



ELSEVIER

# The genetics of HLA-associated disease

Charles E Larsen<sup>1,2</sup> and Chester A Alper<sup>1,3\*</sup>

Type 1 diabetes mellitus (T1D) remains the most intensively studied, and thus the best paradigm, of MHC-associated diseases. Accumulating evidence suggests that MHC susceptibility for T1D is recessive, with susceptibility alleles more common than protective alleles. Updated allele-level and nucleotide sequence analysis of MHC class II T1D susceptibility markers of conserved extended haplotypes underscore the uncertainty surrounding the actual T1D MHC susceptibility locus. Recent studies have established that disease concordance in dizygotic twins is the same as that in siblings generally, for both T1D and the MHC-associated autoimmune disease gluten-sensitive enteropathy, leaving little room for a differential environmental trigger. Epigenetic mechanisms are probably involved in many MHC-associated phenomena, including autoimmunity, and appear to be the best explanation for incomplete penetrance.

## Addresses

<sup>1</sup>The CBR Institute for Biomedical Research, 800 Huntington Avenue, Boston, MA 02115, USA

<sup>2</sup>Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

<sup>3</sup>Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA

\*e-mail: alper@cbr.med.harvard.edu

**Current Opinion in Immunology** 2004, **16**:660–667

This review comes from a themed issue on Immunogenetics

Edited by Chester A Alper

Available online 19th August 2004

0952-7915/\$ – see front matter

© 2004 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.coi.2004.07.014

## Abbreviations

<b>CEH</b>	conserved extended haplotype
<b>GSE</b>	gluten-sensitive enteropathy
<b>IgAd</b>	IgA deficiency
<b>Igd</b>	immunoglobulin deficiency
<b>IgDd</b>	IgD deficiency
<b>LD</b>	linkage disequilibrium
<b>MHC</b>	major histocompatibility complex
<b>MZT</b>	monozygotic twins
<b>T1D</b>	type 1 diabetes mellitus
<b>TDT</b>	transmission disequilibrium test

## Introduction

Type 1 diabetes mellitus (T1D) remains the most intensively studied and thus the best paradigm of major histocompatibility complex (MHC)-associated diseases. T1D is still, however, largely a ‘geneticist’s nightmare’ [1]. Here, we use T1D as a model for exploring the complex

genetics of several HLA-associated autoimmune diseases and we investigate several of their genetic mysteries: the evidence for the mode of their inheritance, the uncertainty of the actual T1D MHC susceptibility locus, the presence and aggregate frequency of non-MHC T1D susceptibility genes, and the stochastic nature of penetrance of MHC susceptibility genes.

We make the explicit assumption that, although complex, each of the HLA-associated diseases has its own specific genetic determinants and mechanism, and that these do not vary in unrelated individuals.

## Evidence for MHC and non-MHC susceptibility genes for T1D

That T1D is genetically determined is most convincingly supported by the fact that about 40 to 50% of monozygotic twins (MZT) of T1D patients also have the disease [2] compared with a rate of approximately 0.4% in the general US population. The lack of complete concordance for T1D in genetically identical MZT implies that only 40 to 50% of all genetically susceptible individuals (as MZT or totally unrelated persons in the general population) will manifest T1D. The higher rate of concordance in MHC-identical sibs of patients (12 to 16%) compared with sibs in general (5 to 6%) [3] is clear evidence for an MHC T1D susceptibility gene. Nevertheless, the much higher concordance in MZT of patients than in MHC-identical sibs indicates that genes in addition to those in the MHC are involved in genetic susceptibility.

## HLA allele and haplotype associations with T1D

About a quarter of a century ago, it was observed that the frequencies of certain (susceptibility) HLA antigens were increased and those of other (protective) antigens were decreased in patients with T1D as compared with unrelated ethnically matched healthy control subjects [4]. Still other (neutral) antigens had similar frequencies in patients and controls or were partially susceptibility conferring or protective. In initial results in Caucasians, HLA-B8, HLA-B18 and HLA-B15(B62) were found to be increased in frequency among patients [4,5]. Later, as genetic testing methods expanded and improved, HLA-DR3 (*HLA-DRB1\*0301*) and HLA-DR4 (*HLA-DRB1\*04*) and *HLA-DQB1\*0201* and *HLA-DQB1\*0302* were found to be susceptibility markers [6<sup>\*\*</sup>,7<sup>\*\*</sup>].

Reports of HLA allele associations with T1D all too often fail to consider the genetic fixity of the MHC regions surrounding the *HLA-DRB1* and *-DQB1* genes. Relatively large stretches (1.5 Mb or more) of shared conserved

**Table 1****Type 1 diabetes susceptibility, neutral and protective conserved extended haplotypes (CEHs).**

CEH type	CEHs [ <i>HLA-B</i> , <i>complotype</i> , <i>DR</i> ]	Ratio (T1D:normal)
Susceptibility	B62, SC31, DR4	17.0
	B18, F1C30, DR3	14.6
	B62, SB42, DR4	5.8
	B62, SC33, DR4	4.4
	B8, SC01, DR3	2.1
Neutral	B60, SC02, DR6	1.0
	B35, FC(3,2)0, DR1	1.0
	B44, SC30, DR4	0.77
	B35, SC31, DR5	0.45
Protective	B7, SC31, DR2	0.22
	B44, FC31, DR7	0.21
	B57, SC61, DR7	0.14

CEHs are grouped as susceptibility, neutral or protective haplotypes for T1D based on the ratio of their occurrence (in frequency) among 344 T1D haplotypes to their frequency among 2000 normal family control haplotypes (haplotypes found only in non-diseased subjects in families having an index case with an MHC-associated disease [12]). The ratios for CEH types are defined by their T1D:normal ratio as: susceptibility >1.5; neutral = 0.4–1.5; protective <0.4.

DNA sequence, termed ‘conserved extended haplotypes’ (CEHs) [8] or ‘ancestral haplotypes’ [9], characterize the MHC. CEHs comprise at least 30% of MHC haplotypes in normal Caucasians [10\*\*]. Operationally, MHC CEHs are identified by identical markers at *HLA-B*, *complotype* and *HLA-DRB1*. Other markers within this conserved region are also fixed [10\*\*,11]. All of the widely accepted susceptibility and protective MHC markers for T1D are parts of CEHs [12] or ancestral haplotypes [13] or their fragments. Recognized T1D susceptibility CEHs in European Caucasians, several ‘neutral’ CEHs and the major protective CEHs are shown in Table 1.

It was recognized early that the distribution of HLA-DR3 (B8 or B18) and DR4 [B15(62)] homozygotes and double

heterozygotes did not fit the Hardy-Weinberg equilibrium [14]. Rather, there was an excess of heterozygotes over either homozygote, as predicted from their allele frequencies among many, but not all [15], Caucasian patient populations. Although the observed excess of HLA-DR3/DR4 heterozygotes is puzzling, it is unlikely to reflect directly the genetic or immunologic mechanism of MHC genes or gene products in T1D, because it is true for only some populations.

### The MHC locus for T1D susceptibility is unknown

A widely, although not universally, held belief is that *HLA-DQB1* (or *HLA-DRB1*, *-DQA1*, *-DQB1*) is the MHC susceptibility locus (or loci) for T1D. However, several lines of evidence show that the MHC susceptibility locus for T1D is still unknown. For example, the odds ratios for (*HLA-DRB1\*0301*, *DQB1\*0201*) and (*HLA-DRB1\*04*, *DQB1\*0302*) susceptibility markers are actually lower than some of the *complotype* susceptibility markers (Table 2). This probably has more to do with the relative polymorphism at specific loci and with the frequencies of specific alleles in background subpopulations than with localization because all of these markers are on CEHs.

Furthermore, if inheritance of T1D MHC susceptibility genes is recessive, why are some protective alleles, such as *HLA-DRB1\*1501* and/or *-DQB1\*0602* [16–18], occasionally found in T1D patients? One possibility is that an ancient crossover produced a *HLA-DRB1\*1501*-susceptibility *DQB1* gene haplotype [16]. Another is that there was a mutation in the usually protective *DQB1\*0602* [17]. However, the latter possibility was ruled out for several T1D patients by complete sequence analysis [18]. Yet another problem is the fact that some *HLA-DQB1* alleles, for example *DQB1\*0303*, that are protective in one population [19,20] are susceptibility conferring in others [21,22].

The simplest explanation of these findings is that *DRB1/DQB1* alleles are markers for T1D susceptibility but these

**Table 2****MHC marker odds ratios (ORs) for T1D.**

HLA-A	OR	HLA-B	OR	Complotype	OR	HLA-DR	OR	<i>HLA-DRB1</i> , <i>-DQB1</i>	OR
A1	0.99	B7	0.52	SC31	0.53	DR1	0.90	<i>DRB1*0301</i> , <i>DQB1*0201</i>	<b>3.55</b>
A2	1.36	B8	<b>2.09</b>	SC01	<b>2.03</b>	DR2	<u>0.16</u>	<i>DRB1*04</i> , <i>DQB1*0301</i>	0.62
A3	1.05	B18	<b>2.22</b>	F1C30	<b>8.49</b>	DR3	<b>3.55</b>	<i>DRB1*04</i> , <i>DQB1*0302</i>	<b>3.95</b>
A11	0.63	B50	1.00	SC33	<b>3.10</b>	DR4	<b>2.91</b>	<i>DRB1*0101</i> , <i>DQB1*0501</i>	0.90
A24	1.58	B57	<u>0.15</u>	SB42	<b>3.20</b>	DR5	<u>0.24</u>	<i>DQB1*0301</i>	0.36
A26	1.53	B60	<b>2.06</b>	SC2(1,2)	0.63	DR6	<u>0.46</u>	<i>DQB1*0302</i>	<b>3.95</b>
A28	<u>0.28</u>	B62	<b>2.01</b>	FC(3,2)0	0.82	DR7	<u>0.28</u>	<i>DQB1*0602</i>	<u>0.10</u>
A30	2.17	B65	<u>0.68</u>	S1C2(1,17)	<b>2.97</b>	DR8	0.76	<i>DQB1*0303</i>	0.23
A33	<u>0.43</u>			SC61	<u>0.08</u>	DR9	0.72		

HLA-A, -B, and -DR are serotypes. *Complotypes* are defined by BF, C2, C4A, and C4B variants (in that arbitrary order) defined by electrophoresis and immunofixation. ORs are determined from haplotypes in T1D patients and family controls [12]. ORs < 0.5 are underlined and those > 1.5 are in bold. The highest OR (that for the *complotype* F1C30) is both underlined and in bold. The same database as that described in Table 1 was used in the compilation of this table.

loci are not themselves the susceptibility genes. In our view, therefore, molecular studies of specific *HLA-DQB1* allele interactions with T1D-associated peptides cannot provide direct insight into the disease mechanism. Some investigators have interpreted the results described above differently to suggest that there is no question that *HLA-DRB1*, *-DQA1* and *-DQB1* genes are T1D susceptibility loci but that the involvement of HLA-DQ in T1D is 'complex' [23].

We do not have space here to explore the vast literature of other proposed T1D susceptibility loci in or near the human MHC or their putative complex interaction with *HLA-DQB1*, but the literature of one locus that has received significant attention recently, the microsatellite *D6S2223*, is illustrative of the difficulty in defining such associations. Allele 3 of the microsatellite *D6S2223* appeared to be in linkage disequilibrium (LD) with an allele at a locus that affects the T1D susceptibility status of (*HLA-DRB1\*03*, *DQA1\*0501*, *DQB1\*0201*) haplotypes using a transmission disequilibrium test (TDT) on the T1D and healthy sibling offspring of parents homozygous for that haplotype from Denmark, Norway and the UK [24] as well as from Sweden and France [25]. However, this result was not found using the homozygous TDT using parents homozygous for either (*HLA-DR4*, *DQ8*) haplotypes in any population or (*HLA-B18*, *DRB1\*03*, *DQB1\*02*) haplotypes in French families [25], nor was this result found using LD analysis in Sardinian families [26]. The dependence of the results on the ethnic origin, the specific haplotype under study and/or the particular analytical technique used is an indicator of possible population stratification effects. In any case, no T1D susceptibility allele was found at the *D6S2223* marker except on (*HLA-B18*, *DRB1\*03*, *DQB1\*02*) haplotypes [25]. Finally, because *D6S2223* alleles are not generally in LD with those of the classic HLA loci, it is difficult to consider *D6S2223* anything more than one of the many candidate non-MHC *IDDM* loci, even if the observed phenomena are not due to population stratification effects.

Another, and related, mystery with respect to the genetics of MHC susceptibility to T1D is why some *HLA-DRB1*, *-DQB1* (so-called 'neutral') haplotypes do not clearly carry susceptibility or protective genes. For example, haplotypes containing the fragment (*HLA-DRB1\*0101*, *DQA1\*0101*, *DQB1\*0501*) have similar frequencies between patient and family control haplotypes (Table 1; [23]). Our view is that such haplotypes must not be fixed between the *HLA-DRB1/DQB1* region and the actual susceptibility locus.

### What is the mode of inheritance of the MHC susceptibility gene for T1D?

To begin to understand the complex genetics of T1D, we must determine the basic mode of inheritance of the MHC susceptibility gene based on available evidence

from MHC allele association and family studies. Before considering more complicated or mixed modes, such as epistasis, overdominance or part recessive, part dominant (all of which have been proposed at one time or another), we must first rule in or out simple recessive or dominant inheritance, if at all possible. Most of the impetus for considering complicated modes of inheritance has been the attempt to explain the excess of HLA-DR3/DR4 heterozygotes among T1D patients in some Caucasian populations [14] (or DR3/DR9 in some Chinese populations [27]).

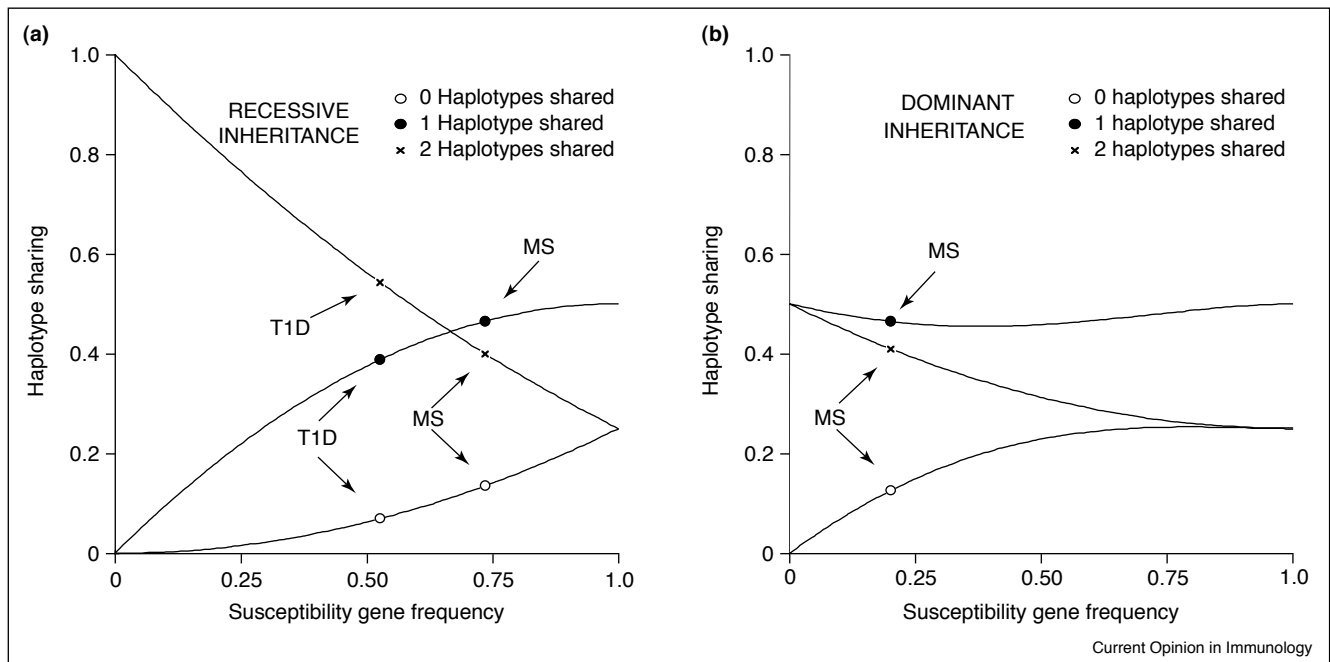
Far more compelling (with respect to the mode of inheritance) are those populations in which homozygotes and heterozygotes for HLA-DR3 and DR4 are in accord with the Hardy-Weinberg equilibrium [15]. Such populations provide strong evidence for simple recessive inheritance of the MHC T1D susceptibility gene. Furthermore, using the *BF\*F1* marker [28], one obtains homozygote and heterozygote frequencies consistent with recessive but not dominant inheritance. The dominant nature of protective MHC alleles is also strong evidence for simple recessive inheritance. By analogy, the *HbA* allele dominantly protects against sickle cell anemia. As for the excess of DR3/DR4 heterozygotes among some populations, other explanations may include population inhomogeneities and admixture (ZL Awdeh *et al.*, unpublished), superimposed on recessive inheritance.

The analysis of haplotype sharing in pairs of affected sibs is a potent probe of genetic disease [29] that also points to a simple recessive model of inheritance for T1D. For a monogenic recessive disease, such as sickle cell anemia, in families with two carrier (heterozygous or *HbA/HbS*) parents, affected sibs will constitute 25% of offspring. All affected pairs (100%) will have the same two haplotypes carrying *HbS* and any genetic marker very closely linked (i.e. no genetic recombination at meiosis) to the  $\beta$  globin susceptibility locus. Penetrance of susceptibility genes is irrelevant to the extent of haplotype sharing since penetrance in the affected sibs is 100%. Of sib pairs affected with T1D, 55% shared both, 38% shared one and 7% shared no MHC haplotype [30]. Although this was offered as evidence for recessive inheritance [29], the deviation from 100% complete MHC haplotype sharing was yet another mystery.

### The MHC susceptibility markers exist at an extraordinarily high normal population frequency

The key point is that the 7% of affected sibs who share no MHC haplotype [30] must be the offspring of parents who are themselves each homozygous for T1D MHC susceptibility markers [31]. From observed MHC haplotype sharing by affected sibs and a simple genetic model [31], dominant inheritance of MHC susceptibility is ruled out; recessive inheritance is consistent with the observed

Figure 1



Predicted frequencies of 2, 1 and 0 haplotypes shared by affected sibs for (a) recessive and (b) dominant inheritance of an MHC disease susceptibility gene at various frequencies of the disease allele. Observed haplotype distribution frequencies for multiple sclerosis (MS) and T1D are marked. Note that for MS, the distribution fits either form of inheritance, whereas there is no dominant solution for T1D. (Modified with permission from S Karger AG, Basel from [31]).

haplotype sharing (Figure 1). If the MHC susceptibility gene frequency is high, it should come as no surprise that many patients are offspring of non-affected and often non-susceptible parents who are nevertheless homozygous for MHC susceptibility haplotypes. What made this into a mystery is our bias to think in terms of rare monogenic recessive disease where essentially all parents of patients are heterozygotes for the fully penetrant susceptibility gene.

If the combined frequency of all MHC susceptibility genes ( $p$ ) is high, the frequency of MHC susceptible homozygotes ( $p^2$ ) will also be high. In matings where both parents are homozygous for MHC susceptibility genes ( $p^4$ ; i.e. 7% of all mating pairs), a quarter of all offspring (who are all MHC susceptible) share no MHC haplotypes with the index patient (sib). We demonstrated [31] the simple relationship:  $0.25p^2 =$  the frequency of non-sharing MHC susceptible sibs = 0.07.

This means that the sum of the frequencies of all MHC susceptibility alleles is 0.53 (the square root of  $4 \times 0.07$ ). Thus, MHC susceptibility alleles are more common (53%) in many Caucasian populations than MHC protective alleles. Fully 28% ( $p^2$ ; over a quarter) of the overall Caucasian population is MHC-susceptible to T1D (i.e. carries two T1D susceptibility MHC haplotypes)!

### What is the frequency of non-MHC gene susceptibles for T1D?

Although we cannot, without more information, know the number or the mode of expression of non-MHC susceptibility genes, we can predict the aggregate frequency of non-MHC gene susceptible persons in the population [31,32]. If the prevalence of T1D is 0.003 to 0.004 and penetrance is 50%, then the prevalence of susceptible individuals is 0.006 to 0.008 (about 0.7% of the population). That frequency, 0.007, represents the product of MHC susceptible individuals (0.276) multiplied by the decimal fraction of non-MHC susceptibles (assuming multiplicative interaction of susceptibility genes [31,32]). Thus, susceptibles by virtue of non-MHC genes are 0.02 to 0.03 of the general population.

The identity of non-MHC T1D susceptibility genes is unclear with no agreement (other than HLA) in large-scale studies in the UK [33] and the USA [34]. Although more than 20 non-MHC loci have been proposed, there is some evidence from both linkage and association studies for only two markers: *INS* (the insulin gene) [35] and *CTLA4* [36]. The subject has recently been reviewed [37]. The problem is to distinguish susceptibility genes that can also be population markers from population markers only [38\*].

### Incomplete penetrance and non-genetic factors associated with T1D

Since MZT are thought to be genetically identical, if one twin has T1D, why doesn't the other? Without elucidating the mechanism, the phenomenon has been termed 'incomplete penetrance of susceptibility genes'. We make a distinction between the 'baseline' penetrance rate of MZT and the lower 'apparent penetrance' rates of MHC-identical sibs, sibs in general, parents or children of patients or in the general population [39], all of which have various frequencies of non-MHC susceptibility genes.

#### What is the basis for incomplete penetrance?

Most attempts to explain incomplete penetrance have invoked an environmental trigger, with a common viral infection being a favorite candidate [40]. This explanation is made attractive by the fact that T1D (as well as other MHC-determined autoimmune diseases) has its onset well after birth and sometimes in adult life. Probably the best argument for an environmental trigger for T1D is the marked difference in disease incidence in the NOD mouse depending upon diet or almost any manipulation [41]. More relevant to human T1D is the fact that dizygotic twins have the same concordance rate [42\*\*] as sibs overall [3,43,44], which strongly suggests that differential environmental effects are minimal, if they exist at all.

In general, there are two kinds of mechanisms that might give rise to the appearance of T1D in fully genetically susceptible persons [39]. One is an extrinsic trigger, such as infection, exposure to chemicals in the environment or the like. The other is an intrinsic (epigenetic) mechanism affecting gene expression. There are many examples of extrinsic precipitating factors, such as cigarette smoking and cardiovascular disease and carcinoma of the lung. Among autoimmune diseases, celiac disease (or gluten-sensitive enteropathy [GSE]) and dermatitis herpetiformis only manifest themselves when gluten (gliadin) is ingested. The relevance of these diseases to the usually irreversible T1D is, however, unclear.

The well-documented seasonal variation in onset of T1D has been considered evidence for a viral trigger. However, strong evidence that pancreatic  $\beta$  cell destruction may begin many years before the onset of overt diabetes [45] casts considerable doubt on this hypothesis. Examples of intrinsic triggers of variable expression of genes include parental imprinting, allelic exclusion in expression of B and T cell receptors, and monoallelic expression of cytokine genes in lymphocytes, and receptors on olfactory cells and natural killer cells. The difference between the extrinsic and intrinsic mechanisms of penetrance is that the first acts on the whole organism, whereas the second would be expected to act independently on the two alleles at a relevant susceptibility locus [39].

### Other MHC-associated diseases as a model for incomplete penetrance in T1D

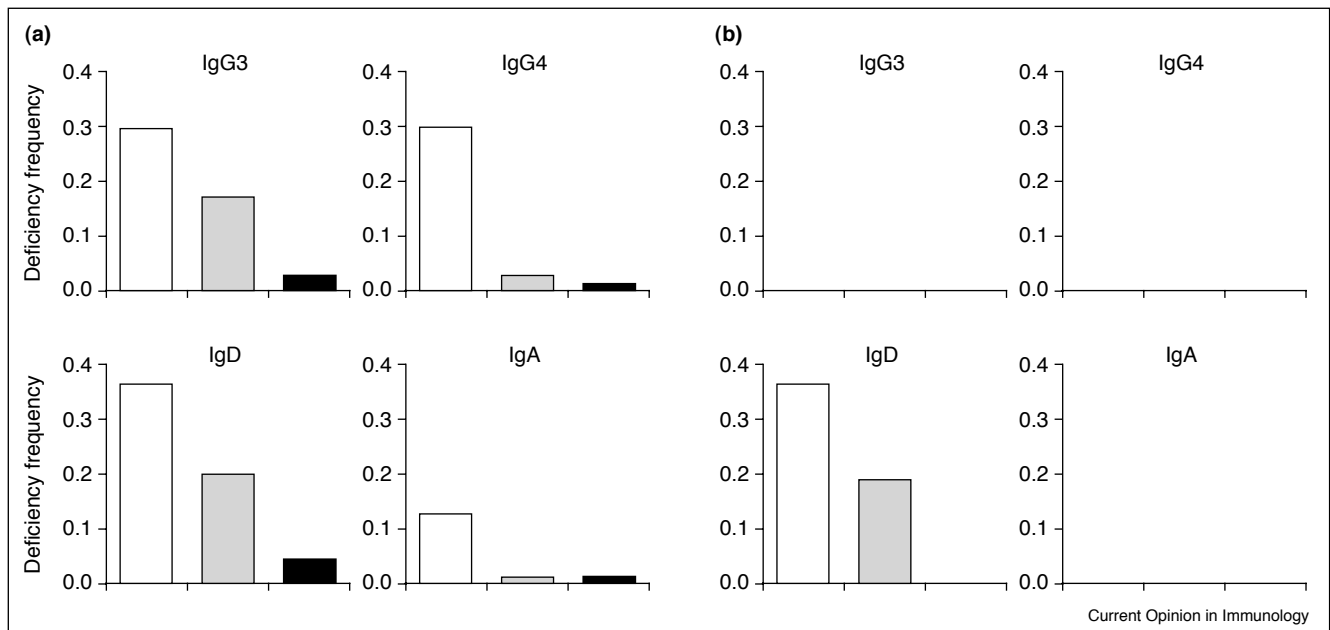
MZT discordant for IgA deficiency (IgAd) have been reported [46], and we recently observed discordance in other MHC-determined immunoglobulin deficiencies (IgD) (CA Alper *et al.*, unpublished). A concordance rate of 70–86% has been found in MZT (30% in MHC-identical siblings) of patients with GSE [47], another MHC-associated genetic disease [48\*]. Thus, incomplete penetrance, known for many years in T1D and other MHC allele-associated autoimmune disorders, could be a general intrinsic property of some MHC genes.

We recently examined the question of the mechanism of penetrance in MHC-susceptibility gene-associated immunoglobulin deficiencies [39,49]. Homozygotes, heterozygotes and non-carriers of the MHC CEH [*HLA-B8, SC01, DR3*] were studied for serum immunoglobulin levels. This haplotype is known to be elevated in frequency among patients with IgAd [50] and IgD deficiency (IgDd) [51]. There was a significantly increased frequency (4 out of 30) of IgAd among the haplotype homozygotes but not heterozygotes or non-carriers, consistent with IgAd being an MHC recessive trait (Figure 2a; [49]). Apparent penetrance (that is due to baseline penetrance and the presence or absence of all susceptibility genes) was thus 13% in the haplotype homozygotes compared with 0.14% in the general population. Owing to recessive inheritance of MHC susceptibility to IgAd, the data did not allow discrimination between intrinsic and extrinsic triggers [39].

By contrast, IgDd was found in both homozygotes (37%) and heterozygotes (19%) for the same haplotype (Figure 2a; [49]). Similar frequencies for IgDd (but no other IgD) were observed in studies of the CEH [*HLA-B18, F1C30, DR3*] among the Basques (Figure 2b; [52]). No IgAd but higher levels of IgDd were also found in both homozygotes (60%) and heterozygotes (35%) for the C2 deficiency CEH [*HLA-B18, S042, DR2*] [53\*]. These findings permit some conclusions about the nature of penetrance of MHC susceptibility genes for immunoglobulin deficiencies and, perhaps, by extension, to T1D. Since [*HLA-B8, SC01, DR3*] carries susceptibility for IgAd as well as IgDd, but [*HLA-B18, F1C30, DR3*] and [*HLA-B18, S042, DR2*] carry only IgDd, there must be distinct MHC susceptibility genes for these two IgD.

The occurrence of IgDd in both homozygotes and heterozygotes for susceptibility haplotypes suggests that the MHC susceptibility genes for IgDd are dominantly expressed. If penetrance is extrinsic, the rate of IgDd should be the same in heterozygotes and homozygotes (i.e. whether one has one or two susceptibility genes does not affect disease incidence). The different observed rates of apparent penetrance in homozygotes and heterozygotes suggest that the mechanism of penetrance is intrinsic,

Figure 2



Homozygotes, heterozygotes and non-carriers of either the MHC CEH [*HLA-B8, SC01, DR3*] (a) or haplotypes containing at least the (*F1C30, HLA-DR3*) fragment of the CEH [*HLA-B8, SC01, DR3*] (b) were studied for serum immunoglobulin levels. The prevalence of immunoglobulin deficiency in haplotype homozygotes (open bars), heterozygotes (grey bars) and non-carriers (black bars) is shown. The height of each bar represents the relative frequency of individuals deficient in that immunoglobulin class or subclass as a fraction of the number of subjects in that category tested. (a) modified with permission from Blackwell Publishing, London from [49] and (b) with permission from Springer, New York from [52].

being nearly twice in homozygotes with two susceptibility genes what it is in heterozygotes with only one susceptibility gene. Moreover, since the only systematic difference between MHC haplotype homozygotes and heterozygotes is the number of MHC susceptibility genes, baseline penetrance is a function of the MHC susceptibility gene for IgDd [39,49,52,53]. We do not know if it is a property of other susceptibility genes or not.

## Conclusions

Many mysteries remain with respect to the genetics of T1D and other diseases controlled by MHC susceptibility genes in concert with other non-MHC genes. We propose here a new paradigm for T1D MHC genetics, on the basis of our analysis of both the historical and recent literature on this disease. There is good evidence that the T1D MHC susceptibility gene is recessive, but that other, non-MHC susceptibility genes are required for T1D to occur. Given the incomplete penetrance of T1D susceptibility genes and the polygenic nature of this disease, both MHC and non-MHC susceptibility genes must have extraordinarily high general population frequencies. No susceptibility gene for T1D (MHC or non-MHC) has yet been identified with certainty nor has the number or mode(s) of inheritance of the non-MHC T1D susceptibility gene(s) been determined. Finally, there is suggestive evidence that penetrance is an intrinsic property of MHC susceptibility genes.

Clearly, the identification of all the genes involved in MHC-associated autoimmune diseases and the elucidation of their pathogenetic roles has the highest priority for future work. Results of sequence determination of the MHC of individuals who carry two identical versions of different (largely CEH) MHC extended haplotypes [54] should prove extremely useful in this quest. Moreover, we believe that future work will also require new approaches and probably new techniques. Once the MHC and non-MHC genes for autoimmune diseases and related phenomena are known, exploration of the basis for incomplete penetrance should prove relatively simple.

## Acknowledgements

This work was supported by grant no. HL-29583 from the National Heart, Lung, and Blood Institute of the National Institutes of Health. We are indebted to many colleagues for stimulating discussions over many years and for past collaborative studies, in particular Zuheir Awdeh, Stuart Brink, George Eisenbarth, Kenneth Gabbay, Donald Raum and Edmond Yunis.

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Neel JV: **Diabetes mellitus – a geneticist's nightmare**. In *The Genetics of Diabetes Mellitus*. Edited by Creutzfeldt W, Kibberling J, Neel JV. Berlin: Springer Verlag; 1977:1-11.

2. Barnett AH, Eff C, Leslie RDG, Pike DA: **Diabetes in identical twins: a study of 200 pairs.** *Diabetologia* 1981, **20**:87-93.
3. Raffel LJ, Rotter JI: **Diabetes mellitus.** In *Emery and Rimoin's Principles and Practice of Medical Genetics*, edn 4. Edited by Rimoin DL, Connor JM, Pyeritz RE, Korf BR. London: Churchill Livingstone; 2002:2231-2276.
4. Nerup J, Platz P, Ortvad-Andersen O, Christy M, Lyngsoe J, Poulsen JE, Ryder LP, Nielsen LS, Thomsen M, Svejgaard A: **HL-A antigens and diabetes mellitus.** *Lancet* 1974, **2**:864-866.
5. Svejgaard A, Platz P, Ryder LP: **Insulin dependent diabetes mellitus.** In *Histocompatibility Testing 1980*. Edited by Terasaki PI. Los Angeles: UCLA Tissue Typing Laboratory; 1981:638-656.
6. Eisenbarth GS: **Immunogenetics/immunopathogenesis of type 1 diabetes.** *Ann NY Acad Sci* 2003, **1005**:109-118.  
This is a general overview of the immunopathogenesis of T1D. It includes the status of antibody testing in the prediction of T1D.
7. Hirschhorn JN: **Genetic epidemiology of type 1 diabetes.**  
•• *Ped Diabetes* 2003, **4**:87-100.  
A well-reasoned and lucid review of the currently held views of the genetic epidemiology of diabetes.
8. Awdeh ZL, Raum D, Yunis EJ, Alper CA: **Extended HLA/complement allele haplotypes: Evidence for T/t-like complex in man.** *Proc Natl Acad Sci USA* 1983, **80**:259-263.
9. Degli-Esposti MA, Leaver AL, Christiansen FT, Witt CS, Abraham LJ, Dawkins RL: **Ancestral haplotypes: conserved population MHC haplotypes.** *Hum Immunol* 1992, **34**:242-252.
10. Yunis EJ, Larsen CE, Fernandez-Viña M, Awdeh ZL, Romero T, Hansen JA, Alper CA: **Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks.** *Tissue Antigens* 2003, **62**:1-20.  
A recent review of MHC CEHs and smaller haplotype blocks, and an update of their allele compositions in the world's populations.
11. Dawkins R, Lellayuwat C, Gaudieri S, Tay G, Hui J, Cattley S, Martinez P, Kulski J: **Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease.** *Immunol Rev* 1999, **167**:275-304.
12. Raum D, Awdeh Z, Yunis EJ, Alper CA, Gabbay KH: **Extended major histocompatibility complex haplotypes in type 1 diabetes mellitus.** *J Clin Invest* 1984, **74**:449-454.
13. Dawkins RL, Martin E, Saueracker G, Kay PH, Leaver A, Christiansen FT: **Supratypes and ancestral haplotypes in IDDM: potential importance of central non-HLA MHC genes.** *J Autoimmun* 1990, **3**:63-68.
14. Rotter JI, Anderson CE, Rubin R, Congleton JE, Terasaki PI, Rimoin DL: **HLA genotypic study of insulin-dependent diabetes the excess of DR3/DR4 heterozygotes allows rejection of the recessive hypothesis.** *Diabetes* 1983, **32**:169-174.
15. Rubinstein P, Walker M, Mollen N, Carpenter C, Beckerman S, Suci-Foca N, McEvoy R, Ginsberg-Fellner F: **No excess of DR\*3/4 in Ashkenazi Jewish or Hispanic IDDM patients.** *Diabetes* 1990, **39**:1138-1143.
16. Erlich HA, Griffith RL, Bugawan TL, Ziegler R, Alper C, Eisenbarth G: **Implication of specific DQB1 alleles in genetic susceptibility and resistance by identification of IDDM siblings with a novel HLA-DQB1 allele and unusual DR2 and DR1 haplotypes.** *Diabetes* 1991, **40**:478-481.
17. Hoover ML, Marta RT: **Molecular modelling of HLA-DQ suggests a mechanism of resistance in type 1 diabetes.** *Scand J Immunol* 1997, **45**:193-202.
18. Pugliese A, Kawasaki E, Zeller M, Yu L, Babu S, Solimena M, Moraes CT, Pietropaolo M, Friday RP, Trucco M *et al.*: **Sequence analysis of the diabetes-protective human leukocyte antigen-DQB1\*0602 allele in unaffected, islet cell antibody-positive first degree relatives and in rare patients with type 1 diabetes.** *J Clin Endocrin Metab* 1999, **84**:1722-1728.
19. Hermann R, Turpeinen H, Laine AP, Veijola R, Knip M, Simell O, Sipila I, Akerblom HK, Ilonen J: **HLA DR-DQ-encoded genetic determinants of childhood-onset type 1 diabetes in Finland: an analysis of 622 nuclear families.** *Tissue Antigens* 2003, **62**:162-169.
20. Melanitou E, Fain P, Eisenbarth GS: **Genetics of type 1A (immune mediated) diabetes.** *J Autoimmun* 2003, **21**:93-98.
21. Park Y, Eisenbarth GS: **Genetic susceptibility factors of type 1 diabetes in Asians.** *Diab/Metab Res Rev* 2001, **17**:2-11.
22. Kawabata Y, Ikegami H, Kawaguchi Y, Fujisawa T, Shintani M, Ono M, Nishino M, Uchigata Y, Lee I, Ogihara T: **Asian-specific HLA haplotypes reveal heterogeneity of the contribution of HLA-DR and -DQ haplotypes to susceptibility to type 1 diabetes.** *Diabetes* 2002, **51**:545-551.
23. Undlien DE, Lie BA, Thorsby E: **HLA complex genes in type 1 diabetes and other autoimmune diseases. Which genes are involved?** *Trends Genet* 2001, **17**:93-100.
24. Lie BA, Todd JA, Pociot F, Nerup J, Akselsen JE, Joner G, Dahl-Jørgensen K, Rønningen KS, Thorsby E, Undlien DE: **The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene.** *Am J Hum Genet* 1999, **64**:793-800.
25. Johansson S, Lie BA, Todd JA, Pociot F, Nerup J, Cambon-Thomsen A, Kockum I, Akselsen HE, Thorsby E, Undlien DE: **Evidence of at least two type 1 diabetes susceptibility genes in the HLA complex distinct from HLA-DQB1, -DQA1 and -DRB1.** *Gen Immunol* 2003, **4**:46-53.
26. Zavattari P, Lampis R, Motzo C, Loddio M, Mulargia A, Whalen M, Maioli M, Angius E, Todd JA, Cucca F: **Conditional linkage disequilibrium analysis of a complex disease superlocus, IDDM1 in the HLA region, reveals the presence of independent modifying gene effects influencing the type 1 diabetes risk encoded by the major HLA-DQB1, -DRB1 disease loci.** *Hum Mol Genet* 2001, **10**:881-889.
27. Huang HS, Peng JT, She JY, Zhang LP, Chao CC, Liu KH, She JX: **HLA-encoded susceptibility to insulin-dependent diabetes mellitus is determined by DR and DQ genes as well as their linkage disequilibrium in a Chinese population.** *Hum Immunol* 1995, **44**:210-219.
28. Raum D, Awdeh Z, Alper CA: **BF types and the mode of inheritance of insulin-dependent diabetes mellitus.** *Immunogenetics* 1981, **12**:59-74.
29. Rubinstein P, Suci-Foca N, Nicholson JF: **Genetics of juvenile diabetes mellitus. A recessive gene closely linked to HLA-D and with 50 per cent penetrance.** *N Engl J Med* 1976, **297**:1036-1040.
30. Payami H, Thomson G, Motro U, Louis EJ, Hudes E: **The affected sib method. IV. Sib trios.** *Ann Hum Genet* 1985, **49**:303-314.
31. Alper CA, Dubey DP, Awdeh Z: **A simple estimate of the general population frequency of the MHC susceptibility gene for autoimmune polygenic disease.** *Exp Clin Immunogenet* 2000, **17**:138-147.
32. Thomson G: **A two locus model for juvenile diabetes.** *Ann Hum Genet* 1980, **43**:383-398.
33. Mein CA, Esposito L, Dunn MG, Johnson GCL, Timms AE, Goy JV, Smith AN, Sebag-Montefiore L, Merriman ME, Wilson AJ *et al.*: **A search for type 1 diabetes susceptibility genes in families from the United Kingdom.** *Nat Genet* 1998, **19**:297-300.
34. Concannon P, Gogolin-Ewens KJ, Hinds DA, Wapelhorst B, Morrison VA, Stirling B, Mitra M, Farmer J, Williams SR, Cox NJ *et al.*: **A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus.** *Nat Genet* 1998, **19**:292-296.
35. Bennett ST, Wilson AJ, Cucca F, Nerup J, Pociot F, McKinney PA, Barnett AH, Bain SC, Todd JA: **IDDM2-VNTR-encoded susceptibility to type 1 diabetes: dominant protection and parental transmission of alleles of the insulin gene-linked minisatellite locus.** *J Autoimmun* 1996, **9**:415-421.
36. Nisticò L, Buzzetti R, Pritchard LE, Van der Auwera B, Giovannini C, Bosi E, Larrad MTM, Rios MS, Chow CC, Cockram CS *et al.*: **The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes.** *Hum Mol Genet* 1996, **5**:1075-1080.

37. Rich SS, Concannon P: **Challenges and strategies for investigating the genetic complexity of common human diseases.** *Diabetes* 2002, **51**:S288-S294.
38. Marchini J, Cardon LR, Phillips MS, Donnelly P: **The effects of human population structure on large genetic association studies.** *Nat Genet* 2004, **36**:512-517.  
This paper stresses the importance of avoiding population differences between test and control samples in association studies to avoid spurious results.
39. Alper CA, Awdeh Z: **Incomplete penetrance of MHC susceptibility genes: prospective analysis of polygenic MHC-determined traits.** *Tissue Antigens* 2000, **56**:199-206.
40. Jaeckel E, Manns M, Von Herrath M: **Viruses and diabetes.** *Ann NY Acad Sci* 2002, **958**:7-25.
41. Scott FW, Marliss EB: **Conference summary: diet as an environmental factor in development of insulin-dependent diabetes mellitus.** *Can J Physiol Pharmacol* 1991, **69**:311-319.
42. Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J: **Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs.** *Diabetes* 2003, **52**:1052-1055.  
An extensive study that establishes definitively what had been found earlier in Finland: the prevalence of T1D is no higher in dizygotic twins of patients than siblings in general.
43. Wagener DK, Sacks JM, LaPorte RE, Macgregor JM: **The Pittsburgh study of insulin-dependent diabetes mellitus: Risk for diabetes among relatives of IDDM.** *Diabetes* 1982, **31**:1115-1118.
44. Kaprio J, Tuomilehto J, Koskenvuo M, Romanov K, Reunanen A, Eriksson J, Stengård J, Kesänemi YA: **Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland.** *Diabetologia* 1992, **35**:1060-1067.
45. Krimpimäki T, Kupila A, Hämäläinen A-M, Kukko M, Kulmala P, Savola K, Simell T, Keskinen P, Ilonen J, Simell O, Knip M: **The first signs of  $\beta$ -cell autoimmunity appear in infancy in genetically susceptible children from the general population: the Finnish Type 1 Diabetes Prediction and Prevention Study.** *J Clin Endocrinol Metab* 2001, **86**:4782-4788.
46. Lewkonen RM, Gairdner D, Doe WF: **IgA deficiency in one of identical twins.** *Br Med J* 1976, **1**:311-313.
47. Greco L, Romino R, Coto I, Di Cosmo N, Percopo S, Maglio M, Paparo F, Gasperi V, Limongelli MG, Cotichini R *et al.*: **The first large population based twin study of coeliac disease.** *Gut* 2002, **50**:624-628.
48. Louka AS, Sollid LM: **HLA in coeliac disease: unravelling the complex genetics of a complex disorder.** *Tissue Antigens* 2003, **61**:105-117.  
This is a comprehensive summary of the current view of the genetics and immunopathology of GSE. This autoimmune disease, just as with T1D, is MHC allele-associated and appears to be polygenic.
49. Alper CA, Marcus-Bagley D, Awdeh Z, Kruskall MS, Eisenbarth GS, Brink SJ, Katz AJ, Stein R, Bing DH, Yunis EJ, Schur PH: **Prospective analysis suggests susceptibility genes for deficiencies of IgA and several other immunoglobulins on the [HLA-B8, SC01, DR3] conserved extended haplotype.** *Tissue Antigens* 2000, **56**:207-216.
50. Hammarström L, Axelsson U, Björkander J, Hanson LÅ, Möller E, Smith CI: **HLA antigens in selective IgA deficiency: distribution in healthy donors and patients with recurrent respiratory tracts infections.** *Tissue Antigens* 1984, **24**:35-39.
51. Fraser PA, Schur PH: **Hypoimmunoglobulinemia D: frequency, family studies, and association with HLA.** *Clin Immunol Immunopathol* 1981, **19**:67-74.
52. Calvo B, Castaño L, Marcus-Bagley D, Fici DA, Awdeh Z, Alper CA: **The [HLA-B18, F1C30, DR3] conserved extended haplotype carries a susceptibility gene for IgD deficiency.** *J Clin Immunol* 2000, **20**:216-220.
53. Alper CA, Xu J, Cosmopoulos K, Dolinski B, Stein R, Uko G, Larsen CE, Dubey DP, Densen P, Truedsson L *et al.*: **Immunoglobulin deficiencies and susceptibility to infection among homozygotes and heterozygotes for C2 deficiency.** *J Clin Immunol* 2003, **23**:297-305.  
This study partially dissects the parts played by complement deficiency *per se* and other MHC genes in the origin of immunoglobulin deficiencies and increased susceptibility to bacterial infections among C2-deficient patients.
54. Stewart CA, Horton R, Allcock RJN, Ashurst JL, Atrazhev AM, Coggill P, Dunham I, Forbes S, Halls K, Howson JMM *et al.*: **Complete MHC haplotype sequencing for common disease gene mapping.** *Genome Res* 2004, **14**:1176-1187.  
This important work reports 4.75 Mb of contiguous DNA sequence for two common (largely CEH) haplotypes, (HLA-A3, B7, DR2) and (HLA-A1, B8, DR3), associated with protection from and susceptibility to T1D, respectively. This is the first of what will undoubtedly be many reports of the fine structure of MHC haplotypes to map the genetic variants that confer disease susceptibility.