

ORIGINAL ARTICLE

Conserved extended haplotypes discriminate HLA-DR3-homozygous Basque patients with type 1 diabetes mellitus and celiac disease

JR Bilbao^{1,2,9}, B Calvo^{1,9}, AM Aransay³, A Martin-Pagola¹, G Pérez de Nanclares¹, TA Aly⁴, I Rica¹, JC Vitoria^{1,5}, S Gaztambide^{1,6}, J Noble⁷, PR Fain⁴, ZL Awdeh⁸, CA Alper⁸ and L Castaño^{1,5}

¹Endocrinology and Diabetes Research Group, Hospital de Cruces, Barakaldo, Bizkaia, Spain; ²Department of Nursing, University of the Basque Country, Bilbao, Spain; ³CIC Biogune Research Center, Derio, Bizkaia, Spain; ⁴Barbara Davis Center for Childhood Diabetes, Aurora, CO, USA; ⁵Department of Pediatrics, University of the Basque Country, Bilbao, Spain; ⁶Department of Medicine, University of the Basque Country, Bilbao, Spain; ⁷The Children's Hospital Oakland Research Institute, Oakland, CA, USA and ⁸The CBR Institute for Biomedical Research, Boston, MA, USA

*The major susceptibility locus for type 1 diabetes mellitus (T1D) maps to the human lymphocyte antigen (HLA) class II region in the major histocompatibility complex on chromosome 6p21. In southern European populations, like the Basques, the greatest risk to T1D is associated with DR3 homo- and heterozygosity and is comparable to that of DR3/DR4, the highest risk genotype in northern European populations. Celiac disease (CD) is another DR3-associated autoimmune disorder showing certain overlap with T1D that has been explained by the involvement of common genetic determinants, a situation more frequent in DR3-rich populations, like the Basques. As both T1D- and CD-associated HLA alleles are part of conserved extended haplotypes (CEH), we compared DR3-homozygous T1D and CD patients to determine whether CEHs were equally distributed between both disorders or there was a differential contribution of different haplotypes. We observed a very pronounced distribution bias ($P < 10^{-5}$) of the two major DR3 CEHs, with DR3-B18 predominating in T1D and DR3-B8 in CD. Additionally, high-density single nucleotide polymorphism (SNP) analysis of the complete CEH [A*30-B*18-MICA*4-F1C30-DRB1*0301-DQB1*0201-DPB1*0202] revealed extraordinary conservation throughout the 4.9 Mbp analyzed supporting the existence of additional diabetogenic variants (other than HLA-DRB1*0301-DQB1*0201), conserved within the DR3-B18 CEH (but not in other DR3 haplotypes) that could explain its enhanced diabetogenicity.*

Genes and Immunity (2006) 7, 550–554. doi:10.1038/sj.gene.6364328; published online 24 August 2006

Keywords: conserved extended haplotypes; type 1 diabetes mellitus; celiac disease; HLA-DR3; single nucleotide polymorphism; basque

Introduction

The major contributors to the genetic risk to type 1 diabetes mellitus (T1D) map to the major histocompatibility complex (MHC) region on chromosome 6p21, and a strong association of certain human lymphocyte antigen (HLA) class II (DR and DQ) subtypes has been very consistently demonstrated across all the population groups analyzed, although particular risk alleles (and especially their frequencies) vary across ethnic groups and/or geographical regions.¹ For example, DR3/DR4 is the highest risk genotype among Caucasians of northern European origin, whereas in certain southern European populations, like the Basques, the frequency of HLA-DRB1*0301 is markedly increased, making it responsible

for most of the genetic susceptibility to the disease, so that DR3/DR3, DR3/DR4 and DR3/X (X being any DR subtype other than DR3 or DR4) are equivalent in terms of their representation among T1D patients.²

Relative frequencies of risk alleles vary from north to south in European population, and this undoubtedly conditions the genotype distribution within each population and can explain, at least in part, the differences of T1D incidence found across the continent.^{3,4} Additionally, it has been proposed that environmental triggering factors may also be implicated in the expression of clinical disease and thus influence its frequency.⁵ But, from the genetic viewpoint, there have been a number of observations accumulated during the last few decades that taken together, can help to clarify this issue. First of all, it has been known for a long time that linkage disequilibrium between genes in the MHC region is very strong, reflecting the high frequency of certain haplotypes that are highly conserved, and termed 'conserved extended haplotypes' (CEHs) or 'ancestral haplotypes' (AHs).^{6,7} All of the MHC markers of T1D risk are parts of CEHs.⁸ One such CEH carries the diabetogenic DR3

Correspondence: Dr L Castaño, Endocrinology and Diabetes Research Group, Hospital de Cruces, Barakaldo, E48903 Bizkaia, Spain.

E-mail: lcastano@hcr.uosakidetza.net

⁹These authors have equally contributed to this work.

Received 17 April 2006; revised 15 June 2006; accepted 21 June 2006; published online 24 August 2006

subtype and is flanked by alleles *HLA-A*30*, *MICA*4*, *HLA-B*18* and the complotype *F1C30* on the telomeric side and by *HLA-DQA1*0501*, *HLA-DQB1*0201* and *HLA-DPB1*0202* on the centromeric side of *HLA-DRB1*0301*. This DR3-B18 extended haplotype accounts for around 50% of DR3 haplotypes in the Basque general population, whereas its frequency is quite low in northern Europe. Another DR3 CEH, abbreviated DR3-B8, or [*A*01-B*08-MICA*5.1-SC01-DRB1*0301-DQA1*0501-DQB1*0201*] is relatively more frequent in northern Caucasians and represents 25% of Basque DR3 chromosomes.⁹

In Basques, and in other Caucasians, both DR3 haplotypes are associated not only with T1D but also with other MHC-related autoimmune disorders, including gluten-sensitive enteropathy or celiac disease (CD), and the existence of common risk alleles has been put forward to explain the overlap between both disorders.¹⁰ Over the last years, a large body of evidence has accumulated, suggesting the presence of at least a second T1D locus (and another CD-associated gene) within or close to the MHC, and markers located within the HLA class II, III and I regions, as well as the extended class I region have been shown to influence the genetic susceptibility of developing T1D.^{11–16} A recent report suggests that DR3-B8 does not confer increased risk to T1D compared to other less conserved DR3 chromosomes,¹⁷ but, in contrast, there is evidence supporting a stronger contribution of DR3-B18 to the genetic susceptibility to T1D; indeed, several predisposing alleles of these HLA class II-independent genes are present in the DR3-B18 haplotype.^{14,16} It can be inferred that particular DR3 CEHs carrying different variants of these risk-modifying genes would contribute unequally to the risk of developing one or another autoimmune disorders. On the other hand, the strong LD within the entire MHC region makes it extremely difficult to distinguish etiological variants from hitchhiking polymorphisms, and this is particularly true in the case of DR3-B18, which has recently been shown to harbor longer haplotype blocks, compared to other DR3 and DR4 haplotypes.¹⁸

In the present study, we have analyzed the relative contribution of DR3 extended haplotypes to T1D and CD in a Basque isolated population, and additionally, we have performed an analysis of 2360 single nucleotide polymorphisms (SNPs) in the MHC region to contribute in the search of the hypothetical additional susceptibility locus in the MHC area and to further define the structure of the DR3-B18 extended diabetogenic haplotype.

Results and discussion

To determine whether there is any bias in the relative distribution of the DR3-B18 and DR3-B8 CEHs among patients with T1D and CD, which could indicate a differential contribution to risk by component genes within each haplotype, we compared the relative frequencies of these two DR3 CEHs among patients suffering from either T1D ($n=20$) or CD ($n=16$) who were homozygous for (*HLA-DRB1*0301-DQB1*0201*). As shown in Figure 1, the distribution of DR3 haplotypes showed very significant differences between these two disease subsets of the same, highly uniform Basque population. All patients with T1D carried at least one copy of DR3-B18, indicative of a possible increased

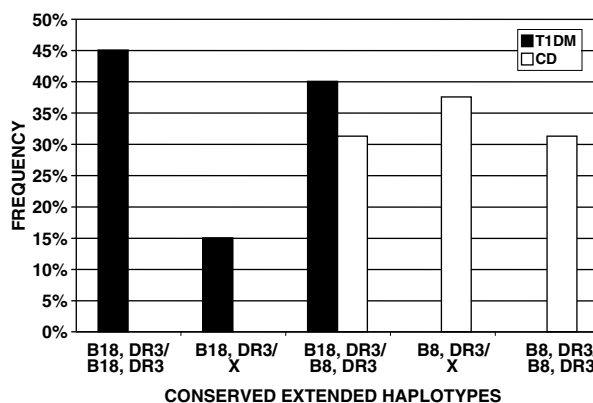


Figure 1 Comparison of the relative frequencies of the DR3-B18 and DR3-B8 CEHs among *HLA-DRB1*0301-DQB1*0201* homozygous patients with T1D ($n=20$) or CD ($n=16$). 'X' corresponds to any HLA-DR3 haplotype different from DR3-B18 and DR3-B8. An extended 2×5 Fisher's exact test¹⁹ revealed significant differences in haplotype distribution between the two groups ($P=0.000006$).

degree of diabetogenicity, whereas all CD patients harbor at least one copy of DR3-B8. In contrast to recent observations on the DR3-B8 haplotype, which appears equivalent to other 'non-conserved' DR3 chromosomes in terms of T1D risk,¹⁷ DR3-B18 is the major genetic contributor to this disease in Basques, a fact that is particularly evident in this population, where this haplotype is relatively frequent and the role of DR4 chromosomes in both T1D and CD susceptibility is reduced. Our results also confirm the existence of a haplotype-dependent bias towards pancreatic or intestinal autoimmunity, probably influenced by a genetic locus different from *HLA-DRB1-DQB1* in the MHC region, and suggest that the T1D-predisposing variant of this putative gene would be present more frequently in DR3-B18 chromosomes compared to other, less predisposing haplotypes. In fact, at least in Basques, and despite the recombination hotspot present centromeric to *HLA-DQB1*, most DR3-B18 haplotypes carry the *HLA-DPB1*0202* allele, and this allele has been shown to independently contribute to diabetes risk in certain population groups.²⁰ Other possible candidate genes include HLA class I genes and complement proteins, and again, alleles that seem to modulate the risk to the disease (including *HLA-A*30*, *HLA-B*18*, *F1C30*), although not always in an HLA-class II-independent manner, are part of the complete DR3-B18 CEH.^{9,14,15,21}

The analysis of 2322 SNPs showed that identity between HLA-haplotype siblings was above 99% and transmission from homozygous parents to their offspring was consistent with Mendelian inheritance. No evidence suggestive of recombination was observed between disease-discordant, HLA-haplotype siblings. Extensive SNP analysis of the two unrelated T1D patients who were homozygous for the [*A*30-B*18-MICA*4-F1C30-DRB1*0301-DQA1*0501-DQB1*0201-DPB1*0202*] CEH showed homozygosity across the entire 4.9 Mbp analyzed with the 2322 SNPs, with only a few occasional heterozygous positions scattered throughout the region, probably artifacts due to technical genotyping errors (Figure 2a). The same genotype was present in two haplotype siblings of one of these patients. In the case of the second patient, his wife and three children

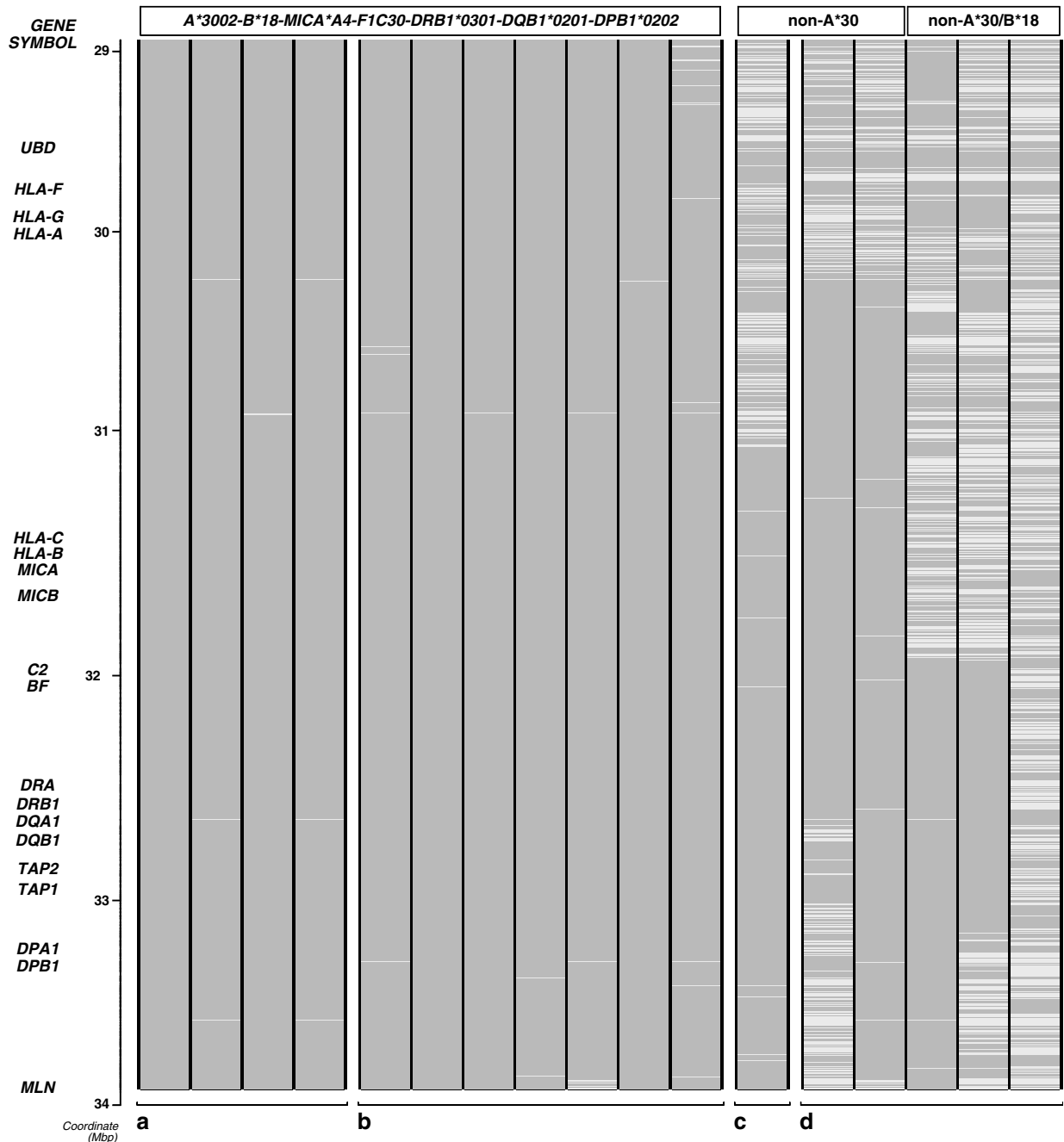


Figure 2 Graphical description of extended HLA-DR3 haplotypes from T1D patients. Each column corresponds to a chromosome and represents the 4.91 Mbp region covered by the 2360 SNPs analyzed. Rows highlighted in white indicate differences from the consensus allele in the Basque [*A*30-B*18-MICA*4-B*18-MICA*4-F1C30-DRB1*0301-DQA1*0501-DQB1*0201-DPB1*0202*] CEH. (a) Four haplotypes from two individuals homozygous for the entire extended haplotype; (b) haplotypes deduced from seven carriers of the extended Basque haplotype; (c) haplotype [*A*03-B*18-MICA*4-F1C30-DRB1*0301-DQA1*0501-DQA1*0501-DQB1*0201-DPB1*0202*], present in a homozygous parent and in all her offspring (see text). (d) Subtracted haplotypes from individuals heterozygous for the Basque haplotype.

were also genotyped because two of them also had T1D. The wife, who was homozygous for the [*A*03-B*18-MICA*4-F1C30-DRB1*0301-DQA1*0501-DQB1*0201-DPB1*0202*] CEH, also presented an extraordinarily homozygous genotype throughout the entire genomic region. Their offspring, carrying a copy of each haplotype, showed a high degree of heterozygosity in the telomeric-most 2 Mbp, and were homozygous thereafter, because both haplotypes are identical centromeric from SNP *rs2530699* (nucleotide 31 063 338) (Figure 2c). Hap-

lotype [*A*03-B*18-DRB1*0301*] appears only in one of our 71 Basque diabetic families and was not included in the SNP analysis. In contrast, [*A*30-B*18-MICA*4-F1C30-DRB1*0301-DQA1*0501-DQB1*0201-DPB1*0202*] is the most frequent extended haplotype among Basque T1D patients, and represents 41% of all diabetic DR3 chromosomes in this group.

A simple Microsoft Excel spreadsheet equation could easily separate *A*03* and *A*30* extended haplotypes in the heterozygous offspring of the family described

above, and was also used to assign haplotypes from the remaining seven unrelated, diabetic individuals who carried one copy of the [A*30-B*18-MICA*4-F1C30-DRB1*0301-DQA1*0501-DQB1*0201-DPB1*0202] CEH (Figure 2b). Five of these seven individuals were homozygous for HLA-DRB1*0301-DQB1*0201, and their subtracted haplotype showed different degrees of similarity compared to the A30-DR3-B18 CEH (Figure 2d). Our results reveal that this CEH is extraordinarily conserved at least throughout the 4.9 Mbp analyzed, and constitutes the longest conserved haplotype block reported so far for the MHC region. A recent study of the extended MHC region has found a higher degree of LD and longer haplotype blocks in this particular CEH, compared to other diabetogenic haplotypes, and we have previously found that conservation of the DR3-B18 CEH extends beyond HLA-DPB1 to RB.^{18,22} The mechanisms responsible for this remarkable conservation are unknown, but plausible explanations include both structural characteristics and selective advantage of certain alleles within this particular haplotype, or could result from a population bottleneck and/or expansion. Nevertheless, the potential impact of these results on the genetics of autoimmune diseases, including T1D, should also be considered; if it is true that another major susceptibility locus is present within the MHC region, it is plausible to hypothesize that the risk variant of this etiological gene would also be conserved (and fixed) in the major Basque DR3 CEH. This would explain why it appears to be more diabetogenic than other DR3 haplotypes, which in turn seem quite uniform in terms of T1D susceptibility.^{15–17} In addition, the relative frequencies of different DR3 haplotypes among populations would also contribute to the geographical differences in risk associated with HLA-DR/DQ genotypes. Finally, this exceptional conservation, together with the enhanced diabetogenicity of the Basque haplotype, also supports the hypothesis that, apart from disease-predisposing (HLA-DRB1*0301-DQB1*0201), these chromosomes carry another etiological gene elsewhere, whose susceptibility allele is fixed in certain CEHs but may not be present in other non-predisposing DR3 chromosomes.²¹ We and others have very recently shown that the diabetogenic DR3-B8 CEH is conserved throughout approximately 3.6 Mb, between HLA-A and HLA-DQ,^{23,24} and those findings, together with our results, might help to define more precisely the susceptibility region. As the risk allele will invariably be present in the CEH, comparison with other, less conserved, predisposing haplotypes (e.g. DR3-B8) will be very helpful for determining the location and nature of the etiological gene. More extensive studies are necessary to identify this susceptibility gene within the MHC region, which could well be outside the region covered by the SNP panel used in this study. Detailed comparison of risk and protective haplotypes in different groups (including non-Caucasian populations) will probably also contribute to this task.

Research design and methods

Sample population

Our group has been collecting and sequentially analyzing putative genetic susceptibility markers (including HLA class I, II and III genes) in families of Basque

descent suffering from T1D (71 families including 81 T1D patients and 249 first-degree relatives) and CD (81 patients from 81 family trios) for more than a decade.^{25,26} In the present work, we have performed additional studies in certain subsets of those families: first of all, to determine the relative frequency of CEHs in both diseases, we have compared T1D and CD patients who were homozygous for the risk haplotype HLA-DRB1*0301-DQB1*0201, and second, a high-density SNP panel was genotyped in 16 T1D patients with at least one DR3 haplotype and 25 of their non-diabetic siblings who were haploidentical by descent (HLA-A through HLA-DPB1). Both parents of one of the families, homozygous and identical for all MHC genes analyzed except HLA-A (A*3002 in the father and A*0301 in the mother), were also genotyped. The study was approved by the Institutional Research Ethics Committee. All genetic analyses have been performed after informed consent of the patients and/or a parent.

SNP typing and analysis

SNP-based analysis of the MHC Panel Set (Cat. No. GT-17-170, Illumina Inc., San Diego, CA, USA) was performed using the *Golden Gate* protocol (Illumina Inc.) to genotype markers across a 4.91 Mbp region spanning the MHC (human chromosome 6, 28 970 148–33 883 424 bp (National Center for Biotechnology Information build 35)). The SNP set consists of two oligonucleotide pools (MHC Mapping Panel and MHC Exon-Centric Panel, for 1228 and 1293 SNP loci, respectively) covering 2360 independent loci spaced at an average of 2 kb. The genotype confidence (GC) score of the assay for keeping allele calls was set to 0.25. Data quality was assessed by controlling for discrepancies between overlapping SNPs in the two panels (161 SNPs present in both pools) and between haploidentical sib-pairs as well as by checking transmission of haplotypes from the homozygous parents to their offspring. In every case, SNPs with more than two discrepant calls were removed from further analysis. Genotypes were obtained from 2322 SNPs: no result was collected from 35 SNPs (because GC score was below the user-defined threshold) and out of the 161 overlapping SNPs in the two oligonucleotide pools, three polymorphisms showed more than two discrepant samples and were also removed. In the remaining overlapping SNPs, consistency was 99.9%.

Individuals who were homozygous for all SNPs served to initially determine haplotypes, which were then estimated by subtraction from the genotypes of other individuals sharing the same HLA CEH (HLA-A through HLA-DPB1). SNP alleles along independent examples of the [A*30-B*18-MICA*4-F1C30-DRB1*0301-DQA1*0501-DQB1*0201-DPB1*0202] CEH were analyzed to determine allele frequencies and a consensus sequence was established. Haplotypes were compared to this consensus sequence and discordant alleles were highlighted. Alleles that were not called in a sample were not accounted for in this analysis.

Acknowledgements

We thank George Eisenbarth and Henry Erlich for their constructive discussions over the years. This work was

funded by Grants RGDM-G03/212 and RCMN-C03/08 from the Instituto de Salud Carlos III of the Spanish Ministry of Health and HL-29583 from the National Heart, Lung, and Blood Institute, NIH, USA (ZLA and CAA). JRB, BC and GPN are FIS researchers supported by the Spanish Ministry of Health Fellowships no. 99/3076, 03/0062 and 03/0064, respectively. AMP is supported by a predoctoral fellowship from the University of the Basque Country.

References

- Pugliese A. Genetics of type 1 diabetes. *Endocrinol Metab Clin N Am* 2004; **33**: 1–16.
- Caillat-Zucman S, Djilali-Saiah I, Timsit J, Bonifacio E, Sepe V, Collins P *et al*. Insulin-dependent diabetes mellitus (IDDM): 12th International Histocompatibility Workshop Study. In: Charon D (ed). *HLA: Genetic Diversity of HLA Functional and Medical Implication*, vol. 1. EDK: Paris 1997, pp 389–398.
- Karvonen M, Viik-Kajander M, Moltchanova E, Libman I, LaPorte R, Tuomilehto J. Incidence of childhood type 1 diabetes worldwide. Diabetes Mondiale (DiaMond) Project Group. *Diabetes Care* 2000; **23**: 1516–1526.
- Ronningen KS, Keiding N, Green A, EURODIAB ACE Study Group (Europe and Diabetes). Correlations between the incidence of childhood-onset type I diabetes in Europe and HLA genotypes. *Diabetologia* 2001; **44** (Suppl 3): B51–B59.
- Åkerblom H, Vaarala O, Hyöty H, Ilonen J, Knip M. Environmental factors in the etiology of type 1 diabetes. *Am J Med Genet* 2000; **115**: 18–29.
- Alper CA. Extended MHC haplotypes and disease markers. *ISI Atlas of Science* 1988; **1**: 79–83.
- Awdeh Z, 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.
- 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.
- Cambon-de Mouzon A, Ohayon E, Hauptmann G, Sevin A, Abbal M, Sommer E *et al*. HLA-A, B, C, DR antigens, Bf, C4 and glyoxalase I (GLO) polymorphisms in French Basques with insulin-dependent diabetes mellitus (IDDM). *Tissue Antigens* 1982; **19**: 366–379.
- Bilbao JR, Martín-Pagola A, Pérez De Nanclares G, Calvo B, Vitoria JC, Vazquez F *et al*. HLA-DRB1 and MICA in autoimmunity: common associated alleles in autoimmune disorders. *Ann NY Acad Sci* 2003; **1005**: 314–318.
- Moghaddam PH, de Knijf P, Roep BO, Van der Auwera B, Naipal A, Gorus F *et al*. Genetic structure of IDDM1: two separate regions in the major histocompatibility complex contribute to susceptibility or protection. Belgian Diabetes Registry. *Diabetes* 1998; **47**: 263–269.
- Lie BA, Sollid LM, Ascher H, Ek J, Akselsen HE, Ronningen KS *et al*. A gene telomeric of the HLA class I region is involved in predisposition to both type 1 diabetes and coeliac disease. *Tissue Antigens* 1999; **54**: 162–168.
- Nejentsev S, Gombos Z, Laine AP, Veijola R, Knip M, Simell O *et al*. Non-class II HLA gene associated with type 1 diabetes maps to the 240-kb region near HLA-B. *Diabetes* 2000; **49**: 2217–2221.
- Zavattari P, Lampis R, Motzo C, Loddo M, Mulargia A, Whalen M *et al*. 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.
- Johansson S, Lie BA, Todd JA, Pociot F, Nerup J, Cambon-Thomsen A *et al*. Evidence of at least two type 1 diabetes susceptibility genes in the HLA complex distinct from HLA-DQB1, -DQA1 and -DRB1. *Genes Immun* 2003; **4**: 46–53.
- Valdes AM, Wapelhorst B, Concannon P, Erlich HA, Thomson G, Noble JA. Extended DR3-D6S273-HLA-B haplotypes are associated with increased susceptibility to type 1 diabetes in US Caucasians. *Tissue Antigens* 2003; **65**: 115–119.
- Ide A, Babu SR, Robles DT, Wang T, Erlich HA, Bugawan TL *et al*. 'Extended' A1, B8, DR3 haplotype shows remarkable linkage disequilibrium but is similar to nonextended haplotypes in terms of diabetes risk. *Diabetes* 2005; **54**: 1879–1883.
- Blomhoff A, Olsson M, Johansson S, Akselsen HE, Pociot F, Nerup J *et al*. Linkage disequilibrium and haplotype blocks in the MHC vary in an HLA haplotype specific manner assessed mainly by DRB1*03 and DRB1*04 haplotypes. *Genes Immun* 2006; **7**: 130–140.
- Uitenbroek DG. SISA-Fisher 2 by 5. Available from <http://home.clara.net/sisa/fiveby2.htm>, accessed 6 December 2005, 1997.
- Noble JA, Valdes AM, Thomson G, Erlich HA. The HLA class II locus DPB1 can influence susceptibility to type 1 diabetes. *Diabetes* 2000; **49**: 121–125.
- Larsen CE, Alper CA. The genetics of HLA-associated disease. *Curr Opin Immunol* 2004; **16**: 660–667.
- Rajsbaum RD, Fici D, Boggs A, Fraser PO, Flores-Villanueva PO, Awdeh ZL. Linkage disequilibrium between HLA-DPB1 alleles and retinoid X receptor b haplotypes. *Hum Immunol* 2002; **63**: 771–778.
- Smith WP, Vu Q, Li SS, Hansen JA, Zhao LP, Geraghty DE. Toward understanding MHC disease associations: partial resequencing of 46 distinct HLA haplotypes. *Genomics* 2006; **87**: 561–571.
- Aly TA, Eller E, Ide A, Gowan K, Babu SR, Erlich HA *et al*. Multi-SNP analysis of MHC region: remarkable conservation of HLA-A1-B8-DR3 haplotype. *Diabetes* 2006; **55**: 1265–1269.
- Bilbao JR, Martín-Pagola A, Calvo B, de Nanclares GP, GEPV-N, Castaño L. Contribution of MIC-A polymorphism to type 1 diabetes mellitus in Basques. *Ann NY Acad Sci* 2002; **958**: 321–324.
- Bilbao JR, Martín-Pagola A, Vitoria JC, Zubillaga P, Ortiz L, Castaño L. HLA-DRB1 and MHC class I chain-related A haplotypes in Basque families with celiac disease. *Tissue Antigens* 2002; **60**: 71–76.