HLA-DPB1*04:01 Protects Genetically Susceptible Children from Celiac Disease Autoimmunity in the TEDDY Study

David Hadley, PhD[1,2], William Hagopian, MD, PhD[3], Edwin Liu, MD[4,5], Jin-Xiong She, PhD[6], Olli Simell, MD, PhD[7], Beena Akolkar, PhD[8], Anette-G. Ziegler, MD[9,10,11], Marian Rewers, MD, PhD[4], Jeffrey P. Krischer, PhD[12], Wei-Min Chen, PhD[13], Suna Onengut-Gumuscu, PhD[13], Teodorica L. Bugawan, BS[14], Stephen S. Rich, PhD[13], Henry Erlich, PhD[15], Daniel Agardh, MD, PhD[12,16] and the TEDDY Study Group

OBJECTIVES: **Tissue transglutaminase autoantibodies (tTGAs) represent the first evidence of celiac disease (CD) development. Associations of HLA-DR3-DQA1*05:01-DQB1*02:01 (i.e., DR3-DQ2) and, to a lesser extent, DR4-DQA1*03:01-DQB1*03:02 (i.e., DR4-DQ8) with the risk of CD differ by country, consistent with additional genetic heterogeneity that further refines risk. Therefore, we examined human leukocyte antigen (HLA) factors other than DR3-DQ2 for their contribution to developing tTGAs.**

METHODS: **The Environmental Determinants of Diabetes in the Young (TEDDY) study enrolled 8,676 infants at an increased HLA-DR-DQ risk for type 1 diabetes and CD into a 15-year prospective surveillance follow-up. Of those followed up, 21% (*n*=1,813) carried DR3-DQ2/DR3-DQ2, 39% (*n*=3,359) carried DR3-DQ2/DR4-DQ8, 20% (*n*=1701) carried DR4-DQ8/DR4-DQ8, and 17% (*n*=1,493) carried DR4-DQ8/DQ4. Within TEDDY, a nested case–control design of 248 children with CD autoimmunity (CDA) and 248 matched control children were genotyped for HLA-B, -DRB3, -DRB4, -DPA1, and -DPB1 genes, and the entire cohort was genotyped for single-nucleotide polymorphisms (SNPs) using the Illumina ImmunoChip. CDA was defined as a positive tTGA test at two consecutive clinic visits, whereas matching in those with no evidence of tTGAs was based on the presence of HLA-DQ2, country, and sex.**

RESULTS: **After adjustment for DR3-DQ2 and restriction to allele frequency (AF) ≥5%, HLA-DPB1*04:01 was inversely associated with CDA by conditional logistic regression (AF=44%, odds ratio=0.71, 95% confidence interval (CI)=0.53–0.96, *P*=0.025). This association of time to CDA and HLA-DPB1*04:01 was replicated with statistical significance in the remainder of the cohort using imputation for specific HLA alleles based on SNP genotyping (hazard ratio=0.84, 95% CI= 0.73–0.96, *P*=0.013).**

CONCLUSIONS: **HLA-DPB1*04:01 may reduce the risk of tTGAs, an early marker of CD, among DR3-DQ2 children, confirming that additional variants in the HLA region influence the risk for CDA.**

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at http://www.nature.com/ajg

Am J Gastroenterol 2015; 110:915–920; doi:10.1038/ajg.2015.150; published online 26 May 2015

[1]Division of Population Health Sciences and Education, St. George's, University of London, London, UK; [2]TransMed Systems, Pleasanton, California, USA; [3]Pacific Northwest Diabetes Research Institute, Seattle, Washington, USA; [4]Digestive Health Institute, Children's Hospital Colorado, Aurora, Colorado, USA; [5]Barbara Davis Center, University of Colorado Denver, Aurora, Colorado, USA; [6]Center for Biotechnology and Genomic Medicine, Georgia Regents University, Augusta, Georgia, USA; [7]Department of Pediatrics, University of Turku, Turku, Finland; [8]DDEMD, NIDDK, Bethesda, Maryland, USA; [9]Institute of Diabetes Research, Helmholtz Zentrum München, Oberschleissheim, Germany; [10]Klinikum rechts der Isar, Technische Universitaet Muenchen, Munich, Germany; [11]Forschergruppe Diabetes e.V., Neuherberg, Germany; [12]Pediatric Epidemiology Center, University of South Florida, Tampa, Florida, USA; [13]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA; [14]Department of Human Genetics, Roche Molecular Systems, Alameda, California, USA; [15]Children's Hospital of Oakland Research Institute, Oakland, California, USA; [16]Department of Clinical Sciences, Lund University, Malmo, Sweden. **Correspondence:** David Hadley, PhD, Division of Population Health Sciences and Education, St. George's, University of London, Tooting, London SW17 0RE, UK. E-mail: dhadley@sgul.ac.uk
Received 13 January 2015; accepted 13 April 2015

INTRODUCTION

Celiac disease (CD) is a common small bowel disorder with autoimmune features, typically preceded by the development of autoantibodies against tissue transglutaminase autoantibodies (tTGAs) (1). CD has strong genetic associations linked to the human leukocyte antigen (HLA) genes coding for the major histocompatibility complex class II complex on chromosome 6, and a similar pattern of genetic association has recently been found in relation to the development of tTGAs.

Individuals developing CD are mainly carriers of HLA-DR3 in linkage disequilibrium (LD) with the DQA1*05:01-DQB1*02:01 (abbreviated as DR3-DQ2) haplotype, and a minority carry HLA-DR4-DQA1*03:01-DQB1*03:02 (abbreviated as DR4-DQ8) (2). The DR3-DQ2/DR3-DQ2 genotype is known as having the highest risk for CD, with these homozygous individuals at two-fold risk compared with those individuals with one copy of DR3-DQ2, implying an HLA gene-dose effect (3–5). This genetic association was also recently published in relation to tTGAs, with a gene-dose effect of DR3-DQ2 on the risk of developing tTGAs in young children (6).

Both the DR3-DQ2 and DR4-DQ8 haplotypes are common in the Caucasian population (7), suggesting that, apart from environmental factors, additional genes contribute to CD susceptibility (8). Recent genome-wide association studies (GWAS) have identified risk variants (single-nucleotide polymorphisms (SNPs)) in 40 non-HLA loci that contribute to additional risk, although of minor effect, for CD (9). Others have fine-mapped the HLA complex for nearby risk alleles and have identified four putatively independent risk loci for CD (10).

In the ongoing multicenter The Environmental Determinants of Diabetes in the Young (TEDDY) study, the risk association for CD varied among the DR3-DQ2/DR3-DQ2 carriers across our study population, indicating that HLA alleles other than DR3-DQ2 might affect the risk (6). Given the similar genetic association between both tTGA autoimmunity and overt disease with the DR3-DQ2 locus, we examined whether other HLA variants are associated with autoimmunity development, as has been done for CD in the past, with various levels of association previously reported (11,12). To dissect this difference, we performed non-HLA-DR-DQ analysis on those children who developed tTGAs and in HLA-DR-DQ-matched controls.

METHODS

TEDDY cohort

TEDDY consortium of six clinical centers in Germany, Finland, Sweden, and the United States screened 424,788 newborns for HLA-DR-DQ genotypes associated with T1D and CD between 2004 and 2010 (13). Certain HLA genotypes were used as an eligibility criterion to select subjects with an increased risk for developing T1D (**Table 1**). Of the 21,589 screened newborns with an increased genetic risk, 8,676 were enrolled in prospective follow-up (**Table 2**). TEDDY was approved by local Ethical Institutional Review Boards and is monitored by an External Evaluation Committee formed by the National Institutes of Health.

Table 1. Eligible HLA genotypes in the TEDDY study

Letter	Full genotype	Abbreviated genotype
A	DRB1*04-DQA1*03-DQB1*03:02/ DRB1*03-DQA1*05-DQB1*02:01	DR4-DQ8/DR3-DQ2
B	DRB1*04-DQA1*03-DQB1*03:02/ DRB1*04-DQA1*03-DQB1*03:02	DR4-DQ8/DR4-DQ8
C	DRB1*04-DQA1*03-DQB1*03:02/ DRB1*08-DQA1*04-DQB1*04:02	DR4-DQ8/DR8-DQ4
D	DRB1*03-DQA1*05-DQB1*02:01/ DRB1*03-DQA1*05-DQB1*02:01	DR3-DQ2/DR3-DQ2
E	DRB1*04-DQA1*03-DQB1*03:02/ DRB1*04-DQA1*03-DQB1*02:02	DR4-DQ8/DR4*
F	DRB1*04-DQA1*03-DQB1*03:02/ DRB1*01-DQA1*01-DQB1*05:01	DR4-DQ8/1
G	DRB1*04-DQA1*03-DQB1*03:02/ DRB1*13-DQA1*01-DQB1*06:04	DR4-DQ8/13
I	DRB1*04-DQA1*03-DQB1*03:02/ DRB1*09-DQA1*03-DQB1*03:03	DR4-DQ8/9
J	DRB1*03-DQA1*05-DQB1*02:01/ DRB1*09-DQA1*03-DQB1*03:03	DR3-DQ2/9

HLA, human leukocyte antigen; TEDDY, The Environmental Determinants of Diabetes in the Young.

Genotypes A, B, C, and D confer general population eligibility but exclude DRB1*0403. Genotypes A through J confer eligibility to a first-degree relative with type 1 diabetes. Where DQB1*03:02 is noted, either this allele or DQB1*03:04 is allowed (13).

Table 2. Demographic summary of participating children in the TEDDY cohort

Country	*n*	Female (%)	Non-Caucasian (%)	DQ2-DR3/ DQ2-DR3 (%)
Finland	1,833	49.1	0	14.9
Germany	596	50.2	0.4	20.4
Sweden	2,525	49.4	0	21.8
USA	3,722	49.4	28.7	23.4
Total	8,676	49.4	87.6	20.9

TEDDY, The Environmental Determinants of Diabetes in the Young.

Only self-reported ethnicity was available for US subjects, with European subjects observed as being ethnically homogeneous by principal components analysis (17).

Screening for CD

Annual screening for CD autoimmunity with tTGAs starts at 2 years of age in TEDDY, as previously described (6). All tTGA-positive children at 2 and 3 years of age are retested after 3 months, whereas children initially positive at 4 years or older are retested after 6 months. For children who tested positive at 2 years of age, we retrospectively analyzed their banked serum samples (collected over multiple clinic visits earlier) for tTGA to estimate the first time point of antibody positivity (seroconversion). Children

who tested persistently positive for tTGA in two consecutive samples are defined as having CD autoimmunity (CDA). Guardians to children with a positive test result of tTGA were informed and, if confirmed on the next sample, were recommended to seek consultation of a pediatric gastroenterologist at their local hospital for further evaluation of CD.

tTGA measurements

tTGAs were detected by standardized radiobinding assay using recombinant ^{35}S-labeled tTG measured in two laboratories, as previously described (14). Serum samples from the US participants were screened for IgA-tTG at the Barbara Davis Center for Childhood Diabetes at the University of Colorado in Denver (cutoff for normal <0.05 relative units). Serum samples from European participants were assayed at the University of Bristol in Bristol for both IgA-tTG and IgG-tTG (cutoff for normal <1.3 relative units). All samples with IgA-tTG levels ≥0.01 units assessed in Denver were sent to Bristol for reanalysis, which was used as the reference laboratory (15).

Nested case–control design and genotyping

High-resolution sequence-based typing of the HLA-B, -DRB3, -DRB4, -DPA1, and -DPB1 genes was performed in 250 children (cases) who developed persistent positive tTGAs (i.e., CDA) and 250 control children matched on country, sex, and the presence of a DR3-DQ2 haplotype. Genomic DNA was typed based on Roche 454 sequencing. After excluding subjects with missing genotype data, 248 matched pairs remained for analysis (**Table 3**).

SNP genotyping

SNPs were genotyped by the Center for Public Health Genomics at University of Virginia, using the Illumina ImmunoChip SNP microarray (Illumina Inc., San Diego, CA) of ~196,000 SNPs selected from 186 regions associated with 12 autoimmune diseases. Individuals with low call rate (>5% SNPs missing), or discordance with reported sex and prior genotyping, were not considered in the analysis. In addition, SNP markers with low call rates (<95%) or with allele distributions strongly deviating from Hardy–Weinberg equilibrium ($P<10^{-6}$) were discarded (except for chromosome 6 due to HLA eligibility requirements). This resulted in a total of 7,023 subjects with genotype data on 176,586 SNPs.

Table 3. Proportions of nested case–control groups and replication cohort, at the time of data freeze

	n	tTGA (%)	Female (%)	Caucasian (%)	DR3-DQ2/DR3-DQ2 (%)
Cases	248	100	61.3	95.6	52.8
Controls	248	0	61.3	97.2	24.2
Replication cohort	4,514	9.6	47.34	100	20.8

tTGA, tissue transglutaminase autoantibody. The case–control group included 232 subjects with DR3-DQ2/DR4-DQ8, 51 with DR4-DQ8/DR4-DQ8, 18 with DR4-DQ8/DR8-DQ4, and 191 with DR3-DQ2/DR3-DQ2.

Cases were children with celiac disease autoimmunity and defined as being persistently positive to tTGAs.

Imputation

In addition to the 496 subjects in the nested case–control data set, SNP data on individuals from the remainder of the TEDDY cohort were used for the imputation of HLA alleles using the method developed by Jia *et al.* (16) Reference HLA data from the Type 1 Diabetes Genetics Consortium were used to impute values for the TEDDY cohort. We chose to limit the cohort to the more common HLA screening genotypes (**Table 1**, categories A–D) in order to make interpretation of the analysis as generalizable as possible. In addition to the nested case–control groups, this imputation provided HLA information on 4,514 unrelated subjects (median follow-up age=4.2 years, interquartile range=2.9 years) in addition to the nested case–control groups, of which 432 had developed CDA at the time of the data freeze (31 August 2013).

Statistical analysis

For the HLA-B, -DRB3, -DRB4, -DPA1, and -DPB1 genes, only alleles with frequency ≥5% were considered (**Table 4**). Association tests used conditional logistic regression models for nested case–control analyses, and Cox proportional hazard models for the replication data set from the remainder of the cohort with covariates as indicated in the results section below. Analyses accounted for the known covariates of HLA category, sex, and country.

RESULTS

We first analyzed the 496 subjects in the nested case–control data set, which includes 248 individuals with CDA. Of the alleles tested, only the DPB1*04:01 allele was significantly associated with CDA risk (odds ratio=0.71, 95% confidence interval (CI)=0.53–0.96, P=0.025) in a conditional logistic regression model. The model included the number of HLA-DR3-DQ2 copies as the only additional covariate, as the controls were already selected to match on sex and country.

Having imputed the HLA-DPB1*04:01 in a majority of the TEDDY cohort using ImmunoChip SNP genotypes, we were able to compare the results from imputation with the directly genotyped HLA. The concordance between the imputed and genotyped HLA allele was high, with only 11/481 (2.3%) results being discordant for HLA-DPB1*04:01.

Although the imputation method was validated using Type 1 Diabetes Genetics Consortium and 1958 British Birth Cohort subjects of European ancestry, we explored whether the restriction to a similar group in TEDDY would improve concordance. There was no evidence of improved concordance when restricting to non-Hispanic White subjects from the US (self-reported ethnicity, where available). Of the 11 discordant results, only 1 observation was dropped when limiting by ethnicity. Yet, when using the ethnic information, an additional 15 concordant results were included.

Table 4. Allele frequencies of HLA-B, -DRB3, -DRB4, -DPA1, and -DPB1 genes considered in the analysis, based on alleles with frequency ≥5%

HLA-B (%)	DPA1 (%)	DPB1 (%)	DRB3 (%)	DRB4 (%)
*07:02 (5.2)	*01:03 (70.1)	*01:01 (21.8)	*01:01 (54.4)	*01:01 (34.6)
*08:01 (46.6)	*02:01 (25.0)	*02:01 (12.5)	*02:02 (5.4)	
*15:01 (12.6)		*04:02 (8.4)		
*18:01 (4.5)		*03:01 (9.1)		
*40:01 (6.5)		*04:01 (35.7)		

Table 5. Results of Cox proportional hazards model (*n*=4,514; 432 failures), with HLA-DPB1*04:01 being the only continuous variable in the model

Covariate	Hazard ratio (95% CI)	*P*-value
HLA-DPB1*04:01	0.84 (0.73–0.96)	0.013
Country		
Finland	1.35 (1.04–1.74)	0.024
Germany	1.04 (0.66–1.62)	0.872
Sweden	1.38 (1.10–1.73)	0.006
Sex (female)	1.71 (1.41–2.07)	0.000
HLA category		
B	0.61 (0.46–0.83)	0.001
C	0.19 (0.12–0.32)	0.000
D	2.29 (1.86–2.82)	0.000

CI, confidence interval: HLA, human leukocyte antigen.
Reference country is the USA, reference sex is male, and reference HLA category is A (see **Table 1**).

Table 6. Frequency of HLA-DPB1*04:01 allele among the four common HLA-DR-DQ risk groups

HLA category	Number of copies and genotype frequency of HLA-DPB1*04:01			
	0	1	2	Total
A: DR4-DQ8/DR3-DQ2	630 (34.5%)	897 (49.1%)	300 (16.4%)	1,827
B: DR4-DQ8/DR4-DQ8	247 (26.1%)	463 (49.0%)	235 (24.9%)	945
C: DR4-DQ8/DR8-DQ4	247 (30.8%)	398 (49.6%)	157 (19.6%)	802
D: DR3-DQ2/DR3-DQ2	398 (42.3%)	425 (45.2%)	117 (12.4%)	940
Total	1,522 (33.7%)	2,183 (48.3%)	809 (17.9%)	4,514

HLA, human leukocyte antigen.

Therefore, the imputation method appeared to work effectively in individuals of both European and non-European ancestry.

Regardless of the high imputation concordance in different ethnic groups, to avoid the possibility of association owing to population stratification, we limited our analysis to the imputed results in the 4,514 subjects with a self-reported Caucasian ancestry. Ethnicity information was only available for the subjects at US centers, which makes up 43% of the TEDDY population. In those subjects, 73% self-reported to be Caucasian (**Table 2**). In the few instances in which self-reported ethnicity was available at European centers, non-Caucasian subjects were removed from the analysis. For the majority of subjects at the European centers, we have previously reported homogeneity in a principal components analysis within each country, as well as strong similarity between the European countries and the US Caucasians (17).

Proportional hazards modeling was performed to test for association with time to seroconversion of tTGAs in CDA children in those unrelated TEDDY subjects who were not part of the nested case–control group. Using a Cox proportional hazards model, with sex, country, and HLA genotype category as covariates, we tested for the association between HLA-DPB1*04:01 and time to tTGA development (**Table 3**). To account for any potential differences between DR4-DQ8/DR4-DQ8 and DR4-DQ8/DR8-DQ4, we adjusted for HLA category, rather than the number of copies of DR3-DQ2. The association between HLA-DPB1*04:01 and time to seroconversion was significantly replicated in the larger cohort (hazard ratio=0.84, 95% CI=0.73–0.96, *P*=0.013; **Table 5**).

The HLA-DPB1*04:01 allele was common among the four HLA-DR-DQ risk categories used in the study eligibility criteria (**Table 6**). We observed that the HLA categories correspond to the familiar pattern of DR3 conferring risk: a doubling of risk between category A (DR4-DQ8/DR3-DQ2) to category D (DR3-DQ2/DR3-DQ2) and a decrease in risk with no copies of DR3-DQ2 (categories B and C). In addition, as expected, the effect of female sex is associated with a significant increase in the risk of CDA. Despite accounting for the above covariates, and including a new HLA marker for risk to the model, there is also a significant difference in risk between countries: Swedish and Finnish subjects both having an increased risk of CDA compared with subjects in the USA.

DISCUSSION

In this study, we analyzed genetic associations with the phenotype of persistent tTGAs, which we termed CDA. This represents an early stage of CD, which generally precedes the clinical diagnosis by months to years, likely owing to the rarity of general pediatric antibody screening, as well as obvious logistical difficulties associated with endoscopic biopsy. It is also true that some children

may develop transient tTGAs that fade in titer without leading to overt CD, even in the absence of dietary alterations. Therefore, this analysis complements those using the clinical CD phenotype and may provide information on earlier steps in the pathway leading to development of CD.

HLA-DR-DQ is the major genetic determinant of CD risk, and the paucity of published genetic associations with CDA at other loci emphasizes the central role that HLA-DR-DQ has in disease risk and development (18). This paucity is even greater when considering non-DR-DQ genetic loci in the HLA region, where analyses must also be corrected for the effect of DR-DQ.

In a previous study, the prediction of CD was improved by 11% by adding a group of non-HLA risk variants to a model including HLA (19). This suggests that additional non-HLA genes may contribute to only small changes to disease risk (20). However, two findings first focused attention on the independent associations of *HLA-DPB1* alleles on the DR3 background. Kagnoff *et al.* (8) described a significantly increased frequency of both the DPB1 and DPB-3 alleles in CD patients. The same year, a paper by Bugawan *et al.* (11) described an increased frequency of DPB-3 and DPB-4.2 in CD.

In the current study, the HLA genotyping found no common alleles, in addition to DR3-DQ2, that were associated with an increased risk for CDA. Instead, the HLA allele DPB1*04:01 was associated with a 29% reduction in risk, suggesting that variants in the HLA region may influence the risk for CDA across all of the HLA groups included in the TEDDY study. Our results therefore support a biological interaction between DPB1*04:01 and DR3-DQ2 on the genetic susceptibility for developing CDA.

Although increases in other *HLA-DPB1* alleles are compatible with a decreased frequency of DPB1*04:01 in CD patients, neither of the papers on HLA-DP associations with CD risk (8,11) assessed the DPB1*04:01 frequency directly, perhaps owing to genotyping limitations at the time. Interestingly, the DPB1*04:01 allele has been associated with other autoimmune diseases showing both a protective effect for type 1 diabetes (21) and for multiple sclerosis (22), as well as an increased risk for rheumatoid arthritis (23), pemphigus (24), and vasculitis (25). To our knowledge, the association between DPB1*04:01 and CDA or CD has not been described before.

To put the magnitude of the DPB1 association in context, we compare our finding with the published literature for protective effect alleles that have been discovered through GWAS for overt CD. Our observed odds ratio for DPB1*04:01 of 0.71 is the same as the magnitude of the largest published protective finding for CD risk in a compilation of GWAS studies (rs13132308, MAF 0.166) (20). In comparison, the allele frequency of HLA-DPB1*04:01 was 0.372 in the T1DGC imputation reference data set (26) and 0.421 in the TEDDY imputed data set. Although the comparison of odds ratios is between different outcomes, the combination of substantial effect size along with relatively high allele frequency suggests that our finding could be as important in disease development as any of the non-HLA SNP associations found in GWAS data sets.

The HLA-DPB1*04:01 allele may be in LD with the 8.1 ancestral multi-gene haplotype, known to be associated with various autoimmune diseases (27): HLA A*0101: Cw*0701: B*0801: DRB1*0301: DQA1*0501: DQB1*0201. Given that DR3-DQ2 was included as a covariate in the modeling, it is likely that the observed association of HLA-DPB1*04:01 with the risk of CDA development is independent of the 8.1 ancestral haplotype in this population. Using the TEDDY imputed data, the LD between DR3-DQ2 from study screening and imputed HLA-DPB1*04:01 in 4514 European ancestry subjects was found to be r^2=0.01, D'=0.14.

The TEDDY study is enriched for children with CDA and CD compared with other birth cohorts, owing to the screening based on increased risk HLA haplotypes. Although this offers insights into effects that may otherwise be difficult to observe in the general population, one drawback is a lack of Hardy–Weinberg equilibrium in the HLA screening polymorphisms (and other genetic variants in LD with them). Departure from Hardy–Weinberg equilibrium violates the assumptions of many useful statistical techniques, such as haplotype imputation from genotypes. Another statistical technique that requires Hardy–Weinberg equilibrium is the calculation of LD between DR3-DQ2 and DPB1*04:01. Because of these limitations, and as the HLA region has particularly high and complex LD, future studies may wish to investigate the haplotype effect of DPB1*04:01 and its LD with the 8.1 ancestral haplotype, although the LD appears to be low in our data.

In summary, we believe that this is the first publication of HLA allele associations with the development of CDA beyond the DR3-DQ2 haplotype. Because CDA appears earlier in disease development than clinical CD, HLA-DPB1*04:01 typing may increase the ability to define children at risk for CD, allowing for better prospective prediction of CD risk in birth cohorts. For prospective studies, such as TEDDY, better definition of disease risk should enable improved focus on environmental risk factors that may act as triggers for CDA in children at genetic risk. Further, these findings may allow improvement of population-based strategies for tTGA surveillance of young children based on HLA genotyping.

ACKNOWLEDGMENTS

This work was funded by U01 DK63829, U01 DK63861, U01 DK63821, U01 DK63865, U01 DK63863, U01 DK63836, U01 DK63790, UC4 DK63829, UC4 DK63861, UC4 DK63821, UC4 DK63865, UC4 DK63863, UC4 DK63836, UC4 DK95300, and UC4 DK100238, and Contract No. HHSN267200700014C from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Child Health and Human Development (NICHD), National Institute of Environmental Health Sciences (NIEHS), Juvenile Diabetes Research Foundation (JDRF), and Centers for Disease Control and Prevention (CDC). This work supported in part by the NIH/NCATS Clinical and Translational Science Awards to the University of Florida (UL1 TR000064) and the University of Colorado (UL1 TR001082). The complete TEDDY Study Group is listed in the **Supplementary file**.

CONFLICT OF INTEREST

Guarantor of the article: David Hadley, PhD.

Specific author contributions: Conception and design: D.H., W.H., E.L., J.X.S., O.S., B.A., A.G.Z., M.R., J.P.K., W.M.C., H.E., and D.A.; acquisition of data: W.H., J.X.S., O.S., A.G.Z., J.P.K., S.O.G., T.L.B., S.S.R., H.E., and D.A.; statistical analysis and interpretation of data: D.H., W.H., A.G.Z., J.P.K., S.S.R., H.E., and D.A.; obtaining funding: W.H., J.X.S., O.S., A.G.Z., and J.P.K.; drafting of the manuscript: D.H., W.H., and D.A.; revision and approval of the final manuscript: D.H., W.H., E.L., J.X.S., O.S., B.A., A.G.Z., M.R., J.P.K., W.M.C., S.O.G., S.S.R., H.E., and D.A.

Financial support: National Institute of Health (NIH), Department of Health and Human Services (DHHS), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Disease (NIAID), National Institute of Environmental Health Sciences (NIEHS), National Institute of Child Health and Human Development (NICHD), Juvenile Diabetes Research Foundation (JDRF), Centers for Disease Control and Prevention (CDC).

Potential competing interests: None.

Study Highlights

WHAT IS CURRENT KNOWLEDGE

- ✓ HLA-DR-DQ is well known for increasing the risk of celiac disease (CD).
- ✓ Other studies have investigated SNP associations with CD, finding numerous with lesser effect than HLA-DR-DQ.
- ✓ HLA-DR-DQ has recently been shown to also have a strong association with the development of persistent tissue transglutaminase (CD autoimmunity).

WHAT IS NEW HERE

- ✓ We explored HLA genes and CD autoimmunity in a young genetically high-risk population, controlling for HLA-DR-DQ.
- ✓ In this population, HLA-DPB1*04:01 was found, and replicated, as having a protective association with CD autoimmunity.
- ✓ This effect size was similar to the largest protective association found thus far in non-HLA SNP meta-analysis for CD.
- ✓ The understanding of the large role that HLA has in CD can be better understood.
- ✓ Future studies of CD autoimmunity can better account for this novel protective genetic association, to elucidate other environmental and genetic determinants of disease.

REFERENCES

1. Dieterich W, Ehnis T, Bauer M *et al.* Identification of tissue transglutaminase as the autoantigen of celiac disease. Nat Med 1997;3:797–801.
2. Lundin KE, Scott H, Hansen T *et al.* Gliadin-specific, HLA-DQ (alpha 1*0501,beta 1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. J Exp Med 1993;178:187–96.
3. Ploski R, Ek J, Thorsby E *et al.* On the HLA-DQ (alpha 1*0501, beta 1*0201)-associated susceptibility in celiac disease: a possible gene dosage effect of DQB1*0201. Tissue Antigens 1993;41:173–7.
4. Murray JA, Moore SB, Van Dyke CT *et al.* HLA DQ gene dosage and risk and severity of celiac disease. Clin Gastroenterol Hepatol 2007;5:1406–12.
5. Margaritte-Jeannin P, Babron MC, Bourgey M *et al.* HLA-DQ relative risks for coeliac disease in European populations: a study of the European Genetics Cluster on Coeliac Disease. Tissue Antigens 2004;63:562–7.
6. Liu E, Lee HS, Aronsson CA *et al.* Risk of pediatric celiac disease according to HLA haplotype and country. N Engl J Med 2014;371:42–9.
7. Gujral N, Freeman HJ, Thomson AB. Celiac disease: prevalence, diagnosis, pathogenesis and treatment. World J Gastroenterol 2012;18:6036–59.
8. Kagnoff MF, Harwood JI, Bugawan TL *et al.* Structural analysis of the HLA-DR, -DQ, and -DP alleles on the celiac disease-associated HLA-DR3 (DRw17) haplotype. Proc Natl Acad Sci USA 1989;86:6274–8.
9. Smyth DJ, Plagnol V, Walker NM *et al.* Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med 2008;359:2767–77.
10. Ahn R, Ding YC, Murray J *et al.* Association analysis of the extended MHC region in celiac disease implicates multiple independent susceptibility loci. PLoS One 2012;7:e36926.
11. Bugawan TL, Angelini G, Larrick J *et al.* A combination of a particular HLA-DP beta allele and an HLA-DQ heterodimer confers susceptibility to coeliac disease. Nature 1989;339:470–3.
12. Caffrey C, Hitman GA, Niven MJ *et al.* HLA-DP and coeliac disease: family and population studies. Gut 1990;31:663–7.
13. Hagopian WA, Erlich H, Lernmark A *et al.* The environmental determinants of diabetes in the young (TEDDY): genetic criteria and international diabetes risk screening of 421 000 infants. Pediatr Diabetes 2011;12:733–43.
14. Vehik K, Fiske SW, Logan CA *et al.* Methods, quality control and specimen management in an international multicentre investigation of type 1 diabetes: TEDDY. Diabetes Metab Res Rev 2013;29:557–67.
15. Williams AJ, Annis P, Lock RJ *et al.* Evaluation of a high-throughput second antibody radiobinding assay for measuring IgA antibodies to human tissue transglutaminase. J Immunol Methods 1999;228:81–5.
16. Jia X, Han B, Onengut-Gumuscu S *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One 2013;8:e64683.
17. Törn C, Hadley D, Lee HS *et al.* Role of Type 1 diabetes associated SNPs on risk of autoantibody positivity in the TEDDY Study. Diabetes 2014;64:1818–29.
18. Sollid LM, Pos W, Wucherpfennig KW. Molecular mechanisms for contribution of MHC molecules to autoimmune diseases. Curr Opin Immunol 2014;31C:24–30.
19. Romanos J, Rosén A, Kumar V *et al.* Improving coeliac disease risk prediction by testing non-HLA variants additional to HLA variants. Gut 2014;63:415–22.
20. Trynka G, Hunt KA, Bockett NA *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet 2011;43:1193–201.
21. Noble JA, Valdes AM, Thomson G *et al.* The HLA class II locus DPB1 can influence susceptibility to type 1 diabetes. Diabetes 2000;49:121–5.
22. Yoshimura S, Isobe N, Yonekawa T *et al.* Genetic and infectious profiles of Japanese multiple sclerosis patients. PLoS One 2012;7:e48592.
23. Perdriger A, Guggenbuhl P, Chales G *et al.* The role of HLA-DR-DR and HLA-DR-DP interactions in genetic susceptibility to rheumatoid arthritis. Hum Immunol 1996;46:42–8.
24. Koc CK, Sallakci N, Akman-Karakaş A *et al.* Human leukocyte antigens class I and class II in patients with pemphigus in southern Turkey. Int J Dermatol 2013;52:53–8.
25. Watts RA, MacGregor AJ, Mackie SL. HLA allele variation as a potential explanation for the geographical distribution of granulomatosis with polyangiitis. Rheumatology (Oxford) 2014;54:359–62.
26. Barrett JC, Clayton DG, Concannon P *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009;41:703–7.
27. Price P, Witt C, Allcock R *et al.* The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. Immunol Rev 1999;167:257–74.