in SLE cohorts of Spanish and Filipino ancestry using a custom illumina chip in order to fine-map association signals in these haplotypically diverse populations. In addition, comparative analyses were performed between these two datasets and a northern European UK SLE cohort. A total of 1433 cases and 1458 matched controls were examined.

Results Using this transancestral SNP mapping approach, novel independent loci were identified within the MHC region in UK, Spanish and Filipino patients with SLE with some evidence of interaction. These loci include HLA-DPB1, HLA-G and MSH5 which are independent of each other and HLA-DRB1 alleles. Furthermore, the established SLE-associated HLA-DRB1*15 signal was refined to an interval encompassing HLA-DRB1 and HLA-DQA1. Increased frequencies of MHC region risk alleles and haplotypes were found in the Filipino population compared with Europeans, suggesting that the greater disease burden in non-European SLE may be due in part to this phenomenon.

Conclusion These data highlight the usefulness of mapping disease susceptibility loci using a transancestral approach, particularly in a region as complex as the MHC, and offer a springboard for further fine-mapping, resequencing and transcriptomic analysis.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic multisystem autoimmune disease characterised by the production of autoantibodies to nuclear and cellular antigens, tissue inflammation and organ damage. There is a strong but complex genetic component to SLE susceptibility, whereby many polymorphisms each with a small or modest effect contribute to disease susceptibility. Genome-wide association studies have shown that variants within the major histocompatibility complex (MHC) region on chromosome 6 confer the greatest genetic risk for SLE in European and Chinese populations.¹-³ The extended MHC spans almost 8 Mb and is divided into five subregions: extended class I (telomeric), class I, class III, class II and extended class II (centromeric). One of the most complex regions of the genome, this locus harbours two copy variable regions (the HLA-DRB genes in class II and the RCCX module containing complement component C4 in class III), some of the most polymorphic genes in the genome and conserved haplotypes where linkage disequilibrium (LD) extends over 2 Mb in some instances. The region has been the subject of extensive study given the importance of MHC alleles in the pathogenesis of tissue incompatibility, drug sensitivity, autoimmune, infectious and inflammatory diseases.

In European SLE cohorts, well-established associations are observed with highly conserved and extended haplotypes bearing the class II alleles HLA-DRB1*03:01 and HLA-DRB1*15:01.⁴ More recent high-density single nucleotide polymorphism (SNP) genotyping studies have demonstrated multiple independent signals across the MHC in northern European cohorts.⁵ ⁶ However, the causal variants remain elusive due to tight LD across disease-associated MHC haplotypes, the highly polymorphic nature of associated variants and the heterogeneity of the SLE phenotype. Fine-mapping studies across the MHC region in other European and non-European SLE populations are lacking. The haplotypic diversity consequent on differing ancestry and environment demonstrated by these populations at the MHC region as well as non-MHC loci should allow further refinement of known association signals together with the identification of novel susceptibility variants. Given these known haplotypic differences, we undertook a high-density case-control SNP study of the MHC region in SLE cohorts of Spanish and...
Filipino ancestry using a custom Illumina chip in order to fine-map established association signals and potentially uncover novel susceptibility loci. In addition, we have performed comparative analyses between these two datasets and a northern European UK SLE cohort. In total we examined 1433 cases and 1458 matched controls.

**METHODS**

**Spanish cohort**

The cohort comprised 464 cases and 468 controls. All cases were recruited from rheumatology clinics throughout Spain. Control samples were obtained from the Blood Bank Units of the hospitals where the cases originated.

**Filipino cohort**

The cohort comprised 335 SLE probands and 247 unrelated controls. We also included 26 trios (father, mother and affected child) to allow checks for Mendelian inheritance. All probands attended the Rheumatology and Clinical Immunology Clinics at the University of Santo Tomas Hospital, Manila, Philippines. Unrelated controls were recruited from spouses and acquaintances of the probands.

**UK cohort**

The cohort comprised 632 SLE probands and 742 unrelated controls from a previous study. All SLE probands fulfilled the American College of Rheumatology criteria for the classification of SLE. Written consent was obtained from all study participants.

**Sample collection**

DNA was obtained from whole blood using phenol-chloroform extraction. Native genomic DNA was used for the Spanish study. For the Filipino cohort, 100 ng (5 μl at 20 ng/μl) native DNA was whole genome amplified using the Qiagen REPLI-g Midi Kit (Cat. No. 150045) according to the manufacturer’s written instructions.

**Custom iSelect Illumina SNP array and genotyping**

The samples were genotyped at the Feinstein Institute, USA using a custom Illumina iSelect chip comprising 10 788 SNPs: 6045 SNPs within the MHC region (29–33.5 Mb) and 4743 SNPs informative for major and European ancestry (see online supplement for SNP selection criteria). In addition, we have performed comparative analyses between these two datasets and a northern European UK SLE cohort. In total we examined 1433 cases and 1458 matched controls.

**Figure 1** High-density transancestral single nucleotide polymorphism (SNP) mapping of the major histocompatibility complex (MHC) region in UK, Spanish and Filipino systemic lupus erythematosus (SLE). The panels show MHC region association plots for (A) UK, (B) Spanish and (C) Filipino SLE cohorts where genomic position (Mb) is shown on the horizontal axis with –log10 p values on the vertical axis. The black squares represent genotyped SNPs and the blue squares indicate imputed SNPs in the UK cohort. The red squares represent classically typed HLA alleles in the UK (HLA-DRB1 only) and Spanish (HLA-B, HLA-DRB1 and HLA-DQB1) cohorts. The panels beneath each association plot demonstrate the recombination rate for each cohort calculated using control haplotypes only generated with the program rhomap. A scaled map of the MHC region with relevant genes is shown in the bottom panel. RCCX represents the copy variable RCCX module containing the complement C4 gene (R=RP1/STK19, C=C4A/C4B, C=CYP21A2/CYP21A1P, X=TNXA/TNXB).
HLA genotyping
All HLA typing was performed using Luminex One Lambda SSO. Four-digit genotyping for HLA-B, HLA-DRB1 and HLA-DQB1 was performed in 82%, 99% and 44% of the Spanish cohort, respectively, following quality control (QC) at Hospital Virgen del Rocio, Seville, Spain and Hospital Virgen de las Nieves, Granada, Spain.

In order to assess LD relationships, four-digit HLA-DRB1 typing was performed in a subset of the Filipino cohort of known genotype for the top SNP, rs9274366, where DNA was available (n=89). Four-digit HLA-DRB1 typing was performed in 606 of 632 UK cases of SLE (96%). The Filipino and UK typing was performed at the Anthony Nolan Trust, London, UK. Two-digit HLA-DRB1 data were obtained for 694 of the 742 UK controls (92%) from the 1958 British birth cohort.

Quality control (QC) filters
All QC analyses except principal components analyses were performed using PLINK.10 Samples and SNPs were put forward for analysis if they met the following quality control filters: SNPs greater than 95% genotyping efficiency, minor allele frequency (MAF) >1% (failed Spanish n=271, Filipino n=758), non-deviation from Hardy-Weinberg equilibrium in controls on the basis of a false discovery rate of 0.05 (failed Spanish n=61, Filipino n=21). SNPs were excluded if they showed >10% Mendel error rate in the post-QC Filipino trios (n=7). Samples required >95% genotyping efficiency, minor allele frequency (failed Spanish n=50, Filipino n=50) and PI-Hat scores >0.2 on identity-by-descent analysis using ancestry informative markers (AIMs) in order to exclude cryptic relatedness and duplicate samples (failed Spanish n=27, Filipino n=3). In order to correct for population stratification, samples were excluded if they were outliers on principal components analysis using post-QC AimSNPs (performed using EIGENSTRAT and defined as >4 SDs from the mean)11 (failed Spanish n=62, Filipino n=68). The genomic inflation factor ($\lambda_{GC}$) was calculated using the post-QC AIMs after correction for population stratification (Spanish $\lambda_{GC}=1.04$ and Filipino $\lambda_{GC}=1.09$).

UK SLE cohort imputation
Genotypes were imputed using IMPUTE12 on the initial set of directly genotyped SNPs (n=1230) up to the Wellcome Trust Case-Control Consortium 2 (WTCCC2) study (n=7119). The WTCCC2 data were used as reference genotypes in the imputation, with dbSNP build 126 defining the genome map.13 No reference haplotypes were used in the imputation. Of the 7119 imputed SNPs in the UK SLE cohort, 3314 overlapped with the 6045 MHC SNPs genotyped in the Spanish and Filipino cohorts in this study and were used for analysis.

Statistical analyses
Single marker association analyses using logistic regression and stepwise logistic regression analyses were performed using PLINK and SNPTEST.14 We took the genotypes for the most associated SNP as a covariate and conditioned on this in the search for other independently associated SNPs in each dataset. If this analysis yielded further SNPs that passed our threshold of significance (see below), we added the top SNP to further stepwise logistic regression models and continued the process until no further SNPs passed our threshold of significance. Haplotypic association analyses were performed using PLINK and R statistical package. Data for SNP rs409558 in the Spanish and Filipino cohorts were meta-analysed using

Table 1 Primary and secondary single marker association in the major histocompatibility complex region in UK, Spanish and Filipino systemic lupus erythematosus

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Position</th>
<th>F_U</th>
<th>Associated allele</th>
<th>OR (95% CI)</th>
<th>p</th>
<th>Other disease</th>
<th>Location (LD)*</th>
<th>Gene expression†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UK SLE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1269852</td>
<td>32188169</td>
<td>0.11</td>
<td>C</td>
<td>2.68 (2.16 to 3.33)</td>
<td>2.48×10⁻¹⁹</td>
<td>TNXB-ATF6B (760 kb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3906272</td>
<td>31370003</td>
<td>0.04</td>
<td>A</td>
<td>2.84 (2.05 to 3.92)</td>
<td>2.95×10⁻¹⁰</td>
<td>HLA-C-HLA-B (22 kb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3129868</td>
<td>32512355</td>
<td>0.11</td>
<td>T</td>
<td>1.67 (1.29 to 2.14)</td>
<td>8.65×10⁻⁵</td>
<td>BTNL2 – DRA (1501 tag SNP)</td>
<td>(375 kb)</td>
<td></td>
</tr>
<tr>
<td>rs47713419</td>
<td>31101215</td>
<td>0.10</td>
<td>C</td>
<td>0.40 (0.25 to 0.63)</td>
<td>7.55×10⁻⁵</td>
<td>MUC21 - PSORS1C1 (9 kb)</td>
<td>HLA-Ct</td>
<td></td>
</tr>
<tr>
<td><strong>Spanish SLE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3130480</td>
<td>31847099</td>
<td>0.04</td>
<td>A</td>
<td>2.96 (1.95 to 4.51)</td>
<td>3.94×10⁻³</td>
<td>MSH5 (20 kb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3129768</td>
<td>32700361</td>
<td>0.13</td>
<td>C</td>
<td>1.91 (1.44 to 2.53)</td>
<td>7.57×10⁻⁶</td>
<td>CBorf27 (820 kb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3117213</td>
<td>33172583</td>
<td>0.18</td>
<td>A</td>
<td>1.76 (1.37 to 2.25)</td>
<td>7.18×10⁻⁶</td>
<td>DBP1-DBP2 (29 kb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Filipino SLE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs409558</td>
<td>31816126</td>
<td>0.20</td>
<td>G</td>
<td>0.57 (0.43 to 0.77)</td>
<td>2.2×10⁻⁴</td>
<td>MSH5 (20 kb)</td>
<td></td>
<td>MSH5‡</td>
</tr>
<tr>
<td>rs9271386</td>
<td>32694832</td>
<td>0.34</td>
<td>G</td>
<td>2.46 (1.83 to 3.30)</td>
<td>1.97×10⁻⁸</td>
<td>DRB1-DAQ1 (1502 tag SNP)</td>
<td>(87 kb)</td>
<td></td>
</tr>
<tr>
<td>rs2571391</td>
<td>30031817</td>
<td>0.16</td>
<td>C</td>
<td>0.36 (0.22 to 0.59)</td>
<td>6.06×10⁻⁵</td>
<td>MSH5 (300 kb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2507987</td>
<td>31450102</td>
<td>0.48</td>
<td>A</td>
<td>0.48 (0.35 to 0.66)</td>
<td>6.80×10⁻⁴</td>
<td>HLA-G-HLA-H (300 kb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs409558</td>
<td>31816126</td>
<td>0.30</td>
<td>G</td>
<td>0.56 (0.37 to 0.86)</td>
<td>7.47×10⁻³</td>
<td>MSH5 (100 kb)</td>
<td></td>
<td>MSH5‡</td>
</tr>
<tr>
<td>rs2071345</td>
<td>33151908</td>
<td>0.17</td>
<td>A</td>
<td>0.48 (0.28 to 0.83)</td>
<td>8.44×10⁻³</td>
<td>DBP1 (19 kb)</td>
<td></td>
<td>HLA-DBP1†</td>
</tr>
</tbody>
</table>

*LD surrounding each SNP calculated in control population of each cohort using r² cut-off >0.8.

T1D, type 1 diabetes; SSc, systemic sclerosis; PBC, primary biliary cirrhosis; SLE, systemic lupus erythematosus; SNP, single nucleotide polymorphism; SSc, systemic sclerosis; T1D, type 1 diabetes.

Meta-analysis of Spanish and Filipino data for the MSH5 SNP rs409558 revealed a locus-wide level of significance at p = 1.92×10⁻⁸.


ACPA + RA, anti-citrullinated protein antibody positive rheumatoid arthritis; CBD, chronic beryllium disease; F_U, frequency in unaffected controls; Hep B, hepatitis B; LD, linkage disequilibrium; MS, multiple sclerosis; p, corrected/conditional p value (see Methods); PBC, primary biliary cirrhosis; SLE, systemic lupus erythematosus; SNP, single nucleotide polymorphism; SSc, systemic sclerosis; T1D, type 1 diabetes.
the standard inverse variance method. We performed tests of heterogeneity using the Breslow-Day test in PLINK. p values are represented following adjustment for the first principal component in each dataset or following adjustment for the first principal component and additional SNPs as covariates in SLR analyses in each dataset. A significance threshold of \( p=7\times10^{-8} \) was set, given that genome-wide significance thresholds based on haplotype structure are typically in the range of \( 5\sim7\times10^{-8} \) and that the MHC region constitutes approximately 1/1000th of the genome. The LD structure of the MHC region has been shown to be similar to that of the genome in general, but there appears to be greater LD between haplotype blocks in the MHC region so our significance threshold is likely to be conservative. In the Spanish cohort, separate logistic regression and conditional logistic regression analyses were performed for HLA-DRB1 alleles in order to assess relative predispositional effects. In order to account for multiple testing, Bonferroni-corrected p values were used as follows: HLA-DRB1, \( p=0.0023 \) (0.05/22 alleles tested). We examined LD relationships between SNPs and HLA alleles in each cohort by calculating the correlation coefficient (\( r^2 \)) using the Tagger algorithm in Haploview.\(^{15}\)

**RESULTS**

In all three datasets under study there was significant SNP association across the entire MHC region (figure 1). The UK SLE data confirmed previously published reports in northern European cohorts demonstrating principal SNP association within the class II and class III regions of the MHC.\(^5\)\(^6\)\(^16\) The most significantly associated SNP was rs1269852, located intergenic TNXB-ATF6B in the class III region of the MHC. Stepwise logistic regression demonstrated independent association at additional MHC loci including SNPs tagging the HLA-DRB1*1501 haplotype in class II, as well as class I SNPs located between HLA-B and HLA-C and 5′ PSORS1C1 (table 1).

**Association of MHC class II and class III variants with Spanish SLE**

In the Spanish cohort, 399 cases and 394 controls were put forward for analysis following QC measures. Logistic regression analysis of 4924 post-QC SNPs showed that the peak signals in this southern European SLE cohort also arise from the class II and class III regions of the MHC (figure 1, table 1 and table S1 in online supplement). The most significantly associated SNP, rs9268382, was located in the class II pseudo-gene HLA-DRB9 (OR 1.80, CI 1.45 to 2.23, \( p=7.64\times10^{-8} \)) and showed moderate/weak LD with HLA-DRB1 alleles (figure S1 in online supplement). Serial stepwise logistic regression revealed a number of independent signals around HLA-DBP4 (best SNP rs3130490; best protective SNP rs409558). Interestingly, the SNPs with the best OR in this Spanish dataset were the aforementioned variants in and around the class III genes MSH5/C6orf27. The most associated of these SNPs was rs3103490 (OR 3.08, CI 2.03 to 4.66, \( p=1.04\times10^{-7} \)) (table 1, figure 2 and figure S2 in online supplement). This SNP showed strong LD with the top UK MHC SNP rs1269852 (\( r^2=0.97 \)). Thus, the primary MHC signal in Spanish SLE replicated that observed in the previously published UK dataset.\(^5\) In contrast to the UK data where the SNP rs3103490 showed strong LD with HLA-DRB1*03:01 (\( r^2=0.71 \)), the Spanish signal showed only moderate LD with HLA-DRB1*03:01 (\( r^2=0.23 \)), suggesting that variants in the class III region of the MHC may play a more important role than previously recognised. Conditioning on rs3103490 also revealed a number of potentially independent signals in the Spanish cohort, the best of which was the class II SNP rs129768 located between HLA-DRB1 and HLA-DQA1 (OR 1.91, CI 1.44 to 2.53, \( p=7.57\times10^{-8} \)). This SNP showed moderate LD with HLA-DRB1*15:01 (\( r^2=0.62 \)). Again this contrasts with our northern European data where one of the main
secondary association signals was observed with variants in strong LD with HLA-DRB1*15:01 ($r^2=0.95$). Further stepwise logistic regression revealed association with the previously mentioned SNPs rs3117213 (HLA-DRB1) and rs409558 (MSH5) (table 1). LD analysis revealed that the association underlying rs9268832 probably represents a composite effect of rs3103490 and rs3129768, resulting in its greater statistical significance (figure S1 in online supplement).

**MHC region SNPs show association independent of HLA-DRB1 alleles in Spanish SLE**

Analysis of HLA-DRB1 alleles alone demonstrated principal association with HLA-DRB1*03:01 (OR 1.89, CI 1.43 to 2.48, $p=5.53 \times 10^{-6}$) (table S2 in online supplement). Conditioning on HLA-DRB1*03:01 in order to assess relative predispositional effects, we found association with HLA-DRB1*15:01 (OR 1.83, CI 1.31 to 2.55; $p=0.00045$). Conditioning on these top two HLA-DRB1 alleles, we found association with HLA-DRB1*08:01 (OR 3.52, CI 1.55 to 8.01; $p=0.0027$). No other HLA-DRB1 alleles showed significant disease association following further stepwise logistic regression. Next we used the three principally associated HLA-DRB1 alleles as covariates in a serial stepwise logistic regression in the entire SNP dataset. We found that all the aforementioned SNPs showed some evidence of association independent of HLA-DRB1 alleles except rs3129768 (table S3 in online supplement). Similar results were obtained when conditioning the UK SNP data for HLA-DRB1*03:01 and HLA-DRB1*15:01 (table S4 in online supplement).

**Major role for MHC class II and class I variants in Filipino SLE**

Following QC measures, 275 cases and 166 controls were put forward for analysis in the Filipino SLE cohort. The overall pattern of association showed that, of the 3704 post-QC SNPs, the major signal arises from the class II region of the MHC and therefore differs from that observed in European SLE cohorts where principal associations are seen in class II and class III (figure 1, table 1 and table S5 in online supplement). The top SNP, rs9271366, was located between HLA-DRB1 and HLA-DQA1 (OR 2.46, CI 1.83 to 3.30, $p=1.97 \times 10^{-9}$). Furthermore, HLA-DRB1 typing in a subset of this cohort showed that the most highly associated SNP, rs9271366, was a perfect proxy for HLA-DRB1*15:02 in the Filipino population ($r^2=1$) and suggests a major role for variants on this haplotype in Filipino SLE (table S6 in online supplement). These data are consistent with the known high allele frequency of HLA-DRB1*15:02 in Filipino reference and other Pacific rim populations where the allele frequency ranges from 37% to 48%. Furthermore, the association of HLA-DRB1*15:01 is well established in East Asian SLE cohorts from Japan and Korea, while the association of HLA-DRB1*15:02 with SLE has been reported in South East Asians from Thailand.

SLR analyses on the top SNP (rs9271366) revealed independent signals in the class I region of the MHC between HLA-G and HLA-A (best SNP rs2571391: OR 0.56, CI 0.22 to 0.59, $p=6.06 \times 10^{-5}$). Further stepwise logistic regression revealed additional independent signals that replicate those observed in the Spanish cohort: MSH5 (best SNP rs409558) and HLA-DRB1 (best SNP rs2071351) (table 1 and figure S5 in online supplement). Meta-analysis of Spanish and Filipino data for the MSH5 SNP rs409558 revealed a locus-wide significance level at $p=1.92 \times 10^{-5}$ (OR$_{meta}$ 0.58, CI$_{meta}$ 0.33 to 0.85).

**Effect size of SNP rs9271366 is significantly greater in Filipino SLE than in European SLE**

The most associated Filipino MHC SNP, rs9271366, which acts as a surrogate marker for HLA-DRB1*15:02 in this population, tags HLA-DRB1*15:01 in populations of European ancestry ($r^2=0.94$ (UK controls); $r^2=0.77$ (Spanish controls)) and, as such, shows disease association in European SLE with ORs of approximately 1.4 (table S7 in online supplement). The frequency of HLA-DRB1*15:02 is low in European populations (1–2%). The SNP rs9271366 also tags the SNP's demonstrating disease association following primary conditional analysis in the UK ($r^2$ with rs3129668=0.98) and Spanish ($r^2$ with rs3129768=0.77) cohorts because these SNPs (rs312968 and rs3129768) are also in LD with HLA-DRB1*15:01 (table 1). As the effect size of the SNP rs9271366 is significantly greater in the Filipino SLE cohort than in the Europeans (Filipino OR 2.46, European OR ~1.4, Breslow-Day $p=5.25 \times 10^{-4}$, table S7 in online supplement), it is interesting to speculate that this genetic variant or variants in LD may predispose to a more severe disease.
phenotype such as renal involvement, as is often observed in non-European populations.25–24

Refinement of SLE-associated HLA-DRB1*15 signal

Previous attempts to fine-map the SLE-associated HLA-DRB1*15 haplotype signal could only delimit the region to approximately 500 kb of the MHC class II region in Europeans-Americans using microsatellite typing.25 When the SNP rs9271366 was used as a surrogate marker for HLA-DRB1*15 haplotypes, we found that the LD surrounding this SNP varied in the different populations studied from a 375 kb region in UK SLE to a 182 kb region in Spanish SLE and to an 87 kb region encompassing HLA-DRB1 and the intergenic interval between HLA-DRB1 and HLA-DQA1 in Filipino SLE (figure 3). Hence, the transancestral mapping approach used in this study allowed refinement of the SLE-associated HLA-DRB1*15 signal.

Evidence for genetic interaction between the top independent MHC region SNPs in SLE

Haplotype analyses were performed on the top three independent SNPs from each cohort to look for evidence of interaction. The top SNP was chosen from each cohort, together with the next two independently associated SNPs which were selected following stepwise logistic regression (table 2). Despite the relatively modest cohort sizes, these analyses suggested evidence of genetic interaction (non-additive effects) in all three populations studied (table S8 in online supplement). For example, in the Spanish cohort, a multiple logistic regression model was fitted using the top three independent SNPs (rs130490, rs3129768 and rs3117213) as explanatory variables and interaction terms were tested for. Interestingly, the best model (difference in Akaike Information Criterion=4.8, difference in Bayesian Information Criterion=6.8) had an rs130490*rs3129768 interaction where the effect on the OR was positive (5.1 (95% CI 1.12 to 23.08), p=0.03), plus an independent additive term for rs3117213 (see table S9 and figure S4 in online supplement).

Increased frequency of MHC region risk alleles and haplotypes in Filipino SLE compared with European SLE

Next we examined haplotype frequency and association using the top three independent SNPs in all three cohorts studied (table 2). We found that the haplotype harbouring the risk alleles of the top three SNPs was rare in European SLE cohorts. However, in Filipino SLE, the risk haplotype was common while the protective haplotype was rare (risk OR 3.45, CI 2.24 to 5.33, p=5.69×10−11; protective OR 0.002, CI 1×10−4 to 0.05, p=2.10×10−4). Thus, the population frequency of the top ranked risk alleles and risk haplotypes increases from the UK and Spain to the Philippines (risk haplotype frequency cases 0, 0.004 and 0.278, respectively), suggesting that the greater disease burden in non-European SLE populations may be due in part to this phenomenon. Furthermore, the frequency of protective haplotypes in each population decreases through the same gradient (protective haplotype frequency cases 0.525, 0.505 and 0.005, respectively).

### Table 2: Haplotypic association using the top three SNPs following serial step-wise logistic regression in each cohort is shown (see table 1 and table S8 in the online supplement for further details).

<table>
<thead>
<tr>
<th>SNP1</th>
<th>SNP2</th>
<th>SNP3</th>
<th>HAPLOTYP</th>
<th>F_A</th>
<th>F_U</th>
<th>F_Uexp</th>
<th>OR</th>
<th>95% CI</th>
<th>p</th>
<th>p_perm</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKrisk</td>
<td>rs1269852</td>
<td>rs3906272</td>
<td>rs3129688</td>
<td>CAT</td>
<td>&lt;0.001</td>
<td>0.0004</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>UKrisk</td>
<td>rs1269852</td>
<td>rs3906272</td>
<td>rs3129688</td>
<td>GGG</td>
<td>0.524</td>
<td>0.729</td>
<td>0.7604</td>
<td>0.37</td>
<td>0.31 to 0.44</td>
<td>3.72×10−28</td>
</tr>
<tr>
<td>Spanish</td>
<td>rs3130490</td>
<td>rs3129768</td>
<td>rs3117213</td>
<td>ACA</td>
<td>0.004</td>
<td>&lt;0.001</td>
<td>0.001</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Spanish</td>
<td>rs3130490</td>
<td>rs3129768</td>
<td>rs3117213</td>
<td>CAC</td>
<td>0.506</td>
<td>0.688</td>
<td>0.6849</td>
<td>0.443</td>
<td>0.36 to 0.55</td>
<td>2.22×10−13</td>
</tr>
<tr>
<td>Filipino</td>
<td>rs9271366</td>
<td>rs2571391</td>
<td>rs2507987</td>
<td>GAT</td>
<td>0.273</td>
<td>0.092</td>
<td>0.1485</td>
<td>3.45</td>
<td>2.24 to 5.33</td>
<td>5.69×10−11</td>
</tr>
<tr>
<td>Filipino</td>
<td>rs9271366</td>
<td>rs2571391</td>
<td>rs2507987</td>
<td>ACA</td>
<td>0.005</td>
<td>0.031</td>
<td>0.0507</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

F_A, haplotype frequency in affected SLE cases; F_U, haplotype frequency in unaffected controls; F_Uexp, expected haplotype frequency calculated from minor allele frequency of SNPs in controls; NA, not applicable (unable to perform statistical analyses as frequencies too low); p, p value; p_perm, permuted p value (10000 permutations); prot, haplotype comprising all three protective alleles; risk, haplotype comprising all three risk alleles; SLE, systemic lupus erythematosus; SNP, single nucleotide polymorphism.

DISCUSSION

We present the results of the first high-density transancestral mapping study of the MHC region in SLE using cohorts from the UK, Spain and the Philippines. Despite the modest sample sizes, we have identified and replicated new independent loci with evidence of interaction across this complex region, some of which appear to be SLE-specific (MSH5) while others suggest shared mechanisms across autoimmune/inflammatory diseases (HLA-DPB1, HLA-C, HLA-B/C) (box 1). In particular, we were able to demonstrate a considerable effect from MHC variants in Filipino SLE using single marker and haplotype analyses due to the high frequency of disease-associated variants in this cohort. There are no accurate prevalence data for SLE in the Philippines and most parts of Asia, even in the most recent literature. In general, published prevalence rates for SLE in Asia are broadly similar to those observed in Europeans and range between 30 and 50 per 100 000.22 24 26 Interestingly, prevalence rates appear to be higher in Asian migrant populations.24

The primary class III risk haplotype tagged by rs1269852 and rs310490 in Europeans displays extended LD such that it encompasses most of the MHC class III region. Stepwise logistic regression analyses demonstrated an identical protective haplotype encompassing the class III gene MSH5 alone in Filipino and Spanish SLE. These data suggest that dysregulation of MSH5 may underlie some of the risk attributable to the conserved class III risk haplotype. Further fine-mapping will be required to elucidate the nature of this signal. The class III variants that confer the greatest risk in SLE cohorts of European ancestry are rare in Filipino and Han Chinese SLE where MAF are approximately 0.001. These data suggest that either these variants are not important in SLE cohorts of south-east Asian ancestry or that different class III SNPs that are uncommon in Europeans and hence not typed in this study show association in these populations.

Two recent genome-wide association scans in SLE case-control cohorts of Chinese ancestry have shown that the most highly associated SNPs were located in the class II region of the MHC, between HLA-DRB1 and HLA-DQA2.2,3 We observed a similar pattern of association in the Filipino cohort under study. The most highly associated SNP in the Filipino cohort, rs9271366, is a surrogate marker for HLA-DRB1*15:02. This SNP showed the greatest overall association in the Hong Kong Chinese genome-wide association study and is ranked 8 of the top 15 MHC SNPs in the Han Chinese genome-wide association study in SLE, implicating HLA-DRB1*15:02 haplotypes in SLE susceptibility in these populations as well.
Box 1 Major histocompatibility complex region susceptibility genes and haplotypes in systemic lupus erythematosus

- **HLA-DPB1**: For the first time we report an association with single nucleotide polymorphisms (SNPs) in the region of HLA-DPB1 and systemic lupus erythematosus (SLE). The most associated SNP in this region is rs3117213, located between the pseudogenes HLA-DPA2 and COL11A2P. This SNP has previously shown association with ACPA-positive rheumatoid arthritis (RA) in northern European cohorts and is independent of the known RA HLA-DRB1 risk alleles.29 30 The association of HLA-DPB1 alleles with chronic beryllium disease is well established.31 Recent studies have also detected associations in the HLA-DPB1 region with chronic hepatitis B infection, Wegener’s granulomatosis, primary biliary cirrhosis, Graves’ disease, Takayasu’s arteritis, juvenile idiopathic arthritis, systemic sclerosis and multiple sclerosis in European and non-European populations.29 32–36 The association of this region with these autoimmune, infectious and inflammatory diseases suggests that it is likely to represent a shared autoimmune/inflammatory locus.

- **MSH5**: The MSH5 (MutS homologue 5) gene, located in the major histocompatibility complex (MHC) class III region, comprises 26 exons and spans 25 kb.37 MSH5 belongs to a family of proteins involved in DNA mismatch repair and meiotic recombination. Recent evidence also suggests a role for human MSH5 in promoting ionising radiation-induced apoptosis.38

- **HLA-G**: The principal signal after stepwise logistic regression analysis on the top Filipino SNP rs9271366 is observed in the class I region of the MHC between HLA-G and HLA-H. HLA-H is transcribed but not translated. HLA-G is a non-classical HLA class I molecule that has been implicated in immune tolerance mechanisms and autoimmune disease.39 Unlike other classical HLA molecules, HLA-G exhibits limited nucleotide and protein polymorphism and shows marked tissue restriction in that the molecule is primarily expressed on cytotoxic lymphocytes of the placenta where it is thought to mediate maternal-fetal tolerance.40 41 (http://hla.alleles.org/proteins/class1.html). HLA-G expression can be induced in tumours, transplanted tissue and plaques from patients with multiple sclerosis. HLA-G is able to suppress immune responses through binding inhibitory receptors such as ILT2, ILT4 and KIR2DL4 which are expressed on a variety of immune cells including NK cells (KIR2DL4), CD4+ and CD8+ T cells, B cells, monocytes and dendritic cells.41 Recent studies in multiple sclerosis and SLE have suggested HLA-G as a putative disease susceptibility gene.39 A 14 bp insertion located in the 3′UTR of the gene has been associated with lower levels of HLA-G mRNA. Interestingly, the HLA-G*01:01:02 allele which is the most common allele carrying the 14 bp insertion shows strong linkage disequilibrium with the SLE-associated HLA-A*01:HLA-B*08:HLA-DRB1*13:HLA-DQA1*05:HLA-DQB1*02 haplotype in European populations.42

- **RCCX module**: The MHC class III risk haplotype, tagged by rs1269852 and rs3130490 showing primary association in the Spanish cohort, spans the copy variable RCCX module containing complement component C4. A similar haplotype, also extending across the RCCX module, harbours the most highly associated SNPs in our northern European SLE cohort.5 Despite the haplotypic diversity observed in the southern European Spanish population compared with northern Europeans, we have not been able to fine-map this signal using SNPs. The reasons underlying the conservation of this haplotype remain unclear but include selective pressures and co-regulation of gene expression.

- **HLA-DRB1*15 alleles**: The HLA-DRB1*15:01 and HLA-DRB1*15:02 alleles, which show association in European and Filipino SLE respectively, differ by only one amino acid—a valine (DRB1*15:01) to glycine (DRB1*15:02) substitution at position 86 of the amino acid sequence in pocket P1 of the peptide-binding cleft.53 Haplotypes harbouring HLA-DRB1*15:01 alleles confer risk for SLE and multiple sclerosis but also demonstrate protection in diseases such as type 1 diabetes and IgA deficiency.5 The mechanisms underlying these phenomena are unknown.

- **HLA-C / HLA-B region**: Data from the Filipino and UK SLE cohorts under study reveal association in the class I region of the MHC encompassing HLA-C/HLA-B/MICA. A recent genome-wide association study comparing HIV-1 controllers and progressors found that the same region harbours the major genetic determinants of HIV-1 control. Further analysis implicated specific HLA-B allele peptide groove amino acids, in addition to an independent HLA-C effect in the control of HIV infection.44

Using transancestral SNP mapping of the MHC region in SLE, we have successfully refined the established SLE-associated HLA-DRB1*15 signal to an interval encompassing HLA-DRB1 and HLA-DQA1. We have identified and replicated association in the genes MSH5 and HLA-DPB1 in Filipino and Spanish SLE cohorts, and also demonstrated association at HLA-G in Filipino SLE. These signals are independent of each other and HLA-DRB1 alleles and show some evidence of genetic interaction. These data highlight the usefulness of mapping disease susceptibility loci using a transancestral approach, particularly in a region as complex as the MHC, and offer a springboard for further fine-mapping, resequencing and transcriptomic analysis.
Basic and translational research


Clinicians who provided access to SLE samples: Norberto Ortego Centeno, Klareskog, Peter K Gregersen, Jorge R Oksenberg and Stephen L Hauser.

A full list of the investigators who contributed to the generation of the WTCCC data is available from www.wtccc.org.uk.

Acknowledgments

The authors would like to thank all study participants. Funding: MMAF and LB were funded through an Arthritis Research UK grant (18239). The IMAGEN consortium was supported by grant A0876152 from the National Institutes of Allergy and Infectious Diseases. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. Funding for the project was provided by the Wellcome Trust under awards 076113 and 085475. We acknowledge the use of DNA from the British 1958 Birth Cohort collection (D Strachan, S Ring, W McArdle and M Pembrey) funded by the Medical Research Council grant G0000934 and Wellcome Trust grant 068545/2/02.

Competing interests

None.

Ethics approval

This study was approved by the London Research Ethics Committee, UK (Ref: 06/MRE02/9) and the Comité de Ética del CSIC, Granada, Spain.

Provenance and peer review

Not commissioned; externally peer reviewed.

REFERENCES

Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at MSH5, HLA-DPB1 and HLA-G

Michelle M A Fernando, Jan Freudenberg, Annette Lee, David Lester Morris, Lora Boteva, Benjamin Rhodes, María Francisca Gonzalez-Escribano, Miguel Angel Lopez-Nevot, Sandra V Navarra, Peter K Gregersen, Javier Martin, IMAGEN and Timothy J Vyse

Ann Rheum Dis published online January 10, 2012

Updated information and services can be found at: http://ard.bmj.com/content/early/2012/01/10/annrheumdis-2011-200808

These include:

Supplementary Material Supplementary material can be found at: http://ard.bmj.com/content/suppl/2012/04/18/annrheumdis-2011-200808.DC1

References This article cites 40 articles, 11 of which you can access for free at: http://ard.bmj.com/content/early/2012/01/10/annrheumdis-2011-200808#BIBL

Open Access This paper is freely available online under the BMJ Journals unlocked scheme, see http://ard.bmj.com/info/unlocked.dtl

Email alerting service Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections Articles on similar topics can be found in the following collections

Open access (593)
Connective tissue disease (4253)
Genetics (968)
Immunology (including allergy) (5144)
Systemic lupus erythematosus (571)
Inflammation (1251)

Notes

To request permissions go to: http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to: http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to: http://group.bmj.com/subscribe/