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] María Francisca Gonzalez-Escribano,[3] Miguel Angel Lopez-Nevot,[4] Sandra V Navarra,[5] Peter K Gregersen,[2] Javier Martin,[6] IMAGEN,* Timothy J Vyse[1]

► Additional data are published online only. To view the files please visit the journal online (http://ard.bmj.com/content/71/5.toc).

For numbered affiliations see end of article

Correspondence to
Michelle M A Fernando and Timothy J Vyse, Division of Genetics and Molecular Medicine and Division of Immunology, Infection and Inflammatory Disease, King's College London, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK;
michelle.fernando@kcl.ac.uk
timothy.vyse@kcl.ac.uk

Received 21 August 2011
Accepted 30 November 2011
Published Online First
10 January 2012

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.[5]

All SLE probands fulfilled the American College of Rheumatology criteria for the classification of SLE.[7] 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).[8 9]

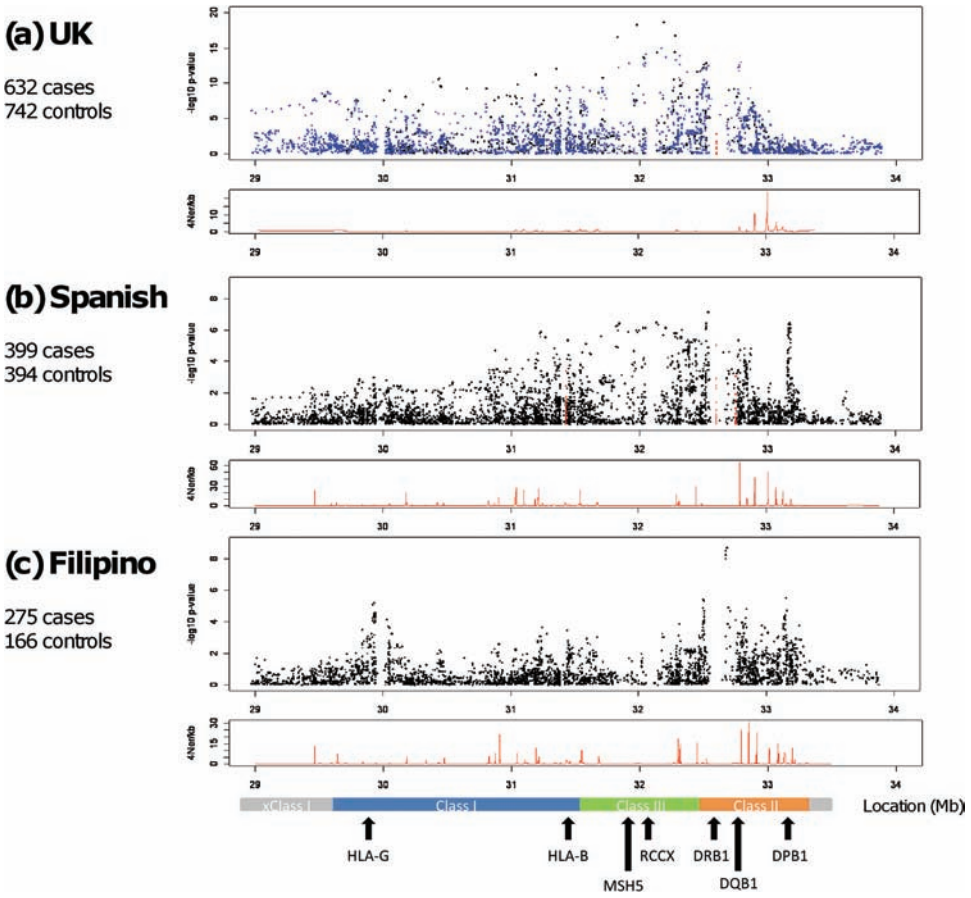


Figure 1 High-density transancestral single nucleotide polymorphism (SNP) mapping of the major histocompatibility complex (MHC) region in UK, Spanish and Filipino systemic lupus erythematosus (SLE). The panels show MHC region association plots for (A) UK, (B) Spanish and (C) Filipino SLE cohorts where genomic position (Mb) is shown on the horizontal axis with $-\log_{10}$ 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*.[27] 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 Rocío, 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, *rs9271366*, 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 (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 AIMs (performed using EIGENSTRAT and defined as >4 SDs from the mean)[11] (failed Spanish n=62, Filipino n=88). The genomic inflation factor (λ_{GC}) was calculated using the post-QC AIMs after correction for population stratification (Spanish λ_{GC}=1.04 and Filipino λ_{GC}=1.09).

UK SLE cohort imputation

Genotypes were imputed using IMPUTE[12] 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

SNP ID	Position	F_U	Associated allele	OR (95% CI)	p	Other disease	Location (LD)*	Gene expression¶
UK SLE								
rs1269852	32188169	0.11	C	2.68 (2.16 to 3.33)	2.48×10^{-19}		TNXB-ATF6B (760 kb)	
rs3906272	*31370903*	*0.04*	*A*	*2.84 (2.05 to 3.92)*	2.95×10^{-10}		*HLA-C-HLA-B (22 kb)*	
rs3129868	*32512355*	*0.11*	*T*	*1.67 (1.29 to 2.14)*	8.65×10^{-5}		*BTNL2 – DRA (1501 tag SNP) (375 kb)*	
rs4713419	*31101215*	*0.10*	*C*	*0.40 (0.25 to 0.63)*	7.55×10^{-5}		*MUC21 – PSORS1C1 (9 kb)*	*HLA-C*‡
Spanish SLE								
rs3130490	31847099	0.04	A	2.96 (1.95 to 4.51)	3.94×10^{-7}		C6orf27 (620 kb)	
rs3129768	*32703061*	*0.13*	*C*	*1.91 (1.44 to 2.53)*	7.57×10^{-6}	*MS*	*BTNL2-DRA (1501 tag SNP) (182 kb)*	
rs3117213	*33172583*	*0.18*	*A*	*1.76 (1.37 to 2.25)*	7.18×10^{-6}	*ACPA+ RA CBD, SSc, PBC, Hep B*	*DPB1-DPB2 (29 kb)*	
rs409558	*31816126*	*0.20*	*G*	*0.57 (0.43 to 0.77)*	2.2×10^{-4}		*MSH5 (20 kb)*	*MSH5*§
Filipino SLE								
rs9271366	32694832	0.34	G	2.46 (1.83 to 3.30)	1.97×10^{-9}	MS	DRB1-DQA1 (*1502 tag SNP*) (87 kb)	
rs2571391	*30031817*	*0.16*	*C*	*0.36 (0.22 to 0.59)*	6.06×10^{-5}	*MS*	*HLA-G-HLA-H (300 kb)*	
rs2507987	*31451012*	*0.48*	*A*	*0.48 (0.35 to 0.66)*	6.80×10^{-6}		*HLA-B-MICA (6 kb)*	
rs409558	*31816126*	*0.30*	*G*	*0.56 (0.37 to 0.86)*	7.47×10^{-3}		*MSH5 (100 kb)*	*MSH5*§
rs2071351	*33151908*	*0.17*	*A*	*0.48 (0.28 to 0.83)*	8.44×10^{-3}		*DPB1 (19 kb)*	*HLA-DPB1*†

Stepwise logistic regression association results are shown in italics.
Meta-analysis of Spanish and Filipino data for the *MSH5* SNP rs409558 revealed a locus-wide level of significance at $p=1.92\times10^{-5}$.
*LD surrounding each SNP calculated in control population of each cohort using r^2 cut-off >0.8.
†All HapMap populations, CEU, YRI, CHB and JPT.
‡CEU and YRI only.
§CEU only.
¶Data from GEO database (http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE6536).[28]
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.

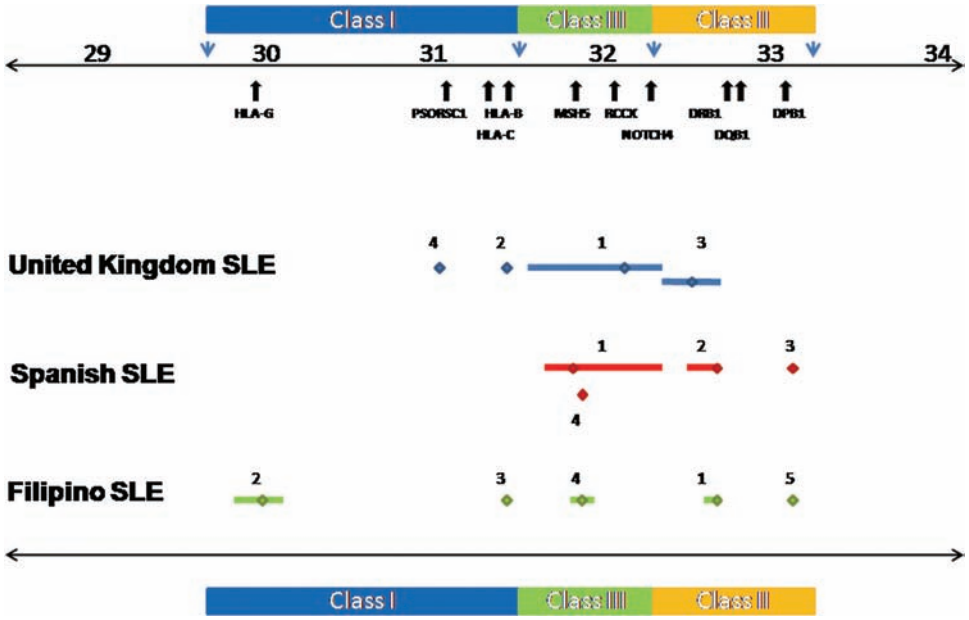


Figure 2 Primary and secondary major histocompatibility complex (MHC) region association signals in UK, Spanish and Filipino systemic lupus erythematosus (SLE). This figure shows the primary and secondary association signals in UK, Spanish and Filipino SLE denoted in blue, red and green, respectively. The primary signals are labelled 1 and the secondary signals obtained by stepwise logistic regression are labelled 2–5 and correspond to the single nucleotide polymorphism (SNPs) shown in table 1. An indication of linkage disequilibrium (LD) surrounding each marker (calculated in the control population of each cohort using r^2 cut-off >0.8) is shown by the bars flanking each marker. LD is <30 kb where flanking bars are absent. The genomic position is shown above the plot together with the positions of the relevant MHC region genes.

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 $p=7\times10^{-5}$ was set, given that genome-wide significance thresholds based on haplotype structure are typically in the range of $5–7\times10^{-8}$ 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, Bonferonni-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, rs9268832, was located in the class II pseudogene *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-DPB1* (best SNP rs3117213) as well as risk and protective signals in and surrounding *MSH5* (best risk 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 rs3130490 (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 rs3130490 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 rs3130490 also revealed a number of potentially independent signals in the Spanish cohort, the best of which was the class II SNP rs3129768 located between *HLA-DRB1* and *HLA-DQA1* (OR 1.91, CI 1.44 to 2.53, $p=7.57\times10^{-6}$). 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=0.93). Further stepwise logistic regression revealed association with the previously mentioned SNPs rs3117213 (*HLA-DPB1*) and rs409558 (*MSH5*) (table 1). LD analysis revealed that the association underlying rs9268832 probably represents a composite effect of rs3130490 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, $p=5.53\times10^{-6}$) (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, $p=1.97\times10^{-9}$). 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%.[17 18] 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.[19–21]

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 0.36, CI 0.22 to 0.59, $p=6.06\times10^{-5}$). Further stepwise logistic regression revealed additional independent signals that replicate those observed in the Spanish cohort: *MSH5* (best SNP rs409558) and *HLA-DPB1* (best SNP rs2071351) (table 1 and figure S3 in online supplement). Meta-analysis of Spanish and Filipino data for the *MSH5* SNP rs409558 revealed a locus-wide significance level at $p=1.92\times10^{-5}$ (OR_{meta} 0.58, CI_{meta} 0.33 to 0.83).

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 SNPs demonstrating disease association following primary conditional analysis in the UK (r^2 with rs3129868=0.98) and Spanish (r^2 with rs3129768=0.77) cohorts because these SNPs (rs3129868 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 $p=5.25\times10^{-4}$, 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

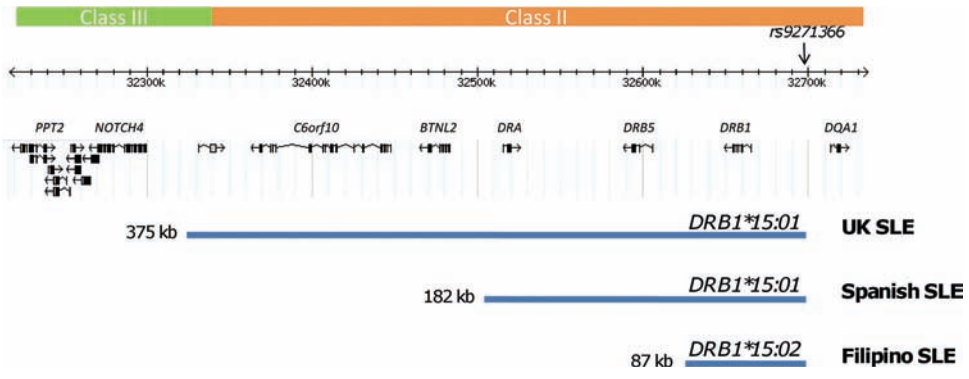


Figure 3 Transancestral fine-mapping of the *HLA-DRB1*15* signal in UK, Spanish and Filipino systemic lupus erythematosus (SLE). The frequencies of *HLA-DRB1*15* alleles show geographical variability. For example, in Europeans the common allele is *HLA-DRB1*15:01*, in Pacific and South East Asians it is *HLA-DRB1*15:02*, while in African populations it is *HLA-DRB1*15:03* (http://allelefrequencies.net/). It is well established that haplotypes harbouring *HLA-DRB1*15* alleles show primary or secondary association with SLE and data from this study support this view. However, the identity of causal variation has remained elusive due to the strong linkage disequilibrium present on the common disease-associated *HLA-DRB1*15:01* haplotype in northern Europeans. Using the single nucleotide polymorphism (SNP) *rs9271366* as a surrogate marker for the common *HLA-DRB1*15* allele in each population studied, we have refined the disease-associated region from 375 kb in northern Europeans to 87 kb in the Filipino population. The latter region encompasses the *HLA-DRB1* gene itself as well as part of the intergenic interval between *HLA-DRB1* and *HLA-DQA1*.

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).

	SNP1	SNP2	SNP3	HAPLOTYPE	F_A	F_U	F_Uexp	OR	95% CI	p	p_{perm}
UK_{risk}	rs1269852	rs3906272	rs3129868	CAT	<0.001	<0.001	0.0004	NA	NA	NA	NA
UK_{prot}	rs1269852	rs3906272	rs3129868	GGG	0.524	0.729	0.7604	0.37	0.31 to 0.44	3.72×10^{-28}	1×10^{-4}
$Spanish_{risk}$	rs3130490	rs3129768	rs3117213	ACA	0.004	<0.001	0.001	NA	NA	NA	NA
$Spanish_{prot}$	rs3130490	rs3129768	rs3117213	CAC	0.506	0.688	0.6849	0.443	0.36 to 0.55	2.22×10^{-13}	1×10^{-4}
$Filipino_{risk}$	rs9271366	rs2571391	rs2507987	GAT	0.273	0.092	0.1485	3.45	2.24 to 5.33	5.69×10^{-11}	1×10^{-4}
$Filipino_{prot}$	rs9271366	rs2571391	rs2507987	ACA	0.005	0.031	0.0507	NA	NA	NA	NA

F_A, haplotypic frequency in affected SLE cases; F_U, haplotypic frequency in unaffected controls; F_U_{exp}, expected haplotypic 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.

phenotype such as renal involvement, as is often observed in non-European populations.[22–24]

Refinement of SLE-associated HLA-DRB1*15 signal

Previous attempts to fine-map the SLE-associated *HLA-DRB1*1501* haplotypic signal could only delimit the region to approximately 500 kb of the MHC class II region in European-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

Haplotypic 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 (rs3130490, 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 rs3130490*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 haplotypic 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, $p=5.69\times10^{-11}$; protective OR 0.002, CI 1×10^{-4} to 0.05, $p=2.10\times10^{-4}$). 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 0.273, 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).

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-G, 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 haplotypic 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 rs3130490 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-DRA* 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 13 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 cytotrophoblastic cells 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*03-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.[43] 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.[4 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.

Author affiliations [1]Division of Genetics and Molecular Medicine and Division of Immunology, Infection and Inflammatory Disease, Guy's Hospital, King's College London, London, UK
[2]The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, New York, USA
[3]Department of Immunology, Hospital Virgen del Rocío, Seville, Spain
[4]Department of Immunology, Hospital Virgen de las Nieves, Granada, Spain
[5]Section of Rheumatology, Clinical Immunology and Osteoporosis, University of Santo Tomas, Manila, Philippines
[6]Instituto de Parasitologia y Biomedicina 'Lopez-Neyra', IPBLN-CSIC, Granada, Spain

*Membership of the International MHC and Autoimmunity Genetics Network (IMAGEN) is provided at the end of the paper.

International MHC and Autoimmunity Genetics Network (IMAGEN): John D Rioux, Philippe Goyette, Timothy J Vyse, Lennart Hammarström, Michelle M A Fernando, Todd Green, Philip L De Jager, Sylvain Foisy, Joanne Wang, Paul I W de Bakker, Stephen Leslie, Gilean McVean, Leonid Padyukov, Lars Alfredsson, Vito Annese, David A Hafler, Qiang Pan-Hammarström, Ritva Matell, Stephen J Sawcer, Alastair D Compston, Bruce A C Cree, Daniel B Mirel, Mark J Daly, Tim W Behrens, Lars Klareskog, Peter K Gregersen, Jorge R Oksenberg and Stephen L Hauser.

Clinicians who provided access to SLE samples: Norberto Ortego Centeno, Department of Internal Medicine, Hospital San Cecilio, Granada, Juan Jimenez Alonso, Department of Internal Medicine, Hospital Virgen de las Nieves, Granada, Enrique de Ramon Garrido, Department of Internal Medicine, Hospital Carlos Haya, Malaga, Maria Teresa Camps Garcia, Department of Internal Medicine, Hospital Carlos Haya, Malaga, Julio Sanchez Roman, Department of Internal Medicine, Hospital Virgen del Rocio, Seville, Spain.

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 AI067152 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/Z/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

1. **Gateva V**, Sandling JK, Hom G, *et al.* A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat Genet* 2009;**41**:1228–33.
2. **Han JW**, Zheng HF, Cui Y, *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet* 2009;**41**:1234–7.
3. **Yang W**, Shen N, Ye DQ, *et al.* Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet* 2010;**6**:e1000841.
4. **Fernando MM**, Stevens CR, Walsh EC, *et al.* Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet* 2008;**4**:e1000024.
5. **Rioux JD**, Goyette P, Vyse TJ, *et al.* Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc Natl Acad Sci USA* 2009;**106**:18680–5.
6. **Barcellos LF**, May SL, Ramsay PP, *et al.* High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. *PLoS Genet* 2009;**5**:e1000696.
7. **Tan EM**, Cohen AS, Fries JF, *et al.* The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982;**25**:1271–7.
8. **Tian C**, Plenge RM, Ransom M, *et al.* Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet* 2008;**4**:e4.
9. **Tian C**, Kosoy R, Nassir R, *et al.* European population genetic substructure: further definition of ancestry informative markers for distinguishing among diverse European ethnic groups. *Mol Med* 2009;**15**:371–83.
10. **Purcell S**, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–75.
11. **Price AL**, Patterson NJ, Plenge RM, *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;**38**:904–9.
12. **Howie BN**, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;**5**:e1000529.
13. **Burton PR**, Clayton DG, Cardon LR, *et al.* Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 2007;**39**:1329–37.
14. **Marchini J**, Howie B, Myers S, *et al.* A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;**39**:906–13.
15. **Barrett JC**, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;**21**:263–5.
16. **Harley JB**, Alarcón-Riquelme ME, Criswell LA, *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet* 2008;**40**:204–10.
17. **Mack SJ**, Bugawan TL, Moonsamy PV, *et al.* Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions. *Tissue Antigens* 2000;**55**:383–400.
18. **Bugawan TL**, Klitz W, Alejandrino M, *et al.* The association of specific HLA class I and II alleles with type 1 diabetes among Filipinos. *Tissue Antigens* 2002;**59**:452–69.
19. **Hashimoto H**, Nishimura Y, Dong RP, *et al.* HLA antigens in Japanese patients with systemic lupus erythematosus. *Scand J Rheumatol* 1994;**23**:191–6.
20. **Hong GH**, Kim HY, Takeuchi F, *et al.* Association of complement C4 and HLA-DR alleles with systemic lupus erythematosus in Koreans. *J Rheumatol* 1994;**21**:442–7.
21. **Sirikong M**, Tsuchiya N, Chandanayingyong D, *et al.* Association of HLA-DRB1*1502-DQB1*0501 haplotype with susceptibility to systemic lupus erythematosus in Thais. *Tissue Antigens* 2002;**59**:113–17.
22. **Osio-Salido E**, Manapat-Reyes H. Epidemiology of systemic lupus erythematosus in Asia. *Lupus* 2010;**19**:1365–73.
23. **Navarra SV**, Ishimori MI, Uy EA, *et al.* Studies of Filipino patients with systemic lupus erythematosus: autoantibody profile of first-degree relatives. *Lupus* 2011;**20**:537–43.
24. **Thumboo J**, Wee H-L. Systemic lupus erythematosus in Asia: is it more common and more severe? *APLAR J Rheumatol* 2006;**9**:320–6.
25. **Graham RR**, Ortmann WA, Langefeld CD, *et al.* Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am J Hum Genet* 2002;**71**:543–53.
26. **Danchenko N**, Satia JA, Anthony MS. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus* 2006;**15**:308–18.
27. **Auton A**, McVean G. Recombination rate estimation in the presence of hotspots. *Genome Res* 2007;**17**:1219–27.
28. **Stranger BE**, Forrest MS, Dunning M, *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007;**315**:848–53.
29. **Ding B**, Padyukov L, Lundström E, *et al.* Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region. *Arthritis Rheum* 2009;**60**:30–8.
30. **Orozco G**, Barton A, Eyre S, *et al.* HLA-DPB1-COL11A2 and three additional xMHC loci are independently associated with RA in a UK cohort. *Genes Immun* 2011;**12**:169–75.
31. **Dai S**, Murphy GA, Crawford F, *et al.* Crystal structure of HLA-DP2 and implications for chronic beryllium disease. *Proc Natl Acad Sci USA* 2010;**107**:7425–30.
32. **Wu XM**, Wang C, Zhang KN, *et al.* Association of susceptibility to multiple sclerosis in Southern Han Chinese with HLA-DRB1, -DPB1 alleles and DRB1-DPB1 haplotypes: distinct from other populations. *Mult Scler* 2009;**15**:1422–30.
33. **Zhou X**, Lee JE, Arnett FC, *et al.* HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. *Arthritis Rheum* 2009;**60**:3807–14.
34. **Wieczorek S**, Holle JU, Epplen JT. Recent progress in the genetics of Wegener's granulomatosis and Churg-Strauss syndrome. *Curr Opin Rheumatol* 2010;**22**:8–14.
35. **Kamatani Y**, Wattanapokayakit S, Ochi H, *et al.* A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. *Nat Genet* 2009;**41**:591–5.
36. **Radstake TR**, Gorlova O, Rueda B, *et al.* Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet* 2010;**42**:426–9.
37. **Her C**, Doggett NA. Cloning, structural characterization, and chromosomal localization of the human orthologue of Saccharomyces cerevisiae MSH5 gene. *Genomics* 1998;**52**:50–61.
38. **Tompkins JD**, Wu X, Chu YL, *et al.* Evidence for a direct involvement of hMSH5 in promoting ionizing radiation induced apoptosis. *Exp Cell Res* 2009;**315**:2420–32.
39. **Larsen MH**, Hviid TV. Human leukocyte antigen-G polymorphism in relation to expression, function, and disease. *Hum Immunol* 2009;**70**:1026–34.
40. **Apps R**, Gardner L, Moffett A. A critical look at HLA-G. *Trends Immunol* 2008;**29**:313–21.
41. **Carosella ED**, Favier B, Rouas-Freiss N, *et al.* Beyond the increasing complexity of the immunomodulatory HLA-G molecule. *Blood* 2008;**111**:4862–70.
42. **Kolte AM**, Steffensen R, Nielsen HS, *et al.* Study of the structure and impact of human leukocyte antigen (HLA)-G-A, HLA-G-B, and HLA-G-DRB1 haplotypes in families with recurrent miscarriage. *Hum Immunol* 2010;**71**:482–8.
43. **Marsh SG**, Bodmer JG. HLA class II nucleotide sequences, 1992. *Immunogenetics* 1993;**37**:79–94.
44. **Pereyra F**, Jia X, McLaren PJ, *et al.* The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 2010;**330**:1551–7.

Supplementary Information

Trans-ancestral mapping of the MHC region in SLE identifies new independent and interacting loci at *MSH5, HLA-DPB1* and *HLA-G*

Supplementary Methods

Custom iSelect SNP selection

The MHC region SNPs were chosen according to the following criteria: all SNPs put forward were required to have an Illumina designability score >0.8. All Illumina MHC-Panel SNPs were chosen (~2,360). Additional SNPs were selected from [1] and [2]. Further SNPs were added from HapMap Phase II. In this latter dataset there are 7,574 SNPs that are both designable (score>0.8) and common (minor allele frequency, MAF>5%) in any population. SNPs were included if they were (i) not picked in the above steps, (ii) not in perfect linkage disequilibrium (LD) with any SNP from the above steps and (iii) not in perfect LD with any other picked SNP (LD estimates across CEU, YRI and CHB+JPT). To resolve the phase, haplotypes from the HapMap website were used. Only SNPs with designability-score >= 1.0 were allowed to tag other SNPs. SNPs identified from the resequencing of eight homozygous MHC haplotypes were also included [3]. From the 37,535 SNPs called between the eight haplotypes, 11,112 were designable and had their minor allele on at least two haplotypes. From these, a set 309 SNPs were not in perfect LD (on the eight haplotypes) with any HapMap SNP or another SNP from the resequencing project and were added to the dataset. Owing to the small sample size, there will be many false positive instances of perfect LD. To account for this, coding SNPs and SNPs in recombination hotspots were also added.

Supplementary Table 1: Single marker association analysis in Spanish SLE cohort (SNPs with $p<10^{-5}$)

MARKER	BP	A1	OR	L95	U95	P
RS9268832	32535767	A	1.80	1.45	2.23	7.64E-08
RS1150758	32136127	C	2.61	1.81	3.78	3.34E-07
RS3117213	33172583	A	1.86	1.46	2.35	3.55E-07
RS3130490	**31847099**	**A**	**2.96**	**1.95**	**4.51**	**3.94E-07**
RS7195	32520517	A	1.75	1.41	2.17	3.96E-07
RS2213586	32521072	A	1.75	1.41	2.17	3.96E-07
RS9277545	33163301	A	1.86	1.46	2.37	4.42E-07
RS3117231	33182886	G	1.85	1.46	2.36	4.54E-07
RS1150755	32146528	A	2.58	1.79	3.73	4.71E-07
RS2395314	33170651	A	1.86	1.46	2.36	4.78E-07
RS3131379	31829012	A	2.87	1.90	4.32	5.13E-07
RS3117574	31833209	A	2.87	1.90	4.32	5.13E-07
RS3763327	32521808	G	1.73	1.39	2.14	5.79E-07
RS3131381	31816442	A	2.85	1.89	4.30	5.87E-07
RS2227139	32521437	G	1.73	1.40	2.15	5.96E-07
RS7194	32520458	G	1.72	1.39	2.14	7.14E-07
RS2213585	32521128	G	1.72	1.39	2.14	7.14E-07
RS1270942	32026839	G	2.83	1.87	4.26	7.38E-07
RS389884	32048876	G	2.83	1.87	4.26	7.38E-07
RS3130288	32203979	A	2.82	1.87	4.26	7.46E-07
RS3128972	33166752	G	1.83	1.44	2.32	7.59E-07
RS3130212	33182367	C	1.87	1.46	2.40	7.72E-07
RS9277555	33163583	A	1.83	1.44	2.33	7.87E-07
RS8084	32519013	A	1.70	1.38	2.10	8.45E-07
RS3117222	33168927	A	1.82	1.43	2.31	8.62E-07
RS3130320	32331236	A	1.76	1.41	2.21	9.59E-07
RS2179919	33167240	G	1.82	1.43	2.30	9.72E-07
RS3128919	33169604	A	1.81	1.43	2.30	9.78E-07
RS558702	31978305	A	2.79	1.85	4.20	1.06E-06
RS497309	32000463	C	2.79	1.85	4.20	1.06E-06
RS3117577	31835453	G	2.78	1.84	4.20	1.08E-06
RS3091284	33165222	C	1.82	1.43	2.32	1.25E-06
RS3130455	31233957	T	1.99	1.51	2.63	1.30E-06
RS130065	31230479	A	1.93	1.48	2.52	1.31E-06
RS3130679	31915519	G	2.76	1.83	4.17	1.35E-06
RS3117234	33181962	G	1.83	1.43	2.35	1.39E-06
RS519417	31986412	A	2.74	1.82	4.12	1.46E-06
RS3117226	33165637	A	1.83	1.43	2.34	1.49E-06
RS3094662	31229924	C	1.98	1.50	2.61	1.63E-06
RS3130191	33169849	G	1.79	1.41	2.28	1.64E-06
RS3134942	32276749	A	2.39	1.67	3.42	1.66E-06
RS3132956	32287416	A	2.39	1.67	3.42	1.66E-06

RS1269852	32188169	C	2.76	1.82	4.20	2.02E-06
RS6899657	33158201	A	1.79	1.41	2.27	2.07E-06
RS3132450	31704117	G	2.66	1.77	4.00	2.54E-06
RS3115572	32328462	C	1.67	1.35	2.07	2.64E-06
RS3131296	32280971	A	2.35	1.65	3.36	2.65E-06
RS130076	31230461	A	1.96	1.48	2.60	2.66E-06
RS1367730	33166092	A	1.79	1.40	2.28	2.75E-06
RS1265181	31263764	G	2.00	1.50	2.68	2.90E-06
RS1265178	31269208	A	2.00	1.50	2.68	2.90E-06
RS3117223	33168042	A	1.77	1.39	2.24	3.30E-06
RS6936204	32325070	A	1.72	1.37	2.17	3.44E-06
RS3134796	32297899	G	2.27	1.61	3.21	3.55E-06
RS3130210	33180707	A	1.80	1.40	2.31	3.72E-06
RS910051	32423489	C	2.55	1.71	3.79	4.24E-06
RS3129927	32441805	C	2.55	1.71	3.79	4.24E-06
RS9268165	32380305	G	1.89	1.44	2.48	4.63E-06
RS2523544	31441541	A	2.02	1.49	2.72	4.71E-06
RS1794282	32774504	A	2.51	1.69	3.73	4.73E-06
RS3129890	32522251	G	1.75	1.38	2.23	5.02E-06
RS3129950	32466179	C	2.43	1.66	3.55	5.03E-06
RS2395149	32433540	A	2.52	1.69	3.76	5.47E-06
RS7775397	32369230	C	2.51	1.69	3.75	5.81E-06
RS9268235	32398186	A	2.51	1.69	3.75	5.81E-06
RS1265757	32410360	A	2.51	1.69	3.75	5.81E-06
RS9268208	32388569	G	2.61	1.72	3.95	6.22E-06
RS2064478	33180244	A	1.76	1.38	2.25	6.61E-06
RS3130614	31584437	A	2.54	1.69	3.82	7.79E-06
RS3129962	32487361	A	2.46	1.66	3.65	8.11E-06
RS9268212	32389867	G	1.88	1.43	2.49	8.25E-06
RS1018433	32389488	A	1.86	1.41	2.44	8.71E-06
DRB1_[301]	**32600002**	**C**	**1.88**	**1.42**	**2.48**	**9.39E-06**
RS409558	31816126	G	0.53	0.40	0.71	1.07E-05
RS2293861	31819103	A	0.53	0.40	0.71	1.07E-05
RS2075788	31820160	C	0.53	0.40	0.71	1.07E-05
RS3828922	31821433	A	0.53	0.40	0.71	1.07E-05
RS3864299	32379652	T	1.84	1.40	2.41	1.13E-05
RS9268213	32390059	G	1.84	1.40	2.41	1.13E-05
RS6909790	32390957	G	1.84	1.40	2.41	1.13E-05
RS6915455	32391472	A	1.84	1.40	2.41	1.13E-05
RS9268215	32390449	G	1.83	1.40	2.40	1.21E-05
RS3117230	33183613	G	1.77	1.37	2.29	1.27E-05
RS3821236	191611003	A	1.69	1.34	2.15	1.32E-05
RS2075801	31836246	A	0.54	0.40	0.71	1.34E-05
RS1548306	32535157	T	1.57	1.28	1.93	1.36E-05
RS1986997	31336389	A	1.66	1.32	2.09	1.39E-05
RS9276447	32823547	C	1.60	1.29	1.98	1.51E-05

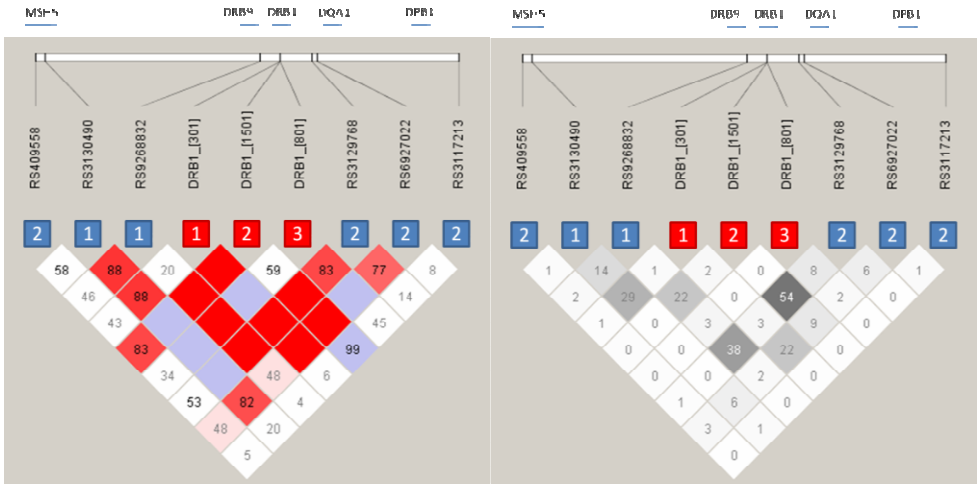
RS2143462	32443182	A	1.90	1.42	2.54	1.52E-05
RS2179920	33166852	A	1.76	1.36	2.28	1.52E-05
RS4947350	32875598	G	1.82	1.39	2.39	1.52E-05
RS7341328	32383172	A	1.83	1.39	2.40	1.55E-05
RS9268200	32386648	A	1.82	1.39	2.38	1.58E-05
RS9268168	32380488	A	1.82	1.39	2.39	1.59E-05
RS9268192	32385189	A	1.82	1.39	2.39	1.63E-05
RS6934546	32387930	G	1.81	1.38	2.37	1.64E-05
RS3132931	32343873	C	1.82	1.39	2.40	1.84E-05
RS9268137	32363247	A	1.82	1.39	2.40	1.84E-05
RS9268176	32382057	A	1.82	1.39	2.40	1.84E-05
RS6934776	32387794	A	1.82	1.39	2.40	1.84E-05
RS9275602	32790790	A	2.03	1.47	2.81	1.85E-05
RS7742654	32379421	G	1.83	1.39	2.42	1.85E-05
RS7775332	32378341	A	1.78	1.37	2.33	2.03E-05
RS9277385	33158451	G	1.67	1.32	2.11	2.09E-05
RS3115560	32344120	A	1.78	1.36	2.32	2.10E-05
RS3096673	32345991	G	1.78	1.36	2.32	2.10E-05
RS3115553	32353805	A	1.78	1.36	2.32	2.10E-05
RS1264372	30877705	A	1.96	1.44	2.67	2.13E-05
RS652888	31959213	G	1.69	1.33	2.16	2.26E-05
RS3129768	32703061	C	1.82	1.38	2.40	2.37E-05
RS6909427	32376679	C	1.77	1.36	2.31	2.47E-05
RS1018430	32389666	A	1.77	1.36	2.31	2.48E-05
RS9268198	32386165	A	1.80	1.37	2.36	2.50E-05
RS3115563	32341616	A	1.77	1.36	2.31	2.54E-05
RS3864302	32386770	A	1.77	1.36	2.31	2.64E-05
RS1059615	32657541	A	1.84	1.38	2.44	2.71E-05
RS3099844	31556955	A	2.12	1.49	3.01	2.75E-05
RS2734583	31613459	G	2.29	1.56	3.38	2.85E-05
RS9277366	33158085	A	1.60	1.29	2.00	2.86E-05
RS9268167	32380415	A	1.77	1.35	2.30	2.92E-05
RS3024886	191608694	A	1.66	1.31	2.11	2.98E-05
RS9268861	32537872	A	1.71	1.33	2.19	3.03E-05
RS7762279	32863268	G	2.22	1.52	3.22	3.07E-05
RS3130557	31202682	A	2.18	1.51	3.15	3.21E-05
RS6457499	32306926	G	0.65	0.53	0.80	3.28E-05
RS9267488	31622226	G	2.29	1.55	3.38	3.36E-05
RS3129924	32441277	A	1.87	1.39	2.52	3.42E-05
RS3129925	32441397	A	1.87	1.39	2.52	3.42E-05
RS2143461	32443325	A	1.87	1.39	2.52	3.42E-05
RS3129937	32444342	A	1.87	1.39	2.52	3.42E-05
RS7750783	32376058	A	1.78	1.35	2.33	3.49E-05
RS2524054	31360375	A	1.76	1.35	2.29	3.55E-05
RS2647044	32775888	A	1.66	1.31	2.11	3.56E-05
RS9268177	32382860	A	2.35	1.57	3.52	3.67E-05

RS2213567	32819633	C	1.55	1.26	1.90	4.14E-05
RS3129939	32444744	G	1.86	1.38	2.50	4.15E-05
RS2239803	32519811	G	1.52	1.24	1.85	4.20E-05
RS2844559	31448054	A	2.14	1.49	3.08	4.25E-05
RS9276731	32873563	A	2.20	1.51	3.20	4.27E-05
RS9276430	32820160	A	1.55	1.26	1.91	4.32E-05
RS3130349	32255674	A	1.83	1.37	2.45	4.39E-05
RS3134940	32257794	G	1.83	1.37	2.45	4.39E-05
RS3129926	32441458	G	1.86	1.38	2.50	4.43E-05
RS2856674	32767623	G	1.77	1.34	2.33	4.89E-05
RS9268197	32385912	A	1.78	1.35	2.35	5.24E-05
RS3134608	32225949	C	1.70	1.31	2.20	5.68E-05
RS1793891	31329677	A	1.62	1.28	2.05	5.80E-05
RS440454	32035321	A	1.63	1.28	2.06	5.96E-05
RS419788	32036778	A	1.63	1.28	2.06	5.96E-05
RS9277565	33164875	A	1.72	1.32	2.25	6.29E-05
RS9277567	33164991	C	1.72	1.32	2.25	6.29E-05
RS931	33162528	A	1.56	1.25	1.94	6.53E-05
RS2064476	33181300	G	1.55	1.25	1.93	6.61E-05
RS2523554	31439808	G	1.60	1.27	2.01	6.87E-05
RS6911419	32517765	A	1.50	1.23	1.84	6.96E-05
RS2239804	32519501	A	1.50	1.23	1.84	6.96E-05
RS3097652	33165813	A	1.56	1.25	1.94	7.31E-05
RS1960278	31377853	C	1.51	1.23	1.84	7.36E-05
RS11860650	31234207	A	1.65	1.29	2.11	7.55E-05
RS3130985	31193335	A	2.07	1.44	2.96	7.90E-05
RS9268220	32392318	A	1.84	1.36	2.50	7.93E-05
RS1264326	30959888	A	2.40	1.55	3.71	7.95E-05
RS4143332	31456344	A	2.13	1.46	3.11	8.06E-05
RS2844531	31461150	G	2.13	1.46	3.11	8.06E-05
RS6931646	32517759	G	1.50	1.23	1.83	8.27E-05
RS9268658	32518694	G	1.50	1.23	1.83	8.31E-05
RS3131788	31132775	A	2.05	1.43	2.94	8.41E-05
RS630379	32030233	A	1.64	1.28	2.10	8.45E-05
RS9268127	32361537	G	1.84	1.36	2.50	8.54E-05
RS9276435	32821845	A	1.73	1.32	2.27	8.58E-05
RS3117225	33165689	A	1.55	1.25	1.93	8.59E-05
RS2858331	32789255	G	0.67	0.55	0.82	8.74E-05
RS3130562	31208953	G	2.08	1.44	3.00	8.87E-05
RS9277394	33158948	A	1.54	1.24	1.91	9.35E-05
RS9277378	33158257	G	1.55	1.24	1.93	9.45E-05
RS9277393	33158855	G	1.54	1.24	1.91	9.49E-05
RS1042544	33162435	G	1.54	1.24	1.91	9.49E-05
RS9277542	33163225	G	1.54	1.24	1.91	9.49E-05
RS3117228	33164413	A	1.54	1.24	1.91	9.49E-05
RS3130188	33165154	G	1.54	1.24	1.91	9.49E-05

RS9277386	33158477	G	1.54	1.24	1.92	9.57E-05
RS3129933	32444139	A	1.83	1.35	2.48	9.70E-05
RS9276689	32859940	A	2.24	1.49	3.36	9.70E-05
RS3134931	32298598	G	0.66	0.54	0.81	9.81E-05

Supplementary Figure 1: Linkage disequilibrium between top independent MHC SNPs and *HLA-DRB1* alleles in Spanish controls

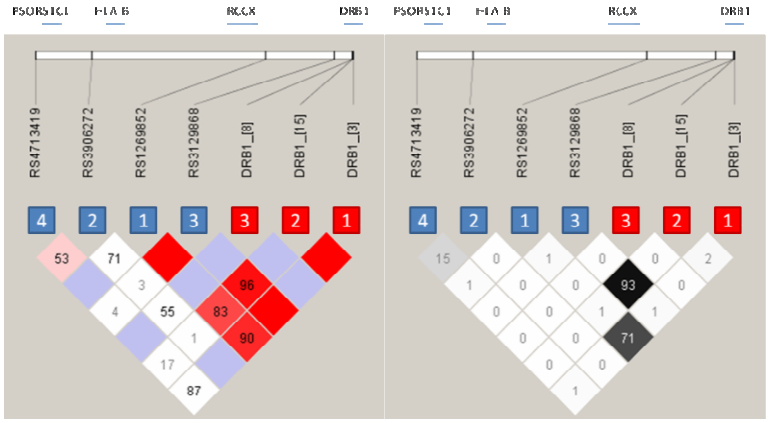
(a) D' plot (b) r^2 plot



Approximate location of MHC region genes is shown above the plots. SNPs and HLA allele positions are shown to scale. Numbers in boxes denote SNP (blue) and *HLA-DRB1* allele (red) primary (1) and secondary (2-3) signals.
Numbers in diamonds denote D' and r^2 values respectively.

Supplementary Figure 2: Linkage disequilibrium between top independent MHC SNPs and *HLA-DRB1* alleles in United Kingdom controls

(a) D' plot (b) r^2 plot



Supplementary Table 2: *HLA-DRB1* association in Spanish SLE using step-wise conditional logistic regression to assess relative predispositional effects

SNP ID	Position	F_U	OR (95% CI)	P	Other disease
			Spanish SLE		
DRB1*03:01	32600000	0.12	1.88 (1.43-2.48)	5.53x10-6	many
*DRB1*15:01*	*32600000*	*0.09*	*1.82 (1.30-2.54)*	*5.00x10-4*	*MS, T1D*
*DRB1*08:01*	*32600000*	*0.01*	*3.67 (1.60-8.38)*	*0.0021*	

Conditional association results are italicised; F_U, frequency in unaffected controls; OR, odds ratio; CI, confidence interval; p, p value (see Methods); MS, multiple sclerosis; T1D, type 1 diabetes.

SLR analyses at the *HLA-DRB1* locus in Spanish SLE reveal significant predisposing effects for the alleles, *HLA-DRB1*03:01*, *HLA-DRB1*15:01* and *HLA-DRB1*08:01*. These data contrast with those of northern European lupus cohorts where *HLA-DRB1*08:01* is uncommon and not associated with SLE [4-5]. However, the *HLA-DRB1*08:01-DQB1*0401* haplotype, inferred using microsatellite markers, was found to be associated with SLE in a previous European-American study. The current data suggest that this association is most likely due to a southern European component within the European-American cohort [6].

Supplementary Table 3: SNP associations following stepwise logistic regression using PC1, *HLA-DRB1*03:01, HLA-DRB1*15:01, HLA-DRB1*08:01* as covariates in Spanish SLE

MARKER	BP	A1	OR	L95	U95	P
RS3117213	33172583	A	1.81	1.42	2.31	2.28E-06
RS409558	31816126	G	0.55	0.41	0.73	5.31E-05
RS3130490	31847099	A	2.28	1.38	3.77	1.26E-03
RS9268832	32535767	A	1.48	1.14	1.92	3.32E-03
RS3129768	32703061	C	1.72	1.08	2.74	2.22E-02

Supplementary Table 4: SNP associations following stepwise logistic regression using PC1, *HLA-DRB1*03:01, HLA-DRB1*15:01* as covariates in United Kingdom SLE

MARKER	BP	A1	OR	L95	U95	P
RS3906272	31370903	A	3.01	2.70	3.32	3.25E-12
RS4713419	31101215	C	0.53	0.18	0.88	3.47E-04
RS1269852	32188169	C	1.66	1.25	2.07	1.65E-02
RS3129868	32512355	T	1.17	0.15	2.20	7.57E-01

Supplementary Table 5: Single marker association analysis in Filipino SLE cohort (SNPs with $p<10^{-5}$)

MARKER	BP	A1	OR	L95	U95	P
RS9271366	32694832	G	2.46	1.83	3.30	1.97E-09
RS1966002	32689662	C	0.42	0.31	0.55	2.27E-09
RS9270986	32682038	C	0.42	0.31	0.56	2.96E-09
RS615672	32682149	C	0.41	0.30	0.55	5.66E-09
RS2858867	32683303	G	0.41	0.31	0.56	1.01E-08
RS2071351	33151908	A	0.32	0.20	0.52	3.21E-06
RS984778	32508066	A	0.52	0.39	0.69	3.89E-06
RS3135338	32509195	A	0.52	0.39	0.69	4.52E-06
RS3135395	32513170	C	0.52	0.40	0.69	5.01E-06
RS2395173	32512837	G	0.52	0.40	0.69	5.04E-06
RS9258738	29935369	A	0.44	0.31	0.63	6.37E-06
RS2523767	29920786	A	0.45	0.32	0.64	8.17E-06
RS3129768	32703061	A	0.53	0.40	0.71	1.31E-05
RS9501624	32507264	G	3.26	1.91	5.56	1.40E-05
RS6457656	32844956	A	0.37	0.23	0.58	1.58E-05
RS9501259	33163529	G	0.55	0.41	0.72	1.96E-05
RS3129878	32516713	C	0.49	0.35	0.68	2.10E-05
RS9272689	32717083	A	0.54	0.41	0.72	2.18E-05
RS765649	29939037	T	0.47	0.33	0.67	2.90E-05
RS6926336	32515300	G	3.55	1.95	6.44	3.19E-05
RS1611717	29937556	A	0.48	0.34	0.68	3.32E-05
RS9258651	29930334	G	0.48	0.34	0.68	3.46E-05
RS2523756	29928696	A	0.48	0.34	0.68	3.55E-05
RS3094175	29930902	A	0.48	0.34	0.68	3.61E-05
RS1431399	33149012	A	0.42	0.28	0.64	3.85E-05
RS1611737	29939550	C	0.48	0.34	0.68	3.94E-05
RS1611710	29936895	G	0.48	0.34	0.68	4.13E-05
RS9268644	32516022	A	0.44	0.29	0.65	4.69E-05
RS2428510	29931006	G	0.49	0.34	0.69	4.74E-05

RS6919513	29931973	G	0.49	0.34	0.69	4.74E-05
RS9258679	29932038	A	0.49	0.34	0.69	4.74E-05
RS3077	33141000	A	0.40	0.25	0.62	5.03E-05
RS3129877	32516575	A	0.44	0.29	0.65	5.13E-05
RS1611732	29938987	G	0.49	0.34	0.69	5.17E-05
RS1978029	32839688	A	0.42	0.27	0.64	5.51E-05
RS9276586	32840915	G	0.42	0.27	0.64	5.51E-05
RS2844827	29919216	A	0.50	0.35	0.70	5.82E-05
RS9469341	33143855	A	0.41	0.26	0.63	6.67E-05
RS1611723	29938484	A	0.50	0.35	0.70	7.04E-05
RS2571391	30031817	C	0.38	0.24	0.62	7.43E-05
RS3129875	32515446	G	0.48	0.33	0.69	7.86E-05
RS1431400	33149154	G	0.41	0.26	0.64	7.92E-05
RS10214910	33145653	C	0.41	0.27	0.64	7.93E-05
RS9348904	33148813	A	0.42	0.27	0.64	8.15E-05
RS7755224	32760295	G	2.48	1.58	3.89	8.39E-05
RS4084096	29919847	C	0.51	0.36	0.71	8.52E-05
RS2395181	32515382	C	0.49	0.34	0.70	9.46E-05
RS9258766	29938390	G	0.51	0.36	0.71	9.48E-05

Supplementary Table 6: *HLA-DRB1* data stratified according to genotype of SNP, *rs9271366*, in 89 subjects of the Filipino SLE cohort

AA HOMs	ALLELES 1&2*
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---

AG HETs	ALLELE 1	ALLELE 2
case	15:02	04:05
case	15:02	14:04
case	15:02	07:01
case	15:02	09:01
case	15:02	07:01
case	15:02	04:05
case	15:02	08:03
case	15:02	12:02
case	15:02	07:01
case	15:02	12:02
case	15:02	03:01/34/50/51
control	15:02	09:01
control	15:02	12:02
control	15:02	11:01/75/81
control	15:02	08:03
control	15:02	12:02
control	15:02	12:02
control	15:02	12:02
control	15:02	12:02
control	15:02	09:01
control	15:02	16:02
control	15:02	11:01/75/81
control	15:02	10:01
control	15:02	13:02

GG HOMs	ALLELE 1	ALLELE 2
case	04:01	07:01
case	07:01	12:02
case	12:02/20	12:02/20
case	12:02/20	12:02/20
case	03:01/20/45/50/51	12:02/19
case	04:03/52	08:03
case	12:02	07:01
case	12:02/19	15:01/35/36/41/42/43
case	14:04	07:01
case	04:01	12:02
case	09:01/02	09:01/02
case	04:03/52	08:03
case	04:05	12:02
case	04:03/52	08:03
control	03:01/45/50/51	09:01
control	04:03/52	11:01/81
control	08:03	09:01
control	04:05	09:01
control	12:02/20	12:02/20
control	11:01	14:07
control	12:01/06/10/17	10:01
control	04:05/83/89	04:05/80/81N/83/89
control	04:06	12:02
control	03:01/45/50/51	11:01/75/81
control	03:01/45/48/50/51	1501/36/41/42/43
control	12:02/20	12:02/20

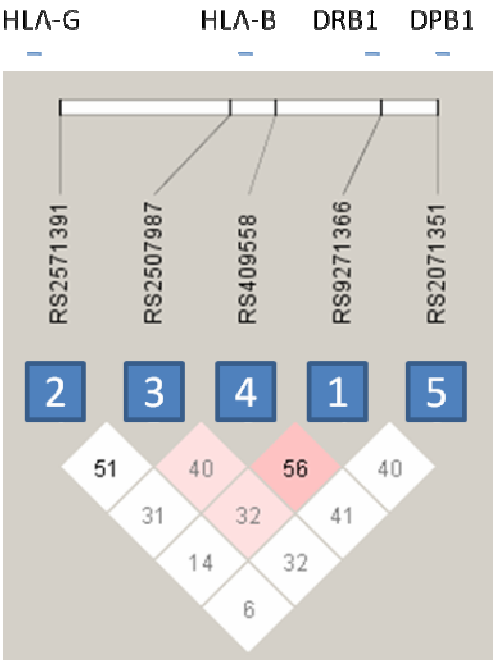
case	15:02, ---
case	15:02, ---
case	15:02, ---
case	15:02, ---
control	15:02, ---
control	15:02, ---
control	15:02, ---
control	15:02, ---
control	15:02, ---
control	15:02, ---
control	15:02, ---
control	15:02, ---
control	15:02, ---

The associated minor allele is A, the major allele is G

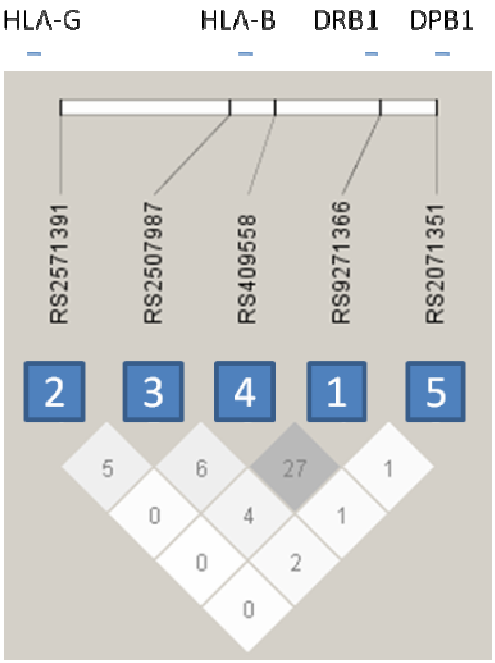
* all samples are homozygous for *HLA-DRB1*15:02*, some sample are homozygous for the string *HLA-DRB1*15:01/02/08*, however as the alleles *HLA-DRB1*15:01* and *HLA-DRB1*15:08* are rare in this cohort, the most likely genotype is homozygous *HLA-DRB1*15:02*

Supplementary Figure 3: Linkage disequilibrium between top independent MHC SNPs and *HLA-DRB1* alleles in Filipino controls

(a) D' plot



(b) r^2 plot



Supplementary Table 7: Association of SNP, rs9271366, in UK, Spanish and Filipino SLE

	UK SLE			Spanish SLE			Filipino SLE			All
	F_U	OR (95% CI)	P	F_U	OR (95% CI)	P	F_U	OR (95% CI)	P	BD P
Primary SMA	0.11	1.32 (1.04-1.66)	0.021	0.10	1.48 (1.09-2.01)	0.012	0.34	2.46 (1.83-3.30)	1.97x10-9	5.25x10-4
Conditional SMA	-	1.63 (1.28-2.07)	7.79x10-5	-	1.55 (1.13-2.11)	0.006	-	NA	NA	-

F_U, frequency in unaffected controls; OR, odds ratio; CI, confidence interval; P, univariate or conditional p-value; BD P, Breslow-Day p value; SMA, single marker association

Supplementary Table 8: Haplotypic association between MHC region SNPs in (a) United Kingdom, (b) Spanish and (c) Filipino SLE

(a) United Kingdom SLE

SNP1	SNP2	SNP3	HAPLOTYPE	F_A	F_U	OR	95% CI	P
RS1269852	RS3906272	RS3129868	CAT	0.000	0.000	NA	NA	NA
RS1269852	RS3906272	RS3129868	GAT	0.016	0.007	NA	NA	NA
RS1269852	RS3906272	RS3129868	CGT	<0.001	0.000	NA	NA	NA
RS1269852	RS3906272	RS3129868	GGT	0.123	0.106	1.2	0.93-1.54	0.15
RS1269852	RS3906272	RS3129868	CAG	0.004	0.001	NA	NA	NA
RS1269852	RS3906272	RS3129868	GAG	0.091	0.048	2.26	1.60-3.18	3.07E-06
RS1269852	RS3906272	RS3129868	CGG	0.239	0.109	2.71	2.17-3.78	8.37E-19
RS1269852	RS3906272	RS3129868	GGG	0.524	0.729	0.37	0.31-0.44	3.72E-28
	RS3906272	RS3129868	AT	0.017	0.006	NA	NA	NA
	RS3906272	RS3129868	GT	0.129	0.105	1.25	0.97-1.61	0.08
	RS3906272	RS3129868	AG	0.097	0.045	2.35	1.64-3.36	2.81E-06
	RS3906272	RS3129868	GG	0.757	0.844	0.59	0.47-0.72	6.52E-07
RS1269852		RS3129868	CT	0.004	<0.001	NA	NA	NA
RS1269852		RS3129868	GT	0.141	0.111	1.29	1.02-1.48	0.03
RS1269852		RS3129868	CG	0.247	0.109	2.74	2.18-3.44	2.96E-18
RS1269852		RS3129868	GG	0.608	0.779	0.44	0.37-0.53	7.94E-19
RS1269852	RS3906272		CA	0.004	<0.001	NA	NA	NA
RS1269852	RS3906272		GA	0.108	0.049	2.37	1.72-3.27	1.59E-07
RS1269852	RS3906272		CG	0.242	0.120	2.42	1.93-3.03	1.48E-14
RS1269852	RS3906272		GG	0.646	0.830	0.36	0.29-0.44	1.23E-23
RS1269852			C (G)		0.110	2.68	2.16-3.33	2.48E-19
	RS3906272		A (G)		0.040	2.84	2.05-3.92	2.95E-10
		RS3129868	T (G)		0.110	1.67	1.29-2.14	8.65E-05

(b) Spanish SLE

SNP1	SNP2	SNP3	HAPLOTYPE	F_A	F_U	OR	95% CI	P
RS3130490	RS3129768	RS3117213	ACA	0.004	0.000	NA	NA	NA
RS3130490	RS3129768	RS3117213	CCA	0.048	0.021	2.46	1.16-5.21	0.02
RS3130490	RS3129768	RS3117213	AAA	0.048	0.014	3.86	2.10-7.08	1.27E-05
RS3130490	RS3129768	RS3117213	CAA	0.190	0.142	1.39	1.07-1.80	0.01
RS3130490	RS3129768	RS3117213	ACC	0.013	0.001	NA	NA	NA
RS3130490	RS3129768	RS3117213	CCC	0.143	0.107	1.65	1.23-2.21	7.12E-04
RS3130490	RS3129768	RS3117213	AAC	0.048	0.027	1.97	1.12-3.47	0.02
RS3130490	RS3129768	RS3117213	CAC	0.506	0.688	0.47	0.38-0.58	3.29E-13
RS3130490	RS3129768		AC	0.017	0.001	NA	NA	NA
RS3130490	RS3129768		CC	0.193	0.128	1.78	1.36-2.38	3.50E-05
RS3130490	RS3129768		AA	0.097	0.041	2.9	1.91-4.42	6.88E-07
RS3130490	RS3129768		CA	0.693	0.830	0.45	0.35-0.57	8.80E-11
RS3130490		RS3117213	AA	0.053	0.014	3.95	2.16-7.23	8.74E-06
RS3130490		RS3117213	CA	0.238	0.164	1.49	1.16-1.91	1.65E-03
RS3130490		RS3117213	AC	0.061	0.027	1.99	1.14-3.46	0.02
RS3130490		RS3117213	CC	0.649	0.795	0.48	0.38-0.61	2.13E-09
	RS3129768	RS3117213	CA	0.053	0.020	2.57	1.22-5.42	1.34E-02
	RS3129768	RS3117213	AA	0.237	0.157	1.75	1.37-2.24	9.58E-06
	RS3129768	RS3117213	CC	0.156	0.109	1.67	1.25-2.23	5.59E-04
	RS3129768	RS3117213	AC	0.553	0.714	0.51	0.42-0.62	6.43E-11
RS3130490			A (C)		0.04	2.96	1.95-4.51	3.94E-07
	RS3129768		C (A)		0.13	1.91	1.44-2.53	7.57E-06
		RS3117213	A (C)		0.18	1.76	1.37-2.25	7.18E-06

(c) Filipino SLE

SNP1	SNP2	SNP3	HAPLOTYPE	F_A	F_U	OR	95% CI	P
RS9271366	RS2571391	RS2507987	GCA	0.017	0.019	0.84	0.20-3.56	0.82
RS9271366	**RS2571391**	**RS2507987**	**ACA**	**0.005**	**0.031**	**0.002**	**0.0001-0.06**	**2.10E-04**
RS9271366	RS2571391	RS2507987	GAA	0.262	0.201	1.48	1.04-2.11	0.03
RS9271366	**RS2571391**	**RS2507987**	**AAA**	**0.150**	**0.271**	**0.44**	**0.30-0.63**	**1.14E-05**
RS9271366	RS2571391	RS2507987	GCT	0.029	0.029	1.05	0.31-3.59	0.94
RS9271366	**RS2571391**	**RS2507987**	**ACT**	**0.026**	**0.084**	**0.14**	**0.06-0.33**	**5.45E-06**
RS9271366	**RS2571391**	**RS2507987**	**GAT**	**0.273**	**0.092**	**3.45**	**2.24-5.33**	**5.69E-11**
RS9271366	RS2571391	RS2507987	AAT	0.237	0.274	0.81	0.59-1.12	0.21
RS9271366	RS2571391		GC	0.045	0.048	0.89	0.41-1.95	0.78
RS9271366	**RS2571391**		**AC**	**0.030**	**0.116**	**0.14**	**0.07-0.29**	**1.10E-07**
RS9271366	**RS2571391**		**GA**	**0.528**	**0.296**	**2.78**	**2.02-3.83**	**3.80E-10**
RS9271366	**RS2571391**		**AA**	**0.397**	**0.539**	**0.56**	**0.42-0.75**	**7.37E-05**
	RS2571391	RS2507987	CA	0.022	0.051	0.20	0.07-0.60	3.90E-03
	RS2571391	RS2507987	AA	0.419	0.477	0.80	0.62-1.05	0.11
	RS2571391	**RS2507987**	**CT**	**0.054**	**0.113**	**0.34**	**0.19-0.62**	**3.88E-04**
	RS2571391	**RS2507987**	**AT**	**0.505**	**0.359**	**1.77**	**1.33-2.35**	**7.74E-05**
RS9271366		RS2507987	GA	0.283	0.218	1.45	1.04-2.03	0.03
RS9271366		**RS2507987**	**AA**	**0.157**	**0.307**	**0.39**	**0.27-0.56**	**3.30E-07**
RS9271366		**RS2507987**	**GT**	**0.300**	**0.118**	**4.35**	**2.73-6.93**	**6.16E-10**
RS9271366		RS2507987	AT	0.260	0.358	0.63	0.46-0.85	2.32E-03
RS9271366			G (A)		0.340	2.46	1.83-3.30	1.97E-09
	RS2571391		C (A)		0.160	0.36	0.22-0.59	6.06E-05
		RS2507987	A (T)		0.480	0.48	0.35-0.66	6.80E-06

F_A, frequency in affected SLE cases; F_U, frequency in unaffected controls; OR, odds ratio; CI, confidence interval; P, nominal p-value; Allele in parenthesis is the unassociated allele in the primary single marker analysis.

Logistic regression interaction and haplotype analysis of top three independent Spanish SNPs

A multiple logistic regression model was fitted with rs3130490, rs3129768 and rs3117213 as explanatory variables. All three SNPs had significant effects ($1.45^{*}\ 10^{-06}$, 4.04×10^{-06} and 7.28×10^{-06} respectively) with odds ratios of 2.90 (95% CI 1.88-4.48) , 1.95 (95% CI 1.47-2.59) and 1.75 (95% CI 1.37-2.24).

Testing for all three 2x2 interaction effects resulted in a significant (p=0.03, unadjusted) interaction for rs3130490*rs3129768 where the effect on the odds ratio was positive = 5.1 (95% CI 1.12-23.08). While the raw p-value for this interaction does not withstand a multiple testing adjustment (for the three interaction terms tested), a further analysis looking at all possible models and evaluating using the AIC showed that the model with rs3130490*rs3129768 plus an independent additive term for rs3117213 had the lowest AIC (1017.3). The multiple regression model with three additive terms had AIC=1022.1, while an interaction model with only rs3130490*rs3129768 had AIC=1040.6 and the null model (no markers in the model) had AIC=1101.3.

Evaluated using the BIC, we get the 'rs3130490*rs3129768 plus rs3117213' model has BIC =1060.6 while the multiple regression model with three additive terms has BIC = 1067.435. This equates to a Bayes Factor of 30 in favour of the 'rs3130490*rs3129768 plus rs3117213' model.

We also asked whether a haplotype model for rs3130490*rs3129768 with an independent effect for rs3117213 was a better fit than the interaction model for rs3130490*rs3129768 with an independent effect for rs3117213. To achieve this we phased the data using fastPHASE and coded the four haplotypes for rs3130490*rs3129768 as variables (we counted the observed number of haplotypes - 0, 1, 2). We then fitted the following three models:

1) interaction model for rs3130490*rs3129768 with an independent effect for rs3117213
2) Protective haplotype for rs3130490*rs3129768 (CA) with an independent effect for rs3117213
3) Four haplotypes for rs3130490*rs3129768 (protective used as baseline) with an independent effect for rs3117213

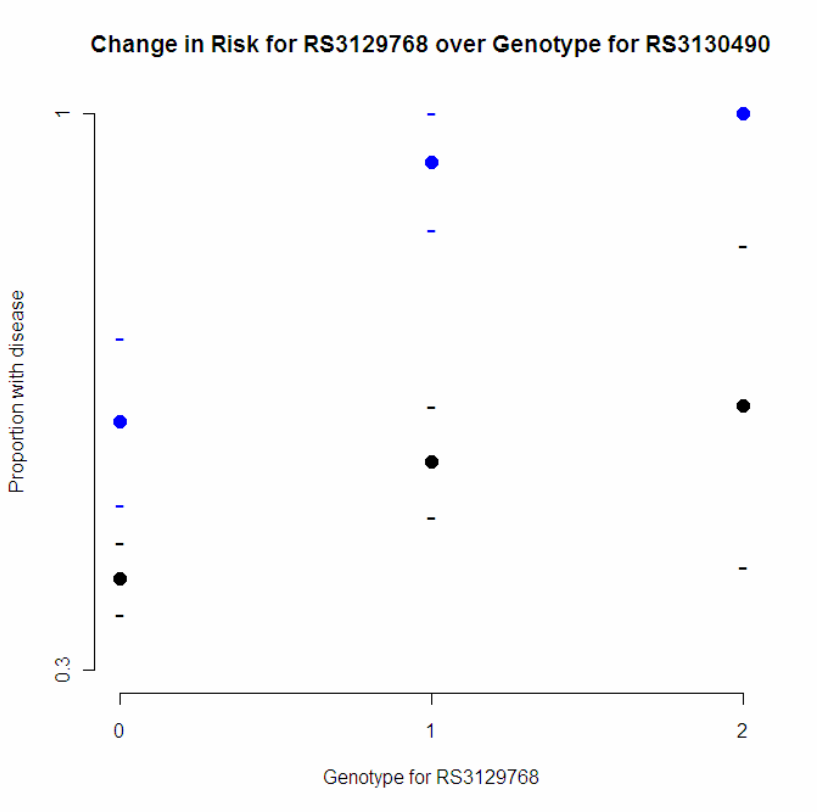
Model -1 has five parameters (intercept + rs3130490 + rs3129768 + rs3130490*rs3129768 + rs3117213), model-2 has three parameters (intercept+protective haplotype+ rs3117213) and model-3 has five parameters(intercept+ 3-non-baseline-haplotypes + rs3117213). We judged the fit of each model by the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Model-1 has the lowest AIC and BIC and the difference is >10. Supplementary Table 7 contains these model choice results. We also present the AIC and BIC for the simple 3-snp additive model and the null model (no SNPs included).

	AIC	BIC	Model Details
Model-1	**1017.3**	**1060.6**	rs3130490*rs312976 + rs3117213
Model-2	1032.8	1080.2	H(CA) + rs3117213
Model-3	1033.4	1076.8	H(CA) + H(CC) + H(AA) + H(AC) + rs3117213
Model-4	1022.1	1067.4	rs3130490 + rs312976 + rs3117213
Model-5	1101.3	1152.6	Null

Supplementary Table 9: Model choice assessment for top three independently associated Spanish SNPs

Model-1, which has an interaction between rs3130490*rs312976 and an additive effect for rs3117213 has the lowest AIC and BIC. For the haplotype models, H(CA) is a variable with values of 0, 1 or 2 if an individual has 0,1 or 2 copies of the haplotype CA. In model 3 the four haplotypes over determine the model, so we set H(AC) as the baseline.

For a graphical representation of the interaction between rs3130490*rs3129768, see Supplementary Figure 4, where the change in risk for rs3129768 is shown to increase greatly between homozygous and heterozygous individuals for rs3130490. We do not show the points for individuals homozygous for the risk rs3130490 allele as they were very close or equal to 1. In fact an interaction model for rs3130490*rs3129768 with rs3130490 as a dominant term (still with an independent additive effect for rs3117213) could also be fitted (AIC=1019.0, OR for rs3129768 when baseline for rs3130490 = 1.75, OR for rs3129768 when rs3130490 has 1-or-2 risk alleles = 9.92



Supplementary Figure 4: Interaction between rs3129768 and rs3130490: plot of disease risk over genotypes

Black points represent individuals homozygous for the non-risk rs3130490 allele, while the blue points denote individuals heterozygous for rs3130490. The relative risk over rs3129768 changes between the black (OR=1.76) and blue (OR=9.49) points (p=0.03 for change in regression coefficient). Points for individuals homozygous for the risk rs3130490 allele are not shown as they were very close or equal to 1.

References

1. de Bakker PI, McVean G, Sabeti PC, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 2006;38:1166-72.
2. Leslie S, Donnelly P, McVean G. A statistical method for predicting classical HLA alleles from SNP data. Am J Hum Genet 2008;82:48-56.
3. Horton R, Gibson R, Coggill P, et al. Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. Immunogenetics 2008;60:1-18.
4. Fernando MM, Stevens CR, Sabeti PC, et al. Identification of Two Independent Risk Factors for Lupus within the MHC in United Kingdom Families. PLoS Genet 2007;3:e192.
5. Barcellos LF, May SL, Ramsay PP, et al. High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. PLoS Genet 2009;5:e1000696.
6. Graham RR, Ortmann WA, Langefeld CD, et al. Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. Am J Hum Genet 2002;71:543-53.