Review article

Immunogenetics of type 1 diabetes: A comprehensive review

Janelle A. Noble*

Children’s Hospital Oakland Research Institute, Oakland, CA 94609, USA

Contents

1. Introduction ............................................. 102
2. HLA and autoimmune disease ............................. 102
   2.1. HLA structure ........................................ 102
   2.2. HLA genes ........................................... 102
   2.3. HLA polymorphism and nomenclature ............... 103
   2.4. HLA genotyping ...................................... 104
3. HLA association with T1D ................................. 105
   3.1. DRB1-DQA1-DQB1 haplotype association .......... 105
   3.2. DP and T1D risk .......................................... 106
   3.3. HLA class I association with T1D risk .............. 106
   3.4. HLA variation among populations ................. 107
4. Association of other immune genes with T1D ............... 108
   4.1. Genes in the HLA region .............................. 108
   4.2. Genes outside the HLA region ..................... 108
   4.3. KIR association with T1D ........................... 108
5. Genetic prediction of T1D .................................. 109

Abstract

Type 1 diabetes (T1D) results from the autoimmune destruction of insulin-producing beta cells in the pancreas. Prevention of T1D will require the ability to detect and modulate the autoimmune process before the clinical onset of disease. Genetic screening is a logical first step in identification of future patients to test prevention strategies. Susceptibility to T1D includes a strong genetic component, with the strongest risk attributable to genes that encode the classical Human Leukocyte Antigens (HLA). Other genetic loci, both immune and non-immune genes, contribute to T1D risk; however, the results of decades of small and large genetic linkage and association studies show clearly that the HLA genes confer the most disease risk and protection and can be used as part of a prediction strategy for T1D. Current predictive genetic models, based on HLA and other susceptibility loci, are effective in identifying the highest-risk individuals in populations of European descent. These models generally include screening for the HLA haplotypes “DR3” and “DR4.” However, genetic variation among racial and ethnic groups reduces the predictive value of current models that are based on low resolution HLA genotyping. Not all DR3 and DR4 haplotypes are high T1D risk; some versions, rare in Europeans but high frequency in other populations, are even T1D protective. More information is needed to create predictive models for non-European populations. Comparative studies among different populations are needed to complete the knowledge base for the genetics of T1D risk to enable the eventual development of screening and intervention strategies applicable to all individuals, tailored to their individual genetic background. This review summarizes the current understanding of the genetic basis of T1D susceptibility, focusing on genes of the immune system, with particular emphasis on the HLA genes.

© 2015 Elsevier Ltd. All rights reserved.
1. Introduction

Susceptibility to type 1 diabetes, in which the immune system destroys insulin-producing beta cells, has a strong genetic component, as evidenced by strong disease concordance in monozygotic twin pairs [1]. Four decades of study have shown that immune genes, specifically, those that encode classical Human Leukocyte Antigens (HLA); confer the strongest genetic risk for disease. This review describes the complex genetics of the HLA region of the human genome (Section 2) and summarizes the current state of knowledge of HLA-associated T1D susceptibility, including differences among populations (Section 3). Associations of other immunologically relevant genes are described, and recent studies investigating genes encoding Killer-cell Immunoglobulin-like Receptors (KIR) on natural killer (NK) cells are presented in the context of T1D susceptibility (Section 4). Finally, the utility of immunogenetics in the prediction of T1D risk is addressed (Section 5).

2. HLA and autoimmune disease

The association of Human Leukocyte Antigens (HLA) with type 1 diabetes was first reported in the 1970s [2–4]. Studies of HLA association with disease were prompted by the striking observation of the association of HLA-B*27 (called HL-Aw27 at the time) with ankylosing spondylitis [5]. At about the same time, HLA association was observed for several other autoimmune diseases, including rheumatoid arthritis (RA), celiac disease (CD), systemic lupus erythematosus (SLE), and multiple sclerosis (MS) [6–9], underscoring the importance of HLA in immune regulation. More than forty years and thousands of genetic association studies later, genes in the HLA region of the human genome remain the primary genetic risk determinants for T1D, accounting for approximately half of the genetic susceptibility to disease, estimated from studies of affected sibling pairs [10,11]. The genetics of HLA are complex and can be difficult to interpret. Understanding of HLA association with T1D requires a basic understanding of the biology of HLA. Details about the polymorphisms, structure, organization, and nomenclature of the classical HLA genes, as well as descriptions of available HLA genotyping methods, are presented in the following sections (2.1 through 2.4).

2.1. HLA structure

Classical HLA molecules are cell-surface proteins that bind and present peptide antigens for recognition by the T cell receptor (TCR). HLA is categorized into two classes, with three types of antigens in each class. Class I molecules are called A, B, and C and consist of a polypeptide chain that forms a heterodimer with the relatively invariant β-2 microglobulin. Class II molecules are called DR, DQ, and DP and consist of a heterodimer created from two polypeptides (α and β). Despite these differences, the overall structure of HLA class I and class II molecules is quite similar and is illustrated schematically in Fig. 1. The structure of HLA-A2 was solved in 1987, illustrating that the portion of the protein distal to the cell surface folds into a β-pleated sheet on which two α-helices create a groove into which peptides bind [12]. The shape of the peptide binding groove and charges within it determine the repertoire of peptides that can bind to a given HLA. The TCR recognizes the combination of HLA and peptide, creating the trimolecular complex that initiates the immune response (Fig. 2).

2.2. HLA genes

The genes that encode HLA molecules are located on
chromosome 6p21. The organization of the exons and introns is similar among the HLA-encoding genes, with exon 1 containing 5’ untranslated sequence and encoding the signal peptide, and exons 2 and 3 (and 4 for class I genes) each encoding an immunoglobulin-like domain that contributes to the extracellular portion of the molecule (Fig. 3). The peptide binding groove is formed by the domains encoded by exons 2 and 3 for class I (e.g., HLA-A) and by exon 2 from each of the two genes encoding a class II molecule (e.g., HLA-DQA1 and -DQB1); thus, DNA-based HLA genotyping systems minimally include those exons. Each class I HLA (A, B, and C) is encoded by a single gene (HLA-A, -B, and -C, respectively). For class II antigens, the heterodimer is formed from the products of two genes, e.g., HLA-DQA1 encodes the α chain and HLA-DQB1 encodes the β chain of the DQ antigen, HLA-DPA1 encodes the α chain and HLA-DPB1 encodes the β chain of the DP antigen, and HLA-DRA1 encodes the α chain and HLA-DRB1 encodes the β chain of the DR antigen. Additional loci encoding DRβ chains are found on some chromosomes. Some additional DRB genes are pseudogenes, but HLA-DRB3,-DRB4, and -DRB5 all produce a functional DRβ chain that can dimerize with the product of the DRA1 gene. HLA-DRB3,-DRB4, and -DRB5 genes segregate as a single locus, i.e., a maximum of one of them is found on any given chromosome. HLA-DRB1 is found on all copies of chromosome 6, but the other three functional genes are only present on some chromosomes, dependent on the DRB1 gene on the chromosome. Table 1 shows which of the three additional DRB coding genes is found on the same chromosome with each DRB1 gene group.

2.3. HLA polymorphism and nomenclature

HLA genes are the most polymorphic known in the human genome. While most genes are monomorphic or have a few variant sequences, classical HLA genes can have thousands (Table 2) [13,14]. As the table shows, most of the polymorphism for class II is found in the -B, rather than the -A gene. This is particularly true for DR-encoding genes, where the DRA1 gene is essentially invariant, and the DRB1 gene is highly polymorphic. In cases where a second DRB coding gene is present, its pairing with the DRA1 gene presents a mechanism to increase the diversity of the DR antigens on the cell surface. For DQ, the genes encoding both polypeptides are polymorphic, increasing the number of potential cell-surface DQ molecules to four via combinatorial diversity (see Section 3.1). For DP, nearly all of the diversity is encoded in the DPB1 gene. The DPA1 gene has limited polymorphism; in general, only small number of DPA1 alleles is observed in a given population. Most of the polymorphic sites in HLA genes are found in the exons encoding the peptide-binding groove. The basis for the polymorphism in these exons is thought to result from selection. The rapid increase in the number of recognized HLA alleles has led to an evolution in the nomenclature used to describe them. The current nomenclature was adopted in 2010 and is depicted in Fig. 4. Numeric fields separated by colons describe four levels of resolution. Previous versions of HLA nomenclature did not employ colon delimiters, which were added because the number of variants in some allele groups exceeded 99. In general, the first field describes a serologically defined allele group, and the second field identifies a unique protein sequence encoded by the allele within that group. Fields 3 and 4 describe silent polymorphisms and non-exonic polymorphisms, respectively. The extremely large number of alleles leads to the inability to make allele-level genotype calls with all but the highest-resolution genotyping technology (discussed in Section 2.4, below), which can lead to difficulty in data interpretation among studies performed with different genotyping methods. Fig. 5 illustrates that a given allele can have multiple designations,

![Fig. 3](image-url)  
**Fig. 3.** Organization of the HLA-encoding genes. The structure of the single chain of HLA class I genes is shown, along with the class II A and B genes, encoding α and β chains of the HLA molecule, respectively. *For class I, the HLA-B gene lacks exon 8, and for class II, the DPB1 gene lacks exon 6.*

<table>
<thead>
<tr>
<th>Locus</th>
<th>Number of alleles*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3107</td>
</tr>
<tr>
<td>B</td>
<td>3887</td>
</tr>
<tr>
<td>C</td>
<td>2623</td>
</tr>
<tr>
<td>DRA1</td>
<td>7</td>
</tr>
<tr>
<td>DRB1</td>
<td>1726</td>
</tr>
<tr>
<td>DRB3</td>
<td>59</td>
</tr>
<tr>
<td>DRB4</td>
<td>15</td>
</tr>
<tr>
<td>DRB5</td>
<td>21</td>
</tr>
<tr>
<td>DQA1</td>
<td>780</td>
</tr>
<tr>
<td>DQB1</td>
<td>39</td>
</tr>
<tr>
<td>DPA1</td>
<td>520</td>
</tr>
<tr>
<td>DPB1</td>
<td></td>
</tr>
</tbody>
</table>

* Numbers are taken from the IMGT/HLA database release 3.20.0, April, 2015.
depending on the level of resolution of the genotyping method, and that certain allele pairs, such as HLA-A:01:01:01:01 and HLA-A:01:01:01:02N, can only be distinguished by the highest level of genotyping resolution, including sequencing of intronic and untranslated sequence. Fortunately, for most disease association studies, the relevant information for function resides in the first two fields (usually four-digits), which include information on the amino acid sequence of the encoded protein. In most cases, silent polymorphisms and polymorphisms in intronic and untranslated regions do not affect function; thus, two-field resolution is sufficient. In the example given, however, the designation “N” in the allele name HLA-A:01:01:01:02N designates that the allele is not expressed (null); thus, HLA-A:01:01:01:01 and HLA-A:01:01:01:02N are expected to be functionally different.

2.4. HLA genotyping

The technology for HLA typing is evolving at a rapid rate. Histocompatibility testing began in the 1960s and was performed by cell-based methods using serum from multi-parous women, particularly those with pregnancies from more than one father, which contained antibodies to one or more paternal HLA molecules that were present in the fetus. Extensive standardization of reagents and protocols among HLA typing laboratories was required, and typing assays resulted in very low resolution with frequent blanks. Adoption of DNA-based genotyping technology in the 1980s allowed better ability to identify and resolve the individual HLA alleles present in a subject and began an explosion in the number of known alleles that is still going on today. Early DNA technology used restriction enzymes and Southern blotting and was known as Restriction Fragment Length Polymorphism (RFLP) [15]. The advent of the Polymerase Chain Reaction allowed the development of PCR-based genotyping methods, including sequence-specific oligonucleotide probe hybridization (SSOP) [16], sequence-specific priming (SSP) [17], and sequence-based typing (SBT) [18]. Currently, next-generation sequencing (NGS)-based methods are available for HLA genotyping, using either exons only or larger gene segments, up to and including whole HLA genes [19,20]. Each of these advances in technology has increased the level of resolution of the resulting genotype data. Fig. 6 depicts the relative levels of resolution of various HLA genotyping methods. The method selected for a given study depends on the intended use of the data. For example, full sequencing of HLA-B to support a diagnosis of ankylosing spondylitis would not be cost-effective when a simple clinical test for HLA-B*27 could be used [21]. On the other hand, the severe drug hypersensitivity to the nucleoside reverse transcriptase inhibitor abacavir exhibited by HIV patients carrying HLA-B*57:01 is not seen in patients carrying HLA-B*57:03, underscoring the need for at least two-field resolution of HLA genotyping for treatment decisions for HIV positive patients [22].

For a disease association study, the decision of which of the many available HLA genotyping methodologies to use greatly affects the results. Low resolution genotyping results in a small number of categories with large numbers of data points in each category, leading to good statistical power, with the caveat that each category may contain alleles with different effects on disease susceptibility, possibly masking true effects of individual alleles within the group. Conversely, allele-level resolution will allow the effects of each individual allele, haplotype, or genotype to be examined; however, the number of data points for each category, especially when analyzing genotype associations, may be so low that statistical power will be insufficient to reveal disease association effects. Particularly for studies involving large numbers of samples, cost is a major factor in determining the genotyping method. In general, higher resolution means higher cost. A well-designed study will incorporate a method that provides adequate resolution to address the scientific hypotheses being tested while remaining within the resources available for the study. For these reasons, HLA genotype resolution varies widely among published studies, and the interpretation of and comparisons among studies must take this into account.

In recent years, methods have been developed to impute genotype data, i.e., assigning a genotype call to a given polymorphic position without direct experimental testing using data from nearby single nucleotide polymorphisms (SNPs) and knowledge of haplotype patterns within a given population. This is a highly cost-effective method of data generation, since it does not depend on expensive laboratory genotyping procedures. Imputation can be

---

**Fig. 4.** HLA nomenclature. The figure illustrates the designation of an allele with the current standard nomenclature.

**Fig. 5.** Alternate designations for HLA alleles. Five different designations are shown that might be reported for each of two HLA-A alleles. In this case, only the highest level HLA genotyping can distinguish the two alleles from one another.

**Fig. 6.** Resolution of HLA genotyping methods. HLA genotyping methods are listed in order of increasing resolution. NGS – Next generation sequencing.
effective in cases where a limited number of known haplotypes are well defined in a homogeneous population. Both complexity of the number of haplotypes and admixture or substructure of populations can greatly decrease the efficacy of these methods. Many groups have tried to impute HLA alleles from SNPs in the HLA region, with varying levels of success [23–28]. However, all groups realize the need for large datasets of classically determined HLA sequences from multiple populations to inform imputation strategies. Even in a Finnish population, for which HLA alleles are very well characterized, imputation methods varied among loci, with the worst performance observed for DRB1 and HLA-B genes [23]. Although calling HLA genotypes from existing genome-wide association study (GWAS) data is attractive, one must remember that these tag SNPs are simply surrogates for given HLA alleles. Very few SNPs on GWA chips lie within the HLA alleles themselves because of the tremendous amount of variation within the genes. As with low-resolution HLA genotyping, most imputed HLA calls are consistent with a group of alleles (an ambiguity string), rather than a single allele. Resolution for HLA genotyping by imputation may be as low or lower than that for serologic methods. In addition, although HLA alleles and haplotypes are well established for populations of European descent, information for most other populations is far more sparse, and, in most cases, not adequate to inform imputation algorithms properly.

Current efforts in the HLA community are directed toward standardization of data reporting in HLA studies to allow comparisons to be made among studies [29]. In the future, the time may come when whole genome sequencing becomes standard for every individual, negating the need for specific genotyping of any locus, including those encoding HLA. Currently, however, an understanding of HLA genotyping methods and nomenclature is essential for correct interpretation of disease association studies.

3. HLA association with T1D

The earliest reported HLA association with T1D was with HL-A antigen “specificity W15” [2], which corresponds to what we now call HLA-DRB1*04:01, a known risk allele for T1D. Subsequently, Nerup et al. demonstrated association with W15 and what was then called “HL-A8” [3], which corresponds to what is now called HLA-B*08:01 and is found on the common, conserved “A1-B8-DR3” haplotype that is positively associated with T1D [54]. Family-based studies confirmed these associations [4]. By 1980, the very high risk of having a heterozygous genotype with a haplotype including DRB1*03 on one chromosome and DRB1*04 on the other (commonly referred to as DR3/4), was reported [30]. Thousands of reports have been published on the association of HLA with T1D. One of the largest T1D genetics studies was the Type 1 Diabetes Genetics Consortium (T1DGC), an international effort to establish resources to discover all of the genes responsible for susceptibility to T1D [31]. The T1DGC, conceived in 2000 and started in 2001, collected over 16,000 subjects, including families, cases, and controls, from populations around the globe. All samples were genotyped for classical HLA loci, including HLA-DRB1, -DQA1, -DQB1, -DP1, -DPB1, -A, -B, and -C using PCR-SSOP methods [32] to provide HLA context for the discovery of other T1D associated genes. Much of the association data described below (Sections 3.1–3.3) come from the T1DGC.

3.1. DRB1-DQA1-DQB1 haplotype association

The highest HLA-associated T1D risk is attributable to class II DR- and DQ-encoding loci. Genes in the HLA region exhibit extensive linkage disequilibrium (LD), which is the tendency for specific alleles at two loci to be found together more often than expected, given the distance between them and expected recombination frequencies. This is particularly true for the genes encoding DR and DQ, such that, for a given population, an individual allele is usually found in only one or a few haplotype combinations. In many cases, the presence of one allele allows assumption of other alleles on the haplotype. For example, DRB1*03:01 is found coupled to DQA1*05:01 and DQB1*02:01 almost exclusively, to create the haplotype DRB1*03:01-DQA1*05:01-DQB1*02:01, commonly called “DR3” and known to be a risk haplotype for T1D. DRB1*04:01 is found with DQA1*03:01 but can have either DQB1*03:01 or DQB1*02:02 included in the haplotype. Both are referred to as “DR4” haplotypes; however, the T1D risk for the two haplotypes is very different, as shown in Table 3. This might lead to the assumption that the DQ-encoding genes are the main risk factor for T1D, and that the DRB1 locus is less, or not at all, important. However, comparison of risk for the same DRB1*04:01-DQA1*03:01-DQB1*03:02 haplotype to that for DRB1*04:03-DQA1*03:01-DQB1*03:02 haplotype could lead to the conclusion that the DRB1 locus is the primary contributor to T1D susceptibility, and the DQB1 locus matters less (Table 3). In fact, the total T1D risk is most likely due to a combination of DR and DQ molecules [10,33,34]. Thus, assessing a single locus for T1D susceptibility can be misleading; analysis of DRB1-DQA1-DQB1 haplotypes is far more informative and desirable. Like most studies of T1D genetics, the T1DGC was primarily comprised of samples from individuals of European descent. Examples of the most predisposing and most protective DRB1-DQA1-DQB1 haplotypes are given in Table 4 with predisposing Odds Ratios shown in red font, and protective shown in green.

<table>
<thead>
<tr>
<th>DRB1</th>
<th>DQA1</th>
<th>DQB1</th>
<th>Odds Ratio</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>04:01</td>
<td>03:01</td>
<td>03:01</td>
<td>0.35</td>
<td>protective</td>
</tr>
<tr>
<td>04:01</td>
<td>03:01</td>
<td>03:02</td>
<td>8.39</td>
<td>susceptible</td>
</tr>
<tr>
<td>04:03</td>
<td>03:01</td>
<td>03:02</td>
<td>0.27</td>
<td>protective</td>
</tr>
</tbody>
</table>

Table 3
HLA T1D risk can be attributable to either DRB1 or DQB1.
Table 4
Selected susceptible and protective DRB1-DQA1-DQB1 haplotypes from the Type 1 Diabetes Genetics Consortium data [34].

<table>
<thead>
<tr>
<th>haplotype</th>
<th>DRB1</th>
<th>DQA1</th>
<th>DQB1</th>
<th>OR</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR3</td>
<td>03:01</td>
<td>05:01</td>
<td>02:01</td>
<td>3.64</td>
<td>2 x 10^{-22}</td>
</tr>
<tr>
<td>DR4</td>
<td>04:05</td>
<td>03:01</td>
<td>03:02</td>
<td>11.37</td>
<td>4 x 10^{-5}</td>
</tr>
<tr>
<td>DR4</td>
<td>04:01</td>
<td>03:01</td>
<td>03:02</td>
<td>8.39</td>
<td>6 x 10^{-36}</td>
</tr>
<tr>
<td>DR4</td>
<td>04:02</td>
<td>03:01</td>
<td>03:02</td>
<td>3.63</td>
<td>3 x 10^{-4}</td>
</tr>
<tr>
<td>DR2</td>
<td>15:01</td>
<td>01:02</td>
<td>06:02</td>
<td>0.03</td>
<td>2 x 10^{-29}</td>
</tr>
<tr>
<td>DR6</td>
<td>14:01</td>
<td>01:01</td>
<td>05:03</td>
<td>0.02</td>
<td>1 x 10^{-6}</td>
</tr>
<tr>
<td>DR7</td>
<td>07:01</td>
<td>02:01</td>
<td>03:03</td>
<td>0.02</td>
<td>3 x 10^{-4}</td>
</tr>
<tr>
<td>DR7</td>
<td>07:01</td>
<td>02:01</td>
<td>02:01</td>
<td>0.32</td>
<td>2 x 10^{-9}</td>
</tr>
<tr>
<td>DR4</td>
<td>04:03</td>
<td>03:01</td>
<td>03:02</td>
<td>0.27</td>
<td>0.017</td>
</tr>
</tbody>
</table>

at 5% [37]. In most studies of DP association with T1D, only the DPB1 locus is genotyped. The nomenclature for DPB1 alleles differs from that of other classical loci in that few serologic reagents were developed for DP typing prior to the availability of DNA-based genotyping. Thus, while alleles for other loci are clustered by serologic reactivity into groups designated in the first field of nomenclature, DPB1 alleles were numbered in order of their discovery, regardless of reactivity. With the exceptions of DPB1*02:02 and DPB1*04:02, allele names were designated by the next available number in the first field, followed by “01” in the second field (e.g., ... “06:01,” “07:01,” “08:01,” ...). The discovery of more than 99 DPB1 alleles was one of the factors leading to the adoption of the colon delimiters in the latest nomenclature convention (Fig. 4).

Like DR and DQ, DP is an HLA class II molecule that can bind exogenous peptides and present them to CD4+ T cells. DPB1 alleles have been reported to be associated with several diseases, including pauciarticular juvenile rheumatoid arthritis and chronic beryllium disease [38,39]. For T1D, analyses of disease association for DPB1, or for other loci in the region, must account for LD with the DR- and DQ-encoding genes on the chromosome [40,41]. Many studies report DPB1 association with T1D. Taken together, the data from these studies show consistently that DPB1*04:02 is protective for T1D, while DPB1*03:01 and DPB1*02:02 increase T1D risk [37,42–47].

3.3. HLA class I association with T1D risk

The autoimmunity that leads to T1D results in the immune destruction of the insulin-producing beta cells in the pancreas, a process that is thought to occur by cytotoxic (CD8+) T cell killing. In addition to helping shape the T cell repertoire, class I HLA molecules present antigens to CD8+ T cells to initiate cytotoxic T cell killing; thus, specific combinations of class I HLA and peptide are likely to influence beta cell destruction. In support of this notion, HLA-A*24 correlates with low residual beta cell function in T1D patients [48]. Although HLA class II genes, specifically, those encoding DR and DQ, are the most strongly associated with T1D risk, well-powered studies, including the T1DGC, have demonstrated class I association with T1D, even after accounting for LD with the DR and DQ-encoding genes [49–52]. HLA-B*39:06, with an odds ratio (OR) of 10.31 after accounting for LD, appears to have the strongest predisposing effect on T1D risk, while HLA-B*57:01 is strongly significantly T1D protective (OR = 0.19, p = 4 × 10^{-11}) [51].

Fig. 7. Trans-encoded DQ heterodimers of the DR3/4 genotype. The DRB1-DQA1-DQB1 haplotypes in the common, high T1D risk DR3/4 genotype, found in Europeans, are shown. The trans-encoded DQA1*05:01-DQB1*03:01 haplotype is shown in the blue ellipse, and the trans-encoded DQA1*03:01-DQB1*02:01 haplotype is shown in the red ellipse. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

encodes an α polypeptide chain from the DQA1 gene and a β polypeptide chain from the DQB1 gene. The α and β chains encoded on the same chromosome can form heterodimers, and, in most cases, the α and β chains encoded on different chromosomes can form heterodimers as well, thus allowing the possible expression of four different DQ molecules on the surface of the cell. Fig. 7 illustrates the DQA1 and DQB1 pairs found in trans on a typical T1D-predisposing DR3/DR4 genotype. The combination of DQA1*05:01 and DQB1*03:02 has not been observed encoded in cis but has been hypothesized to confer very high T1D risk [34]. The other trans-encoded combination, DQA1*03:01 and DQB1*02:01, is encoded in cis on some haplotypes, particularly in African Americans, and appears to increase risk for T1D on all haplotypes on which it is found [36]. Of note, DQB1*02:02 can only be distinguished from DQB1*02:02 by a polymorphism in exon 3 which does not contribute to the peptide binding groove. This ambiguity is sometimes referred to as “02:01g” with “g” representing “group.” DQB1*02:01 and DQB1*02:02 are presumed to have identical immune function.

3.2. DP and T1D risk

The DP molecule is encoded by the DPA1 and DPB1 genes. The DPA1 gene is far less polymorphic than most other HLA genes. In fact, in the T1DGC data, only three alleles accounted for 98% of nearly 15,000 genotypes (30,000 allele calls), with DPA1*01:03 present at 76%, DPA1*02:01 present at 17%, and DPA1*02:02 present
A B*39:06 association has been observed in multiple studies, with one study suggesting that it can improve T1D risk prediction, particularly in patients who carry DRB1*04:04-DQA1*03:01-DQB1*03:02, a moderately predisposing DR4 haplotype [53]. The class I allele HLA-A*01:01 provides an illustrative example of the requirement for accounting for LD in T1D association analyses for genes in the HLA region. HLA-A*01:01 is part of the conserved, extended HLA haplotype known as “A1-B8-DR3,” which is associated with T1D risk [54]. Thus, A*01:01 appeared predisposing for disease. However, when expected allele frequencies were adjusted based on the strong LD of this allele with the DR3 haplotype, A*01:01 was actually significantly protective for T1D [50]. In addition to effects on T1D risk, HLA class I alleles have been shown to be associated with age of onset for T1D [52,55–59].

3.4. HLA variation among populations

Among world populations, prevalence (proportion of population affected) of T1D is thought to be highest in people of European descent at approximately 1 in 300. Not surprisingly, to date, the vast majority of T1D genetic studies have been performed on subjects of European descent. However, T1D is a worldwide problem. Incidence (cases per year) estimates for T1D in children under 15 years old vary from as high as 64 per 100,000 per year in Finland [60] to as low as 0.1 or less per 100,000 children per year in parts of China, Venezuela [61], and Papua New Guinea [62]. Among more than 13,000 HLA alleles have been reported [13,14], more than 90% of these alleles are rare, having been observed only once or twice, with the remaining alleles designated common and well-documented (CWD) [63]. Some common alleles appear in nearly every population studied, e.g., DRB1*03:01; however, others are population specific. The set of alleles present in any given population differs from, but usually overlaps with, the set found in other populations. Even for alleles common to multiple populations, allele frequencies and haplotype combinations vary among populations.

The study of HLA association with T1D in varied populations can reveal effects that are not apparent from the study of a single population. To understand the effect of an allele or haplotype on T1D, the study population must include a high enough frequency of the allele or haplotype to reveal the susceptibility effect. For example, unlike most DRB1*04 alleles, the allele DRB1*04:03 is actually protective for T1D (Tables 3 and 4) [64], but that effect was difficult to detect in studies of European populations until the large T1DGC data set was published [34,65,66]. DRB1*04:03 is seen at far higher frequency in individuals of Asian descent (~3.5%) than for individuals of European descent (~0.6%) [67]. The protective effect of DRB1*04:03 on T1D is an excellent example of why any T1D genetic screening program should resolve genotypes to at least two fields.

African Americans are a recently-admixed population (<20 generations) that has been informative in the study of HLA-associated T1D risk [36]. Haplotypes categorized as DR3 and DR7 illustrate why the study of this population is particularly informative. DR7, a designation for haplotypes including the allele DRB1*07:01, is commonly seen as DRB1*07:01-DQA1*02:01-DQB1*02:02 in Europeans and is known to be protective for T1D [34]. However, a closely-related version of DR7, DRB1*07:01-DQA1*03:01-DQB1*02:02, found in African populations, is not protective for T1D but is, in fact, predisposing for T1D [36,68,69]. Similarly, an African-Specific DR3, DRB1*03:02-DQA1*04:01-DQB1*04:02, has the opposite T1D susceptibility effect (protection) to that of the common, predisposing DR3 haplotype, DRB1*03:01-DQA1*05:01-DQB1*02:01, seen in most other populations [36,70]. The opposite effects of European and African DR3 and DR7 haplotypes are illustrated in Table 5.

Genetic screening for T1D risk has been a goal of T1D researchers for decades. Many large studies (e.g., DAISY [71], TEDDY [72], TrialNet [73], TRIGR [74], and others) use a limited genetic screen to select subjects for follow up autoantibody testing and intervention studies. World populations are becoming increasingly admixed, especially in the United States. The genetic heritage of an individual is not necessarily apparent from phenotype, nor can it be determined, in some cases, from interview. The potential danger of a limited, low-resolution genetic screen is illustrated by Table 6, in which a possible DR3/4 genotype is shown for a hypothetical individual of mixed Asian and African ancestry. If this individual were genotyped with a limited, low-resolution genetic screening test for DR, (s)he would be classified as very high diabetes risk, based on the finding of a DR3/4 genotype. However, both of the haplotypes that comprise this particular DR3/4 genotype are protective for T1D. Genotyping of both DR- and DQ-encoding genes at two-field resolution would clearly show that this individual is highly unlikely to get T1D.

### Table 5
Comparison of T1D susceptibility effects for DR3 and DR7 haplotypes in two populations.

<table>
<thead>
<tr>
<th>Population</th>
<th>A. DR3</th>
<th>B. DR7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DRB1</td>
<td>DQA1</td>
</tr>
<tr>
<td>European</td>
<td>03:01</td>
<td>05:01</td>
</tr>
<tr>
<td>African American</td>
<td>03:02</td>
<td>04:01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>07:01</td>
<td>02:01</td>
</tr>
<tr>
<td>African American</td>
<td>07:01</td>
<td>03:01</td>
</tr>
</tbody>
</table>

*red font denotes T1D susceptible effect, and green font denotes T1D protective effect.
The DR3/4 heterozygous genotype only accounts for up to approximately 40% of T1D patients; therefore, even for populations of European descent, simply searching for high-risk DR3/4 individuals will exclude the remaining 60% of future patients that do not have the highest-risk genotype. Understanding the risk conferred by individual alleles in haplotypic and genotypic context will allow risk assessment for all individuals, regardless of racial or ethnic background.

The need for the study of non-European populations to fully understand HLA susceptibility to T1D has been noted for more than 20 years [75]. The number of published studies on non-European populations and T1D is increasing; however, the need for study of underserved and understudied populations remains a critical component of a full understanding of HLA-associated T1D risk.

4. Association of other immune genes with T1D

More than 40 genetic loci have been implicated in T1D risk [76,77]. Many of these genes are relevant to immune function, and they lie both within and outside the HLA region. The associations of these genes are of far lower magnitude than those of HLA genes; however, their contribution to TID autoimmunity is reproducible. Most of these genes are associated with other autoimmune disorders as well. HLA context must be considered in association analyses for these genes, especially for those located in the HLA region, where LD with T1D associated DR-DQ haplotypes can be mistaken for true disease association.

4.1. Genes in the HLA region

The arrangement of HLA class I and class II genes on chromosome 6 is represented on Fig. 8. Class I genes are telomeric of class II genes, and between them lies a region commonly referred to as class III, although it contains no genes encoding classical HLA. The class III region is very gene dense and encodes several genes of immunological relevance including TNFA, MIC-A, and genes encoding complement proteins. T1D associations for these genes have been reported; however, their location in the HLA region precludes simple association analysis. As for HLA-DPB1 or class I genes, LD with DR- and DQ-encoding genes must be accounted for when analyzing data. Polymorphisms in the promoter of the TNFA gene at positions −238 (rs361525) and −308 (rs1800629) have been heavily studied for T1D association with conflicting results [78–83]. Most observed genetic association for TNFA with T1D is likely attributable to LD effects of TNFA promoter alleles with DR3 haplotype [81,83]. MHC class I chain-related gene A (MIC-A), a stress-induced antigen in gut epithelium, has also been studied for T1D association. Several reports implicate MIC-A5 as a T1D susceptibility allele and MIC-A5.1 as a T1D protective allele [84–87]. Complement C4 genes (C4A and C4B) exhibit copy number variation that is associated with autoimmunity [88]. Both the C4A null allele and the C4B “short” allele have reported association with T1D [89–93].

4.2. Genes outside the HLA region

Cytotoxic T-lymphocyte-associated protein 4 (CTLA4) (gene located at 2q33) is a down-regulator of the immune response. T1D association for CTLA4 in multiple ethnic groups was noted nearly two decades ago [94]. A recent meta-analysis of 58 studies confirmed CTLA4 association with T1D but noted that the associations vary among ethnic populations [95]. The immune regulator PTPN22 (gene located at 1p13) was shown to be associated with T1D in 2004 [96,97]. This association was replicated in many T1D studies as well as in studies of other autoimmune diseases [98].

4.3. KIR association with T1D

An emerging literature is focused on finding association of Killer-cell Immunoglobulin-like receptor (KIR) genes with T1D. KIRs are a family of cell-surface receptors that are found on natural killer (NK) cells and regulate their function [99–101]. HLA class I molecules are ligands for KIR, and HLA class I is associated with T1D; thus, KIR genes might be expected to exhibit T1D association. Most KIR T1D association studies to date are small case-control studies that focus on individual KIR genes and test only presence or absence of KIR loci, and results are largely inconsistent [102–112]. A large, recent study reported a method for imputation of KIR copy number for one KIR locus (3DL1/3DS1) using Genome-Wide Association Study (GWAS) SNP data but saw no significant type 1 diabetes association for that locus [113].

The genes that encode KIR are highly polymorphic [13,114]. In contrast to HLA, where diversity is almost entirely due to extensive allelic polymorphism in the genes, much of the known KIR genetic diversity comes from variations in gene content, i.e., the presence or absence of a gene, in addition to allelic polymorphism within each gene [115]. The KIR region lies on chromosome 19q13 in the Lymphocyte Receptor Complex (LRC) and is comprised of 14 genes;
reported numbers of KIR loci vary due to uncertainty about whether some pairs of named loci are separate genes or multiple alleles of a single pair. With the exception of 4 “framework” genes, other KIR genes may or may not be present on a chromosome. Haplotypes of KIR genes fall into two categories: A, containing mostly inhibitory genes; and B, containing one or more activating KIR genes, schematically represented in Fig. 9. KIR haplotypes include centromeric and telomeric gene groups, exhibiting strong linkage disequilibrium within, but not between, groups [116]. KIR gene content varies among populations [117]. The extent of KIR allelic diversity is just beginning to be understood [13].

Not all KIR ligands have been elucidated; however, HLA class I molecules are among known ligands for KIR [100]. Based on the identity of the amino acid residue at position 80, HLA-C alleles can be categorized into 2 groups, C1 (80N) and C2 (80K). Similarly, HLA-B alleles can be classified as Bw4 (80I or 80T) or Bw6 (80N). Effects of individual KIR loci on disease are, of course, not only dependent on whether or not that particular KIR locus is present in an individual but also on whether or not an allele encoding the HLA ligand for that KIR is present. For example, KIR2DL2 binds HLA-C alleles in the C1 group. Some studies have reported increased T1D risk for KIR2DL2 [103,105,108]. However, in individuals lacking a C1 allele (i.e., both alleles have HLA-C alleles in the C2 group rather than the C1 group), any risk from KIR2DL2 is necessarily absent. The need to stratify based on disease, KIR, and HLA genes creates a requirement for large population sizes to address the association of KIR with T1D. As KIR genotyping methodology evolves, more and larger cohorts will be genotyped, eventually allowing sufficient sample size for robust T1D association analysis.

5. Genetic prediction of T1D

The overarching goal for the study of type 1 diabetes is to find a cure, or better yet, to prevent the disease altogether. However, prevention requires prediction. Effective identification and use of disease-specific prevention or intervention strategies before overt disease presentation depends on the ability to identify future patients, unless the intervention is benign and inexpensive enough for use in the entire population. For T1D, autoimmunity starts months to years before overt disease is observed; identification of future patients is not straightforward. To develop therapies for T1D prevention, demonstration of a significant effect on T1D development in a well-powered study period is necessary to show efficacy of the novel therapeutic being tested [118]. Given the relatively low incidence of T1D, trials aimed at preventing disease require 1) screening a very large population to find at-risk individuals and 2) having a sufficient number of “progressors” in the control arm of the trial who will get disease in a reasonable time frame to enable sufficient power to show the benefit of a tested intervention. To reduce the cost, scope, and length of such clinical studies, optimized initial screening is essential. Genetic testing represents a cost-effective choice for an initial population screen due to the ease and convenience of procuring DNA samples and the need for only a single test. HLA genes are the obvious place to start for T1D.

T1D prediction is becoming feasible for high-risk Europeans, with reports demonstrating good clinical predictive value for the highest-risk HLA alleles, in combination with selected SNPs [119–121]. In fact, only 4 SNPs, 2 tagging HLA (DR3 and DR4), and one each in the insulin and CTLA4 genes, can give a clinically relevant predictive value (receiver operating characteristic AUC = 0.72) to predict the highest risk T1D patients for potential intervention [119]. A model incorporating up to 40 SNPs was able to achieve even greater predictive value (AUC = 0.87) [120]. However, the highest-risk HLA genotype in Europeans, HLA-DR3, only accounts for a maximum of about 40% of patients. Predicting risk from other HLA alleles and haplotypes is essential to create predictive models that include all. We cannot predict genetic T1D risk for non-European populations without performing HLA disease association studies in those populations, again emphasizing the value of genetic studies in non-European populations.

6. Conclusions

T1D is a devastating disease that, in the best case, with state of the art care, is a chronic condition requiring intense management and, in the worst case, with limited resources, leads to early death. Despite decades of study and much recent progress, an effective intervention or cure remains elusive. Modulation of the immune response is necessary to slow or stop beta cell destruction and represents a key component of any prevention or intervention strategy. A thorough understanding of the immunogenetics of T1D will allow the informed selection of subjects, at early stages of disease with most or all beta cell function preserved, for enrollment in clinical trials. Thorough immunogenetic characterization will also enhance the ability to predict response to a given therapeutic, thereby allowing selection of appropriate interventions for at-risk individuals of any racial or ethnic background. Immunogenetic characterization represents a key factor on the path to T1D prevention for all.

Acknowledgments

The author thanks Brenda Stallworth for assistance with figures and tables and Julie Lane for critical reading of the manuscript.

References


