Published in final edited form as:

Diabetes Obes Metab. 2009 February ; 11(Suppl 1): 67–73. doi:10.1111/j.1463-1326.2008.01005.x.

Effect of linkage status of affected sib-pairs on the search for novel type 1 diabetes susceptibility genes in the HLA complex

G. Morahan¹, M. Mehta¹, E. McKinnon², I. James², and the Type 1 Diabetes Genetics Consortium

¹Western Australian Institute for Medical Research and Centre for Medical Research, University of Western Australia, Perth, Western Australia, Australia

²Centre for Clinical Immunology and Biomedical Statistics, Murdoch University and Royal Perth Hospital, Perth, Western Australia, Australia

Abstract

Aim—Type 1 diabetes susceptibility is influenced by a number of genes, of which those with the strongest effects map to the human leucocyte antigen (HLA) complex. Evidence for linkage of non-HLA genes in several affected sib-pair analyses was increased if the HLA or gender status of the sibs was considered independently. We investigated whether linkage status at the HLA complex differentially affected transmission of alleles to sibs depending on their HLA or gender linkage status.

Methods—Genotypes of 2437 markers typed on 11 279 individuals from 2363 families, stratified according to HLA and gender, were examined using the transmission disequilibrium test.

Results—Several significant single nucleotide polymorphisms (SNPs) loci were found near the class II genes; no significant SNPs were found by these analyses near the class I or class II genes. Other significant effects were found when the gender of the sibs or the parents was considered. There was not a significant difference in HLA-DRB1 genotypes between the stratified sets.

Conclusions—These results suggest the presence of novel recessive susceptibility gene(s) within the HLA complex. Keywords: heterogeneity, HLA, linkage, SNPs, type 1 diabetes

Introduction

Type 1 diabetes (T1D) is an autoimmune disease in which the insulin-producing beta cells are destroyed (see Pugliese and Eisenbarth [1], and Pociot and McDermott [2] for reviews). Both genetic and environmental influences are involved in determining an individual's susceptibility to develop T1D [3]. The human leucocyte antigen (HLA) complex contains the major genetic determinant of risk, and this has been recognized for some time as being because of the carriage of particular alleles of the class II genes, *HLA-DRB1* and *-DQB1*. The HLA complex is among the most gene-dense regions in the human genome and contains many genes that encode products that play roles in the immune system [4]. Alleles of these genes often show strong linkage disequilibrium (LD) with alleles of the HLA class II genes. This LD has led to the definition of HLA haplotypes that have been associated with different levels of risk or protection against T1D (reviewed in 5).

Correspondence: Prof. Grant Morahan, Western Australian Institute for Medical Research, MRF Building Level 6, 50 Murray Street, Perth, WA 6000 Australia. gem@waimr.uwa.edu.au.

Conflict of interest: The authors declare that they have no conflicts of interest in publishing this report.

Identification within these haplotypes of non-class II HLA genes that may contribute to T1D risk has been difficult, although evidence has been accumulated implicating a number of genes, including in the central HLA region, class I genes, *MICA* and another locus even more telomeric to HLA [6–11]. To maximize power to identify T1D risk genes, the Type 1 Diabetes Genetics Consortium (T1DGC) was established as a worldwide collaboration to recruit and characterize 4000 affected sib-pair families [11]. DNA from these families is used in ongoing genetic studies to map and identify T1D susceptibility genes [12]. The T1DGC also instigated a detailed investigation of the HLA complex, in which 3000 single nucleotide polymorphisms across 4 Mb of the HLA complex were genotyped in approximately 3000 families, as discussed elsewhere in this supplement. These data provide the opportunity to discover evidence for novel non-class II HLA T1D susceptibility genes.

Several studies have shown that the evidence for non-HLA T1D susceptibility genes is enhanced when one examines subsets of sibs based on relevant parameters, such as linkage at the HLA locus [13–17] and gender [16]. For example, linkage may be compared in sibs who are HLA identical (HLA_i) with those who are HLA mismatched (HLA_m) or in sibs based on their HLA haplotypes (DR3/4, DR4/x, etc.). The rationale for such stratification is that it decreases genetic heterogeneity within this diverse disease. Nevertheless, HLA_m and HLA_i sets of sibs, considered as a whole, did not differ in the proportions of the DR risk haplotypes, at least in the data sets studied by us [16,17]. Therefore, finding increased evidence for linkage in HLA_i vs. HLA_m sibs suggests that non-class II genes contribute to T1D risk, and these genes may act in a recessive fashion, such that HLA_i but not HLA_m sib-pairs have increased risk of T1D.

The above rationale inspired this study to test the hypotheses that sets of sibs, stratified according to relevant features such as HLA linkage status or gender, would show preferential transmission of alleles of SNPs within the HLA complex.

Materials and Methods

Data

The T1DGC obtained DNA samples with full informed consent from participants and with relevant approval from institutional review boards of each recruitment centre. All data were provided de-identified by the T1DGC according to its Access Committee Policy. The Royal Perth Hospital Ethics Committee approved the analyses of these data. SNP genotype data, determined at the Sanger Centre, were provided in two separate sets, ‘OPA1’ and ‘OPA2’, each comprising 1536 markers.

Analyses

The OPA1 and OPA2 data sets were merged, and instances of non-Mendelian inheritance were ‘wiped’ using *PLINK* [18] and *MERLIN* [19] programs. The wiped pedigree files contained data for 2437 markers typed on 11 279 individuals from 2363 families. Pedigree files were recoded as haplotypes using *MERLIN* [19]. Stratification and other pedigree manipulation were performed using the program *SIBSHIPPER* (in prep.). Analyses of allelic transmission were performed using the transmission disequilibrium test (TDT) [20] as implemented in *PLINK* [18]. Comparisons of preferential allelic transmissions between sib-sets were assessed using the likelihood ratio statistic (LRS) for log-linear models, $G^2 = 2 \times \sum (O \times \ln(O/E))$, which was compared with a chi-squared statistic with 1 d.f. Estimates of LD between SNPs were calculated using the *HAPLOVIEW* program [21]. All analyses were performed on a Quad-Core Macintosh Pro under *MAC OS 10.4.10*.

Results

Determination of HLA Linkage Status and Stratification of the Data Set

Two data sets of genotypes provided by the T1DGC were merged using *PLINK* [18]. The merged pedigree file was then analysed using *MERLIN*'s data-checking algorithms, producing 'wiped' datafiles to eliminate Mendelian inconsistencies. After this data cleaning was completed, information on genotypes of 2437 markers from 11 279 individuals was obtained. *MERLIN* was also used to calculate the likelihood of the odds (LOD) score at each SNP and to construct haplotypes within each family. Significant results were obtained both for linkage and by TDT, as also reported elsewhere in this supplement.

The program 'SIBSHIPPER' was developed to manage linkage and other relevant status within sibships and to select sibs for 'shipping' to stratified pedigree files for further analyses. SIBSHIPPER is available at <http://www.sibshipper.waimr.uwa.edu.au>. Sib-pair families were separated by SIBSHIPPER into two groups on the basis of their identity by descent for haplotypes of the four SNPs (rs9275184, rs2856726, rs9275495 and rs7454108) at which sharing was greatest (i.e. having the highest LOD scores as calculated by *MERLIN*). In the case of individuals with recombinant haplotypes, sharing was determined on the basis of genotype at the SNP rs9275495, which showed the highest overall LOD score (121.9) in the entire data set. Finally, only the first affected sib from each family was considered for subsequent analyses to eliminate potential bias in calculating TDT scores for each data set.

The resulting pedigree files contained the first sib from each family that otherwise had HLA identical sib-pairs (HLA_i) or sibs who were mismatched at HLA (HLA_m). These sib-pairs would have had two haplotypes identical-by-descent (in the case of HLA_i families) or either one or zero haplotypes identical-by-descent (HLA_m). After excluding individuals in *PLINK* for data missingness (*GENO* > 0.1), and low minor allele frequencies (<0.01), the HLA_i set included 1620 trio families, while the HLA_m set had 590 trios. It should be noted that these two sets did not differ substantially in the representation of traditional HLA-DR-based haplotypes. The proportion of HLA-DR x/x, HLA-DR3, HLA-DR3/4, and HLA-DR4 subjects in each set were as follows: 0.06, 0.38, 0.30, 0.25, and 0.10, 0.36, 0.33, and 0.20, respectively.

Comparison of HLA_i and HLA_m Sib-pair Data Sets

Transmission disequilibrium test (TDT) *p* values (and $-\log p$ scores) were calculated for each SNP for three sets of parents/first sibs: probands of the total (filtered) data set, the HLA_i probands and the HLA_m probands. The results are summarized in figure 1. The peak TDT score coincides with the location of the class II genes, covered by SNPs numbered in the range ~1860 to ~2280. The highest scoring SNP (rs6927022, $-\log p = 148.9$) is located immediately centromeric of *HLA-DQA1*. The HLA_i sib families contributed most to the total TDT score; the peak value (127.8) was observed for the same SNP. Although the HLA_m sibs had much lower scores overall, in part because of the smaller sample size, they nevertheless showed highly significant TDT scores for SNPs in the class II region, many with *p* values <10⁻¹⁰. These results confirm the importance of the class II genes in conferring susceptibility to T1D.

Preferential Transmission of Alleles Between HLA_i and HLA_m Sibs

We next addressed our major hypothesis that the HLA identical, but not mismatched, sibs would have additional susceptibility gene(s). The null hypothesis was that these groups would show no significant difference in the transmission ratio of alleles. This was tested by determining the number of transmitted and untransmitted alleles of each SNP in both sets of sibs and then comparing the sets using the LRS, *G*². Significance was expressed by the score $-\log p$ and is displayed in figure 2. Following Bonferroni correction, scores greater than 4.7 would be sufficient to reject the null hypothesis with nominal *p* < 0.05.

There was a total of 56 SNPs that exceeded the 4.7 threshold value. Again, most of these were found to reside within the class II interval. The peak score was found at SNP rs7454108, which is located in an intergenic region upstream of *HLA-DQB1*. Two other SNPs with $-\log P > 20$ were also in this region and were in almost LD with rs7454108.

Effect of Gender on Transmission of SNP Alleles

Although equal proportions of males and females are affected by T1D, we have shown previously [16] that linkage to non-HLA genes may be affected by gender status. Therefore, we tested whether there was also any effect of gender on inheritance of SNP alleles within the HLA region. *SIBSHIPPER* was used to create a pedigree file in which there were only affected brothers and also another of only affected sisters. There were a total of 766 all-male affected sib families and 611 all-female affected sib families. Remaining families, in which there was an affected male and an affected female, were not tested further.

Overall, the TDT scores in the two data sets showed a similar distribution to that shown in figure 1, with a peak over the class II region. Following a similar approach to that above, we tested for differential allelic transmission between the all-male and all-female families. The results, summarized in figure 3, show no SNP withstood Bonferroni correction, and that there were only 4 SNPs $-\log P > 3$ scores for differences in transmission between families of affected males and those having affected females. The SNPs with the greatest difference between these groups were on the border of the class II region between *HLA-DRA* and *NOTCH4*.

Finally, we tested whether any SNP showed a parent-of-origin effect, that is whether there was a difference in transmissions to affected offspring from heterozygous fathers vs. heterozygous mothers. No SNP yielded sufficiently low p values to withstand Bonferroni correction. However, one SNP did provide a potentially interesting observation. This marker (rs6931763, near *TRIM39*) had $p = 0.00013$ with preferential transmission from fathers of female affected sibs. This SNP showed no parent-of-origin effect among parents of male affected sibs. It may be interesting to evaluate this observation in another data set.

LD of Selected HLA SNPs

It is possible that the results obtained regarding preferential transmission of SNPs was due to their carriage by particular susceptibility haplotypes and thus LD with susceptibility alleles of the class II genes. Such LD would invalidate our data implicating additional non-class II genes in T1D susceptibility. Therefore, we examined the relationships between each of the SNPs that showed Bonferroni-corrected significant results in the *HLA_I* vs. *HLA_m* and male vs. female comparisons, as well as selected SNPs throughout the HLA complex. LD was evaluated by calculating r^2 values, and LD blocks were plotted using *HAPLOVIEW* [21]. As shown in figure 4, the major haplotype block encompasses the *HLA-DQA2* and *-DQB2* genes. There appears to be little LD overall between the non-class II SNP clusters and the class II genes.

Discussion

The HLA complex has been known for decades as the major risk factor for T1D. The principle genes affecting susceptibility have been identified by genetic, molecular and immunological studies as the class II genes. Whether other HLA genes within or linked to the HLA complex also play a role in determining disease risk has been a moot point, with studies implicating genes in the class I and class III regions, as well as those more telomeric (*MICA*) and at the furthest end of the complex itself [5–10]. The difficulty in identifying additional HLA susceptibility genes has until now been complicated by problems of obtaining sufficient numbers of DNA samples from families affected by T1D to deal with the further problem of high LD within the HLA complex.

The T1DGC provided the resources to recruit 4000 families of affected sib-pairs as well as to characterize these families for genetic, clinical and immunological markers relevant to T1D [10,11]. As part of this effort, the T1DGC sponsored genotyping of ~3000 SNPs in over 11 000 individuals. The resulting data provide new opportunities for insights into the genetic risk factors for T1D; we chose to use these data to test our hypotheses that additional II gene acted in a recessive fashion to increase susceptibility and that other genes had preferential effects on at-risk males and females.

It is important to note that the tests we performed simply address the null hypothesis that there would be no difference in the ratio of transmitted : untransmitted SNP alleles when one compares sets of families selected by proband HLA linkage status or their gender. That is alleles of susceptibility genes would be transmitted in the same proportion in these families, regardless of their HLA linkage or gender status. These null hypotheses could conservatively be rejected if the $-\log(p)$ value exceeded a 4.7 threshold. Regarding the HLA_i vs. HLA_m sibs, 65 SNPs yielded results with $-\log(p)$ values as high as 27.7. Comparison of male and female affected sib families resulted in no SNPs able to reject the null hypothesis, but nevertheless, four SNPs were found to provide suggestive results with $-\log(p) < 3$. Males aged 15–40 years are known to be affected by T1D in ~50% excess in populations of European origin (reviewed in 22); whether any of these four SNPs are found in males with later age-at-diagnosis could be the subject of a further study.

The SNPs we identified appear to be in LD with a distinct risk gene rather than being passively carried by ancestral haplotypes [6] associated with the high-risk alleles of the DR/DQ genes. The most preferentially transmitted SNPs were found on the HLA-DR4 haplotype; approximately one-third of all HLA-DR4 haplotypes did not carry the over-transmitted SNP alleles. Genes near these SNPs could be examined for playing a role in T1D.

In the course of performing these analyses, we needed to be able to stratify the families by a number of criteria, especially gender status and identity by descent at HLA loci. As there was no suitable software available for this, we developed *SIBSHIPPER*, which will be made available publicly, and described in more detail elsewhere.

Our strategy was successful in identifying SNPs in or near class II HLA genes, which were preferentially transmitted in families in which affected sib-pairs were HLA identical, as well as other SNPs, which may have different effects in females and males. The results suggest that there are indeed additional genes that affect T1D susceptibility. The studies reported in this study are a first test of our hypotheses and strategies, and our preliminary data will need to be taken further in more detailed examinations.

Acknowledgments

This work was supported by funding from the Diabetes Research Foundation of Western Australia (G. M.), and by separate Program Grants from the National Health and Medical Research Council of Australia (G. M. and I. J.). M. M. was supported by a grant from the Strategy Fund of the University of Western Australia. The T1DGC was established and supported by NIDDK (U01 DK062418) and JDRF.

References

1. Pugliese A, Eisenbarth GS. Type 1 diabetes mellitus of man: genetic susceptibility and resistance. *Adv Exp Med Biol* 2004;552:170–203. [PubMed: 15622964]
2. Pociot F, McDermott MF. Genetics of type 1 diabetes mellitus. *Genes Immun* 2002;3:235–249. [PubMed: 12140742]
3. Hirshhorn JN. Genetic epidemiology of type 1 diabetes. *Pediatr Diabetes* 2003;4:87–100. [PubMed: 14655265]

4. The MHC Sequencing Consortium. Complete sequence and gene map of a human major histocompatibility complex. *Nature* 1999;401:921–923. [PubMed: 10553908]
5. Tait BD, Harrison LC. Overview: the major histocompatibility complex and insulin dependent diabetes mellitus. *Baillieres Clin Endocrinol Metab* 1991;5:211–228. [PubMed: 1892463]
6. Degli-Esposti MA, Abraham LJ, McCann V, Spies T, Christiansen FT, Dawkins RL. Ancestral haplotypes reveal the role of the central MHC in the immunogenetics of IDDM. *Immunogenetics* 1992;36:345–356. [PubMed: 1356098]
7. Johansson S, Lie BA, Todd JA, 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. [PubMed: 12595901]
8. Tait BD, Colman PG, Morahan G, et al. HLA genes associated with autoimmunity and progression to disease in type 1 diabetes. *Tissue Antigens* 2003;61:146–153. [PubMed: 12694582]
9. Gambelunghie G, Ghaderi M, Tortoioli C, et al. Two distinct MICA gene markers discriminate major autoimmune diabetes types. *J Clin Endocrinol Metab* 2001;86:3754–3760. [PubMed: 11502807]
10. Lie BA, Akselsen HE, Joner G, et al. HLA associations in insulin-dependent diabetes mellitus: no independent association to particular DP genes. *Hum Immunol* 1997;55:170–175. [PubMed: 9361969]
11. Rich SS, Concannon P, Erlich HA, et al. Sanjeevi CB, Hanafusa T. The Type 1 Diabetes Genetics Consortium. *Immunology of Diabetes IV. Annals of the New York Academy of Sciences*, New York 2006;1079:1–8.
12. Concannon P, Erlich HA, Julier C, et al. Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. *Diabetes* 2005;54:2995–3001. [PubMed: 16186404]
13. Davies JL, Kawaguchi Y, Bennett ST, et al. A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 1994;371:130–136. [PubMed: 8072542]
14. Owerbach D, Gabbay KH. The HOXD8 locus (2q31) is linked to type I diabetes. Interaction with chromosome 6 and 11 disease susceptibility genes. *Diabetes* 1995;44:132–136. [PubMed: 7813807]
15. Luo DF, Buzzetti R, Rotter JJ, et al. Confirmation of three susceptibility genes to insulin-dependent diabetes mellitus: IDDM4, IDDM5 and IDDM8. *Hum Mol Genet* 1996;5:693–698. [PubMed: 8733139]
16. Morahan G, Huang DX, Tait BD, Colman PG, Harrison LC. Markers on distal chromosome 2q linked to insulin-dependent diabetes mellitus. *Science* 1996;272:1811–1813. [PubMed: 8650584]
17. Morahan G, Huang DX, Ymer SI, et al. Linkage disequilibrium of a type 1 diabetes susceptibility locus with a regulatory *IL12B* allele. *Nat Genet* 2001;27:218–221. [PubMed: 11175794]
18. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575. [PubMed: 17701901]
19. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin – rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002;30:97–101. [PubMed: 11731797]
20. Spielman R, McGinnis R, Ewens W. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;52:506–516. [PubMed: 8447318]
21. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–265. [PubMed: 15297300]
22. Gale EA, Gillespie KM. Diabetes and gender. *Diabetologia* 2001;44:3–15. [PubMed: 11206408]

Fig. 1.

Transmission disequilibrium test (TDT) of single nucleotide polymorphisms (SNPs) in human leucocyte antigen (HLA) identical and HLA mismatched sib-pair families. The TDT [20] was performed on trio families from the total data set (triangles) as well as the first affected sib from each of the HLA identical (HLA_i, open circles) and HLA mismatched (HLA_m, filled circles) sib-pairs. SNP position is indicated by the number of each SNP in the study, ordered by position on chromosome 6. TDT results are shown as the negative log of the corresponding p value.

Fig. 2. Preferential inheritance of single nucleotide polymorphism (SNP) alleles by sibs showing linkage to human leucocyte antigen (HLA). To test the null hypothesis of no difference in transmission between sets of families with different HLA linkage status, the numbers of transmitted and untransmitted alleles of each SNP were determined for the HLA identical (HLA_i) and HLA mismatched (HLA_m) sib families. These were used to derive G^2 scores from which the negative logp values were calculated. The grey line indicates a statistically significant deviation in transmission between data sets, at nominal $p < 0.05$, corrected for 2437 tests.

Fig. 3. Preferential inheritance of single nucleotide polymorphism (SNP) alleles according to sibs' gender status. SNPs with significantly different transmission between male and female affected sibs were sought by calculating G^2 as above and converting to the negative log of the resulting p value.

Fig. 4. Linkage disequilibrium (LD) between preferentially transmitted human leucocyte antigen (HLA) single nucleotide polymorphisms (SNPs). The set of SNPs showing significantly different transmission between HLA identical (HLA_i) and HLA mismatched (HLA_m) sib families (56 SNPs) between male and female sib pairs (four SNPs) and the SNP showing the strongest parent-of-origin effect were selected together with other selected SNPs across the HLA region. Genotypes of these SNPs in all families were assessed using HAPLOVIEW [21]. The upper lines indicate the positions of the SNPs on chromosome 6 from *rs6931763* to *rs241404*. The lower part displays r^2 values between each SNP, where $r^2 = 0$ is shown in white, $r^2 = 1$ in black and intermediate values in shades of grey. The major block of LD is within the class II region, from *rs3916765* to *rs9276598*, and this includes *HLA-DQA2* and *-DQB2*.