An Exploration of HLA Variants associated with Type-1 Diabetes Progression

James Najera

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington 2021

Committee:

Alison Fohner

Lue Ping Zhao

Program Authorized to Offer Degree:

Public Health Genetics

©Copyright 2021

James Najera

University of Washington

Abstract

An Exploration of HLA Variants associated with Type-1 Diabetes Progression

James Najera

Chair of the Supervisory Committee:

Alison Fohner

Department of Epidemiology

The Diabetes Prevention Trial-1(DPT-1) assembled genome sequences on 711 participants with the goal of identifying HLA variants associated with Type 1 Diabetes(T1D) progression. We assessed HLA alleles and haplotypes across 9 distinct HLA genes for correlation between time of patient seroconversion to clinically diagnosed T1D. Using a Cox proportional hazards regression model for each approach, we performed an allelic association analysis across 302 HLA alleles and ran a haplotype association analysis across all possible gene pairs within the study. The results identified one allele, DQB1*03:02:01, and the three following haplotypes as significantly associated with T1D progression :DQA1*03:01:01-DQB1*03:02:01, DQA1*03:01:01-B*18:01:01, and DQA1*03:01:01-B*35:01:01. Our results support prior significant findings within T1D literature regarding the association of the identified HLA variants within our study and the risk for developing T1D.

Acknowledgements

I want to express my gratitude to my thesis chair, Alison Fohner, and my thesis advisory committee member, Lue Ping Zhao. Dr. Zhao spent a great deal of time guiding my thesis progression, working with me on coding the haplotypes, and helping me navigate the association analysis. His investment in my work and education went above and beyond. Alie has offered me invaluable advice and guidance for my thesis and academic planning.

I am beyond grateful for the support I have received from my friends, family, and colleagues during my time at the University of Washington, particularly as I have worked toward the completion of my thesis.

I. INTRODUCTION

Type-1 diabetes(T1D) is becoming an increasingly more common autoimmune, chronic disease with a global prevalence of 9.5% and is estimated to make up approximately 10% of all people with diabetes.[1,2] Among those cases, children and young adults are particularly vulnerable to accelerated T1D onset. Specifically, those within the age brackets of 5-14 years old have the highest incidence rates and tend to have the fastest, most severe onset of T1D.[3] With rising annual incidence rate as high as 4%, as indicated in a 2015 analysis of French children T1D progression, challenges for managing the disease among affected households and the financial and resource burden on healthcare systems is a growing concern.[4,5]

T1D is an organ-specific autoimmune disease that leads to the destruction of beta cells in the insulin-producing pancreatic islets. Insulin induces the uptake of glucose into muscle and adipose cells, stimulates the liver to store glycogen and fatty acid, and is responsible for regulation of potassium absorption in cells.[6] Many long-term complications arise from T1D, including severe cardiovascular disease, nephropathy, neuropathy, and retinopathy.[7] Diabetic patients require life-long insulin replacement therapy in order to keep blood sugar levels in an acceptable, healthy range.

Environmental and lifestyle risk factors play an important role in the etiology of T1D. Multiple retrospective and cohort studies have found significant association between a variety of exposures, such as infection, diet, and toxins that affect children in utero, perinatally, or during early childhood, and islet autoimmunity.[8] However, our understanding of how environmental exposures trigger the initial immune activation in T1D is not entirely understood.

Whether triggered by viral infection or other environmental exposures, the initial immune activation leads to an immune response that results in the development of autoantibodies that

either target pancreatic beta cells directly or interfere with the cascade pathway that leads to the production of insulin.[9] The five most popular and well-researched autoantibodies are the following: Islet cell antibodies (ICA), Insulin autoantibodies (IAA), glutamic acid decarboxylase autoantibodies (GADA), islet-antigen 2(IA-2), and zinc transporter 8 autoantibodies (ZNT8).[10] ICA is an autoantibody that targets the beta cells directly, while the other four target specific proteins found in the islet and are considered part of the insulin production pathway.

Once 2 or more autoantibodies have been identified, subjects are considered to have undergone islet-targeted autoantibody seroconversion. The International Society for Pediatric and Adolescent Diabetes has characterized type 1 diabetes progression after seroconversion into four distinct stages.[11] The layout of the T1D progression stages include: seroconversion to two or more autoantibodies with normal glycemic levels (stage 1), progression to dysglycemia (stage 2), clinical diagnosis of symptomatic T1D (stage 3), and long-standing T1D (stage 4).

While our understanding of the etiology is incomplete, particularly due to the lack of evidence for what triggers the immune activation and autoimmune response, we do have a firm understanding of the pathogenesis that leads to T1D development. Professional antigen presenting cells, such as dendritic cells, drive the activation of beta cell-targeted T cells. When CD4 helper cells become activated, they can lead to the activation of additional B cells and CD8 T cells. As a result, CD8 cells will begin targeting beta cells for destruction and B cells will produce autoantibodies that also target the islet. Ultimately, the destruction of beta cells within the pancreatic islet lead to the onset of symptomatic T1D. Islet-targeted autoantibodies appear months or years before the total destruction of beta cells. Therefore, they are considered an ideal diagnostic and prognostic biomarker for T1D.

While an environmental trigger is necessary for initiating disease progression, T1D is a multifactorial disease that requires the interaction between multiple genes and environmental factors. Candidate gene studies have revealed many genes, such as CTLA4 and IL2RA; However, the advent of GWAS led to an explosion of novel genes associated with T1D exceeding 60 by 2012.[11] One of the more prominent genetic determinants discovered has been the HLA loci on chromosome 6p21. HLA genes have shown to account for 40-50% familial aggregation of T1D.[12] Among the HLA genes, DQ and DR haplotypes confer some of the highest risk. These HLA class II genes encode major histocompatibility complex (MHC) proteins, surface proteins that are responsible for regulating our adaptive immune system and therefore may play a role in T1D autoimmunity. Past studies have found that HLA-DQA**0102/DQB**0602 to be associated with a slower T1D progression among islet cell antibody (ICA)-positive relatives of diabetic patients, suggesting a possible protective haplotype.[13] However, other studies have shown HLA-DRB1*03 (DR3) or HLA-DRB1*04 (DR4) with DQB1*03:02 (DQ8) for significantly increasing the risk of T1D among relatives of diabetic patients.[14]

HLA allele typing is a practical tool for assessing the risk for T1D and further helps expand our understanding of T1D pathogenesis. In particular, HLA typing can eventually be useful in prevention and intervention trials that test potential preventative treatments in high-risk subjects.[15] Acquired through the Fred Hutchinson Cancer Research Center, the Diabetes Prevention Trial of Type 1 Diabetes (DPT-1) data, stored within the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Central repository, was originally designed for determining whether insulin could prevent or delay the onset of overt diabetes in relatives of patients with diabetes.[16] However for the purpose of our study, we will be assessing

which HLA variants among the DPT-1 participants are associated with the progression of T1D. To do this, we performed a survival analysis by looking at time from identified patient seroconversion to T1D onset in order to determine which individual alleles and haplotypes are correlated with accelerated or decelerated T1D progression.

II. METHODS

Data:

-DPT-1:

The data was acquired from the Diabetes Prevention Trial-1(DPT-1).[16] Two randomized clinical trials took place from 1994-2003 and were investigating whether insulin during prediabetic stages prolonged the transition from pre-diabetes to stage 3 symptomatic T1D. Participants were recruited across 380 recruitment centers and screened based on first and second-degree relatives who have been diagnosed with T1D. The total participant enrollment was 711 participants. Potential subjects underwent islet-cell antibody titer screening in order to detect the presence of autoantibody biomarkers, indicating seroconversion among individuals. Those with an islet-cell antibody (ICA) titer ≥ 10 were included in the study with the exemption of those who were carriers of HLA-DQA1*0102/DQB1*0602(a protective haplotype known to slow the progression of T1D). After further excluding patients with no or poor DNA samples, we are left with 676 patients who qualified to participate in the investigation. However, we needed to exclude 6 patients during the analysis stage because they lacked any clinical variable data required for covariate adjustments. Based on insulin autoantibody(ICA) status, first phase insulin response to intravenous glucose, and oral glucose tolerance(OGT), subjects were categorized into two

groups: high and low risk. High risk subjects were those who tested positive for ICA and those who fell below a clinical threshold for first phase insulin response to intravenous glucose, whereas low risk subjects were those who tested positive for ICA and had normal first phase insulin response to intravenous glucose. Those labeled with high risk had a projected risk of >50% over five years for developing T1D while those labeled with low risk had a projected risk of 25-50% over five years for developing T1D. Those in high risk groups either received parenteral insulin in the treatment arm or were placed under observation in the control. The low risk group either received oral insulin or a placebo.

-Subjects, Risk factors, and clinical outcome:

Age eligibility for the study was 3 to 45 years old with an immediate family member with T1D or 3 to 20 years old with an extended family member with T1D. The exclusion criteria were the following: those with no diabetes, no previous history of insulin or diabetic medication, have not received other forms of T1D preventative medication or immunosuppressive drugs, and have no serious disease. The DPT-1 data included a range of demographic and clinical characteristics including: race, gender, age, treatment type, risk level, and use of insulin. All subjects were followed from time of seroconversion to clinical diagnosis of symptomatic T1D or end of the study. During this duration, the subjects underwent follow-up every six months at a clinic to perform OGT to assess glycemic status, the primary study endpoint. The subjects were followed until the onset of T1D or termination of the investigation. The progression from stage 1 islet autoimmunity to stage 3 clinical diagnosis was quantified through censored time-to-onset of T1D.

-Biomarker Measurement:

Autoantibodies were used to measure the seroconversion of subjects to stage 1 and to classify subjects into high and low risk groups. During the 6 month follow-ups, autoantibody assays and blood glucose levels were used to determine which stage the subjects were in, including whether or not subjects could be clinically diagnosed with T1D.

-Genotyping:

HLA genotypes were collected from Next generation targeted sequencing technologies(NGTS). HLA genotyping was carried using the ScisGo HLA v6 typing kit (Scisco Genetics Inc., Seattle, WA) The procedure uses an amplicon-based 2-stage PCR, followed by sample pooling and sequencing using a MiSeq v2 PE500 (Illumina, San Diego, CA). The protocol yielded 3-field coverage of the HLA genes. Phase within each gene was determined in part by overlapping sequences for HLA class I and database lookup table for HLA class II.[17] As NGTS sequenced DNA nucleotides of selected exons and used sequences to infer DQA1 and DQB1 alleles, we determined HLA-DQA1 and -DQB1 genotypes at the high resolution of 6 digits or higher, based on the HLA nomenclature of IMGT.

Statistical Analysis:

Before performing data analysis, the HLA data was subjected to data integration, transformation, and quality control(QC) steps. We started by merging the genotype data with the subject's clinical data. Duplicates needed to be removed from the clinical data set prior to the merging step. Both the genotype and clinical data had undergone quality control for sample and genotype missingness. HLA is an extremely polymorphic and highly multiallelic loci; therefore, we

decided to merge rare alleles rather than applying a minor allele frequency(MAF) QC step. Finally, we manipulated the gene columns into columns of individual unique alleles and haplotypes. We originally had 18 columns of 9 gene pairs: HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-DRA1, HLA-DRB1, HLA-DQA1, HLA-DQB1, and HLA-DRB3/4/5. After transforming the data, we had a total of 302 unique allele columns. For the haplotype transformation, we created every combination of haplotypes between two different genes from the genes mentioned above.

Descriptive data analysis was performed in order to determine the frequencies and distributions of the clinical variables of interest and select which covariates to include in the association analysis. We selected the Cox proportional hazards regression model for the allelic association analysis. The predictor variables were the individual HLA alleles, and the survival time was the duration from seroconversion to clinical diagnosed T1D. The cox regression model was chosen for survival analysis because of its ability to evaluate simultaneously the effect of several factors on T1D progression. The model is expressed as a hazards function:

$$h(t)=h0(t)\times exp(b_1x_1+b_2x_2+...+b_px_p),$$

in which “t” represents survival time, “h(t)” was the hazard function determined by the predictor variables “x_p”, “h(0)” was the baseline hazard function estimated from the data, and the coefficient “b_p” quantified the allelic association or measures the effect size of the covariates used for adjustments. For each allele, the subjects carry 0, 1, or 2 of the given allele. The null hypothesis is that no relationship exists between the allele and survival time or that b_1 equals zero. The alternative hypothesis is that we have a significant association or that b_1 is not equal to zero. The significance level, or alpha, is 0.05. The model behaves similarly for the haplotype association analysis with the only difference being that the predictor variable is the unique

haplotype rather than an allele genotype. The cox regression model will be run once per allele or haplotype within the study. Normally this would be problematic because we are at an increased risk for false positives when running the association model multiple times. In order to resolve this problem, we will be using the false discovery rate (FDR), specifically the Benjamini-Hochberg Procedure, to adjust for multiple testing comparisons. This approach is ideal because it is less conservative than the traditional Bonferroni method.[18] We selected FDR because we are concerned that a more conservative approach would lead to missed findings. However, FDR consequently leads to a higher rate of false positives.

III. RESULTS

Sample Description

The final allelic and haplotype association analysis included 670 participants. As shown in table 1, the sample is multi-ethnic, but the majority of the subjects (90.7%) identified as non-Hispanic white race/ethnicity. Table 1 shows the mean age enrolled (11.4 years), the percentage of participants who developed T1D within the observed time of the study (35.9%), and the average time, in days, of onset or when last contacted (1170.5). The table also includes the distribution of males in the study (56.4%), the distribution of subjects with a 5-year risk of $\geq$50% defined as "high-risk" (47.2%), and those who were administered insulin prior to the onset of T1D (50.1%). The original DPT-1 study found no evidence that insulin altered the progression towards T1D or impacted the risk of developing T1D. In figure 1, we see that the age distribution is heavily skewed to the right with more young people in the study and a fairly balanced gender percentage

across all age groups. We have a higher count of young people in the study because infants and adolescents tend to have more rapid onset of T1D. Previous research shows that children progress more rapidly from autoantibody positivity to T1D and have lower C-peptide levels compared to adults.[20] In fact, prevalence among children has been growing over the last few decades, with The SEARCH for Diabetes in Youth (SEARCH) study showing a 21.1% rise in prevalence between 2001-2009 among ages 0-19.[19]

Some of these variables are potentially confounding or effect modifiers. With that in mind, we will select covariates for adjustment in our association analysis.

Covariate Selection

To further explore possible covariates, we ran cox regression analyses on the following variables: treatment, age, gender, and race. If we look at table 2, we see that race, gender, and insulin treatment were not found to be associated with T1D progression. However, age at a cutoff of 8 years old was found to be significantly associated with T1D progression. We also adjusted for the risk level as it is a variable that groups the subjects into either low (<50%) or high risk (≥ 50%) for developing T1D in a 5-year period. The degree of risk was calculated based on testing positive for islet-cell antibodies or insulin antibodies, first-phase insulin response to intravenous glucose, and assessed oral glucose tolerance.

Allelic Association Analysis

We removed rare alleles and created one individual column titled "rare" that grouped all alleles that were present 1% or less of the time within the 670 participants. The final merged "rare" column grouped 191 rare alleles together, leaving 112 common alleles out of the original 302

unique alleles to undergo association testing. Looking at table 3, the allelic association analysis showed that one significant allele, DQB1*03:02:01, was significantly associated with the progression of T1D. The FDR adjusted p-value was 0.01062222 with a count allele frequency of 66%. The hazards ratio was 1.49 with a 95% CI of (1.29,1.70). Figure 2 shows the kaplan meier curve for the subjects who either carried 2, 1, or 0 of the DQB1*03:02:01 allele. The curve shows that those who are heterozygous or homozygous for the allele have an accelerated T1D progression compared to those who lacked the allele. We can also see that homozygous individuals had a steeper survival probability compared to heterozygous individuals.

Haplotype Association Analysis

Within each pair of genes, we removed rare haplotypes and created one individual column titled "rare_10" that grouped all the haplotypes that had less than 10 haplotype counts within the 670 participants. Of the paired genes that had significant haplotypes, the DQA1-DQB1 haplotypes had 15 unique haplotypes and DQA1-B haplotypes had 21 unique haplotypes. Looking at table 4 and 5, we see that the following haplotypes are significantly associated with T1D progression: DQA1*03:01:01-DQB1*03:02:01, DQA1*03:01:01-B*18:01:01, and DQA1*03:01:01-B*35:01:01. DQA1*03:01:01-DQB1*03:02:01 has a count allele frequency(CAF) of 33.2% with an FDR adjusted p-value of 0.03021067. The hazards ratio was 1.36 with a 95% CI of (1.16, 1.56). DQA1*03:01:01-B*18:01:01 has a CAF of 2.3% with an FDR adjusted p-value of 0.04977038. The hazards ratio was 1.99 with a 95% CI of (1.52, 2.46). DQA1*03:01:01-B*35:01:01 has a CAF of 1.3% with an FDR adjusted p-value of 0.04977038.The hazards ratio was 2.13 with a 95% CI of (1.60, 2.66). Figure 3-5 show the kaplan meier curves for the subjects who either carried 2, 1, or 0 counts of the DQA1*03:01:01-DQB1*03:02:01, DQA1*03:01:01-

B*18:01:01, and DQA1*03:01:01-B*35:01:01 haplotypes, respectively. The curve for all three haplotypes shows that those who are heterozygous or homozygous for the haplotype have an accelerated T1D progression compared to those who lacked the haplotype.

IV. DISCUSSION

The key finding of the allelic association analysis was DQB1*03:02:01, a gene from the MHC class II family. MHC class II genes encode information for producing proteins that are presented on the surface of professional antigen-presenting cells within the body. These cells engulf extracellular proteins, break them down to peptides, present them on surface proteins, and ultimately activate antigen specific CD4 T cells. The HLA-DQB1 gene is associated with a handful of autoimmune diseases, including T1D. In a case-control study published in 2014, 85 unrelated Egyptian children with T1D recruited consecutively from the Pediatric Diabetes Endocrinology Outpatients Clinic were genotyped for HLA-DQB1 alleles.[21] The findings showed that DQB1*03:02:01, among other alleles within the gene, were associated with an increased risk for T1D.[21] Our study provides further evidence and support for the fact that DQB1*03:02:01 plays a role in the development of T1D. Unlike the Egyptian Study, however, our study highlighted that DQB1*03:02:01 leads to an accelerated progression towards T1D, a temporal component that has not been captured through previous case-control studies.

Within our haplotype association analysis, the following three haplotypes were positively associated with the progression of T1D: DQA1*03:01:01-DQB1*03:02:01, DQA1*03:01:01-B*18:01:01, and DQA1*03:01:01-B*35:01:01. DQA1*03:01:01-DQB1*03:02:01 is a gene

pairing of two MHC class II, whereas the other two haplotypes are a gene pairing of an MHC class II(HLA-DQA1) and MHC class I (HLA-B).

The Type 1 Diabetes Genetics Consortium(T1DGC) has published lists of haplotypes that are significantly associated with T1D, with DRB*04:01-DQA1*03:01:01-DQB1*03:02:01 haplotype having been identified as highly predisposing for T1D.[22] While our haplotype association analysis did not investigate DRB1-DQA1-DQB1 haplotypes, our findings highlight that the combination of the DQA1*03:01:01 and DQB1*03:02:01 alleles appear to contribute to a complex interaction of other possible HLA alleles that lead to accelerated T1D progression.[22] DQ and DR often represent the major genetic determinant of T1D, but interestingly our study did not find any significant DQ-DR haplotypes associated with T1D progression.

For future exploration, we would encourage running DRB1-DQA1-DQB1 haplotypes association analysis in order to clarify whether the presence of specific DRB1 alleles is essential for significant association of the identified DQA1-DQB1 haplotype. In addition, we recommend running conditional testing on each allele and haplotype, stratified by significant DR/DQ alleles and haplotypes either identified within the study or those found to be clinically significant in the literature. This procedure would be useful for accounting for any effects due to LD with any significant HLA variants within our allelic or haplotype association analysis.

In addition to the significant allele, the DQA1*03:01:01-B*18:01:01 and DQA1*03:01:01-B*35:01:01 haplotypes were considered interesting findings. Historically, it was believed that HLA-B alleles were only associated with T1D because of strong LD with more prominent genetic determinants for T1D, such as HLA class II loci.[23] However, with larger studies that offered substantial increase in power researchers were able to adjust for LD of HLA class II loci. These studies provided strong evidence for HLA-B alleles associated with T1D, including

B*18:01:01.[24] While our study did not find any significance for the B*18:01:01 allele alone, perhaps due to a lack of power attributed by the rarity of the allele (6.3% CAF) with our study, our findings may suggest that an interaction with DQA1*03:01:01 may exist. While multiple HLA-B alleles are correlated with T1D, no existing evidence for B*35:01:01 is currently available. Therefore, it possible that DQA1*03:01:01-B*35:01:01 haplotypes may be a false positive or in LD with additional DQ or DR alleles. Further investigation of more complex haplotypes could be performed on this dataset.

Our study has several limitations. One of them is that we did not have a large enough sample size. The number of alleles within the HLA loci exceeds 23,000 with over 200 genes.[23, 25] Compared to larger studies, such as the T1DGC through which >14,000 samples were collected and genotyped, our study sample of 670 after QC is far too underpowered for alleles with rare population frequencies.[22] Diversity of participants is another serious issue with our study containing 90.7% non-Hispanic white participants. With the abundance of alleles within the HLA loci being exceedingly high, it is imperative that we gather samples from across different populations. For example, HLA-B*15:02, a clinically important HLA allele for immune-mediated drug hypersensitivity reactions, has drastically different allele frequencies for Europeans(1/1427), Western Asian(1/3774), Eastern Asian(1/9), and Northern and Central America(1/240).[26] In addition, admixture populations make for unique combinations of alleles resulting in novel haplotypes associated with T1D. Without proper representation within studies, we risk not identifying these significant alleles and haplotypes.

Future studies into T1D progression have many different avenues to explore. While autoantibodies were measured for identifying seroconversion among qualified participants, we did not investigate the association between HLA variants and autoantibody types, nor did we

explore an association between autoantibody type and T1D progression. Many studies have begun exploring these associations and coming up with findings. For example, it has been shown that children predisposed with *DR3-DQ2,* an HLA haplotype, have an immune response that is primarily against GAD signaled by GAD autoantibodies (GADA) as the first-appearing autoantibody; additionally, GADA positive children have an overall lower risk for developing T1D(HR 0.60) compared to IA-2A positive children (HR 1.97).[27,28] Being able to predict when and which autoantibodies develop in recently seroconverted subjects will be extremely useful for the purposes of advancing our understanding of T1D progression and risk stratification among patients.

In conclusion, identifying T1D-associated HLA variants benefits the clinical setting and someday will transform how we approach diabetes on a population level. Earlier detection for genetic predisposition already creates an opportunity for subjects to participate in clinical trials that are specifically targeted for those who are not positive for islet autoantibodies but are considered high risk for developing T1D. These clinical trials have great potential for identifying therapeutics that effectively halt T1D progression in pre-diabetic individuals. As our understanding of T1D etiology and pathogenesis continues to progress, the discovery of novel HLA alleles/haplotypes and their association to islet-targeted autoantibody and T1D progression will lead to the improvement of models which will ultimately support new strategies in the prevention, classification, and treatment of T1D.

References:

1. Mobasseri, Majid, et al. "Prevalence and Incidence of Type 1 Diabetes in the World: a Systematic Review and Meta-Analysis." *Health Promotion Perspectives*, Tabriz University of Medical Sciences, 30 Mar. 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7146037/.
2. *Type 1 Diabetes*. https://www.idf.org/aboutdiabetes/type-1-diabetes.html. Accessed 23 Apr. 2021.
3. 10Piffaretti C, Mandereau□Bruno L, Guilmin□Crepon S, Choleau C, Coutant R, Fosse□Edorh S. Trends in childhood type 1 diabetes incidence in France, 2010□2015. *Diabetes Res Clin Pract*. 2019;149:200□207.
4. Patterson CC, Dahlquist GG, Gyurus E, Green A, Soltesz G, Group ES: Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. Lancet 2009;373:2027.
5. Noble, Janelle A, and Ana M Valdes. "Genetics of the HLA Region in the Prediction of Type 1 Diabetes." *Current Diabetes Reports*, U.S. National Library of Medicine, Dec. 2011, www.ncbi.nlm.nih.gov/pmc/articles/PMC3233362/.
6. Lucier, Jessica. "Diabetes Mellitus Type 1." *StatPearls [Internet]*., U.S. National Library of Medicine, 19 Nov. 2020, www.ncbi.nlm.nih.gov/books/NBK507713/.
7. "Diabetes - Symptoms and Causes." *Mayo Clinic*, https://www.mayoclinic.org/diseases-conditions/diabetes/symptoms-causes/syc-20371444. Accessed 23 Apr. 2021.
8. Rewers, Marian, and Johnny Ludvigsson. "Environmental Risk Factors for Type 1 Diabetes." *Lancet (London, England)*, vol. 387, no. 10035, June 2016, pp. 2340–48. *PubMed Central*, doi:10.1016/S0140-6736(16)30507-4.

9. Filippi, Christophe M, and Matthias G von Herrath. "Viral Trigger for Type 1 Diabetes: Pros and Cons." *Diabetes*, American Diabetes Association, Nov. 2008, www.ncbi.nlm.nih.gov/pmc/articles/PMC2570378/.
10. University of Bristol. "Islet Cell Autoantibodies." *Islet Cell Autoantibodies | Bristol Medical School: Translational Health Sciences | University of Bristol*, University of Bristol, 20 Sept. 2017, www.bristol.ac.uk/translational-health-sciences/research/diabetes/research/autoantibodies/.
11. Bakay, Marina, et al. "Genes Involved in Type 1 Diabetes: An Update." *Genes*, vol. 4, no. 3, Sept. 2013, pp. 499–521. *PubMed Central*, doi:10.3390/genes4030499.
12. Couper, Jenny J; Haller, Michael J; Greenbaum, Carla J; Ziegler, Anette-G; Wherrett, Diane K; Knip, Mikael; Craig, Maria E (2018). *ISPAD Clinical Practice Consensus Guidelines 2018 Stages of type 1 diabetes in children and adolescents. Pediatric Diabetes, (), –.* doi:10.1111/pedi.12734
13. Greenbaum CJ;Schatz DA;Cuthbertson D;Zeidler A;Eisenbarth GS;Krischer JP; "Islet Cell Antibody-Positive Relatives with Human Leukocyte Antigen DQA1*0102, DQB1*0602: Identification by the Diabetes Prevention Trial-Type 1." *The Journal of Clinical Endocrinology and Metabolism*, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/10720072/.
14. Nguyen, Cao, et al. "Definition of High-Risk Type 1 Diabetes HLA-DR and HLA-DQ Types Using Only Three Single Nucleotide Polymorphisms." *Diabetes*, vol. 62, no. 6, June 2013, pp. 2135–40. *diabetes.diabetesjournals.org*, doi:10.2337/db12-1398.
15. Van Belle, Tom L., et al. "Type 1 Diabetes: Etiology, Immunology, and Therapeutic Strategies." *Physiological Reviews*, vol. 91, no. 1, Jan. 2011, pp. 79–118. *journals.physiology.org (Atypon)*, doi:10.1152/physrev.00003.2010.

16. *NIDDK: Diabetes Prevention Trial of Type 1 Diabetes (DPT-1)*. https://repository.niddk.nih.gov/studies/dpt-1/.
17. Nelson WC, Pyo CW, Vogan D, Wang R, Pyon YS, Hennessey Cet al. : An integrated genotyping approach for HLA and other complex genetic systems. Hum Immunol 2015;76:928.
18. Narum, Shawn R. "Beyond Bonferroni: Less Conservative Analyses for Conservation Genetics." *Conservation Genetics*, vol. 7, no. 5, Oct. 2006, pp. 783–87. *Springer Link*, doi:10.1007/s10592-005-9056-y.
19. Chiang, Jane L., et al. "Type 1 Diabetes in Children and Adolescents: A Position Statement by the American Diabetes Association." *Diabetes Care*, vol. 41, no. 9, Sept. 2018, pp. 2026–44. *care.diabetesjournals.org*, doi:10.2337/dci18-0023.
20. Leete P;Mallone R;Richardson SJ;Sosenko JM;Redondo MJ;Evans-Molina C; "The Effect of Age on the Progression and Severity of Type 1 Diabetes: Potential Effects on Disease Mechanisms." *Current Diabetes Reports*, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/30259209/.
21. Mosaad, Youssef M., et al. "HLA-DQB1* Alleles and Genetic Susceptibility to Type 1 Diabetes Mellitus." *World Journal of Diabetes*, vol. 3, no. 8, Aug. 2012, pp. 149–55. *PubMed Central*, doi:10.4239/wjd.v3.i8.149.
22. Noble, Janelle A., and Henry A. Erlich. "Genetics of Type 1 Diabetes." *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 1, Jan. 2012. *PubMed Central*, doi:10.1101/cshperspect.a007732.

23. Rodey, Glenn E., et al. “HLA-DR Specificities among Black Americans with Juvenile-Onset Diabetes.” *New England Journal of Medicine*, vol. 301, no. 15, Oct. 1979, pp. 810–12. *DOI.org (Crossref)*, doi:10.1056/NEJM197910113011503.
24. Varney MD, Valdes AM, Carlson JA, Noble JA, Tait BD, Bonella P, Lavant E, Fear AL, Louey A, Moonsamy P, Mychaleckyj JC, Erlich H. HLA DPA1 DPB1 alleles haplotypes contribute to the risk associated with type 1 diabetes: analysis of the type 1 diabetes genetics consortium families. Diabetes. 2010;59:2055–2062.
25. Gonzalez-Galarza, Faviel F., et al. “Allele Frequency Net Database (AFND) 2020 Update: Gold-Standard Data Classification, Open Access Genotype Data and New Query Tools.” *Nucleic Acids Research*, vol. 48, no. D1, Jan. 2020, pp. D783–88. *Silverchair*, doi:10.1093/nar/gkz1029.
26. Zhou, Yitian, et al. “Global Frequencies of Clinically Important HLA Alleles and Their Implications For the Cost-Effectiveness of Preemptive Pharmacogenetic Testing.” *Clinical Pharmacology & Therapeutics*, vol. 109, no. 1, 2021, pp. 160–74. *Wiley Online Library*, doi:https://doi.org/10.1002/cpt.1944.
27. Regnell, Simon E, and Åke Lernmark. “Early prediction of autoimmune (type 1) diabetes.” Diabetologia vol. 60,8 (2017): 1370-1381. doi:10.1007/s00125-017-4308-1
28. Jacobsen, Laura M., et al. “The Risk of Progression to Type 1 Diabetes Is Highly Variable in Individuals with Multiple Autoantibodies Following Screening.” *Diabetologia*, vol. 63, no. 3, Mar. 2020, pp. 588–96. *Springer Link*, doi:10.1007/s00125-019-05047-w.

Figures

Figure 1. Distribution of Age and Gender within the Participants. The continuous age variable has a right skewed distribution. The gender percentage are approximately even across all age groups.

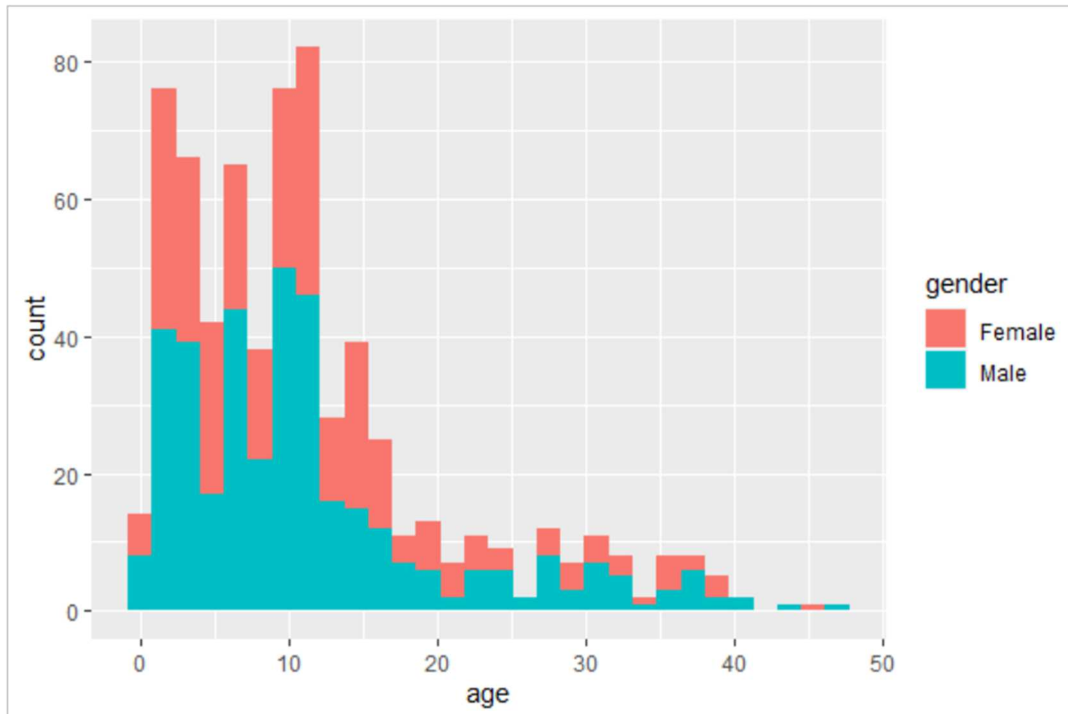


Figure 2. Kaplan Meier Curve of HLA DQB1*03:02:01 allele and time to onset of T1D. Allele Count: 2= homozygous, 1=heterozygous, 0= no allele. The curve is adjusted for the following variables: gender, age, and risk group.

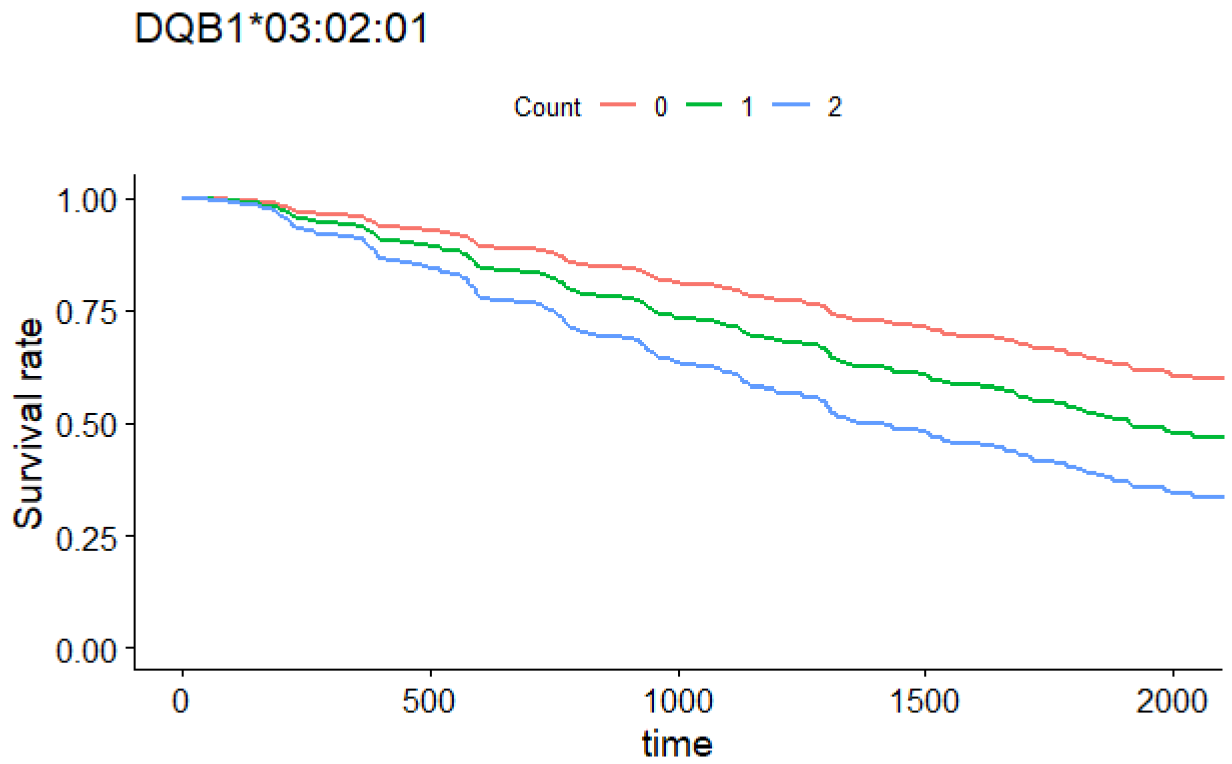


Figure 3. Kaplan Meier Curve of DQA1*03:01:01-DQB1*03:02:01 haplotype and time to onset of T1D. Allele Count: 2= homozygous, 1=heterozygous, 0= no allele. The curve is adjusted for the following variables: gender, age, and risk group.

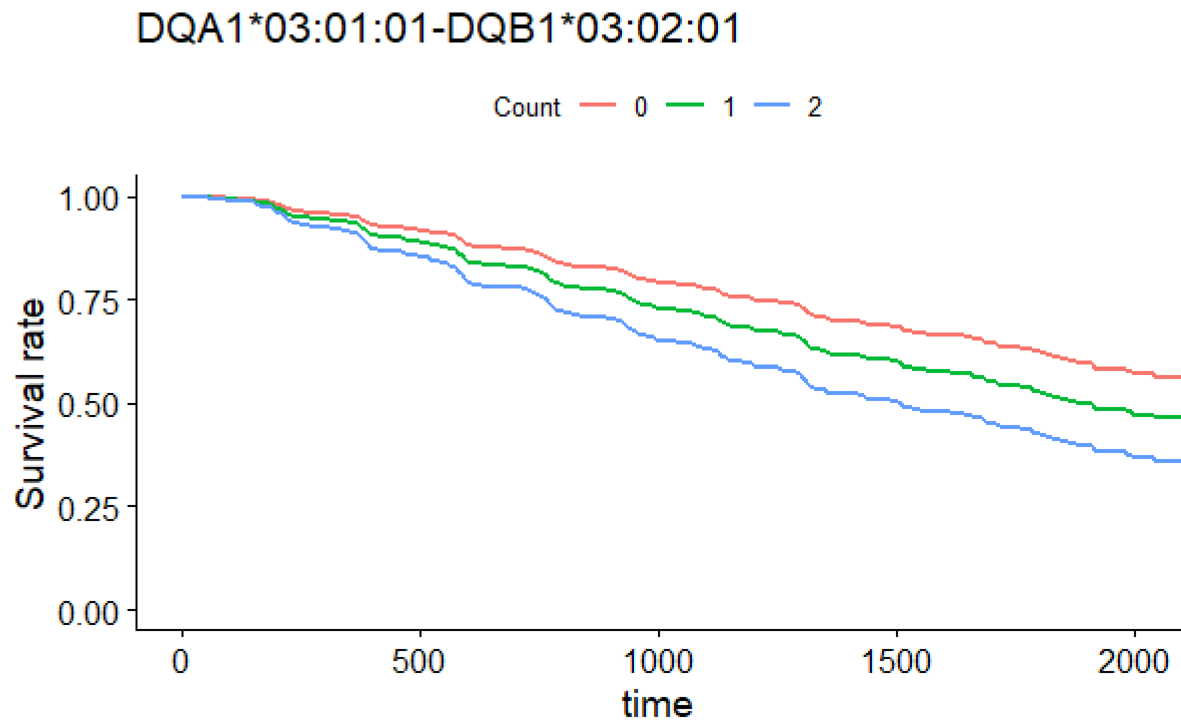


Figure 4. Kaplan Meier Curve of DQA1*03:01:01-B*18:01:01 haplotype and time to onset of T1D. Allele Count: 1=heterozygous, 0= no allele. The curve is adjusted for the following variables: gender, age, and risk group.

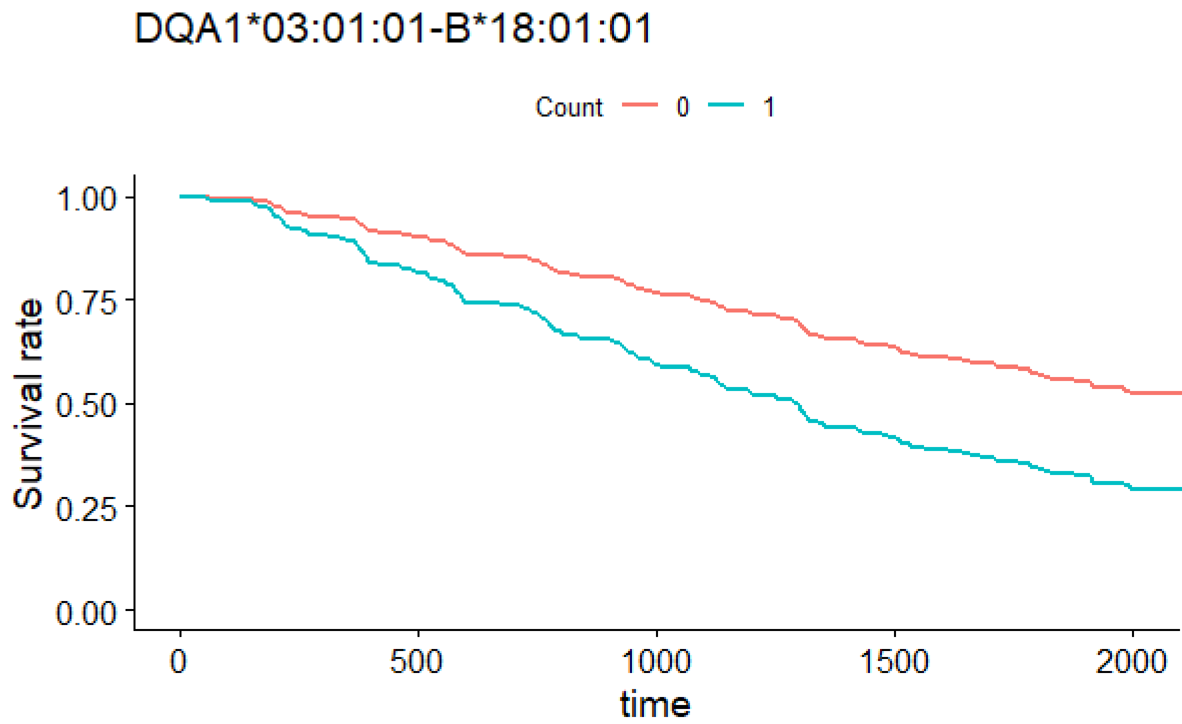
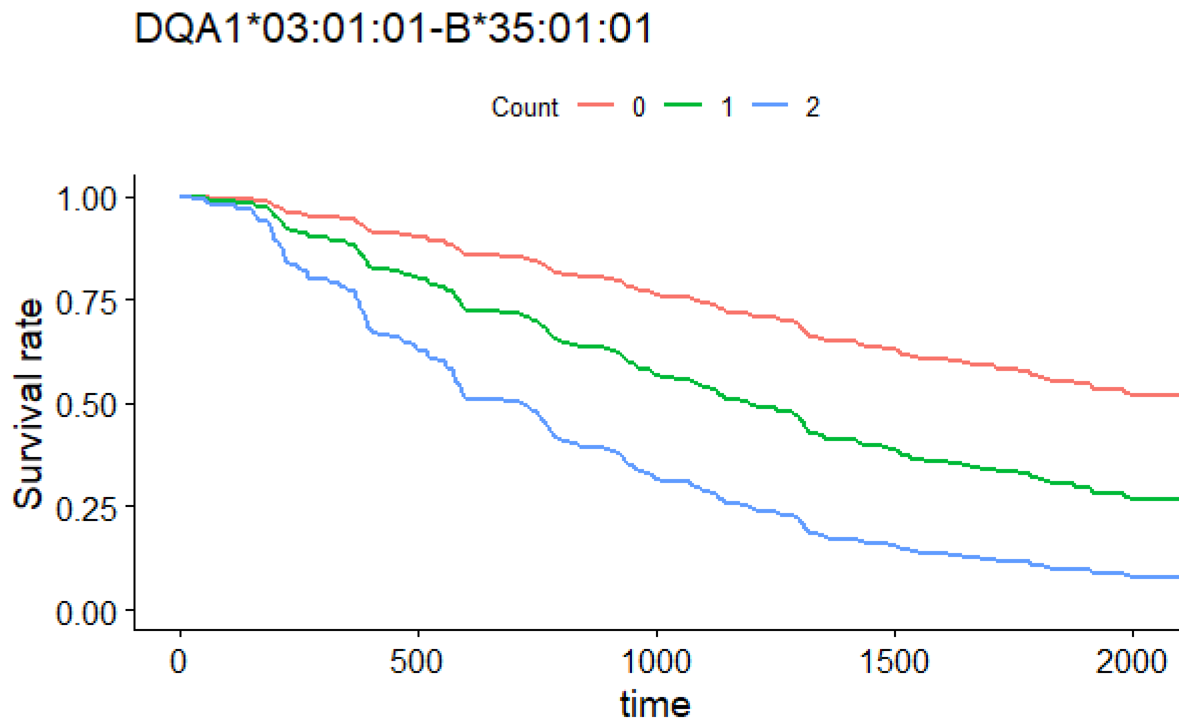


Figure 5. Kaplan Meier Curve of DQA1*03:01:01-B*35:01:01 haplotype and time to onset of T1D. Allele Count: 2= homozygous, 1=heterozygous, 0= no allele. The curve is adjusted for the following variables: gender, age, and risk group.



Tables

Table 1: Characteristics of subjects included in analysis

Study	The Diabetes Prevention Trial of Type 1 Diabetes (DPT-1)
Institution	Fred Hutch
Diagnosed, %	35.9
Mean Age Enrolled, Years	11.4
Mean Time of Onset or Last Time Contacted, Days	1170.5
High Risk Patients, %	47.2
Received Insulin, %	50.1
Male, %	56.4
Race/Ethnicity	
White (%)	608 (90.7)
Asian (%)	3 (0.4)
Black (%)	0 (0)
Hispanic (%)	29 (4.3)
Other (%)	7 (1)
Unkown (%)	14 (2.1)
Total	670

Table 2. Estimated coefficients in the Cox proportional hazards regression model for the following covariates: Race, Gender, Treatment, and Age. **Note:** coef=regression coefficient, exp.coef=Hazards Ratio, se.coef= Standard Error, and z=Z-score, and Blue Highlight=Reference Group

Variables	Frequency	coef	exp.coef	se.coef	z	p-value
Race						
White	608	-	-	-	-	-
Asian	0	-	-	-	-	-
Black	0	-	-	-	-	-
Hispanic	29	-0.36	0.7	0.41	-0.86	0.39
Other	7	0.66	1.93	0.51	1.3	0.19
Unknown	14	-0.53	0.5	0.5	-0.91	0.36
Gender						
Male	378	-	-	-	-	-
Female	292	0.11	1.2	0.13	0.82	0.41
Insulin treatment						
Oral.placebo	177	-	-	-	-	-
Oral.Insulin	177	-0.23	0.79	0.21	-1.08	0.3
Age						
Age>8	369	-	-	-	-	-
Age<8	301	0.39	1.48	0.13	3	0.02*10^-1

Table 3. Estimated coefficient in the Cox proportional hazards regression model for the significant HLA alleles. The frequencies for the significant alleles. aa= subjects who do not carry the allele ; Aa=heterozygous for the allele ; AA=homozygous for the allele **Note:** coef=regression coefficient, HR=Hazards Ratio, SE= Standard Error, Z=Z-score, P=p-value, and q=FDR adjusted p-value.

HLA alleles and their associated results on T1D Progression

	allele	aa	aA	AA	coef	HR	SE	Z	p	q
1	DQB1*03:02:01	227	369	74	0.4	1.49	0.1	3.9	9.48*10^-5	0.01

Table 4. HLA DQA1/DQB1 Haplotype Association Analysis

Note: coef=regression coefficient, exp.coef=Hazards Ratio, se.coef= Standard Error, z=Z-score, Pr..z..=p-value, and q=FDR adjusted p-value. Green highlight are significant haplotypes.

HLA DQA1 and DQB1 Haplotypes and their associated results on T1D Progression

	coef	exp.coef.	se.coef.	z	Pr...z..	q
DQA1*01:01:01-DQB1*05:01:01	-0.03	0.96	0.20	-0.18	0.85	0.91
DQA1*01:01:02-DQB1*05:01:01	0.09	1.09	0.42	0.21	0.83	0.91
DQA1*01:02:01-DQB1*06:04:01	-0.19	0.82	0.26	-0.74	0.45	0.79
DQA1*01:02:02-DQB1*05:02:01	-0.27	0.75	0.50	-0.55	0.58	0.79
DQA1*01:03:01-DQB1*06:03:01	-0.18	0.83	0.28	-0.64	0.52	0.79
DQA1*02:01:01-DQB1*02:02:01	-0.39	0.67	0.25	-1.50	0.13	0.39
DQA1*03:01:01-DQB1*03:02:01	0.31	1.36	0.10	3.08	0.02*10^-1	0.03
DQA1*03:02:01-DQB1*03:03:02	0.64	1.89	0.45	1.41	0.15	0.39

DQA1*03:03:01-DQB1*02:02:01	-0.38	0.67	0.69	-0.55	0.57	0.79
DQA1*03:03:01-DQB1*03:01:01	-0.47	0.62	0.22	-2.08	0.036	0.276
DQA1*03:03:01-DQB1*03:02:01	0.32	1.38	0.19	1.65	0.09	0.36
DQA1*04:01:01-DQB1*04:02:01	0.11	1.11	0.36	0.30	0.75	0.91
DQA1*05:01:01-DQB1*02:01:01	-0.008	0.99	0.10	-0.08	0.93	0.93
DQA1*05:05:01-DQB1*03:01:01	-0.48	0.61	0.27	-1.76	0.07	0.36
rares_10	-0.27	0.76	0.26	-1.04	0.29	0.63

Table 5. HLA DQA1/B Haplotype Association Analysis

Note: coef=regression coefficient, exp.coef=Hazards Ratio, se.coef= Standard Error, z=Z-score, Pr..z..=p-value, and q=FDR adjusted p-value. Green highlight are significant haplotypes.

HLA DQA1 and B Haplotypes and their associated results on T1D Progression

	coef	exp.coef.	se.coef.	Z	Pr...z..	Q
DQA1*01:01:01-B*35:01:01	-0.15	0.85	0.41	-0.37	0.70	0.86
DQA1*01:01:02-B*14:02:01	0.44	1.55	0.40	1.09	0.27	0.81
DQA1*01:02:01-B*40:01:02	-0.09	0.91	0.45	-0.19	0.84	0.91
DQA1*02:01:01-B*44:03:01	-0.54	0.57	0.45	-1.20	0.22	0.81
DQA1*03:01:01-B*07:02:01	0.25	1.29	0.26	0.97	0.33	0.81
DQA1*03:01:01-B*15:01:01	0.05	1.05	0.15	0.37	0.70	0.86
DQA1*03:01:01-B*18:01:01	0.69	1.99	0.24	2.86	0.04*10^-1	0.04
DQA1*03:01:01-B*35:01:01	0.75	2.13	0.26	2.82	0.04*10^-1	0.04

DQA1*03:01:01-B*38:01:01	-0.64	0.52	0.56	-1.13	0.25	0.81
DQA1*03:01:01-B*40:01:02	0.28	1.32	0.22	1.27	0.20	0.81
DQA1*03:01:01-B*44:02:01	-0.32	0.72	0.41	-0.77	0.44	0.81
DQA1*03:01:01-B*51:01:01	-0.13	0.87	0.41	-0.33	0.73	0.86
DQA1*03:01:01-B*55:01:01	-0.33	0.71	0.58	-0.57	0.56	0.86
DQA1*03:03:01-B*27:05:02	-0.28	0.75	0.39	-0.73	0.46	0.81
DQA1*03:03:01-B*44:02:01	0.03	1.03	0.29	0.10	0.91	0.91
DQA1*03:03:01-B*45:01:01	-0.26	0.76	0.58	-0.45	0.64	0.86
DQA1*05:01:01-B*07:02:01	-1.82	0.16	1.00	-1.81	0.06	0.48
DQA1*05:01:01-B*08:01:01	-0.06	0.93	0.13	-0.49	0.61	0.86
DQA1*05:01:01-B*18:01:01	0.24	1.27	0.27	0.87	0.38	0.81

DQA1*05:05:01-B*44:02:01	-0.07	0.92	0.50	-0.15	0.88	0.92
rares_10	-0.09	0.92	0.11	-0.82	0.41	0.81