Published in final edited form as:

Diabetes Obes Metab. 2009 February ; 11(0 1): 8–16. doi:10.1111/j.1463-1326.2008.00998.x.

Haplotype-based search for SNPs associated with differential type 1 diabetes risk among chromosomes carrying a specific HLA DRB1-DQA1-DQB1 haplotype

R McGinnis¹, W McLaren¹, V Ranganath¹, P Whittaker¹, S Hunt¹, P Deloukas¹, and the T1DGC²

¹Wellcome Trust Sanger Institute, Cambridge, UK

SUMMARY

Aim: To test chromosomes carrying the same DRB1-DQA1-DQB1 haplotype for SNPs in the major histocompatibility complex (MHC) that might mark subgroups of the haplotype with different risks for type 1 diabetes (T1D).

Methods: Chromosomes from T1D children, their parents, and non-diabetic siblings in families of the Type 1 Diabetes Genetics Consortium (T1DGC) were analyzed by two haplotype-based methods: (1) logistic regression analysis restricted to phased chromosomes carrying the same DRB1-DQA1-DQB1 haplotype but differentiated by the two alleles at MHC SNPs which were individually tested for association with T1D; (2) homozygous parent TDT (hpTDT) testing for biased transmission of a SNP allele to diabetic children from parents who are heterozygous at the SNP but homozygous for the specific DRB1-DQA1-DQB1 haplotype being evaluated.

Results: A number of SNPs gave nominally significant ($p < 0.05$) evidence of marking two subsets of the 301-501-201 haplotype that might differ with respect to their diabetogenic potency. However, none of the SNPs achieved experiment-wide significance and hence may be false-positive associations.

Conclusions: We discuss limitations and possible deficiencies of our study suggesting further work which might yield more robust SNP associations marking two subgroups of a DRB1-DQA1-DQB1 haplotype with different T1D risks.

Keywords

type1 diabetes; HLA; MHC; haplotype; SNP; association

INTRODUCTION

The primary genetic determinants of susceptibility to type 1 diabetes (T1D) in Caucasians appear to be haplotypes defined by alleles at the Human Leukocyte Antigen (HLA) class II loci DRB1, DQA1 and DQB1 [1]. In particular, the common DRB1-DQA1-DQB1 haplotypes (0301-0501-0201, 0401-0301-0302, 0402-0301-0302, 0405-0301-0302) account for a major portion of T1D susceptibility conferred by the HLA region, and furthermore other DRB1-DQA1-DQB1 haplotypes are also associated with lesser degrees of T1D

Correspondence to: Ralph McGinnis, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA United Kingdom; email: rm2@sanger.ac.uk.

²www.t1dgc.org

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interests in publishing this article.

susceptibility or resistance [1,2]. However there are a number of important unresolved issues about the role of DRB1-DQA1-DQB1 haplotypes in T1D etiology. One unresolved question is the degree to which susceptibility to type 1 diabetes is altered by additional HLA variants other than those at DRB1, DQA1 and DQB1. Such additional HLA variants might act independently of DRB1-DQA1-DQB1 haplotypes and perhaps be located at a substantial distance from the class II loci. Alternatively, the additional HLA variants might depend on the presence of a specific DRB1-DQA1-DQB1 haplotype, perhaps because the additional variant(s) physically modulate the diabetes risk of the haplotype or simply because the variant(s) “tag” two or more subsets of the haplotype that differ in their diabetogenic potency.

Here we describe our use of the Type 1 Diabetes Genetics Consortium (T1DGC) MHC fine-mapping data to search for evidence of this second type of HLA variant, namely variants whose association with diabetes risk may depend on the presence or activity of a particular DRB1-DQA1-DQB1 haplotype. To search for such variants we pursued two independent strategies which each statistically controlled for T1D risk contributed by the DRB1-DQA1-DQB1 haplotype being considered while simultaneously testing other HLA variants for additional association with T1D beyond that explained by the DRB1-DQA1-DQB1 haplotype. One of these two strategies involved phasing of genotypes across the MHC region into extended haplotypes followed by logistic regression analysis that considered only a single DRB1-DQA1-DQB1 haplotype and tested individual SNPs carried on the haplotype for additional association with T1D. Besides being a first step in this strategy, phasing of MHC chromosomes also provided us with a valuable overview of DRB1-DQA1-DQB1 haplotype frequencies across the T1DGC datasets, enabling population differences and similarities to be noted. Until very recently, it would have been impossible to phase the ~2900 genotyped MHC SNPs into haplotypes for the ~11000 T1DGC samples since the available haplotype-phasing software could not process more than a few hundred subjects with reasonable speed and computer memory. However, the newly released Beagle software can accurately resolve unphased genotypes at thousands of SNPs into haplotypes for datasets of several thousand individuals in a reasonable (3-4 hour) timeframe [3]. By comparison, our attempt to phase chromosomes in one T1DGC cohort using FASTPHASE took on the order of days. We therefore capitalized on Beagle's availability by phasing genotypes at 8 classical HLA genes together with high-density MHC SNP genotypes in each of the T1DGC datasets in order to apply the haplotype regression-based analysis we describe.

Our second strategy to search for HLA variants that mark differential disease risk within a specific DRB1-DQA1-DQB1 haplotype was to apply the “homozygous parent” TDT (hpTDT), a technique that appears to have been first adopted by Li et al [4]. The hpTDT counts transmissions of each marker allele from heterozygous parents to affected offspring, but only includes transmissions from parents who are also homozygous at one or more additional polymorphisms. By requiring both parental chromosomes to be identical (i.e. homozygous) at the additional polymorphisms, the hpTDT removes any influence of these polymorphisms on biasing transmissions at other nearby markers on the same chromosome that are being tested for association with disease. The polymorphisms required to be homozygous are usually known to be associated with disease and the goal of removing their influence is to detect transmission bias emanating from *other undiscovered* polymorphisms in the same region that also alter disease susceptibility. For our application of the hpTDT, we required that DRB1, DQA1, and DQB1 be homozygous for a specific combination of classical HLA alleles and hence SNP alleles being tested by the hpTDT are carried on the DRB1-DQA1-DQB1 haplotype defined by the required combination of homozygous alleles. Consequently, the hpTDT is a haplotype-based analysis which, like the regression analyses we applied, only considers the subset of chromosomes that carry a specific DRB1-DQA1-

DQB1 haplotype and searches within this subset for SNP allele variation that exhibits association with T1D susceptibility. However unlike the regression analysis, the hpTDT haplotype-based analysis does not rely on prior phasing of haplotypes by inferential steps which might sometimes be inaccurate or yield artefacts peculiar to the haplotype-phasing algorithm. Thus, the hpTDT analysis provided an important complementary strategy to our haplotype-based logistic regression analyses.

METHODS

Subjects

We analyzed subjects genotyped for the T1DGC MHC Fine Mapping project and identify the specific cohorts by the abbreviations adopted by the T1DGC (AP=Asia Pacific, BDA=British Diabetic Association, DAN=Danish, EUR=European, HBDI=Human Biological Data Interchange, JOS=Joslin Diabetes Center, NA=North American, SAR=Sardinian, UK=United Kingdom). Due to differences in coding of classical HLA genes in the BDA cohort compared to the other cohorts, BDA subjects were omitted from some of the analyses. The samples and genotype data in the T1DGC MHC fine-mapping study are described in greater detail in [5].

Phasing of MHC chromosomes

Phasing of chromosomes genotyped at 2919 MHC SNPs and at 8 classical HLA Class I and II genes was executed in Beagle software [3]. We phased chromosomes across the MHC region separately for parents and children, and also separately for each T1DGC cohort. Beagle phasing was blind to affection status of subjects. These steps produced two phased MHC chromosomes for each T1DGC subject defined by alleles at the 2919 SNPs and 8 classical HLA genes. Combined phasing of the 2919 MHC SNPs and of the 8 classical HLA loci was anticipated to provide enhanced accuracy in the phased haplotypes since Beagle was able to incorporate information from the strong linkage disequilibrium which extends across the ~4.5 Mb region covered by the MHC SNPs [6].

Haplotype-based logistic regression analyses

Our logistic regression analyses aimed to test individual SNPs for association with T1D which might be due to: (1) altering the diabetes risk of a specific DRB1-DQA1-DQB1 haplotype or (2) marking a specific DRB1-DQA1-DQB1 haplotype for two subgroups with different diabetes risk. We therefore conducted logistic regression analyses for each common DRB1-DQA1-DQB1 haplotype that only considered the subset of phased MHC chromosomes carrying the DRB1-DQA1-DQB1 haplotype being evaluated and which tested individual SNPs carried on the haplotype for association with diabetes. Phased chromosomes were designated as “affected” (diabetic) or “unaffected/control” and coded as 1 or 0, respectively, to form the dependent variable of the regression equation. There were two predictor variables in the equation: (1) T1DGC cohort (coded as a factor) so that SNP allele frequencies for diabetic and control chromosomes were compared within cohort; and (2) an indicator variable to code for the presence (or absence) on the haplotype of a particular allele at the SNP being tested for association with diabetes. Note that this regression analysis treats the two haplotypes in each subject as independent observations analogous to standard *allele* tests for disease association by the Pearson chi-square or Armitage Trend statistics and hence no genetic model (recessive, dominant, etc.) is specified. Such tests of allelic rather than genotypic association with disease are often conducted in genetic investigations [7] and have been shown to be both valid and more powerful than genotypic tests for many disease models (for example, see [8,9]). Each SNP's association with diabetes was evaluated in the regression analysis by the standard likelihood ratio test which is distributed as chi-square with one degree of freedom. Logistic regression

was carried out in R software (<http://www.r-project.org/>) with separate analyses being conducted for the following six common DRB1-DQA1-DQB1 haplotypes: 0401-0301-0302, 0405-0301-0302, 0701-0201-0303, 0301-0501-0201, 0101-0101-0501, 0701-0201-0201. Two relatively rare, diabetes-associated haplotypes (0801-0401-0402, 01601-0102-0502) were not evaluated due to their small sample size and likely limited statistical power. Prior to the regression analyses, we removed from the 11,279 total subjects approximately 11% who were missing genotypes at DRB1, DQA1, or DQB1.

In our initial efforts to apply logistic regression to Beagle-phased chromosomes, we assigned “affected” status only to parental chromosomes that were transmitted to *all* diabetic children in a family and “unaffected” status to the alternate parental chromosome not transmitted to any diabetic children in the family. We also explored a simpler approach in which, for each diabetic child, a parent's two chromosomes were tabulated as affected or unaffected depending on which was transmitted to the diabetic child. However, for many parental chromosomes (represented by extended, possibly recombinant, haplotypes), there were difficulties in unambiguously determining which chromosome had been transmitted to a diabetic offspring which resulted in substantial missing data whose missingness might not be random. We therefore ultimately settled on two other approaches to the haplotype-based regression analysis that did not require determination of which parental chromosome had been transmitted to a diabetic child. One of these analyses was to select one affected and one unaffected sibling from each T1DGC family that had an unaffected child and designate the two chromosomes in the diabetic sib as “affected” and those in the non-diabetic as “unaffected”. This regression analysis considered the 824 pairs of diabetic and non-diabetic sibs which were available in the T1DGC families. The second analysis was to select, from each T1DGC family, one diabetic child not used in the first regression analysis and to designate their chromosomes as “affected” while chromosomes from the child's parents were designated as “unaffected”. This second regression analysis was based on a sample set of 1387 diabetic children and their 2774 parents. Since these two regression analyses evaluated different subjects, we regard them as providing independent tests of whether a particular SNP is associated with diabetes when carried on chromosomes with a specific DRB1-DQA1-DQB1 haplotype.

Homozygous parent TDT (hpTDT) analyses

We applied a homozygous parent TDT (hpTDT) analysis similar to that described by Lie et al [4] by creating our own software. In our version of hpTDT, we only considered SNP allele transmissions from parents who were homozygous for a specific haplotype at DRB1-DQA1-DQB1. To apply the hpTDT, we first identified parents homozygous for the particular DRB1-DQA1-DQB1 haplotype and among these parents identified the subset heterozygous at the SNP to be evaluated by the TDT for preferential transmission of one of its alleles to diabetic children. For this subset of heterozygous parents, we counted transmissions of the two SNP alleles to diabetic children and calculated the standard TDT chi-square statistic with one degree of freedom [$\chi^2 = (a-b)^2 / (a+b)$ where “a” and “b” are total counts of alleles A and B transmitted from heterozygous A/B parents to affected children].

Some DRB1-DQA1-DQB1 haplotypes are not sufficiently frequent to be found as homozygotes or are not carried as homozygotes by sufficient parents who are also heterozygous at the SNP to be evaluated by the hpTDT. To provide sufficient statistical power and to reduce multiple testing in our hpTDT analyses, we only considered SNPs with at least 10 allele transmissions from heterozygous parents to affected children. Due to total TDT transmissions falling below this threshold, a number of DRB1-DQA1-DQB1 haplotypes were not evaluated and, for other DRB1-DQA1-DQB1 haplotypes, we only considered a subset of the genotyped SNPs. In total, 5288 hpTDT tests were conducted

representing ~30% of the ~17,500 regression tests that could be conducted if the 6 common HLA haplotypes (see above) were tested at all of the 2919 genotyped SNPs.

Combining of the two regression results to achieve an overall p-value for each SNP

To combine chi-square results from the two regression analyses for each SNP and DRB1-DQA1-DQB1 haplotype, we applied the meta-analysis method of Stouffer [10]. In this method, a chi-square value is converted to its corresponding Z-score and appropriate direction of association (positive or negative) followed by summing of the Z-scores and division by the standard deviation to achieve an overall Z-score for the independent components (Z-scores) of a study. Since the Z-scores are independent random variables, the standard deviation of the overall Z-score is the square root of the number of independent study components (hence we divided the summed Z-scores for the two T1DGC regression analyses by the square root of 2). This enabled us to achieve an overall p-value for each SNP which combined the regression results for the two independent analyses.

Further inspection of SNPs showing a strong association signal

SNPs providing highly significant evidence for association were inspected for cluster plots and also assessed in relation to other genotyped SNPs if pairwise linkage disequilibrium with the statistically significant SNP was very high ($r^2 > 0.9$) as determined from online MHC SNP genotypes in Caucasians [11]. For example, SNP rs2040410 was highly significant within the 0301-0501-0201 haplotype when tested by the hpTDT ($p < 4.6 \times 10^{-9}$); but we subsequently concluded this was due to genotyping artefact (see Appendix I). Evidence for genotyping error was partly provided by another genotyped SNP (rs6914950) with a pairwise r^2 value of 0.93 with rs2040410. Following cluster plot inspection, rs2040410 was removed from further analysis. The cluster plots of rs2040410 and of another SNP (rs763026) having an unconfirmed association signal were suggestive of the presence of a nearby copy number variant (CNV) which is consistent with an elevated density of CNVs in this region (<http://projects.tcag.ca/variation/>).

RESULTS

Frequencies of DRB1-DQA1-DQB1 haplotypes in T1DGC parents

For each T1DGC dataset, Table 1 shows the frequencies of DRB1-DQA1-DQB1 haplotypes in T1DGC parents as phased by Beagle software. The table lists all DRB1-DQA1-DQB1 haplotypes with at least 5 counts in the T1DGC parents considered in aggregate, and names of haplotypes known to be predisposing to type 1 diabetes [1,2] are highlighted in black. In general, haplotype frequencies are similar across the datasets except for those in the Sardinia (SAR). Sardinians appear to completely lack the diabetes-predisposing 0401-0301-0302 haplotype which is very frequent ($> 14\%$) in all of the other populations. Furthermore, other diabetes-predisposing haplotypes (0301-0501-0201, 0405-0301-0302, 1601-0102-0502) appear to be much more frequent in Sardinia than elsewhere. Marked differences in DRB1-DQA1-DQB1 haplotype frequencies between Sardinia and mainland Europe have previously been described [12].

Search for SNPs marking heterogeneity within specific DRB1-DQA1-DQB1 haplotypes

As described above, we applied logistic regression to phased chromosomes carrying the *same* DRB1-DQA1-DQB1 haplotype in order to detect MHC SNPs that might exhibit association with type 1 diabetes because the SNP marks two subgroups of the DRB1-DQA1-DQB1 haplotype that have different diabetes risks. For each DRB1-DQA1-DQB1 haplotype that we evaluated, separate regression analyses were applied to two different datasets derived from the T1DGC families: (1) 824 pairs of diabetic and non-diabetic siblings

identified by selecting one discordant sib-pair from each T1DGC family that had an unaffected child; (2) 1387 diabetic children and their 2774 parents identified by selecting, from each T1DGC family, one diabetic child *not* included in dataset (1). The results of regression analyses conducted on these two datasets can be regarded as being independent since the datasets consist of different subjects. Therefore, as a strategy to identify true-positive associations and filter out false positives and artefacts, we decided to focus on SNPs that gave evidence of significant association ($p < 0.05$) in the regression analyses of *both* data sets. Although hpTDT results are not independent of the regression analysis results, we also applied the additional filter of requiring a SNP be statistically significant ($p < 0.05$) in hpTDT analysis of the DRB1-DQA1-DQB1 haplotype in order to cross-check and validate the associations observed in the two regression analyses.

In applying these filters, we found that very few SNPs were statistically significant in both regression analyses conducted for the same DRB1-DQA1-DQB1 haplotype. Of the six common DRB1-DQA1-DQB1 haplotypes we evaluated, three haplotypes (0401-0301-0302, 0405-0301-0302, 0701-0201-0303) did not yield a single SNP with $p < 0.05$ in both regression analyses. For the other three DRB1-DQA1-DQB1 haplotypes, most of the SNPs with significant ($p < 0.05$) regression and hpTDT results were found for the 0301-0501-0201 haplotype and these results are displayed in Table 2. In some instances, the significant results in Table 2 are for two or more SNPs in strong pairwise LD ($r^2 > 0.9$, see column 2 of the table) which increases confidence that the significant result is not due to genotyping error. The few remaining SNPs that gave significant results for 0101-0101-0501 or for 0701-0201-0201 in both regression analyses are shown in Table 3. Only one SNP in Table 3 (rs2523676 for the 0101-0101-0501 haplotype) was also significant in the hpTDT analysis.

The convergence of significant regression results for the particular SNPs listed in Tables 2 and 3 raises the question of whether any of these SNPs exhibit experiment-wide statistical significance that accounts for multiple testing. To provide an approximate p-value cutpoint for experiment-wide significance, we take 5288 as a lower bound on the number of statistical tests since this was the number of hpTDT tests conducted across the six DRB1-DQA1-DQB1 haplotypes after discarding hpTDT analyses for SNPs with fewer than 10 total TDT transmissions (see Methods). This lower bound for statistical tests implies a Bonferroni-corrected p-value of 9.4×10^{-6} for experiment-wide significance. To determine if any SNPs in Tables 2 or 3 yield a p-value below 9.4×10^{-6} , we note that the two regression analyses conducted for each DRB1-DQA1-DQB1 haplotype can be considered independent and hence we combine the two results to achieve an overall p-value for each SNP (see Methods). The lowest such overall p-value from Tables 2 and 3 is 0.00011 (for rs3130685 in Table 2) and hence none of the combined regression analysis results appear to have experiment-wide significance. Nor do any of the hpTDT results, considered on their own, achieve experiment-wide significance. So, in summary, the SNPs highlighted in Tables 2 and 3 provide provisional evidence of marking subsets of DRB1-DQA1-DQB1 haplotypes with different diabetes risk; but none of the SNPs provided convincing experiment-wide significance and hence the SNPs may be false positives.

DISCUSSION

In this paper, we described results of a novel approach that searched for MHC SNPs that might mark chromosomes having different diabetes risk despite carrying the same haplotype at DRB1-DQA1-DQB1. This novel approach relied heavily on the newly released Beagle software which is able to rapidly phase genotypes from several hundred thousand SNPs in datasets consisting of several thousand subjects. We identified a number of SNPs that gave nominally significant evidence for marking subsets of the 301-501-201 haplotype that might

differ with respect to their diabetogenic potency. However, none of the SNPs achieved experiment-wide significance and hence the SNPs may be false-positive associations.

To try to determine if any of these SNPs might be true-positives, we followed up these results by attempting to replicate them in the BDA cohort. The BDA families were not included in our primary analyses because DQA1 was not genotyped in BDA samples and the genotypes at DRB1 and DQB1 were coded somewhat differently than in the other T1DGC cohorts. However the BDA coding was unambiguous for carriers of the 301-501-201 haplotype enabling us to analyze this haplotype in the BDA samples using the same two regression analyses performed in the primary analysis (i.e. one diabetic versus non-diabetic sib per family; one affected child per family versus the parents). To increase the power of this attempted replication, we combined the results of the two BDA regression analyses by the same method used in the primary study (see Methods). When we compared these combined BDA regression results with those of the SNPs listed in Table 2, we found that none of the SNPs were replicated (i.e. $p < 0.05$) in the BDA cohort. Though the smaller size of the BDA dataset might lack sufficient power to replicate the modest effect sizes in Table 2, the failure to replicate reinforces the impression that the SNPs may be false-positives.

This raises the issue of whether failure to identify more robust findings might be due to limitations or deficiencies of the analysis we conducted. Several items should be mentioned in this regard: The regression comparisons of affected versus non-affected sib, and of affected child versus parents are likely to be less powerful than a comparison of diabetic cases versus unrelated controls [13]. Another possible weakness might arise from an unidentified feature of the Beagle software or its phased haplotypes which are not well suited to the haplotype-based regression analysis we conducted. We included the hpTDT analysis to guard against this possibility and to provide an alternative test for association that did not depend on haplotype phasing; however, a limitation of the hpTDT is that many combinations of heterozygous SNP and homozygous DRB1-DQA1-DQB1 haplotype were found in too few parents to test. Further work to address these possible deficiencies and limitations might identify more robust associations in which SNPs mark subgroups of DRB1-DQA1-DQB1 haplotypes with different diabetes risk. In addition, phasing of MHC haplotypes in large numbers of subjects may help answer a related question we did not address: namely whether any MHC variants modulate type 1 diabetes independent of specific DRB1-DQA1-DQB1 haplotypes.

Acknowledgments

This research utilizes 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), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418.

This work was supported by the Wellcome Trust.

APPENDIX I

As noted in the main text, our initial hpTDT analyses identified SNP rs2040410 as providing highly significant TDT evidence ($p < 4.6 \times 10^{-9}$) of marking two subsets of the 0301-0501-0201 haplotype with very different diabetes risks. However, further examination of this result indicated that it is likely to be due to a limited number of genotype errors in which homozygous parents who are “1/1” at rs2040410 have been miscalled as “1/3” heterozygotes; and hence TDT allele transmissions to affected children are tabulated for

these (apparently) heterozygous parents. Since the other parent in the 24 families generating TDT significance is almost always a “3/3” homozygote, it is important to note that *all* children in the 24 families (affected *and* unaffected) have been called as “1/3” heterozygotes – this is exactly what their genotypes should be if their parent's genotypes are actually “1/1” and “3/3” at rs2040410. The affected “1/3” children appear to always inherit allele 1 (and never allele 3) from the likely miscalled “1/3” parent which, in turn, generates the highly significant hpTDT result. However, evidence that the “1/3” parents are miscalled is provided by the fact that all *non-diabetic* children in the 24 families also exhibit the same extremely skewed inheritance of allele 1 as diabetic children. Furthermore, strong evidence for miscalling of rs2040410 in the “1/3” parents is also provided by rs6914950 which is in strong LD with rs2040410 in HapMap CEU subjects ($r^2=0.93$). All of the “1/3” parents at rs2040410 in the 24 families are called as “1/1” homozygotes at rs6914950. Consequently, no significant hpTDT result is generated at rs6914950 because there are almost no parents in the 24 families who are heterozygous at rs6914950 and homozygous for 301-501-201 at DRB1-DQA1-DQB1. This again indicates that the correct genotype for the “1/3” parents at rs2040410 is “1/1”.

REFERENCES

1. Thomson G, Valdes AM, Noble JA, et al. Relative predispositional effects of HLA class II DRB1-DQA1 haplotypes and genotypes on type 1 diabetes: a meta-analysis. *Tissue Antigens*. 2007; 70:110–127. [PubMed: 17610416]
2. Melanitou E, Fain P, Eisenbarth GS. Genetics of Type 1A (immune mediated) diabetes. *Journal of Autoimmunity*. 2003; 21:93–98. [PubMed: 12935776]
3. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am J Hum Genet*. 2007; 81:1084–1097. [PubMed: 17924348]
4. Lie BA, Todd JA, Pociot F, et al. The predisposition to type 1 diabetes linked to Human Leukocyte Antigen complex includes at least one non-class II gene. *Am J Hum Genet*. 1999; 64:793–800. [PubMed: 10053014]
5. www.t1dgc.org – description of samples, genotypes
6. Miretti M, Walsh E, Ke X, et al. A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am J Hum Genet*. 2005; 76:634–646. [PubMed: 15747258]
7. Sham, P. Chapter 4 in *Statistics in Human Genetics*. Arnold; London: 1998. The analysis of allelic associations.
8. Balding D. A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*. 2006; 7:781–791.
9. Page G, Amos C. Comparison of linkage-disequilibrium methods for localization of genes influencing quantitative traits in humans. *Am J Hum Genet*. 1999; 64:1194–1205. [PubMed: 10090905]
10. Rosenthal, R. *Meta-Analytic Procedures for Social Research*. Sage Publications Inc.; 1991.
11. de Bakker P, McVean G, Sabeti P, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nature Genetics*. 2006; 38:1166–1172. [PubMed: 16998491]
12. Lampis R, Morelli L, Congia M, et al. The inter-regional distribution of HLA class II haplotypes indicates the suitability of the Sardinian population for case-control association studies in complex diseases. *Human Molecular Genetics*. 2000; 9:2959–2965. [PubMed: 11115839]
13. Risch N. Searching for genetic determinants in the new millennium. *Nature*. 2000; 405:847–856. [PubMed: 10866211]

Table 1

Within-dataset frequencies of DRB1-DQA1-DQB1 haplotypes in T1DGC parents¹

DRB1-DQA1-DQB1 ²	AP (574)	BDA ³ (1668)	DAN (458)	EUR (1510)	HBDI (1728)	JOS (270)	NA (1060)	SAR (256)	UK (408)
0101-0101-0501	7	10.3	7.6	8.4	7.1	6.3	7.5	2	6.4
0101-0102-0504	0.2	0	0	0.2	0.1	0.4	0.1	0	0
0102-0101-0501	0.9	0	0.2	0.6	0.8	1.1	0.9	1.2	0.7
0103-0101-0501	0.5	0	0.4	0.4	1	0.4	0.4	0.4	1.5
0301-0501-0201	26.7	26.8	24	21.7	23.1	23.7	26.5	41	31.6
0401-0301-0301	4.7	5.8	2.8	1.5	3.1	3.3	3.4	0	2.5
0401-0301-0302	17.4	17.6	24.2	17.1	15.5	14.4	14.8	0	18.9
0402-0301-0302	1.7	0.7	0.4	3.8	2.8	3	2.5	2	1
0403-0301-0302	0.9	0.4	0	0.6	0.3	0.4	0.8	0.4	0.5
0404-0301-0302	5.6	6.4	2.6	2.8	6.1	7.4	5.8	0.8	5.6
0405-0301-0201	0	0	0	0.5	0.5	1.1	0.7	3.5	0
0405-0301-0302	1.6	1.3	0.4	1.5	1.6	3	2.1	15.6	1.5
0407-0301-0301	0.5	0	0.2	0.6	0.6	0.7	0.3	0	0
0407-0301-0302	0	0	0	0	0.1	0	0.4	0	0.2
0408-0301-0301	0.3	0	0	0.3	0.3	0	0.5	0	0.2
0408-0301-0304	0.2	0	0	0.8	0.2	0	0.1	0	0.2
0701-0201-0201	6.1	5.7	4.6	7.8	5.7	5.2	5.5	2.7	3.9
0701-0201-0303	1.6	1.4	1.7	1.7	1.6	2.2	1.2	1.2	1.7
0801-0401-0402	1.4	0	1.7	2.9	2.8	1.5	2.7	0	1
0803-0601-0301	0.2	0	0	0.2	0.1	0.4	0.1	0	0
0804-0401-0402	0.2	0	0	0.2	0.1	0.4	0.1	0.8	0
0901-0301-0201	0	0	0	0.1	0.2	0	0.2	0	0.2
0901-0301-0303	1.7	1.3	1.1	1	1	1.1	1	0.4	2.7
01001-0101-0501	0.3	0	0.2	0.3	0.3	0.7	0.8	0	0.2
01101-0501-0301	2.3	2.8	2.8	4	3	3.3	3.5	2.7	1.7
01102-0501-0301	0.2	0	0	0.1	0.1	0	0.7	0.4	0
01103-0501-0301	0.3	0	0.2	0.9	0.5	0.4	0.3	0.4	0.2
01104-0501-0301	0.7	0	1.1	1.3	1.5	1.1	0.8	3.9	1

DRB1-DQA1-DQB1 ²	AP (574)	BDA ³ (1668)	DAN (458)	EUR (1510)	HBDI (1728)	JOS (270)	NA (1060)	SAR (256)	UK (408)
01104-0103-0603	0	0	0	0.1	0.2	0	0.2	0	0
01201-0501-0301	0	0.6	1.7	0.7	0.8	0.4	0.9	0.8	2
01301-0103-0603	3.7	2.6	2.2	3.6	2.7	1.1	2.5	0.4	1
01302-0102-0604	2.4	3.0	8.5	3	3.6	3.7	2.5	0.4	3.2
01302-0102-0609	0	0.7	0.2	0.1	0.8	0.4	0.3	0	0.2
01303-0501-0301	0.2	0	0.2	0.7	0.7	1.1	0.4	1.2	0.2
01401-0101-0503	1.2	0.9	0.4	0.7	1.1	1.9	0.9	2	0.2
01404-0101-0503	0.3	0	0.2	0.1	0.1	0.4	0.4	0	0.2
01501-0102-0502	0	0.2	0	0.1	0.2	0	0.1	0.8	0.2
01501-0102-0602	5.2	4.7	7	5.5	6.2	8.5	4.6	1.2	6.9
01501-0102-0603	0	0.1	0	0.1	0.1	0	0.1	0	0.2
01502-0103-0601	0.3	0.4	0.2	0.5	0.3	0	0.4	0.4	0
01601-0102-0502	1	0.6	0.4	2.3	1.7	0	0.9	11.7	0.2

¹ Haplotypes with 5 or more total counts in T1DGC parents are displayed. HLA genotypes were phased by Beagle and frequencies are displayed as a percentage of total parental haplotype counts in each data set (shown in parenthesis)

² Names of haplotypes known to be diabetes-predisposing [1,2] are highlighted in black

³ BDA haplotype categories could be usually be inferred despite the absence of DQA1 genotypes and a coding scheme for DRB1 and DQB1 that differed somewhat from the other datasets

Table 2
Logistic regression and hpTDT results for haplotypes carrying 0301-0501-0201 at HLA-DRB1-DQA1-DQB1¹

SNP	LD Group ⁴	Assoc Allele ⁵	Regression 1 affd child - 1 unaffd child per family ²		Regression 1 affd child - 2 parents per family ³		hpTDT all affd children - 2 parents per family			
			Likelihood Ratio ChiSq	p-value	Likelihood Ratio ChiSq	p-value	Trans	Not Trans	ChiSquare	p-value
rs879882	*	1 (3)	6.91	0.0086	3.95	0.0468	68	34	11.3	0.00078
rs1265112	A	1 (3)	6.1	0.0135	6.12	0.0133	73	48	5.2	0.02259
rs1576	A	2 (3)	5.82	0.0159	6.21	0.0127	78	48	7.1	0.00771
rs1265078	A	3 (2)	5.82	0.0159	5.95	0.0148	78	48	7.1	0.00771
rs746647	A	1 (3)	5.82	0.0159	5.95	0.0148	78	48	7.1	0.00771
rs3135363	*	1 (3)	6.01	0.0142	3.86	0.0494	63	37	6.8	0.00912
rs7766843	B	3 (1)	5.8	0.0161	5.82	0.0158	61	33	8.3	0.00396
rs4428528	B	3 (2)	5.51	0.0189	5.13	0.0235	61	35	7	0.00815
rs2187823	B	2 (4)	5.51	0.0189	4.71	0.03	61	41	3.9	0.04829
rs6901541	*	4 (2)	5.78	0.0163	4.97	0.0258	59	35	6.1	0.01352
rs3130685	*	3 (1)	5.33	0.021	10.04	0.0015	69	37	9.7	0.00184
rs7452076	*	3 (1)	5.17	0.023	4.4	0.0359	64	34	9.2	0.00242
rs720465	*	3 (4)	5.09	0.024	6.8	0.0091	71	48	4.4	0.03594
rs7194	C	4 (2)	4.93	0.0264	5.16	0.0232	61	32	9	0.0027
rs7192	C	2 (1)	4.41	0.0358	4.61	0.0318	62	29	12	0.00053
rs3135461	*	3 (1)	4.85	0.0276	4.23	0.0397	71	40	8.7	0.00318
rs3128982	*	4 (2)	4.81	0.0284	4.91	0.0267	74	46	6.5	0.01079
rs2239804	*	3 (1)	4.7	0.0301	3.92	0.0478	65	26	16.7	0.00004
rs1736980	D	1 (3)	4.6	0.032	7.3	0.0069	87	56	6.7	0.00964
rs1633021	D	4 (2)	4.6	0.032	7.22	0.0072	88	56	7.1	0.00771
rs1265095	*	2 (4)	4.59	0.0321	6.86	0.0088	76	44	8.5	0.00355
rs1633030	E	1 (3)	4.41	0.0357	6.84	0.0089	86	52	8.4	0.00375
rs1614309	E	3 (1)	4.41	0.0357	6.84	0.0089	87	55	7.2	0.00729
rs1737049	E	4 (2)	4.14	0.0432	6.57	0.0104	87	55	7.2	0.00729
rs3130573	*	4 (2)	4.09	0.042	4.56	0.0328	83	45	11.3	0.00078
rs2073717	*	2 (3)	4.02	0.0449	4.33	0.0375	71	34	13	0.00031

SNP	LD Group ⁴	Assoc Allele ⁵	Regression 1 affd child - 1 unaffd child per family ²		Regression 1 affd child - 2 parents per family ³		hpTDT all affd children - 2 parents per family			
			Likelihood Ratio ChiSq	p-value	Likelihood Ratio ChiSq	p-value	Trans	Not Trans	ChiSquare	p-value
rs241439	*	1 (2)	3.94	0.0471	6.45	0.0111	82	48	8.9	0.00285
rs1611635	*	2 (4)	3.88	0.0488	3.88	0.0487	80	57	3.9	0.04829
rs3131012	*	1 (3)	3.87	0.0491	5.16	0.0231	73	36	12.6	0.00039

¹ Analyses include all cohorts except BDA. Regression analyses consider only phased haplotypes carrying 0301-0501-0201 at HLA-DRB1-DQA1-DQB1. Homozygous parent TDT (hpTDT) considers only transmissions from heterozygous parents who are also homozygous for 0301-0501-0201 at DRB1-DQA1-DQB1. Results are shown only if allelic associations are in the same direction and significant ($p < 0.05$) in all three analyses.

² Haplotypes are drawn from data set consisting of one affected and one unaffected child per family

³ Haplotypes are drawn from data set consisting of unaffected T1DGC parents and one affected child per family (where the affected child is different from the affected child in the 1 affd-1 unaffd child regression analysis)

⁴ SNPs with pairwise $r^2 > 0.9$ are denoted as belonging to same LD group; "*" indicates SNP not in strong LD with others in the table

⁵ Diabetes-associated allele and its alternate (in parenthesis); alleles are numbered as in T1DGC data files (1=A, 2=C, 3=G, 4=T)

Table 3

Logistic regression and hpTDT results for haplotypes carrying 0101-0101-0501 or 0701-0201-0201 at HLA-DRB1-DQA1-DQB1¹

0101-0101-0501 haplotype results			Regression 1 affd child - 1 unaffected child per family ²		Regression 1 affd child - 2 parents per family ³		hpTDT all affd children - 2 parents per family ⁴			
SNP	LD Group ⁵	Assoc Allele ⁶	Likelihood Ratio ChiSq	p-value	Likelihood Ratio ChiSq	p-value	Trans	Not Trans	ChiSquare	TDT p-value
rs2844746	*	2(4)	5.7	0.0170	5.8	0.016	7	5	0.33	0.5637
rs422544	*	4(3)	4.7	0.0310	6.8	0.009	--	--	--	--
rs2523676	*	4(2)	3.9	0.0480	4.5	0.034	13	1	10.29	0.00135
0701-0201-0201 haplotype results			Regression 1 affd child - 1 unaffected child per family ²		Regression 1 affd child - 2 parents per family ³		hpTDT all affd children - 2 parents per family ⁴			
SNP	LD Group ⁴	Assoc Allele ⁵	Deviance	p-value	Deviance	p-value	Trans	Not Trans	ChiSquare	TDT p-value
rs4713438	*	2(4)	9.7	0.0018	3.9	0.0477	--	--	--	--
rs9276529	*	3(1)	4.5	0.0339	4.7	0.0306	--	--	--	--
rs449643	*	3(1)	4.2	0.0404	4.4	0.0349	--	--	--	--
rs387608	A	2(4)	4.2	0.0404	5.0	0.0250	--	--	--	--
rs389512	A	3(2)	4.2	0.0404	5.0	0.0250	--	--	--	--
rs3104401	*	1(3)	4.1	0.0418	4.3	0.0380	--	--	--	--
rs743400	*	4(2)	4.0	0.0459	4.9	0.0259	--	--	--	--

¹ Analyses include all cohorts except BDA. Separate analyses were conducted for haplotype 0101-0101-0501 and haplotype 0701-0201-0201. Results are shown only for SNPs exhibiting allelic association in the same direction and a significant ($p < 0.05$) result in *both* regression analysis for conducted for the same DRB1-DQA1-DQB1 haplotype and SNP

² Haplotypes are drawn from data set consisting of one affected and one unaffected child per family

³ Haplotypes are drawn from data set consisting of unaffected T1DGC parents and one affected child per family (where the affected child is different from the affected child in the 1 affd-1 unaffected child regression analysis)

⁴ Homozygous parent TDT (hpTDT) only considers SNP allele transmissions from heterozygous parents who are also homozygous for 0101-0101-0501 or 0701-0201-0201 at DRB1-DQA1-DQB1. SNPs with fewer than 10 SNP allele transmissions were not evaluated by hpTDT and are denoted by two dashes ("--") in the table.

⁵ SNPs with pairwise $r^2 > 0.9$ are denoted as belonging to same LD group; "*" indicates SNP not in strong LD with others in the table

⁶ Diabetes-associated allele and its alternate (in parenthesis); alleles are numbered as in T1DGC data files (1=A, 2=C, 3=G, 4=T)