Epistasis Between *HLA-DRB1*16:02:01* and *SLC16A11 T-C-G-T-T* Reduces Odds for Type 2 Diabetes in Southwest American Indians

Robert C. Williams,[1] Robert L. Hanson,[1] Bjoern Peters,[2] Kendall Kearns,[2] William C. Knowler,[1] Clifton Bogardus,[1] and Leslie J. Baier[1]

Diabetes 2024;73:1002–1011 | https://doi.org/10.2337/db23-0925

We sought to identify genetic/immunologic contributors of type 2 diabetes (T2D) in an indigenous American community by genotyping all study participants for both high-resolution *HLA-DRB1* alleles and *SLC16A11* to test their risk and/or protection for T2D. These genes were selected based on independent reports that *HLA-DRB1*16:02:01* is protective for T2D and that *SLC16A11* associates with T2D in individuals with BMI <35 kg/m^2. Here, we test the interaction of the two loci with a more complete data set and perform a BMI sensitivity test. We defined the risk protection haplotype of *SLC16A11*, *T-C-G-T-T*, as allele 2 of a diallelic genetic model with three genotypes, *SLC16A11*11*, **12*, and **22*, where allele 1 is the wild type. Both earlier findings were confirmed. Together in the same logistic model with BMI ≥35 kg/m^2, *DRB1*16:02:01* remains protective (odds ratio [OR] 0.73), while *SLC16A11* switches from risk to protection (OR 0.57 [22*] and 0.78 [**12*]); an added interaction term was statistically significant (OR 0.49 [**12*]). Bootstrapped (b = 10,000) statistical power of interaction, 0.4801, yielded a mean OR of 0.43. Sensitivity analysis demonstrated that the interaction is significant in the BMI range of 30–41 kg/m^2. To investigate the epistasis, we used the primary function of the *HLA-DRB1* molecule, peptide binding and presentation, to search the entire array of 15-mer peptides for both the wild-type and ancient human *SLC16A11* molecules for a pattern of strong binding that was associated with risk and protection for T2D. Applying computer binding algorithms suggested that the core peptide at *SLC16A11* D127G, FSAFASGLL, might be key for moderating risk for T2D with potential implications for type 1 diabetes.**

ARTICLE HIGHLIGHTS

- This study enlarged our sample of high-resolution *HLA-DRB1* alleles and 5 individually typed mutations for the *SLC16A11* locus and used these to test for protection, risk, and interaction for type 2 diabetes.
- We confirmed our earlier reports of protection (*DRB1*16:02*) and risk (*SLC16A11*) and used all genotypes in a sensitivity analysis for BMI.
- *HLA-DRB1*16:02* was found to be protective, and sensitivity analysis demonstrated that *SLC16A11* is a risk in lower BMI strata and protective in higher ones.
- Epistasis for individuals with *DRB1*16:02* and *T-C-G-T-T* reduces the odds for type 2 diabetes in a BMI range of 30–41 kg/m^2, and binding studies implicate core peptide FSAFASGLL.

In 2011, we reported the protective effect of *HLA-DRB1*02* on the susceptibility to type 2 diabetes (T2D) via an effect on insulin secretion in a population of southwestern indigenous Americans that suggested a role for the immune system in T2D as well as type 1 diabetes (T1D) (1). In a recent multiethnic study of T2D that included a stratum of American Indians, it was reported that 41.3% of patients had cellular islet autoimmunity and 13.5% had humoral islet autoimmunity (2). The natural correlate of these findings is to search for risk and protection *HLA* alleles for T2D in our southwestern indigenous sample with a very high prevalence of the disease (3,4). Since our 2011 article (1), we

[1]Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, AZ
[2]La Jolla Institute for Immunology, La Jolla, CA
Corresponding author: Robert C. Williams, williamsr@mail.nih.gov

Received 22 November 2023 and accepted 18 March 2024

This article contains supplementary material online at https://doi.org/10.2337/figshare.25460452.

have *HLA* typed nearly the entire study group with targeted next-generation sequencing and computer algorithms that revealed high-resolution alleles. Our *HLA-DRB1*02* resolves to *DRB1*16:02:01* (5–7). In the present article, we present an analysis of the role of this protective molecular allele and its interaction with the monocarboxylate transporter *SLC16A11*, a known T2D risk locus specific to American ancestral groups, and determine their combined contributions to the genetic immunological risk for T2D. We used binding prediction algorithms to identify 15-mer *SLC16A11* peptides that might work synergistically with *DRB1*16:02:01* in the mechanics of protection, develop polygenic statistical models that reflect their joint action, and test for interaction and epistasis.

A haplotype of *SLC16A11*, defined by five single nucleotide polymorphisms (SNPs), four missense and one synonymous, including *rs117767867(T)*, *rs13342692(C)*, *rs13342232(G)*, *rs75418188(T)*, and *rs75493593(T)*, was reported by the Slim Initiative Genomic Medicine for the Americas (SIGMA) Type 2 Diabetes Consortium to be strongly associated with T2D in Mexicans and Latin Americans, with age and weight effects (8). In addition, this ancient human haplotype dates from the origins of modern humans and the genus *Homo* as much as 500,000–750,000 years ago (9). We subsequently used the G340S missense mutation and reported risk for T2D in leaner people and protection from the disease in very heavy individuals (10). The association was replicated in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) in a large Mexican American study but not in other Latino samples with little indigenous American admixture (11). The search for the association of the ancient human haplotype with other expressions of T2D and with clinical markers of the disease has continued to be an active area of research (12–21). The ancient human *SLC16A11* haplotype shares features with the concept of a private allele that was first defined by James Neel (22).

RESEARCH DESIGN AND METHODS

Population

A sample of 6,669 Southwest American Indians who have some portion of American Indian heritage, who participated at some time in a longitudinal study of T2D conducted from 1965 to 2007, and who had exome sequencing data available were chosen for this study.

Diabetes Diagnosis

Diabetes was diagnosed using 1997 American Diabetes Association criteria: fasting plasma glucose $\geq$7.0 mmol/L, 2-h plasma glucose $\geq$11.1 mmol/L, or a diagnosis made in routine clinical care as previously described (10).

Genotyping Method

The five coding SNPs that define the ancient *SLC16A11* haplotype were genotyped for association analyses using the TaqMan Allelic Discrimination (AD) Assay (Applied Biosystems) on an ABI 7900HT (Applied Biosystems). For genotyping using AD-PCR, we followed the manufacturer's instructions (Applied Biosystems). AD-PCR primers and probes were custom designed from a published sequence using the Custom TaqMan Assay Design Tool (Thermo Fisher Scientific) (23–25).

HLA Typing

HLA typing was performed from exome sequencing data, as previously described (5–7). *HLA* alleles were resolved at the four-field level of resolution from whole-genome sequences and at the three-field level from exomes. All analyses were performed at three-field resolution.

Statistical Analysis

Allele frequencies were calculated either by gene counting or the estimator-maximum (EM) method (26). Haplotype frequencies were computed by the EM algorithm (27). Haplotypes were defined for each functional *DRA-DRB1* heterodimer assigned to genotypes by a weighted probability function based on the estimated EM haplotype frequencies (5–7). Descriptive statistics and general linear models for the association of T2D with covariates were calculated using SAS 9.4 software (28). Covariates for the linear models included age, sex, and the first 5 principal components (PCs) derived from a genome-wide association study (PC1–PC5) (29). Each observation in the logistic regression was corrected for sibship using the REPEATED option in the SAS PROC GENMOD procedure that corrects for the familial correlation of the members of the sibship. The statistical power of the interaction term was calculated by a 10,000-iteration bootstrap of the fully parametrized logistic regression, with the power defined by the proportion of iterations <0.05. The sensitivity analysis was calculated from logistic regressions in strata defined by BMI categories $<$BMI and $\geq$BMI in BMI range 25–45 kg/m^2.

MHC Binding Prediction

MHC binding prediction for *HLA-DRB1* was performed using the algorithm NetMHCIIpan 4.0 EL from the Immune Epitope Database and Analysis Resource (30–32). For input, the program uses 15-mer peptides from a FASTA file for a given protein and a vector of class II alleles and calculates for each a probability of strong binding score and a percentile rank (rank%). The rank% compares the peptide's score against the scores of 5 million random 15-mers selected from the Swiss-Prot database. A small rank% indicates high affinity of the allele with the 15-mer peptide. The program moves over one amino acid at a time and calculates the two parameters for each MHC allele, and then, by this method, continues to the end of the amino acid sequence.

Data and Resource Availability

Data and resource availability will be considered following joint guidelines developed with the indigenous community.

RESULTS

The Ancient Human *SLC16A11* Haplotype Has a High Allele Frequency

In Supplementary Table 1, the EM-estimated frequencies of *DRB1* alleles are presented. *DRB1*16:02:01(DRB1*02)*, has an EM-estimated allele frequency of 0.077. The haplotype frequencies for *SLC16A11* reveal the segregation of two common specificities, the wild type (*C-T-A-C-G*, the 1 allele) and the ancient human risk (*T-C-G-T-T*, the 2 allele), and results in a simple diallelic, additive genetic model (*SLC16A11*11*, **12*, **22*). In the all-people stratum, the frequency of the wild-type allele is 0.603 and the risk allele, 0.397.

Association of *SLC16A11* With T2D Is a Function of BMI

The allele *DRB1*16:02:01*, in a dominant model, and the locus *SLC16A11* were first incorporated in fully specified logistic models (LMs) with T2D as the dependent variable in three strata: all people, BMI <35 kg/m^2, and BMI ≥ 35 kg/m^2 (Table 1). Furthermore, four logistic regressions were run within each stratum: each locus alone, together, and with interaction (Table 2). For all people, and when included with *SLC16A11*, *DRB1*16:02:01* is significantly associated with T2D and protective, with an odds ratio (OR) of 0.80 (95% CI 0.66, 0.97); when tested alone or with the *DRB1* allele, *SLC16A11* has ORs not significantly different from 1.0. In the BMI <35 kg/m^2 stratum, *DRB1*16:02:01* has similar ORs to the larger sample, with and without adjustment for *SLC16A11* (0.86 and 0.83, respectively) but with marginal statistical significance. However, carriers of the *SLC16A11* haplotype have a strong risk for T2D, with and without adjustment for *DRB1*16:02:01*. Genotype *SLC16A11*22* has an OR in both models of $\sim$1.8 (1.3, 2.3; $P < 0.0001$), while among heterozygotes, it is $\sim$1.3 (1.1, 1.6; $P \sim 0.01$). In the interaction model LM8, the ORs for *SLC16A11* have a similar magnitude and significance, but the interaction term has an OR not significantly different from 1.0.

*DRB1*16:02:01*, *SLC16A11* Interact to Amplify Protection for People With BMI ≥ 35 kg/m^2

In the BMI ≥ 35 kg/m^2 stratum, the early human *SLC16A11 T-C-G-T-T* haplotype switches from risk to strong protection for T2D (Table 2 and Fig. 1), with an OR of 0.58 (95% CI 0.43, 0.78) and 0.78 (0.64, 0.96) for *SLC16A11*22* and *SLC16A11*12*, respectively, in both LM10 and LM11. Allele *DRB1*16:02:01* shows T2D protection, without adjustment for *SLC16A11* (OR 0.73; 0.56, 0.95) and with adjustment for *SLC16A11* (OR 0.73; 0.56, 0.95). The surprising result is model LM12, in which the interaction term between *DRB1*16:02:01* and the *SLC16A11*12* is statistically significant (OR 0.49; 0.28, 0.86) and the *DRB1* allele and the early human *SLC16A11* haplotype together reduce the risk of T2D by 51% (Fig. 1 and Supplementary Figs. 1 and 2). In contrast, there is no significant interaction between the *DRB1* allele and *SLC16A11*22* in LM12.

Joint, Polygenic, Functional Analysis

We designed a different, joint model for the combined protective alleles at *DRB1*16:02:01* and *SLC16A11* in the BMI ≥ 35 kg/m^2 stratum where we observed the protective effect of *SLC16A11*. People with three or four protective alleles also had an OR of 0.49, with a model significance $P < 0.0001$ (Supplementary Table 2).

Statistical Power Estimation

To estimate the power of the interaction term in LM12 (Table 2) with the *SLC16A11*12* and *DRB1*16:02:01+* we performed a 10,000-iteration bootstrap of the model. The sample size for the BMI ≥ 35 kg/m^2 stratum is 2,513. For the interaction term, 4,801 iterations had $P < 0.05$ (Supplementary Fig. 3). Therefore, the power estimate is 0.4801, or 48%. The distribution of these significant P values has a mean very close to that of the term in the logistic model of $\sim$0.01. The bootstrap also allows us to estimate the interaction term's bootstrap-mean or moment for the β and OR with 95% confidence interval (OR 0.43; 0.22, 0.80). These are empirical CIs that are free of any assumptions about the shape of the distribution of the data.

Sensitivity Analysis

We first chose BMI ≥ 35 kg/m^2 as the stratum because the mean in the entire sample is 34.8 and we had used this cutoff in an earlier article (10). However, we decided to perform a sensitivity analysis to further define the role of BMI in the associations and interaction. Such an analysis usually tracks a single event, i.e., risk, whereas we have two events to monitor, i.e., risk and protection. Therefore, we performed three separate sets of logistic regressions with BMI values from 25 to 45 kg/m^2: protection ($\geq$BMI strata in Table 3 and Supplementary Table 3) and risk ($<$BMI strata in Supplementary Table 4). In Table 3, between BMIs of 31 and 39 kg/m^2, both the *SLC16A11*22* and the interaction term have statistical significance with

Table 1—Description of the Southwest American Indian sample

Stratum	*n*	BMI, mean (SD)	Age, mean (SD)	T2D, %	Female sex, %	*DRB1 16:02:01*, %	*SLC16A11 T-C-G-T-T*, %
All people	5,707	34.8 (8.6)	36.4 (15.2)	36.4	57.9	14.7	63.8
BMI <35 kg/m^2	3,194	28.8 (4.0)	36.7 (16.5)	33.8	52.3	14.4	64.7
BMI ≥ 35 kg/m^2	2,513	42.4 (6.7)	36.1 (13.4)	39.6	65.0	15.6	62.7

Table 2—Association of *HLA-DRB1*16:02:01* and *SLC16A11* (risk allele *T-C-G-T-T* = 2) with T2D in Southwest American Indians[a]

	Explanatory loci						Interaction				
	*DRB1*16:02:01*		*SLC16A11*22*		*SLC16A11*12*		**16:02:01 × *12*		**16:02:01 × *22*		
LM	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	Overall *P*[b]
All individuals (*N* = 5,707)											
1	0.80 (0.66, 0.97)	0.0214									
2			1.02 (0.84, 0.125)	>0.05	1.00 (0.87, 1.16)	>0.05					
3	0.80 (0.66, 0.97)	0.0210	1.03 (0.84, 1.25)	>0.05	1.01 (0.87, 1.17)	>0.05					
4	0.94 (0.67, 1.30)	>0.05	1.01 (0.82, 1.25)	>0.05	1.06 (0.90, 1.24)	>0.05	0.71 (0.46, 1.07)	>0.05	1.06 (0.63, 1.80)	>0.05	0.105
Last BMI < 35 kg/m^2 (*n* = 3,194)											
5	0.86 (0.66, 1.13)	>0.05									
6			1.76 (1.34, 2.31)	<0.0001	1.32 (1.07, 1.64)	0.0097					
7	0.83 (0.64, 1.09)	>0.05	1.78 (1.36, 2.33)	<0.0001	1.33 (1.08, 1.65)	0.0083					
8	0.71 (0.41, 1.21)	>0.05	1.72 (1.28, 2.30)	0.0003	1.30 (1.04, 1.63)	0.0227	1.22 (0.64, 2.33)	>0.05	1.29 (0.61, 2.74)	>0.05	0.755
Last BMI ≥35 kg/m^2 (*n* = 2,513)											
9	0.73 (0.56, 0.95)	0.0183									
10			0.58 (0.43, 0.78)	0.0003	0.78 (0.64, 0.96)	0.0173					
11	0.73 (0.56, 0.95)	0.0178	0.58 (0.43, 0.80)	0.0009	0.78 (0.64, 0.96)	0.0173					
12	1.03 (0.68, 1.57)	>0.05	0.58 (0.43, 0.80)	0.0009	0.87 (0.70, 1.08)	>0.05	0.49 (0.28, 0.86)	0.0133	0.93 (0.42, 2.06)	>0.05	0.024

[a]Also controlled for relatedness, sex, age at last biennial examination, and PC1–PC5. [b]Overall *P* value for interaction on 2 df.

the joint maximum approximately at BMI ≥35 kg/m^2. We performed a second set of regressions with *SLC16A11* genotypes alone in the ≥BMI strata, which maximizes protection in stratum BMI ≥35 kg/m^2 but exhibits a range of protections in BMI strata 31–41 kg/m^2 (Supplementary Table 3 and Fig. 2). For the risk analyses, we performed regressions in the <BMI strata where the risk OR for *SLC16A11*22* maximizes in the BMI <26 kg/m^2 stratum (OR 2.95; *P* = 0.0019) but has statistical significance in the BMI 26–39 kg/m^2 strata (Supplementary Table 4 and Fig. 3).

Protective Allele *DRB1*16:02* Strongly Binds *SLC16A11* 15-mer Peptides at D127G

To explore potential mechanisms that may explain the observed interactions, we analyzed the predicted binding of *HLA* class II alleles with *SLC16A11*. For *HLA* class II, the seven polymorphic *DRB1* alleles in Table 1 were applied to the NetMHCIIpan EL 4.0 algorithm with the 15-mers of the ancient human *SLC16A11* protein, which yielded 3,192 records (Supplementary Table 5). The program provides two numbers: the probability of a strong binding score and rank% for each 15-mer peptide. A total of

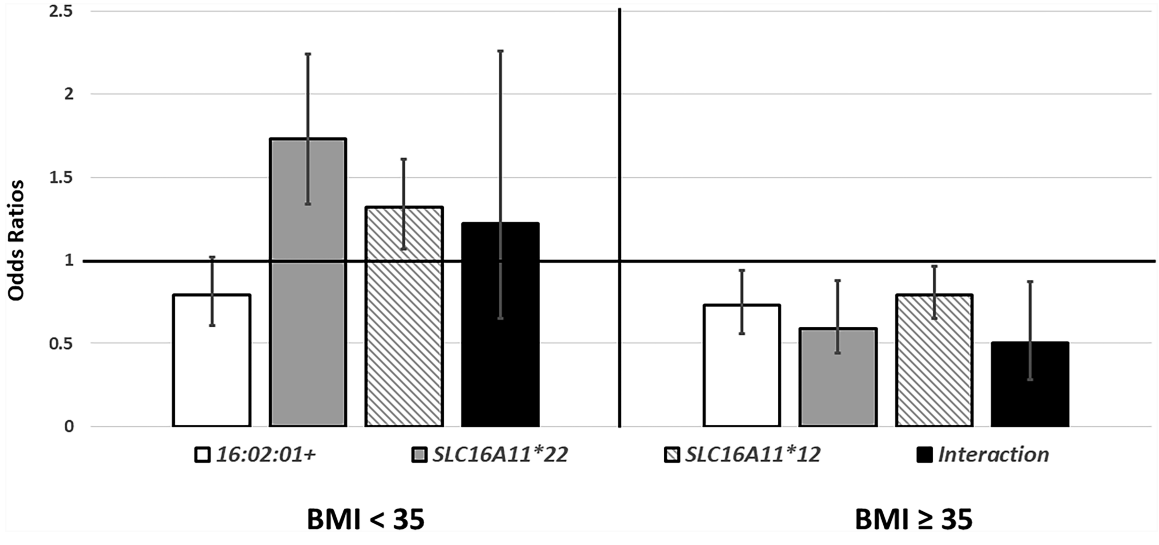

Figure 1—Logistic regression of genotypes on T2D in Southwestern American Indians. Logistic regressions with T2D as the dependent variable included both *HLA-DRB1*16:02* and the *SLC16A11* polymorphisms controlled for age at last biennial examination, sex, first-degree relationship, and PC1–PC5, stratified by BMI (Table 2). When last BMI <35 kg/m^2, *HLA-DRB1*16:02:01* is protective, and the ancient human haplotype *T-C-G-T-T* is risk for T2D, with no significant interaction. However, when last BMI ≥35 kg/m^2, *HLA-DRB1*16:02:01* remains a protective allele while *SLC16A11 T-C-G-T-T* becomes protective and the two loci have significant interaction, epistasis. This switch in risk/protection in the heavier stratum demonstrates that the function of *SLC16A11* in the physiology of the body makes it a separate source of risk and protection, while in the heavier stratum, the two loci work together to amplify protection for T2D.

Table 3—Sensitivity analysis of *SLC16A11-HLA-DRB1*16:02:01* interaction ORs across the ≥BMI strata in a Southwest American Indian sample for the fully parametrized logistic regression

			Genotype *SLC16A11*22*			Interaction: *SLC16A11*12* and **16:02:01*		
≥BMI	Mean BMI	*n*	β	OR	*P*	β	OR	*P*
25	36.22	5,116	−0.02294	0.98	>0.05	−0.40134	0.67	>0.05
26	36.61	4,939	−0.04939	0.95	>0.05	−0.31729	0.73	>0.05
27	37.06	4,727	−0.07589	0.93	>0.05	−0.35868	0.70	>0.05
28	37.63	4,463	−0.09437	0.91	>0.05	−0.38616	0.68	>0.05
29	38.15	4,222	−0.17126	0.84	>0.05	−0.38228	0.68	>0.05
30	38.75	3,952	−0.18612	0.83	>0.05	−0.49026	0.61	0.0313
31	39.46	3,638	−0.28460	0.75	0.0330	−0.54627	0.58	0.0219
32	40.12	3,363	−0.38201	0.68	0.0058	−0.59205	0.55	0.0151
33	40.82	3,078	−0.43205	0.65	0.0032	−0.67306	0.51	0.0074
34	41.55	2,800	−0.45732	0.63	0.0028	−0.60496	0.55	0.0205
35	42.35	2,513	−0.53720	0.58	0.0009	−0.71882	0.49	0.0100
36	43.32	2,203	−0.44240	0.64	0.0104	−0.68365	0.50	0.0238
37	44.22	1,947	−0.45043	0.64	0.0131	−0.71460	0.49	0.0300
38	45.05	1,735	−0.41760	0.66	0.0300	−0.85719	0.42	0.0140
39	45.90	1,536	−0.44075	0.64	0.0305	−0.92799	0.40	0.0114
40	46.80	1,345	−0.39334	0.67	>0.05	−0.82461	0.44	0.0361
41	47.65	1,187	−0.30461	0.74	>0.05	−0.92529	0.40	0.0260
42	48.61	1,028	−0.35996	0.70	>0.05	−1.08041	0.34	>0.05
43	49.47	902	−0.37587	0.69	>0.05	−0.87119	0.42	>0.05
44	50.57	762	−0.30607	0.74	>0.05	−0.74239	0.48	>0.05
45	51.47	664	−0.36346	0.70	>0.05	−0.81343	0.44	>0.05

89.5% of the binding scores were in the interval 0.0–0.1, whereas only 1.4% of the 15-mers had a strong binding >0.4 (Supplementary Table 6). To assess the predicted binding in a larger sample of proteins, we directed the seven polymorphic *DRB1* alleles against 113 proteins, which yielded 624,358 records (Supplementary Tables 5 and 6). Predicted strong binding scores >0.3 represented only 2.4% of the peptides. The rank% of binding is based on a library of >5 million random 15-mers in the SwissProt database. The top 2% of binding values are considered as strong binding, 2.0–10% as weak binding, and >10% as what we classify as null binding.

The presence of four missense mutations (V113I, D127G, G340S, and P443T) in the ancient human haplotype of *SLC16A11* allows the comparison of the wild-type amino acid and its substitution. Table 4 presents the 15 15-mers at D127G for predicted binding results with *DRB1*16:02* (the prediction algorithm allows two levels of allele precision) and contrasts the wild-type *SLC16A11* protein with aspartic acid (D) with the ancient human mutation with a substituted glycine (G). For the core peptide FSAFASDLL the *DRB1*16:02* exhibits a maximum binding of 0.4717 for peptide 6 that includes the D amino acid, which is within the top 2% of values in Supplementary Table 6. Compared with the ancient human mutation, the effect is amplified. With the G amino acid in the core peptide FSAFASGLL, the maximum binding increases to 0.5650, with a range of ∼0.52–0.57, which is now within the top 1% of scores (Supplementary Tables 5 and 6). In addition, *DRB1*16:02* recognizes a consistent core of amino acids in 15-mers 3–8, FSAFASDLL in the wild type and FSAFASGLL in the ancient human haplotype. A search with the UniProt peptide finder for *Homo sapiens* found that the wild-type peptide is unique in the human genome for *SLC16A11*. A search for the G mutation returned no match, meaning that the minor allele and its amino acid substitution are not represented in the protein database. Patterns of peptide binding for *DRB1*16:02* and the three remaining missense mutations are found in Supplementary Tables 7–9. Supplementary Table 10 presents the 442 15-mer peptides for the ancient human protein when tested against *DRB1*16:02* in amino acid order, excluding the leader sequence.

DISCUSSION

There is increasing recognition of the complexity of the human immune system (33,34). However, a fundamental

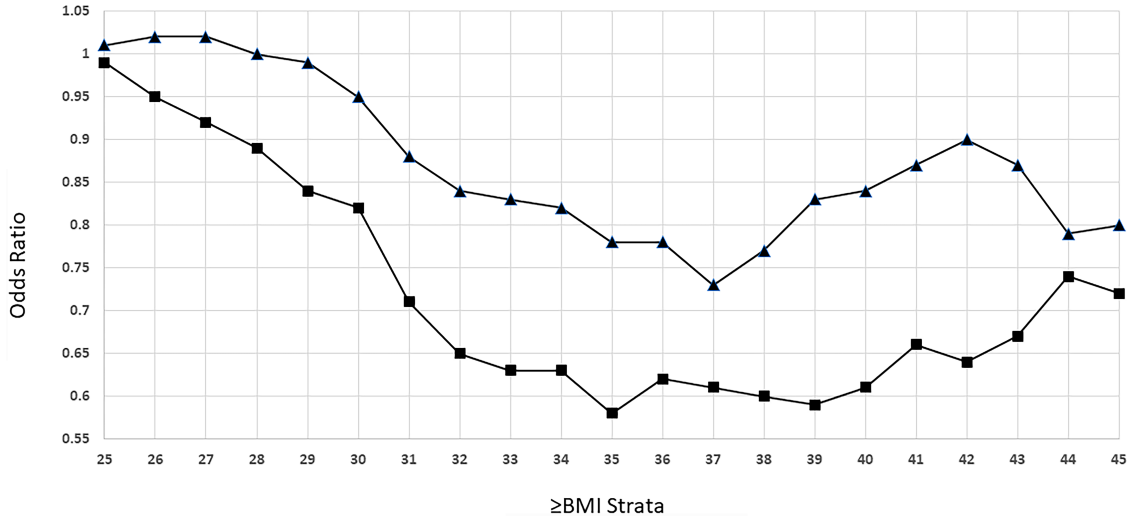

Figure 2—Odds ratios for logistic regression of *SLC16A11* on T2D within cumulative ≥BMI strata for Southwestern American Indian sample (protection allele = "2"). Twenty-one sample strata were created from the total number of 5,707 by partitioning the data into those groups greater than or equal to BMIs ranging from 25 to 45 kg/m^2, e.g., for a BMI of 25 kg/m^2, all members of the sample who have a BMI ≥25 kg/m^2, then a sample with BMI ≥26 kg/m^2, and so forth (see Supplementary Table 3). Within each stratum, a logistic regression was performed with no interaction, controlled for age, sex, relationship, and PC1–PC5, and the ORs for the *SLC16A11*12* (triangles) and *SLC16A11*22* (squares) genotypes were captured and plotted. The most protective OR was for the BMI ≥35 kg/m^2 stratum.

reaction among the cascades is the presentation of peptide by the class I and class II *HLA* heterodimers. In the past, many studies of the association of *HLA* alleles alone have shown a statistical relation, risk or protective, with diseases. We expanded this model to identify and include a potential peptide and its protein source and to search for an epistatic relationship between them.

Since the 1977 Oxford HLA Workshop and Conference, when the serological *HLA-DR* locus was first elaborated and applied to disease association studies, the antigen *HLA-DR2* has been found to be protective for T1D (35). In 1985, Tiwari and Terasaki (36) compiled data on *HLA* disease associations and reported results for *HLA-DR2* from 14 studies from seven countries with a total of 1,268 patients with T1D and 2,768 control individuals. The mean relative risk was 0.18. Antigen *HLA-DR2,* in later serology, split into *HLA-DR15*, *DR16*. In 2008, as part of the Type 1 Diabetes Genetics Consortium, Erlich et al. (37) identified in a sample of individuals primarily of European, Caucasian heritage the protective molecular haplotype from *HLA-DR15* as *DRB1*15:01-DQA1*01:02-DQB1*06:02*, with an OR of 0.03 (95% CI 0.01, 0.07; $P = 2 \times 10^{-29}$). People with this haplotype did not develop the disease.

In 2011, we reported a protective effect in this population for the *A* allele at *rs9268852* that tagged *HLA-DRB1*02*, which we had earlier typed as *DRB1*16:02* with an OR of 0.723 ($P = 0.002$) (1,5). We have now extended this finding of protection for T2D and identify the associated haplotype as *DRA*01:01:01-DRB1*16:02:01-DQA1*05:03:01-DQB1*03:01:01* (6). In an earlier report, we also showed the interaction between *SLC16A11* and BMI such that in people with BMI ≥35 kg/m^2, the ancient human haplotype at *SLC16A11* was protective for T2D in Southwest American Indians (10). The natural question is, what is the relationship between the two loci? This is a three-part question: Is there epistasis in which the loci interact; if yes, can we identify a 15-mer and core peptide in the protective allele with a strong binding pattern for *DRB1*16:02*? Furthermore, with the *HLA-DR2* protective effect being a property common to T1D and T2D, can we identify a potential common immunological component and link among *HLA-DRB1*, the two diseases, and *SLC16A11*?

Our sensitivity analysis shows that there is strong epistasis between *HLA-DRB1* and *SLC16A11* in the fully parametrized logistic regression with a power of 0.4801 and a bootstrap estimate OR of 0.43. It is not, however, only a property of the BMI ≥35 kg/m^2 stratum, but the interaction term for *DRB1*16:02:01+* and *SLC16A11*12* has statistical significance between BMI 30 and 41 kg/m^2, with mean BMI values of 38.75–47.65 kg/m^2 (Table 3).

As for the second question, our hypothesis is that this epistatic effect comes from the binding of 15-mer peptides of *SLC16A11* by the *HLA-DRB1*16:02:01* heterodimer. Evidence comes from the monocarboxylate transporter mutation being both risk and protection for T2D in this population, while the *DRB1* allele has consistent protection and strong predicted binding to both the wild-type and ancient human haplotype. In Table 4, the core peptide FSAFASDLL at mutation site D127G has a maximum predicted binding probability of 0.4717 for the basic aspartic amino acid, while for the nonpolar glycine mutation and core peptide FSAFASGLL, the predicted binding score increases to 0.5650. This is the seventh largest

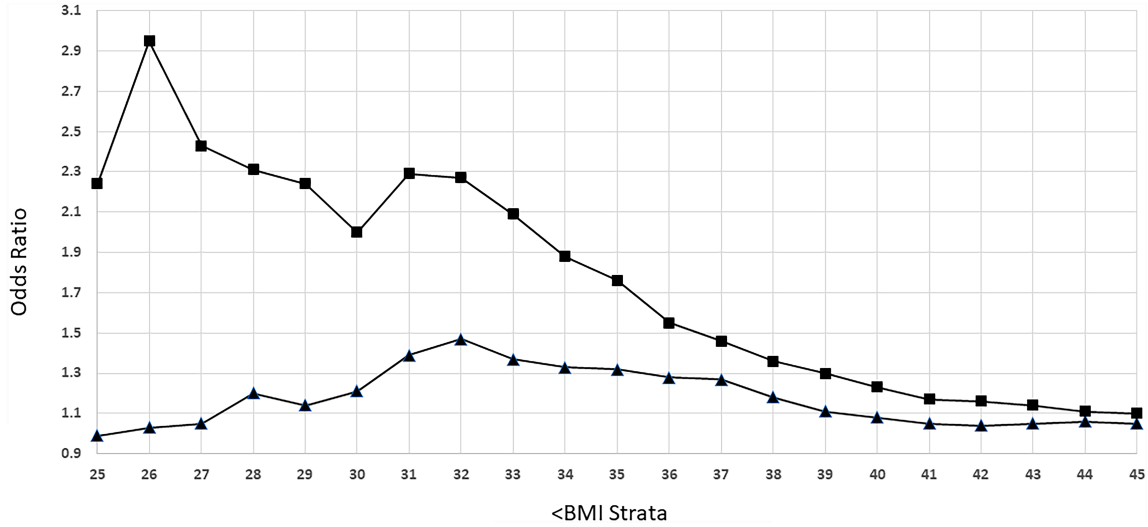

Figure 3—Odds ratios for logistic regression of *SLC16A11* on T2D within cumulative ≥BMI strata for Southwestern American Indian sample (risk allele = "2"). Twenty-one sample strata were created from the total number of 5,707 by partitioning the data into those groups less than the BMIs ranging from 25 to 45 kg/m^2, e.g., for a BMI of 25 kg/m^2, all members of the sample who have BMI <25 kg/m^2, then a sample <26 kg/m^2, and so forth (see Supplementary Table 4). Within each stratum, a logistic regression was performed with no interaction, controlled for age, sex, relationship, and PC1–PC5, and the ORs for the *SLC16A11*12* (triangles) and *SLC16A11*22* (squares) genotypes were plotted. The highest risk OR of 2.95 was for *SLC16A11*22* in the BMI <26 kg/m^2 stratum. Only in the BMI 31–37 kg/m^2 strata were the ORs for both genotypes significantly different from 1.0 in the model, with mean BMI values of 26.6–29.9 kg/m^2.

binding probability over the entire set of 15-mers from *SLC16A11* (Supplementary Table 10). *HLA-DRB1*16:02:01* demonstrates protection for the all-people stratum in Table 2, while the increase in predicted binding with the glycine mutation might contribute to its enhancement of protection for the BMI ≥35 kg/m^2 stratum.

What we have, then, is the observation of epistasis for the two molecules and the observation of strong predicted binding for the *DRB1* protective allele; what we lack is the detailed mechanics of risk and protection observed here. They are almost certainly independent, with *SLC16A11* having both risk and protection mechanisms deriving from its action as a monocarboxylate transporter. Rusu et al. (14) suggested two mechanisms by which this proton-coupled monocarboxylate transporter's function could be changed by mutations: *1*) The expression of *SLC16A11* could be reduced in the liver, and *2*) its interaction with basigin (CD147, BSG), a multifunctional protein with two immunoglobulin-like domains that plays a role in the orientation of monocarboxylic acid transporters, could disrupt the expression and/or orientation of *SLC16A11* on the cell surface. Further study is needed to make more precise the effects of the mutations. It likely plays a passive role in epistasis with *HLA-DRB1*16:02:01* in that it supplies a 15-mer peptide to the molecule when the protein is being turned over in the cell. What immunological cascade is initiated by the binding of peptide and heterodimer that leads to protection is also unknown. When the protective *HLA* allele and mutant transporter both are present, then the 15-mer will be a self-peptide, and its strong recognition might play a role in maintaining self-recognition and the prevention of autoimmune antibodies that might otherwise contribute to T2D.

With the historic role of protection for *HLA-DRB1*15* and T1D, and now *HLA-DRB1*16* for T2D, is there a common pattern in their binding for *SLC16A11*? Table 5 presents the probability of a strong binding score for *HLA-DRB1*15:01*, the most protective of the alleles in European-derived, Caucasian populations among whom the ancient human mutation haplotype at *SLC16A11* is mostly absent. For 15-mers 4–6, the allele has binding scores of 0.2783–0.3526 and recognizes the same core peptide, FSAFASDLL, as does *DRB1*16:02:01* for the wild-type protein. The ranks% of binding are all in the top 5% of binding scores compared with the Swiss-Prot database. In addition, *DRB1*15:01* has very strong binding for peptides 13 and 14, with scores of 0.3035 and 0.4140, that recognize the core peptide LHLYLGLGL and for which the ranks are also in the top 5% of scores; the aspartic acid at position D127 lies outside the vector of amino acids of this core.

There is virtually no T1D in this Southwest American Indian population, which is consistent with the absence of the risk alleles for the disease in full indigenous American heritage people: *HLA-DRB1*04:01*, **04:02*, **04:05*, and **03:01* (37) (Supplementary Table 1). Furthermore, in contrast with *HLA-DRB1*15:01* and **16:02*, the highest probability of binding for these T1D risk alleles at the D127G missense site, >60 15-mers, is only 0.1139 with a rank% of 11. However, outside of the four missense sites and in the parts of the molecule common to the wild-type and mutant forms of *SLC16A11*, there is strong predicted binding for the T1D risk alleles (Supplementary Table 11).

With respect to locus *SLC16A11*, the private ancient human haplotype and risk allele has among the highest

Table 4—Predicted binding of *HLA-DRB1*16:02* with wild-type and early human mutation from *SLC16A11* at D127G polymorphism

	Wild type					Ancient human haplotype				
	Peptide	Core peptide	Score	Rank%		Peptide	Core peptide	Score	Rank%	
1	VLASLGFVFSAFASD	LGFVFSAFA	0.0129	54.0	NL	ILASLGFVFSAFASG	LGFVFSAFA	0.0088	62.0	NL
2	LASLGFVFSAFASDL	FVFSAFASD	0.0129	54.0	NL	LASLGFVFSAFASGL	FVFSAFASG	0.0079	65.0	NL
3	ASLGFVFSAFASDLL	FSAFASDLL	0.1062	16.0	NL	ASLGFVFSAFASGLL	FSAFASGLL	0.1188	15.0	NL
4	SLGFVFSAFASDLLH	FSAFASDLL	0.4303	2.7	WB	SLGFVFSAFASGLLH	FSAFASGLL	0.5183	1.7	SB
5	LGFVFSAFASDLLHL	FSAFASDLL	0.4503	2.4	WB	LGFVFSAFASGLLHL	FSAFASGLL	0.5438	1.5	SB
6	GFVFSAFASDLLHLY	FSAFASDLL	0.4717	2.2	WB	GFVFSAFASGLLHLY	FSAFASGLL	0.5650	1.3	SB
7	FVFSAFASDLLHLYL	FSAFASDLL	0.2054	8.6	WB	FVFSAFASGLLHLYL	FSAFASGLL	0.2440	7.0	WB
8	VFSAFASDLLHLYLG	FSAFASDLL	0.0655	23.0	NL	VFSAFASGLLHLYLG	FSAFASGLL	0.0697	22.0	NL
9	FSAFASDLLHLYLGL	FASDLLHLY	0.0041	78.0	NL	FSAFASGLLHLYLGL	FASGLLHLY	0.0014	94.0	NL
10	SAFASDLLHLYLGLG	FASDLLHLY	0.0019	90.0	NL	SAFASGLLHLYLGLG	FASGLLHLY	0.0010	97.0	NL
11	AFASDLLHLYLGLGL	LHLYLGLGL	0.0215	43.0	NL	AFASGLLHLYLGLGL	LHLYLGLGL	0.0214	43.0	NL
12	FASDLLHLYLGLGLL	LHLYLGLGL	0.0404	31.0	NL	FASGLLHLYLGLGLL	LHLYLGLGL	0.0380	32.0	NL
13	ASDLLHLYLGLGLLA	LHLYLGLGL	0.0783	21.0	NL	ASGLLHLYLGLGLLA	LHLYLGLGL	0.0734	22.0	NL
14	SDLLHLYLGLGLLAG	LHLYLGLGL	0.1204	15.0	NL	SGLLHLYLGLGLLAG	LHLYLGLGL	0.1150	15.0	NL
15	DLLHLYLGLGLLAGF	LHLYLGLGL	0.0465	29.0	NL	GLLHLYLGLGLLAGF	LHLYLGLGL	0.0433	30.0	NL

NL, null binding; SB, strong binding; WB, weak binding.

reported frequencies in this Southwest American Indian sample, with 0.397 in all people and 0.421 in full indigenous American heritage people. Of more importance than the allele frequency is the proportion of people with at least one early human mutant haplotype of 64%. This community has one of the highest frequencies of T2D and a distribution of *SLC16A11* risk haplotypes that blankets two thirds of the population. The SIGMA Type 2 Diabetes Consortium article (8) on the association of *SLC16A11* with T2D in Mexico reported a risk haplotype frequency in the entire SIGMA sample of 30%, while within people who were >95% Native American ancestry (n = 290), 48%, although it is not clear whether these are allele or combined genotype frequencies. The 1000 Genomes Project reported a frequency of 0% for the African sample (n = 185), <2% for the European sample (n = 379), 12% for the East Asian

Table 5—Predicted binding of wild-type *SLC16A11* 15-mer peptides by *HLA-DRB1*15:01* at D127G missense site

	Start	End	Peptide	Core peptide	Probability	Rank%
1	113	127	VLASLGFVFSAFASD	GFVFSAFAS	0.0080	44.0
2	114	128	LASLGFVFSAFASDL	GFVFSAFAS	0.0060	50.0
3	115	129	ASLGFVFSAFASDLL	FSAFASDLL	0.0293	24.0
4	116	130	SLGFVFSAFASDLLH	FSAFASDLL	0.2783	4.5
5	117	131	LGFVFSAFASDLLHL	FSAFASDLL	0.3157	3.9
6	118	132	GFVFSAFASDLLHLY	FSAFASDLL	0.3526	3.4
7	119	133	FVFSAFASDLLHLYL	FSAFASDLL	0.1095	11.0
8	120	134	VFSAFASDLLHLYLG	FSAFASDLL	0.0284	24.0
9	121	135	FSAFASDLLHLYLGL	FSAFASDLL	0.0019	74.0
10	122	136	SAFASDLLHLYLGLG	LLHLYLGLG	0.0007	90.0
11	123	137	AFASDLLHLYLGLGL	LHLYLGLGL	0.0626	15.0
12	124	138	FASDLLHLYLGLGLL	LHLYLGLGL	0.1700	7.4
13	125	139	ASDLLHLYLGLGLLA	LHLYLGLGL	0.3035	4.1
14	126	140	SDLLHLYLGLGLLAG	LHLYLGLGL	0.4140	2.7
15	127	141	DLLHLYLGLGLLAGF	LHLYLGLGL	0.1776	7.1

sample (n = 286), and 28% for a small sample of Mexican Americans from Los Angeles (n = 66). The frequency of the ancient human risk haplotype in Mexican Americans and admixed Mexicans and Latin Americans is likely derived by genetic admixture from indigenous American populations.

Our observation that enhanced protection from T2D associated with *HLA DRB1*16:02:01* and *SLC16A11* in the most obese stratum may seem counterintuitive given that obesity increases the risk of T2D. However, the current analyses are based on cross-sectional associations and do not reflect longitudinal risk. Our previous longitudinal analyses in this population suggested that the ancient *SLC16A11* haplotype associates with increased risk in the leanest stratum and that the lower prevalence of diabetes in the heavier stratum was partially the result of increased weight loss that occurs after diabetes onset (10). However, in the present analysis, the heavier stratum (BMI $\geq$35 kg/m^2) has a higher percentage of T2D than the leaner one (39.6 vs. 33.8%) (Table 1). The risk-to-protective switch is a key piece of evidence of the mechanics of *SLC16A11*'s overall function when further epidemiological studies and in vitro or animal experiments address it (38–40).

In summary, the demonstration of epistasis between *HLA-DRB1* and *SLC16A11* in the amplification of protection in the heavier stratum suggests that the two loci together have a complicated interplay of functions. The ancient human haplotype is a strong risk allele for T2D in the leaner stratum of people, where the combined risk genotypes' frequency closely mimics the prevalence of the disease in the population. We suggest that the likely missense mutation in this risk is D127G because of the amplification of *HLA-DRB1*16:02* binding in protection for the disease, while the independent mechanism of risk for *SLC16A11* has yet to be completely elaborated. Absence of T1D in Southwest American Indians, and in full heritage indigenous people throughout the New World, is likely because of the absence of the *HLA-DRB1* risk alleles. However, the strong protection of *HLA-DRB1* 15:01*, and its shared binding pattern with *DRB1*16:02* and a shared core peptide, suggests that the epistasis might not be exclusive to T2D in protection from diabetes.

Acknowledgments. The authors thank the Southwest American Indian community in this study for their cooperation and participation and thank the staff of the Diabetes Epidemiology and Clinical Research Section, National Institute of Diabetes and Digestive and Kidney Diseases, for conducting the examinations.

Funding. This research was supported in part by the Intramural Research Program of the National Institutes of Health National Institute of Diabetes and Digestive and Kidney Diseases.

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. R.C.W. designed the study, performed the statistical analyses, and wrote the manuscript. R.L.H. and W.C.K. performed statistical analyses and reviewed and edited the manuscript. B.P. and K.K. contributed to the design and reviewed and edited the manuscript. C.B. and L.J.B. reviewed and edited the manuscript. R.C.W. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Williams RC, Muller YL, Hanson RL, et al. HLA-DRB1 reduces the risk of type 2 diabetes mellitus by increased insulin secretion. Diabetologia 2011;54:1684–1692

2. Brooks-Worrell B, Hampe CS, Hattery EG, et al.; GRADE Beta-Cell Ancillary Study Network. Islet autoimmunity is highly prevalent and associated with diminished β-cell function in patients with type 2 diabetes in the GRADE study. Diabetes 2022;71:1261–1271

3. Knowler WC, Pettitt DJ, Saad MF, Bennett PH. Diabetes mellitus in the Pima Indians: incidence, risk factors and pathogenesis. Diabetes Metab Rev 1990;6:1–27

4. Knowler WC, Bennett PH, Hamman RF, Miller M. Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minnesota. Am J Epidemiol 1978;108:497–505

5. Williams R, Chen YF, Endres R, et al. Molecular variation at the HLA-A, B, C, DRB1, DQA1, and DQB1 loci in full heritage American Indians in Arizona: private haplotypes and their evolution. Tissue Antigens 2009;74:520–533

6. Williams RC, Knowler WC, Shuldiner AR, et al.; Regeneron Genetics Center. Next generation sequencing and the classical HLA loci in full heritage Pima Indians of Arizona: defining the core HLA variation for North American Paleo-Indians. Hum Immunol 2019;80:955–965

7. Williams RC, Koroglu C, Knowler WC, et al. Next generation sequencing for *HLA* loci in full heritage Pima Indians of Arizona, part II: *HLA-A, -B*, and *-C* with selected non-classical loci at 4-field resolution from whole genome sequences. Hum Immunol 2021;82:385–403

8. SIGMA Type 2 Diabetes Consortium; Williams AL, Jacobs SBR, Moreno-Macías H, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature 2014;506:97–101

9. Prüfer K, Racimo F, Patterson N, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 2014;505:43–49

10. Traurig M, Hanson RL, Marinelarena A, et al. Analysis of SLC16A11 variants in 12,811 American Indians: genotype-obesity interaction for Type 2 diabetes and an association with RNASEK expression. Diabetes 2016;65:510–519

11. Hidalgo BA, Sofer T, Qi Q, et al. Associations between *SLC16A11* variants and diabetes in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). Sci Rep 2019;9:843–849

12. Huerta-Chagoya A, Vázquez-Cárdenas P, Moreno-Macías H, et al. Genetic determinants for gestational diabetes mellitus and related metabolic traits in Mexican women. PLoS One 2015;10:e0126408

13. Lara-Riegos JC, Ortiz-López MG, Peña-Espinoza BI, et al. Diabetes susceptibility in Mayas: evidence for the involvement of polymorphisms in HHEX, HNF4α, KCNJ11, PPARγ, CDKN2A/2B, SLC30A8, CDC123/CAMK1D, TCF7L2, ABCA1 and SLC16A11 genes. Gene 2015;565:68–75

14. Rusu V, Hoch E, Mercader JM, et al.; MEDIA Consortium; SIGMA T2D Consortium. Type 2 diabetes variants disrupt function of SLC16A11 through two distinct mechanisms. Cell 2017;170:199–212.e20

15. Miranda-Lora AL, Cruz M, Molina-Díaz M, Gutiérrez J, Flores-Huerta S, Klünder-Klünder M. Associations of common variants in the SLC16A11, TCF7L2, and ABCA1 genes with pediatric-onset type 2 diabetes and related glycemic traits in families: a case-control and case-parent trio study. Pediatr Diabetes 2017;18:824–831

16. Almeda-Valdes P, Gómez Velasco DV, Arellano Campos O, et al. The SLC16A11 risk haplotype is associated with decreased insulin action, higher transaminases and large-size adipocytes. Eur J Endocrinol 2019;180:99–107

17. Tan YX, Hu SM, You YP, Yang GL, Wang W. Replication of previous genome-wide association studies of *HKDC1, BACE2, SLC16A11* and *TMEM163* SNPs in a gestational diabetes mellitus case-control sample from Han Chinese population. Diabetes Metab Syndr Obes 2019;12:983–989

18. Bello-Chavolla OY, Bahena-López JP, Vargas-Vázquez A, et al.; Metabolic Syndrome Study Group; Group of Study CAlPaDi. Clinical characterization of

data-driven diabetes subgroups in Mexicans using a reproducible machine learning approach. BMJ Open Diabetes Res Care 2020;8:e001550
19. Kimura Y, Higuchi I, Kobayashi M, et al. The association between SLC16A11 haplotype and lipid metabolism in Japanese patients with type 2 diabetes. Drug Metab Pharmacokinet 2021;37:100376
20. Srinivasan S, Chen L, Todd J, et al.; ProDiGY Consortium. The first genome-wide association study for type 2 diabetes in youth: the Progress in Diabetes Genetics in Youth (ProDiGY) Consortium. Diabetes 2021;70:996–1005
21. Mardones L, Petermann-Rocha F, Martinez-Sanguinetti MA, et al.; ELHOC Group (Epidemiology of Lifestyle and Health Outcomes in Chile). Genetic variants in the *SLC16A11* gene are associated with increased BMI and insulin levels in nondiabetic Chilean population. Arch Endocrinol Metab 2021;65:305–314
22. Neel JV. "Private" genetic variants and the frequency of mutation among South American Indians. Proc Natl Acad Sci U S A 1973;70:3311–3315
23. Shen GQ, Abdullah KG, Wang QK. The TaqMan method for SNP genotyping. Methods Mol Biol 2009;578:293–306
24. Schleinitz D, Distefano JK, Kovacs P. Targeted SNP genotyping using the TaqMan assay. Methods Mol Biol 2011;700:77–87
25. Hui L, DelMonte T, Ranade K. Genotyping using the TaqMan assay. Curr Protoc Hum Genet 2008;Chapter 2:10
26. Weir B. *Genetic Data Analysis.* Sunderland, MA, Sinaurer Associates, 1990
27. Long JC, Williams RC, Urbanek M. An E-M algorithm and testing strategy for multiple-locus haplotypes. Am J Hum Genet 1995;56:799–810
28. *SAS/STAT 14.2 User's Guide.* Cary, NC, SAS Institute Inc., 2016
29. Piaggi P, Masindova I, Muller YL, et al.; SIGMA Type 2 Diabetes Consortium. A genome-wide association study using a custom genotyping array identifies variants in *GPR158* associated with reduced energy expenditure in American Indians. Diabetes 2017;66:2284–2295
30. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res 2020;48:W449–W454
31. Jensen KK, Andreatta M, Marcatili P, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology 2018;154:394–406
32. Andreatta M, Karosiene E, Rasmussen M, Stryhn A, Buus S, Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. Immunogenetics 2015;67:641–650
33. Vinuesa CG, Linterman MA, Yu D, MacLennan ICM. Follicular helper T cells. Annu Rev Immunol 2016;34:335–368
34. Deng Q, Luo Y, Chang C, Wu H, Ding Y, Xiao R. The emerging epigenetic role of CD8+T cells in Autoimmune Diseases: a systematic review. Front Immunol 2019;10:856
35. Batchelor JR, Morris PJ. HLA and disease. In *Histocompatibility Testing 1977.* Bodmer WF, Batchelor JR, Bodmer JG, Festenstein H, Morris PJ, Eds. Copenhagen, Munksgaard, 1978, pp. 205–258
36. Tiwari JL, Terasaki PI. *HLA and Disease Associations.* New York, Springer-Verlag, 1985
37. Erlich H, Valdes AM, Noble J, et al.; Type 1 Diabetes Genetics Consortium. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. Diabetes 2008;57:1084–1092
38. Zhao Y, Feng Z, Zhang Y, et al. Gain-of-function mutations of SLC16A11 contribute to the pathogenesis of type 2 diabetes. Cell Rep 2019;26:884–892.e4
39. Hoch E, Florez JC, Lander ES, Jacobs SBR. Gain-of-function claims for type-2-diabetes-associated coding variants in SLC16A11 are not supported by the experimental data. Cell Rep 2019;29:778–780
40. Zhao Y, Feng Z, Ding Q. Type 2 diabetes variants in the *SLC16A11* coding region are not loss-of-function mutations. Cell Rep 2019;29:781–784