Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk

Xinli Hu[1-6]*, Aaron J. Deutsch[1-5]*, Tobias L. Lenz[2,7], Suna Onengut-Gumuscu[8], Buhm Han[2,4,9], Wei-Min Chen[8], Joanna M.M. Howson[10], John A. Todd[11], Paul I.W. de Bakker[12,13], Stephen S. Rich[8], Soumya Raychaudhuri[1-4,14]

1. Division of Rheumatology, Immunology and Allergy, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA
2. Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
3. Partners Center for Personalized Genetic Medicine, Boston, MA, USA
4. Program in Medical and Population Genetics, The Broad Institute, Cambridge, MA, USA
5. Harvard-MIT Division of Health Sciences and Technology, Boston, MA USA
6. Division of Medical Sciences, Harvard Medical School, Boston, MA, USA
7. Evolutionary Immunogenomics, Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Ploen, Germany
8. Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA
9. Asan Institute for Life Sciences, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea
10. Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
11. The Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, NIHR Cambridge Biomedical Research Centre, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Cambridge Biomedical Campus, Cambridge, UK
12. Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, the Netherlands;
13. Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands
14. Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK

* These authors contributed equally to this work.

Correspondence should be address to:
Soumya Raychaudhuri
77 Avenue Louis Pasteur
Harvard New Research Building, Suite 250D
Boston, Massachusetts 02115, United States of America
soumya@broadinstitute.org
617-525-4484 (tel) 617-525-4488 (fax)

Format: Article
Organization: Title, Authors, Affiliations, Abstract, Introduction, Results, Discussion, Acknowledgements, Author Contributions, References for main text, Figure Legends for main text, Tables, Online Methods
Running title: Fine-mapping HLA associations in T1D
Keywords: Type 1 diabetes, MHC, HLA, amino acids, interaction

ABSTRACT

Variation in the human leukocyte antigen (HLA) genes accounts for one-half of the genetic risk in type 1 diabetes (T1D). Amino acid changes in the HLA-DR and -DQ molecules mediate most of the risk; but extensive linkage disequilibrium complicates localization of independent effects. Using 18,832 case-control samples, we localized the signal to three amino acid positions in HLA-DQ and -DR. DQβ1 position 57 (previously known, $p=10^{-1355}$) alone explained 15.2% of the total phenotypic variance. Independent effects at DRβ1 positions 13 ($p=10^{-721}$) and 71 ($p=10^{-95}$) increased the proportion to 26.9%. The three positions together explained 90% of the phenotypic variance in *HLA-DRB1-DQA1-DQB*1. Additionally, we observed significant interactions in 11 of 21 pairs of common *HLA-DRB1-DQA1-DQB1* haplotypes ($p=1.6\text{x}10^{-64}$). DRβ1 positions 13 and 71 implicate the P4 pocket in the antigen-binding groove, thus pointing to another critical protein structure for T1D risk, in addition to the DQ P9 pocket.

INTRODUCTION

Type 1 diabetes (T1D) is a highly heritable autoimmune disease that results from T cell-mediated destruction of the insulin-producing pancreatic β cells. The worldwide incidence of T1D ranges from 0.1 per 100,000 persons in China to >36 per 100,000 in parts of Europe, and has been steadily increasing[1]. Many autoimmune diseases, including T1D, rheumatoid arthritis (RA), celiac disease, and multiple sclerosis, have more genetic risk attributed to variants in the human leukocyte antigen (HLA) genes within the major histocompatibility complex (MHC) region located on 6p21.3[2-4] than any other locus. HLA genes encode surface proteins that display antigenic peptides to effector immune cells in order to regulate self-tolerance and downstream immune responses. Autoimmune risk conferred by HLA is likely the result of variation in amino acid residues at specific positions within the antigen-binding grooves, which may then alter the repertoire of presented peptides[5-8]. In T1D, the largest allelic associations are in *HLA-DRB1-DQA1-DQB1*, a three-gene "superlocus" that encodes HLA-DR and -DQ proteins[9,10]; additional associations have been identified in the genes encoding HLA-A, -B, -C, and -DP[11-14].

Todd *et al.* initially identified strong T1D risk conferred by non-aspartate residues at position 57 of HLA-DQβ1[15]. However, this amino acid position alone does not fully explain the HLA risk in T1D. Subsequently, many amino acid positions in DQβ1 and DRβ1 have been hypothesized to modify risk[16]; but extensive linkage disequilibrium (LD) spanning the 4 Mb MHC region makes it challenging to pinpoint the specific risk variants. In addition, certain heterozygous genotypes confer the greatest disease risk [13,17-19], consistent with synergistic interactions between classical HLA alleles. Despite evidence of non-additive effects within the MHC on autoimmune disease risk, interactions have not been comprehensively examined in T1D. If the risk-conferring amino acid positions and their interactions were understood, mechanistic investigation of how autoantigens interact with HLA proteins could become feasible. In this study, we utilized recently

established accurate genotype imputation methods to examine a large case-control sample, and rigorously identified independent amino acid positions as well as interactions within the HLA that account for T1D risk (see **Supplementary Figure 1** for a schematic of analyses).

RESULTS

HLA Imputation and association testing

We fine-mapped the MHC region in a collection of 8,095 T1D cases and 10,737 controls genotyped with the ImmunoChip array, provided by the Type 1 Diabetes Genetics Consortium (T1DGC)[20-22]. The dataset included (1) case-control samples collected in the United Kingdom (UK), and (2) a pseudocase-control set derived from European families (Eur) (see **Online Methods** in **and Supplementary Table 1**). Using a set of 5,225 individuals with classical HLA typing as a reference[22], we accurately imputed 8,617 binary markers (with minor allele frequency > 0.05%) between ~29 Mb and ~33 Mb (the 4 Mb classical MHC region) on chromosome 6p21.3 with SNP2HLA software[21]. The resulting data included 7,242 SNPs, 260 2- and 4-digit classical alleles, and amino acid residues at 399 positions in eight HLA genes (*HLA-A, -B, -C, -DRB1, -DQA1, -DQB1, -DPA1,* and *-DPB1*) with high imputation quality (INFO score > 0.96; see **Supplementary Table 2** for the list of variants and imputation quality). We have previously independently benchmarked the imputation strategy employed in this study for accuracy, using a set of 918 samples with gold-standard HLA typing data. Starting with SNPs from the ImmunoChip genotyping platform and using the T1DGC reference panel, SNP2HLA obtained an accuracy of 98.4%, 96.7% and 99.3% for all 2-digit alleles, 4-digit alleles, and amino acid polymorphisms, respectively[21].

To test for T1D association of a given variant, we used a logistic regression model, assuming the log-odds of disease to be proportional to the allelic dosage of the variant. We also included covariates to adjust for sex and region of origin (see **Supplementary Note** and **Supplementary**

Figure 2). As expected, the strongest associations with T1D were within the *HLA-DRB1-DQA1-DQB1* locus. We confirmed that the leading risk variant was the presence of alanine at DQβ1 position 57 (DQβ1#57, $p=10^{-1090}$, OR=5.17; **Figure 1A, Supplementary Table 2**). In contrast, the single most significantly associated classical allele was *DQB1*03:02* ($p=10^{-840}$), which has an alanine at DQβ1#57, but was much more weakly associated than the amino acid residue itself. **Table 1** lists common classical alleles tagged by each residue at key amino acid positions.

Three amino acid positions independently drive T1D risk

Given the strength and complexity of the association within *HLA-DRB1-DQB1-DQA1*, we aimed to first identify independent effects in this locus before examining the rest of the MHC. We assessed the significance of multi-allelic amino acid positions using conditional analysis by forward-search (see **Online Methods**). Unsurprisingly, the most strongly associated position with T1D was DQβ1#57 (omnibus $p=10^{-1355}$, **Figure 2, Tables S3** and **S4A**). At this position, alanine conferred the strongest risk (OR=5.17; **Figure 3**), while the most common residue in controls, aspartic acid, was the most protective (OR=0.16). Conditioning on DQβ1#57, the second independent association was at DRβ1#13 (omnibus $p=10^{-721}$, **Figure 2**). At this position, histidine (OR=3.64) and serine (OR=1.28) conferred the strongest risk, while arginine (OR=0.08) and tyrosine (OR=0.28) were protective (**Figure 3, Supplementary Table 4A**). The DRβ1#71 position was the third independently associated signal (omnibus $p=10^{-95}$, **Figure 2**); lysine conferred strong risk (OR=4.70), and alanine was strongly protective (OR=0.04, **Figure 3**, **Supplementary Table 4A**). We note that at these positions, the risk-conferring amino acid residues indeed tag the *DR3* and *DR4* haplotypes, which confer the strongest risk among haplotypes. Histidine at position 13 tags *DRB1*04:01* and *04:04*, while serine tags *DRB1*03:01*. Lysine at position 71 tags both *DRB1*03:01* and *DRB1*04:01*. **Tables 1** and **S6** list the classical alleles tagged by residues at each key amino acid position, and multivariate odds ratios of

haplotypes defined by these positions.

Given the reported deviation from log-scale additivity of T1D risk effects in the HLA[19,23], we wanted to confirm that their contribution did not alter the risk-driving amino acid positions. By repeating the forward-search analysis while including non-additive terms in the regression model, we confirmed that DQβ1#57, DRβ1#13, and DRβ1#71 were the top three independent signals under the non-additive model as well as the additive model (see **Supplementary Note** and **Supplementary Figure 3**).

We exhaustively tested all possible combinations of two, three, and four amino acid positions in *HLA-DRB1-DQA1-DQB1*, and confirmed that DQβ1#57, DRβ1#13, and DRβ1#71 were the best of all 457,450 combinations of three amino acids ($p=10^{-2161}$; see **Supplementary Table 7**). Conditioning on these three positions, more than 80 other positions and classical alleles remained highly significant ($p < 10^{-8}$; see **Supplementary Table 4B** and **S4C**), suggesting the presence of other independent associations. DQβ1#–18 (located within the signal peptide) emerged as the fourth most significant association ($p=10^{-40}$) through forward-search; however, in the exhaustive test, many other combinations of four amino acid positions exceeded the goodness-of-fit of DQβ1#57/DRβ1#13/DRβ1#71/DQβ1#–18 (**Supplementary Table 7**). Therefore, we do not report subsequent positions that emerged through conditional analysis, as we could not confidently claim additional positions as independent drivers of T1D risk.

We wanted to confirm that these three amino acid positions were not simply tagging effects of specific haplotypes. To this end, we performed a permutation analysis in which we randomly reassigned amino acid sequences corresponding to each *HLA-DRB1, -DQB1,* and *-DQA1* classical allele, and retested for the best amino acid positions (see **Online Methods**). This approach preserved haplotypic associations; if certain amino acids were tagging associated

haplotypes, equally significant amino acid associations would be found in the permuted data. After 10,000 permutations, no combination of permuted amino acids resulted in a model that equaled or exceeded the goodness-of-fit of DQβ1#57/DRβ1#13/DRβ1#71 in our data, as measured by either deviance or *p*-value (see **Supplementary Figure 4**).

Finally, to ensure that the observed effects were not the results of heterogeneity between the UK and the European subsets, we separately repeated the association analysis in the two subsets. The two sets yielded highly correlated effect sizes for all binary markers (Pearson r=0.952, **Supplementary Figure 5A**), as well as for all haplotypes formed by residues at DQβ1#57, DRβ1#13, and DRβ1#71 (Pearson r=0.989, **Supplementary Figure 5B**).

Key amino acids are located in the peptide-binding grooves

DQβ1#57, DRβ1#13, and DRβ1#71 are each located in the peptide-binding grooves of the respective HLA molecule (**Figure 4**). DRβ1#13 and #71 line the P4 pocket of HLA-DR, which has been previously implicated in seropositive RA[2], seronegative RA[24], and follicular lymphoma[25]. While DRβ1#13 and DRβ1#71 are both involved in T1D and RA, the effects of individual residues at each position are discordant between the diseases ($p<10^{-232}$; see **Supplementary Figure 6** and **Online Methods**).

The three amino acid positions explain over 90% of T1D phenotypic variance in *HLA-DRB1-DQA1-DQB1*

We quantified the proportion of phenotypic variance captured by the three amino acid positions using the liability threshold model[26] (see **Supplementary Note**). Assuming a T1D prevalence of 0.4%[27], the additive effects of all 67 haplotypes in *HLA-DRB1-DQA1-DQB1* explained 29.6% of the total phenotypic variance (**Supplementary Table 6**). DQβ1#57 alone explained 15.2% of the total variance, while the addition DRβ1#13, and DRβ1#71 increased the proportion

explained by 11.7%. Therefore, these three amino acid positions together capture 26.9% of the total variance, accounting for over 90% of the T1D-HLA association in this locus (**Figure 5**).

Independent HLA associations in *HLA-B, -DPB1, and -A*

We then sought to identify HLA associations to T1D independent of those in *HLA-DRB1-DQA1-DQB1.* We conservatively conditioned on all *HLA-DRB1, DQA1,* and *DQB1* 4-digit classical alleles to eliminate all effects at these loci. We observed the next strongest association across the MHC in *HLA-B,* where the classical allele *HLA-B*39:06* was the most significant signal (OR=6.64, p=$10^{-75,}$ **Figure 1B, Supplementary Table 5A**)[11]. After adjusting for *B*39:06*, other classical alleles and amino acid positions in *HLA-B* remained significantly associated, including *B*18:01* and *B*50:01*. Upon additionally adjusting for all *HLA–B* alleles, *HLA-DPB1*04:02* was the next strongest independent signal (OR=0.47, $p < 10^{-55}$, **Figure 1C**), which is nearly perfectly tagged by methionine at amino acid position 178. Conditioning on *DPB1*04:02*, additional associations were present in *HLA-DPB1*, including position 65 and *DPB1*01:01* (**Supplementary Table 5B**). After conditioning on *DPB1* alleles as well, we observed independent effects in *HLA-A* led by position 62 (p=10^{-45}, **Figure 1D**); additional signals included *A*03* and *A*24:02* (**Supplementary Table 5C**). We observed no independent association with T1D in *HLA-C* or *-DPA1* (**Figure 1E**). The independent effects of all haplotypes in *HLA-B, -DPB1, and -A* together explained ~4% of the total phenotypic variance. The total T1D risk variance explained by additive effects in the eight HLA genes was ~34%, consistent with the estimates by Speed *et al.*[28].

HLA haplotypic interaction effects are common in T1D

The previously observed excess risk of T1D in *HLA-DR3/DR4* (*DRB1*03:01-DQA1*05:01-DQB1*02:01/DRB1*04:xx-DQA1*03:01-DQB1*03:02)* heterozygotes may represent a synergistic interaction between two distinct alleles[23]. Here, we conducted an unbiased search

for interactions among all haplotypes within the *HLA-DRB1-DQA1-DQB1* locus (see **Online Methods**). As interactions cannot be observed reliably with rare genotypes, we focused this analysis on the seven *HLA-DRB1-DQA1-DQB1* haplotypes with frequencies > 5%; all of these haplotypes had very imputation accuracies (INFO score > 0.94; see **Supplementary Table 8 and Supplementary Note**).

We tested for interactions between all possible pairs of haplotypes using a global multivariate regression model that included 21 interactive terms as well as seven additive terms. The inclusion of interactions in the model produced a statistically significant improvement in fit over the additive model (p=1.6x10^{-64}). Of the 21 potential interactions, 11 were significant after correcting for 21 tests (p<0.05/21=2.4x10^{-3}**; Figure 6, Table 2, Supplementary Table 9**). Consistent with previous reports[9,19], we observed a significant interaction between the *HLA-DR3* haplotype (*DRB1*03:01-DQA1*05:01-DQB1*02:01*) and a *DR4* haplotype (*DRB1*04:01-DQA1*03:01-DQB1*03:02*) (p=1.2 x 10^{-5}). This interaction resulted in an odds ratio of 30.42, compared to an expected odds ratio of 15.51 due to only additive contributions. Likewise, we confirmed an independent interaction between *DRB1*03:01-DQA1*05:01-DQB1*02:01*and *DRB1*04:04-DQA1*03:01-DQB1*03:02* (p=1.9x10^{-4}).

We observed many other significant haplotypic interactions beyond the well-studied *DR3/DR4* heterozygote effect (**Table 2** and **Supplementary Table 9**). Most interactions increased T1D risk. For example, the combination of *DRB1*04:01-DQA1*03:01*DQB1*03:02* and *DRB1*07:01-DQA1*02:01-DQB1*02:02* dramatically increased risk by 5.09-fold (beyond the risk predicted by the additive model). Other pairs significantly reduced risk. Notably, while *DRB1*04:01-DQA1*03:01-DQB1*03:02* and *DRB1*04:04-DQA1*03:01-DQB1*03:02* each conferred risk, the heterozygote combination elicited a 3-fold reduction from the expected risk. Since we restricted

our analysis to haplotypes with at least 5% allele frequency, other interactive effects are likely present but unobserved[19].

Interaction effects are mediated by DQβ1#57 and DRβ1#13

The HLA-DQ α/β *trans* heterodimer formed by proteins encoded by *DQA1*05:01* and *DQB1*03:02* may confer a particularly high risk for individuals with the *DR3/DR4* genotype due to its unique antigen binding properties[29]. In order to identify the possible drivers of this haplotypic interaction, we tested pairwise interactions among the *HLA-DRB1, -DQA1,* and -*DQB1* 4-digit alleles. We observed a significant interaction *between DQA1*05:01 and DQB1*03:02* (p=1.71x10^{-25}). However, due to high LD across the locus, several pairs of classical alleles (including *DQB1*02:01/DQB1*03:02 and DRB1*03:01/DQB1*03:02*, **Supplementary Table 10**) achieved similarly significant *p*-values. Therefore, while our model was consistent with a risk-conferring interaction between *DQA1*05:01* and *DQB1*03:02*, we could not eliminate the possibility that interactions between other alleles within the two haplotypes are driving this specific interaction.

We next assessed whether these haplotypic interactions could be explained by amino acid positions. We exhaustively tested for all pairwise interactions among amino acid residues in *HLA-DRB1-DQA1-DQB1*, again limiting the analysis to residues with at least 5% frequency. Of the 3,773 pairs of amino acid positions tested, we observed that interactions between DQβ1#57 and DRβ1#13 yielded the largest improvement over the additive model (**Supplementary Table 11**). We note that two other pairs of amino acid positions achieved similarly significant *p*-values. These analyses suggest that the same amino acid positions that explain the greatest proportion of the additive risk may also be the positions that mediate interactive effects within this locus.

DISCUSSION

Fine-mapping the MHC locus in T1D demonstrates that amino acid polymorphisms at DQβ1#57, DRβ1#13, and DRβ1#71 independently modulate T1D risk, and capture over 90% of the phenotypic variance explained by the *HLA-DRB1-DQA1-DQB1* locus (and 80% of the variance explained by the entire MHC). Previous studies have suggested that other amino acid positions within the HLA class II molecules confer T1D risk; for example, DRβ1#86, DRβ1#74, and DRβ1#57 in the P1, P4, and P9 pockets, respectively[16]. While our analysis highlights the top three amino acid positions as the main contributors of T1D risk, there is also evidence of other allelic effects within the *HLA-DRB1-DQA1-DQB1* locus; however their relative effect sizes were very modest compared to the three leading positions identified. We note that our results are derived from cases and controls from a relatively homogeneous population (the United Kingdom), and our ability to interrogate rare alleles in this population may be limited. For instance*, HLA-DRB1*04:03*, a common protective allele in the Sardinian population highlighted by Cucca *et al.*[16], is rare in this dataset with an allele frequency of 0.3%. As such, the observed effect of amino acid positions which best define this allele (DRβ1#74 and #86) may have been less pronounced than what might be observed in a more diverse dataset. Additional variants may be conclusively identified in the future with increased sample size. Finally, although coding variants contribute to the majority of the phenotypic variance in T1D, there is the possibility that there are other mechanisms, such as gene and protein expression, that further modulate susceptibility[30,31].

Beyond the previously described *HLA-DR3/DR4* interactions, we find nine additional pairwise interactions between HLA haplotypes that contribute to T1D risk, suggesting that non-additive effects are common within this locus. Notably, we showed that amino acid positions in DQβ1#57 and DRβ1#13 were the strongest contributors to both additive and interactive risk effects. Interestingly, the two strongest interacting amino acid positions are in separate HLA molecules (DQ and DR, respectively). *HLA-DQA1*, which is in strong LD to *HLA-DRB1* and *HLA-DQB1*,

appears to play a minimal role in modulating T1D risk. This suggests that the interactive effects are possibly due to the alteration in antigen-presentation repertoire created by the combination of different HLA molecules, rather than the consequence of specific DQα/β heterodimers with particular structural features that confer extreme binding affinities.

The HLA amino acid variants identified in our study may mediate recognition of one or more autoantigens and cause autoimmunity through different mechanisms. In particular, our findings implicate the HLA-DR P4 pocket in T1D in addition to the known role of the HLA-DQ P9 pocket; this is the first instance to our knowledge where the DR P4 pocket plays an important but secondary role to a different locus (DQβ1#57). The DR P4 pocket has been shown to play primary roles in other autoimmune diseases. For example, in RA, the risk-conferring amino acid residues in P4 likely facilitate the binding of citrullinated peptides[7]. In T1D, the anti-islet autoantibody reactivity in patients' sera is largely accounted for by four autoantigens: preproinsulin, glutamate decarboxylase (GAD), islet antigen 2 (IA-2), and ZnT8; although the identification of specific peptides that affect autoreactivity is still work in progress[8,32-37]. Cucca *et al.* implicated signal peptide sequences of preproinsulin as potentially important in T1D, by modeling the associations of HLA class II alleles and their polymorphic amino acid positions with structural features of the peptide-binding pockets[16]. The discovery of critical variants that drive T1D risk enable future functional investigations. Synthesis of HLA molecules containing single-residue alterations at risk-modulating positions may reveal their effects on the physicochemical properties of the antigen-binding pockets. Furthermore, the use of peptide display or small molecule libraries may directly identify and characterize peptides that differentially bind to HLA molecules that differ at risk-modulating positions, thereby revealing the essential pathogenic peptides and the mechanisms through which they evoke autoimmunity.

ACKNOWLEDGMENT

This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. This work is supported in part by funding from the National Institutes of Health (5R01AR062886-02 (PIdB), 1R01AR063759 (SR), 5U01GM092691-05 (SR), 1UH2AR067677-01 (SR), R01AR065183 (PIWdB)), a Doris Duke Clinical Scientist Development Award (SR), the Wellcome Trust (JAT) and the National Institute for Health Research (JAT and JMMH), and a Vernieuwingsimpuls VIDI Award (016.126.354) from the Netherlands Organization for Scientific Research (PIWdB). TLL was supported by the German Research Foundation (LE 2593/1-1 and LE 2593/2-1).

AUTHOR CONTRIBUTIONS

XH and SR conceived and the study. XH, AJD, TLL, SR, BH, PIWdB, and SRS contributed to the study design and analysis strategy. XH, AJD, TLL, and SR conducted all analyses. XH and AJD wrote the initial manuscript. BH contributed critical analytical methods. SG-O, WC, and SRS organized and contributed subject samples and provided SNP genotype data. JMMH, JAT, PIWdB, SRS and SR contributed critical writing and review of the manuscript. All authors contributed to the final manuscript.

REFERENCES

1. Maahs, D.M., West, N.A., Lawrence, J.M. & Mayer-Davis, E.J. Epidemiology of type 1 diabetes. *Endocrinol Metab Clin North Am* **39**, 481-97 (2010).
2. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* **44**, 291-6 (2012).
3. Trynka, G., Wijmenga, C. & van Heel, D.A. A genetic perspective on coeliac disease. *Trends Mol Med* **16**, 537-50 (2010).
4. Gourraud, P.A., Harbo, H.F., Hauser, S.L. & Baranzini, S.E. The genetics of multiple sclerosis: an up-to-date review. *Immunol Rev* **248**, 87-103 (2012).
5. Lee, K.H., Wucherpfennig, K.W. & Wiley, D.C. Structure of a human insulin peptide-HLA-DQ8 complex and susceptibility to type 1 diabetes. *Nat Immunol* **2**, 501-7 (2001).
6. Astill, T.P., Ellis, R.J., Arif, S., Tree, T.I. & Peakman, M. Promiscuous binding of proinsulin peptides to Type 1 diabetes-permissive and -protective HLA class II molecules. *Diabetologia* **46**, 496-503 (2003).
7. Scally, S.W. *et al.* A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. *J Exp Med* **210**, 2569-82 (2013).
8. van Lummel, M. *et al.* Posttranslational modification of HLA-DQ binding islet autoantigens in type 1 diabetes. *Diabetes* **63**, 237-47 (2014).
9. Erlich, H. *et al.* HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. *Diabetes* **57**, 1084-92 (2008).
10. Noble, J.A. *et al.* The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am J Hum Genet* **59**, 1134-48 (1996).
11. Nejentsev, S. *et al.* Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* **450**, 887-92 (2007).
12. Cucca, F. *et al.* The HLA-DPB1--associated component of the IDDM1 and its relationship to the major loci HLA-DQB1, -DQA1, and -DRB1. *Diabetes* **50**, 1200-5 (2001).
13. Deschamps, I. *et al.* HLA genotype studies in juvenile insulin-dependent diabetes. *Diabetologia* **19**, 189-93 (1980).
14. Howson, J.M., Walker, N.M., Clayton, D., Todd, J.A. & Type 1 Diabetes Genetics, C. 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 (2009).
15. Todd, J.A., Bell, J.I. & McDevitt, H.O. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* **329**, 599-604 (1987).
16. Cucca, F. *et al.* A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. *Hum Mol Genet* **10**, 2025-37 (2001).
17. Thomson, G. *et al.* Genetic heterogeneity, modes of inheritance, and risk estimates for a joint study of Caucasians with insulin-dependent diabetes mellitus. *Am J Hum Genet* **43**, 799-816 (1988).
18. Svejgaard, A. & Ryder, L.P. HLA genotype distribution and genetic models of insulin-dependent diabetes mellitus. *Ann Hum Genet* **45**, 293-8 (1981).
19. Koeleman, B.P. *et al.* Genotype effects and epistasis in type 1 diabetes and HLA-DQ trans dimer associations with disease. *Genes Immun* **5**, 381-8 (2004).
20. Onengut-Gumuscu, S. *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* (2015).

21. Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* **8**, e64683 (2013).
22. Brown, W.M. *et al.* Overview of the MHC fine mapping data. *Diabetes Obes Metab* **11 Suppl 1**, 2-7 (2009).
23. Noble, J.A. & Erlich, H.A. Genetics of type 1 diabetes. *Cold Spring Harb Perspect Med* **2**, a007732 (2012).
24. Han, B. *et al.* Fine mapping seronegative and seropositive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. *Am J Hum Genet* **94**, 522-32 (2014).
25. Foo, J.N. *et al.* Coding variants at hexa-allelic amino acid 13 of HLA-DRB1 explain independent SNP associations with follicular lymphoma risk. *Am J Hum Genet* **93**, 167-72 (2013).
26. Witte, J.S., Visscher, P.M. & Wray, N.R. The contribution of genetic variants to disease depends on the ruler. *Nat Rev Genet* **15**, 765-76 (2014).
27. Sivertsen, B., Petrie, K.J., Wilhelmsen-Langeland, A. & Hysing, M. Mental health in adolescents with Type 1 diabetes: results from a large population-based study. *BMC Endocr Disord* **14**, 83 (2014).
28. Speed, D., Hemani, G., Johnson, M.R. & Balding, D.J. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* **91**, 1011-21 (2012).
29. Reichstetter, S., Kwok, W.W. & Nepom, G.T. Impaired binding of a DQ2 and DQ8-binding HSV VP16 peptide to a DQA1*0501/DQB1*0302 trans class II heterodimer. *Tissue Antigens* **53**, 101-5 (1999).
30. Ettinger, R.A., Liu, A.W., Nepom, G.T. & Kwok, W.W. Exceptional stability of the HLA-DQA1*0102/DQB1*0602 alpha beta protein dimer, the class II MHC molecule associated with protection from insulin-dependent diabetes mellitus. *J Immunol* **161**, 6439-45 (1998).
31. Miyadera, H., Ohashi, J., Lernmark, A., Kitamura, T. & Tokunaga, K. Cell-surface MHC density profiling reveals instability of autoimmunity-associated HLA. *J Clin Invest* (2014).
32. Knight, R.R. *et al.* A distinct immunogenic region of glutamic acid decarboxylase 65 is naturally processed and presented by human islet cells to cytotoxic CD8 T cells. *Clin Exp Immunol* **179**, 100-7 (2015).
33. Kronenberg, D. *et al.* Circulating preproinsulin signal peptide-specific CD8 T cells restricted by the susceptibility molecule HLA-A24 are expanded at onset of type 1 diabetes and kill beta-cells. *Diabetes* **61**, 1752-9 (2012).
34. Nakayama, M. *et al.* Prime role for an insulin epitope in the development of type 1 diabetes in NOD mice. *Nature* **435**, 220-3 (2005).
35. Baekkeskov, S. *et al.* Identification of the 64K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. *Nature* **347**, 151-6 (1990).
36. Karlsen, A.E. *et al.* Recombinant glutamic acid decarboxylase (representing the single isoform expressed in human islets) detects IDDM-associated 64,000-M(r) autoantibodies. *Diabetes* **41**, 1355-9 (1992).
37. Roep, B.O. & Peakman, M. Antigen targets of type 1 diabetes autoimmunity. *Cold Spring Harb Perspect Med* **2**, a007781 (2012).

FIGURE LEGENDS

Figure 1. Independently associated HLA loci to T1D. Each binary marker was tested for T1D association, using the imputed allelic dosage (between 0 and 2). In each panel, the horizontal dashed line marks p=5x10^{-8}. Color gradient of the diamond indicates LD (r^2) to the most strongly associated variant; the darkest shade is r^2=1. A) The strongest associations were located in *HLA-DRB1-DQA1-DQB1*. The single strongest risk variant was alanine at DQβ1#57 (OR=5.17; p=10^{-1090}). See **Supplementary Table 2** for unadjusted associations for all markers. B) Adjusting for all *DRB1, DQA1,* and *DQB1* 4-digit classical alleles, the strongest independent signals were in *HLA-B*. The strongest association was to *B*39:06* (OR=6.64, p=10^{-75}). C) Adjusting for *HLA-DRB1-DQA1-DQB1* and *HLA-B,* the next associated variant was *DPB1*04:02 (*OR=0.48, p=10^{-55}). D) The final independent association was in *HLA-A*, led by glutamine at A#62 (OR=0.70, p=10^{-25}). E) We found no residual independent association in the *HLA-C* or *HLA-DPA1*.

Figure 2. Amino acid positions DQβ1#57, DRβ1#13, and DRβ1#71 independently drive T1D risk associated to the *HLA-DRB1-DQA1-DQB1* locus. To identify each independently associated position, we used conditional haplotypic analysis by forward search, using the phased best-guess genotypes. In each panel, the dots mark amino acid positions along the gene (x-axis) and their $\log_{10}$ association p-values on the y-axis. The horizontal dashed line marks the $\log_{10}$ p-value of most strongly associated classical allele. The most strongly associated signals are circled. The colored arrows indicate positions that have been conditioned on. The most strongly associated position was DQβ1#57 (p=10^{-1355}). Conditioning on it, DRβ1#13 was the next independently associated position (p=10^{-721}), followed by DRβ1#71 (p=10^{-95}). Each position was much more strongly associated than the best classical allele (*DQB1*03:02, DQA1*02:01,* and *DRB1*04:01*, respectively).

Figure 3. Amino acid residue effect sizes. Case (colored) and control (unfilled) frequencies, as well as unadjusted univariate odds ratio, of each residue, at DQβ1#57, DRβ1#13 and DRβ1#71.

Figure 4. DQβ1#57, DRβ1#13 and DRβ1#71 are each located in the respective molecule's peptide-binding groove. DRβ1#13 and #71 line the P4 pocket of the DR molecule.

Figure 5. DQβ1#57, DRβ1#13 and DRβ1#71 explain over 90% of the phenotypic variance explained by the *HLA-DRB1-DQA1-DQB1* locus. Assuming the liability threshold model and a global T1D prevalence of 0.4%, all haplotypes in *HLA-DRB1-DQA1-DQB1* together explain 29.6% of total phenotypic variance. DQβ1#57 alone explains 15.2% of the variance; the addition of DRβ1#13 and #71 increases the explained proportion to 26.9%. Therefore, these three amino acid positions together capture over 90% of the signal within *HLA-DRB1-DQA1-DQB1.* In contrast, variation in *HLA-A, -B,* and *-DPB1* together explain approximately 4% of total variance. Genome-wide independently associated SNPs outside the HLA together explain about 9% of variance; rs678 (in the *INS* gene) and rs2476601 (in *PTPN22*) each explain 3.3% and 0.78%, respectively.

Figure 6. Interactions between common *HLA-DRB1-DQA1-DQB1* haplotypes lead to observed non-additive effects. We exhaustively tested the seven common haplotypes for pairwise interactions. Of the 21 possible pairs, eleven of them showed significant interactive effects. Along the perimeter, each segment represents one haplotype; red or blue color indicates risk or protective additive effect for each haplotype, respectively. Each arch connecting two haplotypes represents a significant interaction. Red indicates additional risk due to the

interaction beyond the additive effects; while blue indicates reduced risk (protection) due to the interaction beyond the additive effects. Thickness of the arches represents the effect size of the interaction (thicker red means larger risk while thicker blue means more protective.) See **Tables 2** and **S9** for *p*-values and effect sizes of all pairwise haplotypic interactions

Table 1. Haplotypes defined by DQβ1#57, DRβ1#13, and DRβ1#71 (control frequency > 0.1%). The three amino acid positions define 21 common haplotypes. We list their multivariate ORs, frequencies in controls and cases, as well as classical 4-digit alleles tagged by each haplotype. See **Supplementary Table 5B** for multivariate ORs and p-values of all 31 haplotypes formed by DQβ1#57, DRβ1#13, and DRβ1#71.

Haplotype	OR	CtrlFreq	CaseFreq	Classical *HLA-DQB1* Alleles	Classical *HLA-DRB1* Alleles
A-H-K	2.13	0.050	0.248	0201,0202,0302,0304,0305	0401,0409
A-H-E	1.33	0.005	0.016	0201,0202,0302,0304,0305	0402,0437
A-S-K (Ref)	1.00	0.145	0.332	0201,0202,0302,0304,0305	0301,0302,0304,1303
A-H-R	0.89	0.054	0.107	0201,0202,0302,0304,0305	0403,0404,0405,0406,0407,0408,0410,0411
A-S-E	0.53	0.001	0.001	0201,0202,0302,0304,0305	1102,1103,1301,1302,1304
D-F-R	0.48	0.012	0.014	0301,0303,0401,0402,0503,0601,0602,0603	0101,0102,0901,1001
A-F-R	0.43	0.001	0.001	0201,0202,0302,0304,0305	0101,0102,0901,1001
S-R-R	0.37	0.008	0.007	0502,0504	1601,1602
V-F-R	0.35	0.106	0.085	0501,0604,0609	0101,0102,0901,1001
V-S-E	0.34	0.040	0.030	0501,0604,0609	1102,1103,1301,1302,1304
D-G-R	0.32	0.039	0.029	0301,0303,0401,0402,0503,0601,0602,0603	0801-0806,1201,1202,1404,1415
D-H-K	0.27	0.068	0.042	0301,0303,0401,0402,0503,0601,0602,0603	0401,0409
V-F-E	0.24	0.013	0.006	0501,0604,0609	0103
A-Y-R	0.18	0.103	0.043	0201,0202,0302,0304,0305	0701
D-S-E	0.11	0.058	0.015	0301,0303,0401,0402,0503,0601,0602,0603	1102,1103,1301,1302,1304
D-F-E	0.08	0.004	0.001	0301,0303,0401,0402,0503,0601,0602,0603	0103
D-H-R	0.06	0.017	0.002	0301,0303,0401,0402,0503,0601,0602,0603	0403-0408,0410,0411
D-S-K	0.06	0.010	0.001	0301,0303,0401,0402,0503,0601,0602,0603	0301,0302,0304,1303
D-S-R	0.05	0.083	0.010	0301,0303,0401,0402,0503,0601,0602,0603	1101,1104,1106,1108,1305,1401,1402,1405, 1406,1407
D-Y-R	0.03	0.041	0.003	0301,0303,0401,0402,0503,0601,0602,0603	0701
D-R-A	0.02	0.140	0.005	0301,0303,0401,0402,0503,0601,0602,0603	1501

Table 2. Pairwise haplotypic interactions in *HLA-DRB1-DQA1-DQB1*. The table shows, for each given pair of haplotypes, the fold change in odds ratio (from additive effect-only) due to interaction. The "amino acids" column/row denote the residues at DQβ1#57, DRβ1#13, and DRβ1#71 corresponding to each haplotype. For each pair, the p-value of the interaction term is shown in parenthesis. Cells in **bold** indicate interactions that are significant after Bonferroni correction ($p < 0.05/21=0.0024$). Cells with underlines indicate the known *DR3/DR4* heterozygote effect. The odds ratio of a given diploid genotype is calculated as $\text{additive}_{\text{haplotype1}} \times \text{additive}_{\text{haplotype2}} \times \text{interaction}_{1,2}$ (see **Supplementary Table 9**).

		DRB1 DQA1 DQB1	**15:01 01:02 06:02**	**07:01 02:01 02:02**	**04:04 03:01 03:02**	**04:01 03:01 03:02**	**04:01 03:01 03:01**	**03:01 05:01 02:01**	**01:01 01:01 05:01**
		Amino acids	**D-R-A**	**A-Y-R**	**A-H-R**	**A-H-K**	**D-H-K**	**A-S-K**	**V-F-R**
DRB1 DQA1 DQB1	**Amino acids**	**Additive OR**	0.16	0.19	2.77	5.49	1.40	2.83	1.00 (Ref)
01:01 01:01 05:01	**V-F-R**	1.00 (Ref)	0.14 (0.004)	2.32 (0.04)	0.71 (0.19)	**2.16 (1.2×10^{-4})**	**1.95 (7.7×10^{-4})**	1.04 (0.77)	
03:01 05:01 02:01	**A-S-K**	2.83	**0.09 (1.2×10^{-5})**	2.24 (0.03)	**2.12 (1.9×10^{-4})**	**1.96 (1.2×10^{-5})**	**0.32 (9.2×10^{-11})**		
04:01 03:01 03:01	**D-H-K**	1.40	0.26 (0.03)	**4.78 (1.1×10^{-4})**	**0.23 (2.4×10^{-6})**	**0.48 (1.1×10^{-3})**			
04:01 03:01 03:02	**A-H-K**	5.49	0.36 (0.03)	**5.09 (4.2×10^{-5})**	**0.33 (3.5×10^{-5})**				
04:04 03:01 03:02	**A-H-R**	2.77	0.62 (0.03)	2.16 (0.08)					
07:01 02:01 02:02	**A-Y-R**	0.19	0.76 (0.74)						
15:01 01:02 06:02	**D-R-A**	0.16							

ONLINE METHODS

Sample collection

The dataset was provided by the Type 1 Diabetes Genetics Consortium[20], and consisted of (1) a UK case-control dataset (UK) and (2) a European family based dataset (Eur). All samples were collected after obtaining informed consent. The UK case-control dataset consisted of a total of 16,086 samples (6,670 cases and 9,416 controls) from 3 collections: (1) cases from the UK-GRID, (2) shared controls from the British 1958 Birth Cohort and (3) shared controls from Blood Services controls (data release February 4, 2012, hg18). The UK samples were collected from 13 regions, listed in **Supplementary Table 1**. The European Family based dataset consisted of 10,791 samples (5,571 affected children and 5,220 controls) from 2,699 European-ancestry families (data release January 30, 2013, hg18). All samples were genotyped on the ImmunoChip array. After quality control, 6,223 and 6,608 markers, respectively, were genotyped in the MHC region between 29 and 45Mb on Chromosome 6 in the two datasets. From the family data we constructed 1,662 pairs of pseudocase and pseudocontrol samples (see **Supplementary Note**).

HLA Imputation

We used SNP2HLA (default input parameters) to impute SNPs, amino acid residues, indels, and 2- and 4-digit classical alleles in eight HLA genes in the MHC between 29602876 and 33268403bp on Chromosome 6. We used the reference panel provided by T1DGC, which included 5,225 European samples classical typed for *HLA-A, B, C, DRB1, DQA1, DQB1, DPB1*, and *DPA1* 4-digit alleles[21,22]. The imputed genotype dataset included 8,961 binary markers prior to frequency tresholding. For each marker and each individual, two types of output were produced: a phased best-guess genotype (*e.g.* “AA/AT/TT”); and a dosage, which accounts for imputation uncertainty and can be continuous between 0 (0 copies of the alternative allele) and 2 (2 copies of the alternative allele).

We imputed the UK case-control dataset and European family dataset independently; within each set, cases and controls were imputed together to avoid disparity in imputation quality. 4,604 and 5,125 SNPs in the MHC region were used for imputation in the UK and Eur datasets, respectively. After combining the UK and Eur datasets, we excluded a total of 344 binary markers due to allele missingness or rareness (allele frequency < 0.05%); we then removed individuals who carried the missing or rare alleles. The post-quality control final dataset consisted of 18,832 samples, including 8,095 cases (including 1,662 pseudo-cases) and 10,737 controls (including 1,662 pseudo-controls).

Statistical framework

We test a given variant's association to disease status using the logistic regression model:

$$\log(odds_i) = \beta_0 + \sum_{j=1}^{m-1} \beta_{i,j} x_{i,j} + \sum_{k=1}^{n-1} \beta_{2,k} y_{i,k} + \beta_3 z_i$$

where variant x_i may be an imputed dosage or the best-guess genotype for a SNP, classical allele, amino acid, or haplotype. β_0 is the logistic regression intercept and $\beta_{1,j}$ is the additive effects of allele *j* of variant x_i. The number of alleles at each variant is *m*; for a binary variant (presence or absence of x_i), *m* equals 2. The covariate $y_{i,k}$ denotes each region of sample collection (*n*=14). We included sex as a covariate *z*. β_2 and β_3 are the effect sizes of the region and gender covariates, respectively.

To account for population stratification, we included the region codes as covariates (see **Supplementary Note**). Samples from the Eur dataset were considered as the 14th region. To assess the statistical significance of a tested variant, we calculated the improvement of fit of the model containing the test variant over the null model (only region and gender covariates). We calculate the model improvement as deviance defined by

$\Delta\text{deviance}_{alt-null} = -2\ln(\text{likelihood}_{alt}/\text{likelihood}_{null})$, which follows a χ^2 distribution with $m-1$ degrees of freedom, from which we calculate the p-value. We considered $p = 5\text{x}10^{-8}$ as the significance threshold.

Analysis of amino acid positions

To test amino acid effects within *HLA-DRB1-DQA1-DQB1*, we applied conditional haplotypic analysis. We tested each single amino acid position by first identifying the *m* amino acid residues occurring at that position, and partitioning all samples into *m* groups with identical residues at that position. We estimated the effect of each of the *m* groups using logistic regression model (including covariates as above), and assessed the significance of model improvement by Δdeviance compared to the null model, with *m–1* degrees of freedom. This is equivalent to testing a single multi-allelic locus for association with *m* alleles. To test the effect of a second amino acid position while conditioning on the first, we further update the model to include all unique haplotypes created by residues at both positions. We then test whether the updated model improves upon the previous model based on Δdeviance, taking into consideration the increased degrees of freedom.

Exhaustive test

To ensure that the independently associated amino acids were not emerging only as the result of forward-search which might possibly converge on local minima, we exhaustively tested of all possible combinations of one, two, three, and four amino acid positions in *HLA-DRB1, DQA1,* and *DQB1*. For each number of amino acid combination, we select the best model based on Δdeviance from the null (gender and region covariates only).

Haplotype-amino acid permutation analysis

Given the polymorphic nature of the HLA genes and the strong effect sizes in the DRB1-DQA1-

DQB1 locus, we wanted to assess whether the observed associations at DQβ1#57, DQβ1#13 and DQβ1#71 could emerge by chance, due to these positions' ability to tag classical alleles with different risks. To eliminate this possibility, we conducted a permutation test. In each permutation, for each of the three genes (e.g *HLA-DQB1, -DRB1,* and *-DQA1)* we preserved the sample's case/control status and gender/region covariates. To preserve allelic associations, we preserved groups of samples with the same amino acid sequence (4-digit classical allele) at each gene. We then randomly reassigned the amino acid sequence corresponding to each classical allele in each permutation, and repeated the forward-search analysis. We repeated this permutation 10,000 times, each time selecting the combination of two, three, and four amino acid positions that produce the best model (as measured by deviance). If the amino acids were merely tagging the effects of certain haplotypes, the effects we observed in the real data would not be more significant compared to those generated from permutations. To obtain the permutation-based *p*-value, we calculate the proportion of permutated models that exceeds the goodness-of-fit of the best model in the unpermuted data.

Testing for non-additivity and interactions

We defined haplotypes across the *HLA-DRB1-DQA1-DQB1* locus based on unique combinations of amino acid residues across the three genes. As non-additive effects can be observed only when sufficient numbers of homozygous individuals are present, we limited the interaction analysis to a subset of common haplotypes or classical alleles with frequencies greater than 5%. We excluded all individuals with one or more haplotypes that fell below this threshold.

We constructed an interaction model, which included additive terms for each common haplotype and interaction terms between all possible pairs of common haplotypes.

$$\log(odds_i) = \beta_0 + \sum_{j=1}^{m-1} \beta_{i,j} x_{i,j} + \sum_{j=1}^{m-1} \sum_{l=j+1}^{m} \emptyset_{j,l} x_{i,j} x_{j,l} + \sum_{k=1}^{n-1} \beta_{2,k} y_{i,k} + \beta_3 z_i$$

where $\phi$is the interaction effect size. We determined the improvement in fit with each successive model by calculating the change in deviance, and used a significance threshold of $p = 0.05/h$, where h is the total number of interactive parameters added to the original additive model.

HLA-DR3/DR4 classical allele interactions

To characterize the DR3/DR4 interaction, we defined 12 interaction terms, where each term represents a potential interaction between a classical allele on the DR3 haplotype (*DRB1*03:01*, *DQA1*05:01*, *DQB1*02:01*) and a classical allele on the DR4 haplotype (*DRB1*04:01* or *DRB1*04:04*, *DQA1*03:01*, *DQB1*03:02*). We only looked at *trans* interactions, since haplotype analyses already account for classical alleles that occur together in *cis*. We began with a null model that included additive effects for all haplotypes. Then, we individually tested each of the 12 interaction terms by adding each term to the null model separately. Once again, we used the change in deviance to assess the improvement in fit, using p=0.05/21=2.4x10^{-3} as the threshold.

Amino acid interaction analysis

To determine whether amino acid positions can explain haplotypic interactions, we defined haplotypes across the *DRB1-DQA1-DQB1* locus based on the 141 amino acid positions imputed in this locus. To ensure that a significant number of homozygous individuals were present, we excluded all amino acid residues with less than 5% frequency prior to creating the haplotypes. We also excluded any individual who had one or more amino acids that fell below this threshold.

We began with a null model that included additive effects for each amino acid haplotype. Then, for each pair of amino acid positions {*q, r*}, we added a set of $n_q \times n_r$ interaction terms, where each term specifies a *trans* interaction between one variant at each position, and n_p represents the total number of variants at position *p*. Each pair of amino acids was tested in a separate model, and we calculated the change in deviance to determine the improvement in fit. Monomorphic amino acid positions were excluded from this analysis, since they were constant across all individuals.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.