UCSF

UC San Francisco Previously Published Works

Title

MHC associations with clinical and autoantibody manifestations in European SLE

Permalink

https://escholarship.org/uc/item/0q93f98p

Journal

Genes & Immunity, 15(4)

ISSN

1466-4879

Authors

Morris, DL
Fernando, MMA
Taylor, KE
et al.

Publication Date

2014-06-01

DOI

10.1038/gene.2014.6

Peer reviewed

MHC associations with clinical and autoantibody manifestations in European SLE

DL Morris[1], **MMA Fernando**[1,17], **KE Taylor**[2,17], **SA Chung**[2], **J Nititham**[2], **ME Alarcón-Riquelme**[3,4], **LF Barcellos**[5], **TW Behrens**[6], **C Cotsapas**[7], **PM Gaffney**[4], **RR Graham**[6], **BA Pons-Estel**[8], **PK Gregersen**[9], **JB Harley**[10], **SL Hauser**[11], **G Hom**[6], **CD Langefeld**[12], **JA Noble**[13], **JD Rioux**[14], **MF Seldin**[15], **Systemic Lupus Erythematosus Genetics Consortium**[16], **TJ Vyse**[1], and **LA Criswell**[2]

[1]Department of Medical & Molecular Genetics, King's College London School of Medicine, Guy's Hospital, London, UK

[2]Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California San Francisco, San Francisco, CA, USA

[3]Department of Human DNA Variability, GENYO, Centro de Genómica e Investigación Oncológica Pfizer-Universidad de Granada—Junta de Andalucía, Granada, Spain

[4]Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma, OK, USA

[5]Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, CA, USA

[6]Immunology Biomarkers Group, Genentech, South San Francisco, CA, USA

[7]Department of Neurology, Yale School of Medicine, Connecticut, CT, USA

[8]Rheumatology Service, Hospital Provincial de Rosario, Sanatorio Parque, Rosario, Argentina

[9]The Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, North Shore LIJ Health System, Manhasset, NY, USA

[10]Cincinnati Children's Hospital Medical Center and US Department of Veterans Affairs Medical Center, Cincinnati, OH, USA

[11]Department of Neurology, University of California San Francisco, San Francisco, CA, USA

Correspondence: Dr DL Morris, Department of Medical & Molecular Genetics, King's College London School of Medicine, Guy's Hospital, London, UK. david.l.morris@kcl.ac.uk.
[16]The members of the Systemic Lupus Erythematosus Genetics Consortium who participated are listed at the end of the article.
[17]These Authors contributed equally to this work.

MEMBERS OF THE SYSTEMIC LUPUS ERYTHEMATOSUS GENETICS CONSORTIUM
John B Harley, Marta E Alarcón-Riquelme, Lindsey A Criswell, Patrick M Gaffney, Chaim O Jacob, Robert P Kimberly, Kathy L M Sivils, Betty P Tsao, Timothy J Vyse and Carl D Langefeld.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

[12]Department of Biostatistical Sciences, Wake Forest University Health Sciences, Wake Forest, NC, USA

[13]Children's Hospital Oakland Research Institute, Oakland, CA, USA

[14]Department of Medicine, Université de Montréal and Research Center, Montreal Heart Institute, Montreal, QC, Canada

[15]University of California Davis, Davis, CA, USA

Abstract

Systemic lupus erythematosus (SLE) is a clinically heterogeneous disease affecting multiple organ systems and characterized by autoantibody formation to nuclear components. Although genetic variation within the major histocompatibility complex (MHC) is associated with SLE, its role in the development of clinical manifestations and autoantibody production is not well defined. We conducted a meta-analysis of four independent European SLE case collections for associations between SLE sub-phenotypes and MHC single-nucleotide polymorphism genotypes, human leukocyte antigen (HLA) alleles and variant HLA amino acids. Of the 11 American College of Rheumatology criteria and 7 autoantibody sub-phenotypes examined, anti-Ro/SSA and anti-La/SSB antibody subsets exhibited the highest number and most statistically significant associations. HLA-DRB1*03:01 was significantly associated with both sub-phenotypes. We found evidence of associations independent of MHC class II variants in the anti-Ro subset alone. Conditional analyses showed that anti-Ro and anti-La subsets are independently associated with HLA-DRB1*0301, and that the HLA-DRB1*03:01 association with SLE is largely but not completely driven by the association of this allele with these sub-phenotypes. Our results provide strong evidence for a multilevel risk model for HLA-DRB1*03:01 in SLE, where the association with anti-Ro and anti-La antibody-positive SLE is much stronger than SLE without these autoantibodies.

Keywords

Sub-phenotype analysis; MHC; meta-analysis; genetics; systemic lupus erythematosus; Europeans

INTRODUCTION

Systemic lupus erythematosus (SLE; OMIM 152700) is a complex autoimmune disease that can affect multiple organ systems. Processes involving both the innate and adaptive immune systems contribute to its development.[1] The disease is clinically heterogeneous, and affected individuals only need 4 out of 11 of the American College of Rheumatology (ACR) criteria to be classified as having SLE. Although patients may differ in their clinical manifestations, patients do share a propensity to develop autoantibodies directed against nucleic acids and associated nuclear and cellular proteins.

There is overwhelming evidence of a genetic component to SLE risk with higher concordance rates observed between monozygotic twins (20–40%) compared with dizygotic twins (2–5%).[2] The familial aggregation for SLE (sibling risk ratio, $\lambda_s = 8–29$)[2,3] is higher than other autoimmune diseases, and the estimate of heritability is approximately 66%.[4]

Genetic association studies of SLE have been successful in identifying multiple loci.[5–11] However, relatively few studies have investigated the genetic association with specific SLE sub-phenotypes.[12–15] These studies focused mainly on major histocompatibility complex (MHC) class II genes, and found evidence that class II alleles such as *HLA-DRB1*03:01* are associated with auto-antibody production.[13] Our study substantially expands this work by not only analysing imputed classical human leukocyte antigen (HLA) alleles, but also examining variant HLA amino-acid positions in conjunction with single-nucleotide polymorphism (SNP) genotypes across the extended MHC region (chromosome 6: 26–34 Mb). Our aim was to discover genetic loci within the MHC region that are associated with specific clinical and/or immunological manifestations within SLE cases and hence to find evidence of genetic variants that may drive specific forms of the disease. For complex heterogeneous diseases such as SLE, comprehensive sub-phenotype studies are critical in order to understand how previously identified genetic associations contribute to disease pathogenesis and specific disease manifestations.

RESULTS

Study sample

For this study, we collected genetic and sub-phenotype data from 3070 SLE cases of European descent characterized in four genetic association studies of SLE. These SLE cases were previously examined in a large meta-analysis that examined the association between MHC genetic variation and SLE susceptibility.[16] Table 1 describes the genotyping platform, number of genotyped MHC SNPs, and sample size of each case collection in the study. Given the strong genetic associations observed with anti-Ro/SSA and anti-La/SSB autoantibody production described below, Table 1 also provides the frequency of these antibodies for each case collection. Genetic (SNP) imputation was performed previously[16] for each case collection, resulting in a total of 7119 SNPs common between the four collections. In addition, classical HLA class I and II alleles as well as their corresponding variant amino acids (AAs, see Materials and Methods) were imputed and analysed.

Selection of sub-phenotypes for analysis

We examined the 11 ACR classification criteria[17] and 7 SLE-related autoantibodies (anti-double-stranded DNA, anti-Sm, anti-RNP, anti-Ro/SSA, anti-La/SSB, anti-cardiolipin IgG and anti-cardiolipin IgM) as candidate sub-phenotypes for this study. Single-marker associations for each candidate sub-phenotype with all variants were assessed using logistic regression adjusted for population substructure and case collection (Supplementary Table 1). We analysed 7656 variants in total (7119 SNPs, 199 HLA alleles and 338 HLA amino-acid positions (see methods)). The specific sub-phenotypes comprising anti-Ro and anti-La antibodies demonstrated by far the most associations: 1635 and 1828 variants, respectively, at $P<0.00001$. For all other sub-phenotypes, there were fewer than 30 variants that were significant at this level. Thus, we targeted anti-Ro and anti-La antibody subsets for detailed investigation as they have the strongest evidence for a genetic aetiology.

Anti-Ro antibody sub-phenotype

Stepwise conditional analysis—The most associated marker (in terms of *P*-value as a single marker) was the class III SNP rs3129962 in *BTNL2* ($P = 9.47 \times 10^{-27}$; odds ratio (OR) = 2.44, 95% confidence interval (CI) = 2.08–2.94; Table 2A). This marker is in linkage disequilibrium (LD) with *HLA-DRB1*03:01* ($R^2 = 0.84$, $D' = 0.99$). When conditioning on this SNP as a covariate in forward stepwise regression, the next most associated marker was the class II SNP, rs9271731, between *HLA-DRB1* and *HLA-DQA1* ($P = 9.56 \times 10^{-07}$; OR = 1.54, 95% CI = 1.30–1.85). This SNP is in LD with *HLA-DRB1*15:01* ($R^2 = 0.72$, $D' = 1$). When using rs9271731 as an additional covariate, one further association signal was detected at the class II SNP, rs3957146, between *HLA-DQB1* and *HLA-DQA2* ($P = 5.70 \times 10^{-06}$; OR = 0.52, 95% CI = 0.39–0.69). Of note, the effect sizes (ORs) and *P*-values that we present here are estimated from the multivariate models returned by stepwise regression (columns 2–3 in Table 2). The association results for a given variant from single marker analyses can be seen in the last two columns of Table 2.

The most associated amino acid (AA) was at position 77 in HLA-DRB1 with the common AA threonine having a protective effect ($P = 2.72 \times 10^{-13}$; OR = 0.49; 95% CI = 0.41–0.60). *HLA-DRB1*03:01* and *HLA-DRB1*03:02* encode the single alternative AA, asparagine, ($R^2 = 1$). *HLA-DRB1*03:02* is not significantly associated with this sub-phenotype ($P = 0.37$) possibly because of this allele being rare (frequency of 0.01% in our data). We cannot be certain that this lack of association applies to the general population and this needs to be investigated to address this uncertainty. All other *HLA-DRB1* alleles code for threonine. The single marker *P*-value for this AA was very close to that of the most strongly associated SNP (see last column in Table 2). Therefore, we ran a stepwise regression starting from this marker. When conditioning on this AA, the next most associated marker was the class II SNP, rs9271731, between *HLA-DRB1* and *HLA-DQA1* ($P = 4.5 \times 10^{-08}$; OR = 1.63, 95% CI = 1.37–1.95). When using rs9271731 as an additional covariate, one further association signal was detected at the class III SNP, rs3130781, in *DPCR1* ($P = 1.76 \times 10^{-05}$; OR = 1.44, 95% CI = 1.22–1.71). The SNP rs3130781 is in LD with *HLA-DRB1*03:01* ($R^2 = 0.29$, $D' = 0.64$) and *HLA-B*08:01* ($R^2 = 0.29$, $D' = 0.72$). One final association signal was detected at *HLA-DQB1*03:02* ($P = 2.49 \times 10^{-05}$; OR = 0.56, 95% CI = 0.42–0.73). The results from this analysis can be seen in Table 2B.

Owing to the correlation between the most associated SNPs with known associated *HLA-DRB1* alleles (rs3129962 tags *HLA-DRB1*03:01*/Thr77 in DRB1 ($R^2 = 0.84$); rs9271731 tags *HLA-DRB1*15:01*), we performed stepwise regression conditioning on these HLA alleles as covariates. When conditioning on *HLA-DRB1*03:01* and *HLA-DRB1*15:01*, the next most associated marker (rs9275582) was in class II between *HLA-DQB1-HLA-DQA2* ($P = 2.99 \times 10^{-06}$; OR = 0.61; 95% CI = 0.5–0.76). The most significant HLA allele was *HLA-DQB1*03:02*, which is in LD with rs9275582 ($R^2 = 0.29$, $D' = 0.80$). These two sets of results can be seen in Tables 2C and 2D. We note that *HLA-DQB1*03:02* is in LD ($R^2 = 0.58$) with rs3957146 (the third associated SNP in the first stepwise regression presented in Table 2).

A simple stepwise regression analysis including only AA variants indicated associations with Thr77, Leu67 and Gln96 in *HLA-DRB1* (Table 2E). The *HLA-DRB1* AA glutamine at position 96 is in LD with *HLA-DRB1*15:01* (R^2 =0.82, D' =1.00).

Model choice using the bayesian information criterion (BIC)—Owing to the extended LD, an analysis of the MHC using stepwise regression to find evidence for multiple independently associated variants can lead to many models depending on the first marker conditioned on (used as a covariate for further association analysis). This was discussed previously[16] and here we also used the BIC as an aid to model choice; the lower the BIC, the better fit the model is to the data (see methods). In our analysis of sub-phenotype data, there was not much difference between models A, C, D and E in Table 2 in terms of the BIC, which represents the relative belief in a model given the data. However, model B, which began the forward stepwise regression with threonine at position 77 in *HLA-DRB1,* had the lowest BIC. This model does have one more term than the other four models. Our extended model search (see methods) did not result in a model with a lower BIC.

Haplotype analysis—There are two main extended MHC haplotypes associated with SLE in northern Europeans that contain the class II alleles *HLA-DRB1*03:01* and *HLA-DRB1*15:01*.[18] These extended haplotypes are comprised of the following HLA alleles: *HLA-A*03:01—HLA-B*07:02—HLA-C*07:02—**HLA-DRB1*15:01**—HLA-D QA1*01:02 —HLA-DQB1*06:02* and *HLA-A*01:01—HLA-B*08:01—HLA-C*07:01—**HLA-DRB1*03:01**—HLA-DQA1*05:01—HLA-DQB1*02:01*. We tested for association of these extended haplotypes with anti-Ro antibody status with the hypothesis that the association signals at *HLA-DRB1*03:01 and HLA-DRB1*15:01* are independent of these haplotypes. We observed significant effects for both haplotypes (*HLA-DRB1*03:01: P* =1.02×10^{-12}, OR =2.17; *HLA-DRB1*15:01: P* =0.02, OR =1.71). We found evidence that *HLA-DRB1*03:01* is associated independently of the *HLA-B*08:01-DRB1*03:01* haplotypic background ($P = 3.05 \times 10^{-07}$), whereas we fail to find evidence that *HLA-DRB1*15:01* (P = 0.17) is independent of the *HLA-B*07:02-DRB1*15:01* haplotype.

Anti-La antibody subphenotype

Stepwise conditional analysis—The most strongly associated marker with the anti-La autoantibody sub-phenotype was the SNP rs2894254, in the class III region (P =3.40×10^{-30}; OR = 3.38, 95% CI =2.74–4.16). This SNP is in LD (R^2 =0.84, D' = 0.99) with *HLA-DRB1*03:01*. We do not find further associations when conditioning on this SNP as a covariate. However, if we condition on *HLA-DRB1*03:01*, we find a further association with rs9268832, located between *HLA-DRA* and *HLA-DRB5* in class II (P =6.53×10^{-06}; OR =1.64; 95% CI =1.32–2.04). Results from these two models can be seen in Table 3. The *HLA-DRB1* AA threonine at position 77 was observed to have a protective effect, consistent with the anti-Ro analyses. However, this AA was not the most associated marker (P =2.4×10^{-28}). Conditioning on Thr77, we find an additional association with rs2227139, located in *HLA-DRA* in class II (P =6.47×10^{-06}; OR = 1.64; 95% CI =1.32–2.04). The SNP, rs2227139, is in LD with rs9268832 ($R^2 = 0.91$, D' =0.96).

Model choice using the BIC—As with the analysis of anti-Ro, we used the BIC as an aid to model comparison. The model including AA variation has the lowest BIC (model C in Table 3) but is only slightly lower than the model conditioning on *HLA-DRB1*03:01*. Therefore, we cannot choose between the AA and the HLA allele as the best explanation for the data; however, conditional on either of these we find an independent association in class II. Both of these models have a lower BIC than model A, which only has the single most associated SNP (rs2894254). These data therefore favour two independent associations in class II, one of which is most likely *HLA-DRB1*03:01* or the *HLA-DRB1* AA threonine at position 77. Our extended model search (see methods) returned the same models as in Table 3.

Haplotype analysis—We observed significant effects for the *HLA-DRB1*03:01* haplotype but not the *HLA-DRB1*15:01* haplotype with anti-La antibody status (*HLA-DRB1*03:01: P* $=1.19 \times 10^{-16}$, OR $=3.12$; *HLA-DRB1*15:01: P* $=0.63$). We found evidence that *HLA-DRB1*03:01* is associated independently of the *HLA-B*08:01-DRB1*03:01* haplotype ($P = 6.42 \times 10^{-13}$).

Independence of anti-Ro and anti-La autoantibody associations with *HLA-DRB1*03:01*

Thus far, we have observed strong evidence of association between *HLA-DRB1*03:01* and both anti-Ro and anti-La autoanti-body subsets. As these two phenotypes are correlated (R^2 $=0.27$), we performed conditional analyses to determine whether the associations for each sub-phenotype were independent of each other. We performed logistic regression analysis with each sub-phenotype as an outcome and the other sub-phenotype as a covariate. Table 4 displays the sample sizes and *HLA-DRB1*03:01* frequencies for these case only analyses.

When conditioning on anti-La as a covariate, *HLA-DRB1*03:01* continues to be strongly associated with anti-Ro antibody status ($P = 1.23 \times 10^{-07}$, OR $=1.60$ 95% CI $=1.02-2.54$). Also, when conditioning on anti-Ro, *HLA-DRB1*03:01* continues to be strongly associated with anti-La antibody status ($P =1.66 \times 10^{-12}$, OR $=2.57$ 95% CI $=1.98-3.34$). To assess the robustness of these conditional regression results, we examined the anti-Ro association in only anti-La-negative cases and found that *HLA-DRB1*03:01* was still strongly associated with anti-Ro ($P =6.79 \times 10^{-07}$, OR $=1.58$ 95% CI $=1.32-1.89$). In anti-La antibody-positive SLE cases, *HLA-DRB1*03:01* is weakly associated with anti-Ro ($P = 0.055$, OR $= 2.37$ 95% CI $=0.98-5.74$). We performed the same analyses for the anti-La antibody subset, stratifying on the anti-Ro phenotype. In anti-Ro-positive SLE cases, *HLA-DRB1*03:01* is strongly associated with anti-La ($P =6.18 \times 10^{-12}$, OR $=2.81$ 95% CI $=2.09-3.77$). Among anti-Ro-negative SLE cases, *HLA-DRB1*03:01* is weakly associated with anti-La ($P =0.06$, OR $=1.96$ 95% CI $=0.97-3.73$). Therefore, we conclude that the association signal for *HLA-DRB1*03:01* with anti-La is not due to this sub-phenotype's correlation with anti-Ro, and vice-versa.

The *HLA-DRB1*03:01* association with SLE susceptibility is independent of the association with anti-Ro and anti-La antibody subsets

We have provided strong evidence for the association between *HLA-DRB1*03:01* and both anti-Ro and anti-La antibody subsets. This *HLA-DRB1* allele has been consistently and

strongly associated with SLE susceptibility in European populations,[16] and this is confirmed in our current data ($P = 3.38 \times 10^{-49}$; OR =1.86 95% CI =1.71–2.02). However, the association of *HLA-DRB1*03:01* with anti-Ro/anti-La antibody subsets and SLE susceptibility may not be independent—the *DRB1*03:01* association with SLE may be purely secondary to its association with anti-Ro and anti-La antibody status.

If the association between *HLA-DRB1*03:01* and SLE status is not driven entirely by sub-phenotype then one could hypothesize a three-level model of disease type (unaffected; sub-phenotype-negative case; sub-phenotype-positive case) based on increasing *HLA-DRB1*03:01* frequency. Figure 1 plots the change in *HLA-DRB1*03:01* dosage over levels of disease; the average dosage appears to increase over all three levels. Therefore, we examined (see methods) the hypothesis that the *HLA-DRB1*03:01* association with anti-Ro and anti-La antibody sub-phenotypes explains the association of *DRB1*03:01* with SLE in general. We also tested whether the risk was additive over the three levels of disease.

Anti-Ro antibody sub-phenotype

We found a significant difference in *HLA-DRB1*03:01* dosage between healthy controls and anti-Ro antibody negative cases ($P = 1.97 \times 10^{-14}$). The estimated change in dosage was 0.1 (95% CI =0.08–0.13), equivalent to a change in allele frequency of 0.05 (95% CI =0.04–0.06).

We also found a significant increase in dosage between anti-Ro-negative cases and anti-Ro positive cases ($P = 2.97 \times 10^{-33}$). The estimated change in dosage (see Table 5) is 0.27 (95% CI =0.22–0.31), equivalent to a change in frequency of 0.13 (95% CI = 0.11–0.16).

We found evidence against the hypothesis that the increase in dosage is additive over the three disease levels ($P = 0.008$). Our final test against the additive model implies that the difference in *HLA-DRB1*03:01* dosage between anti-Ro(−)/anti-Ro(+) status (increase of 0.27) in the cases is more than double that of the difference between cases and healthy controls (increase of 0.10).

Anti-La antibody subphenotype

We found a significant difference in *HLA-DRB1*03:01* dosage between healthy controls and anti-La-negative cases ($P = 3.57 \times 10^{-25}$). The estimated change (see Table 5) in dosage is 0.13 (95% CI = 0.11–0.15), equivalent to a change in frequency of 0.06 (95% CI =0.05–0.08).

We also found a significant increase in dosage between anti-La-negative and anti-La-positive cases ($P = 2.45 \times 10^{-39}$). The estimated change in dosage (see Table 5) is 0.41 (95% CI =0.35–0.47), equivalent to a change in frequency of 0.21 (95% CI =0.18–0.24).

We found evidence against the hypothesis that the increase in dosage is additive over the three disease levels ($P = 1.5 \times 10^{-04}$). Table 5 displays the effect sizes and *P*-values for this analysis. Our final test against the additive model implies that the difference in *HLA-DRB1*03:01* dosage between anti-La(−)/anti-La(+) status (increase of 0.41) in the cases is more than triple that of the difference between cases and healthy controls (increase of 0.13).

Double positive and double negative anti-Ro and anti-La antibody sub-phenotypes

Our study was large enough to determine whether the frequency of *HLA-DRB1*03:01* differs between SLE cases who are double negative for anti-Ro and anti-La antibodies (*N* =1781) and healthy controls (*N* =9782). It is known that these antibodies are present in approximately 2% of the healthy population; however, we do not have this phenotype data for the controls. The following results therefore assume that all controls are negative for antinuclear antibodies. We found a significant association of *HLA-DRB1*03:01* with the double negative SLE cases/healthy controls status (OR =1.49, 95% CI =1.35–1.65; *P* = 2.23 $\times 10^{-14}$). Further analysis demonstrated a stronger association with the double positive (*n* =259)/double negative SLE case status (OR =3.71, 95% CI = 2.97–4.64; *P* =2.00 $\times 10^{-16}$). To test whether these two odds ratios differ, we ran the same analysis for a three-stage risk model as we did for anti-Ro and anti-La antibody subsets separately (see above, Table 5 results). We found very strong evidence against the hypothesis that the increase in dosage is additive over the three disease levels (*P* = 6.65 $\times 10^{-06}$). The non-additive effect leads to a very large odds ratio between double positive SLE cases and healthy controls, which we found to be 5.27 (95% CI = 4.31–6.44; *P* =3.14 $\times 10^{-59}$; Figure 1).

DISCUSSION

Our results confirm, in the largest SLE sub-phenotype genetic association study to date, that the often replicated genetic association at *HLA-DRB1*03:01* does not just influence SLE susceptibility but is also associated with anti-Ro and anti-La autoantibody production. For the first time, we have shown that *HLA-DRB1*03:01* is associated with SLE *per se*, independent of anti-Ro and anti-La antibody subsets. These data implicate *HLA-DRB1*03:01* and variants in LD with it in the predisposition to anti-Ro and anti-La autoantibody production as well as processes outside of this manifestation.

We do not find conclusive evidence that variant HLA AAs explain the majority of the MHC association signal in anti-Ro and anti-La autoantibody subsets in SLE. This is largely due to the confounding effects of extended LD displayed by the associated DRB1*03:01 and to a lesser extent, the DRB1*15:01 haplotypes in our study cohorts. These results contrast with those of a recent study in anti-CCP-positive rheumatoid arthritis, where five HLA AA variants were suggested to largely explain the MHC association with disease status.[19] In this case, the disease-associated variants generally reside on a diversity of haplotypes. Studies in other autoimmune/inflammatory diseases have either not shown robust association signals with variant HLA AA data or like the present study have shown association with AAs in strong LD with previously associated HLA alleles. It may be that HLA amino association signals are more complex than the single-variant testing method we and others have used.

Limitations of the present study include the heterogeneity in autoantibody testing procedures and sub-phenotype data collection between the four studies. As a result, data were tabulated and analysed in an essentially binary format (that is, individual cases were classified as positive, negative or missing for each trait), to allow meta-analysis. However, in so doing, a degree of noise is inevitable, which would reduce our power to detect true association signals particularly in the less common sub-phenotypes. We were also limited by the imputation required to analyse a consistent set of SNPs across studies and the reliance on

HLA imputation. In addition, we are constrained in our conclusions on differences in results for anti-Ro and anti-La antibody subsets given the much smaller sample size available for the anti-La phenotype. Thus, we have confined some of our analyses to the most robust association; that of *HLA-DRB1*03:01* with both anti-Ro and anti-La antibody sub-phenotypes. We must also allow for the possibility that associations with *HLA-DRB1*03:01* could exist with other SLE subsets that overlap with anti-Ro/La, but have not been detected in our study. This highlights the need for extension of this work to other cohorts with sub-phenotype data in order to increase sample size and power across as wide a range of phenotypes as possible.

In both anti-Ro and anti-La sub-phenotypes, we find evidence of secondary independent associations in the class II region of the MHC after conditioning on *HLA-DRB1*03:01*, and we find additional signals in class II and class III for anti-Ro. We have shown that the association of *HLA-DRB1*03:01* with anti-Ro antibody status is independent of the association with anti-La and vice-versa. We have also shown that the association between SLE case/healthy control and *HLA-DRB1*03:01* is not purely due to the association with anti-Ro and anti-La antibody sub-phenotypes. This implies a three-level model of risk for increasing dosage of *HLA-DRB1*03:01,* where the frequency of this allele is higher in anti-Ro-negative cases than in healthy controls and higher still in anti-Ro-positive cases than anti-Ro-negative cases. The same is true for anti-La. In fact, we find very strong evidence that the *HLA-DRB1*03:01* risk of anti-Ro/anti-La double positive within SLE patients is much greater than the risk of anti-Ro/anti-La double negative (other lupus phenotypes without these anti-bodies present) in the general population. We can conclude that the association of *HLA-DRB1*03:01* with SLE is driven to a large extent but not entirely by anti-Ro and anti-La auto-antibody sub-phenotypes.

Although we do find evidence of an independent class III association with anti-Ro, there is some uncertainty. We find a significant association with the class III SNP rs3130781 conditional on the AA Thr77-DRB1. However, when conditioning on the markers in model C in Table 2 (*HLA-DRB1*03:01* +*HLA-DRB1*15:01* + rs9275582; BIC =2829.8) in a forward stepwise regression, the association with rs3130781 is not significant ($P = 4.2 \times 10^{-05}$). This is also the case for model D in Table 2 (*HLA-DRB1*03:01* + *HLA-DRB1*15:01* +*HLA-DQB*03:02;* BIC =2829.6). So conditional on *HLA-DRB1*03:01* and *HLA-DRB1*15:01,* we find an independent association in class II but not class III. However, we did consider conditioning on *HLA-DRB1*03:01* alone, where a stepwise regression returned a class II SNP (rs9271731; R^2 with HLA-DRB1*15:01 =0.72) and the class III SNP rs3130781. This model has a BIC =2929.00. Hence, there is uncertainty as to whether there is an independent class III effect when conditioning on *HLA-DRB1*03:01*; all three models fit the data equally well (not much difference in the BIC). Nevertheless, the best model in Table 2 does suggest that there is an independent class III effect conditional on the class II AA Thr77-DRB1. This model has a much lower BIC than any others. There is some evidence, therefore, of a class III association with anti-Ro; however, we believe that more data, and ideally across diverse populations (to help remove effects due to LD), are required to be more definitive about this.

The results of the present study while enlightening are confounded by the strong and extended LD present on the principally associated *HLA-DRB1*03:01* and *HLA-DRB1*15:01* haplotypes. Complementary studies in accurately phenotyped southern European and non-European SLE cohorts, which show haplotypic diversity at the MHC, will allow refinement of the sub-phenotype association signals found in the predominantly northern European populations studied thus far.[20] These efforts may still yield association intervals that harbour several genes/variants. Therefore, future work will inevitably require re-sequencing, transcriptomic and epigenetic studies in order to tease out these complex association signals.

MATERIALS AND METHODS

Study design

This study is a meta-analysis of four studies taken from work described in a previous paper.[16] We only included four of the six previous studies in this work as sub-phenotype data were not available from the other two studies (named 'Affy500K' and 'Affy100K' in the previous paper). We refer to the previous meta-analysis of SLE case–control data as the 'parent study' in this work. The number of SLE cases and controls in this paper for the four included studies are the same as in the parent study, and quality control (QC) procedures for these data are described in full in the previous paper, including tests for relatedness and adjustments for population structure. We include some QC descriptions below for clarity in this paper.

QC and imputation

SNPs—We only analysed SNPs that passed QC in our previous paper,[16] which utilized these data: 90% genotyping for all subjects and SNPs, minor allele frequency >0.01 and Hardy–Weinberg equilibrium (false discovery rate of 0.05).

HLA imputation—We imputed HLA genotypes using HLA*IMP V2.[21] Only genotyped SNPs in each case collection were used for this imputation. We used posterior probabilities of HLA genotypes, rather than most likely genotypes, in order to allow for uncertainty in imputation. From these probabilities, we calculated dosages for each allele (expected number of alleles $0 < x < 2$). We had HLA-DRB1 typed data in two studies: the 'Illumina Combined MHC panel' study ($N = 1608$) and the 'Illumina Custom panel' study ($N = 605$). This allowed for assessment of accuracy, which for the two main reported positive associations in this paper were as follows: for *HLA-DRB1*03:01,* we achieved sensitivity of 0.992/0.999 and specificity of 0.995/0.993 for the Illumina Combined MHC panel and Illumina Custom panel' respectively. For *HLA-DRB1*15:01,* we achieved sensitivity of 0.980/0.992 and specificity of 0.996/0.997.

AA translation

AA sequences for each HLA allele were extracted from the European Bioinformatics Institute HLA database (http://www.ebi.ac.uk/ipd/imgt/hla/). HLA allele dosages were converted to AA dosages at each position; the dosage for a particular amino acid 'A' at position 'p' would be the sum of HLA alleles' dosage that coded for amino acid 'A' at

position 'p'. The total dosage for each position is therefore equal to 2 and this total is split between each possible AA at the position.

We had data at 338 AA positions that had variable AAs (HLA-A =67, HLA-B = 75, HLA-C =71, HLA-DPB1 = 21, HLA-DQA1 = 41, HLA-DQB1 =61, HLA-DRB1 = 52). Owing to multiple possible AAs at each position, we actually had 1255 possible position/AA variants in total.

Adjustment for population structure—We analysed the data with the statistical computing language R[22] using logistic regression. All analyses were adjusted for ancestry utilizing the first principal component (PC) or percentage of northern European ancestry, as previously described[16] and included a covariate for project. As the PCs were computed specifically for each case collection, we also included interaction terms between projects and ancestry to allow for different effect sizes in the adjustment for population structure.

Single-marker analysis of candidate sub-phenotypes and analysis of SLE as a simple disease outcome—We examined the 11 ACR criteria[17] and presence of 7 SLE-related auto-antibodies (anti-Ro/SSA, anti-La/SSB, anti-double-stranded DNA, anti-RNP, anti-Sm and anticardiolipin IgG and IgM) as candidate sub-phenotypes for detailed analysis. To determine which sub-phenotypes were most strongly influenced by genetic variation in the MHC, we tested each sub-phenotype for association with all variants (SNPs, HLA alleles and HLA AAs) in single-variant association tests using logistic regression adjusted for population substructure and case collection. We also tested the association between markers and SLE as a simple disease outcome for the four studies considered here. Results for association with *HLA-DRB1*03:01* are discussed in the beginning of the section titled 'The *HLA-DRB1*03:01* association with SLE susceptibility is independent of the association with anti-Ro and anti-La antibody subsets'.

Conditional association analysis of anti-Ro and anti-La—Owing to numerous single-marker associations within the extended LD of the MHC, we used conditional analyses to narrow these associations to those with the best evidence for strength and independence. All analyses utilized logistic regression with ancestry and project covariates (see above) and were halted when the evidence for association with a new term was $P > 3 \times 10^{-05}$. We performed classic forward stepwise regression, conditioning on the top variant to find the second variant, and so on.

A simple forward stepwise approach can lead to over-fitting (selecting many correlated markers) and the results may be misleading because of selected markers potentially tagging two or more independently associated markers.[16] Therefore, we also performed a model search using the BIC[16] as the inclusion metric in a stepwise regression using the R[22] 'step()' function, first starting with no prior model (other than covariates above) and also starting from *HLA-DRB1*03:01* and *HLA-DRB1*15:01* as initial model terms. Although BIC optimization was used to select model terms, we terminated the selection when it would result in a term with $P > 3 \times 10^{-5}$. The BIC[23,24] is a penalized likelihood model choice criterion similar to the Akaike Information Criterion[24] except there is a stronger penalty for additional model parameters that increases with sample size. The BIC is therefore more

conservative and favours smaller models than the Akaike Information Criterion. As with the Akaike Information Criterion, the smaller the BIC the better the model is judged to fit the data.

Haplotype analysis of anti-Ro and anti-La—Given the high degree of correlation between the associated variants identified from the model searches described above, we conducted a haplotype analysis of these variants using PLINK[25] using the best-guess genotypes estimated from HLA*IMP2. We used PLINK to phase haplotypes and perform multivariate logistic regression where terms are haplotypes rather than individual variants, optionally controlling for individual variants or haplotypes.

Multiple testing—In the MHC, a Bonferroni adjustment for multiple testing is inappropriate because of the extensive LD and hence correlated variants. In order to determine the number of independent variants, we performed a PC analysis of all SNPs. In our data, we found that 374 PCs had eigenvalues >1 and these PCs explained 96% of the variance. Thus, we used a multiple-testing threshold of $P<0.01/374 = 3 \times 10^{-5}$.

Testing for independence between the SLE association and sub-phenotype association with HLA-DRB1*03:01—We fitted a linear regression model with dosage for *HLA-DRB1*03:01* as the outcome and both case/control status and sub-phenotype status as explanatory variables. We therefore tested each effect conditional on the other. A significant association for case/control status conditional on sub-phenotype implies that we reject the hypothesis that sub-phenotype is solely driving the case/control association. This is equivalent to setting the three-level status as a factor in the regression in terms of model fit. But rather than obtaining an estimate of dosage change between healthy controls and sub-phenotype positive as we would in a three-level factor (where the baseline is healthy control), we get an estimate of change between sub-phenotype positive and sub-phenotype negative. In both models, we also get an estimate of change between healthy controls and sub-phenotype negative.

Furthermore, we tested the hypothesis that the increase in dosage is additive over the three disease levels (Healthy-Control Case sub-phenotype negative/Case sub-phenotype positive). This is achieved by fitting a model with an additive effect for dosage over the three phenotype levels. This additive model is nested within our model used to test independence of sub-phenotype association with SLE-case/healthy control, so we performed a likelihood ratio test. A rejection of this additive model, in favour of the three-level factor model (described in the previous paragraph), is evidence that the change in dosage over sub-phenotype within cases is different than the change in dosage between healthy controls and SLE without the sub-phenotype.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Web Resources: The URLs for data presented herein are as follows:

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

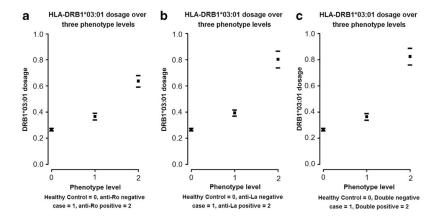Online Mendelian in MAN (OMIM): http://www.omim.org

References

1. Guerra SG, Vyse TJ, Graham DSC. The genetics of lupus: a functional perspective. Arthritis Res Ther. 2012; 14:211. [PubMed: 22640752]

2. Deapen D, Escalante A, Weinrib L, Horwitz D, Bachman B, Royburman P, et al. A revised estimate of twin concordance in systemic lupus-erythematosus. Arthritis Rheum. 1992; 35:311–318. [PubMed: 1536669]

3. Alarcon-Segovia D, Alarcon-Riquelme ME, Cardiel MH, Caeiro F, Massardo L, Villa AR, et al. Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1177 lupus patients from the GLADEL cohort. Arthritis Rheum. 2005; 52:1138–1147. [PubMed: 15818688]

4. Lawrence JS, Martins CL, Drake GLA. Family survey of lupus-erythematosus.1. Heritability. J Rheumatol. 1987; 14:913–921. [PubMed: 3430520]

5. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, 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; 40:204–210. [PubMed: 18204446]

6. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, et al. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. New Engl J Med. 2008; 358:900–909. [PubMed: 18204098]

7. Yang W, Shen N, Ye DQ, Liu Q, Zhang Y, Qian XX, et al. Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. PLoS Genetics. 2010; 6:e1000841. [PubMed: 20169177]

8. Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, Hu Z, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genetics. 2009; 41:1234–1237. [PubMed: 19838193]

9. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genetics. 2009; 41:1228–1233. [PubMed: 19838195]

10. Graham DSC, Morris DL, Bhangale TR, Criswell LA, Syvanen AC, Ronnblom L, et al. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with Systemic Lupus Erythematosus. PLoS Genetics. 2011; 7:e1002341. [PubMed: 22046141]

11. Li R, Yang W, Zhang J, Hirankarn N, Pan HF, Mok CC, et al. Association of CD247 with systemic lupus erythematosus in Asian populations. Lupus. 2012; 21:75–83. [PubMed: 22004975]

12. Taylor KE, Chung SA, Graham RR, Ortmann WA, Lee AT, Langefeld CD, et al. Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. PLoS Genetics. 2011; 7:e1001311. [PubMed: 21379322]

13. Hamilton RG, Harley JB, Bias WB, Roebber M, Reichlin M, Hochberg MC, et al. 2 Ro (Ss-a) autoantibody responses in systemic lupus-erythematosus—correlation of Hla-Dr/Dq specificities with quantitative expression of Ro (Ss-a) autoantibody. Arthritis Rheum. 1988; 31:496–505. [PubMed: 2451920]

14. Arnett FC, Hamilton RG, Reveille JD, Bias WB, Harley JB, Reichlin M. Genetic studies of Ro (SS-A) and La (SS-B) autoantibodies in families with systemic lupus erythematosus and primary Sjogren's syndrome. Arthritis Rheum. 1989; 32:413–419. [PubMed: 2706027]

15. Harley JB, Sestak AL, Willis LG, Fu SM, Hansen JA, Reichlin M. A model for disease heterogeneity in systemic lupus erythematosus. Relationships between histocompatibility antigens, autoantibodies, and lymphopenia or renal disease. Arthritis Rheum. 1989; 32:826–836. [PubMed: 2787639]

16. Morris DL, Taylor KE, Fernando MM, Nititham J, Alarcon-Riquelme ME, Barcellos LF, et al. Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans. Am J Hum Genet. 2012; 91:778–793. [PubMed: 23084292]

17. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1997; 40:1725–1725. [PubMed: 9324032]

18. Fernando MM, Stevens CR, Sabeti PC, Walsh EC, McWhinnie AJ, Shah A, et al. Identification of two independent risk factors for lupus within the MHC in United Kingdom families. PLoS Genetics. 2007; 3:e192. [PubMed: 17997607]

19. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee H-S, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genetics. 2012; 44:291–296. [PubMed: 22286218]

20. Fernando MMA, Freudenberg J, Lee A, Morris DL, Boteva L, Rhodes B, et al. Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at MSH5, HLA-DPB1 and HLA-G. Ann Rheum Dis. 2012; 71:777–784. [PubMed: 22233601]

21. Dilthey A, Leslie S, Moutsianas L, Shen JD, Cox C, Nelson MR, et al. Multi-population classical HLA type imputation. Plos Comput Biol. 2013; 9:e1002877. [PubMed: 23459081]

22. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2013.

23. Andel J, Perez MG, Negrao AI. Estimating the dimension of a linear-model. Kybernetika. 1981; 17:514–525.

24. Lazic SE. Model based inference in the life sciences: a primer on evidence. J Roy Stat Soc a Sta. 2011; 174:506–506.

25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]

26. Barcellos LF, May SL, Ramsay PP, Quach HL, Lane JA, Nititham J, et al. High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. PLoS Genetics. 2009; 5:e1000696. [PubMed: 19851445]

27. Rioux JD, Goyette P, Vyse TJ, Hammarstrom L, Fernando MM, Green T, et al. Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. Proc Nat Acad Sci USA. 2009; 106:18680–18685. [PubMed: 19846760]

Figure 1.

*HLA-DRB1*03:01* dosage (average number of alleles observed) over levels of disease (**a**): (healthy controls/anti-Ro(−)/anti-Ro(+)); (**b**): (healthy controls/anti-La(−)/anti-La(+)); (**c**) (healthy controls/anti-Ro(−) AND anti-La(−)/anti-Ro(+) AND anti-La(+)/). Average dosage is represented by a square, whereas upper and lower 95% confidence intervals are represented by ' −'. Note that dosage ranges from 0 to 2 for each subject and so to convert to allele frequency you must divide by 2. All three plots have been truncated at 1.

Table 1

Individual studies with number of genotyped SNPs, number of SLE cases and sample sizes for anti-Ro/La within cases

Study case collection	N genotyped MHC SNPs[a]	N cases[b]	Sample sizes (+/–/missing)
Illumina HumanHap550[6]	2380	1123	Anti-Ro: 319/796/8 anti-La: 137/978/8
Illumina HumanHap317[5]	1522	398	Anti-Ro: 36/107/225 anti-La: 17/126/255
Illumina Combined MHC panel[26]	2360	917	Anti-Ro:158/454/305 anti-La:79/531/307
Illumina custom panel[27]	1230	632	Anti-Ro: 168/446/18 anti-La: 48/565/19

Abbreviations: MHC, major histocompatibility complex; SLE, systemic lupus erythematosus; SNP, single-nucleotide polymorphism. The last column denotes the number of SLE cases who were positive, negative or had missing data for each sub-phenotype.

[a]Number of SNPs on the genotyping platform located on chromosome 6 between 26 000 and 34 000 kb.

[b]See original paper[16] for a description of SLE case recruitment.

Table 2

Forward stepwise regression models for anti-Ro

Marker	Estimates from the multiple regression model				Single marker	
	OR (95% CI)	P-value	Class	Gene	OR (95% CI)	P-value
(A) Forward stepwise regression (BIC = 2826.1)						
rs3129962 (A<G)	2.44 (2.08–2.94)	9.47×10^{-27}	Class III	BTNL2	2.32 (1.98–2.71)	2.02×10^{-25}
rs9271731 (A<G)	1.54 (1.30–1.85)	9.56×10^{-07}	Class II	DRB1-DQA1	1.27 (1.09–1.49)	2.58×10^{-03}
rs3957146 (G<A)	0.52 (0.39–0.69)	5.70×10^{-06}	Class II	DQB1-DQA2	0.38 (0.29–0.50)	3.10×10^{-12}
(B) Forward stepwise regression beginning from Thr77 DRB1 (BIC =2815.9)						
Thr77 DRB1	0.49 (0.41–0.60)	2.72×10^{-13}	Class II	DRB1	0.45 (0.39–0.52)	5.26×10^{-25}
rs9271731 (A<G)	1.63 (1.37–1.95)	4.50×10^{-08}	Class II	DRB1-DQA1	1.27 (1.09–1.49)	2.58×10^{-03}
rs3130781 (G<A)	1.44 (1.22–1.71)	1.76×10^{-05}	Class III	DPCR1	1.95 (1.69–2.25)	2.19×10^{-20}
HLA-DQB1*03:02	0.56 (0.42–0.73)	2.49×10^{-05}	Class II	DQA1	0.42 (0.33–0.55)	1.67×10^{-10}
*(C) Forward stepwise regression conditioning on HLA-DRB1*03:01 + HLA-DRB1*15:01 (BIC = 2829.8)*						
HLA-DRB1*03:01	2.22 (1.88–2.66)	1.11×10^{-20}	Class II	DRB1	2.22 (1.91–2.59)	9.29×10^{-25}
HLA-DRB1*15:01	1.54 (1.28–1.85)	4.11×10^{-06}	Class II	DRB1	1.32 (1.13–1.57)	7.42×10^{-04}
rs9275582 (A<G)	0.61 (0.50–0.75)	2.99×10^{-06}	Class II	DQB1-DQA2	0.45 (0.38–0.55)	6.12×10^{-16}
(D) Forward stepwise regression analysing HLA alleles only (BIC =2829.6)						
HLA-DRB1*03:01	2.38 (2.02–2.80)	1.21×10^{-25}	Class II	DRB1	2.22 (1.91–2.59)	9.29×10^{-25}
HLA-DRB1*15:01	1.63 (1.36–1.95)	9.71×10^{-08}	Class II	DRB1	1.32 (1.13–1.57)	7.42×10^{-04}
HLA-DQB1*03:02	0.53 (0.41–0.70)	4.84×10^{-06}	Class II	DQB1	0.42 (0.33–0.55)	1.67×10^{-10}
(E)Forward stepwise regression analysing amino acid data only (BIC =2827.2)						
Thr77 DRB1	0.29 (0.24–0.36)	4.00×10^{-32}	Class II	DRB1	0.45 (0.39–0.52)	5.26×10^{-25}
Leu67 DRB1	0.64 (0.53–0.77)	2.81×10^{-06}	Class II	DRB1	1.05 (0.93–1.20)	4.22×10^{-01}
Gln96 DRB1	1.47 (1.23–1.76)	2.29×10^{-05}	Class II	DRB1	1.29 (1.11–1.51)	1.28×10^{-03}

Abbreviations: BIC, bayesian information criterion; CI, confidence interval; HLA, human leukocyte antigen; OR, odds ratio; SNP, single-nucleotide polymorphism. SNPs have their minor and major alleles noted in brackets (A<G where A is the minor allele (A<G, for example), the OR is with respect to the minor allele. For BIC see Materials and Methods.

Table 3

Forward stepwise regression models for anti-La

Marker	Estimates from the multiple regression model					Single marker	
	OR	95% CI	*P*-value	Class	Gene	OR	*P*-value
(A) Forward stepwise regression (BIC = 1677.92)							
rs2894254 (C<A)	3.38	2.74–4.16	3.40×10^{-30}	Class III	*c6orf10—BTNL2*	3.38	3.40×10^{-30}
(B) Conditional on HLA-DRB1*03:01 as a covariate (BIC = 1673.31)							
HLA-DRB1*03:01	2.50	2.00–3.13	1.40×10^{-15}		*HLA-DRB1*	3.15	3.31×10^{-28}
rs9268832 (C<A)	1.64	1.32–2.04	6.53×10^{-06}	Class II	*DRA-DRB5*	2.31	2.46×10^{-17}
(C) Forward stepwise regression starting from AA-DRB1*77T (BIC =1672.53)							
Thr77 DRB1	0.40	0.32–0.50	1.33×10^{-15}		*HLA-DRB1*	0.32	2.40×10^{-28}
rs2227139 (C<A)	1.64	1.32–2.04	6.47×10^{-06}	Class II	*HLA-DRA*	2.32	1.86×10^{-17}

Abbreviations: BIC, bayesian information criterion; CI, confidence interval; OR, odds ratio. SNPs have their minor and major alleles noted in brackets (A<G where A is the minor allele, for example), the OR is with respect to the minor allele. For BIC see Materials and Methods.

Table 4

Allele frequencies for *HLA-DRB1*03:01* in case only association analysis of anti-Ro and anti-LA when conditioning on the status of each sub-phenotype

Status	Anti-La(+), *N*	Anti-La(−), *N*
Anti-Ro(+)	0.41 (259)	0.26 (418)
Anti-Ro(−)	0.28 (22)	0.18 (1781)

Frequencies for the *HLA-DRB1*03:01* allele are shown with the sample sizes in brackets.

Table 5

Multi-level model for *HLA-DRB1*03:01* dosage over phenotype

Phenotype	Effect (change in dosage)	95% CI	*P*-value
(A) Results for anti-Ro			
Anti-Ro(−)/control	0.10	0.08–0.13	1.97×10^{-14}
Anti-Ro(+)/anti-Ro(−)	0.27	0.22–0.31	2.97×10^{-33}
(B) Results for anti-La			
Anti-La(−)/control	0.13	0.11–0.15	3.57×10^{-25}
Anti-La(+)/anti-La(−)	0.41	0.35–0.47	2.45×10^{-39}

Abbreviation: CI, confidence interval. Effect is the change in dosage for *HLA-DRB1*03:01* between positive and negative for specified phenotype. (A) Results for healthy control/anti-Ro-negative/anti-Ro-positive. (B) Results healthy control/anti-La-negative/anti-La-positive.