Novel HLA Class I Alleles Outside the Extended DR3 Haplotype Are Protective against Autoimmune Hepatitis

Craig Lammert, MD1, Elizabeth J. McKinnon, PhD2, Naga Chalasani, MD1 and Elizabeth J. Phillips, MD3

INTRODUCTION: HLA class II allele, DRB1*03:01, is the most common genetic risk factor for autoimmune hepatitis (AIH), but other unrecognized HLA related risks exist.

METHODS: We compared the HLA class I (A, B, C) and class II (DR, DQ, DP) typing between patients with well-characterized AIH and healthy controls by high resolution sequencing of the HLA region. Seventy-three patients with AIH and 87 healthy controls were included. Association between HLA alleles and AIH was considered singly and in clusters and adjusted for age, gender, and DRB1*03:01.

RESULTS: DRB1*03:01 was singly associated with AIH among whites (odds ratio [OR]: 3.09, P = 0.002) and carriers of DRB1*03:01 also carried DQA*05:01 and DQB1*02:01. Significant HLA class I alleles were associated with AIH including those belonging to the A03 (OR: 0.4, P = 0.01) and B44 supertype (OR: 0.44, P = 0.03). Further refinement of HLA-A by binding pocket structure revealed that the sequence Y(F/T)AVMENV(H/Q)Y, corresponding to HLA-A alleles A*03:01-02; *31:01; *32:02, was protective for AIH (OR: 0.3, P = 0.002). A protective association also existed for alleles belonging to the HLA-B binding pocket structure Y(H/Y)TVKEISNY (OR: 0.35, P = 0.01), corresponding to HLA-B alleles: B*40:01-02; *41:02; *44:02-03; *45:01; *49:01; *50:01-02. Associations with specific class I alleles belonging to the 8.1 ancestral haplotype (HLA-A*01:01, HLA-B*08:01, HLA-C*07:01) were not significant when considered jointly with DRB1*03:01 and reported protective class I alleles.

DISCUSSION: Our study identified novel supertypes and HLA-A and B peptide binding structures protective against AIH. Further risk assessment of class I molecules remains important in AIH as they are key mediators of adaptive immunity.


INTRODUCTION
Autoimmune hepatitis (AIH), despite being a rare disease with estimated prevalence at 16–18 people per 100,000 (1–3), affects all population demographics and is increasing in incidence (4). It is characterized by immune-mediated destruction of hepatocytes along with serum antibodies such as anti-nuclear antibody and anti-smooth muscle antibody (5) and can result in progressive fibrosis, liver failure, and death (6). There is marked phenotypic heterogeneity associated with AIH as exemplified by a spectrum of autoantibody profiles (7), clinical symptoms at presentation (8), and response to treatment medication (9).

The HLA locus is currently the most important genetic contribution to AIH (10). The HLA region on chromosome 6 is comprised of major histocompatibility complex class I and II genes and has been strongly associated with risk for many autoimmune diseases (11). The results from the only AIH genome-wide association study have reported one strict association rooted in the HLA locus (rs2187668, odds ratio [OR]: 2.9, P = 1.3 × 10^-48), an efficient tagging SNP for DRB1*0301 (10). Alleles at this position encode the antigen-binding groove of class II molecules and can influence binding of antigens presented to native CD4+ helper T lymphocytes (12). Although historical studies (13) have linked HLA-DRB1 with AIH risk, a thorough investigation of the HLA locus has not been conducted. All HLA-associated disease risk cannot be explained by DRB1, as different alleles can portray risk or protection in different populations (14). In addition, top risk alleles encode identical residues (DRB71) and yet have been associated with different clinical outcomes (15).

Poor locus resolution in prior genetic studies and high linkage disequilibrium within the region are a major challenge to clarify the role of HLA-associated risk in AIH (16). Furthermore, class I molecules within the HLA region also remain understudied in...
many autoimmune diseases. These molecules can interact closely with killer immunoglobulin-like receptors (KIRs) to activate natural killer cells, key mediators of adaptive immunity (17–19), and hepatotropic viral replication (20). Therefore, we aimed to compare HLA class I (A, B, C) and class II (DR, DQ, DP) typing between patients with well-characterized AIH and healthy controls by high resolution sequencing of the HLA region.

METHODS
Genetic repository of autoimmune liver diseases and contributing exposures cohort
The genetic repository of autoimmune liver diseases and contributing exposures (GRACE) cohort (21), established in 2014 at Indiana University (IU), was developed in order to strategically warehouse and link biospecimens, high quality clinical data, and environmental exposure histories among autoimmune liver disease patients. AIH patients (cases) included in the GRACE cohort included both man and woman more than 18 years of age with a diagnosis of AIH meeting the criteria established by the International Group for the Study of AIH (22). Cases were excluded if other etiologies of liver disease were present (e.g., viral hepatitis, alcoholic liver disease, nonalcoholic steatohepatitis, and hemochromatosis, etc.). Cases were serially recruited from patients with well-defined AIH who were attending IU liver clinics and hospitalized within the IU hospital system.

Controls
Study controls were identified from patients enrolled in the Indiana Biobank (IB). The IB, established in 2010, is an established repository of biologic samples linked to the Indiana Network for Patient Care, a health information exchange that contains data contributed by multiple Indiana healthcare institutions, by medical record number, and other demographics.

Controls had historically normal alanine aminotransferase levels per the Prati criteria (23) and normal aspartate aminotransferase, alkaline phosphatase, and total bilirubin at time of IB collection.

DNA collection and isolation
Study DNA was extracted from stored saliva collected in the Oragene Saliva Kit which contains a stabilizing solution. DNA was extracted using the AutoGen Flex STAR. It provides a pure DNA sample using Flexigene chemistry that employs an initial process that pellets the white cell nuclei and mitochondria while simultaneously lysing the red cells. During the next step, it denatures and eliminates the proteins and cellular material utilizing a protease enzyme.

HLA sequencing and binding pocket structure
All samples required at least 2.5 μg of DNA at a concentration of 50 ng/μL in EDTA buffer to complete sequencing. HLA loci were polymerase chain reaction amplified using sample specific multiplex identifier (MID)-tagged primers that amplify polymorphic exons from class I (A, B, C exons 2 and 3) and class II (DQ, exons 2 and 3; DRB and DPB1, exon 1) major histocompatibility genes. Amplified DNA products from unique MID-tagged products (up to 48 MIDs) were pooled in equimolar ratios and subjected to library preparation. Libraries will be quantified using the KAPA library quantitation kit (Roche Kapa Biosystems, Wilmington, MA) and High sensitivity D1000 sceptermap on an Agilent 2,200 Tapestation (Agilent Scientific Instruments, Savage, MD) for concentration and size distribution. Normalized libraries were sequenced on the Illumina MiSeq platform using the MiSeq V3 600-cycle kit (Illumina, Inc., San Diego, CA). Sequences were separated by MID tags and alleles called using an accredited HLA allele caller software pipeline that minimizes the influence of sequencing errors. Alleles were called using the latest International ImMunoGeneTics HLA allele database as a reference library. Samples to report integrity were checked using proprietary and accredited laboratory information and management system. Furthermore, the HLA analyse reporting software performs comprehensive allele balance and contamination checks on the final dataset.

Statistical considerations
Descriptive statistics were reported as means and SDs or percentages. HLA allelic associations with cases and controls were considered both singly and in clusters defined by haplotype, supertype, and binding pocket structure. Initially, HLA-specific differences between cases and controls were assessed by Fisher’s exact test. Multivariable logistic regression analyses adjusted for age, gender, and ethnicity were then utilized to further explore these associations and provide Odds ratios (OR) with 95% confidence intervals.

RESULTS
Cohort characteristics
Seventy-three well-phenotyped cases, with an average age at time of AIH diagnosis of 44.4 years, were included in the analysis. A majority of the cases were women (81%) and white (90.4%). Eighty-seven controls were included in this study. The controls were no different than cases according to sex (87%) and ethnicity (94.3%) (Table 1); however, the age of enrollment into the GRACE cohort was lower for the controls (45 vs 53 [mean age in years], P = 0.05). Among cases, anti-nuclear antibody was positive in 59%, 94% had a positive (1:40) anti-smooth muscle antibody, and no patients had a positive liver kidney microsomal antibody within 6 months of AIH diagnosis. The mean of the alanine aminotransferase at AIH diagnosis was 546 U/L (638), aspartate aminotransferase: 458 U/L (552), total bilirubin: 4.9 mg/dL (7.1), international normalized ratio: 1.16 (0.16), and immunoglobulin G: 2,014 mg/dL (883).

HLA alleles, supertype, and binding pocket
In order to eliminate ethnically derived polymorphic allele contributions to HLA, only white cases were included in this analysis (66 cases and 82 controls) (Table 2). Forty-nine cases (74%) had at least one of the previously identified risk alleles HLA *03:01 or *04:01 compared to 34 (41%) controls (P = 0.05). Individually DRB1*03:01 was strongly associated with AIH (OR: 3.09, P = 0.002), yet DRB1*04:01, was only observed to have marginal additional risk in cases (OR: 1.6, P = 0.1). All carriers of DRB1*03:01 also carried DQA*05:01 and DQB1*02:01, as is consistent with the DR3 haplotype.

Table 1. Demographics of AIH cases and healthy controls

<table>
<thead>
<tr>
<th></th>
<th>Cases (n = 73)</th>
<th>Control (n = 87)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (% woman)</td>
<td>81</td>
<td>87</td>
<td>NS</td>
</tr>
<tr>
<td>Age (yr) at enrollment, mean (SD)</td>
<td>53 (16)</td>
<td>45 (13)</td>
<td>0.05</td>
</tr>
<tr>
<td>Race (% white)</td>
<td>90.4</td>
<td>94.3</td>
<td>NS</td>
</tr>
<tr>
<td>Age (yr) at AIH diagnosis, mean (SD)</td>
<td>44.4 (17)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>AIH, autoimmune hepatitis; NS, not significant.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
HLA class 1 allelic differences between cases and controls included the consideration of the supertype grouping and peptide binding structure of the HLA molecule. Alleles belonging to the A03 (37.9% vs 57.3%, \( P = 0.02 \)) and B44 (28.8% vs 50%, \( P = 0.01 \)) HLA supertype were observed less frequently among cases compared to controls. Further refinement of the class I molecules by binding pocket structure revealed that the amino acid sequence Y(F/T)AVMENV(H/Q)Y, corresponding to HLA-A alleles A*03:01-02; *31:01; *32:02, was protective for AIH (OR: 0.3, \( P = 0.002 \)). Another protective association also existed for alleles belonging to the HLA-B binding pocket structure Y(H/Y) TVKEISNY (OR: 0.35, \( P = 0.01 \)), corresponding to HLA-B alleles: B*40:01-02; *41:02; *44:02-03; *45:01; *49:01; *50:01-02. Common amino acids across key positions of the binding pocket structure revealed that the amino acid sequence at 67–72 in the antigen-binding groove of class II molecules. Class II molecules are highly prevalent on antigen presenting cells, thus are likely to exert a strong influence on the diversity of antigens presented to native CD4+ helper T lymphocytes (8,12,24). However, not all HLA-associated AIH risk can be solely explained by DRB1 as different alleles can portray risk or protection in different populations (27,28). In addition, these top risk alleles encode identical residues (DRB71) and have been associated with different clinical outcomes (15). Despite current and previously identified HLA risk alleles, the impact of HLA architecture on AIH heterogeneity remains incompletely elucidated. An integrative approach will be required to overcome these disease gaps and will require high-quality exposure data with longitudinal measures, detailed phenotyping, and inclusion of diverse ethnic and geographic populations.

**DISCUSSION**

We utilized a well-described AIH cohort to confirm the historically observed (24,25) associations between HLA-DRB1 alleles and AIH. We provide an initial in depth assessment of HLA class I genes that identify an independent yet disease protective effect of HLA-A alleles *03:01-02;*31:01;*32:02 and HLA-B alleles *40:01-02;*41:02;*44:02-03;*45:01;*49:01;*50:01-02. Furthermore, we describe related class I supertypes and binding pocket structures that encompass these protective alleles. Despite a well-described history of variable HLA-DRB1 disease risk and protection across ethnically diverse populations with AIH (8,10,24), our data suggest genetic risk with the HLA locus expands beyond the class II genes in whites. Further dissection and clarification of a more complete HLA risk association with AIH will help provide a comprehensive understanding of AIH pathogenesis and assist with future strategies aimed at improving diagnosis and individualized treatment strategies.

AIH is undoubtedly a polygenic disorder for which the HLA locus is the most important genetic contribution. The results from the only AIH genome-wide association study revealed only one strict genetic association rooted in the HLA locus (rs2187668, OR: 2.9; \( P = 1.3 \times 10^{-48} \)) (10). Historically, candidate gene studies have also supported this finding as well as DRB1*0401 as major susceptibility alleles in AIH among both European and North American patients (26). We confirmed this association within our AIH cohort, as nearly 75% of our cases had at least one copy of alleles DRB1*0301 or *0401 (Table 2). These alleles share the 71K amino acid polymorphism, encoding the LLEQKR amino acid sequence at 67–72 in the antigen-binding groove of class II molecules. Class II molecules are highly prevalent on antigen presenting cells, thus are likely to exert a strong influence on the diversity of antigens presented to native CD4+ helper T lymphocytes (8,12,24). However, not all HLA-associated AIH risk can be solely explained by DRB1 as different alleles can portray risk or protection in different populations (27,28). In addition, these top risk alleles encode identical residues (DRB71) and have been associated with different clinical outcomes (15). Despite current and previously identified HLA risk alleles, the impact of HLA architecture on AIH heterogeneity remains incompletely elucidated. An integrative approach will be required to overcome these disease gaps and will require high-quality exposure data with longitudinal measures, detailed phenotyping, and inclusion of diverse ethnic and geographic populations.

**Table 2. HLA associations with autoimmune hepatitis among whites**

<table>
<thead>
<tr>
<th>HLA carriage</th>
<th>Cases N (%)</th>
<th>Controls N (%)</th>
<th>Unadjusted OR</th>
<th>P</th>
<th>Adjusteda OR</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-DRB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>03:01(^a)</td>
<td>34 (51.5)</td>
<td>21 (25.6)</td>
<td>3.09</td>
<td>0.002</td>
<td>3.25</td>
<td>0.002</td>
</tr>
<tr>
<td>04:01(^a)</td>
<td>15 (22.7)</td>
<td>13 (15.9)</td>
<td>1.56</td>
<td>0.3</td>
<td>2.06</td>
<td>0.1</td>
</tr>
<tr>
<td>HLA-A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A03 supertype</td>
<td>25 (37.9)</td>
<td>47 (57.3)</td>
<td>0.45</td>
<td>0.02</td>
<td>0.4</td>
<td>0.01</td>
</tr>
<tr>
<td>B pocket: Y(F/T)AVMENV(H/Q)Y(^b)</td>
<td>16 (24.2)</td>
<td>40 (48.8)</td>
<td>0.34</td>
<td>0</td>
<td>0.3</td>
<td>0.002</td>
</tr>
<tr>
<td>HLA-B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B44 supertype</td>
<td>19 (28.8)</td>
<td>41 (50.0)</td>
<td>0.4</td>
<td>0.01</td>
<td>0.44</td>
<td>0.03</td>
</tr>
<tr>
<td>B pocket: Y(H/Y) TVKEISNY(^c)</td>
<td>14 (21.2)</td>
<td>36 (43.9)</td>
<td>0.34</td>
<td>0.01</td>
<td>0.35</td>
<td>0.01</td>
</tr>
</tbody>
</table>

\(^a\)Logistic regression, adjusting for carriage of DRB1*03:01, age and sex.


\(^c\)HLA-B alleles: B*40:01-02; *41:02; *44:02-03; *45:01; *49:01; *50:01-02. Common amino acids across key positions of the B binding pocket: p7-p9-p24-p34-p45-p63-p66-p70-p99.
increased frequencies of HLA-C*02, 04, and 06 (31). Thus, it not only remains highly likely that HLA class I molecules are critical for the immunologic differentiation of self vs nonself via a traditional CD4/CD8 T cell-mediated model, but also may impact innate immunity and regulation of adaptive immune responses via NK cell activation in differing phenotypic types of AIH.

We observed key differences between cases and controls according to A03 and B44 supertypes that correlated with specific peptide binding structures within the class I molecules (Table 2). Our observation supports an additive risk to AIH in white patients beyond the traditional class II genetic risks and necessitates further dissection of the class I role in AIH pathogenesis. Disease susceptibility linked to Kir genes has been reported in other autoimmune liver diseases as well, such as primary sclerosing cholangitis (32). Karlsen et al. (32) reported similar Kir gene frequencies among primary sclerosing cholangitis (PSC) and controls but with a reduction of ligands to inhibitory Kirs 3DL1 and 2DL1. AIH and PSC both share an association with the 8.1 ancestral haplotype in whites, which includes DRB1*0301. Other key HLA associations within PSC, specifically genes HLA-B8 and HLA-DR13, have not been observed as independent risk factors for AIH (5). There is also risk overlap at the DRB1 gene in primary biliary cholangitis; however, primary biliary cholangitis has also been historically associated with risk alleles in HLA-DQB1 in whites (33).

Environmental exposures, an integral part to overall disease risk in many complex autoimmune diseases (34,35), also fit well with this model given strong associations with hepatrophic viral exposure and autoimmune liver diseases (36,37).

We admit our study does have few limitations, yet we strongly believe these have little impact on the key experimental findings. Despite an overall low study number, we enriched the cohort by ensuring a highly accurate phenotype of all AIH cases and controls. In order to eliminate other genetic confounding, we also only included patients that were white in the final analysis. We do not believe that a study with this scope will ultimately disentangle HLA contributions to AIH, but we do believe it is an integral step in the clarification of HLA-associated disease risk. We only completed deep sequencing of the HLA region, as we believe data to date suggests the HLA region is the area most likely to yield the largest effect size for the risk of AIH. We plan to expand this study cohort, include all ethnicities, and integrate other key environmental contributions and HLA-environment interactions, as this likely more accurately describes the underpinning to AIH pathogenesis.

In conclusion, broken tolerance and failure of immunologic homeostasis in AIH are likely the result of permissive HLA architecture within a wide array of environmental triggers, but this pathogenesis paradigm remains poorly understood. We report results of HLA class I/II sequencing within a well-described adult white AIH cohort. Our data support historical associations between AIH and HLA-DRB1 alleles as well as identifies new findings of independent and protective HLA-A and HLA-B alleles. Clarification of the HLA architecture in AIH will certainly provide a more comprehensive understanding of disease pathogenesis but may also promote an individualized approach to treatment and management.

**CONFLICTS OF INTEREST**

**Guarantor of the article:** Craig Lammert, MD.

**Specific author contributions:** C.L.: manuscript drafting, cohort development, and sample attainment. E.J.M.: statistical support, data interpretation, and critical review of manuscript. N. C: experiment design and critical review of manuscript. E.J.P.: experimental design, sample preparation, data interpretation, and critical review of manuscript.

**Financial support:** This publication was made possible with support provided to C.L. from Grant Numbers, KL2TR001106, and UL1TR001108 (A. Shekhar, PI) from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award. Also in part, with support from the Indiana Clinical and Translational Sciences Institute funded, in part by Award Number UL1TR002529 from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award.

**Potential competing interests:** None.

**Study Highlights**

<table>
<thead>
<tr>
<th>WHAT IS KNOWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔ AIH is the result of broken immunologic homeostasis.</td>
</tr>
<tr>
<td>✔ Genetics and environment are central to AIH pathogenesis.</td>
</tr>
<tr>
<td>✔ The ancestral haplotype is the most studied genetic risk in AIH.</td>
</tr>
<tr>
<td>✔ Investigative resolution of HLA in prior studies is limited.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WHAT IS NEW HERE</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔ We identify new independent HLA-A and -B alleles protective against AIH.</td>
</tr>
<tr>
<td>✔ HLA class I effects via killer immunoglobulin-like are important in pathogenesis.</td>
</tr>
</tbody>
</table>

**TRANSLATIONAL IMPACT**

| ✔ Further study and dissection of HLA class I alleles is needed to clarify AIH heterogeneity and pathogenesis. |

**REFERENCES**


