The handle http://hdl.handle.net/1887/33221 holds various files of this Leiden University dissertation

**Author:** Heemst, Jurgen van  
**Title:** Towards a molecular basis for the association of HLA, rheumatoid arthritis and autoantibodies  
**Issue Date:** 2015-06-17
CHAPTER 6

Fine-mapping the HLA locus in rheumatoid arthritis and other rheumatic diseases: Identifying causal amino acid variants.

Jurgen van Heemst, Tom W.J. Huizinga, Diane van der Woude and René E.M. Toes.

Curr Opin Rheumatol. 2015 May;27(3):256-261
ABSTRACT

Purpose of review To provide an update on- and context of the recent findings obtained with novel statistical methods on the association of the Human leucocyte (HLA) locus with rheumatic diseases.

Recent findings Novel SNP fine-mapping data obtained for the HLA locus have indicated the strongest association with aa positions 11 and 13 of HLA-DRB1 molecule for several rheumatic diseases. On the basis of these data, a dominant role for position 11/13 in driving the association with these diseases is proposed and the identification of causal variants in the HLA-region in relation to disease-susceptibility implicated.

Summary The human leukocyte antigen (HLA) class II locus is the most important risk factor for several rheumatic diseases. Recently, new statistical approaches have identified previously unrecognized amino acid positions in the HLA-DR-molecule that associate with anti-citrullinated protein antibody negative and positive rheumatoid arthritis. Likewise, similar findings have been made for other rheumatic conditions such as giant cell arteritis and systemic lupus erythematosus. Interestingly all these studies point towards an association with the same amino acid positions: amino acid position 11 and 13 of the HLA-DR beta chain. As both these positions influence peptide binding by HLA-DR and have been implicated in antigen presentation, the novel fine-mapping approach is proposed to map causal variants in the HLA-region relevant to RA and several rheumatic diseases. If these interpretations are correct, they would direct the biological research aiming to address the explanation for the HLA-disease association. Here, we provide an overview of the recent findings and evidence from literature that, while relevant new insights have been obtained on HLA-disease associations, the interpretation of the biological role of these amino acids as causal variants explaining such associations should be taken with caution.
INTRODUCTION

Rheumatoid Arthritis (RA) is a chronic systemic inflammatory disorder predominantly affecting synovial joints. In the past pivotal pathophysiological insight has been obtained by the identification of the anti-citrullinated protein antibodies (ACPA), a group of autoantibodies specifically directed to citrullinated proteins[1]. ACPA are highly specific for RA and directed against antigens that are also expressed in the inflamed joint[2].

It is now clear that RA represents two main syndromes, ACPA-positive- and ACPA-negative disease, that are most likely etiologically distinct as different genetic risk factors are underlying the two distinct disease phenotypes. This was first evidenced by the finding that different Human Leukocyte Antigen (HLA)-genes are predisposing to either ACPA-positive or ACPA-negative disease, and boosted by the results of Whole Genome Association (WGA)-studies[3-5]. In the recent past, the genetics of RA has undergone a revolution through the identification of several novel genetic risk factors. Although the identification of genetic regions triggering human disease represents a major step forward, one of the biggest challenges to date is to elucidate the biological pathways underlying these associations to obtain a thorough understanding of disease pathogenesis. At least 20% of all genetic variance contributing to RA is explained by the HLA class II region, which is considerably more than that of all known non-HLA common genetic variants together[6]. Thus, the HLA-region is by far the most important known genetic risk factor for RA. Nonetheless, the biological pathway(s) causing this association is/are still unknown. Therefore, one of the major questions in the field is how the HLA class II region is contributing to disease pathogenesis, how the immune response underlying the disease is generated and which role it is playing in disease induction and progression.

HLA-class II-proteins present antigenic peptides to T-cells, which will result in T-cell activation and induction of antigen-specific immune reactions. Activated T-cells can mediate a plethora of effects including provision of help to B-cells necessary for antibody isotype-switching, affinity-maturation and memory B-cell-formation. Although the association between HLA and RA is known for over 30 years, the biology explaining this association and the nature of the presented antigen is poorly understood. One of the prevailing ideas to date is the notion that the HLA-molecules that predispose to RA have an enhanced ability, in comparison to the HLA-alleles that do not predispose to ACPA-positive disease, to present citrullinated epitopes[7-8]. The HLA-molecules associating with RA are generally referred to as HLA-Shared-Epitopes (HLA-SE) as the HLA-DR-molecules encoded by the predisposing HLA-haplotypes share a common sequence in the peptide-binding groove of the molecule[9]. Through the ability of the HLA-SE-molecules to present citrullinated antigens, these molecules are best suited to activate T-cells reactive to citrullinated antigens, thereby facilitating the induction of T-cell responses that can provide help to B cells recognizing the same antigens. Whether this hypothesis
is correct is not known, but it has been shown that the pocket in the HLA-antigen-binding groove (the P4-pocket) that is, in part, shaped by the SE-sequence, is able to accommodate citrulline, but not arginine, the amino acid from which citrulline is formed. Nonetheless, it is not known whether these findings apply for all HLA-SE-molecules or for all citrullinated epitopes. To better understand the contribution of the HLA-SE-alleles in ACPA-positive RA, more refined insights the HLA-RA connection is important. Obviously, a similar notion applies for other rheumatic diseases where the contribution of the HLA-region to disease pathogenesis is also not well understood.

FINE MAPPING THE HLA LOCUS

The HLA locus is complex as it contains many genes that can be highly polymorphic[10]. Also, this region is characterized by extensive linkage disequilibrium (LD) making it difficult to identify the responsible genes and/or causal variants. Recent studies have revealed relevant new insights into the association of the HLA-region with rheumatoid arthritis and several other rheumatic diseases[4, 8-12]. By fine-mapping of the HLA locus and imputation of amino acids (aa) sequences of HLA-molecules from SNP genotype data, HLA-genetics have been made feasible in large groups of patients and controls. This allowed new insights into the HLA-association for many different diseases (detailed description in [4]).

Analysis of the association of individual imputed amino acids (aa) throughout the HLA-locus revealed specific positions in the HLA-DRB1-allele that associated best with disease. As the identified HLA-variants are part of peptide-binding grooves, these new findings were proposed to map causal variants in the disease-associating HLA-molecules. This approach has now been applied to both ACPA-negative[11, 12] and ACPA-positive RA [4, 11, 13], but also to other rheumatic diseases including Systemic Lupus Erythematosus (SLE) and Giant Cell Arteritis (GCA)[14-15]. Previously, dominant risk genotypes of these diseases were documented on the basis of conventional HLA-genotyping. Susceptibility to ACPA-negative RA was shown to be associated with HLA-DRB1*03[3, 11], whereas ACPA-positive RA and GCA are most strongly associated with HLA-DRB1*04[16, 17]. SLE susceptibility in Koreans was reported to be linked to HLA-DRB1*15[18].

Intriguingly, the novel SNP fine-mapping data obtained for the HLA locus showed the strongest association with aa positions 11 and 13 of HLA-DRB1 for all these rheumatic diseases (summarized in Table 1). Importantly, positions 11 and 13, which are in tight LD, are critically relevant for shaping several peptide-binding pockets: Aa position 11 is involved in shaping peptide-binding pocket 6, whereas aa position 13 affects the shape of pocket 4[19]. Peptide-binding pockets are part of the peptide-binding groove of the HLA-molecule that are involved in epitope-presentation. The interpretations of these combined findings often imply a dominant
role for position 11/13 in driving the association with disease or implicate the identification of causal variants in the HLA-region in relation to disease-susceptibility. For instance, when these aa positions where found in association with SLE, it was concluded that these aa positions at the epitope-binding groove of HLA-DRβ1 are responsible for most of the association between SLE and MHC. Likewise, when it was shown that the variants that associate with RA are the same variants that associated with risk for development of follicular lymphoma, the presence of a common HLA-DR antigen-driven mechanism for the pathogenesis of follicular lymphoma (FL) and RA was suggested[20].

**PERSPECTIVE**

If the interpretations, such as the one mentioned above, are correct, it would have an important impact on the design of experiments aiming to address the biological explanation for the HLA-disease association. For example, in the case of RA, less efforts should be devoted to the elucidation of the role of the shared epitope, but instead, the contribution of aa position 11/13 should be scrutinized. Aa positions 11 is not part of pocket 4 and does not interact with SE-residues. Therefore, these findings would imply that, at the molecular level, two different mechanisms might be operative that contribute to the development of ACPA-positive RA. Similar implications would apply for other rheumatic diseases. To better understand the possible contributions of these aa associations and the findings generated by the described method, we reviewed variation between HLA-DRB1-alleles in closer detail.

An aa position can only be significantly associated with a disease if the particular position is polymorphic. HLA-DR molecules are heterodimers consisting of an alpha and a beta chain. For HLA-DR all the variation between donors is found in the beta chain. Figure 1A schematically depicts the variation in the HLA-DR b1 domain per amino acid of the top-15 most prevalent HLA-DRB1 chains present in the Caucasian population (together covering 91% of the variation in Caucasians[21]). Black dots indicate aa residues that are involved in shaping the peptide-binding pockets. As illustrated, most variation in HLA-DRB1 resides in residues critically engaged

<table>
<thead>
<tr>
<th></th>
<th>Pos1</th>
<th>Pos2</th>
<th>Pos3</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACPA- RA</td>
<td>11/13</td>
<td>-</td>
<td>-</td>
<td>[9, 10]</td>
</tr>
<tr>
<td>ACPA+ RA (Caucasians)</td>
<td>11/13</td>
<td>71</td>
<td>74</td>
<td>[6, 9, 11]</td>
</tr>
<tr>
<td>Giant cell arteritis</td>
<td>11/13</td>
<td>-</td>
<td>-</td>
<td>[13]</td>
</tr>
<tr>
<td>Systemic Lupus Erythematosus</td>
<td>11/13</td>
<td>26</td>
<td>-</td>
<td>[12]</td>
</tr>
</tbody>
</table>

Table 1: Summary of amino acid positions associated with risk to develop rheumatic diseases.
in shaping peptide-binding pockets. The finding that the statistically significant amino-acid positions are those involved in peptide presentation is therefore not surprising and does not, per se, support a more important pathophysiological role of these particular amino acid positions compared to other non-associated positions.

HLA-crystallization and peptide-elution studies have shown that peptide-binding pockets are shaped by the combination of multiple different (polymorphic) amino acids (schematically depicted in Figure 1B) and that the combination of all these amino acids will eventually determine the size, charge and hydrophobicity of each peptide-binding pocket and thereby the ligands that can be presented by the HLA-molecule [22-27]. It is therefore surprising that aa positions 11 and 13 are so strongly associated with different rheumatic diseases compared to other aa positions that are also involved in shaping these pockets as one could assume that, in case a particular pocket is crucial to the presentation of a particular (set of) antigens, also the other aa involved in the formation of the pocket would influence the binding of such antigen(s). Recently, these studies were also conducted for non-rheumatic diseases: ulcerative colitis and FL [20, 28]. Again, aa positions 11 and 13 were most strongly associated with risk for these two diseases. Together,
these findings and considerations make it conceivable that the applied approach does not identify causal variants as implied, but may contain a bias towards these two aa residues. Such bias could be introduced by the fact that some aa positions are more polymorphic than others. As depicted in Figure 1A, aa positions 11 and 13 are the most polymorphic amino acid positions in the HLA-DRB1-gene. It is conceivable that the applied approach biases towards amino acid positions that can subdivide the different HLA class II molecules into the largest number of groups, i.e. the most polymorphic amino acid positions. Thus, positions 11 and 13 are likely to be the best surrogate markers for defining HLA-genotype as they contain the most discriminative ability. This is illustrated in Figure 1C for ACPA-positive RA which is known to associate most strongly with HLA-DR4[29]. Stratification on position 13 generates 6 different groups on the basis of the 6 different aa that can be present in this position. HLA-DR4 will be in a group without other HLA-DR subtypes. Stratification on a less polymorphic position, e.g. position 9, subdivides the HLA class II molecules in only 2 groups. HLA-DR4 will now be in a subgroup with HLA-DR3, HLA-DR8, HLA-DR11 and HLA-DR13 that will all dilute the risk effect of HLA-DRB1*04. Therefore, a statistical approach based upon the presence of individual aa in certain positions will unlikely identify aa at position 9, as this position is relatively non-polymorphic. Nonetheless, in biological terms, this position shapes one of the pockets: pocket 9, and hence will impact HLA-ligand binding. Likewise, giant cell arteritis is also associated with HLA-DRB1*04 and therefore similar aa position will likely associate with this disease although the biological basis is likely different[16]. For ACPA-negative RA, HLA-DRB1*03 is the major risk determinant[3, 11, 30]. Stratification on the basis of aa position 13 will generate an HLA-DR3 subgroup with HLA-DRB1*11 and HLA-DR13, whereas stratification based on aa position 9 will generate a subgroup of HLA-DR3 with HLA-DR4, HLA-DR8, HLA-DR11 and HLA-DR13 (Figure 1C). Hence, stratification on aa position 13 will again be favored over less polymorphic positions, e.g. position 9. Finally, SLE is most strongly associated with HLA-DR15 in Asians and stratification on position 13 will generate an HLA-DR15 subgroup with DR16, whereas stratification on position 9 will generate a subgroup of HLA-DR15 with HLA-DR1, HLA-DR7 and HLA-DR16[18]. Stratification on aa positions 11 and 13 does not only generate the largest number of different groups, but also provides the best resolution due to the fact that the various aa variants predisposing to different rheumatic diseases are quite evenly distributed in size. This could further explain why these aa residues are so strongly associated with the different diseases using this statistical aa fine-mapping approach (Figure 1D).
CONCLUSION

The development of new methods and tools for genetic data analyses in the HLA-region have been rewarding as they further refined the HLA-diseases associations by allowing inclusion of large datasets that lacked conventional HLA-genotyping data. Based on the analyses described above we feel, however, that conclusions on the biological role of amino acid positions obtained via this novel fine-mapping approach should be taken with caution as it appears that this approach may contain a bias towards particular aa positions that, at a single amino acid level, best mark the associated HLA-DRB1 genotype. These considerations are in agreement with MHC-ligand studies that clearly show that the combination of all residues shaping peptide-binding pockets will eventually determine the type of ligands that are presented by these molecules.

REFERENCES


