Supplementary Materials

Supplementary description to study population

We used a case-control study design to evaluate genetic associations with T1D. A total of 962 patients (cases) were from the nationwide Swedish Better Diabetes Diagnosis (BDD) study(1-4), which involves participation during 2005-2010 by all 42 pediatric clinics in Sweden. The American Diabetes Association and World Health Organization criteria were used for the diagnosis of diabetes and to classify the disease (5). Patients diagnosed with diabetes at 9 months – 18 years of age were sequentially registered in the BDD study (2-4). At the time of clinical diagnosis, the patients had one or several autoantibodies against either insulin (IAA), GAD65 (GADA), IA-2 (IA-2A), and three variants (residue 325 being either R, W or Q) (ZnT8RA, ZnT8WA or ZnT8QA, respectively, as previously reported (3; 6; 7). A total of 636 geographically representative controls were analyzed in our analyses *(6; 7)*. The Karolinska Institute Ethics Board approved the BDD study (2004/1:9).

The HLA-DQA1/B1 original genotype and clinical data will be availabel for collaborative research upon approval from BDD investigators and the Karolinska Institute Ethics Board. Interested investigators are encouraged to contact Dr. Ake Lernmark, the sponsor of this project.

Supplementary description to DNA extraction

The plasmid Max isolation kit (Qiagen, Bothell, Washington, USA) was used to isolate DNA according to the manufacturer's instructions from frozen whole blood samples of cases and controls.

Supplementary description to HLA next generation targeted sequencing (NGTS) analysis

The NGTS HLA typing approach utilized PCR based amplification of HLA and sequencing using Illumina MiSeq technology as described in detail (8; 9). Briefly, the laboratory steps consisted of consecutive PCR reactions with bar coding incorporated in the PCRs for individual sample tracking followed by application to the MiSeq. Robust assays for each of the target loci for all class II genes were developed. The depth of genotyping was extended to HLA-DQA1 and DQB1 to include exons 1-4. The analytical tools to define haplotypes and genotypes were developed in collaboration with Scisco Genetics (Seattle, WA). To date these tools have been tested – with 100% accuracy – on >2,000 control samples genotyped with the present NGS approach (8; 9).

Despite high polymorphisms, both DQA1 and DQB1 are in high LD, and thus their haplotypes can be inferred accurately for 97% of subjects. For 13 subjects with less than perfect haplotyping, the posterial probabilities of inferring corresponding haplotypes exceed 95%, and thus are used to infer their haplotypes with the best possible haplotypes.

Supplementary description to residue sequences of DQA1 and DQB1 alleles

As NGTS sequenced DNA nucleotides of selected exons and used sequences to deduce DQA1 and DQB1 alleles, HLA-DQA1 and -DQB1 genotypes were obtained at the high resolution of 6 digits or higher, based on the HLA nomenclature of IMGT (https://raw.githubusercontent.com/ANHIG/IMGTHLA/Latest/alignments/DQA1). After

translating nucleotides to predict residues, we further determined physical positions of individual residues in the alpha chain from codon α1 to α232, and in the beta chain from codon β1 to β237. The numbering system first suggested (10) and later improved (11) allowed for structural equivalence among residues of various MHC II alleles, regardless of gene locus or species, and based on the structure of the first published MHC II allele HLA-DR1(12; 13). There were also 23 and 32 residues in the signaling peptides of DQA1 and DQB1, respectively, for which their positions were labeled as, respectively, position (-23) and (-32) to the position (-1).

Supplementary description to islet autoantibodies

GADA, IA-2A, IAA, and three variants of ZnT8A (ZnT8-RA, ZnT8-WA or ZnT8-QA, respectively) were determined in quantitative radio-binding assays using in house standards to determine levels as previously described in detail, which are then used to call positive or negative by pre-established assay-specific threshold values (3; 14).

Supplementary description to molecular depiction of HLA-DQ structures

All depictions of determined structures of various HLA-DQ alleles have been carried out using the WebLabViewer v. 3.5 and the DSViewerPro v. 6.0, 3-D molecular rendering software of Accelrys (currently Dassault Systèmes, BIOVIA, San Diego, CA, USA), based on coordinates freely available in the Protein Data Bank. The complex of DQ B1*06:02—Insulin B4-16 was determined and reported based on the structure of DQ B1*06:02-hypocretin peptide (15; 16).

Supplementary description to statistical analysis and software

Upon extracting residues for each DQ haplotype, we computed distances between all unique sequences pairwise by the stringdist function of the R package, with the Levenshtein distance, and the applied hierarchical clustering algorithm by hclust function of R package with the agglomeration method of ward.D2, i.e., clustering two sequences closer together, if their sum of squared distances was relatively small. Upon identifying clustered haplotypes of interest, e.g., DQ5 cluster, we focused on the cluster by re-coding all cluster-specific haplotypes and performed haplotype association of T1D with individual residues sharing the same haplotype cluster. In the diplotypic association analysis, we compared frequency counts of one diplotype versus the rest, and computed corresponding odds ratios, 95% confidence interval, and the Fisher's exact p-values. This analytical strategy tended to produce more robust inferences when sample sizes were relatively small.

With respect to p-values and multiple comparison issues, we chose to present the p-values without multiple comparison correction. While the conventional genetic analysis tended to compute corrected p-values for number of comparisons by, say Bonferroni correction or False discovery rate (17), its intension was to control the overall false positive error rates of all comparisons. In the current context with the analysis of HLA-DQ haplotypes/diplotypes, roles of both genes in T1D were implicated in multiple empirical and functional studies. Further, the empirical haplotype association analysis suggested that DQ associations were highly significant at 5% level, even if multiple comparisons with 45 comparisons were considered. Hence, the haplotypic and diplotypic association analyses aimed to uncover which haplotypes/diplotypes may explain the overall association, i.e., our explorations were in the alternative hypothesis domain. Finally, p-values with no multiple comparison corrections had clear and simple interpretations, free from varying haplotypes/diplotypes in various explorations. Throughout analyses, we used the threshold of 0.05 to highlight those p-values to be positively or negatively associated, while being mindful that some p-values, close to 0.05, could be falsely labeled.

Naïve applications of any multiple comparisons without varying correction factors could obscure the interpretations of results (18).

When computing haplotype-specific odds ratios, a “virtual null haplotype” under the null hypothesis is chosen as the reference haplotype, as explained earlier(7).

Table S1 list individual DQ haplotypes in those selected DQ clusters, with respect to their polymorphic residues within clusters.

Here are results obtained from Table S1. The haplotype cluster DQ4 included two neutral haplotypes (OR = 0.93 and 0.88, p-value = 0.70 and 0.87, respectively), one potentially resistant haplotype (DQB1*06:01-*03:01) with 6 observed copies only among controls, and two rare haplotypes (font colored gray). Other than the nearly monomorphic residue α150 (Table 1A, font colored blue for that column), there were 17 residues in total, ten of which (β9, β26, β55, β56, β66, β67, β70, β71, β74, β75) were in complete LD with each other and so were five residues (α22, β(-18), β13, β45, β167, β185), forming two complete LD blocks $DQ4_1$ and $DQ4_2$ (Table 2A). For simplicity, we used β9 and β13 to represent these two LD blocks. Note that we refer to a haplotype as a resistant or risk, if the corresponding p-value is less than 0.05 and odds ratio was, respectively, less than or greater 1, while it was a neutral haplotype with the corresponding p-value greater than 0.05. Also, if the number of observed haplotypes was less or equal to five (approximate haplotype frequency of ~0.16% in this study population), the haplotype is treated as a “rare haplotype”, since association analysis with fewer than five copies is generally unstable.

The haplotype cluster DQ5 had 8 haplotypes: two resistant haplotypes (OR = 0.08 and 0.11, p-value = $6.38*10^{-3}$ and $1.51*10^{-5}$, respectively, Table 1B, highlighted red), two neutral haplotypes (OR = 0.82 and 0.62, p-value = 0.0518 and 0.175, respectively), and four rare haplotypes. Excluding those residues that were monomorphic among the relatively common haplotypes, there were 7 residues, in which the pairs (α(-7), αb1) and (α31, α204), separately, were in complete LD (table 2B), and were represented by αb1 and α31, respectively. The polymorphism of haplotypes in the cluster DQ5 may be captured by motifs of five residues (αb1, α31, α196, β57, β126).

Similarly, the haplotype cluster DQ6 had four resistant haplotypes, two neutral haplotypes, and two rare haplotypes. Excluding the “nearly monomorphic residue” α81, a total of 28 residues included three LD blocks: the LD block DQ6.1 had six residues (α22F/Y, α38K/R, α126H/Q, α127A/S, α204V/M), LD block DQ6.2 (β3P/S, β(-21)G/D, β(-18)A/V, β(-17)P/A, β(-6)T/S, β(-5)P/L, β(-4)V/L, β13A/G, β26Y/L, β37D/Y, β38V/A, β66D/E, β67I/V, β167H/R, β197N/S, β203I/V) and LD block DQ6.3 (β57V/D, β86G/A, β87Y/F, β130Q/R) (Table 2C). The polymorphism of haplotypes in cluster DQ6 can be represented by the motif of six residues (α22, β3, β9, β30, β57, β70).

The next haplotype cluster DQ7a included two neutral haplotypes and one potentially resistant haplotype with 7 observed copies only among controls, in addition to one rare haplotype, fully represented by three residues (α(-6), α157, β57). Lastly, the haplotype cluster DQ7b included one neutral haplotype, one rare and potentially resistant haplotype with seven copies in controls and none in the patients, and six very rare haplotypes, which could be represented by two residues (α(-13), α157) after excluding 11 nearly monomorphic residues.

Table S2 lists residue-specific associations and Table S3 lists motif associations with T1D in respective DQ clusters

A residue accounts for the potential resistance to T1D among individuals with DQ4: Three residues (α135, β9, β13) represented 17 polymorphic residues in the DQ4 cluster. Clearly, α135 was not significantly associated with T1D, since its two residues (α135I versus α135T) had no associations (OR = 0.88 and 0.82, p-value = 0.868 and 0.238, respectively) and the difference of their associations was insignificant (Fisher's p-value=1.00) (Table S2A). On the other hand, the residue β9F, representing a LD block of 10 residues, denoted as DQ4, was not associated with T1D (OR = 0.93, p-value = 0.653), but its polymorphism had a potential to be resistant with 7 copies observed only among controls. The difference of their association appeared to be significantly different from zero (Fisher's p-value = 0.00288). Similar association results were observed for the residue β13 that represented 6 residues in the LD block, denoted $DQ4_2$. In fact, β9 and β13 themselves were nearly in perfect LD, differing only in one rare haplotype, and thus were represented by a single residue β9.

Together with six other residues (to be explained below), we evaluated two motifs (a1, a2) of all seven residues (α1b, α157, α196, β9, β30, β57, β70) for carriers of DQ4 (Table S3). Note that the residue is annotated to be ".", if it is the same as the reference residue (fonts are highlighted in boarded-brown below under motif m2). Given the single β9 residue selected (highlighted column in gray), it corresponded to two motifs a1 and a2, in which the second polymorphic residue β70 was in the LD block of β9. All other five residues were monomorphic. Among carriers of DQ4, the motif a2, corresponding to DQA1*04:01-B1*03:03 and DQA1*06:01-B1*03:01, appeared among controls only, indicating that they potentially were resistant to T1D. Note that a total of seven residues were identified from five resistant DQ haplotype clusters (below) and were combined for a consistent comparison and evaluation.

Three residues account for the resistance to T1D among individuals with DQ5: Twelve polymorphic residues across haplotypes in the DQ5 cluster could be represented by five residues (αb1, α31, α196, β57, β126), due to pooling residues through LD blocks (Table S2B). The first residue α1b had both residues (D and G) significantly resistant to T1D (OR = 0.71 and 0.07, p-value = $2.19*10^{-4}$ and $2.99*10^{-3}$, respectively), and the difference of their associations was significantly different from zero (Fisher's p-value = $9.97*10^{-3}$). Similar association patterns were observed for two other residues (α196, β57). On the other hand, differences of residues for residues α31, β126 were not significantly different from zero, thus eliminating their functional role of protecting against T1D in DQ5 individuals.

Together with four other residues, these three residues formed four motifs, after excluding two rare motifs (with one observed motif each in the study population). Across seven residues, three selected residues (αb1, α196, β57), highlighted in gray, were polymorphic among carriers of DQ5 (Table S3, right hand side), while five other residues were monomorphic. Residue β57 had three possible residues (D, S, V). The residues β57S and β57V on the motifs b2 and b3, respectively, captured the neutral association (OR = 0.62, and 0.80, p-value = 0.175, and 0.031, respectively), while β57D (on motifs b1 and b4) corresponded to two T1D resistant motifs (OR = 0.13 and 0.08, p-value = $2.21*10^{-4}$, and $1.87*10^{-2}$, respectively).

The residue α157 accounts for the T1D resistance among individuals with DQ7a: DQ7 haplotypes were clustered into DQ7a and DQ7b (Figure 1), in which HLA-DQA1 alleles were different. The DQ haplotypes in the DQ7a cluster represented two neutral, one potentially T1D resistant and one rare haplotype (Table S1D), all with three polymorphic residues (α(-6), α157, β57). The residue-specific association analysis indicated that the residue α157 had a variable association with residue A (alanine) and D (aspartic acid) (Fisher's p-value = 0.0064), listed in Table S2D. The first residue α157A was observed only among controls, implying potential T1D resistance, while the second residue, α157D, had marginally negative association with T1D (OR = 0.68, p-value = 0.0343).

Together with six other residues identified elsewhere, DQ7a carriers had monomorphic residues for all except for (α157, β57), of which β57 was deemed not to associate with variable associations among individuals with DQ7a haplotypes. The motif d1 had seven copies observed only among controls, implying its potential resistance to T1D (Table 3).

DQ haplotypes in DQ7b cluster are resistant to T1D: The DQ7b cluster included one resistant haplotype DQA1*05:05-B1*03:01, one likely resistant haplotype DQA1*05:03-B1*03:01 observed in controls only, and six rare DQ haplotypes (Table S1E). Excluding monomorphic residues in the DQ7b cluster left 14 residues, but 12 of them were nearly monomorphic, since variants occurred only among the rare haplotypes. Thus, two residues (α(-13), α157) were investigated for their associations with T1D. Both residue (A, T) of the residue α(-13) were significantly resistant to T1D (OR = 0.08 and 0.21, p-value = $2.55*10^{-3}$ and $1.94*10^{-22}$, respectively), but the difference of their ORs was not significant (Fisher's p-value = 0.688) (Table S2D). Hence this residue was consistently resistant to T1D, regardless of its residue. For the other residue, α157A was significantly resistant (OR = 0.21, p-value=$2.19*10^{-22}$), and α157S was observed only among controls. The comparison of both resistant associations indicate that they were not significantly different (Fisher's p-value = 0.202). To retain the DQ7b cluster, α157 was still chosen.

Based on seven residues (αb1, α157, α196, β9, β30, β57, β70) selected in other DQ clusters, there were two motifs, indexed by polymorphic α157A and α157S, while other residues are monomorphic. Despite this binary polymorphism, both motifs (e1, e2) showed significant resistance to T1D: the motif e1 had OR = 0.20 (p-value = $3.64*10^{-23}$), and motif e2 had 7 copies observed only among controls, implying that e2 was potentially resistant to T1D (Table S3).

References

1. Delli AJ, Lindblad B, Carlsson A, Forsander G, Ivarsson SA, Ludvigsson J, Marcus C, Lernmark A: Type 1 diabetes patients born to immigrants to Sweden increase their native diabetes risk and differ from Swedish patients in HLA types and islet autoantibodies. Pediatr Diabetes 2010;11:513-520
2. Delli AJ, Lindblad B, Carlsson A, Forsander G, Ivarsson SA, Ludvigsson J, Marcus C, Lernmark A: Type 1 diabetes patients born to immigrants to Sweden increase their native diabetes risk and differ from Swedish patients in HLA types and islet autoantibodies. Pediatric Diabetes 2010;11:513-520
3. Delli AJ, Vaziri-Sani F, Lindblad B, Elding-Larsson H, Carlsson A, Forsander G, Ivarsson SA, Ludvigsson J, Kockum I, Marcus C, Samuelsson U, Ortqvist E, Groop L, Bondinas GP, Papadopoulos GK, Lernmark A, Better Diabetes Diagnosis Study G: Zinc transporter 8 autoantibodies and their association with SLC30A8 and HLA-DQ genes differ between

immigrant and Swedish patients with newly diagnosed type 1 diabetes in the Better Diabetes Diagnosis study. Diabetes 2012;61:2556-2564
4. Carlsson A, Kockum I, Lindblad B, Engleson L, Nilsson A, Forsander G, Karlsson AK, Kernell A, Ludvigsson J, Marcus C, Zachrisson I, Ivarsson SA, Lernmark A: Low risk HLA-DQ and increased body mass index in newly diagnosed type 1 diabetes children in the Better Diabetes Diagnosis study in Sweden. Int J Obes (Lond) 2012;36:718-724
5. Diagnosis and classification of diabetes mellitus. Diabetes Care 2014;37 Suppl 1:S81-90
6. Zhao LP, Alshiekh S, Zhao M, Carlsson A, Larsson HE, Forsander G, Ivarsson SA, Ludvigsson J, Kockum I, Marcus C, Persson M, Samuelsson U, Ortqvist E, Pyo CW, Nelson WC, Geraghty DE, Lernmark A, Better Diabetes Diagnosis Study G: Next-Generation Sequencing Reveals That HLA-DRB3, -DRB4, and -DRB5 May Be Associated With Islet Autoantibodies and Risk for Childhood Type 1 Diabetes. Diabetes 2016;65:710-718
7. Zhao LP, Papadopoulos GK, Kwok WW, Xu B, Kong M, Moustakas AK, Bondinas GP, Carlsson A, Elding-Larsson H, Ludvigsson J, Marcus C, Persson M, Samuelsson U, Wang R, Pyo CW, Nelson WC, Geraghty DE, Lernmark A: Eleven Amino Acids of HLA-DRB1 and Fifteen Amino Acids of HLA-DRB3, 4, and 5 Include Potentially Causal Residues Responsible for the Risk of Childhood Type 1 Diabetes. Diabetes 2019;68:1692-1704
8. Smith AG, Pyo CW, Nelson W, Gow E, Wang R, Shen S, Sprague M, Pereira SE, Geraghty DE, Hansen JA: Next generation sequencing to determine HLA class II genotypes in a cohort of hematopoietic cell transplant patients and donors. Hum Immunol 2014;75:1040-1046
9. Nelson WC, Pyo CW, Vogan D, Wang R, Pyon YS, Hennessey C, Smith A, Pereira S, Ishitani A, Geraghty DE: An integrated genotyping approach for HLA and other complex genetic systems. Hum Immunol 2015;
10. Fremont DH, Monnaie D, Nelson CA, Hendrickson WA, Unanue ER: Crystal structure of I-Ak in complex with a dominant epitope of lysozyme. Immunity 1998;8:305-317
11. Bondinas GP, Moustakas AK, Papadopoulos GK: The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. Immunogenetics 2007;59:539-553
12. Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC: Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 1993;364:33-39
13. Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC: Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature 1994;368:215-221
14. Vaziri-Sani F, Delli AJ, Elding-Larsson H, Lindblad B, Carlsson A, Forsander G, Ivarsson SA, Ludvigsson J, Marcus C, Lernmark A: A novel triple mix radiobinding assay for the three ZnT8 (ZnT8-RWQ) autoantibody variants in children with newly diagnosed diabetes. J Immunol Methods 2011;371:25-37
15. Ettinger RA, Papadopoulos GK, Moustakas AK, Nepom GT, Kwok WW: Allelic variation in key peptide-binding pockets discriminates between closely related diabetes-protective and diabetes-susceptible HLA-DQB1*06 alleles. J Immunol 2006;176:1988-1998
16. Siebold C, Hansen BE, Wyer JR, Harlos K, Esnouf RE, Svejgaard A, Bell JI, Strominger JL, Jones EY, Fugger L: Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy. Proc Natl Acad Sci U S A 2004;101:1999-2004
17. Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and power approach to multiple testing. JRSS B 1995;57:289-300
18. Wasserstein RL, Schirm AL, Lazar NA: Moving to a World Beyond "p < 0.05". Am Stat 2019;73:1-19