Published in final edited form as:

Diabetes Obes Metab. 2009 February ; 11(Suppl 1): 31–45. doi:10.1111/j.1463-1326.2008.01001.x.

Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A

J. M. M. Howson, N. M. Walker, D. Clayton, J. A. Todd, and the Type 1 Diabetes Genetics Consortium

Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK

Abstract

Aim—Until recently, human leucocyte antigen (HLA) class II-independent associations with type 1 diabetes (T1D) in the Major Histocompatibility Complex (MHC) region were not adequately characterized owing to insufficient map coverage, inadequate statistical approaches and strong linkage disequilibrium spanning the entire MHC. Here we test for HLA class II-independent associations in the MHC using fine mapping data generated by the Type 1 Diabetes Genetics Consortium (T1DGC).

Methods—We have applied recursive partitioning to the modelling of the class II loci and used stepwise conditional logistic regression to test ~1534 loci between 29 and 34 Mb on chromosome 6p21, typed in 2240 affected sibpair (ASP) families.

Results—Preliminary analyses confirm that *HLA-B* (at 31.4 Mb), *HLA-A* (at 30.0 Mb) are associated with T1D independently of the class II genes *HLA-DRB1* and *HLA-DQB1* ($P = 6.0 \times 10^{-17}$ and 8.8×10^{-13} , respectively). In addition, a second class II region of association containing the single-nucleotide polymorphism (SNP), rs439121, and the class II locus *HLA-DPB1*, was identified as a T1D susceptibility effect which is independent of *HLA-DRB1*, *HLA-DQB1* and *HLA-B* ($P = 9.2 \times 10^{-8}$). A younger age-at-diagnosis of T1D was found for *HLA-B*39* ($P = 7.6 \times 10^{-6}$), and *HLA-B*38* was protective for T1D.

Conclusions—These analyses in the T1DGC families replicate our results obtained previously in ~2000 cases and controls and 850 families. Taking both studies together, there is evidence for four T1D-associated regions at 30.0 Mb (*HLA-A*), 31.4 Mb (*HLA-B*), 32.5 Mb (rs9268831/*HLA-DRA*) and 33.2 Mb (rs439121/*HLA-DPB1*) that are independent of *HLA-DRB1*/*HLA-DQB1*. Neither study found evidence of independent associations at *HLA-C*, *HLA-DQA1* loci nor in the *UBD/MAS1L* or *ITPR3* gene regions. These studies show that to find true class II-independent effects, large, well-powered sample collections are required and be genotyped with a dense map of markers. In addition,

© 2009 The Authors Journal Compilation © 2009 Blackwell Publishing Ltd

Correspondence: Joanna M. M. Howson, Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0XY, UK. joanna.howson@cimr.cam.ac.uk.

Additional Supporting Information may be found in the online version of this article.

Conflict of Interest:

The authors declare that they have no conflicts of interest in publishing this article.

Please note: Wiley-Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

a robust statistical methodology that fully models the class II effects is necessary. Recursive partitioning is a useful tool for modelling these multiallelic systems.

Keywords

HLA-A; HLA-B; HLA-DPB1; type 1 diabetes

Introduction

The Major Histocompatibility Complex (MHC) is the major susceptibility locus for type 1 diabetes (T1D). The class II loci, *HLA-DRB1* and *HLA-DQB1*, have long been accepted as having the strongest effects [1,2]. However, over the past 10 years, there have been numerous reports of other MHC loci in addition to these, that are also apparently associated with T1D [3–9]. Yet, the MHC is a region that is renowned both for extensive linkage disequilibrium (LD) spanning several megabases (Mb) [10,11] and high levels of variability, with the human leucocyte antigen (HLA) genes (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-DQB1*, *HLA-DQA1*, *HLA-DRB1*, *HLA-DPB1*, *HLA-DPA1*) having hundreds of alleles (http://www.anthonynolan.org.uk/HIG/nomen/nomen_index.html). Despite this, some authors claim the associations they have found are independent and, in fact, add to the established class II associations [4,6,8]. Until recently [12], none of these studies was adequately powered, had genotyped a sufficiently dense marker map across the entire MHC region and had modelled the class II effects appropriately to be confident that the associations were not attributable to either the class II loci themselves or other (untyped) loci with which the locus under study was in LD.

To test for class II-independent effects, a dense marker map needs to be genotyped in a well-powered sample set of thousands of subjects. Small sample sets (<100 trios) are well powered to find the major class II loci, but to detect the smaller effects of non-class II loci in the MHC, after taking class II into account, require much larger data sets [12]. A sparse marker map would leave uncertainty in the assessment of independent associations as the actual locus providing the independent association signal may lie in a region not covered, or worse, the signal may be missed completely as no marker in LD with the independent locus has been genotyped.

To be confident, the associations found on 6p21 are independent of the MHC class II associations, the *HLA-DRB1* and *HLA-DQB1* class II effects have to be accounted for because LD spans the entire MHC region [10,11]. Studies that do not account for class II effects and are testing for non-class II associations are invalid. Accounting for the class II effects is challenging because the number of alleles at both *HLA-DRB1* and *HLA-DQB1* give rise to a huge number of genotypes and haplotypes. Therefore, in T1D, many authors have resorted to matching cases and controls according to their *HLA-DRB1* genotype or *HLA-DQB1* genotype, or perhaps specific *HLA-DRB1*–*HLA-DQB1* haplotypes. These methods have three disadvantages: first, they reduce the size of the data set used and potentially the power of the study; secondly, they require multiple testing correction for all the genotypes and haplotypes considered at the class II loci and thirdly, they limit the amount of information that can be obtained about the test locus as the analysis is confined to a handful of specific genotypes or haplotypes. Another method of accounting for the class II effects is to group together the alleles, genotypes or haplotypes of the class II loci in some way to reduce the dimensionality of the parameter space required to model them. This approach uses all the data and does not have the problems of subgroup analysis inherent in the testing of specific haplotypes or genotypes. The difficulty with this approach is choosing a method of grouping the class II genotypes/alleles/haplotypes. Another approach is to group them together using a frequency threshold, for example, alleles with a frequency of less than 0.10, but this assumes rare alleles have the same

effect. Alternatively, the grouping can be based on risk, that is, group alleles with similar effect together. The disadvantage of this grouping strategy is that the allelic or genotypic risks need to be known *a priori*, which may not be the case. The results of testing for class II-independent effects have been shown to be reliant, in some instances, on how the class II loci have been modelled [12]. Therefore, it is imperative to have a robust model of class II effects.

To evaluate whether there are associations in the MHC that are independent of *HLA-DRB1* and *HLA-DQB1*, the T1DGC have genotyped an extensive panel of loci, ~3000 SNPs between 29 and 35 Mb on chromosome 6p21, the class II loci *HLA-DPA1*, *HLA-DPB1*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1* and the class I loci, *HLA-A*, *HLA-C* and *HLA-B*. These loci have been genotyped in 2300 affected sibpair (ASP) families, and hence the study is well powered to identify class II-independent effects. We have accounted for the confounding effects of *HLA-DRB1* and *HLA-DQB1* by using recursive partitioning to generate a tree model of these genes and used this model in a conditional logistic regression analysis of the remaining loci [12]. This approach has the advantage over the other methods of accounting for class II, discussed above, of using all the data so as to retain power; the genotype risks are not required *a priori* and the multiple testing correction associated with subgroup analysis is not required. We have used this approach previously to show that both the class I genes, *HLA-B* and *HLA-A* are associated with T1D independently of *HLA-DRB1* and *HLA-DQB1* [12] and now replicate the results in the T1DGC family data set.

Materials and Methods

The November 2007 release of the T1DGC's MHC Fine Mapping data was used for all analyses.

Subjects

In total, 2300 families of two parents with two T1D-affected offspring were used for the T1DGC study. These were from nine cohorts, Asia Pacific (AP; 191), North America (NA; 334), Human Biological Data Interchange (HBDI) (431), Joslin (JOS; 112), Europe (EUR; 475), Sardinia (SAR; 78), UK (114), British Diabetes Association (BDA) (418) and Denmark (DAN; 147). The AP, EUR, NA and UK collections were recruited specifically for the T1DGC study, whereas the remainder were part of established collections. All subjects were asked to give their primary, secondary and tertiary ethnic group, these were cross-referenced between parents and offspring and families with inconsistencies dropped (60 families: 18AP; 36 NA; 1 EUR; 5 UK), so that statistical analysis was confined to families of white European origin.

Genotyping

The classical HLA loci (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DRB1*) were genotyped for all cohorts (except BDA) using both sequence-specific oligonucleotide probe-based method and line strips from Roche Molecular Systems (Alameda, CA, USA). In the BDA samples, genotyping of the *HLA-DQB1*, *HLA-DRB1*, *HLA-A*, *HLA-B* and *HLA-C* genes was performed predominantly using Dynal RELI SSO assays (Invitrogen, Paisley, UK). In total, 3072 SNPs were genotyped on two oligonucleotide pool assays (OPA1 and OPA2) using the Illumina Golden Gate technology at the Wellcome Trust Sanger Institute (there was some inbuilt redundancy with 115 SNPs common to both chips). It should be noted that 166 samples typed at the classical loci were not genotyped on either OPA1 or OPA2.

Statistics

All analyses unless otherwise stated were carried out with in the statistical package STATA (www.stata.com) or the R environment (www.r-project.org) [13].

Genotype Calling

The data from OPA1 and OPA2 were called using the Illuminus calling algorithm [14]. The JOS, SAR and DAN collections were genotyped using whole genome amplified (WGA) DNA and so were clustered separately to the DNA samples in accordance with the Illuminus recommendations. Having separated the WGA and genomic DNA samples, signal plots still lacked clarity, with many SNPs showing greater than three clusters. This has previously been observed [15] and we concluded this was because of differential sample preparation and lab-batch processing effects. Samples were therefore separated by DNA source (BDA, HBDI, and the samples collected for the T1DGC study, AP, EUR, NA, UK). Clustering was still not considered satisfactory and this was traced to the late addition of samples from on going collections (plates 11305, 11306, 11307 from NA and 11298, 11299, 11300 from EUR). Thus, these plates were called separately to the remainder of plates.

All SNPs with a call rate of less than 95%, within samples clustered together, were dropped (this cut-off is consistent with other genome-wide association (GWA) studies [16]). SNPs with a minor allele frequency (MAF) across all samples of less than 0.05 were also dropped (607 and 494 for OPA1 and OPA2 respectively). SNPs out of Hardy–Weinberg equilibrium (HWE) in parents not known to have T1D ($P = 3 \times 10^{-5}$) were dropped from the analysis (143 SNPs on OPA1 and 122 SNPs on OPA2). Ideally, one would like to retain SNPs out of HWE as the MHC region is known to be under-going selection and so the genotypes of MHC loci may not segregate according to HWE. However, inspection of the signal clouds revealed that most of these SNPs were poorly clustered (e.g., had four clusters instead of the expected three) or called. Of the original 115 common SNPs, only 31 remained on both chips after the above quality control (QC) steps. Signal clouds for all 31 on both platforms were examined: 16 clustered badly on OPA1 so OPA2 data were used; 12 had bad clustering on OPA2 so OPA1 data were used. The data for three SNPs that clustered well on both chips were combined with inconsistencies recoded to missing. The misinheritance rate of each SNP was considered and the signal clouds of SNPs with greater than 5% misinheritances (11 SNPs) were examined. All 11 SNPs clustered poorly and so were not analysed (table S1).

In total 1535 unique SNPs were included for association testing with T1D. All SNPs found to be strongly associated had their genotyping signal clouds examined [17] (www.t1dbase.org). Even after the QC measures described above, the SNPs rs1633097, rs2524024, rs2023478, rs7756993, rs3901554, rs3093542, rs6914950, rs3117583 and rs6911279 were found to have poor clustering after association testing and so were dropped. We recommend the signal clouds for any SNP that shows a strong association (low p-value) should be examined, as well as further consideration of HWE results (e.g., if an SNP is out of HWE in the parents yet in HWE in offspring).

Single Locus Analyses

Sets of cases and matched pseudo-controls (consisting of the three genotypes that could have been transmitted to offspring but were not) were generated and analysed using conditional logistic regression [18]. Non-independence of offspring was accounted for by using Huber/White sandwich estimators. SNPs were coded as 0, 1 and 2, representing the genotypes 1/1, 1/2 and 2/2, respectively. A Wald test was used to test non-multiplicative effects by including a 'dominance' term (coded 0.5, -0.5 and 0.5) in the regression model and analysing its additional effect. The eight classical *HLA* genes were coded without assuming a specific mode of inheritance, by including an indicator variable for each possible genotype in the conditional logistic regression model. Low-frequency genotypes (MAF <0.001) were grouped together before analysis. The microsatellites and classical *HLA* genes were tested for non-additive effects by comparing the model

$$\text{clogit}(p) \sim \sum_{i=1}^{N-1} \vec{x}_i \beta_i \quad (1)$$

with

$$\text{clogit}(p) \sim \sum_{i=1}^{N-1} (\vec{x}_i \beta_i + (\vec{x}_i \hat{\beta}_i)^2) \quad (2)$$

using a Wald test, where $\vec{x}_i$ is a vector of allele counts and i is used to sum over the N alleles at a given locus. β_i is a vector of coefficients to be estimated in eqn (1) and used for $\hat{\beta}_i$ in eqn (2), and p is the probability that an offspring has genotype $\times$ conditional on the genotypes of the parents and disease status.

Recursive Partitioning: Grouping of HLA-DRB1 and HLA-DQB1 Genotypes

A classification tree approach was used to group the *HLA-DRB1* and *HLA-DQB1* genotypes. The alleles of *HLA-DRB1* and *HLA-DQB1* do not behave multiplicatively in conferring T1D risk, an example of this is the 'DR3/4' genotype effect which has been widely observed. Hence, to allow for non-multiplicative effects, genotype (rather than allele) effects were modelled. The classification method adopted was recursive partitioning as implemented in the *R* rpart library (<http://cran.r-project.org>, [13,19,20]).

Recursive partitioning begins with the full (unsplit) data set, termed the 'root node'. All possible binary splits corresponding to presence or absence of the different genotypes at *HLA-DRB1* and *HLA-DQB1* are considered. The split that maximizes the reduction in the deviance in disease status is accepted, such that two subgroups (or nodes) are created. Each split reduces the fitting error, so that the groups formed are more homogeneous with respect to disease status than the original node. The procedure is then repeated on each of these two new nodes, and subsequent nodes, continuing until no further splits or reduction in deviance is possible. A visual representation of this method looks like a tree, with a single root node connected to the offspring nodes by 'branches'. Terminal nodes appear as 'leaves' and represent optimized groups of the *HLA-DQB1/HLA-DRB1* genotypes, defining strata within which additional MHC loci can be tested.

Cases and matched pseudo-controls were generated (conditioned on transmitted and untransmitted genotypes, regardless of phase, when phase is not inferable and conditioning on phase being inferable otherwise) [18] for use in recursive partitioning. As the default splitting routine does not allow retention of the matching between case and pseudo-controls, an alternative and new set of functions were created (user-splits.R – see Appendix). In addition to an initialization function this provides a function (evaluate) that evaluates how 'splittable' a node is. Here, we fit the current tree model using conditional logistic regression. If L is the likelihood for the model, then the deviance, $D = -2 \ln L$, can be calculated from the conditional logistic regression model. The deviance for the offspring nodes should be less than that of the parents for a node to be 'splittable'. If a node is 'pure' and 'unsplittable', the deviance will be zero. The function 'split' chooses the next split to maximize a goodness-of-split metric, here based on a score test for adding the corresponding binary covariate to the current conditional logistic regression model.

Pruning of the HLA-DQB1/HLA-DRB1 Trees

The tree produced using the above method is too complex to use for analysis and, hence, needed to be pruned. To assess how many leaves to prune, *rpart* uses the minimal cost-complexity pruning metric, $R_\alpha = R_{\text{leaves}} + \alpha S R_{\text{leaves}}$ is the sum of the deviance values in the leaves, the complexity parameter, $\alpha \geq 0$ and S is the size of the tree, that is, the number of leaves. For a given α , the subtree that minimizes R_α is the one that is chosen. Thus, for a small α , a tree that is large with many leaves will minimize R_α ; for a large α , a small tree (which has been pruned) will minimize R_α . A difficulty in this approach is to select the appropriate complexity parameter, α . If α is chosen too large, the tree is pruned to excess and will not adequately model the class II loci, whereas if α is selected to be too small, there will be too many HLA groups to be able to estimate corresponding parameters in the conditional logistic regression models.

Two different measures were used to select a tree model of the HLA class II loci, *HLA-DRB1* and *HLA-DQB1*. One is the Akaike information criteria (AIC) and the other the Bayesian Information Criteria (BIC). The AIC is an asymptotic approximation to cross-validation and given by $AIC = 2S - D$, where D is the deviance (and equals twice the log likelihood). The BIC is a less conservative measure of the cost of the tree and is given by, $BIC = S \ln(n) - D$, where n is the number of observations. For both methods, the deviance was calculated for the subtree corresponding to a given α . The AIC and the BIC were calculated for a range of complexity parameters, $0 \leq \alpha \leq 1$, the subtree which minimized AIC was chosen and used in models to test for class II-independent effects. Likewise, the tree corresponding to the minimum in the BIC was also used to test for HLA class II-independent effects.

Testing for Associations Conditional on HLA-DRB1 and HLA-DQB1

We specifically wished to test the hypothesis that loci within the MHC were associated with T1D independently of the highly associated class II genes *HLA-DRB1* and *HLA-DQB1*. Owing to the established complex non-multiplicative relationship between the alleles of these two genes, extensive LD and epistatic interaction effects [2], we concluded that a joint model was required to explain the observed association. This approach was justified because both loci were necessary to partition the data within *rpart*.

Forward stepwise conditional logistic regression was used to test whether any of the 1541 loci typed in the MHC had an effect in addition to the HLA class II *DRB1/DQB1* effect [18]. Only individuals typed at both the class II loci and the test locus were used for the stepwise analysis, at most 2238 families. The *HLA-DRB1/HLA-DQB1* loci (modelled using the recursive partitioning method described above) were placed in the regression model as confounders and other loci added; whether or not a non-*HLA-DRB1/HLA-DQB1* locus improved on the model was tested by a Wald test because robust variance estimates were applied. The non-*HLA-DRB1/HLA-DQB1* loci were modelled as alleles when the multiplicative model was appropriate, and genotypes otherwise. Nine SNPs were found to be associated but on examination of the signal clouds had poor clustering and so were excluded from all analyses and figures (see the Genotype Calling of the Methods section for the rs numbers).

Testing Age-at-diagnosis Effects

Age at diagnosis of *HLA-B*39* was tested using the cases. Regression was used with *HLA-B*39* as the outcome variable and age-at-diagnosis as the independent variable. Non-independence of family members was accounted for by using Huber/White sandwich estimators.

Results

Single Locus

Unconditional single locus analysis of all 1534 loci that passed QC (eight classical loci and 1526 SNPs), revealed, as expected the two most associated loci, *HLA-DRB1* and *HLA-DQB1* ($P = 10^{-274}$ and $P = 10^{-282}$ respectively) (figure 1). There were a number of other loci also under this peak of association, the *HLA-DQA1* gene at 32.7 Mb ($P = 10^{-224}$) and the two most associated SNPs, rs2395533 and rs660895 with $P = 10^{-112}$ and $P = 10^{-107}$, respectively. There were, as expected from previous results [12], two other peaks of association, one around *HLA-B* at 31.4 Mb ($P = 10^{-93}$) and a second around the *HLA-A* gene at 30.0 Mb ($P = 10^{-22}$). Not unexpectedly, we found that a multiplicative model for the effect of the alleles was inappropriate in the class II region [12] (figure 2). Consequently, caution is required when using conventional transmission disequilibrium test (TDT)-like statistics that assume a multiplicative model, in the class II MHC region, as they are likely to be inappropriate for modelling T1D associations.

Evaluation of the HLA-DRB1 and HLA-DQB1 Model

A tree model of *HLA-DRB1* and *HLA-DQB1* genotypes that consisted of 200 terminal leaves (i.e., HLA class II groups) was produced using recursive partitioning. This number of groups is too many for parameter estimation in the conditional logistic regression model, so the tree was pruned. The appropriateness of two pruned trees to the modelling of the HLA class II loci was assessed. One tree had 17 terminal leaves and corresponded to the minimum BIC. The other, more conservatively pruned tree had 50 terminal leaves, corresponding to the minimum AIC. Both the BIC and AIC trees are subtrees of the full tree model of 200 terminal leaves; the BIC tree is also a subtree of the 50 terminal leaves AIC tree. Figure 3 represents the BIC tree with the relative risks (RR) and the 95% confidence intervals (CI). The vertical spacing is proportional to the error in the fit of the tree and so is a measure of impurity (i.e., how many individuals are misclassified). The first split of the root node has the greatest reduction in error, with susceptible HLA genotypes being put in the right-hand branch and protective genotypes in the left-hand branch. The horizontal axis can be thought of as an axis of T1D risk, with the most protective genotypes forming the groups on the left and the most susceptible HLA genotypes forming groups on the right.

The AIC tree consists of 50 groups, but nine were composed exclusively of pseudo-controls and, therefore, could not be included for model estimation. These 1634 pseudo-controls (of 10515) were dropped from the analysis causing a reduction in sample size, with a potential loss of power. The BIC tree had 17 groups (terminal leaves) only one of which had to be dropped as it consisted exclusively of 160 pseudo-controls. Thus, the BIC tree is much simpler than the AIC tree as it consists of 16 HLA genotype groups compared with the AIC tree's 41 HLA genotype groups. Both trees led to a consistent interpretation of the data despite obtaining different p-values for the same loci (table 1 and table 2). Hence, for figure 4–figure 7, only the results using the simpler BIC tree in the conditional logistic regression are given.

Association Testing Conditional on HLA-DRB1 and HLA-DQB1

Initial analyses considered two models of the MHC class II genes *HLA-DRB1* and *HLA-DQB1*. Both models are based on the same recursive partitioning model but correspond to different levels of pruning.

The most associated locus after conditioning on *HLA-DRB1* and *HLA-DQB1* was *HLA-B*, $P_{\text{BIC}} = 6.02 \times 10^{-17}$ and $P_{\text{AIC}} = 1.99 \times 10^{-15}$ (the subscripts on the P values indicate whether the BIC tree was used to model the class II genes or the more complex AIC tree; figure 4 and table 1). There was also a peak of association at *HLA-A* $P_{\text{BIC}} = 8.81 \times 10^{-13}$, $P_{\text{AIC}} = 2.84 \times$

110^{-10} , and a second peak in the MHC class II region that included *HLA-DPB1* $P_{\text{BIC}} = 5.10 \times 10^{-8}$, $P_{\text{AIC}} = 2.07 \times 10^{-8}$ (table 2). Table 1 shows the RR for the alleles of *HLA-B* conditional on *HLA-DRB1* and *HLA-DQB1*. *HLA-B*39* is the most predisposing *HLA-B* allele, with a RR (95% CI) = 3.42 (2.47–4.73) using *HLA-B*08* as reference. *HLA-B*13*, *HLA-B*50* and *HLA-B*18* are also predisposing to T1D independently of class II. *HLA-B*38* was protective for T1D, while the *HLA-B*27* protection observed in [13] did not reach significance in these T1DGC families.

Next, the associations of the remaining loci were tested conditional on *HLA-DRB1*, *HLA-DQB1* and *HLA-B*, to establish whether the peaks of association in the *HLA-A* and *HLA-DPB1* regions were attributable to LD with *HLA-B*. Both the *HLA-B* alleles (all alleles at frequency >0.001) and the tree model of the class II loci, *HLA-DRB1* and *HLA-DQB1*, were included in the regression model and the test locus added to test for additional independent effects. Evidence for independent effects of *HLA-A* at 30.0 Mb and a peak around *HLA-DPB1* at 33.2 Mb was obtained (figure 5). The RR of the *HLA-A* alleles show that *HLA-A*24* is the most susceptible *HLA-A* allele (table 3), consistent with our earlier work [12]. The most associated loci were the intergenic SNPs, rs439121, $P_{\text{BIC}} = 9.15 \times 10^{-8}$ and rs421446, $P_{\text{BIC}} = 8.87 \times 10^{-7}$ and *HLA-A*, $P_{\text{BIC}} = 8.93 \times 10^{-7}$. These SNPs at 88 and 131 kb centromeric of *HLA-DPB1*, were in LD with $D' = 0.8$ and $r^2 = 0.7$.

Subsequently, rs439121, the alleles of *HLA-B* and the tree model of *HLA-DRB1* and *HLA-DQB1* were included in the conditional logistic regression model and the association of the remaining loci tested. The association peak at *HLA-A* remained convincing, $P_{\text{BIC}} = 6.48 \times 10^{-6}$ (figure 6) and included a SNP ~100 kb telomeric of *HLA-A*, rs1619379, $P = 3.67 \times 10^{-6}$. There still remained an association at the peak containing the *HLA-DPB1* gene, with the most associated locus being rs6457721, $P_{\text{BIC}} = 1.54 \times 10^{-5}$. Finally, the remaining loci were tested for association with T1D independently of *HLA-DRB1*, *HLA-DQB1*, *HLA-B*, rs439121 and *HLA-A*. No convincing evidence of association was obtained ($P > 10^{-4}$; figure 7).

Discussion

The analyses presented here, combined with previous reports in the literature [3,12], clearly demonstrate that *HLA-B* is involved directly in the genetics and aetiology of T1D. In particular, the *HLA-B*39* allele increases T1D susceptibility at a younger age-at-diagnosis. The frequency of *HLA-B*39* was elevated in those under 5.5 years at 0.70 (corresponding to the lowest 25 percentile of the age-at-diagnosis distribution) compared with those over 5.5 years in which the frequency was 0.36. A single copy of the allele was found to lower the average age at diagnosis by 1.7 years from 11.8 years for individuals with zero copies of *HLA-B*39* to 10.1 years for individuals with one copy of *HLA-B*39*, in agreement with earlier findings [12]. This age-at-diagnosis effect was independent of the *HLA-DRB1* and *HLA-DQB1* genotypes ($P = 7.57 \times 10^{-6}$). In contrast to the susceptibility conferred by *HLA-B*39*, the rare *HLA-B*38* allele (frequency equal to 0.008 in the British population [12]) conferred the most protection from T1D. This allele was also found to be the most protective in our earlier work [12]. Comparison of the amino acid sequence of these two alleles with opposite effects (<http://www.anthonynolan.org.uk/research/hlainformaticsgroup/seq/hla-b-data.html>), *HLA-B*390601* and *HLA-B*380101* [21], differ at eight amino acids (positions 99, 102, 105–108, 120 and 122). Association analysis of the amino acids in *HLA-B* may provide insights for future directions in elucidating the role of *HLA-B* in T1D risk.

We have also identified two regions of association, one including *HLA-A* in agreement with our earlier work [12] and one including *HLA-DPB1*. These genes, however, may not be responsible for these associations. *HLA-A* was not the most associated locus, in contrast to the singular association of *HLA-B*, although there was very little difference in disease association

significance between *HLA-A* and the more associated neighbouring SNP, rs1619379. Addition of rs1619379 to the model with *HLA-A*, *HLA-B*, *HLA-DRB1*, *HLA-DQB1* and rs439121 was not significant, whereas addition of the *HLA-A**24 and *HLA-A**02 alleles to the model with rs1619379, *HLA-B*, *HLA-DRB1*, *HLA-DQB1* and rs439121 was nearing significance ($P = 0.0001$). *HLA-A**24 and *HLA-A**02 were chosen as they were the most associated alleles at *HLA-A* (table 3) and because we wished to minimize the number of parameters in the model. Thus, *HLA-A* is more likely to be responsible for this association.

It should be noted, however, that we are reaching the limits of statistical power, so further work using larger sample sizes is required to unravel the T1D association in this region. In contrast, the peak that includes *HLA-DPB1* actually has two intergenic SNPs at its peak and when either SNP is included in the analysis some association remains ($P = 10^{-5}$). However, there are five confounders, *HLA-DRB1*, *HLA-DQB1*, *HLA-B*, rs439121 and *HLA-A*, included in the regression model and given that many hypotheses have been tested, the statistical power is again reduced and the likelihood of a false positive is elevated. So while this result implies that these intergenic SNPs are unlikely to be causal, and instead are markers for the causal variant, we treat the result with caution given the statistical constraints. Consequently, to distinguish the associations around the *HLA-A* and *HLA-DPB1* genes that are independent of *HLA-DRB1*, *HLA-DQB1* and *HLA-B*, an even denser map of markers, as well as the classical loci, will need to be genotyped in a larger collection of several thousand cases and controls or trio families.

The SNP, rs9268831, close to *HLA-DRA* that has been reported to be associated with T1D independently of *HLA-DRB1*, *HLA-DQB1* and *HLA-B* [12], was not genotyped in these T1DGC families nor was it in LD with rs439121, the most associated SNP in the class II region once *HLA-DRB1* and *HLA-DQB1* effects have been removed, $D' = 0.04$ in the European Caucasian (CEU) families from HapMap release 21, nor was it in LD with rs6457721, $D' = 0.13$, $r^2 = 0.003$, which was associated with T1D once all the independent associations were included in the logistic regression model. Thus, rs9268831 needs to be genotyped in these T1DGC families and the possibility of association in the class II region itself independent of all five independent T1D associations identified in the MHC region so far, requires further investigation.

The work presented here clearly highlights the importance of using large sample collections that are well powered if true class II-independent effects are to be found. The recently reported association of rs1233478 at 29.6 Mb in the *UBD* region in a subset of these samples [22], was not replicated in the full data set once the effects of *HLA-DRB1*, *HLA-DQB1* were included ($P = 0.53$) or when *HLA-B* was included in the model ($P = 0.09$). The effect of *UBD* was significantly associated if the class II loci were not included in the model ($P = 2.7 \times 10^{-19}$). Similarly, haplotypes of the known associations (*HLA-A*, *HLA-B*, *HLA-DRB1*, *HLA-DQB1*) were constructed and the additional independent association of rs1233478 tested and found not to be associated ($P = 0.83$).

Aly *et al.* [22] analysed 1240 T1DGC families at rs1233478 conditional on *HLA-DRB1* and *HLA-DQB1*. However, the conditioning was insufficient as they only include the 13 alleles of *HLA-DRB1* and 11 alleles of *HLA-DQB1*, in a multiplicative effects model, a model we have shown (here and [12]) to be an unsuitable approximation. Despite this, they do eliminate rs1233478 as a candidate going from an unconditional OR = 2.0, $P = 1.6 \times 10^{-23}$ to OR = 1.3, $P = 0.01$ after conditioning, which is unconvincing in the circumstances. The SNP, rs1233478, was associated in the full 2240 T1DGC family data set in the absence of the *HLA-B**08:*HLA-A**01 haplotype, consistent with Aly *et al.* [22]. However, by removing individuals carrying this haplotype from the data set, a number of other highly predisposing alleles remain, including *HLA-DRB1**03 and *HLA-DRB1**0401, and it is likely the association of this SNP is

attributable to genotypes formed from these alleles. The approach of removing a susceptible haplotype and reducing the data set size is unsatisfactory and can lead to misinterpretation of results. This reinforces the necessity to use a model that fully accounts for the class II effects in a large data set to test for T1D associations that are independent of the strong class I and II associations.

Similarly, the association of *HLA-DQA1* with T1D can be attributed to the *HLA-DRB1* and *HLA-DQB1* genes as there was no association at this locus once these genes were included in the model. The association observed at *HLA-C* was also attributable to the class II genes in this T1D data set, consistent with our previous results [12]. The SNP rs2296336 in the *ITPR3* gene region has been reported to be associated with T1D [10]. The SNP rs2296343, which is in complete LD with rs2296336 ($r^2 = 1.0$ in the CEU families from HapMap release 21) was genotyped in these families. No evidence of association of rs2296343 was obtained ($P = 0.02$) without conditioning on *HLA-DRB1* and *HLA-DQB1* and ($P = 0.26$) after conditioning on the class II loci. This finding is consistent with our previous null results for *ITPR3* [12].

The analyses presented here, while extensive, remain preliminary as further work is required to elucidate all the T1D associations in this 4 Mb region of chromosome 6. The influence of phase on the associations confirmed in this report is in need of additional investigation as well as any haplotype-specific effects tested. The use of recursive partitioning to model the multidimensional confounding because of *HLA-DRB1* and *HLA-DQB1* has been successful, producing consistent results between multiple family and case-control data sets. Despite not being included in this preliminary report, haplotypes of the class II and I loci can be included in the recursive partitioning procedure to allow for phase if it is necessary. Recursive partitioning easily lends itself to the investigation of MHC associations in other diseases such as Graves' disease [23], or any other region in which there is a large multidimensional confounder. Similarly, it can also be applied when the confounding is attributable to one or more phenotypes, which have to be included in the statistical model as covariates. A recursive partitioning approach that includes all covariates could be constructed and used to model the confounding.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Juvenile Diabetes Research Foundation International, the Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Centre fund the research. The Cambridge Institute for Medical Research is in receipt of a Wellcome Trust Strategic Award (079895). This research uses resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. Ranganath Bangalor Venkatesh at the WTSI efforts in assembling the raw intensity data for OPA1 and OPA2 are gratefully acknowledged.

References

1. Todd JA, Bell JI, McDevitt HO. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* 1987;329:599–604. [PubMed: 3309680]
2. Cucca F, Lampis R, Congia M, et al. A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. *Hum Mol Genet* 2001;10:2025–2037. [PubMed: 11590120]

3. Valdes AM, Erlich HA, Noble JA. Human leukocyte antigen class I B and C loci contribute to Type 1 Diabetes (T1D) susceptibility and age at T1D onset. *Hum Immunol* 2005;66:301–313. [PubMed: 15784469]
4. Valdes AM, Noble JA, Genin E, Clerget-Darpoux F, Erlich HA, Thomson G. Modeling of HLA class II susceptibility to Type I diabetes reveals an effect associated with DPB1. *Genet Epidemiol* 2001;21:212–223. [PubMed: 11668578]
5. 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* 2005;65:115–119. [PubMed: 15663750]
6. Noble J, Valdes A, Apple R, Bugawan T, Thomson G, Erlich H. HLA loci other than DR and DQ can influence susceptibility to type 1 diabetes: analysis of DPB1 and HLA-A in 269 Caucasian, multiplex families. *Diabetes* 1998;47:A395.
7. 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. [PubMed: 10615959]
8. Aly TA, Baschal EE, Jahromi MM, et al. High density SNP analysis of the MHC region reveals multiple loci for type 1A diabetes. *Clin Immunol* 2007;123:S133.
9. Roach JC, Deutsch K, Li S, et al. Genetic mapping at 3-kilobase resolution reveals inositol 1,4,5-triphosphate receptor 3 as a risk factor for type 1 diabetes in Sweden. *Am J Hum Genet* 2006;79:614–627. [PubMed: 16960798]
10. de Bakker PI, McVean G, Sabeti PC, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 2006;38:1166–1172. [PubMed: 16998491]
11. Traherne JA, Horton R, Roberts AN, et al. Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLoS Genet* 2006;2:e9. [PubMed: 16440057]
12. Nejentsev S, Howson JMM, Walker NM, et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* 2007;450:887–892. [PubMed: 18004301]
13. R Development Core Team. Version 2.2.1. 2.3.1 edn. Vienna: R Foundation for statistical computing; 2006. R: A Language and Environment for Statistical Computing.
14. Teo YY, Inouye M, Small KS, et al. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* 2007;23:2741–2746. [PubMed: 17846035]
15. Clayton DG, Walker NM, Smyth DJ, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* 2005;37:1243–1246. [PubMed: 16228001]
16. Wellcome Trust Case Control Consortium. Genomewide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–678. [PubMed: 17554300]
17. Hulbert EM, Smink LJ, Adlem EC, et al. T1DBase: integration and presentation of complex data for type 1 diabetes research. *Nucleic Acids Res* 2007;35:D742–D746. [PubMed: 17169983]
18. Cordell HJ, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 2002;70:124–141. [PubMed: 11719900]
19. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Chapman and Hall. 1984
20. Therneau TM, Atkinson EJ. *An Introduction to Recursive Partitioning Using the RPART Routine*. Mayo Clinic, Division of Biostatistics. Available from URL: <http://mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm>
21. Robinson J, Waller MJ, Parham P, et al. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res* 2003;31:311–314. [PubMed: 12520010]
22. Aly TA, Baschal EE, Jahromi MM, et al. Analysis of single nucleotide polymorphisms identifies major type 1A diabetes locus telomeric of the major histocompatibility complex. *Diabetes* 2008;57:770–776. [PubMed: 18065518]
23. Simmonds MJ, Howson JM, Heward JM, et al. A novel and major association of HLA-C in Graves' disease that eclipses the classical HLA-DRB1 effect. *Hum Mol Genet* 2007;16:2149–2153. [PubMed: 17597093]

Appendix

```

user-splits.R      evaluate <- function (y, wt, parms) {      #
Sets involved, whether as case, and current rr      ntot <- nrow
(ccdata)      at.node <-rep (0, ntot)      at node [y] <- 1
cc <- ccdata$cc      set <- ccdata$set      # Suppress singularity
warnings      warn = options () $warn      options (warn = -1)
if (is.null (mcc.rpart.design))      clfit <- clogit (cc ~ -1 + at.node
+ strata (set))      else      clfit <- clogit (cc ~ -1 +
mcc.rpart.design + at.node +      strata (set))      aliased <-is.na
(clfit$coefficients)      rank <-sum (!aliased)      if (!aliased[length
(aliased)])      mcc.rpart.design <- cbind (mcc.rpart.design,
at.node)      options (warn = warn)      mcc.rpart.work[, "RR"] <- exp
(clfit$linear.predictor) [,1]      mcc.rpart.work[, "Resid"] <- residuals
(clfit, type="martingale")      dev <- sum ((residuals(clfit,
type="deviance")) [y] ^2)      RR <- mcc.rpart.work [y[1], "RR"]
lab <- c(RR, length(y), sum (cc[y]))      cat ("RR = ", lab[1], ", N = ",
lab[2], ", # cases = ", lab[3],      "Deviance = ", dev, "rank = ",
rank, "\n")      list (label=lab, deviance = dev)      }      split
<- function(y, wt, x, parms, continuous) {      if (continuous)
{      stop ("Continuous predictor not yet implemented")      }
ux <- sort (unique (x))      res <- mcc.rpart.work [y, "Resud"]
cc <- ccdata$cc[y]      p <- cc - res      u <- tapply (res, x,
sum)      v <- tapply (p* (1-p), x, sum)      nx <- length (ux)
ord <- order(u/v)      u.left <- cumsum(u[ord]) [-nx]      u.right <-
sum (u) - u.left      v.left <- cumsum(v[ord]) [-nx]      v.right <-
sum (v) - v.left      good <- u.left^2/v.left + u.right^2/v.right
list (goodness = good, direction = ux [ord])      }      init <- function
(y, offset, parms wt) {      N = nrow (ccdata)      if (!is.null
(offset))      stop ("offset not relevent")      if (!is.null (wt) && var (wt) !=
0) {      stop ("weights are not used")      }      N <- nrow (ccdata)
mcc.rpart.work <- matrix (nrow=N, ncol =2,
dimnames=list (NULL, c ("RR", "Resid")))      mcc.rpart.design <- NULL
initSum = function (yval, dev, wt, ylevel, digits) {      paste (" RR = ",
format (signif(yval[,1], digits)),      ", N = ", yval [,2],
", # cases = ", yval[,3], ", Deviance = ", dev, sep="")      }      list (y=matrix
(1:N,ncol=1), parms=NA, numresp=3, numy=1, summary=initSum)      } fcns <- list
(eval=evaluate, split=split, init=init)      tree <- rpart (cc ~ dq + dr, data =
ccdata,      control=rpart.control (minbucket =10, cp=-0.01, xval
=0),      method=fcns)      summary (tree)      cp.try <- c(-0.01, 0, 0.0001,
0.0002, 0.0005, 0.001, 0.002, 0.005,      0.01, 0.02, 0.05, 0.1, 0.2,
0.5, 1)      prune.results <- matrix (nrow =length (cp.try), ncol=4,
dimnames=list (as.character (cp.try), c("chi.square", "df", "AIC", "BIC")))
aic.min <- NA      bic.min <- NA      for (cp in cp.try) {      tree.cp <- prune (tree,
cp=cp)      group <- factor(tree, cp$where)      if (nlevels (group)<1) {      fit
<- clogit (ccdata$cc ~ group + strata (ccdata$set))      N <- fit$n      chi2
<- 2* (fit$loglik[2]-fit$loglik[1])      df <- sum (!is.na (fit
$coefficients))      aic <- -2*fit$loglik[2] + 2*df      bic <- -2*fit$loglik
[2] + df*log (N)      prune.results[as.character(cp), 1] <- chi2
prune.results[as.character(cp), 2] <- df      prune.results[as.character(cp),
3] <- aic      prune.results[as.character(cp), 4] <- bic      if (is.na
(aic.min) || aic<aic.min) {      aic.min <- aic      tree.use <-

```

```
tree.cp      }      if (is.na (bic.min) || bic<bic.min) {      bic. min <-  
bic      tree.bic.use <- tree.cp      }  }}print (prune.results)summary  
(tree.use)summary (tree.bic.use)
```

Fig. 1. Association plot of all 1534 loci typed between 29 and 34 Mb of the major histocompatibility complex in up to 2240 affected sibpair families.

Fig. 2.

Test for dominance at 1526 SNPs spanning 4.6 Mb of the major histocompatibility complex. Note in the class II region that there are strong non-multiplicative effects, so by using TDT or other association tests that only model the alleles of the class II loci, *HLA-DRB1* and *HLA-DQB1*, the effects of class II will be incorrectly modelled and statistical power lost.

Fig. 3.

The Bayesian information criteria (BIC) tree with the relative risks and corresponding 95% confidence intervals for each of the terminal leaves (HLA groups) using a neutral group as reference. The horizontal axis can be thought of as an axis of type 1 diabetes risk, and the vertical spacing is proportional to the error in the trees fit. NA is the group that is pure because it consists of pseudo-controls only.

Fig. 4. Conditional association analysis of all 1532 genotyped loci conditional on the major effects of *HLA-DRB1* and *HLADQB1* using Bayesian information criteria tree model.

Fig. 5. Association analysis of all 1531 genotyped loci conditional on *HLA-B* as well as the Bayesian information criteria tree model of *HLA-DRB1* and *HLA-DQB1*.

Fig. 6. Association analysis of 1530 loci conditional on rs439121 at 33.2 Mb, *HLA-B* and the Bayesian information criteria tree model of *HLA-DRB1* and *HLA-DQB1*.

Fig. 7. Association analysis of 1529 loci conditional on *HLA-A*, rs439121 at 33.2 Mb, *HLA-B* and the Bayesian information criteria (BIC) tree model of *HLA-DRB1* and *HLA-DQB1*.

Table 1

Relative risks of the *HLA-B* alleles (>0.01 frequency) without and with conditioning on *HLA-DRB1* and *HLA-DQB1* genotypes

<i>HLA-B</i> allele	Frequency in affected offspring, N (%)	Frequency in parents, N (%)	Unconditional RR (95% CI)	RR (95% CI) conditional on <i>HLA-DRB1</i> / <i>HLA-DQB1</i> (BIC)	RR (95% CI) conditional on <i>HLA-DRB1</i> / <i>HLA-DQB1</i> (AIC)	OR (95% CI) conditional on <i>HLA-DRB1</i> / <i>HLA-DQB1</i> from [1]
<i>HLA-B*39</i>	417 (4.31)	210 (2.93)	1.44 (1.08–1.93)	3.42 (2.47–4.73)	3.56 (2.53–5.01)	2.53 (1.57–4.07)
<i>HLA-B*13</i>	153 (1.58)	140 (1.95)	0.41 (0.30–0.56)	1.83 (1.30–2.58)	1.94 (1.39–2.73)	2.03 (1.03–4.03)
<i>HLA-B*50</i>	171 (1.77)	115 (1.60)	0.78 (0.55–1.11)	1.83 (1.27–2.63)	1.82 (1.25–2.64)	2.31 (1.10–4.85)
<i>HLA-B*18</i>	946 (9.76)	580 (8.09)	1.01 (0.84–1.23)	1.47 (1.20–1.80)	1.43 (1.16–1.77)	1.92 (1.27–2.90)
<i>HLA-B*49</i>	199 (2.06)	130 (1.81)	0.77 (0.55–1.09)	1.24 (0.87–1.77)	1.24 (0.86–1.80)	3.23 (1.39–7.53)
<i>HLA-B*07</i>	775 (8.01)	746 (10.40)	0.38 (0.33–0.46)	1.21 (1.00–1.46)	1.19 (0.98–1.46)	1.32 (0.94–1.84)
<i>HLA-B*55</i>	92 (0.95)	98 (1.37)	0.41 (0.28–0.58)	1.20 (0.77–1.87)	1.22 (0.77–1.92)	1.35 (0.68–2.67)
<i>HLA-B*51</i>	300 (3.10)	246 (3.43)	0.50 (0.39–0.63)	1.12 (0.85–1.48)	1.32 (0.99–1.76)	0.98 (0.57–1.68)
<i>HLA-B*40</i>	729 (7.54)	512 (7.14)	0.64 (0.54–0.77)	1.12 (0.92–1.37)	1.11 (0.90–1.37)	0.91 (0.66–1.25)
<i>HLA-B*15</i>	1272 (13.15)	739 (10.30)	1.02 (0.87–1.23)	1.07 (0.89–1.29)	1.00 (0.82–1.20)	1.03 (0.76–1.39)
<i>HLA-B*08</i>	2084 (21.54)	1258 (17.54)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
<i>HLA-B*14</i>	181 (1.87)	185 (2.58)	0.33 (0.25–0.44)	0.97 (0.73–1.30)	1.00 (0.74–1.35)	0.70 (0.42–1.16)
<i>HLA-B*35</i>	551 (5.70)	497 (6.93)	0.43 (0.36–0.52)	0.94 (0.75–1.18)	0.98 (0.78–1.28)	0.86 (0.58–1.28)
<i>HLA-B*27</i>	321 (3.32)	259 (3.61)	0.53 (0.42–0.66)	0.92 (0.70–1.19)	0.89 (0.68–1.15)	0.55 (0.36–0.82)
<i>HLA-B*44</i>	868 (8.97)	819 (11.42)	0.36 (0.31–0.42)	0.91 (0.76–1.09)	0.91 (0.76–1.11)	1.05 (0.79–1.39)
<i>HLA-B*57</i>	74 (0.76)	157 (2.19)	0.10 (0.06–0.15)	0.60 (0.37–0.98)	0.62 (0.38–1.01)	0.53 (0.27–1.04)
<i>HLA-B*38</i>	132 (1.36)	115 (1.60)	0.35 (0.24–0.51)	0.43 (0.29–0.65)	0.50 (0.33–0.77)	0.51 (0.14–1.77)

AIC, Akaike information criteria; BIC, Bayesian information criteria; CI, confidence interval; OR, odds ratio; RR, relative risks.

*HLA-B*08* is used as reference. Conditioned RR with corresponding 95% CI are presented using both the AIC tree and the BIC tree, which give results consistent with Nejentsev *et al.* in [1]. The alleles that are significantly protective or susceptible in both studies are highlighted in bold.

* Parents who were known to have type 1 diabetes were removed from the frequency calculations.

Table 2

Tests for HLA class II-independent associations using the BIC tree and the AIC tree to model the effects of *HLA-DRB1* and *HLA-DQB1*

Locus	P_{AIC}	AIC order	P_{BIC}	BIC order	Start position/bp
<i>HLA-B</i>	1.99×10^{-15}	1	6.02×10^{-17}	1	31 429 628
<i>HLA-A</i>	2.84×10^{-10}	2	8.81×10^{-13}	2	30 041 270
rs1619379	4.05×10^{-7}	7	3.09×10^{-8}	3	29 893 214
<i>HLA-DPB1</i>	5.10×10^{-8}	4	5.10×10^{-8}	4	33 151 694
rs439121	7.52×10^{-8}	5	6.21×10^{-8}	5	33 239 754
rs3130161	3.32×10^{-8}	3	8.82×10^{-8}	6	33 233 836
rs3130695	4.08×10^{-7}	8	2.34×10^{-7}	7	31 319 028
rs6457721	1.29×10^{-6}	9	3.13×10^{-7}	8	33 215 728
rs2281389	1.93×10^{-7}	6	3.31×10^{-7}	9	33 167 774
rs421446	1.31×10^{-6}	10	1.68×10^{-6}	10	33 282 760
rs5024431	9.25×10^{-5}	36	3.24×10^{-6}	11	32 792 322
rs2394186	5.20×10^{-5}	29	3.89×10^{-6}	12	29 924 400
rs2855438	1.65×10^{-6}	11	4.17×10^{-6}	13	33 246 816
rs2294479	2.92×10^{-6}	12	4.88×10^{-6}	14	33 206 368
rs1737010	6.64×10^{-5}	32	1.67×10^{-5}	15	29 864 980
rs1736951	1.07×10^{-4}	40	1.86×10^{-5}	16	29 893 756
rs9277678	5.28×10^{-6}	13	2.06×10^{-5}	17	33 134 244
rs1610640	9.81×10^{-5}	38	2.16×10^{-5}	18	29 866 504
rs213209	9.40×10^{-6}	17	2.18×10^{-5}	19	33 284 936
rs1362070	8.63×10^{-5}	34	2.29×10^{-5}	20	29 850 278

AIC, Akaike information criteria; BIC, Bayesian information criteria; HLA, human leucocyte antigen; RR, relative risks; The association results for the top 20 loci from the BIC model are given.

Table 3

RR of the *HLA-A* alleles (>0.01 frequency) without and with conditioning on *HLA-DRB1*, *HLA-DQB1* and *HLA-B*

<i>HLA-A</i> allele	Frequency in affected offspring, N (%)	Frequency in parents N (%)	Unconditional RR (95% CI)	RR (95% CI) conditional on <i>HLA-DRB1</i> / <i>HLA-DQB1</i> (BIC)	RR (95% CI) conditional on <i>HLA-DRB1</i> / <i>HLA-DQB1</i> (AIC)	OR (95% CI) conditional on <i>HLA-DRB1</i> / <i>HLA-DQB1</i> from [1]
<i>HLA-A*24</i>	977 (10.84)	705 (9.72)	1.32 (1.13–1.54)	1.34 (1.13–1.59)	1.22 (1.02–1.46)	1.54 (1.11–2.11)
<i>HLA-A*33</i>	87 (0.97)	81 (1.12)	0.83 (0.57–1.19)	0.90 (0.62–1.30)	1.02 (0.67–1.55)	0.63 (0.25–1.60)
<i>HLA-A*02</i>	3097 (34.36)	2324 (32.03)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
<i>HLA-A*29</i>	258 (2.86)	231 (3.18)	0.73 (0.58–0.91)	0.84 (0.65–1.10)	0.96 (0.73–1.26)	0.85 (0.53–1.35)
<i>HLA-A*03</i>	1130 (12.54)	955 (13.16)	0.86 (0.75–0.99)	0.90 (0.76–1.06)	0.91 (0.76–1.09)	1.17 (0.87–1.58)
<i>HLA-A*23</i>	102 (1.13)	116 (1.60)	0.55 (0.39–0.77)	0.80 (0.56–1.12)	0.81 (0.56–1.18)	0.51 (0.22–1.17)
<i>HLA-A*26</i>	181 (2.01)	171 (2.36)	0.70 (0.54–0.91)	0.67 (0.50–0.91)	0.79 (0.57–1.08)	0.63 (0.34–1.16)
<i>HLA-A*31</i>	192 (2.13)	165 (2.27)	0.94 (0.72–1.23)	0.72 (0.53–0.96)	0.78 (0.57–1.05)	0.41 (0.24–0.69)
<i>HLA-A*68</i>	249 (2.76)	228 (3.14)	0.73 (0.57–0.94)	0.76 (0.57–1.01)	0.74 (0.55–1.01)	0.84 (0.52–1.36)
<i>HLA-A*30</i>	352 (3.91)	239 (3.29)	1.33 (1.05–1.70)	0.99 (0.76–1.30)	0.73 (0.54–0.99)	0.89 (0.50–1.57)
<i>HLA-A*01</i>	1682 (18.66)	1292 (17.81)	1.09 (0.97–1.23)	0.73 (0.63–0.84)	0.73 (0.60–0.88)	0.58 (0.43–0.79)
<i>HLA-A*11</i>	312 (3.46)	352 (4.85)	0.54 (0.44–0.66)	0.67 (0.52–0.85)	0.71 (0.55–0.91)	0.47 (0.30–0.73)
<i>HLA-A*25</i>	202 (2.24)	162 (2.27)	1.01 (0.76–1.35)	0.82 (0.61–1.10)	0.66 (0.48–0.91)	0.62 (0.31–1.23)
<i>HLA-A*32</i>	173 (1.92)	209 (2.88)	0.52 (0.40–0.68)	0.55 (0.41–0.74)	0.59 (0.43–0.82)	0.64 (0.39–1.04)

AIC, Akaike information criteria; BIC, Bayesian information criteria; CI, confidence interval; HLA, human leucocyte antigen; OR, odds ratio; RR, relative risks.

*HLA-A*02* is used as reference. Conditioned RR with corresponding 95% CI are presented using the BIC tree, which give results consistent with the Nejentsev *et al.* in [1]. Alleles that are significantly protective or susceptible in both studies are highlighted in bold.

* Parents who were known to have type 1 diabetes were removed from the frequency calculations.