TASK QUARTERLY 18 No 4, 305–320

ANALYSIS OF SUSCEPTIBILITY LOCI FOR JUVENILE IDIOPATHIC ARTHRITIS IN THE EXTENDED MHC REGION USING HIGH RESOLUTION SNP AND HLA ALLELE TYPING

JACEK BIESIADA[1], MARC SUDMAN[2,3], MICHAEL WAGNER[1], ANDREW RUPERT[1], JILL HOLLENBACH[4], JOHANNES-PETER HAAS[5], JAREK MELLER[1,6,7] AND SUSAN D. THOMPSON[2,3]

[1] *Biomedical Informatics, Children's Hospital Research Foundation*
3333 Burnet Avenue, Cincinnati, OH 45229, USA

[2] *Center for Autoimmune Genomics and Etiology*
Cincinnati Children's Hospital Medical Center
3333 Burnet Avenue, Cincinnati, OH 45229, USA

[3] *Division of Rheumatology, Cincinnati Children's Hospital Medical Center*
3333 Burnet Avenue, Cincinnati, OH 45229, USA

[4] *Department of Neurology, University of California*
675 Nelson Rising Lane, San Francisco CA 94143, USA

[5] *German Centre for Rheumatology in Children and Young People*
Gehfeldstr. 24, 82467 Garmisch-Partenkirchen, Germany

[6] *Department of Environmental Health, University of Cincinnati*
PO Box 670056, Cincinnati, Ohio 45267-0056, USA

[7] *Department of Informatics, Nicolaus Copernicus University*
Grudziadzka 5, 87-100 Torun, Poland

(Paper presented at the CBSB14 Conference, May 25–27, 2014, Gdansk, Poland)

Abstract: Juvenile idiopathic arthritis (JIA) spans several pediatric arthropathies that involve autoimmune responses. Primary genetic risk factors for JIA have been mapped to Major Histocompatibility Complex (MHC) class I and class II genes. Recent genome-wide association studies using SNP arrays have shown that the non-HLA genetic component of JIA involves

multiple low-risk loci, reinforcing the notion that JIA is a complex genetic trait with a strong HLA component. In this work, with the goal of detailed mapping and analysis of linkage disequilibrium (LD) patterns and associations with JIA in the extended MHC (xMHC) region, we combined SNP data from Affymetrix SNP Array 6.0 and Immunochip (IC) platforms with high resolution typing of 8 classical HLA genes. A cohort of about 800 affected individuals and about 500 ethnically matched controls from the Cincinnati region, as well as secondary validation cohorts of affected individuals and matched controls from Germany were used to perform the analysis, and to assess reproducibility of the results across different platforms and populations. Accurate high resolution maps of linkage with classical HLA genes and association with individual HLA alleles were generated. High concordance of results obtained using Affymetrix SNP Array 6.0 and IC was observed, with an additional resolution and improved mapping of associations provided by the latter in some regions. Several new peaks of statistically significant and reproducible association with JIA outside the regions of strong LD with classical HLA genes were observed, including one peak in the class III region, and one in the telomeric end of xMHC. Conditional analysis provided further evidence that these associations appear to be independent of classical HLA genes studied here. The results and observed association patterns are further discussed in the context of other recent studies on autoimmune diseases, including the role of HLA-DRB1 in adult Rheumatoid Arthritis.

Keywords: Juvenile Idiopathic Arthritis, HLA, MHC region, autoimmunity, HLA independent association, SNP genotyping, HLA alleles

1. Introduction

Juvenile idiopathic arthritis (JIA) is a complex genetic disorder characterized by inflammation of the joints and other tissues in children [1–4]. $\lambda_S = 20$ JIA is classified by the International League of Associations for Rheumatology (ILAR) into several clinically distinct JIA subtypes (1–2), including oligoarticular (oligo) and IgM rheumatoid factor negative polyarticular (poly RFneg) subtypes that constitute majority of JIA cases. Other subtypes are classified as IgM rheumatoid factor positive polyarticular JIA, systemic JIA, enthesitis-related arthritis, juvenile psoriatic arthritis and undifferentiated JIA. While in general JIA is more common in children of European ancestry, the distribution of JIA subtypes differs significantly across ethnic groups [4] and most subtypes are more prevalent in females.

JIA shares histopathological features with other autoimmune diseases. It has also been observed that individuals from JIA families have increased risk for other autoimmune disorders [1]. Consistent with these similarities, genetic variation within the HLA region, including the HLA-DR region in particular, has been found to confer the strongest genetic risk for JIA [2–4]. Interestingly, class I and class II HLA alleles and haplotypes that are associated with JIA risk appear to be distinct from those implicated in rheumatoid arthritis [4–8]. At the same time, other loci outside the HLA region have been reported in genome-wide [2, 9–12] or candidate gene association studies [13–15]. Many confirmed and nominally associated JIA susceptibility loci [2, 12] have been shown to also be associated in other autoimmune diseases and include PTPN2, PTPN22, IL2, IL2RA, TNFAIP3 and STAT4 loci among others [16–21]. However, taken together, the associated JIA

loci identified to date explain less than 1/3 of JIA heritability, with the majority of this contributed by the HLA region.

These association findings and SNP datasets available from recent GWAS studies prompted us to further characterize patterns of linkage disequilibrium (LD), correlations between SNPs and HLA gene alleles, and associations with JIA within the xMHC. Immunochip (IC) and Affymetrix SNP6.0 (Affy 6) genotyping platforms, in conjunction with statistical approaches, were used to map and validate high resolution xMHC haplotypes and linkage disequilibrium patterns in the matching Caucasian cohorts of Midwest and European descent (previously described in [18]). Importantly, the availability of HLA alleles typed in these cohorts using standardized DNA sequence-based, quality control and disambiguation approaches for eight classical HLA genes, including HLA-A, -B, -C, -DRB1, -DQA1, -DQB1, -DPA1, -DBP1, provided us with an opportunity to comprehensively analyze observed associations with JIA in the context of associations between SNP and classical HLA gene alleles in this region, as a complement to other recently published studies.

The purpose of the paper is three-fold: (i) to derive and compare high resolution, multiple platform based maps of LD patterns and carefully reconstructed xMHC haplotypes for JIA and control cohorts representative of Caucasian populations; (ii) to delineate, characterize and further validate associations with JIA in these well matched cohorts by comparing the results from Affy 6 and IC platforms; (iii) to provide a detailed map of candidate associations with JIA that appear to be largely independent of the well-studied classical HLA genes for further studies and validation.

2. Materials and Methods

JIA and control cohorts. The first JIA cohort used in this study, referred to as Cincinnati JIA cohort, consists of 790 individuals who were included in the „discovery" cohort in several recent GWAS studies [18, 5, 12]. In order to maximize both homogeneity and sample size, the JIA cases were limited to the two most common subtypes of JIA, *i.e.*, IgM rheumatoid factor negative polyarticular (polyRFneg) and oligoarticular JIA (both persistent and extended). The geographically matched Cincinnati control cohort of 518 healthy children without known major health conditions has also been derived from the controls used in [12] and [18]. The second JIA cohort comprises 572 affected (polyRFneg and oligo) individuals from Germany, with geographically matched set of 501 healthy controls. The details of these cohorts are described in [18]. While the overall size of the JIA cohorts considered here is relatively limited, it should be emphasized that we do not aim here at a genome wide analysis. Instead, the goal is to perform detailed analysis and validation of associations with JIA within xMHC, using specifically assembled cohorts of affected individuals and the healthy controls that are well matched both in terms of size and ethnic background.

SNP and HLA genotyping. Cincinnati JIA and control cohorts considered here were analyzed using both Affymetrix SNP Array 6.0 (Affymetrix Inc., Foster City, CA) and Immunochip (IC) SNP genotyping platforms, providing us with an opportunity to perform cross-platform validation. The Affy 6 platform, which has been used extensively in the recent years (including the recent JIA GWAS studies), provides relatively unbiased genome-wide coverage, with 2681 SNPs within the extended MHC (xMHC) region defined as spanning chromosome 6 21.31p region between roughly 26 and 33.5 Mb (26000508–33539547 using NA30 reference). IC on the other hand, provides a specifically selected panel of SNPs with a higher coverage of xMHC (and with the total number of SNPs in this region equal to 9,181), in particular in the region between 29 and 33.5 Mb. Availability of genotypes obtained using these two platforms for the Cincinnati cohorts, provided us with an opportunity to perform systematic technical replication/variability analysis for SNPs present on both platforms, as well as the assessment of the effects of denser sampling of MHC. In order to avoid potential batch effects and other technical artifacts, samples were arranged and genotyped on individual plates in batches of interspersed samples that each included cases, controls, both genders and diseases subtypes. These batches were selected independently for each of the platforms, thus enabling further validation of observed association peaks by comparing results from the two platforms. Specifically, for Affy 6 genotyping, 768 cases and 277 controls were interspersed evenly by case control status, gender, and JIA subtype on 96-well plates, while the remaining cases and controls were grouped by case or control status and done sequentially. For IC, on the other hand, all cases and all controls were interspersed evenly (and independently) on all plates. The secondary cohorts of JIA cases and controls from Germany were used for validation. Finally, for each sample, high resolution HLA genotype data with 4-digit (*i.e.* at the resolution commonly used for association studies and also transplant matching) alleles for 8 classical HLA genes, including HLA-A, -B, -C and DRB1, DQA1, DQB1, DPA1, and DPB1, were obtained using the PCR SSOP DNA sequencing approach, as described before [5, 12, 18].

Quality control and filtering. For Affymetrix 6.0, the Na30 annotation file was used to project SNPs consistently to the positive strand. BirdSeed ver. 2 was used to process genotype calls, with a call rate of 98%. The genotypic data were used to test for cryptic relatedness for duplicates and first degree relatives, (second degree or higher relatives, if present in either data set, were not removed), autosomal heterozygocity outliers and plate effects, as described in [12]. As described above, in order to avoid technical artifacts, samples were arranged in batches that each included cases, controls, both genders and all disease subtypes. Nevertheless, to test for systematic plate-to-plate variation in genotype calls for a particular SNP, the minor allele frequency (MAF) for each plate was computed and tested whether it was within three standard deviations from the median MAF across all plates. All individuals with data for less than 95% of SNP alleles (missingness of 0.05 or more) were excluded. Subsequently,

a number of SNPs were removed based on several exclusions criteria that included missingness of more than 0.05 (individuals), Hardy-Weinberg equilibrium test (p-value of 0.001, performed separately for cases and controls), and minor allele frequency smaller than 2%. All the remaining missing values were imputed during haplotype reconstruction by Impute2, with a number of problematic SNPs that did not match allele frequency in the reference panel removed at this stage. The resulting set comprised 2,293 SNPs in xMHC region. For Immunochip, SNP were projected to positive strand based on UK annotation file, and similar exclusion criteria were then applied resulting in a set of 7,781 SNPs.

Population stratification analysis. Only self-reported non-Hispanic individuals of European American (EA) ancestry were included in this study. Further analysis of sub-population stratification, using HapMap populations as a reference, led to the exclusion of additional individuals deemed as outliers. To that end, genome wide Affy 6 SNP genotypes for the cohorts of interest were analyzed using principal component analysis (PCA), as implemented in EigenStrat, and multi-dimensional scaling (MDS), as implemented in Plink. Several reference populations from HapMap3, including those of European, Mexican and African-American ancestries, were also used at this stage to enable identification of individuals overlapping with non-European ancestries in terms of projections onto two main PCA or MDS components. In order to identify „outliers" to be excluded from the analysis, one-dimensional distributions of projections onto main components were analyzed. Individuals with a significant distance from the center of the distribution for Caucasian cohorts, and thus potentially representing additional subpopulation structure, were subsequently removed. Here, we conservatively defined outliers as individuals more than 3 standard deviations from the mean of the distribution (or Z-score greater than 3.0) for the HapMap CEU/TSI cohorts [22] in either of the two main components. It should be noted that this approach accounts for main admixture effects and population substructure by effectively removing the first two principal components, while excluding a relatively small number of individuals from the study. To confirm that, for each cohort, the ratio between the observed and expected median chi-square statistic was assessed using Affy 6.0 whole genome data. The resulting genomic inflation factors were very close to 1 (1.04 for the Cincinnati JIA cohort), indicating no significant inflation of the test statistic after removal of the outliers.

Haplotype reconstruction for xMHC. Haplotypes were reconstructed from the genotype data using a statistical population-based approach, as implemented in Phase ver. 2.1 [19, 20]. The advantage of Phase is its well benchmarked high accuracy and its ability to handle multi-allelic loci, such as HLA genes. However, computational cost of Phase limits the size of block of SNPs (or other markers) that can be considered at a time. Therefore, we performed haplotype reconstruction for blocks of 100 SNPs, augmented by a set of additional 181 tagging SNPs covering the whole xMHC region. The use of the latter allowed us to subsequently concatenate such obtained partial haplotypes into xMHC long haplotypes. This procedure was tested on HapMap data, yielding on average about

1% error rate (defined as the fraction of incorrectly assigned SNP alleles) for the whole xMHC region. For multi-allelic HLA genes, a set of 100 SNPs around each gene was used together with HLA genotypes in order to assign individual HLA alleles to chromosomes defined in terms of SNP alleles. The resulting sets of chromosomes (haplotypes) for the entire xMHC region considered here, with both SNP and HLA alleles assigned in each case, are available upon request.

Linkage Disequilibrium analysis. Standard measures, including D' and r^2, were used in order to assess Linkage Disequilibrium between SNP loci. For the analysis of associations between SNPs and individual HLA alleles, in which case the target allele forms one class and all other alleles are classified as the „other" class, the test is effectively reduced to a similar 2 by 2 problem. Following de Bakker *et al.* [21], the strength of LD between a biallelic SNP, i, and a multi-allelic HLA locus, j, with N alleles, was also quantified by relative information gain defined as follows:

$$I_{i,j} = 1 - H^C_{i,j} | H_j \tag{1}$$

In the above equation, H_j represents information entropy for multi-allelic HLA locus defined as:

$$H_j = -\sum_{k=1}^{N} p_k \log p_k \tag{2}$$

$H^C_{i,j}$ is defined as average specific conditional entropy of biallelic SNP:

$$H^C_{i,j} = -\left[f^0_i \sum_{k=1}^{N} p^0_k \log p^0_k + f^1_i \sum_{k=1}^{N} p^1_k \log p^1_k \right] \tag{3}$$

where p^x_k is the relative frequency of the kth HLA allele coinciding with the x allele at the SNP (joint probability of kth HLA allele and x SNP allele), whereas f^x_i is the relative frequency of the x allele in the population data. Thus, the relative information gain can be understood as mutual information between a biallelic SNP and a multi-allelic HLA locus, normalized by information entropy of the HLA.

Association Analysis. Standard tests of genotypic association were computed using the Plink program, including two degrees of freedom tests for 2×3 tables using simple allelic, dominant, and recessive models, with adjustment for False Discovery Rate (FDR) of 0.05. Both odds ratios (which are cohort size independent) and the resulting adjusted p-values, which were transformed using minus log (base 10) were computed for each platform and compared as part of the validation of the results and to assess inter-platform variability. Unless otherwise stated, the results for the simple allelic model, in which genotype based allele counts are compared between the two groups, are reported in the manuscript and shown in the figures.

3. Results

One of the goals of this study is to provide a detailed map of associations with JIA in the context of high resolution linkage disequilibrium pattern within

the extended MHC region. To that end, patterns of LD throughout xMHC were first analyzed by computing relative information between individual SNPs and the corresponding multi-allelic HLA loci, using xMHC haplotypes reconstructed as described in the Methods section.

As expected, the observed peaks of association largely coincide with classical HLA genes previously implicated in JIA [5, 12, 18]. In addition, secondary peaks and persistent low LD is observed throughout xMHC region. The patterns of LD within xMHC are further illustrated in Figure 1, where each color represents the strength of correlation with a particular HLA gene. It should be noted that these patterns are essentially identical for the Cincinnati JIA and Controls cohorts (data not shown).

In order to separate signal (true correlations) from noise and find the strength of spurious LD that could have arisen by chance, we have systematically randomized the data to generate background distributions of LD. In particular, we have repeatedly permuted HLA allele assignment in different cohorts, while re-computing relative information between SNPs and such randomized multi-allelic HLA loci. It should be noted that values of relative information greater than 0.05 have never been observed in our permutation experiments (performed 1,000 times

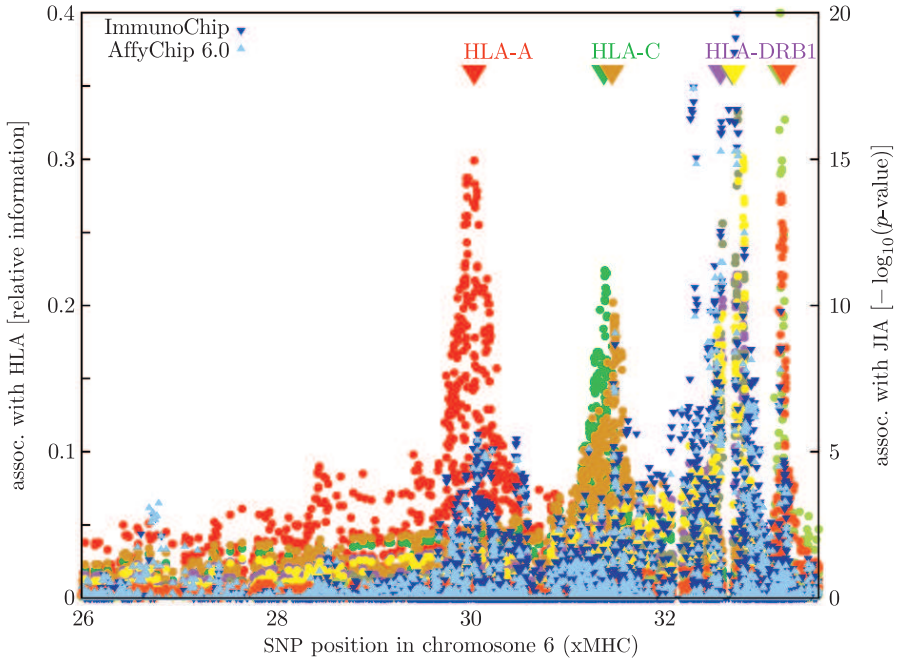

Figure 1. Relative information between HLA gene (HLA-A: red, HLA-C: green, HLA-B: strong amber, HLA-DRB1: pink, HLA-DQA1: grayish olive, HLA-DQB1: yellow, HLA-DPA1: light brilliant lime, HLA-DPB1: orange, respectively) and SNPs in XMHC shown on the left y-axes; adjusted p-value for genotypic association for JIA-Case Cohort in the Discovery Cohort (Oligo + NeqPoly), using Affymetrix (light blue) and Imunnochip (dark blue) platforms, respectively, shown on the right y-axes

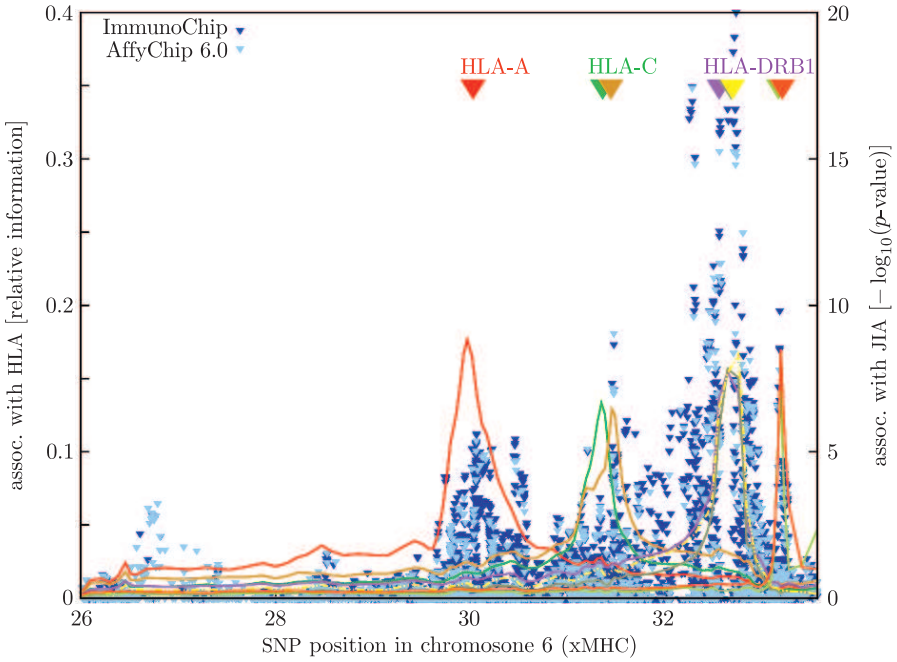

Figure 2. Same as in previous figure, except that association with HLA genes is shown using Bezier smoothing

for all the cohorts considered), and thus values greater than 0.05 likely represent true LD.

The observed LD patterns are subsequently superimposed with peaks of association with JIA, as shown in Figures 1 and 2. The strength of association with JIA is first assessed for each SNP using genotypic allelic test for association with the resulting p-values adjusted for multiple testing using FDR of 0.05. These values are shown as light and dark blue triangles (for Affy 6 and Immunochip platforms, respectively) in the $-\log$ (base 10) scale in Figures 1 and 2. Figure 2 shows the same, but with Bezier polynomial smoothing applied to the association with HLA.

In order to facilitate validation of the observed putative associations, cohort size independent odds ratios are used to compare the results across primary and validation cohorts. Specifically, the pattern of associations with JIA observed in the discovery (Cincinnati) cohort is compared with associations observed in the secondary (German, VC) cohort by using ratios of Odds Ratios computed using allele frequencies in each of these cohorts. These odds ratios and adjusted p-values are also shown in Figures 1 through 5 in order to map and analyze JIA susceptibility loci in the context of LD patterns between SNPs and HLA loci.

As can be seen from Figure 3, there is a high level of concordance between the discovery and validation cohorts. Most ratios of ORs remain within a narrow range close to 1.0 (plus/minus one standard deviation shown by red lines), indicating that most ORs in both cohorts are very similar. In particular, nearly

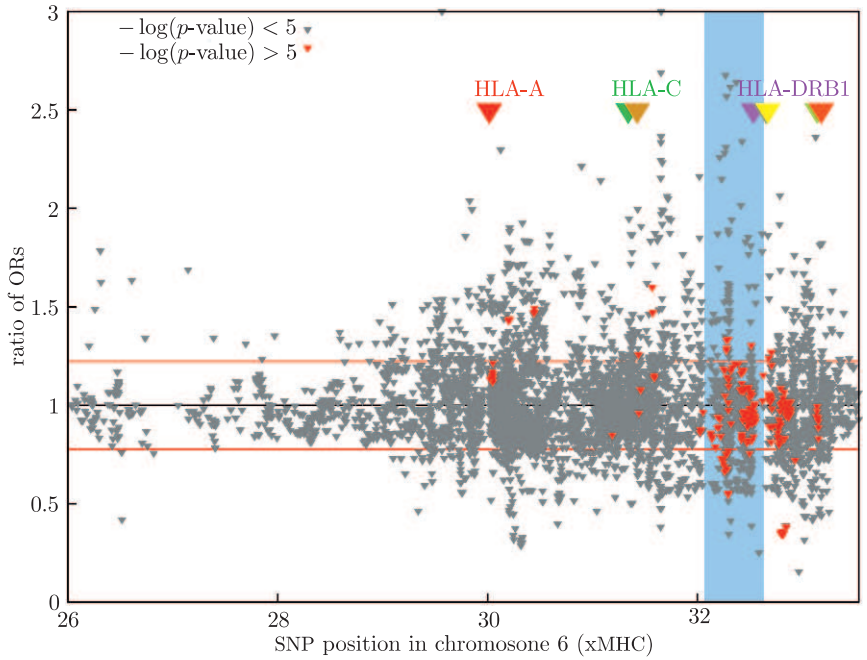

Figure 3. Reproducibility of association with JIA between the discovery (DC) and validation (VC) cohorts using ratio of Odds Ratios for association with JIA in both cohorts for Imunnochip data (similar trends are observed for Affymetrix 6 data); red points represent SNPs with $-\log(p\text{-value})$ bigger than 5 for the discovery cohort (significant associations)

all of the most significant associations (shown as red triangles) are within this narrow band. Thus, majority of SNPs associated with JIA in the discovery cohort reproduce well in the validation cohort.

As can be seen from Figures 1, 2 and 3, there are peaks of association with JIA that are located outside regions of strong linkage with classical HLA genes considered here. The strongest of those peaks is located just outside the peak of strong LD around DRB1, in MHC Class III region. This region (highlighted in blue in Figure 3) includes NOTCH4, PBX2, PRRT1 and TNXB genes and it has been implicated before in other autoimmune disorders, including systemic lupus erythematosus [17].

A more detailed view of associations with JIA and patterns of LD in this region is shown in Figures 4 to 6. The blue vertical band highlighted in Figure 3 is first shown in terms of adjusted p-values (Figure 4) and then also using ORs to factor out sample size effects for both primary and secondary cohorts (Figure 5). Figure 6 shows significantly associated SNPs (adjusted p-values lower than 10^{-5}) and standard r^2 measure of LD, providing further evidence of the lack of strong LD between NOTCH4/PBX2 and DRA1/DRB1 loci.

Another interesting (although much weaker) peak is observed in the telomeric end of xMHC (26.5–27 Mb) that comprises multiple butyrophilin (BTN) family as well as histone related genes, well outside the region of strong LD to any of the

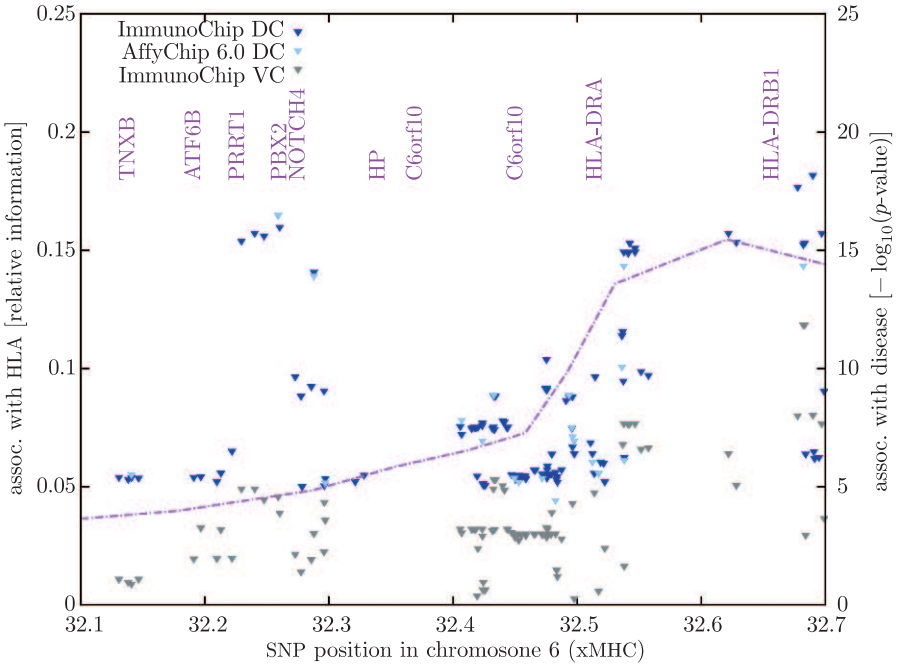

Figure 4. Detailed map of association within the novel Class III peak shown in Figure 3 using a blue vertical band; significantly associated SNPs in the region with $-\log(p\text{-value})$ larger than 5 for the genotypic association test in the discovery cohort (using Imunnochip data) are shown as dark blue triangles, whereas their counterparts using Affymetrix data for DC (if they exist) in light blue and Imunnochip data for validation cohort in grey, respectively; strength of association with JIA is shown on the right y-axes, whereas strength of association with DRB1 on the left y-axis (Bezier smoothing was applied for relative information between HLA-DRB1 and SNPs of interest); the region shown is between 32.1–32.7 Mb, and the genes in the region are: TNXB, ATF6B, PRRT1, PPT2, EGFL8, AGPAT1, AGER, PBX2, GPSM3, NOTCH4, HP (hypothetical protein LOC100294273), C6orf10, BTNL2, HLA-DRA, HLA-DRB5, HLA-DRB1 (left to right)

classical HLA genes. These association peaks are observed irrespective of the platform (Affymetrix 6 vs. Immunochip) used, providing cross-platform validation of the results. As can be seen from Figure 3, these peaks have also been reproduced in the validation cohort, suggesting that these associations should be targeted for further analysis and validation in future studies.

4. Discussion

Major Histocompatibility Complex (MHC) Class I and II genes have long been mapped as major genetic risk factors of Juvenile idiopathic arthritis (JIA). Recent genome-wide association studies further support this view [5, 9–12, 18]. Moreover, similarly to other complex autoimmune diseases, the non-HLA genetic component of JIA appears to be largely mediated by multiple low-risk variants. Some of these polymorphisms are likely to be JIA-specific, and some are shared

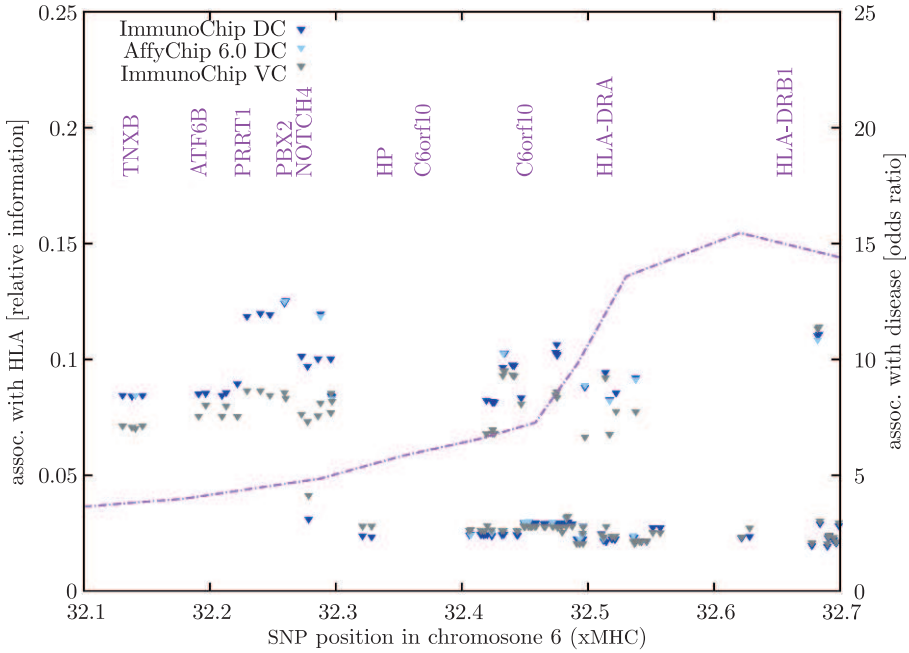

Figure 5. Same as in previous figure, except that sample size independent Odds Ratios are used to measure the strength of association with JIA

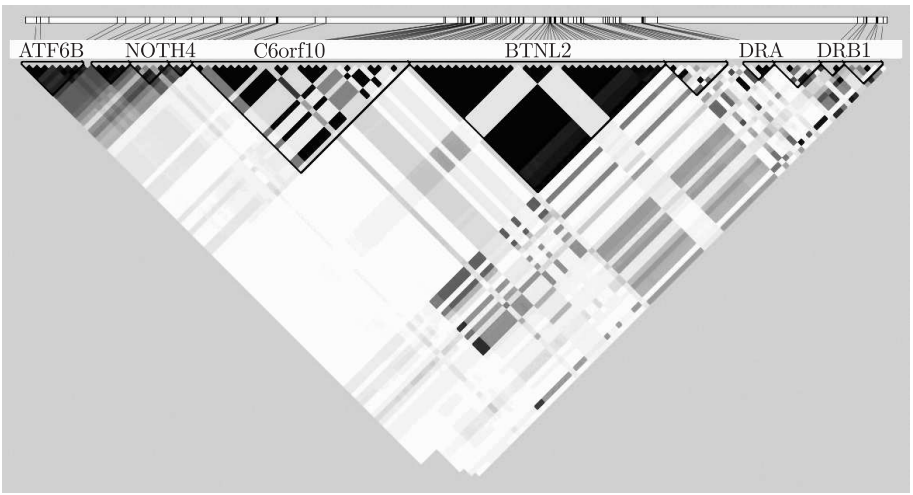

Figure 6. LD (r^2) between Notch4 and DRA/DRB1 loci

with other autoimmune diseases, as reported recently for celiac disease, systemic *lupus* erythematosus and type 1 diabetes mellitus [14–17].

Here, we used well matched JIA and control cohorts representative of Midwestern Caucasian populations (with additional validation in a German cohort) and multiple genotyping platforms, coupled with high resolution HLA typing, in order to further dissect LD patterns and associations with JIA in xMHC

region. Consistent with previous findings, association with JIA largely coincides with LD peaks corresponding to classical HLA genes. However, additional peaks of associations are observed that appear to be largely independent of the well-studied classical HLA genes considered here.

The first peak, located near DRB1 within the MHC III region, includes NOTCH 4 and PBX2 genes. Notch signaling plays an important role in T cell development and differentiation and has been proposed as a target for selective therapy to treat autoimmune disorders. Overlapping associations in the same region have been reported in other autoimmune diseases, including SLE and celiac disease [14–17]. The second interesting peak is located in the telomeric end of the xMHC region, and appears to be independent of classical HLA genes. In particular, several SNPs in this region remained significant when conditioning on major DRB1 susceptibility alleles in our initial conditional analysis (data not shown). Thus, while one cannot exclude a subtle form of an epistatic effect due gene-gene interactions, further investigation of this locus is suggested. It is expected that shedding light on this novel association will be facilitated by newer genotyping platforms that provide a more complete SNP coverage in this telomeric subregion of xMHC.

Importantly, HLA associations in children appear to be largely distinct from those in adult RA. In this context, a recent finding by Raychaudhuri *et al.* of just several amino acid variants of DRB1 explaining most of the variance in RA associations [23], prompted us to (re-) analyze these associations in our well matched pediatric cohorts. It was reported previously that the shared epitope that is strongly associated with adult RA involves select DRB1*01/*04 alleles. On the other hand, it was previously reported that DRB1*08, DRB1*11, and DRB1*13 contribute the risk of inflammatory arthritis in children, while DRB1*04 is in fact protective, depending on the clinical subtype [5, 12, 18].

In [23], as part of the overall analysis of haplotypes contributing to adult RA risk, DRB1 alleles are grouped based on the amino acid residues present at positions 11, 71 and 74 within DRβ1, *e.g.*, merging DRB1 *04:08, *04:05, *04:04, *10:01 alleles (see Table 1). The relative risk for such obtained strata are re-computed from allele frequencies in affected and unaffected cohorts in [23] and compared with the relative risk for the JIA, obtained using our primary Cincinnati cohort.

As can be seen from Table 1, and consistent with previous reports [18], strata with higher relative risks in RA (DRB1 *04:01 and DRB1 *04:08, *04:05, *04:04, *10:01) have the lowest risk in JIA. Conversely, strata with the highest relative risk for JIA (DRB1 *08:01, *08:04) is strongly protective in RA. Thus, JIA appears to be driven by different epitopes and it remains to be seen if a unique combination of just a few amino acid residues can explain most of the observed variance in JIA associations. Finally, we would like to comment that high resolution eMHC haplotypes generated as part of this study can provide valuable reference sets for the development of improved methods for the prediction of HLA alleles from SNP data [24].

Table 1. Relative risk for *DRB1* haplotypes in RA [23] and JIA [this study] cohorts; *DRB1* are grouped based on the amino acid residues present at positions 11 (or 13), 71 and 74 within DRβ1 based on Table 1 from S. Raychaudhuri *et al.*; Unadjusted (joint) allele frequencies are given for cases and controls in S. Raychaudhuri (2 and 3 column) and our JIA study (4 and 5 column); column six and seven report the respective relative risks based on unadjusted frequencies for RA and JIA study; the last column includes 95% confidence intervals for JIA study

DRB1 alleles	Unadjusted frequencies				Relative Risk		CI
	RA		JIA		RA	JIA	
	Controls	Case	Controls	Case			
*04:01	0.106	0.316	0.094	0.044	2.98	0.47	[0.345–0.641]
*04:08, *04:05, *04:04, *10:01	0.056	0.141	0.048	0.028	2.52	0.58	[0.385–0.885]
*11:01, *11:04, *12:01	0.103	0.049	0.089	0.160	0.48	1.80	[1.428–2.274]
*08:01, *08:04	0.028	0.013	0.022	0.114	0.46	5.16	[3.313–8.117]
*03:01	0.128	0.083	0.125	0.119	0.65	0.95	[0.775–1.240]
*11:02, *11:03, *13:01, *13:02	0.112	0.041	0.115	0.160	0.37	1.39	[1.131–1.829]

Table 2. Distribution of batches of samples on individual plates for Affy 6 (left table) and IC (right table) genotyping platforms

Affy		
plate #	sample count	
	control	JIA
1107	23	54
1108	22	55
1109	17	58
1110	21	57
1111	20	58
1112	21	58
1113	19	55
1114	21	59
1115	21	57
1116	20	55
1117	20	56
1118	20	54
1119	21	54
1120	11	36
1121	47	0
1122	80	0
1	0	74
2	0	55
68	87	0
69	50	0

iChip		
plate #	sample count	
	control	JIA
1	38	57
2	38	58
3	38	56
4	38	58
5	38	58
6	38	58
7	38	58
8	38	58
9	38	58
10	38	58
11	38	58
12	38	58
13	38	58
14	37	59
15	2	5
31	0	1
32	0	1
38	0	2
39	0	3
41	0	3
42	0	3

Table 3. Examples of sample arrangement on plates 1167 (Affy 6) and 2 (IC).

Affy 1107	1	2	3	4	5	6	7	8	9	10	11	12
a	Control	JIA	JIA	Control	JIA	JIA	Control	JIA	JIA		JIA	
b	JIA	Control	JIA	JIA	Control	JIA	JIA		JIA	JIA	Control	
c	JIA	JIA	Control	JIA	JIA		JIA	JIA	Control	JIA	JIA	
d		JIA	JIA	Control	JIA	JIA	Control	JIA	JIA		JIA	
e	JIA	Control	JIA	JIA	Control	JIA	JIA		JIA	JIA	Control	
f	JIA	JIA		JIA	JIA	Control	JIA	JIA	Control	JIA	JIA	
g	Control	JIA	JIA		JIA	JIA	Control	JIA	JIA	Control	JIA	
h	JIA		JIA	JIA	Control	JIA	JIA		JIA	JIA	JIA	

Immunochip 2	1	2	3	4	5	6	7	8	9	10	11	12
a	JIA	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA
b	JIA	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA
c	JIA	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA
d	JIA	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA
e	JIA	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA
f	JIA	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA
g	JIA	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA
h	JIA	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA	Control	JIA

Table 4. Affy sample locations for 12 samples located on iChip plate 2 row A

	iChip plate	iChip well	Plate #	Well location	Cohort
5968	2	A01	1116	A02	JIA
6132	2	A02	1110	F06	JIA
60134	2	A03	1112	B05	Control
6135	2	A04	1111	A09	JIA
60174	2	A05	1107	H03	Control
6018	2	A06	1118	C04	JIA
60187	2	A07	1108	D04	Control
6366	2	A08	1114	B02	JIA
60219	2	A09	1109	C04	Control
6141	2	A10	1110	C10	JIA
60277	2	A11	1115	H01	Control
6367	2	A12	1112	E05	JIA

Acknowledgements

This work was supported in part by NIH grants U01AI067150, N01 AR42272, P30 AR47363, UL1RR026314, and P30 ES006096. Computational resources were made available by Cincinnati Children's Hospital Research Foundation and University of Cincinnati College of Medicine. We would like to dedicate this work to David Glass, who passed away recently.

References

[1] Petty R E, Southwood T R, Manners P, Baum J, Glass D N, Goldenberg J, He X, Maldonado-Cocco J, Orozco-Alcala J, Prieur A M *et al.* 2004 *J. Rheumatol.* **31** 390

[2] Cassidy J T and Petty R E 2005 *In Textbook of Pediatric Rheumatology*, Cassidy J T and Petty R E (eds.), WB Saunders 808

[3] Prahalad S, Shear E S, Thompson S D, Giannini E H and Glass D N 2002 *Arthritis Rheum.* **46** 1851

[4] Saurenmann R K, Rose J B, Tyrrell P, Feldman B M, Laxer R M, Schneider R and Silverman E D 2007 *Arthritis Rheum.* **56** 1974

[5] Hollenbach J, Thompson S D, Bugawan T, Ryan M, Sudman M, Marion M, Langefeld C, Thomson G, Erlich H and Glass D N 2010 *Arthritis Rheum.* **62** 1781

[6] Murray K J, Moroldo M B, Donnelly P, Prahalad S, Passo M H, Giannini E H and Glass D N 1999 *Arthritis Rheum.* **42** 1843

[7] Prahalad S, Ryan M H, Shear E S, Thompson S D, Giannini E H and Glass D N 2000 *Arthritis Rheum.* **43** 2335

[8] Behrens E M, Finkel T H, Bradfield J P, Kim C E, Linton L, Casalunovo T, Frackelton E C, Santa E, Otieno F G, Glessner J T *et al.* 2008 *Arthritis Rheum.* **58** 2206

[9] Hinks A, Barton A, Shephard N, Eyre S, Bowes J, Cargill M, Wang E, Ke X, Kennedy G C, John S *et al.* 2009 *Arthritis Rheum.* **60** 258

[10] Prahalad S, Hansen S, Whiting A, Guthery S L, Clifford B, McNally B, Zeft A S, Bohnsack J F and Jorde L B 2009 *Arthritis Rheum.* **60** 2124

[11] Hinks A, Eyre S, Ke X, Barton A, Martin P, Flynn E, Packham J, Worthington J and Thomson W 2009 *Ann. Rheum. Dis.*, ard.2009.110650

[12] Thompson S D, Sudman M, Ramos P S, Marion M C, Ryan M, Tsoras M, Weiler T, Wagner M, Keddache M, Haas J P *et al.* 2010 *Arthritis Rheum.* **62** 3265
[13] Stahl E A, Raychaudhuri S, Remmers E F, Xie G, Eyre S, Thomson B P, Li Y, Kurreeman F A, Zhernakova A, Hinks A *et al.* 2010 *Nat. Genet.* **42** 508
[14] Reveille J D, Sims A M, Danoy P, Evans D M, Leo P, Pointon J J, Jin R, Zhou X, Bradbury L A, Appleton L H *et al.* 2010 *Nat. Genet.* **42** 123
[15] Stuart P E, Nair R P, Ellinghaus E, Ding J, Tejasvi T, Gudjonsson J E, Li Y, Weidinger S, Eberlein B, Gieger C *et al.* 2010 *Nat. Genet.* **42** 1000
[16] Ma W, Du J, Chu Q, Wang Y, Liu L, Song M and Wang W 2011 *Biochem. Biophys. Res. Commun.* **408** 84
[17] Orozco G, Eyre S, Hinks A, Bowes J, Morgan A W, Wilson A G, Wordsworth P, Steer S, Hocking L, Thomson W *et al.* 2011 *Ann. Rheum. Dis.* **70** 463
[18] Hinks A, Cobb J, Marion M C, Prahalad S, Sudman M *et al.* 2013 *Nature Genetics* **45** 664
[19] Stephens M, Smith N J and Donnelly P 2001 *Am. J. Hum. Genet.* **68** (4) 978
[20] Lin S, Qin Z S, Munro H M, Abecasis G R and Donnelly P 2006 *J. Hum. Genet.* **78** (3) 437
[21] de Bakker P I, McVean G, Sabeti P C, Miretti M M, Green T, Marchini J, Ke X, Monsuur A J, Whittaker P, Delgado M *et al.* 2006 *Nat. Genet.* **38** 1166
[22] The International HapMap Consortium 2010 *Nature* **467** 52
[23] Raychaudhuri S *et al.* 2012 *Nat. Genet.* **44** (3) 291
[24] Leslie S *et al.* 2008 *Am. J. Hum. Genet.* **82** (1) 48