Nine residues in HLA-DQ molecules determine with susceptibility and resistance to type 1 diabetes among young children in Sweden

Lue Ping Zhao1,2*, George K. Papadopoulos2,10, Antonis K. Moustakas3, George P. Bondinas2, Annelie Carlsson4, Helena Elding Larsson5, Johnny Ludvigsson6, Claude Marcus7, Martina Persson8, Ulf Samuelsson6, Ruihan Wang9, Chul-Woo Pyo9, Daniel E. Geraghty9 & Åke Lernmark5

HLA-DQ molecules account for over 50% of genetic risk of type 1 diabetes (T1D), but little is known about associated residues. Through next generation targeted sequencing technology and deep learning of DQ residue sequences, the aim was to uncover critical residues and their motifs associated with T1D. Our analysis uncovered (αa1, α44, α157, α196) and (β9, β30, β57, β70, β135) on the HLA-DQ molecule. Their motifs captured all known susceptibility and resistant T1D associations. Three motifs, “DCAA-YSARD” (OR = 2.10, p = 1.96*10−20), “DQAA-YYARD” (OR = 3.34, 2.69*10−72) and “DQDA-YYARD” (OR = 3.71, 1.53*10−6) corresponding to DQ2.5 and DQ8.1 (the latter two motifs) associated with susceptibility. Ten motifs were significantly associated with resistance to T1D. Collectively, homozygous DQ risk motifs accounted for 43% of DQ-T1D risk, while homozygous DQ resistant motifs accounted for 25% protection to DQ-T1D risk. Of the identified nine residues five were within or near anchoring pockets of the antigenic peptide (α44, β9, β30, β57 and β70), one was the N-terminal of the alpha chain (αa1), one in the CD4-binding region (β135), one in the putative cognate TCR-induced αβ homodimerization process (α157), and one in the intra-membrane domain of the alpha chain (α196). Finding these critical residues should allow investigations of fundamental properties of host immunity that underlie tolerance to self and organ-specific autoimmunity.

Type 1 diabetes (T1D) is an autoimmune disease, due to the destruction of normal pancreatic islet beta-cells by the host immune system, and accounts for the majority of all diabetes among young children1,2. Further, earlier twin and family studies underline the major role of genetic factors together with environmental factors3–4. Using modern genotyping technologies, recent genome wide association studies (GWAS) have uncovered ~50 single nucleotide polymorphisms (SNPs) in multiple non-HLA immune genes explaining probably less than 25% T1D associations5–7. In contrast, class II genes in the human leukocyte antigen (HLA) system have been researched for several decades, and strongly contribute to the genetic association with T1D, accounting for ~50% of genetic association with T1D8,9. Class II genes include the multiallelic HLA-DR*AI-B1 haplotype coding for the DRαβ heterodimer (α-chain essentially monomorphic), and the HLA-DQ*AI-B1 haplotype coding the DQαβ heterodimer. Due to their genetic proximity, HLA-DR and -DQ genes are in linkage disequilibrium (LD), and thus empirically separating their individual associations with T1D is challenging8–11. Conventionally, DQ molecules exhibit
stronger empirical associations with T1D. Our recent effort focused on identification of specific DQ risk residues among subjects with high-risk HLA-DQ haplotypes (DQ2.5 and DQ8.1) as well as those resistant residues among subjects with low risk HLA-DQ haplotypes, complementing earlier works on HLA-DR molecules.

Integrating susceptible and resistant DQ residues, the current investigation was to explore their genotypic associations with T1D, especially, assessing heterodimeric associations with DQA1 and DQB1 alleles on the same chromosome (cis) and between homologous chromosomes (trans), with specific focus on the DQ2.5 and DQ8.1 haplotypes. Furthermore, we assessed if DR3 and DR4, two major risk alleles of HLA-DRB1, confounded HLA-DQ associations with T1D through haplotypic association analysis. To gain insight into autoimmunity, we centered our analysis on DQ motif associations with six biochemically defined islet autoantibodies measured at the time of T1D diagnosis. To replicate discovered associations, we utilized a case–control study from the Type 1 Diabetes Genetic Consortium (T1DGC) to carry out T1D associations with discovered motifs. Last, in a final analysis we estimated how much susceptibility and resistant DQ motifs contributed to either the excess or reduction of risk for T1D.

**Results**

**Critical residues among subjects with risk or resistant DQ haplotypes.** DQA1 and DQB1, which are in high LD and their alleles mostly form cis-haplotypes, are translated to form respectively the alpha and beta chain of the heterodimeric HLA-DQ molecule. DQ haplotypes were polymorphic and had 45 unique haplotypes in our case–control study, each of which consisted of unique residues (complete lists of amino acid sequences of mature molecules in Supplementary Figs. S1 and S2). Our phylogenetic analysis of sequence similarities showed that these DQ haplotypes were hierarchically organized into seven haplotype clusters: 1 to 7 (Fig. 1). Clusters 2 and 4 included susceptible motifs DQ2.5 and DQ8.1 (motifs colored green), and carriers of all DQ motifs in these two clusters were at high T1D risk. Among high-risk subjects, hierarchically organized haplotype (HOH) analysis identified three critical residues (α44, β57, β135) that captured disease association with HLA-DQ among subjects with high-risk DQ haplotypes. The remaining clusters included largely resistant DQ haplotypes, and the HOH analysis of subjects revealed seven critical residues (αa1, α156, α196, β9, β30, β57, β70) that accounted for autoimmunity in T1D.

**Figure 1.** Phylogenetic representation of all cis DQ motifs that are risk to (p < 0.05 and OR > 1, colored green), resistant to (p < 0.05, OR < 1, colored red), neutral (p > 0.05, black), and rare (with fewer than 5 observed motif copies, colored blue), together with all trans-DQ motifs (colored gray).
Four alpha and five beta residues were critically associated with T1D risk. Integrating all risk and resistant residues in the DQ molecule led to four alpha residues (αa1, α44, α157, α196) and five beta residues (β9, β30, β57, β70, β135) that were critically associated with either susceptibility or resistance to T1D. To assess roles of these residues in T1D associations, we investigated residue-specific frequencies of corresponding amino acids between controls and patients, who had high and low DQ risk haplotypes (Table 1). Residue α44 took four possible amino acids (C, K, Q, R); α44C and Q were strongly associated with high T1D risks (OR = 2.10 and 3.32, p = 1.96*10−20 and 1.00*10−86, respectively), α44K was strongly associated with resistance to T1D risk (OR = 0.27, p = 1.85)13 and α44R was absent among high-risk DQ subjects. Among low risk DQ subjects, the resistant α44C and Q had significant associations with resistance to T1D risk (OR = 0.39 and 0.60, p = 5.42*10−19 and 2.59*10−3, respectively), and the residue α44R was highly associated with the resistance to T1D (OR = 0.40 and p-value = 5.20*10−59). Interestingly, the amino acid α44K was absent in low risk subjects. When contrasting the associations of α44C and Q between high and low risk subjects, their differences of risk and resistant associations were highly significant (exact Fisher’s p-value = 2.16*10−31 and 1.92*10−19, respectively). On the other hand, the beta chain residues β57A (OR = 2.31; p = 5.46*10−99) and D (OR = 0.46; 3.93*10−4) had reverse associations with T1D risk among high-risk subjects. Among the low risk subjects, the first residue, β57A, had no significant association to T1D risk (OR = 0.93, p = 0.89), while the second residue, β57D, was associated with resistance to T1D (OR = 0.25, p = 2.23*10−96). Two amino acids, β57S/V, were absent among high-risk subjects and had no associations (p = 0.11 and 0.78, respectively). Last, β135D and G had opposite associations with T1D high-risk subjects (OR = 2.47 and 0.40, p = 6.06*10−104 and 2.00*10−7, respectively). Among the low risk subjects, β135D had significant resistance to T1D (OR = 0.41, p = 1.84*10−20), while β135G was absent. The disease associations with residue β135D were completely opposite to each other between high and low risk subjects (Fisher’s exact test; p = 2.88*10−117).

From the perspective of resistant residues, αa1D and G among low risk subjects were significantly associated with resistance to T1D (OR = 0.42 and 0.07, p = 9.79*10−49 and 2.99*10−3, respectively). Among high-risk subjects, residue αa1D had significant association with susceptibility to T1D (OR = 2.10, p = 1.84*10−30), while αa1G was absent. The association with αa1D was significantly different between low and high-risk subjects (exact Fisher’s

| Residue | AA | Control | Patient | OR | HS | p     |
|--------|----|---------|---------|----|----|------|
| 1) αa1 | D  | 441 (34.67) | 1404 (72.97) | 2.10 | 20.28 | 1.84E-91 |
|        | G  | 9 (0.71) | 1 (0.05) | 0.07 | -2.97 | 2.99E-03 |
| 2) α44 | C  | 153 (12.03) | 485 (25.21) | 2.10 | 9.26 | 1.96E-20 |
|        | K  | 114 (8.96) | 46 (2.39) | 0.27 | -8.23 | 1.85E-16 |
|        | Q  | 174 (13.68) | 873 (45.37) | 3.32 | 19.74 | 1.00E-86 |
|        | R  | 559 (43.95) | 337 (17.52) | 0.40 | -16.20 | 5.20E-59 |
| 3) α157 | A  | 420 (33.02) | 1300 (67.57) | 2.05 | 18.15 | 1.27E-73 |
|        | D  | 21 (1.65) | 104 (5.41) | 3.27 | 5.33 | 9.57E-08 |
|        | S  | 7 (0.55) | 1 (0-05) | 0.08 | -2.73 | 3.88E-03 |
| 4) α196 | A  | 441 (34.67) | 1404 (72.97) | 2.10 | 20.28 | 1.84E-91 |
|        | T  | 8 (0.63) | 1 (0.05) | 0.08 | -2.73 | 3.88E-03 |
| 5) β9  | F  | 260 (20.44) | 85 (4.22) | 0.22 | -13.87 | 9.44E-44 |
|        | L  | 9 (0.71) | 1 (0.05) | 0.07 | -3.25 | 1.10E-03 |
|        | Y  | 441 (34.67) | 1404 (72.97) | 2.10 | 20.28 | 1.84E-91 |
|        | S  | 344 (27.04) | 320 (16.63) | 0.61 | -7.34 | 2.09E-13 |
| 6) β30 | H  | 231 (18.16) | 532 (27.65) | 1.52 | 6.37 | 1.85E-10 |
|        | V  | 204 (16.04) | 266 (13.83) | 0.86 | -1.76 | 7.80E-02 |
| 7) β57 | A  | 392 (30.82) | 1370 (71.21) | 2.31 | 21.12 | 5.46E-99 |
|        | D  | 49 (3.85) | 34 (1.77) | 0.46 | -3.54 | 3.93E-04 |
|        | S  | 19 (1.49) | 17 (0.88) | 0.59 | -1.61 | 1.08E-01 |
| 8) β70 | E  | 61 (4.80) | 79 (4.11) | 0.86 | -0.94 | 3.47E-01 |
|        | G  | 493 (38.76) | 248 (12.89) | 0.33 | -16.58 | 3.99E-62 |
|        | R  | 441 (34.67) | 1404 (72.97) | 2.10 | 20.28 | 1.84E-91 |
|        | D  | 363 (28.54) | 1357 (70.53) | 2.47 | 21.65 | 6.06E-104 |
|        | G  | 78 (6.13) | 47 (2.44) | 0.40 | -5.20 | 2.00E-07 |

Table 1. Results from association analysis of nine selected residues (αa1, α44, α157, α196, β9, β30, β57, β70, β135) with T1D in the case-control study with 636 control and 962 patients with estimated allelic frequencies (%), odds ratio, hap-score and p-value among carriers of DQ2.5 and DQ8.1 and non-carriers.
| ID   | cis Motif       | Control | Patient OR | Z  | p     | Sera | Allele Group                  |
|------|----------------|---------|------------|----|-------|------|-------------------------------|
| 1    | DCAA-FYDGD      | 55 (4.32)| 77 (4.00)  | -0.45 | 6.53E-01 | DQ4  | *04:02-04:02-04:01-04:02-04:04-*04:02  | C.A.F.E   |
| 2    | DCAA-YASARD     | 153 (12.03)| 485 (25.21) | 2.10 | 9.36E+02 | DQ2.5 | *05:01-02:01-05:05-02:01      | C.A.S.A   |
| 3    | DCAA-YYDRD      | 139 (10.93)| 40 (2.08) | 0.19 | -10.41 | DQ7  | allele group 3                 | C.A.      |
| 4    | DCSA-YYDRD      | 7 (0.55) |            |      |       |      |                               |            |
| 5    | DKAA-YASARG     | 78 (6.13) | 41 (2.13) | 0.35 | -5.76  | DQ2.2 | *02:01-*02:02                 | K.A.S.A.G  |
| 6    | DKAA-YYDRD      | 36 (2.83) | 5 (0.26)  | 0.09 | -6.36  | DQ2.3 | *02:01-03:03                 | K.A.      |
| 7    | DQAA-YAYRD      | 152 (11.95)| 769 (39.97)| 3.34 | 17.98  | DQ8.1 | *03:01-03:02-03:01-03:02-03:04 | A.A.A.A  |
| 8    | DQAA-YYYRD      | 9 (0.71) |            |      |       |      |                               |            |
| 9    | DQCA-YASARG     | 6 (0.31) |            |      |       |      |                               |            |
| 10   | DQCA-YYaRD     | 13 (1.02) | 73 (3.79) | 3.71 | 4.81   | DQ8.1 | *03:02-03:03-*03:02-*03:04     | A.A.A.    |
| 11   | DQCA-YYDRD      | 67 (5.27) | 88 (4.57) | 0.87 | -8.32  | DQ2.3 | *02:03-02:02-02:03-02:04      |            |
| 12   | DRAA-YFYGD      | 201 (15.8) | 8 (0.42)  | 0.03 | -16.39 | DQ6.2 | *01:02-06:02                  | R.A.F.G.   |
| 13   | DRAA-LYDRD      | 9 (0.71) | 1 (0.05)  | 0.07 | -3.22  | DQ6.2 | *01:03-06:01                  | R.A.L.     |
| 14   | DRAA-YHGDG      | 117 (9.2) | 44 (2.29) | 0.25 | -8.86  | DQ6.3 | allele group 3                 | R.A.H.G.   |
| 15   | DRAA-YHSGD      | 16 (1.26) | 15 (0.78) | 0.62 | -1.36  | DQ6.4 | *01:02-05:02                  | R.A.H.S.G  |
| 16   | DRAA-YHVID      | 149 (11.71)| 180 (9.36) | 0.80 | -2.15  | DQ6.4 | *01:05-01:01-01:02-05:01      | R.A.H.V.G  |
| 17   | DRAA-YHVRD      | 52 (4.09) | 80 (4.16) | 1.02 | 9.22E-01 | DQ6.4 | *01:02-06:04                  | R.A.H.V    |
| 18   | DRAA-YYSED      | 3 (0.24) | 2 (0.10)  | 0.44 | -0.92  | DQ6.4 | *01:02-05:04                  | R.A.S.E.   |
| 19   | DRAA-YYVRD      | 2 (0.16) | 6 (0.31)  | 1.98 | 0.86   | DQ6.4 | *01:02-06:09                  | R.A.A.V.   |
| 20   | DRTA-YHGDG      | 8 (0.63) | 1 (0.05)  | 0.08 | -2.73  | DQ6.5 | *01:04-05:03                  | G.R.A.T.Y.H.G. |

1) allele group *05:05-*03:01-*05:05-*01:19-*05:05-*03:09-*05:10-*03:01-*05:09-*03:01-*06:01-*03:01-*04:01-*03:03
2) allele group *01:01-*05:03-*01:03-*06:03-*01:02-*06:03-*01:07-*05:03-*01:10-*06:03.
3) motif ID=11; DQDA-YYDRD is selected as a reference motif

Table 2. Allelic association analysis of type 1 diabetes with DQA1-DQB1 motifs of (αa1, α44, α157, α196, β9, β30, β57, β70, β135): estimated frequency (%) among control and patient, estimated odds ratios, Z-score, and p-value, corresponding serotype, and equivalent allelic groups, in addition to profiles of amino acids across all motifs. Five rare motifs, with fewer than five observed motif copies, are excluded. In total we have 1266 and 1921 motif copies observed in 636 controls and 962 patients, respectively, after excluding 6 and 3 motif copies from respective groups.

p = 7.87*10^-123. The next residue α157 took three possible amino acids (A, D, S); the first two, α157A/D, had significant resistance associations (OR = 0.40 and 0.68, p = 7.83*10^-89 and 0.034, respectively), while α157S was absent among patients but had seven allelic copies among controls with low risk DQ haplotypes. Among high-risk subjects, α157A/D were significantly associated with susceptibility to T1D (OR = 2.05 and 3.27, p = 1.85*10^-10 and 4.45*10^-71, respectively), but β30H was not observed in such carriers. Among high-risk subjects, β30S/Y had significant associations with T1D risk (OR = 0.40 and 0.68, p = 7.83*10^-89 and 0.034, respectively), while the third residue, α157S, was absent. Again, T1D risk associations with α157A/D were significantly different between low- and high-risk subjects (exact Fisher's p = 9.18*10^-97 and 4.93*10^-8, respectively). The next residue, α157A/D, were significantly associated with susceptibility to T1D (OR = 2.05 and 3.27, p = 1.27*10^-73 and 4.45*10^-71, respectively), but β30H/Y had two resistant amino acids β30H/Y (OR = 0.61 and 0.27, p = 2.09*10^-13 and 1.02*10^-71, respectively), and β30S was not observed in such carriers. Among high-risk subjects, β30S/Y had significant associations with susceptibility to T1D (OR = 2.01, p = 1.84*10^-91), and the association was significantly different between high and low risk (exact Fisher's p = 3.74*10^-44). Among low-risk subjects, residue β30 had two resistant amino acids β30H/Y (OR = 0.61 and 0.27, p = 2.09*10^-13 and 1.02*10^-71, respectively), and β30S was not observed in such carriers. Among high-risk subjects, β30S/Y had significant associations with susceptibility to T1D (OR = 2.05 and 3.27, p = 1.27*10^-73 and 4.45*10^-71, respectively), but β30H was not observed in such subjects. For amino acid β30Y, its association with T1D was significantly different between high and low risk (exact Fisher’s p = 7.62*10^-100). The ninth residue, β70, had two amino acids significantly associated with resistance, G and R (OR = 0.33 and 0.46, p = 7.76*10^-13 and 1.84*10^-9, respectively), and had one amino acid, β70E, with a neutral association (OR = 0.86, p = 0.34) among low-risk subjects. No high-risk subjects were found with β70E or G, while β70R was significantly associated with T1D (OR = 2.10, p = 1.84*10^-9), and its association was significantly different between high- and low-risk subjects.

In summary, all these nine residues had highly significant associations with T1D among subjects with either low- or high-risk DQ haplotypes. Strikingly, the same amino acids could exhibit paradoxical associations between low- and high-risk subjects. Further, some of common amino acids in the high-risk group are absent from the low-risk group, and vice-versa. Note that some motifs listed in Table 2 had fewer than 10 observed copies, e.g., ID 13, 18, 19 and 20, and their associated p-values were not robust enough and should be taken as suggestive values.

Motifs of nine residues account for DQ susceptibility and resistance to type 1 diabetes risks.

Four alpha residues and five beta residues originated from transcription in cis and haplotypes of these nine amino acids are referred to as cis DQ motifs. In total, there were 25 unique DQ motifs: four susceptibility
motifs (highlighted in green), ten resistance motifs (highlighted in red), and six neutral motifs with insignificant associations (p-values > 0.05) (Table 2). Five rare motifs with fewer than five observations were shown only in the Supplementary Table S1. The motif “DCAA-YSARD”, corresponding to DQA1*05:01-B*02:01 and A*05:05-B*07:01, was significantly associated with susceptibility to T1D (OR = 2.10, p = 1.96*10^{-20}), and is equivalent to the traditionally known DQ2.5 serotype. The motifs “DQAA-YYARD” and “DQDA-YYARD”, corresponding to traditionally known DQ8.1 serotypes, had significant association with T1D susceptibility (OR = 3.34 and 3.71, p = 2.69*10^{-72} and 1.53*10^{-6}, respectively). Last, the susceptible motif “DQDA-YDARG” was absent among controls and had six copies with allelic frequency of 0.31% among patients.

Among twenty HLA-DQ motifs, ten were associated with resistance to T1D. The motif “DRAA-FYDGD”, uniquely corresponding to the well-known protective DQ6.2 haplotype, conveyed significant resistance to T1D risk (OR = 0.03, p = 2.25*10^{-60}). The next resistance motif “DCAA-YYDGR”, motif 3, corresponding to a group of HLA-DQ haplotypes listed in the footnote, was significantly resistant to T1D (OR = 0.19, p = 2.32*10^{-23}). The third most protective motif, “DRAA-YHDGD”, was associated with resistance to T1D (OR = 0.25, p = 7.72*10^{-39}). The other five protective motifs were “DKAA-YSARG”, “DKAA-YYARD”, “DRAA-YYDRD”, “DRAA-YHVGD” and “GRAT-YHDGD”, with p-values less than 0.05. Two motifs, “DCSA-YYDRD” and “DQAA-YYARD”, were observed (with more than five copies) only among controls (Table 2).

Given the substantially different genetic distributions of residues between low- and high-risk subjects (Table 1), we assessed motif associations stratified by low or high DQ risks. As expected, distributions of DQ motifs were rather distinct between subjects with low and high DQ risks (Table S1). The most of commonly observed motifs were specific to subjects with high or low risk DQ haplotypes, and their associations with T1D were largely consistent with those in the combined low and high-risk subjects. For example, the motif “DCAA-YSARD” was observed only among high-risk subjects, and “DCAA-YYDGR” was observed only among low-risk subjects. Hence, in all combined data, their respective association with susceptibility or resistance for T1D were largely unchanged. One exception, to this general observation, was the motif “DQDA-YYARD”, which was significantly associated with susceptibility to T1D risk among high-risk subjects (OR = 4.56, p = 4.30*10^{-7}) but was relatively uncommon among low risk subjects and showed no indication of associations with T1D risks (OR = 0.88, p = 0.87).

**Subtle variations of residues alter association profiles with T1D.** The directly constructed profile of disease-associated motifs permitted us to examine how subtle variations of specific residues may alter their T1D associations (right panel of Table 2). The motif “DCAA-YSARD”, corresponding to DQ2.5, differed from the adjacent motif “DCAA-YYDRD” with only two residues at β30 and β57, but their T1D associations were completely opposite and reverted from susceptibility to resistance (OR = 2.10 and 0.19, p = 1.96*10^{-20} and 2.32*10^{-23}, respectively), the difference between the two was significant (Fisher's exact test: p = 4.24*10^{-39}). The next motif pair “DQAA-YYARD”, corresponding to DQ8.1, and “DKAA-YYDRD”, differed only at α44 and β57, and their T1D associations were opposite (OR = 3.34 and 0.09, p = 2.69*10^{-72} and 1.99*10^{-10}, respectively), the difference was significant (Fisher's exact test: p = 5.04*10^{-22}). The next DQ8.1 motif “DQDA-YYARD” differed from the adjacent motif “DQDA-YYDGR” only at residue β57, but the former was associated with susceptibility to T1D (OR = 3.71, p = 1.53*10^{-6}) and the other was neutral (OR = 0.87, p = 0.38). The difference of the associations between these two motifs was significant (Fisher's exact test: p = 7.48*10^{-6}). It was also of interest to contrast the uncommon motif 7, “DQAA-YYDRD”, with motif 8, “DQAA-YYARD”, and they were nearly identical except for β57. In fact, all five alleles in both motifs had an identical alpha chain, DQA1*03:01. Their associations differed significantly from each other (Fisher's exact test: p = 1.16*10^{-7}).

**Genotypes of DQ motifs have profound associations with T1D.** Each subject carries a pair of DQ motifs, i.e., a motif genotype, and two motifs were jointly associated with T1D. The combination of 20 motifs (Tables 2 and S2) resulted in several rare genotypes with fewer than 5 observations. Upon merging of all rare genotypes into a single category, there was a total of 60 distinct genotypes (Table S2), and 25 were deemed to have neutral associations with corresponding p-values greater than 0.05. The genotype association results for 35 motifs were sorted by Z-scores (Table 3). The first eight genotypes were significantly associated with susceptibility to T1D, with OR ranging from 2.46 to 14.54 (p-values ranged from 0.012 to 4.14*10^{-21}). Next thirteen genotypes were significantly associated with resistance to T1D risk (OR ranging from 0.038 to 3.69*10^{-9}). The last fourteen genotypes, which had frequencies of five or more copies were observed among controls and had six copies among patients, and were thus deemed to be highly resistant, perhaps protective, to T1D risks.

**DQ2.5 and DQ8.1 genotypic effects.** Out of 20 common DQ motifs, three motifs “DCAA-YSARD”, “DQAA-YYARD” and “DQDA-YYARD”, and the relatively uncommon “DQQA-YSARG”, were significantly associated with susceptibility to T1D. Given their susceptibilities, it was important to investigate specific genotype associations of these three DQ motifs, with themselves (homozygotes), with each other (doubly heterozygote), with neutral motifs (showing motif-specific effect), and with resistant motifs (showing additive motif effect): neutral motifs included all motifs with associated p-values greater than 0.05 and resistant motifs with p-values less than 0.05 and odds ratios less than 1. We conducted a univariate association analysis to compare each genotype combined through a logistic regression model, estimating coefficient, odds ratio, standard error, Z score and p-value (Table 4).

Pairing the motif “DCAA-YSARD” (DQ2.5) with neutral motifs, itself, two DQ8.1 motifs and resistant motifs resulted in five possible genotypes (Table 4). The heterozygote with neutral motifs showed a significant association with susceptibility to T1D (OR = 2.54, p = 0.035), which was largely due to the motif “DCAA-YSARD”. As
expected, its homozygote had a significant association with susceptibility to T1D with OR = 5.21 (p = 8.97*10−4), with an effect that was nearly doubled that of the heterozygote with neutral motifs (OR = 2.54). The heterozygote “DCAA-YSARD” (DQ2.5)/ “DQAA-YYARD” (DQ8.1) was associated with markedly increased susceptibility to T1D (OR = 14.52, p = 1.99*10 −10). Similarly, the heterozygote “DCAA-YSARD” (DQ2.5)/ “DQDA-YYARD” (DQ8.1) was associated with an even greater susceptibility (OR = 24.93, p = 2.95*10−3), although it had a lower genotype frequency. Last, the heterozygote with resistant motifs had insignificant association with T1D (OR = 0.69, p = 0.35), probably due to additive effect from both susceptibility and resistance motifs.

Pairing the motif “DQAA-YYARD” with neutral motifs, itself, “DQDA-YYARD” and resistant motifs, showed similar effects, albeit with some differences. The heterozygote with the neutral motifs captured the pure

| DQ Motif Genotype          | Control | Patient | OR    | Z     | p        |
|---------------------------|---------|---------|-------|-------|----------|
| 1 DCAA-YSARD/DQAA-YYARD   | 21 (3.3)| 269 (27.96) | 8.47 | 9.43 | 4.14E-21 |
| 2 DQAA-YYARD/DQAA-YYARD   | 5 (0.79)| 97 (10.08)  | 12.83| 5.56 | 2.64E-08 |
| 3 DQAA-YYARD/DAA-YHVDG    | 13 (2.04)| 83 (8.63)  | 4.22 | 4.83 | 1.38E-06 |
| 4 DCAA-FYDED/DQAA-YYARD   | 9 (1.42)| 48 (4.99)   | 3.53 | 3.47 | 5.22E-04 |
| 5 DCAA-YSARD/DCAA-YSARD   | 10 (1.57)| 46 (4.78)  | 3.04 | 3.19 | 1.43E-03 |
| 6 DQAA-YYARD/DAA-YHRD     | 11 (1.73)| 41 (4.26)  | 2.46 | 2.66 | 7.91E-03 |
| 7 DCAA-YSARD/DQAA-YYARD   | 1 (0.16)| 22 (2.29)   | 14.54| 2.62 | 8.83E-03 |
| 8 DQAA-YYARD/DQAA-YYARD   | 1 (0.16)| 20 (2.08)   | 13.22| 2.52 | 1.17E-02 |
| 9 DCAA-YYRD/DCAA-YYRD     | 7 (1.1) | 2 (0.21)    | 0.19 | -2.08| 3.77E-02 |
| 10 DCAA-FYDED/DAA-YHDGD   | 8 (1.26)| 2 (0.21)    | 0.17 | -2.28| 2.28E-02 |
| 11 DKAA-YYRD/DQAA-YYARD   | 8 (1.26)| 1 (0.1)     | 0.08 | -2.35| 1.87E-02 |
| 12 DKAA-YYRD/DAA-YHRD     | 9 (1.42)| 1 (0.1)     | 0.07 | -2.48| 1.32E-02 |
| 13 DCAA-YSARD/DCAA-YYRD   | 16 (2.52)| 6 (0.62)   | 0.25 | -2.91| 3.58E-03 |
| 14 DAA-YHVDG/DAA-YHVDG    | 13 (2.04)| 3 (0.31)   | 0.15 | -2.94| 3.33E-03 |
| 15 DAA-YHDGD/DAA-YHVDG    | 13 (2.04)| 2 (0.21)   | 0.10 | -3.01| 2.62E-03 |
| 16 DCAA-YSARD/DAA-YYRD    | 16 (2.52)| 4 (0.42)   | 0.17 | -3.22| 1.28E-03 |
| 17 DCAA-YYRD/DAA-YHVDG    | 16 (2.52)| 2 (0.21)   | 0.08 | -3.32| 8.86E-04 |
| 18 DCAA-YYRD/DAA-YYRD     | 17 (2.67)| 3 (0.31)   | 0.12 | -3.43| 6.02E-04 |
| 19 DAA-FYGD/DAA-YHVDG     | 25 (3.93)| 1 (0.1)    | 0.03 | -3.56| 3.68E-04 |
| 20 DQAA-YYRD/DAA-FYGD     | 29 (4.56)| 6 (0.62)   | 0.14 | -4.44| 9.18E-06 |
| 21 Rare Genotypes (<5 copies) | 77 (12.11)| 35 (3.64) | 0.30 | -5.90| 3.69E-09 |

Table 3. Genotypic association analysis of type 1 diabetes with DQA1-DQB1 motifs of (αa1, α44, α157, α196, β9, β30, β57, β70, β135): estimated frequency (%) among control and patient, estimated odds ratios, Z-score, and p-value, corresponding serotype, and equivalent allelic groups. Listed are genotypes with p-values less than 0.05. Genotypic association results for all genotypes are listed in Table S2. The logistic regression analysis used the offset to compute odds ratio with the "virtual genotype" under the null as the reference.
contribution of the motif “DQAA-YYARD” to T1D, with a significant association with susceptibility to T1D (OR = 4.07, p = 5.14×10⁻⁴). The homozygote had much elevated association with T1D susceptibility (OR = 21.99, p = 9.64×10⁻⁹). The heterozygote with another DQ8.1 “DQDA-YYARD” was also highly associated with increased susceptibility (OR = 22.67, p = 3.99×10⁻³). Last, the heterozygote with resistant motifs had much weaker association (OR = 2.13, p = 0.014), because resistant motifs compromised some of disease associations but not to the same extent (ORs from 0.03 to 0.35, Table 2).

Examining in the same manner another DQ8.1 motif, motif 10, i.e. “DQDA-YYARD”, its heterozygote with neutral motifs had significant association with T1D susceptibility (OR = 14.73, p = 0.0142), which captured its sole contribution. There were no homozygotes in this relatively large study population, since the expected genotype frequency is in the order of 1.44×10⁻³. The heterozygote with resistant motifs showed insignificant association with T1D risk (OR = 2.04, p = 0.18), probably due to relatively low genotypic frequency.

By estimated odds ratios, the heterozygote “DCAA-YSARD/DQAA-YYARD” had an OR of 14.52, which was slightly greater than the product of individual contributions 10.34 (= 2.54 × 4.07) from “DCAA-YSARD/neutral” and “DQAA-YYARD/neutral”, respectively. Similarly, the heterozygote “DCAA-YSARD/DQDA-YYARD” had an OR of 24.93, which was slightly less than the product of individual contributions 37.41 (= 2.54 × 14.73) from “DCAA-YSARD/neutral” and “DQDA-YYARD/neutral”.

Expression in trans of DQA1 and DQB1 motifs enriches cellular diversity of DQ heterodimers but has limited associations with T1D. Heterodimeric formation of the DQ molecule with doubly heterozygous DQA1 and DQB1 genotypes could produce four different DQ molecules: two cis-formations with alleles on the same chromosome and two trans-formations across homologous chromosomes, which may not be observed experimentally, as some are structurally prohibited18 (see also Fig. 1). For homozygotes or semi-homozygotes, DQA1-B1 motifs on different chromosomes, formed trans-DQ motifs (whenever structurally allowed) that were the same as those cis-DQ motifs. In total, in trans-formation created 73 DQ motifs, of which 48 were new -DQ motifs, several of which were structurally forbidden18–20 (see Supplementary Table S3). In a systematic assessment of all trans-motifs, we evaluated potential conditional association of each trans-motif with T1D, after adjusting for all cis-motifs. After excluding trans-motifs with fewer than five copies, we reported conditional associations with 45 trans-motifs. Most of trans-motifs did not exhibit any meaningful adjusted associations, with a few exceptions, e.g., “DQDA-YYARD” (p = 0.01) (see Supplementary Table S3). The many permitted trans-motifs are indicated in the column “Prmt”. In the table, estimated coefficient quantifies adjusted log odds ratio for cis-motifs, together with estimated standard error, Z-score and p-value. Should there be an interest in marginally estimated odds ratio, it was possible to take the ratio of patient and control trans-motif frequencies. For example, the trans-motif 24 “DQAA-YYARD” had the marginal odds ratio of 8.37, but the adjusted odds ratio was 1.26, which was not significantly different from null (p-value = 0.387), because associated cis-motifs explained most of the risk associations. In summary, trans-motifs did not appear to exhibit additional associations, after accounting for all 25 cis-motifs. However, this empirical result from the conditional adjust-

| First Motif     | Second Motif     | Control | patient | Coef | OR  | SE  | Z      | p      |
|-----------------|------------------|---------|---------|------|-----|-----|--------|--------|
| DCAA-YSARD      | Neutral          | 21 (3.3)| 47 (4.89)| 0.93 | 2.54| 0.44| 2.11   | 3.48E-02|
| (DQ2.5)         | DCAA-YSARD       | 10 (1.57)| 46 (4.78)| 1.65 | 5.21| 0.50| 3.32   | 8.97E-04|
| DQAA-YYARD      | 21 (3.3)         | 269 (27.96)| 2.68 | 14.52| 0.42| 6.36 | 1.99E-10|
| DQDA-YYARD      | 1 (0.16)         | 22 (2.29)| 3.22 | 24.93| 1.08| 2.97 | 2.95E-03|
| (DQ8.1)         | Resistant        | 87 (13.68)| 53 (5.51)| -0.37| 0.69| 0.39| -0.94  | 3.48E-01|
| DQAA-YYARD      | Neutral          | 34 (5.35)| 122 (12.68)| 1.40| 4.07| 0.40| 3.47   | 5.14E-04|
| (DQ8.1)         | DQAA-YYARD       | 5 (0.79)| 97 (10.08)| 3.09| 21.99| 0.58| 5.33   | 9.64E-08|
| DQDA-YYARD      | 1 (0.16)         | 20 (2.08)| 3.12 | 22.67| 1.08| 2.88 | 3.99E-03|
| (DQ8.1)         | Resistant        | 86 (13.52)| 162 (16.84)| 0.76| 2.13| 0.38| 2.00   | 4.51E-02|
| DQDA-YYARD      | Neutral          | 1 (0.16) | 13 (1.35)| 2.69 | 14.73| 1.10| 2.45   | 1.42E-02|
| (DQ8.1)         | DQDA-YYARD       | 0 (0)  | 0 (0)   |      |     |     |        |        |
| Neutral         | Neutral          | 17 (2.67)| 15 (1.56)| -0.13| 0.88| 0.35| -0.35  | 7.24E-01|

Table 4. Results from the univariate association analysis for three high risk DQ motifs (DQ2.5: DCAA-YSARD, DQ8.1: DQAA-YYARD and DQ8.1: DQDA-YYARD), after accounting for all neutral and resistant motifs. Numbers and frequency percentages of genotype groups are listed for controls and patients, and so are estimated coef (log odds ratio), standard error, Z score, and p-value. The reference genotype is the specific genotype versus all others.
ment did not exclude their potential biological associations between trans DQ motifs and T1D as the cellular expression of DQ heterodimers are enriched.

DQ associations may be influenced by its linkage-disequilibrium with DRB1. DQA1 and DQB1 were genetically nearby DRB1, and in high LD with each other, i.e., DR3 and DQ2.5 were linked, and so were DR4 and DQ8.1. With DQ motifs of nine residues, we performed haplotype association analysis of DQ motifs and DRB1, with six DR4 subtypes specified by four-digit resolution. Clearly, DR3 was uniquely linked with the motif “DCAA-YSARD” (DQ2.5), and their joint haplotype had significant association with T1D susceptibility (OR = 2.10, p = 1.96*10^-20) (Table 5). Similarly, DRB1*04:02, *04:03, *04:04, *04:05 and *04:07 were linked DQ motifs with variable associations. Hence, these empirical haplotype associations prohibited separating DRB1 and DQ associations. However, the same DRB1*04:01 was on the same haplotypes with DQ motifs “DQAA-YYARD”, “DQDA-YYARD”, or “DQDA-YYDRD”. These three haplotypes had variable associations (OR = 5.52, 4.25 and 0.89, p = 4.94*10^-70, 7.98*10^-5 and 0.57, respectively), implying that DQ motifs dictated T1D association in the background of DRB1*04:01. DRB1*04:03 reduced T1D susceptibility when in LD with DQ8.1 (OR = 0.12, while for the DQ8.1 alleles collectively OR = 3.34) even though the number of patients and controls with this haplotype was very low. In the case of the DRB1*15 molecule, the OR is identical to the very tightly linked DQA1*01:02-B1*06:02 haplotype. In fact, the number of patients and controls is nearly identical concerning two nearly congruent groups.

Motifs of nine residues account for all DQ associations with T1D. Besides selected residues on alpha and beta chains, there were 95 polymorphic amino acids across DQ haplotypes in this study population. Due to LD, many residues were expected to have significant marginal associations. We evaluated these marginal associations via a logistic regression of T1D with polymorphic residues. By the log-likelihood ratio test, p-values were computed for each residue, and were shown in the top panel of the bar plot in the Fig. 2. Evidently, many p-values greatly exceeded the threshold of 0.001, and some were approaching 10^-90 or even less, showing strong marginal associations with T1D. By adjusting for the motifs of selected nine residues, we repeated the same logistic regression analysis, estimating p-values from log-likelihood ratio test (the lower bar plot of the Fig. 2). Indeed, most p-values were insignificant, i.e., p > 0.05, except that several p-values were slightly less than 0.05.

| ID | DR  | DQ          | Control | Patient | OR   | Z     | p        |
|----|-----|-------------|---------|---------|------|-------|----------|
| 1  | *01 | DRAA-YHVGD  | 136(10.69)| 173(8.99)| 0.84 | -1.59 | 1.11E-01 |
| 2  | *03 | DCAA-YSARD  | 153(12.03)| 485(25.21)| 2.10 | 9.26  | 1.96E-20 |
| 3  | *04:01| DQAA-YYARD  | 71(5.58 )| 593(30.82)| 5.52 | 17.69 | 4.94E-70 |
| 4  | *04:01| DQDA-YYARD  | 7(0.55  )| 45(2.34 )| 4.25 | 3.94  | 7.98E-05 |
| 5  | *04:01| DQDA-YYDRD  | 41(3.22 )| 55(2.86 )| 0.89 | -0.57 | 5.67E-01 |
| 6  | *04:02| DQAA-YYARD  | 7(0.55  )| 27(1.40 )| 2.55 | 2.25  | 4.74E-02 |
| 7  | *04:03| DQAA-YYARD  | 11(0.86 )| 2(1.00  )| 0.12 | -3.31 | 9.18E-04 |
| 8  | *04:04| DQAA-YYARD  | 61(4.8  )| 146(7.59)| 1.58 | 3.10  | 1.91E-03 |
| 9  | *04:05| DQDA-YYARD  | 3(0.24  )| 23(1.20 )| 5.07 | 2.97  | 2.99E-03 |
| 10 | *04:07| DQDA-YYDRD  | 9(0.71  )| 1(0.05  )| 0.07 | -3.28 | 1.05E-03 |
| 11 | *07  | DKAAYSARGD  | 78(6.13 )| 41(2.13 )| 0.35 | -5.76 | 8.47E-09 |
| 12 | *07  | DKAAYYDRD  | 36(2.83 )| 5(0.26  )| 0.09 | -6.36 | 1.99E-10 |
| 13 | *08  | DCAA-FYDED  | 55(4.32 )| 76(3.95 )| 0.91 | -0.52 | 6.00E-01 |
| 14 | *09  | DQDA-YYDRD  | 10(0.79 )| 28(1.46 )| 1.85 | 1.72  | 8.57E-02 |
| 15 | *11  | DCAA-YYDRD  | 89(7.00 )| 20(1.04 )| 0.15 | -8.99 | 2.57E-19 |
| 16 | *12  | DCAA-YYDRD  | 29(2.28 )| 14(0.73 )| 0.32 | -3.67 | 2.45E-04 |
| 17 | *13  | DCAA-YYDRD  | 11(0.87 )| 5(0.26  )| 0.30 | -2.38 | 1.73E-02 |
| 18 | *13  | DRAA-YHGDG  | 91(7.15 )| 39(2.03 )| 0.28 | -7.28 | 3.88E-13 |
| 19 | *13  | DRAA-YHVRD  | 52(4.09 )| 80(4.16 )| 1.02 | 0.10  | 9.22E-01 |
| 20 | *14  | DRAA-YHGDG  | 20(1.60 )| 4(0.21  )| 0.13 | -4.46 | 8.09E-06 |
| 21 | *15  | DRAA-FYGDG  | 197(15.49)| 8(0.42 )| 0.03 | -16.24| 2.49E-59 |
| 22 | *16  | DRAA-YHSGD  | 14(1.10 )| 13(0.68 )| 0.61 | -1.29 | 1.97E-01 |

Table 5. Haplotype association analysis of type 1 diabetes with HLA-DRB1 and DQA1/B1: estimated haplotype frequency (%) among control and patient, estimated odds ratio, Z-score, and p-value, for those haplotypes with five or more copies.
We plotted adjusted p-values against marginal p-values as an x–y plot, showing that adjusting motifs accounted for most of the marginal associations (Fig. 2).

**DQ motifs are less diverse among patients than among controls.**  DQ motifs of selected nine residues retained all susceptibility and resistance to T1D associations that were identified by DQ haplotypes but had reduced polymorphisms due to merging of some DQ haplotypes, especially those rare DQ haplotypes with relatively common ones. The frequencies of 25 DQ motifs among controls and patients (Fig. 3) indicates that the two motifs “DCAA-YSARD” and “DQAA-YYARD”, corresponding to DQ2.5 and DQ8.1, had exceptionally higher frequencies among patients than controls. Further, controls tended to have modest frequencies across many motifs. For this reason, DQ motifs among controls, with a Shannon's entropy value of 2.50, tended to be more diverse than those among patients with an entropy value of 1.85, where the entropy was used to quantify diversity of genetic polymorphisms.

**Motifs of nine residues capture DQ associations with islet autoantibodies.**  Our study measured six autoantibodies among newly diagnosed patients, i.e., IAA, GADA, IA-2A, ZnT8RA, ZnT8WA and ZnT8QA, which were scored as positive or negative. Focusing on 11 DQ motifs observed more than 10 times among patients, we evaluated the association of each DQ motif with the positivity for each type of autoantibody (Table 6). T1D subjects with the motif “DCAA-YSARD”, corresponding to DQ2.5, tended to have an increased...
frequency of GADA (OR = 1.56; p = 6.35 \times 10^{-8}) but a depressed frequency of IA-2A (OR = 0.59, p = 6.55 \times 10^{-11}). Among subjects with the motif “DQAA-YYARD” corresponding to DQ8.1, patients had a reversed association pattern, i.e., depressed GADA (OR = 0.87; p = 4.50 \times 10^{-3}) but elevated IA-2A (OR = 1.52, p = 2.18 \times 10^{-9}) frequency. Importantly, subjects with the motif “DQDA-YYARD”, corresponding to a slightly different DQ8.1 allele, had an elevated IA-2A positivity (OR = 5.67, p = 5.42 \times 10^{-4}), while its association with GADA was neutral (OR = 0.91, p = 0.67). Lastly, the motif “DRAA-YHVRD” corresponding to DQ6.4 (itself a neutral allele regarding T1D susceptibility) seemed to associate with increased frequency of ZnT8A (RA, WA and QA; OR = 1.92, 1.72 and 1.57, p = 6.25 \times 10^{-3}, 0.017 and 0.041, respectively), while having negative associations with GADA and IA-2A frequencies. In order to delineate possible pathways for T1D manifestation and resistance via islet-autoantigen-specific epitope presentation restricted by susceptible and resistant motifs, respectively, we have chosen three examples of: a susceptible (DQ8—InsB11-23), a neutral (DQB1*06:04—InsB5-15) and a resistant (DQB1*06:02—InsB5-15) heterodimer (Fig. 4A–C, and Supplementary Fig. S3A-C).
To quantify an overall T1D risk or protection from DQ motifs, we computed attributable risk and risk fractions, respectively (methodological details of which are given in “Methods”). Briefly, we pooled DQ motifs into three groups: resistant, neutral or risk. Each individual had pairs of resistant, neutral or risk motifs. Using a logistic regression model on genotypes, we estimated regression coefficients corresponding to five genotypes, treating homozygous neutral motifs as the reference (Table 7). Motif frequencies were estimated from the controls (Table 7). Motif frequencies were estimated from the controls (Table 7).

Estimated attributable fractions of risk and protection from respectively susceptibility and resistance motifs. To quantify an overall T1D risk or protection from DQ motifs, we computed attributable risk and risk fraction, respectively (methodological details of which are given in “Methods”).

Table 6. Association analysis of DQ motifs with elevated autoantibody levels among all patients (with five or more subjects): estimated odds ratio, haplotype score, and p-value, across all motifs with six autoantibodies (IAA, GADA, IA2A, ZnT8RA, ZnT8WA and ZnT8WA).
four alpha residues (aa1, a44, a157, a196) and five beta residues (b9, b30, b57, b70, b135) forming motifs that was associated with either susceptibility or resistance to T1D. The importance of both DQα and DQβ chains in determining T1D susceptibility and resistance is demonstrated not just in the respective motif consisting of 4 α-chain and 5 β-chain residues but also in all DQαβ combinations revealed by next generation targeted sequencing. The deep sequencing approach reveals the importance of HLA-DQA-B combinations in T1D both susceptibility and resistance to considerations of –DQB or –DQA alone. This is evident, for example, in motif 5 (DQA1*02:01-B1*02:02), a T1D mildly resistant allele compared to the rare motif 9 (DQA1*03:02-B1*02:02) with only six copies among T1D patients, despite their sharing of an identical β-chain.

When interpreting association results of nine identified residues with respect to their binding properties, we should be mindful of two residue sets: one set includes all monomorphic and conserved residues, and another set includes those monomorphic residues in their respective clusters of DQ haplotypes. The former set of residues essentially defines the overall DQ molecule structure, and the latter set defines substructures of DQ heterodimers, e.g., DQ2.5 or DQ8.1 heterodimers. Hence, they remain structurally and functionally important, even though they may have limited associations with T1D risk.

Specifically, the newly identified DQB1*02:02, previously scored as DQBI*02:01 with low resolution methods, was found in a T1D-resistant combination (motif 5, Table 2, DQA1*02:01-B1*02:02) in 78 controls and 41 patients. The same β-chain by contrast, forms a rare haplotype in combination with a different α-chain, i.e., DQA1*03:02-B1*02:02, found only in six patients but in no controls. Similar striking findings were observed with DQB1*03:01 and DQB1*03:03. When either of them pairs with DQA1*03:01 they form rare haplotypes found only in 9 controls (motif 8, Table 2). Yet when the first of them pairs with DQA1*03:03 and the second with DQA1*03:02 they result in neutral haplotypes, found in 67 controls and 88 patients (motif 11, Table 2).

We also demonstrate that the DQ motifs associated with T1D appeared to be mediated through the frequency of islet autoantibodies. In particular, subjects with the DQA-DQAA-YYARD motif, more often had GADA (p = 6.35*10−10) than IA-2A (p = 6.55*10−11). Meanwhile, subjects with DQ*DQAA-YYARD and DQ*DQDA-YYARD, corresponding to DQ8.1, more often had IA-2A (p = 2.18*10−9 and 5.42*10−4, respectively), but less often GADA (p = 4.50*10−3). Other motifs had more variable associations with the different autoantibodies. To assess motif associations with the number of autoantibodies at diagnosis, we compared patients with 2, 3 or 4 autoantibodies with patients with 0 or 1 autoantibody, which revealed a single significant association (see Supplementary Table S4). In this neutral/mildly resistance motif, DQ*DQDA-YYDRD, the association with the number of autoantibodies was significant only when all four autoantibodies were present. These data suggest that the presence of all four autoantibodies overcame the resistance mechanism(s) of the DQ7 and DQ9 haplotypes (both β57Asp+) represented in this motif29. Hence, it cannot be excluded that other non-HLA genetic factors important to the pathogenesis contributed to autoantigen spreading and appearance of multiple autoantibodies.

It is important to realize that antigen-specificity in HLA-DR- and -DQ-directed autoimmune response (as well as resistance to it) is paramount for an organ-specific autoimmune disease, such as T1D. This is reflected by the observation that 4 of the 9 selected HLA-DQ residues are in anchoring pockets involved directly in antigen binding (b9, b30, b57, b70), another one (aa1) is near pocket 1, and two of the first four residues (b57, b70) in probable cognate TCR engagement and activation30,31,32. The remaining polymorphic residues (except aα1 about which there is hardly any information), are involved in other accessory functions of the HLA-DQ heterodimer: a135 in CD4 co-receptor binding, as part of the β134-148 stretch32,33, a157 in the formation of the putative homodimer of heterodimers that presumably facilitates CD4 binding34,35 and consequent cognate TCR activation and HLA-DQ signal transduction in the specific T cell and APC respectively36. The remaining residues (except aα1) about which there is hardly any information), are involved in other accessory functions of the HLA-DQ heterodimer: a135 in CD4 co-receptor binding, as part of the β134-148 stretch32,33, a157 in the formation of the putative homodimer of heterodimers that presumably facilitates CD4 binding34,35 and consequent cognate TCR activation and HLA-DQ signal transduction in the specific T cell and APC respectively36. The remaining residues (except aα1) about which there is hardly any information), are involved in other accessory functions of the HLA-DQ heterodimer: a135 in CD4 co-receptor binding, as part of the β134-148 stretch32,33, a157 in the formation of the putative homodimer of heterodimers that presumably facilitates CD4 binding34,35 and consequent cognate TCR activation and HLA-DQ signal transduction in the specific T cell and APC respectively36. The remaining residues (except aα1) about which there is hardly any information), are involved in other accessory functions of the HLA-DQ heterodimer: a135 in CD4 co-receptor binding, as part of the β134-148 stretch32,33, a157 in the formation of the putative homodimer of heterodimers that presumably facilitates CD4 binding34,35 and consequent cognate TCR activation and HLA-DQ signal transduction in the specific T cell and APC respectively36.
and T regulatory cells (Tregs) were present at higher frequencies in the peripheral blood of HLA-DRB1*15:01-DQA1*01:02-B1*06:02 positive individuals (i.e. negatively associated with T1D) compared to subjects with neutral or T1D-susceptible alleles44. Furthermore, cloned Tregs carrying these specificities and restricted to HLA-DRB1*15:01 suppressed so-restricted and specific Teffectors11 but will be of interest to determine whether HLA-DQ alleles linked with resistance to T1D can also give rise to similar Tregs. A similar effect has already been noted in the case of Goodpasture’s syndrome with higher frequency of Tregs specific for the Collagen IV autoantigen α3(141-145) in HLA-DRB1*01:01+ (disease-resistant) individuals, compared to HLA-DRB1*15:01+ (disease-susceptible) individuals, regarding the same autoantigen specificity, albeit with binding register shifted by one residue45. This would extend the “epitope-stealing” effect, long hypothesized for T1D and already documented for HLA-DQA1*01:02-B1*06:02 over HLA-DQ8-specific proinsulin epitopes23,37,42. If DQA1*01:02-B1*06:02-restricted Tregs can be shown to have similar suppressive effects as those restricted to DRB1*15:01 above, then the contributions of the former molecule to T1D resistance would be multiple. The amino acid motifs identified in the present study should prove useful to identify these mechanisms.

All our considerations thus far, have assumed equal levels of membrane HLA-DQ protein expression on APCs. It was shown over two decades ago, that the density of APC, the level of expression of HLA-DQ on the membrane of such APC, as well as the level of soluble peptide antigen present, would influence the proliferation of restricted and specific CD4 + T effector cells46. While the upstream regions of HLA-DQα and –DQB genes are replete with regulatory nucleotide sequences, very few studies exist on the regulation of HLA-DQ membrane protein expression by the various different alleles45,46. Notwithstanding the fact that in a disease situation (i.e. the islet β cells under autoimmune attack, as well as the proximal pancreatic lymph nodes) the cytokine and chemokine milieu is an unknown and may differentially influence the expression of HLA-DR-DQ heterodimers even in an allele-specific manner. The situation has become more complex by the fact that human, as well as mouse, islet β cells, are closely approximated by islet macrophages that present peptides to these β cells and possess MHCII molecules through which such antigens are presented47,48. It is also of interest that Tregs, shown to be suppressive over Teffector cells specific for T1D autoantigens, seem to be consistently absent from islets in autopsy material from newly diagnosed T1D patients49,50.

We have limited our discussion to possible autoantigenic epitopes and their interactions with HLA-DQ molecules and alleles that confer susceptibility or resistance to, or are neutral to T1D. Unfortunately nothing can be said about the interaction of diabetogenic pHLA-DQ heterodimers with cognate TCRs, as there is no single complex with solved 3-dimensional coordinates. Such complex would provide information about the possible engagement of the amino acid residue motifs identified in the present study. Using TCR—HLA-DR-DQ complexes with CNS antigenic epitopes in multiple sclerosis as a guide (the only available such complexes from organ-specific autoimmune diseases), the TCR orientation with respect to pMHCII is not the canonical diagonal, but expected to be off-diagonal49. The fact that this interaction, that leads to autoimmunity, is different from the canonical one is also reflected in the irregular form of immunological synapse formed by autoimmune CD4+ T cell clones from patients with multiple sclerosis and T1D, compared to such clones specific for influenza virus peptides: there was no accumulation of pMHCII complexes in the central Supramolecular Activation Cluster of the synapse; nor for that matter, any transport of TCR–pMHCII complexes into this central cluster of the synapse50. In another example, the key HLA class II molecule accounting for over 90% of patients with celiac disease is DQ5.2 (A1*05:01-B1*02:01), while the very closely related DQ2.2 (A1*02:01-B1*02:02) is neutral. The a22/F dimorphism (DQ2.2 compared to DQ2.2, respectively), differentiate between the binding at p3 of gluten peptides in DQ2.2 compared to DQ2.5. These clustered combinations set for a very focused set of cognate TCRs that are also pathogenic for self resulting in small intestinal tissue destruction51. The small set of gluten peptide-DQ2.2 complexes surprisingly resulted in a wide set of cognate TCRs, none of which were pathogenic51.

A potential weakness to the present studies is that all children are Swedish, only 8% in the present cohort were born to parents born in another country along with the grandparents51 and that the environmental exposure would be relatively homogeneous. It cannot be excluded that the pattern of amino acid residues observed as well as the presence of autoantibodies at the time of clinical onset may reflect a Swedish environment. The exposures of these children diagnosed with T1D are not only long term but also multiplex as the incidence rate for IAA-first in DR4-DQ8 children showed a peak at 1–3 years of age in the TEDDY study as related to prolonged shedding of enterovirus40,53,54 and other studies have shown that viral exposure of islet autoantibody positive children may accelerate progression to clinical onset51. It cannot be excluded that the two potential endotypes in T1D etiology and multiple environmental exposures leading to the eventual clinical onset of T1D have been important to the frequency of specific amino acid motifs.

Three additional limitations of the study are worth noting. The current study uses the retrospective case–control approach, drawing patients from clinics around Sweden and geographically matched controls. Inevitably, hospital-referred patterns may bias patient selection, and some controls, if any, may have not developed T1D at the time of the study. However, both issues are unlikely to alter the results of assessing the genetic susceptibility to T1D, because HLA genes are time-invariant, T1D is relatively uncommon among adults and healthcare system provides universal care in Sweden. Another limitation is associated with measuring autoantibodies among patients at the time of diagnosis. As noted above, autoantibodies are important biomarkers predicting T1D onset54,56 but have variable temporal patterns. In the absence of observing such temporalities, the current investigation of genetic associations with autoantibodies is inevitably limited and needs to be cautiously interpreted, even though observed associations still remain meaningful. The third limitation concerns the possibility of genetic heterogeneity due to immigrants that may confound the association results. In the BDD study, 8% of patients are children born to immigrants56,69. Even though immigrant status information is not controlled, it is expected that normal controls may have a comparable percentage of immigrants. Given the relatively small percentage, the mixture of immigrants with native Sweden is not expected to substantially alter association results observed here.
Recently, a novel reductionist conditional approach was used to search for critical residues. The key idea is to repeat residue-specific association tests, while adjusting (conditioning on) selected residues. Their conditional analysis led to identification of β57 as a critical residue for T1D. However, this approach fails to differentiate between empirically observed associations that result from genetic causal associations or from LD-induced associations. Hence, their results have invited much controversy. In contrast to the reductionist approach, our approach, being a holistic, starts with motifs of all polymorphic residues, retaining their sequence structure, and eliminates those residues that do not associate with T1D within their respective clusters. Consequently, our primary results concern T1D associations with motifs of selected residues, and these motifs explain nearly all DQ haplotype associations with T1D. By the same reason, this holistic approach is probably preferred to other residue-specific association methods.

In T1D, recent data in the TEDDY study suggest that prolonged shedding of enterovirus B in young DR4-DQ8.1 positive children precede the appearance of insulin autoantibodies (IAA) as a biomarker for autoimmunity against the beta cells. In older children, prolonged shedding of adenovirus F in DR3-DQ2.5 positive children preceded the appearance of GAD65 autoantibodies (GADA) as a first biomarker for beta cell autoimmunity. The intersection between prolonged virus exposure and appearance of a first autoantibody marker is yet to be dissected. However, it cannot be excluded that the motifs described in the present paper contribute to antigen presentation that involves a competition between a viral peptide or either (pro)insulin or GAD65 to trigger an autoimmune rather than a viral immune response. It is therefore of interest that the susceptibility motifs were more diverse among controls than among patients (entropy = 2.5 versus 1.85, respectively). Out of 25 unique motifs, three motifs "DCAA-YSARD," "DQA1-YYARD," and "DQDA-YYARD" exhibited significant susceptibility to T1D (OR = 2.10, 3.34 and 3.71, p = 1.96*10^-20, 2.69*10^-15 and 1.53*10^-13, respectively) with one uncommon motif "DQDA-YSARG" was observed seven times only among patients. On the other hand, eight motifs "DCAA-YYDRD," "DKAA-YSARG," "DCAA-YYDRD," "DKAA-YYDRD," "DCAA-FYDGD," "DAA-LYDRD," "DRE-ROMGD," and "DQAA-RYGDRD" were negatively associated with autoimmunity against β cells resulting in T1D (OR = 0.19, 0.35, 0.09, 0.07, 0.25, 0.80 and 0.08, p = 2.32*10^-25, 8.47*10^-9, 1.99*10^-10, 2.25*10^-10, 1.14*10^-3, 7.72*10^-19, 0.0312 and 6.38*10^-3, respectively), with two motifs "DCAA-YYDRD" and "DQAA-YYDRD" respectively observed 7 and 9 times among only controls. It is obvious that the lower HLA-DQ motif entropy among T1D patients reflects the fact that susceptibility to T1D requires certain structural features on the part of the susceptibility-conferring HLA-DQ molecules. These certainly would have to do with the relevant autoantigenic epitopes, as well as the molecules that the T1D-causing pMHCII complexes interact with such cognate TCRs and CD4 co-stimulatory molecules, and perhaps other molecules in the respective signal transducing pathways.

**Methods**

**Study design and populations.** We used a case–control study design to evaluate genetic associations with T1D. A total of 962 patients (cases) were from the nation-wide Swedish Better Diabetes Diagnosis (BDD) study, which involves participation in 2005–2010 of 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. Included patients had one or several autoantibodies against either insulin (IAA), GAD65 (GADA), IA-2 (IA-2A), and three variants (amino acid 325 being either R, W or Q, ZnT8-RA, ZnT8-WA or ZnT8-QA, respectively) at the time of clinical diagnosis. Patients diagnosed with diabetes at 9 months–18 years of age were sequentially enrolled in the BDD study. A total of 636 nation-wide and geographically representative samples genotyped with the present NGS approach. To date these tools have been tested—with 100% accuracy—on > 2000 control samples. Analytical tools to define haplotypes and genotypes were made available as part of the genotyping kit assay from Scisco Genetics Inc. (Seattle, WA). To date these tools have been tested—with 100% accuracy—on > 2000 control samples genotyped with the present NGS approach. All experiments (DNA extraction, genotyping, measuring islet autoantibodies) and analytic methods are carried out in clinical laboratories (CLA certified or equivalent in Sweden), following relevant guidelines and regulations.

**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.

**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. Briefly, the laboratory steps consisted of consecutive amplicon-based PCR 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 purchased from Scisco Genetics Inc., Seattle WA (https://sciscogenetics.com). This system employs amplicons individually extending 4–500 bp segments covering each of exons 1 ~ 4 of HLA-DQA1 and -DQB1 and MiSeq read depths of over 100-fold coverage for each amplicon were obtained. The analytical tools to define haplotypes and genotypes were made available as part of the genotyping kit assay from Scisco Genetics Inc. (Seattle, WA). Approximately 8% of BDD participants are immigrants. The informed consent was taken from participants/guardians for young participants (< 18 years old). The Karolinska Institute Ethics Board approved the BDD study (2004/19).

As NGTS sequenced DNA nucleotides of selected exons and used sequences to infer DQA1 and DQB1 alleles, we determined HLA-DQA1 and -DQB1 genotypes 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). Further, we determined physical positions of individual aminoacids in the alpha chain from codon a1 to a232, and in the beta chain from codon β1 to β237.
We have adopted the numbering system as first suggested\(^{25}\), and later improved\(^{20}\), that essentially allows 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-DR\(^{134,66}\).

**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\(^{10,38}\).

**Cis- and trans-heterodimer through interlocus recombination.** High LD enables us to haplotype HLA-DQA1 and DQB1 genotypes with exceptionally high posterior probabilities (minimum value > 0.97). Let \((h_1/b) = (\hat{a}_1\hat{a}_8/\hat{a}_1\hat{a}_8)\) denote a pair of -DQA1 and -DQB1 haplotypes, in which \(\hat{a}_1\hat{a}_8\) and \(\hat{a}_1\hat{a}_8\) are two cis-haplotypes (or cis-heterodimers). In contrast, two trans-haplotypes are \(\hat{a}_1\hat{a}_8\) and \(\hat{a}_1\hat{a}_8\). In case that -DQA1 and -DQB1 are doubly homozygous, all four haplotypes are identical. For a semi-heterozygous, two trans-haplotypes are identical to two cis-haplotypes. Thirdly, a doubly heterozygous DQA1 and DQB1 would lead to four different haplotypes. Under the assumption that the "interlocus recombination" is a completely random process, alleles of DQA1 and DQB1 are always recombined to form cis- and trans-heterodimers within cells, even though some trans-heterodimers may be structurally prohibited or unexpressed\(^{18}\). Lastly, it is expected that some trans-heterodimers, from doubly heterozygotes, are novel and unobserved in the general population, while many others could be observed in the general population. There are still some that were probably expressed at the mRNA level but were not permitted to form proteins\(^{18}\).

**Conditional analysis strategy.** In the empirical association analysis of T1D disease outcome \((Y)\) with one primary covariable \(X\) in light of another correlated variable \(Z\) that may potentially confound the association of interest, the conventional approach is to adjust the confounding variable in the association analysis via the following logistic regression model for modeling the disease probability:

\[
\Pr(Y = 1|X, Z) = 1/[1 + \exp(-\text{adjust}(X, \beta) - \text{test}(Z, \gamma))],
\]

in which the first component “\(\text{adjust}(X, \beta)\)” is a function of the primary covariables to be adjusted for and is indexed by the parameter \(\beta\), and the second component “\(\text{test}(Z, \gamma)\)” is a function of the test covariables indexed by the parameter \(\gamma\). Here the first step is to fit a logistic regression model with the “\(\text{adjust}(X, \beta)\)”. Then fixing this component, the second step is to systematically test the disease association with the second component “\(\text{test}(Z, \gamma)\)”. Despite the intuitive appeal of the conditional analysis for adjusting the primary variable, it also has several intrinsic limitations, since this adjustment is empirical rather than not biological (see “Discussion”). Nevertheless, it is helpful to perform such a conditional analysis for the following scenarios:

1. Assessing disease associations with trans-motifs, after adjusting for the cis-motifs present.
2. Assessing disease associations with individual residues, after adjusting for cis-motifs

**Shannon’s entropy.** HLA-DQ motifs are polymorphic, and have a degree of diversity. Because of their categorical nature, we use the information entropy, a basic information unit in the information theory, which is also known as Shannon's entropy introduced by Claude Shannon\(^{21}\). Let \(f_1, f_2, \ldots, f_q\) denote allelic frequencies for \(q\) alleles. The corresponding Shannon's entropy is defined as

\[
\text{Entropy} = -\sum_{j=1}^{q} f_j \log(f_j).
\]

**Attributable risk/protection fractions.** Attributable fraction is used to quantify an overall fraction of risk (protection) that can be accounted by a binary exposure factor, i.e., exposed versus non-exposed, and depends on the magnitude of the association and frequency of the exposure in the population\(^{46}\) which can be computed as:

\[
\text{AF}_{jk} = \frac{f_j(\text{OR}_{jk} - 1)}{1 + f_j(\text{OR}_{jk} - 1)},
\]

which may be referred to as attributable risk fraction (AF > 0) or attributable protective fraction (AF < 0).

**Structural properties of HLA-DQ residues.** 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, https://www.3ds.com/products-services/biovia), based on coordinates freely available in the Protein Data Bank. The DQ8 (A*03:01-B1*03:02)—insulin B11-23 complex structure is obtained from the determined crystal structure\(^{22}\). The modeled structures of the complexes of DQB1*06:02 (A1*01:02-B1*06:02)—insulin B4-16, and DQB1*06:04 (A1*01:02-B1*06:04)—insulin B4-16 are obtained from the crystal structure of DQB602-hypocretin complex, as already described\(^{21}\).
Statistical analysis and software. Upon extracting aminoacids for each DQ haplotype, we compute distances between all unique sequences pairwise by the stringdist function of R package, with the Levenshtein distance, and then apply the hierarchical clustering algorithm using the hclust function of R package (version 4.0.2, https://www.r-project.org) with the agglomeration method of ward.D2, i.e., clustering two sequences closer together, if their sum of squared distances is relatively small. For the association analysis, we use the logistic regression function in "glm" of R package. To assess their associations with qualitatively determined elevations of islet autoantibodies (IAA, GADA, IA-2A, ZnT8RA, ZnT8WA, and ZnT8QA), we also use haplo.GE. To address the instability of "small sample sizes" in the diploitic association analysis, we compare frequency counts of one diplotype versus those of all other diplotypes combined, and compute the corresponding odds ratios, 95% confidence interval, and the Fisher's exact test p-values.

With respect to p-values and multiple comparison issues, we took an unconventional approach, presenting the p-values without the multiple comparison correction. While the conventional approach was to compute corrected p-values for number of comparisons by, say, the Bonferroni correction or false discovery rate, its intention was to control the overall false positive error rates of all comparisons. In the current context with the analysis of HLA-DQ haplotypes/diplotypes, the roles of both genes in T1D were implicated in multiple empirical and functional studies. Further, the empirical haplotype association analysis suggested that DQ associations are highly significant at the 5% level, even if multiple comparisons with 45 comparisons are considered. Hence, the haplotype and diploitic 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 have clear and simple interpretations, free from varying haplotypes/diplotypes in various explorations. Throughout the 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. Naive applications of any multiple comparisons without varying correction factors could obscure the interpretations of results.

When computing haplotype-specific odds ratios, we intentionally avoided choosing a specific reference haplotype, because any choice of a particular reference haplotype could have undesired implications in the interpretations of the results. Instead, we considered two different strategies. One way was to compute the odds ratio of one haplotype versus all other haplotypes. Alternatively, we choose a "virtual haplotype" that has no association with T1D. As expected, the ratio of its corresponding haplotype frequency among patients over that among controls equals one. Hence, the expected odds ratio for any given haplotype is thus computed as the ratio of two observed haplotype frequencies among patients and controls

Received: 11 September 2020; Accepted: 4 March 2021
Published online: 23 April 2021

References
1. Knip, M. et al. Environmental triggers and determinants of type 1 diabetes. *Diabetes* **54**(Suppl 2), S125–136 (2005).
2. Levin, L. et al. Analysis of HLA genes in families with autoimmune diabetes and thyroiditis. *Hum. Immunol.* **65**, 640–647. https://doi.org/10.1016/j.humimm.2004.02.026 (2004).
3. Redondo, M. J. et al. Heterogeneity of type 1 diabetes: Analysis of monozygotic twins in Great Britain and the United States. *Diabetologia* **44**, 354–362. https://doi.org/10.1007/s001250001626 (2001).
4. Rewers, M. & Ludvigsson, J. Environmental risk factors for type 1 diabetes. *Lancet* **387**, 2340–2348. https://doi.org/10.1016/S0140-6736(16)30507-4 (2016).
5. Bradfield, J. P. et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet.* **7**, e1002293. https://doi.org/10.1371/journal.pgen.1002293 (2011).
6. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature* **447**, 661–678. https://doi.org/10.1038/nature05911 (2007).
7. Todd, J. A. et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.* **39**, 857–864. https://doi.org/10.1038/ng2068 (2007).
8. Howson, J. M., Walker, N. M., Clayton, D. & Todd, J. A. Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. *Diabetes Obes. Metab.* **11**(Suppl 1), 31–45. https://doi.org/10.1111/j.1463-1326.2008.01001.x (2009).
9. Pociot, F. & Lernmark, A. Genetic risk factors for type 1 diabetes. *Lancet* **387**, 2331–2339. https://doi.org/10.1016/S0140-6736(16)30582-7 (2016).
10. Ilonen, J. et al. Genetic susceptibility to type 1 diabetes in childhood: Estimation of HLA class II associated disease risk and class II effect in various phases of islet autoimmunity. *Pediatr. Diabetes* **17**(Suppl 22), S96. https://doi.org/10.1111/pedi.12327 (2016).
11. Mychaleckyj, J. C. et al. HLA genotyping in the international Type 1 Diabetes Genetics Consortium. *Clin. Trials* 7, 575–87. https://doi.org/10.1177/1740774510373494 (2010).
12. Zhao, L. P. et al. Motifs of three HLA-DQ amino acid residues (alpha44, beta57, beta135) Capture Full association with the risk of type 1 diabetes in DQZ and DQ8 children. *Diabetes* **69**, 1573–1587. https://doi.org/10.2337/db20-0075 (2020).
13. Zhao, L. P. et al. Next generation HLA sequence analysis uncovers seven HLA-DQ amino acid residues and six motifs resistant to childhood type 1 diabetes. *Diabetes* **69**, 2523–2533 (2020).
14. Zhao, L. P. et al. 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* **68**, 1692–1704. https://doi.org/10.2337/db17-0273 (2019).
15. Zhao, L. P. et al. Next-generation sequencing reveals that HLA-DRB3, –DRB4, and –DRB5 may be associated with islet autoantibodies and risk for childhood type 1 diabetes. *Diabetes* **65**, 710–718. https://doi.org/10.2337/db15-1115 (2016).
16. Delli, A. J. et al. 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* **61**, 2556–2564. https://doi.org/10.2337/db11-1699 (2012).
17. Rich, S. S. et al. Overview of the Type 1 Diabetes Genetics Consortium. *Genes Immun* **10**(Suppl 1), S1–4. https://doi.org/10.1038/gene.2009.84 (2009).
18. Kwock, W. W., Kovats, S., Thurtle, P. & Nepom, G. T. HLA-DQ allelic polymorphisms constrain patterns of class II heterodimer formation. *J Immmunol* **150**, 2263–2272 (1993).
58. Andersson, C. et al. Triple specificity of ZnT8 autoantibodies in relation to HLA and other idet autoantibodies in childhood and adolescent type 1 diabetes. Pediatr. Diabetes 14, 97–105. https://doi.org/10.1111/j.1399-5448.2012.00916.x (2013).
59. Persson, M. et al. The Better Diabetes Diagnosis (BDD) study: A review of a nationwide prospective cohort study in Sweden. Diabetes Res. Clin. Pract. 140, 236–244. https://doi.org/10.1016/j.diabres.2018.03.057 (2018).
60. Classification and Diagnosis of Diabetes. Diabetes Care 38, 58–516. https://doi.org/10.2337/dci15-5005 (2015).
61. Carlsson, A. et al. 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) 36, 718–724. https://doi.org/10.1038/ijo.2011.122 (2012).
62. Gyllenberg, A. et al. Age-dependent variation of genotypes in MHC II transactivator gene (CITIA) in controls and association to type 1 diabetes. Genes Immun. 13, 632–640. https://doi.org/10.1038/gene.2012.44 (2012).
63. Smith, A. G. et al. Next generation sequencing to determine HLA class II genotypes in a cohort of hematopoietic cell transplant patients and donors. Hum. Immunol. 75, 1040–1046. https://doi.org/10.1016/j.humimm.2014.08.206 (2014).
64. Nelson, W. C. et al. An integrated genotyping approach for HLA and other complex genetic systems. Hum. Immunol. https://doi.org/10.1016/j.humimm.2015.05.001 (2015).
65. Fremont, D. H., Monnaie, D., Nelson, C. A., Hendrickson, W. A. & Unanue, E. R. Crystal structure of I-Ak in complex with a dominant epitope of lysozyme. Immunity 8, 305–317. https://doi.org/10.1016/s1074-7613(00)80536-1 (1998).
66. Brown, J. H. et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364, 33–39. https://doi.org/10.1038/364033a0 (1993).
67. Porta, M. A Dictionary of Epidemiology 5th edn. (Oxford University Press, Oxford, 2008).
68. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. & Golani, I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125, 279–284. https://doi.org/10.1016/s0166-4328(01)00297-2 (2001).
69. Siebold, C. et al. Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy. Proc Natl Acad Sci USA 101, 1999–2004. https://doi.org/10.1073/pnas.0308458100 (2004).
70. Matsuoka, T., Tabata, H. & Matsushita, S. Monocytes are differentially activated through HLA-DR, -DQ, and -DP molecules via mitogen-activated protein kinases. J. Immunol. 166, 2202–2208. https://doi.org/10.4049/jimmunol.166.4.2202 (2001).
71. H. Miyadera, L. B. Bungener, S. Kusano, S. Yokoyama, K. Tokunaga, B. G. Hepkema, (2015) Questionable expression of unstable DQ heterodimer containing HLA-DQA1*01:07. Tissue Antigens 86 (6):413–418

Acknowledgements
The authors would also like to thank Dr. Gun Forssander at Department of Pediatrics, the Queen Silvia Children’s Hospital, Sahlgrenska University Hospital, Gothenburg, Sweden and Dr. Sten A. Ivarsson Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden, for their clinical expertise in conduct of the BDD study. Equally, the authors would like to thank Anita Ramelius, Linda Faxius, Ida Jönsson and Rasmus Bennet all at the Department of Clinical Sciences, Lund University/CRC, in Malmö, Sweden for expert technical assistance.

Author contributions
L.P.Z., Å.L., G.K.P., D.G. and S.A. researched and analyzed the data and wrote the manuscript. I.K. researched data (control subjects) and reviewed the manuscript. M.K., B.X., R.W., C.-W.P. and W.C.N. contributed to the next-generation sequencing, researched data and reviewed the manuscript. A.K.M. and G.P.B. carried out molecular simulations and graphical representations of select HLA-DR molecules. A.C.; H.E.L.; J.L.; M.P.; U.S.; E.Ö. and C.M. designed the BDD study, researched data, contributed to discussion and reviewed the manuscript. All authors used appropriate clinical, epidemiological, genetic, and statistical methods to carry out all investigation in accordance with the best practice in biomedical research per relevant guidelines and regulations. Å.L. is the guarantor of this work, had full access to all the data, and takes full responsibility for the integrity of data and accuracy of data analysis.

Funding
Open access funding provided by Lund University. The study was supported by a grant (1R01DK117276) from the National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), a grant from the European Foundation for the Study of Diabetes (EFSD) Clinical Research Grants Programme 2013 and in part the Swedish Child Diabetes Foundation (Barndiabetesfonden), Swedish Foundation for Strategic Research (Dnr IRC15-0067), Swedish Research Council, Strategic Research Area (Dnr 2009–1039), the Skåne County Council for Research and Development as well as the Swedish Association of Local Authorities and Regions (SKL).

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-86229-8.

Correspondence and requests for materials should be addressed to L.P.Z.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
