Fine-mapping the MHC locus in juvenile idiopathic arthritis (JIA) reveals genetic heterogeneity corresponding to distinct adult inflammatory arthritic diseases


ABSTRACT

Objectives Juvenile idiopathic arthritis (JIA) is a heterogeneous group of diseases, comprising seven categories. Genetic data could potentially be used to help redefine JIA categories and improve the current classification system. The human leucocyte antigen (HLA) region is strongly associated with JIA. Fine-mapping of the region was performed to look for similarities and differences in HLA associations between the JIA categories and define correspondences with adult inflammatory arthritides.

Methods Dense genotype data from the HLA region, from the Immunochip array for 5043 JIA cases and 14 390 controls, were used to impute single-nucleotide polymorphisms, HLA classical alleles and amino acids. Bivariate analysis was performed to investigate genetic correlation between the JIA categories. Conditional analysis was used to identify additional effects within the region. Comparison of the findings with those in adult inflammatory arthritic diseases was performed.

Results We identified category-specific associations and have demonstrated for the first time that rheumatoid factor (RF)-negative polyarticular JIA and oligoarticular JIA are genetically similar in their HLA associations. We also observe that each JIA category potentially has an adult counterpart. The RF-positive polyarthritides association at HLA-DRB1 amino acid at position 13 mirrors the association in adult seropositive rheumatoid arthritis (RA). Interestingly, the combined oligoarticular and RF-negative polyarthritides dataset shares the same association with adult seronegative RA.

Conclusions The findings suggest the value of using genetic data in helping to classify the categories of this heterogeneous disease. Mapping JIA categories to adult counterparts could enable shared knowledge of disease pathogenesis and aetiology and facilitate transition from paediatric to adult services.

INTRODUCTION

Juvenile idiopathic arthritis (JIA), the most common arthritic disease of childhood, is a heterogeneous group of diseases. The current International League of Associations for Rheumatology (ILAR) classification system defines seven categories based on clinical features, including an undifferentiated category for cases that do not fall into one of the defined categories. Genetic data could be used to help define JIA categories and improve the current classification system. Prior studies of the best established genetic risk factor for JIA, the major histocompatibility region (MHC) on chromosome 6, have been in modest sample sizes. The development of methods for imputation of classical human leucocyte antigen (HLA) alleles and amino acids from genotyping array data enables a comprehensive and cost-effective approach for generating HLA typing on much larger JIA cohorts. We sought to use this powerful approach to dissect and refine the HLA associations of the heterogeneous JIA categories.

While there are considerable clinical similarities between some JIA categories and adult inflammatory arthritides, there is also substantial heterogeneity. Hence, we sought to compare the associations across the MHC region observed in JIA cohorts with those observed in adult inflammatory arthritides, such as rheumatoid arthritis (RA). Furthermore, some categories of JIA have obvious adult counterparts (eg, enthesitis-related arthritis (ERA) and adult anklyosing spondylitis (AS), or juvenile psoriatic arthritis (jPsA) with psoriatic arthritis), the most common categories of JIA, oligoarthritis and rheumatoid arthritis (RF)-negative polyarthritides, do not appear to map to any adult form of disease. Mapping each of the JIA categories to RA and other adult inflammatory arthritic diseases could have many benefits including enhanced understanding of the genetic basis and etiopathogenesis of inflammatory arthritis in general, allow extrapolation of results from clinical trials in adult inflammatory arthritis to paediatric counterparts to improve the therapy of JIA, and facilitate smooth transition of paediatric patients to adult care providers with consistent clinical designations.
The goals of this study were threefold, to use comprehensive MHC fine-mapping genetic data to refine HLA associations across each JIA category, to assess correspondences between the JIA categories and finally compare associations with adult inflammatory arthritic diseases.

**METHODS**

**Subjects**

All cohorts comprised individuals from populations of European descent from the USA, UK, Canada, Norway and Germany. Descriptions of the datasets can be found in the online supplementary information. The total dataset prequality control comprised all JIA categories and included 5737 patients with JIA and 16 403 controls genotyped for 191 494 markers.

**Genotyping and quality controls**

Samples were genotyped using ImmunoChip, a custom-made Illumina Infinium array, described previously. The ImmunoChip includes dense coverage of the HLA region and 186 additional non-HLA loci. Genotyping was performed according to Illumina’s protocols at labs in Hinxton, UK, Manchester, UK, Cincinnati, USA, Utah, USA, Charlottesville, USA, New York, USA, Brisbane, Australia and Toronto, Canada. The Illumina GenomeStudio GenTrain V2.0 algorithm was used to reclustering all 22 140 samples together.

Single-nucleotide polymorphisms (SNPs) were initially excluded if they had a call rate <98% and a cluster separation score of <0.4. A SNP was subsequently removed from the primary analysis, if it exhibited significant differential missingness between cases and controls (p<0.05), had significant departure from Hardy-Weinberg equilibrium (p<0.000001 in cases or p<0.01 in controls), or had a minor allele frequency (MAF) <0.01. Based on the SNPs that passed the above quality control thresholds, samples were then excluded for call rate <98%, or if there were inconsistencies between recorded and genotype-inferrered gender or excess heterozygosity on the autosomes. Duplicates and first-degree or second-degree relatives were removed based on identity-by-descent statistics computed using the programme KING. Admixture estimates were computed on the remaining samples while including the HapMap phase III individuals (CEU, YRI and CHB) as reference populations using the software ADMIXTURE. The admixture estimates were then used to identify and remove genetic outliers. Three of these admixture estimates were included as covariates in the logistic regression (association) analysis to account for within-sample variation.

**HLA imputation**

The markers spanning 29–34 Mb (hg build19) on chromosome 6 which encompasses the HLA region were extracted from the post-QC Immunochip dataset. Cases and controls were imputed together using SNP2HLA V1.0 (http://www.broadinstitute.org/mpg/snp2hla/). This is a robust approach which enables imputation of classical HLA alleles as well as specific amino acid positions within HLA alleles, which may play an important functional role. The method uses a large reference dataset collected by the type 1 diabetes genetics consortium (n=52225). This dataset has gold-standard HLA typing and high SNP density, thus using linkage disequilibrium patterns around SNPs and classical HLA alleles enables the inference of classical HLA alleles, amino acids and SNPs across the region based on the SNP data generated from Immunochip, an approach successfully used by a number of researchers. Post-imputation QC included removing variants with a MAF <0.01 and variants with an r² <0.8. The dosage output, which accounts for imputation uncertainty, was used for the association analyses.

To assess the quality of the imputation, a proportion of the UK and the US JIA cases have two-digit and four-digit HLA-DRB1 typing available (n=1562) performed using a semi-automated, reverse dot-blot method, which was used to calculate the proportion of accurately imputed two-digit and four-digit HLA-DRB1 alleles. In addition, the DRB1 two-digit and four-digit allele frequencies were compared between genotyped and imputed HLA allele calls.

**Association analysis of HLA alleles and amino acid polymorphisms**

To compare the differences and similarities of HLA associations across the different JIA categories, genetic correlation of the MHC region between the categories was calculated using bivariate analysis implemented using GCTA. This analysis first calculates the genetic variance (heritability) of each category and then calculates the genetic correlation between the categories across the HLA region. High correlation suggests similarities or pleiotropy between the two categories compared. This analysis requires independent controls for the two categories being compared and therefore the controls were randomly assigned to the two categories, splitting equally, taking into account the proportions of controls from each population.

HLA variants were binary coded (presence or absence) and included SNPs and two-digit and four-digit HLA alleles. Association analysis was performed using logistic regression in R, using dosage data (genotype probabilities), which takes into account imputation uncertainty. For the analysis of each JIA category, the total control dataset was used for each comparison. HLA amino acid polymorphisms have multiple residues at each position and were analysed using the omnibus test. This is a log-likelihood ratio test comparing the likelihood of the null model against the likelihood of the fitted model, which gave a p value assessing the improvement in fit of the model, the deviance is calculated (−2χ² distribution with m−1 degrees of freedom (where m is the number of HLA variant alleles)). Three of the admixture estimates were included as covariates to account for potential population stratification.

To look for independent effects across the HLA region, conditional analysis was performed. Logistic regression, as described above, was performed to identify the most associated marker. Then this marker was used as a covariate in the model and logistic regression repeated. This analysis was continued sequentially in a forward stepwise approach until no variant satisfied the genome-wide significance threshold (conditioned-p<5×10⁻⁸). When the covariate was an amino acid, all multi-allelic variants of the amino acid were included as covariates, excluding the most frequent variant. To look for additional effects outside HLA-DRB1, we included all two-digit and four-digit HLA-DRB1 alleles within the model and looked for any residual effects.

To confirm the results of the conditional analysis and to check that there were no other combinations of variants that better fitted the models derived from the forward stepwise approach, described above, we exhaustively tested all possible combinations of 2, 3 and 4 amino acid positions, including the three admixture estimates as covariates. For each combination we calculated deviation from the null hypothesis, which included only the admixture covariates. To assess the improvement in the model we also calculated the improvement in the Akaike information criterion (ΔAIC), and also the improvement in the Bayesian information criterion (ΔBIC).
We used a disease prevalence of 0.001 to estimate the variance explained (h^2) by the HLA region and some of the independent effects and compared with the estimate calculated for all Immunochip, implemented using GCTA.14

RESULTS

HLA imputation

Post-QC data was available for 6920 SNPs, 335 amino acids and 171 HLA alleles in 5043 JIA cases and 14 390 healthy controls (see online supplementary table S1). A detailed breakdown of the JIA cases by ILAR category is shown in table 1, and by population and gender in online supplementary table S2.

For a proportion of the UK and US JIA cases (n=1562), two-digit and four-digit HLA-DRB1 classical typing was available. The accuracy of the imputed data was calculated as 97.9% for two-digit and 89.3% for four-digit alleles, which is similar to those calculated in previous studies in RA.9 A detailed analysis strategy is shown in online supplementary figure S1.

Bivariate analysis to look for genetic correlation between the JIA categories

We performed bivariate analysis to calculate the estimated HLA region genetic correlation between each pair of JIA categories (figure 1). The heritability for each category estimated from the bivariate analysis was similar to that estimated from univariate analyses performed in the total dataset (see online supplementary tables S3 and S4). The estimates of correlation between each pair of JIA categories showed a surprisingly strong correlation between the most common categories of JIA: RF-negative polyarthritis, persistent and extended oligoarthritis (rG>0.88). In contrast, the correlations between these three categories and the other categories of JIA were lower (figure 1).

Association analysis of HLA markers

After conducting primary association analysis of all 7426 variants, in each of the seven JIA categories (table 2), we observed that for oligoarthritis and both RF-positive and RF-negative polyarthritis the strongest association was with HLA-DRB1 amino acid position 13. However, for oligoarthritis and RF-negative polyarthritis, the most common categories of JIA, glycine13 confers the strongest risk; serine13 also confers a risk effect but histidine13 is protective. By contrast, in RF-positive polyarthritis, it is histidine13 that confers the strongest risk and serine13 confers a strong protective effect (see figure 2, online supplementary table S5 and supplementary figure S2). When the effect estimates for the histidine13 residue in the associated JIA categories were compared using multinomial logistic regression, strong protective effects were observed in persistent and extended oligoarthritis, with no significant difference in the effect estimates (p=0.63). There was a slightly weaker, protective effect for RF-negative polyarthritis compared with that for persistent and extended oligoarthritis (p<0.05). Importantly, there was a significantly different risk effect in RF-positive polyarthritis compared with RF-negative polyarthritis, persistent and extended oligoarthritis. The remaining JIA categories had distinct HLA associations from these common categories. The most significant association in systemic JIA (sJIA) was for HLA-DRB1*11 and for the ERA category was HLA-B*27. For jPsA, no associations reached genome-wide level of significance (p<5×10^-8).

Investigation of multiple effects within the region

Observing that oligoarthritis and RF-negative polyarthritis showed similar HLA associations and evidence for pleiotropy from the bivariate analysis in GCTA,13 14 these categories were combined to increase power for further analyses (total sample size=3934). To look for independent genetic effects across the HLA region, we conditioned on the most associated marker, HLA-DRB1 amino acid position 13 and detected a second independent effect within HLA-DRB1 at amino acid position 67 (omnibus p=7.01×10^-8). Further conditioning revealed separate effects at amino acid positions 181 (omnibus p=3.33×10^-22) and 71

Table 1 Breakdown of the total JIA cohort by ILAR category

<table>
<thead>
<tr>
<th>ILAR category</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>All JIA</td>
<td>5043</td>
</tr>
<tr>
<td>Systemic JIA</td>
<td>373</td>
</tr>
<tr>
<td>Oligoarthritis</td>
<td></td>
</tr>
<tr>
<td>Persistent</td>
<td>1751</td>
</tr>
<tr>
<td>Extended</td>
<td>658</td>
</tr>
<tr>
<td>RF-negative polyarthritis</td>
<td>1525</td>
</tr>
<tr>
<td>RF-positive polyarthritis</td>
<td>337</td>
</tr>
<tr>
<td>Enthesitis-related arthritis</td>
<td>183</td>
</tr>
<tr>
<td>Juvenile psoriatic arthritis</td>
<td>112</td>
</tr>
<tr>
<td>Undifferentiated</td>
<td>104</td>
</tr>
<tr>
<td>Combined oligoarthritis and RF-negative polyarthritis dataset</td>
<td>3934</td>
</tr>
</tbody>
</table>

ILAR, International League of Associations for Rheumatology; JIA, juvenile idiopathic arthritis; RF, rheumatoid arthritis.

Figure 1 Heatmap showing genetic correlation for human leucocyte antigen (HLA) between the juvenile idiopathic arthritis (JIA) categories. Each square shows the level of correlation between each JIA category. With a scale of red colour representing higher correlation through orange to yellow for low correlation. The numbers within the squares represent the correlation values ranging from 0 to 1 for high correlation. Note that this plot is symmetrical. POligo, persistent oligoarthritis; EOligo, extended oligoarthritis; RF, rheumatoid factor; RFnegpoly, RF-negative polyarthritis; RFpospoly, RF-positive polyarthritis; sJIA, systemic JIA; ERA, enthesitis-related arthritis; jPsA, juvenile psoriatic arthritis.
Basic and translational research

Table 2  Primary association of the HLA region in JIA ILAR categories

<table>
<thead>
<tr>
<th>Category</th>
<th>HLA variant primary association</th>
<th>Position</th>
<th>p Value</th>
<th>Amino acid residue</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systemic JIA</td>
<td>HLA-DRB1*11</td>
<td>32660042</td>
<td>3.41×10⁻¹¹</td>
<td>Glycine</td>
<td>2.09</td>
<td>1.67 to 2.59</td>
</tr>
<tr>
<td>Persistent oligoarthritis</td>
<td>HLA-DRB1 AA pos 13</td>
<td>32660109</td>
<td>9×10⁻¹⁵⁶</td>
<td>Glycine</td>
<td>2.72</td>
<td>2.41 to 3.08</td>
</tr>
<tr>
<td>Extended oligoarthritis</td>
<td>HLA-DRB1 AA pos 13</td>
<td>32660109</td>
<td>1.05×10⁻¹⁰⁴</td>
<td>Glycine</td>
<td>2.87</td>
<td>2.4 to 3.42</td>
</tr>
<tr>
<td>RF-negative polyarthritis</td>
<td>HLA-DRB1 AA pos 13</td>
<td>32660109</td>
<td>4.29×10⁻⁹⁹</td>
<td>Glycine</td>
<td>2.02</td>
<td>1.76 to 2.32</td>
</tr>
<tr>
<td>RF-positive polyarthritis</td>
<td>HLA-DRB1 AA pos 13</td>
<td>32660109</td>
<td>3.65×10⁻³¹</td>
<td>Histidine</td>
<td>2.44</td>
<td>2.0 to 2.97</td>
</tr>
<tr>
<td>ERA</td>
<td>HLA-B*27</td>
<td>31431272</td>
<td>8.81×10⁻⁹⁸</td>
<td></td>
<td>1.71</td>
<td>1.32 to 2.24</td>
</tr>
</tbody>
</table>

ERA, enthesitis-related arthritis; HLA, human leucocyte antigen; ILAR, International League of Associations for Rheumatology; JIA, Juvenile idiopathic arthritis; RF, rheumatoid factor.

(omnibus p=1.16×10⁻⁸) (see online supplementary table S6 and supplementary figure S3). Conditioning on all two-digit and four-digit HLA-DRB1 alleles found independent effects at HLA-DRB1*02:01, HLA-A amino acid 95 and HLA-B amino acid 152 (see online supplementary table S6 and figure 3). All possible combinations of 2, 3 and 4 amino acids in HLA-DRB1 were exhaustively tested. HLA-DRB1 amino acids 13 and 67 were the most strongly associated of all possible 2 amino acid combinations. However, for both the 3- amino acid and 4-amino acid combinations, there were two other combinations that had a better model, according to deviance from the null hypothesis, compared with the combination of HLA-DRB1 amino acids at positions 13, 67 and 181 or for HLA-DRB1 at positions 13, 67, 181 and 71, the 3- and 4-amino acid models that might have been expected to be the most significant considering the results of the conditional analysis. Therefore, while there appears to be multiple independent effects within the HLA-DRB1 gene, we can only be confident in the HLA-DRB1 amino acids at positions 13 and 67.

Variance explained by the HLA region

We calculated the proportion of variance explained by the independent HLA variants in the combined oligoarthritis and RF-negative polyarthritis dataset (see table 3 and online supplementary table S4) and found that the total HLA region explained 8% of the total phenotypic variance, with the HLA-DRB1 region, driven by the amino acid at position 13, contributing 50% of variance explained by the HLA region.

Comparison with adult inflammatory arthritic diseases

We compared our HLA association findings across JIA categories with those of adult inflammatory arthritic diseases (see online supplementary table S7). In seropositive RA, Raychaudhuri et al showed multiple independent associations within the HLA-DRB1 gene at three amino acid positions (11, 71 and 74) and also independent associations at amino acid position 9 in HLA-B and amino acid position 9 in HLA-DPB1.3 The DRB1 amino acid at position 11 is in strong linkage disequilibrium with the amino acid at position 13, which makes it difficult to assign causality to one or the other. In this study, oligoarthritis and polyarthritis each showed association with HLA-DRB1 amino acid at position 13. If the ORs of the residues at HLA-DRB1 amino acid position 13 for paediatric and adult arthritic diseases are compared, the combined oligoarthritis and RF-negative polyarthritis dataset shows similar ORs to that seen in seronegative RA4 (see online supplementary figure S4), suggesting that these JIA categories could potentially have an adult counterpart. Likewise, in RF-positive polyarthritis, the histidine residue at position 13 at HLA-DRB1 confers the greatest risk for disease and, unsurprisingly, this mirrors the association in seropositive RA5 (see online supplementary figure S4). For the ERA category, as expected the most significant association was for HLA-B*27, the same HLA allele found in AS.

DISCUSSION

This is the largest investigation of association of the HLA region with JIA and its categories to date, exploiting novel imputation strategies we have observed differences and similarities between HLA associations for the different categories. The most common and also the most clinically homogeneous categories of JIA, oligoarthritis and RF-negative polyarthritis, showed strong genetic correlation across the HLA region supporting our previous approaches of combining these categories for genetic studies.1 11 Combined analysis of these categories show they share association across the HLA region with strong association for HLA-DRB1 amino acid position 13. The results for these combined categories are consistent with previous findings investigating association of classical HLA alleles in JIA. For example, there is previous evidence for association of HLA-DRB1*08 and the HLA alleles that lie on this haplotype, with oligoarthritis and RF-negative polyarthritis.2 3 At HLA-DRB1 amino acid position 13, the glycine residue lies on the HLA-DRB1*08 haplotype. The association with the amino acids is much stronger than that for the classical HLA allele (see online supplementary figure S3A). These combined categories also show multiple independent effects across the region, at HLA-DRB1 amino acid residues at human leucocyte antigen (HLA)-DRB1 position 13 between the combined dataset of oligoarthritis and rheumatoid factor (RF)-negative polyarthritis compared with those for RF-positive polyarthritis.
Figure 3  Analysis in the combined dataset of persistent and extended oligoarthritis and rheumatoid factor (RF)-negative polyarthritis, evidence for multiple independent effects across the major histocompatibility region (MHC). (A) Association results for all human leucocyte antigen (HLA) markers, HLA-DRB1 amino acid position 13 showed the strongest association (p<10^{-377}). (B) Conditioning on all HLA-DRB1 two-digit and four-digit alleles, HLA-DPB1*02:01 was associated (p<10^{-57}). (C) Conditioning on all HLA-DRB1 two-digit and four-digit alleles and HLA-DPB1*02:01, HLA-A amino acid position 95 was associated (p<10^{-37}). (D) Conditioning on all HLA-DRB1 two-digit and four-digit alleles, HLA-DPB1*02:01 and HLA-A amino acid position 95, HLA-B amino acid position 152 was associated (p<10^{-15}).
position 67, additional association at HLA-DPB1*02:01, an effect at HLA-A and at HLA-B. There is previous evidence for association of HLA-DPB1*0201 and HLA-A*02 (we found stronger association considering amino acids, specifically a valine residue at HLA-A amino acid position 95 in HLA-A*02). Previous studies have failed to demonstrate an association with HLA-B, which are apparent only with the additional samples available for this study.

A striking finding has been the shared association of HLA-DRB1 amino acid position 13 for both paediatric and adult diseases. It is known that amino acid position 13 is involved in shaping the peptide-binding pocket 4 of HLA-DRB1. We find that the association in the combined oligoarthritis and RF-negative polyarthritis dataset mirrors the findings seen in seronegative RA and similarly, in RF-positive polyarthritis, the findings correspond to the association in seropositive RA. Interestingly, the magnitudes of associations are stronger in the paediatric diseases compared with adult, suggesting the paediatric disease is more genetically driven.

We then further compared the associations seen in each of the other JIA categories with those of their proposed adult counterparts. Based on clinical features, it is likely that sJIA would map to adult Still’s disease, but there is currently no HLA genetic data to support or refute this. The most significant association in sJIA was for HLA-DRB1*11, consistent with recent findings from a large genome-wide association study for sJIA, which used an overlapping set of samples. Previous studies of a HLA association with sJIA had yielded conflicting results, but there is now clear evidence for association of the HLA region with this category of JIA. Data from the current study also show that the association is distinct to that seen in the other categories. This supports emerging evidence that sJIA is a distinct disease, with amino acid position 95 in HLA-A*02. Previous studies have failed to demonstrate an association with HLA-B, which are apparent only with the additional samples available for this study.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>( h^2 ) (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Immunochip</td>
<td>0.17 (0.006)</td>
</tr>
<tr>
<td>Total HLA region</td>
<td>0.08 (0.007)</td>
</tr>
<tr>
<td>All Immunochip without HLA</td>
<td>0.09 (0.006)</td>
</tr>
<tr>
<td>HLA-DRB1 2D/4D alleles</td>
<td>0.04 (0.008)</td>
</tr>
<tr>
<td>HLA-DPB1 2D/4D alleles</td>
<td>0.02 (0.005)</td>
</tr>
<tr>
<td>HLA-A 2D/4D alleles</td>
<td>0.006 (0.002)</td>
</tr>
<tr>
<td>HLA-DRB1 amino acid position 13</td>
<td>0.04 (0.02)</td>
</tr>
<tr>
<td>HLA-DRB1 amino acid positions 13 and 67</td>
<td>0.04 (0.02)</td>
</tr>
</tbody>
</table>

\( h^2 \), heritability; HLA, human leucocyte antigen; RF, rheumatoid factor.
We gratefully acknowledge contributions from physicians at CCHMC and others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/

---

**REFERENCES**


Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet* 2015;47:898–905.


The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. *Immunogenetics* 2007;59:539–53.

Association between the PTPN22 gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population: further support that PTPN22 is an autoimmunity gene. *Arthritis Rheum* 2005;52:1694–9.


