2014

HLA-DRB1-Associated Rheumatoid Arthritis Risk at Multiple Levels in African Americans Hierarchical Classification Systems, Amino Acid Positions, and Residues

R. J. Reynolds

A. F. Ahmed

M. I. Danila

L. B. Hughes

P. K. Gregersen
Zucker School of Medicine at Hofstra/Northwell

See next page for additional authors

Authors

R. J. Reynolds, A. F. Ahmed, M. I. Danila, L. B. Hughes, P. K. Gregersen, S. Raychaudhuri, R. M. Plenge, S. L. Bridges, and Af Consortium Longitudinal Evaluat

HLA-DRB1 rheumatoid arthritis risk in African Americans at multiple levels: Hierarchical classification systems, amino acid positions and residues

Richard J. Reynolds[1,*], **Altan F. Ahmed**[1,*], **Maria I. Danila**[1], **Laura B. Hughes**[1], **CLEAR Investigators**[2], **Peter K. Gregersen**[3], **Soumya Raychaudhuri**[4], **Robert M. Plenge**[5], and **S. Louis Bridges Jr**[1]

[1]University of Alabama at Birmingham

[2]Richard D. Brasington, Jr. MD (Washington University at St. Louis), Leigh F. Callahan, PhD (University of North Carolina at Chapel Hill), Doyt L. Conn, MD (Emory University), Beth L. Jonas, MD (University of North Carolina at Chapel Hill), Larry W. Moreland, MD (University of Pittsburgh), Edwin A. Smith, MD (Medical University of South Carolina)

[3]Feinstein Institute for Medical Research, Manhasset, NY.

[4]Brigham & Women's Hospital, Harvard Medical School, Division of Genetics & Rheumatology, Boston, MA.

[5]Merck & Co., Inc.

Abstract

Objective—To evaluate African American rheumatoid arthritis *HLA-DRB1* genetic risk by three validated allele classification systems, and by amino acid position and residue. To compare the genetic risk between African American and European ancestries.

Methods—Four-digit *HLA-DRB1* genotyping was performed on 561 autoantibody-positive African American cases and 776 African American controls. Association analysis was performed on Tezenas du Montcel (TdM); de Vries (DV); and Mattey classification system alleles and separately by amino acid position and individual residues.

Results—TdM S2 and S3P alleles were associated with RA (odds ratios (95% CI) 2.8 (2.0, 3.9) and 2.1 (1.7, 2.7), respectively). The DV (P-value=3.2 x 10^{-12}) and Mattey (P-value=6.5 x 10^{-13}) system alleles were both protective in African Americans. Amino acid position 11 (permutation P-value < 0.00001) accounted for nearly all variability explained by *HLA-DRB1*, although conditional analysis demonstrated that position 57 was also significant (0.01<= permutation P-val <=0.05). The valine and aspartic acid residues at position 11 conferred the highest risk for RA in African Americans.

Address correspondence to Richard J. Reynolds, PhD, Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, 1720 2nd Ave South, SHEL 211, Birmingham, AL 35294-2182. rreynolds@uab.edu, Phone: (205) 975-9300, Fax: (205) 996-6734.
*indicates co-first authorship

None of the authors report conflicts of interest with regard to this work.

Conclusion—With some exceptions, the genetic risk conferred by *HLA-DRB1* in African Americans is similar to European ancestry at multiple levels: classification system (e.g., TdM), amino acid position (e.g. 11) and residue (Val 11). Unlike that reported from European ancestry, amino acid position 57 was associated with RA in African Americans, but positions 71 and 74 were not. Asp11 (OR = 1 in European ancestry) corresponds to the four digit classical allele, *09:01, also a risk allele for RA in Koreans.

The largest contributor of RA genetic risk in the major histocompatibility complex (MHC) is *HLA-DRB1*. Stastny (1) first reported an association between the HLA-DRw4 (*HLADRB1*04) alloantigen and RA. The DRw4 alloantigen was found in 70% of rheumatoid factor positive RA patients as compared to 28% of the healthy controls. Gregersen et al. (2) observed that genetic variation among DR4 haplotypes is restricted to the *HLA-DRB1* gene. This variability is restricted to the codons surrounding amino acid position 70 in the third hypervariable region of *HLA-DRB1*. Gregersen et al. also noted a region of conserved amino acids from position 70-74 in the *DRB1*04 and *DRB1*01 alleles associated with RA. This region of conservation was named the shared epitope (SE). Despite validation of the SE hypothesis, the exact mechanism by which the SE contributes to RA remains unknown.

Recent research demonstrates that the majority of the known genetic risk for autoantibody positive rheumatoid arthritis (RA) arises from allelic variants of two class II (*HLADRB1* and *HLA-DPB1*) and one class I (*HLA-B*) MHC genes (3, 4). *HLA-DRB1* in particular has consistently been shown to be highly associated with RA across ethnicities (3, 5, 6). Certain amino acid residues at positions 11, 71, and 74 account for the majority of the association of *HLA-DRB1* with RA risk (3). The residues corresponding to the traditional SE (positions 71 and 74) and the newly defined risk residue (valine11) are all located in the peptide binding groove of the HLA-DRB1 molecule. Therefore, the variable risk conferred by *HLA-DRB1* appears to be explained by specific amino-acid substitutions in the primary structure of its polypeptide chain.

The original description of the SE hypothesis did not address the potential variation in risk among the different SE containing alleles. Subsequent detailed studies in European ancestry populations have shown that there exists a hierarchy of risk among SE alleles (7-10). Tezenas du Montcel et al. (9) proposed an alternative classification system to overcome these limitations. In this system, SE-containing *HLA-DRB1* alleles are further categorized into genotypes based on the amino acids at positions 70 and 71. These amino acids appear to modify the genotypic risk conferred by the conserved motif at positions 72-74. For example, the genotypic risk of <u>KRAA</u> - <u>Q/R RRAA</u> heterozygotes was found to be higher than the homozygotes for each motif. Specifically, genotypes with one allele that differ by a substitution of arginine for lysine at position 71 and the presence of a glutamine or arginine at position 70 appear to have the highest risk for RA. However, the study by Morgan et al. (11) demonstrated that the highest risk genotype was characterized by the presence of the KRAA sequence on both alleles. In addition, de Vries et al. (10) and Mattey et al. (12) have observed that isoleucine and aspartic acid amino acid substitutions at positions 67 and 70, respectively, are less common in cases than in controls of European ancestry. The TdM, DV

and Mattey systems of *HLA-DRB1* reclassification have all been validated in people of European ancestry (9, 10, 12), but have not been examined in African Americans.

Populations of various ethnicities differ in terms of variability of linkage disequilibrium, haplotype structure and allele frequency differentiation, which may result from neutral or selective evolutionary forces (13, 14). Therefore, the genetic risk of reclassified *HLA-DRB1* alleles, e.g. TdM system or the risk evident at amino acid positions or specific residues, may vary among different ancestries. Hughes et al. (5) addressed the hypothesis that African Americans' genetic risk for RA through the HLA-DRB1 SE is indirect (as a marker of increased European genetic admixture); support for this was modest compared to similar studies in systemic lupus erythematosus (15). Nevertheless, this raises the possibility that the specific *HLA-DRB1* alleles responsible for disease risk may differ in African and European ancestral populations. Compared to peoples of European descent, African Americans are thought to have a lower prevalence of RA (16), lower frequency of the highest risk *HLA-DRB1* classical alleles (e.g. *04:01, *04:04), and lower effect size of the high risk alleles with RA (5). However, the well-established hierarchy of risk conferred by SE-containing alleles has not been previously investigated in the African American RA population, nor has the risk been quantified at the amino acid position or residue levels.

In this study we sought to evaluate the applicability of the TdM, DV, and Mattey systems of *HLA-DRB1* alleles and to define genetic risk at individual amino acids across *HLA-DRB1* exon 2 in a large sample of autoantibody positive African American RA patients and healthy controls. The specific objectives of this analysis were to test the following hypotheses: 1) the TdM system allele risk hierarchy correlates with RA in African Americans; 2) the DV and Mattey-classified risk alleles, shown to be "protective" in peoples of European ancestry, i.e. higher frequencies in controls than cases, are also protective in African Americans; 3) the highest risk alleles in African-Americans contain a valine residue at position 11 (V11); and 4) the RA risk conferred by *HLA-DRB1* alleles at all levels of analysis is similar between African Americans and Europeans.

This study provides novel insight into the association of specific amino acid motifs with RA in African Americans in light of the newly defined high risk V11 variant of *HLA-DRB1* in exon 2. We also make conclusions, supported by strong empirical evidence, concerning the level of trans-ethnic variation in the risk conferred by the four classifications of *HLA-DRB1* alleles.

PATIENTS AND METHODS

Study patients and controls

We analyzed *HLA-DRB1* genotype data of 561 African American autoantibody positive RA patients and 776 healthy African American controls. All patients with RA were participants in the Consortium for the Longitudinal Evaluation of African Americans with Early Rheumatoid Arthritis (CLEAR) Registry previously described by our group (5, 16). The healthy African American controls were age-, sex-, and geographic location-matched to the CLEAR RA patients. 358 of the 776 healthy African American controls were from the CLEAR registry; the remaining 418 controls were recruited from the Birmingham, Alabama

area. All protocols were approved by the University of Alabama at Birmingham Institutional Review Board for Human Use.

Sequencing and genotyping

High resolution DRB1 sequencing was completed for all study participants using the AlleleSEQR HLA DRB1 reagent kit and protocol (Abbot) and the ABI 377 automated sequencer (Applied Biosystems, Foster City, CA). Four digit DRB1 genotypes were assigned using Assign software (Conexio Genomics, Fremantle, Western Australia, Australia).

Statistical analysis

In order to test the hypothesis that the TdM classification system allele risk hierarchy correlates with RA in African Americans, we first defined S2, S3P and L alleles based on the classical allele amino acid sequence from position 70-74 (Table 1). We then estimated genome-wide ancestry using the software Admixture (17), using a set of ~30,000 autosomal and genome-wide SNP markers with partial correlation < 0.5 (i.e. approximate linkage equilibrium), and minor allele frequency greater than 0.05 typed on the Illumina ImmunoChip array (18). The Immunochip marker set matched 21,106 markers from HAPMAP representing the 2 founding populations (CEU and YRI). In order to obtain markers informative for ancestry in the admixed sample, only markers (N =3,826) with different minor alleles between the two ancestral populations were extracted from the 21,106 SNP set. The average difference in MAF between the two ancestral populations for these 3,826 markers was 0.35, which is sufficiently informative for admixture estimation for models assuming two founding populations (19). Admixture models were fit with and without individuals from HAPMAP known to be 100% European (N = 3) and 100% African (N = 3). Estimated proportion ancestry for these individuals using the genome wide panel of 3,826 markers were identical to expectations.

We fit base models of RA case affectation status as the logit of proportional European admixture and sex. A second model was fit that included the base model predictors and two additional variables. These variables were numerically coded as the number of copies {0, 1, 2} of the S2 and S3P alleles (TdM classification) or N or P alleles (DV & Mattey classifications). These models were fit using the "glm" function in R (20) assuming a binomial response with a logit link function. ANOVA was used to test the hypothesis of RA association with number of copies of TdM classification, DV and Mattey system alleles by comparing the change in model deviance between the models with and without the genetic variables. The P-values of the test that the model with the genetic variables was a better fit to the case affectation data than the base model were calculated assuming the test statistic was distributed as $\chi^2_{(2)}$.

Genotypic association analysis was performed by fitting a base model of case affectation status as the logit of European admixture proportion and sex, and then a second model, identical with the base except for the addition of a third variable coding the six TdM classification system genotypes; e.g., S2/L, S3P/L, S2/S2, etc. The L/L genotype was the referent. To test the hypothesis that the TdM classification system genotypes were

associated with case affectation status, like the allelic association models described above, ANOVA was used to compare the two nested models, one with and one without the genotype variable. The test statistic is the change in the model deviance which is asymptotically distributed as $\chi^2_{(5)}$.

We tested the hypothesis that the newly discovered valine motif at *HLA-DRB1* amino acid position 11 shown to be associated with RA among European ancestry also is a risk factor among African Americans. To test this hypothesis we created new variables corresponding to the presence or absence of residues at each relevant amino acid position across HLA DRB1 exon 2. Each amino acid position contains a variable number (= p) of residue substitutions. For example position 11 contains six residues. We evaluated the association of each position (containing p residues) by fitting two statistical models, one with and another without the p-1 variables amino residues. The full model contained sex, proportion European admixture and the p-1 variables corresponding to the number of copies of residues each individual had for each amino acid. The reduced model did not contain the p-1 residue variables. The change in residual deviance between the full and reduced models was computed. The models were binomial regressions of case control status fit using the glm function in R. Because all amino acid positions are physically linked, we used permutation test p-values to determine the statistical significance of the residues comprising each amino acid position. The permutation test was performed by permuting the unique combination comprising the set of residues for a given position and the corresponding sex and admixture proportion for each individual. The change in the deviance over the null model was observed for each amino acid position, and the maximum deviance for each iteration (n = 100,000) was selected to generate a null distribution of test statistics. In addition, we assessed the evidence that additional positions beyond the most highly associated position contributed to variation in case control status by conditioning out its variability and testing the effect of each additional position again using binomial regression analysis. We also used permutation test p-values to assess statistical significance of the conditional tests. We compared risk profiles between European and African American ancestries by comparing univariate odds ratios of the residues within the most highly associated amino acid positions estimated in the present study with estimates from European populations reported in the literature (9) (11). With 30 amino acids tested we consider results highly significant at the Bonferroni corrected alpha level = 0.0017.

RESULTS

Of the 561 participants, 465 (83%) were women, with mean (± SD) age 46 ± 13 years; 303 (54%) had ever smoked and 241 (43%) had at least one shared epitope (SE) allele. The mean disease duration (± SD) of RA was 7.3 ± 9 years. The base model contained a statistically significant effect of sex (p=5.6E-06) but proportion European ancestry was not significantly associated with case-control status (p=0.11).

An allelic model that included the dosage (0, 1, or 2) of the S2 and S3P TdM classification system alleles in addition to the covariates sex and genetic admixture proportion explained significantly more of the variation in affectation status than the non-genetic covariates alone (χ^2 = 74.4, DF = 2, p-value = 7.0e-17). The risk hierarchy of the TdM classification system

alleles based on magnitude of effect alone was S2 > S3P > L (Table 2). The S2 susceptibility allele effect, adjusting for S3P had an odds ratio (95% CI) of 2.8 (2.0-3.9). The S3P allele effect, adjusting for S2, had an odds ratio of 2.1 (1.7-2.7).

A genotypic model including a variable defining the six TdM system genotypes in addition to the covariates also explained significantly more variation in affectation status than the covariates alone (χ^2= 77.1, DF=5, p-value = 3.4e-15). The risk hierarchy of the TdM classification system genotypes was S2/S3P > S3P/S3P > S2/S2 > S2/L > S3P/L (Table 2).

The allelic odds ratios of TdM classification system *HLA-DRB1* RA risk alleles from African Americans were nearly identical to the estimates in Europeans reported by Morgan et al. (11) (Table 3, Figure 1). For both ethnicities, the frequency of S2 and S3P alleles in RA cases was much higher than for controls. Among Europeans, the estimated odds ratio for the S2 allele was about 50% higher than for the S3P allele; for African Americans, its estimate was 33% higher. The highest risk genotype in African Americans was S2/S3P, which has also been reported in some European samples (9, 21). The study by Morgan et al. (11) reported that the highest genotypic risk in Europeans was conferred by S2/S2. The lowest risk genotype in African Americans was S3P/L, which is also the lowest risk genotype for Europeans. The pattern of effect sizes according to allelic and genotypic models was remarkably similar between European and African American populations. The greatest variability in genotypic risk between African Americans and Europeans appeared to be for the S2 homozygotes and S2/S3P heterozygotes; the risk associated with these genotypes is also heterogeneous among the studies of Europeans.

Based on permutation test p-values, amino acid positions (in ascending order of p-values) 11, 13, 33, 47, 10, 57, 12, 30, and 86 were significantly associated with RA (Figure 2). When conditioning on amino acid 11, no other position was statistically significant at the alpha = 0.001 level. Position 57 was marginally significant with permutation p-value between 0.05 and 0.01; notably this residue resides near the peptide binding groove on the beta subunit of HLA DR molecule.

Certain residues of amino acid position 11 are similar in risk to that observed in Europeans, but there are exceptions. The valine residue has the highest risk. Its OR (95% CI) is 3.0 (2.3, 4.0) with residue frequencies of 0.17/0.07 in cases/controls respectively, consistent with the observations among Europeans (Figure 3). Aspartic acid at amino acid 11 is also a risk residue in African Americans. The OR (95% CI) is 1.9 (1.3, 2.7) with residue frequency of 0.06/0.03 in cases and controls, respectively, but in Europeans, it is found at very low frequency and is not associated with RA. All African Americans (N = 5) homozygous for aspartic acid at residue 11 are homozygous for *HLA-DRB1* *09:01. This classical allele has been observed to have very low frequency in Europeans. Accordingly, HLA-DRB1 09:01 has a frequency of 0.06/0.03 in African American cases and controls, respectively with a corresponding odds ratio (95% CL) of 1.9 (1.3, 2.7).

DISCUSSION

The *HLA-DRB1* locus is highly polymorphic, confers more risk for RA than any other locus, and the risk varies according to specific amino acid residues in the peptide binding region of the HLA-DRB1 molecule. Notably, the frequencies of these high risk alleles (i.e., specific amino acid residues near position 70) are known to be different among ethnic groups. Here we have endeavored to define the variable genetic risk within African Americans (776 healthy controls and 561 autoantibody positive cases), making this the largest study analyzing the genetic risk conferred by *HLA-DRB1* on RA in African Americans to date. Our major finding is that *HLA-DRB1* SE alleles reclassified according to amino position confer substantial risk for RA in African Americans, whether the scale of analysis is at allele groupings (e.g, Tezenas du Montcel) or specific amino positions (e.g, position 11) and residues (e.g. valine 11). Furthermore, the risk observed in African Americans is remarkably similar to findings from Europeans, no matter the scale of analysis, notwithstanding highly variable allele frequencies between ancestries. We demonstrated that amino acid position 11 harbors the highest source of variation in RA risk in African Americans primarily from valine and aspartic acid residues. We have also defined HLADRB1 09:01 as a novel risk allele for African Americans with RA, similar to findings in Koreans with RA (6).

We found that in this sample of African Americans, HLA-DRB1 risk alleles defined by the TdM classification system were very highly associated with RA, at both the allelic and genotypic levels, over a base model of sex and proportion European ancestry. This result validates the applicability of the TdM hierarchical classification system in this ethnic group. The strength of the relationship between S2 alleles and RA (OR=2.8) is 33% stronger than for the S3P alleles (OR=2.1). The group of high risk alleles defined by S2, containing the amino acid motif KRAA at positions 71-74 and including a lysine at position 71, carries the highest risk. S3P alleles differ from S2 alleles in the presence of an arginine at position 71 and either an arginine or glutamine at position 70. Even between the high risk alleles, it appears that the different hierarchical risk may be attributed to amino acid substitutions at position 71.

The estimated odds ratio of association for S2 and S3P alleles from the present study are essentially identical to those estimated from the Morgan et al. (11) study (Figure 1) and qualitatively similar to the smaller but seminal Tezenas du Montcel et al. (9) study. The effect sizes in the present study also corroborate the findings recently reported by Mackie et al. (n=3657) (22). Furthermore, the N and P alleles of the DV and Mattey classification systems are protective in African Americans, a finding consistent with those of Morgan et al. (11). Therefore, in terms of allelic association statistics, the TdM, DV and Mattey systems of *HLA-DRB1* alleles do not support the hypothesis of genetic heterogeneity in risk profiles between ethnic groups. Furthermore, the homogeneity of allelic effects observed across ethnic groups is also evident in genotypic effects. We found that the S2/S3P heterozygous genotype had the highest risk followed by the S3P homozygote, then the S2 homozygote, the S2/L heterozygote and finally, the S3P/L heterozygote. We found discordance with the Morgan study for the highest genotypic risk (S2/S2), but agreement with the Tezenas du Montcel et al. (9) and Michou et al. (21) studies (S2/S3P). It is highly probable that the differences observed in genotypic risk among these studies is due to

sampling variability, due to very low expected frequencies of subjects homozygous for the S2/S2 genotypes.

We also categorized the classical *HLA-DRB1* alleles according to the presence/absence of amino acid residues for each position across exon 2. We report here that variability explained by position 11 is also highly associated with RA in African Americans, identical to the result reported in peoples of European descent (3). Unlike the result reported in European ancestry, we were not able to demonstrate that either positions 71 and 74 were significant according to the distribution of permutation test statistics. Lysine at position 71 and alanine at position 74, which were the high risk residue in Europeans, had odds ratios (95% CL) of 1.2 (0.96, 1.4) and 1.1 (0.96,1.4), respectively in the African American sample. The two residues were common with the lysine 71 and alanine 74 control frequencies at 0.18 and 0.65, respectively. After accounting for variability due to position 11, there was no more variance explained significantly by any positions, except possibly position 57. This position is also located in the binding groove of the beta chain of HLA-DR molecule. Thus, we conclude that the bulk of the variability of *HLA-DRB1*'s effect on RA in this sample is explained by substitutions of residues within position 11, which is explained primarily by valine and aspartic acid. The aspartic acid residue is specific to the classical allele *HLA-DRB1* *09:01, which has been shown to be a risk allele for RA in Koreans (6). Therefore we conclude that *HLA-DRB1* *09:01 is also a risk allele in African Americans.

Similar to the report by Hughes et al. (5) we did not find that genome-wide proportion European admixture was associated with RA risk (Table 2). There remains uncertainty regarding the conclusion that RA risk among African Americans is due to the introgression of European derived *HLA-DRB1* alleles. Global admixture is known to be a poor proxy for local ancestry of alleles (23). Thus we acknowledge the limitation of the current study in explicitly testing the null hypothesis that RA genetic risk is homogeneous between ethnicities and is independent of European ancestry. It seems unlikely, however, that ancestry explains the consistent hierarchy of risk for SE alleles among the European and African American populations. For this to be the case, one must hypothesize that very strong susceptibility alleles are in linkage disequilibrium with the various *HLA-DRB1* risk alleles, and this is generally not consistent with current data (3).

In conclusion, we have validated the association of several classification systems of *HLA-DRB1* alleles, amino acid position 11, and the specific residue valine11, with risk for RA in African Americans. Previously our group reported that risk conferred by non-MHC alleles tagging validated RA risk loci was also similar to populations of European descent, although there were a few exceptions (24). Three SNPs had risk opposite to that observed among Europeans (25), and the causal *PTPN22* SNP, rs2476601, for example, was nearly monomorphic in the African American sample. There were also differences between the ethnic groups observed in *HLA-DRB1* risk: The aspartic acid at position 11 is a risk allele (neutral in Europeans), and the risk evident among Europeans at position 71 and 74 was not observed in this sample of African Americans. Finally, it appeared that residue substitutions at amino acid position 57 also are associated with RA risk independent from the effects observed at position 11. The conclusions regarding the homogeneity of risk attributed to large and modest effect RA risk loci and the differential allele frequencies observed between

ethnicities are promising in terms of potential gains in statistical power for mapping new RA risk loci using multiethnic sample sets (26, 27). Overall, the accumulating body of evidence demonstrates that the genetic basis of RA is highly uniform between African Americans and Europeans, with certain exceptions, notably the novel finding that *09:01 is a RA risk allele in the African American population. Future directions include estimating local admixture (e.g., using the software HAPMIX (28) or LAMP-LD (29)) in the region of chromosome 6 immediately flanking *HLA-DRB1* in the African American sample in order to obtain more detailed results on *HLA-DRB1* genetic risk stratified by ancestry.

Acknowledgements

References

1. Stastny P. Association of the B-cell alloantigen DRw4 with rheumatoid arthritis. N Engl J Med. 1978; 298(16):869–71. [PubMed: 147420]

2. Gregersen P, Silver J, Winchester R. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis and Rheumatism. 1987; 30:1205–13. [PubMed: 2446635]

3. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. 2012; 44(3):291–6. [PubMed: 22286218]

4. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat Genet. 2012; 44(5):483–9. [PubMed: 22446960]

5. Hughes L, Morrison D, Kelley J, Padilla M, Vaughan L, Westfall A, et al. The HLA-DRB1 shared epitope is associated with susceptibility to rheumatoid arthritis in African Americans through European genetic admixture. Arthritis Rheum. 2008; 58(2):349–58. [PubMed: 18240241]

6. Lee HS, Lee KW, Song GG, Kim HA, Kim SY, Bae SC. Increased susceptibility to rheumatoid arthritis in Koreans heterozygous for HLA-DRB1*0405 and *0901. Arthritis Rheum. 2004; 50(11): 3468–75. [PubMed: 15529363]

7. Weyand CM, Hicok KC, Conn DL, Goronzy JJ. The influence of HLA-DRB1 genes on disease severity in rheumatoid arthritis. Ann Intern Med. 1992; 117(10):801–6. [PubMed: 1416553]

8. Hall FC, Weeks DE, Camilleri JP, Williams LA, Amos N, Darke C, et al. Influence of the HLA-DRB1 locus on susceptibility and severity in rheumatoid arthritis. QJM. 1996; 89(11):821–9. [PubMed: 8977961]

9. du Montcel ST, Michou L, Petit-Teixeira E, Osorio J, Lemaire I, Lasbleiz S, et al. New classification of HLA-DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility. Arthritis Rheum. 2005; 52(4):1063–8. [PubMed: 15818663]

10. de Vries N, Tijssen H, van Riel PL, van de Putte LB. Reshaping the shared epitope hypothesis: HLA-associated risk for rheumatoid arthritis is encoded by amino acid substitutions at positions 67-74 of the HLA-DRB1 molecule. Arthritis Rheum. 2002; 46(4):921–8. [PubMed: 11953968]

11. Morgan AW, Haroon-Rashid L, Martin SG, Gooi HC, Worthington J, Thomson W, et al. The shared epitope hypothesis in rheumatoid arthritis: evaluation of alternative classification criteria in a large UK Caucasian cohort. Arthritis Rheum. 2008; 58(5):1275–83. [PubMed: 18438843]

12. Mattey DL, Dawes PT, Gonzalez-Gay MA, Garcia-Porrua C, Thomson W, Hajeer AH, et al. HLA-DRB1 alleles encoding an aspartic acid at position 70 protect against development of rheumatoid arthritis. J Rheumatol. 2001; 28(2):232–9. [PubMed: 11246655]

13. Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, et al. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat Genet. 2006; 38(11): 1251–60. [PubMed: 17057719]

14. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science. 2002; 296(5576):2225–9. [PubMed: 12029063]

15. Molokhia M, Hoggart C, Patrick AL, Shriver M, Parra E, Ye J, et al. Relation of risk of systemic lupus erythematosus to west African admixture in a Caribbean population. Hum Genet. 2003; 112(3):310–8. [PubMed: 12545274]

16. Bridges SL, Hughes LB, Mikuls TR, Howard G, Tiwari HK, Alarcón GS, et al. Early rheumatoid arthritis in African-Americans: the CLEAR Registry. Clin Exp Rheumatol. 2003; 21(5 Suppl 31):S138–45. [PubMed: 14969066]

17. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009; 19(9):1655–64. [PubMed: 19648217]

18. Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet. 2012; 44(12):1336–40. [PubMed: 23143596]

19. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. Am J Hum Genet. 2003; 73(6):1402–22. [PubMed: 14631557]

20. R Core Development Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2009.

21. Michou L, Croiseau P, Petit-Teixeira E, du Montcel ST, Lemaire I, Pierlot C, et al. Validation of the reshaped shared epitope HLA-DRB1 classification in rheumatoid arthritis. Arthritis Res Ther. 2006; 8(3):R79. [PubMed: 16646982]

22. Mackie SL, Taylor JC, Martin SG, Wordsworth P, Steer S, Wilson AG, et al. A spectrum of susceptibility to rheumatoid arthritis within HLA-DRB1: stratification by autoantibody status in a large UK population. Genes Immun. 2012; 13(2):120–8. [PubMed: 21881596]

23. Shriner D, Adeyemo A, Ramos E, Chen G, Rotimi CN. Mapping of disease-associated variants in admixed populations. Genome Biol. 2011; 12(5):223. [PubMed: 21635713]

24. Hughes LB, Reynolds RJ, Brown EE, Kelley JM, Thomson B, Conn DL, et al. Most common single-nucleotide polymorphisms associated with rheumatoid arthritis in persons of European ancestry confer risk of rheumatoid arthritis in African Americans. Arthritis Rheum. 2010; 62(12): 3547–53. [PubMed: 21120996]

25. Perkins EA, Landis D, Causey ZL, Edberg Y, Reynolds RJ, Hughes LB, et al. Association of single-nucleotide polymorphisms in CCR6, TAGAP, and TNFAIP3 with rheumatoid arthritis in African Americans. Arthritis Rheum. 2012; 64(5):1355–8. [PubMed: 22127930]

26. Pulit SL, Voight BF, de Bakker PI. Multiethnic genetic association studies improve power for locus discovery. PLoS One. 2010; 5(9):e12600. [PubMed: 20838612]

27. Kurreeman FA, Stahl EA, Okada Y, Liao K, Diogo D, Raychaudhuri S, et al. Use of a multiethnic approach to identify rheumatoid- arthritis-susceptibility loci, 1p36 and 17q12. Am J Hum Genet. 2012; 90(3):524–32. [PubMed: 22365150]

28. Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, et al. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genet. 2009; 5(6):e1000519. [PubMed: 19543370]

29. Baran Y, Pasaniuc B, Sankararaman S, Torgerson DG, Gignoux C, Eng C, et al. Fast and accurate inference of local ancestry in Latino populations. Bioinformatics. 2012; 28(10):1359–67. [PubMed: 22495753]
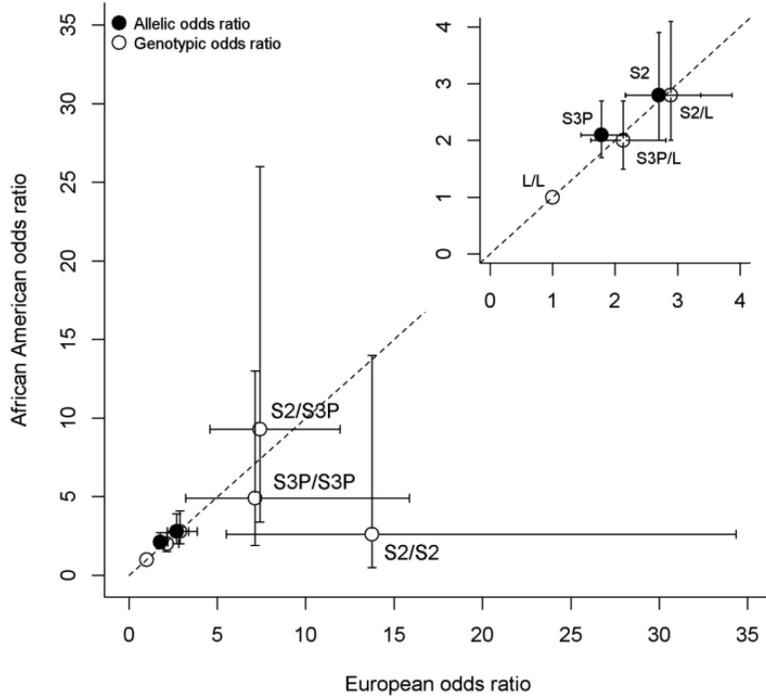
Figure 1.
TdM classified HLA DRB1 genotypic and allelic odds ratios of RA risk compared between Europeans and African Americans. Confidence limits (Cl) along y-axis correspond to African Americans from the CLEAR registry and CI limits along x-axis correspond to European SE OR reported by Morgan et al. The black (white) filled circles are allelic (genotypic) odds ratios. The 1:1 dotted line indicates equivalent odds ratios between ethnicities. All genotypic odds ratios are significantly different from the referent genotype, L/L, except for African American S2/S2. The inserted graph in the upper right hand quadrant of the figure zooms in on the region defined by OR less than four.
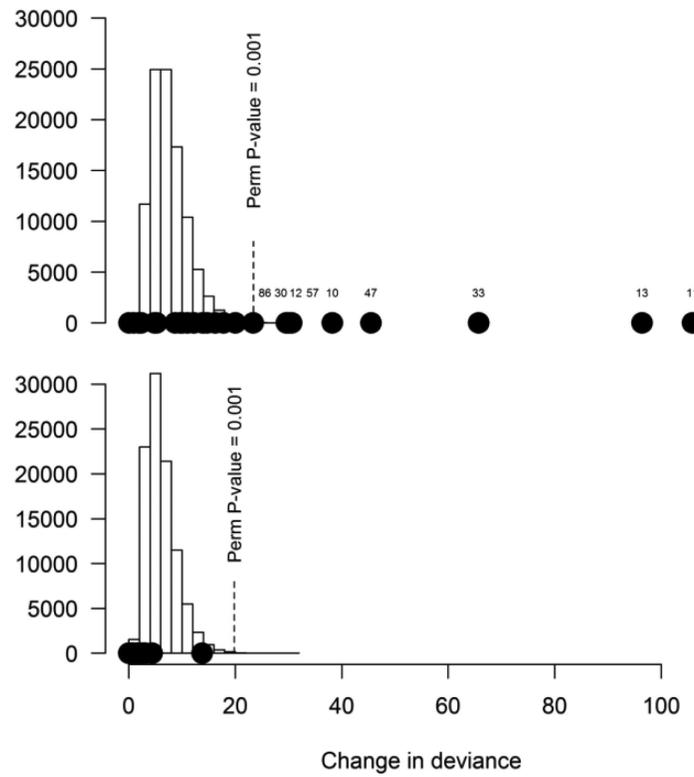
Figure 2.
Permutation test distribution of deviance statistics under the null hypothesis (top) that there is no correlation between any tested amino acid position and RA and (bottom) that there is no correlation between any tested position after accounting for the effect due to position 11.
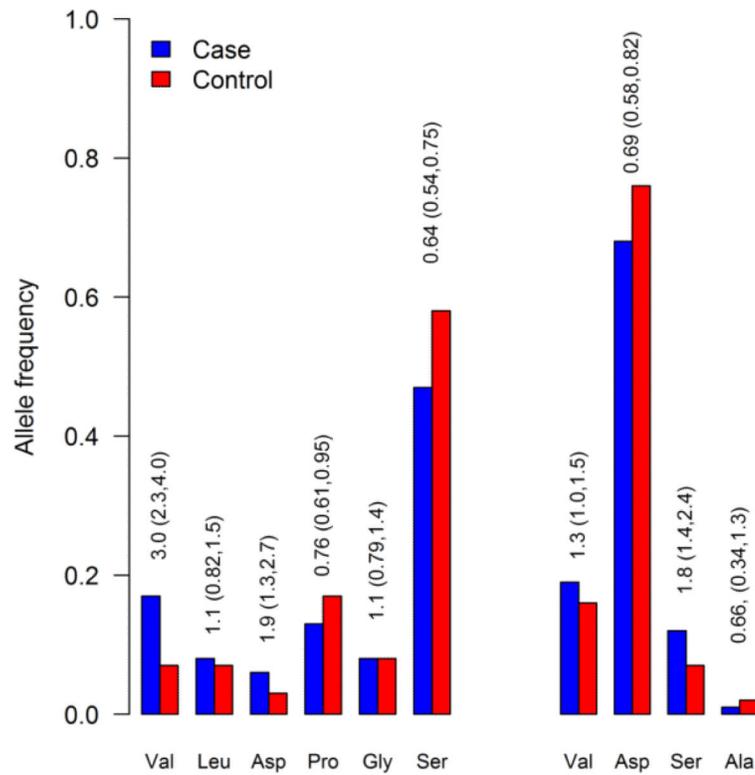
Figure 3.
Univariate odds ratios (95% CI) for amino acid positions 11 and 57 associated significantly with RA in African African Americans. Bar plots show frequency of reference alleles between cases and controls.

Table 1

HLA-DRB1 classification. Not shown are L alleles corresponding to the KRGR (71 - 74) motif (*03:01, *03:02), and RRGQ (*07:01). Shared epitope alleles are shown in bold.

Tezenas du Montcel Allele		L		S3P	S2
Amino Acid Motif (pos 70-74)	**ARAA**	**ERAA**	**DRRAA**	**Q/RRRAA**	**KRAA**
HLA-DRB1 Alleles	*15:01	*01:03	*11:02	***01:01**	***04:01**
	*15:02	*04:02	*11:03	***01:02**	*13:03
	*15:03	*13:01	*12:02	***04:04**	
		*13:02	*13:05	***04:05**	
			*16:01	***04:08**	
				***10:01**	
				***14:02**	

Table 2

Model fit statistics and parameter estimates from 1) the base model of affectation status including the intercept, sex and proportion European admixture and 2) the model including the allelic effects conferred by TdM-classified *HLA-DRB1* alleles[*]. There were 561 cases and 776 controls. P-values reported are from tests that the parameter estimates are significantly different than zero.

	Frequency (cases)	Frequency (controls)	OR	CL	p-value
Base model, N = 1,337, deviance=1794.7, DF = 1,334, AIC= 1800.7					
Intercept			0.92	(0.77,1.1)	0.82
Sex (Male)	0.17	0.28	0.52	(0.40,0.69)	5.6e-06
Sex (Female)	0.83	0.72	--	(--,--)	--
European admixture	--	--	0.43	(0.15,1.2)	0.11
Base + TdM alleles, deviance = 1720.3, DF = 1,332, AIC = 1730.2					
S2	0.10	0.045	2.8	(2.0,3.9)	2.8e-10
S3P	0.19	0.11	2.1	(1.7,2.7)	2.2e-10
L	0.71	0.84	--	(--,--)	--
Base + TdM genotypes, deviance = 1717.6, DF = 1,329, AIC = 1733.6					
S2/S3P	0.040	0.0064	9.3	(3.4,26)	9.0e-06
S3P/S3P	0.026	0.0090	4.9	(1.9,13)	0.00072
S2/S2	0.0052	0.0038	2.6	(0.48,14)	0.26
S2/L	0.15	0.076	2.8	(2.0,4.1)	1.8e-08
S3P/L	0.29	0.19	2.0	(1.5,2.7)	2.5e-07
L/L	0.50	0.71	REF	(--,--)	--
Base +DeVries alleles, deviance = 1734, DF=1,332, AIC = 1744.2					
N	0.32	0.38	0.42	(0.33,0.54)	1.2e-12
P	0.44	0.49	0.45	(0.36,0.57)	3.2e-12
SE	0.24	0.13	REF	(--,--)	--
Base +Mattey alleles, deviance = 1735, DF=1,332, AIC = 1744.8					
N	0.30	0.34	0.44	(0.35,0.57)	2.7e-11
P	0.46	0.53	0.44	(0.35,0.55)	6.5e-13
SE	0.24	0.13	REF	(--,--)	--

[*] OR=Odds ratio; CL= 95% confidence limit; REF=referent, AA11=amino acid 11

Table 3

Comparison of Tezenas du Montcel classification system allele frequencies and allelic and genotypic odds ratios across the present study and selected studies in European populations[*]

	N	Tezenas du Montcel MAF (controls)			OR$_{Allelic}$ (CL)		OR$_{Genotypic}$ (CL)				
		S3P	S2	L	S3P	S2	S3P/S3P	S2/S2	S2/S3P	S2/L	S3P/L
Ahmed et al (this report).	776	.11	.045	.84	2.1	2.8	4.9 (1.9-13)	2.6 (0.48-14)	9.3 (3.4-26)	2.8 (2.0-4.1)	2.0 (1.5-2.7)
Morgan et al.	460	.16	.13	.71	1.8	2.7	7.14 (3-16)	13.74 (5.5-34)	7.4 (4.6-12)	2.9 (2.2-3.9)	2.1 (1.6-2.8)
Tezenas du Montcel et al.	232	--	--	--	--	--	3.3[†]	5.9[†]	6.6[†]	2.7[†]	1.9[†]
Michou et al.	600	.19	.16	.65	--	--	8.2 (3-21)	20.5 (7-59)	22 (10-50)	5.9 (3-12)	4.4 (2.3-8.4)

[*] MAF= minor allele frequency; OR=odds ratio; CL=95% confidence limit

[†] CL not reported