Published in final edited form as:

Pediatr Diabetes. 2018 March ; 19(2): 277–283. doi:10.1111/pedi.12543.

Prediction of type 1 diabetes using a genetic risk model in the Diabetes Autoimmunity Study in the Young (DAISY)

Brigitte I. Frohnert, MD, PhD^{1,*}, Michael Laimighofer^{2,*}, Jan Krumsiek, PhD^{2,3}, Fabian J. Theis, PhD², Christiane Winkler, PhD⁴, Jill M. Norris, PhD⁵, Anette-Gabriele Ziegler, MD⁴, Marian J. Rewers, MD, PhD¹, and Andrea K. Steck, MD¹

¹Barbara Davis Center for Childhood Diabetes, School of Medicine, University of Colorado, Aurora, CO 80045 USA

²Institute of Computational Biology, Helmholtz Zentrum München, München-Neuherberg 85764 Germany

³German Center for Diabetes Research (DZD), München-Neuherberg 85764 Germany

⁴Institute of Diabetes Research, Helmholtz Zentrum München and Forschergruppe Diabetes, Klinikum rechts der Isar, Technische Universität München, Neuherberg 85764 Germany

⁵Department of Epidemiology, Colorado School of Public Health, University of Colorado, Aurora, CO, 80045 USA

Abstract

BACKGROUND—Genetic predisposition for type 1 diabetes is largely determined by HLA genes; however, over 50 other genetic regions confer susceptibility. We evaluated a previously-reported 10-factor weighted model derived from the Type 1 Diabetes Genetics Consortium to predict development of diabetes in the Diabetes Autoimmunity Study in the Young (DAISY) prospective cohort. Performance of the model, derived from individuals with first-degree relatives (FDR) with type 1 diabetes, was evaluated in DAISY general population (GP) participants as well as FDR subjects.

METHODS—The 10-factor weighted risk model (HLA, *PTPN22*, *INS*, *IL2RA*, *ERBB3*, *ORMDL3*, *BACH2*, *IL27*, *GLIS3*, *RNLS*), 3-factor model (HLA, *PTPN22*, *INS*) and HLA alone were compared for prediction of diabetes in children with complete SNP data (n=1941).

RESULTS—Stratification by risk score significantly predicted progression to diabetes by Kaplan-Meier analysis (GP: p=0.00006; FDR: p=0.0022). The 10-factor model performed better in discriminating diabetes outcome than HLA alone (GP, p=0.03; FDR, p=0.01). In GP, the restricted 3-factor model was superior to HLA (p=0.03), but not different from the 10-factor model (p=0.22). In contrast, for FDR the 3-factor model did not show improvement over HLA (p=0.12) and performed worse than the 10-factor model (p=0.02)

Address Correspondence To: Brigitte I. Frohnert MD PhD, Barbara Davis Center, 1775 Aurora Court, A140, Aurora, CO 80045, Phone (303) 724-2323, FAX (303) 724-6779, brigitte.frohnert@ucdenver.edu.

*These authors contributed equally to this work.

CONCLUSIONS—We have shown a 10-factor risk model predicts development of diabetes in both GP and FDR children. While this model was superior to a minimal model in FDR, it did not confer improvement in GP. Differences in model performance in FDR vs. GP children may lead to important insights into screening strategies specific to these groups.

Keywords

Child; Diabetes Mellitus; Type 1; Epidemiology; Prospective study; Risk Factors

INTRODUCTION

The increasing incidence of type 1 diabetes (T1D) in children can be attributed to the action of environmental factors in a context of genetic predisposition. Current screening strategies for at-risk individuals are limited by the low specificity of genetic screening in the general population, who make up about 85% of those who develop T1D (1), and low sensitivity in those with a family history of T1D. More precise identification of children at high risk of developing T1D is important for recruitment into natural history studies to better understand the etiologic factors contributing to islet autoimmunity (IA) and T1D. Moving forward, these strategies will also allow for better identification of individuals who could benefit from both primary and secondary prevention trials. Further, as population screening for islet autoimmunity is being explored (2), better definition of genetic risk could serve as a second line of screening.

The human leukocyte antigen (HLA) region of chromosome 6p21 plays a significant role, conferring up to 50% of the genetic risk for diabetes (3). In addition, however, more than 50 other genetic susceptibility markers have been associated with development of T1D (1,4–9). While any one non-HLA gene may not confer significant risk increase alone, an improvement in prediction strategy can be achieved by giving weight to varying gene contributions (10). We have previously used multivariable logistic regression and Bayesian feature selection to generate a weighted risk model with SNPs selected from 41 genetic susceptibility markers included in data from the Type 1 Diabetes Genetics Consortium (T1DGC) dataset (10). This 10-factor model included HLA genotype plus nine SNPs from the *PTPN22*, *INS*, *IL2RA*, *ERBB3*, *ORMDL3*, *BACH2*, *IL27*, *GLIS3* and *RNLS* genes and was used to predict progression to T1D or multiple islet autoantibody positive status (10). The model was validated in a group of children and young adults less than 20 years of age with new onset diabetes (11), and in a group of T1D parents of children with T1D from the German BABYDIAB study, with the non-T1D parents as the control group (12). In this previous analysis, the 10-factor model showed improved discrimination of those at risk for development of T1D when compared to HLA genotype alone. Further, the risk of development of T1D or multiple IA status was significantly higher in the portion of the prospectively-followed BABYDIAB cohort with risk scores in the upper quintile when compared to the lower quintile, using Kaplan-Meier analysis (10).

The Diabetes Autoimmunity Study in the Young (DAISY) is a cohort consisting of first degree relatives (FDR) of T1D patients, similar to the BABYDIAB study. However, unlike BABYDIAB, the DAISY cohort also includes individuals recruited from the general

population (GP) based on high-risk HLA status. The purpose of the current study was to validate the weighted 10-factor model of HLA plus nine other SNPs for prediction of development of T1D in participants of the DAISY cohort. Validation in this novel cohort allows examination of the performance of a model developed from a T1D FDR cohort in a group of GP individuals. Further, the performance of this 10-factor model was compared to a more parsimonious 3-factor model or HLA alone.

METHODS

Study Participants

The Diabetes Autoimmunity Study in the Young (DAISY) is a prospective cohort study that has followed 2,547 children at increased risk of type 1 diabetes for a median of nine years. The details of screening and follow-up have been previously published (13,14). Recruitment began in 1993 and included two groups of children: FDR of T1D patients, enrolled between birth and seven years of age; and general population (GP) subjects born at a Denver, Colorado hospital. While FDR subjects were enrolled regardless of HLA type, GP subjects were enriched for high risk HLA-DR,DQ susceptibility genotypes for T1D. Specifically, of the 31,881 newborns screened, all children with DR3/4,DQB1*0302, DR3/3, and DR4/4,DQB1*0302 and a sample of those with DR4/DRx, DQB1*0302, or DR3/DRx (where DRx = DR3 or DR4) were invited to participate in DAISY. Distribution of HLA types for the GP and FDR subjects is shown in Supplemental Figure 1. Characteristics of FDR and GP participants are shown in Supplemental Table 1. Follow-up results are available through July of 2015. At this point, 94 participants had developed T1D. Written informed consent was obtained from the parents of study participants. The Colorado Multiple Institutional Review Board approved all study protocols.

Outcome Measures

Children in DAISY were tested for islet autoantibodies during the prospective follow-up, beginning at 9 months, 15 months, 24 months and, if negative, annually thereafter. DAISY subjects were tested for GADA, IAA, and IA-2A, which are performed in the Clinical Immunology Laboratory at the Barbara Davis Center using radio-immunoassays as previously described (15). Since January 2010, GADA and IA-2A have been measured with a harmonized assay (16). Positive antibody tests are confirmed by blinded quality control. If positive for any of these three antibodies, subjects are tested more frequently (every 3–6 months). After the identification of zinc transporter 8 (ZnT8) as a type 1 diabetes-associated antigen (17), all subjects who had ever been antibody positive were retrospectively tested for ZnT8 antibodies, and this is now done prospectively. ZnT8A antibodies were measured in the Clinical Immunology Laboratory at the Barbara Davis Center, as previously described (17,18). Study subjects were considered persistently islet autoantibody positive if they had at least two confirmed positive samples that were not due to maternal islet autoantibody transfer or if they had one confirmed positive sample and developed diabetes prior to the next sample collection. Subjects were considered multiple antibody positive when they were persistently antibody positive and had tested positive for more than one autoantibody. Type 1 diabetes was diagnosed according to ADA criteria. Time to development of type 1 diabetes was defined as time from birth to type 1 diabetes diagnosis.

Genotyping

Typing for HLA class II alleles at HLA-DRB1, HLA-DQA1 and HLA-DQB1 was previously described (19,20). In the DAISY study, genotyping of 9 non-HLA SNPs was as follows: *INS*-23Hph1 (rs689), *PTPN22* R620W (rs2476601) polymorphisms were genotyped using a linear array (immobilized probe) method essentially as described in Mirel et al. The following SNPs were genotyped in the laboratory of Dr. Cisca Wijmenga using Illumina GoldenGate Beadexpress assays (veracode 48-plex): *IL2RA* (rs12251307) and *BACH2* (rs11755527). Taqman SNP genotyping assays (Applied Biosystems, Foster City, CA, USA) were utilized to obtain genotype information for *GLIS3* (rs7020673), *GSDM* (rs2290400), *ERBB3* (rs2292239) and *IL27* (rs4788084) as described previously (19). All but one of the SNPs used in the Winkler *et al.* risk model (10) were present in the DAISY data set. The SNP for *IL2RA* measured in the DAISY cohort differed from the SNP in the T1DGC study used for development of the Winkler *et al.* 10-factor risk model (10). The *IL2RA* SNP from the DAISY data set (rs12251307) is located in a intergenic region, while the SNP used in the Winkler *et al.* model (rs12722495) (10) is from an intron of *IL2RA*. The two SNPs were queried for linkage disequilibrium using the Broad Institute SNP Annotation and Proxy Search Pairwise LD tool on the 1000 Genomes Pilot 1 data set (21), which resulted in an R^2 of 0.543 and a D' of 0.843.

HLA risk genotypes were categorized as: 6 = DR3/DR4-DQ8; 5 = DR4-DQ8/DR4-DQ8; 4 = DR3/DR3; 3 = DR4-DQ8/ x; 2 = DR3/DRx; 1 = DRx/DRx (where x represents the non-DR3 and non-DR4-DQ8 alleles). For other SNPs, a score of 2 was given to persons homozygous for the susceptibility allele, 1 when heterozygous and 0 when homozygous for the non-susceptibility allele.

Statistical Analysis

Demographic characteristics were compared using chi-square analysis for categorical variables and Wilcoxon rank sum tests for continuous variables. Time of follow up was calculated from birth to age at last visit or diagnosis of T1D. The weights generated by previous analysis (10) were used to calculate risk scores in the DAISY children. Specifically, the risk score per patient i in DAISY was calculated as:

$$\text{risk score}_i = \hat{\beta}_0 + \sum_{j=1}^p \hat{\beta}_j \text{SNP}_{i,j}$$

using the weights $\hat{\beta}_j$ and the intercept $\hat{\beta}_0$ derived in [1] and p is the number of SNPs plus the HLA risk categories. The 9 non-HLA SNPs used in the model were: *PTPN22* R620W (rs2476601), *INS*-23Hph1 (rs689), *IL2RA* (rs12722495), *ERBB3* (rs2292239), *ORMDL3*/*GSDM* (rs2290400), *BACH2* (rs11755527), *IL27* (rs4788084), *GLIS3* (rs7020673), and *RNLS/C10orf59* (rs10509540) genes (10).

The discriminative power of the model was assessed using receiver operating characteristic (ROC) analysis and the area under the ROC curve (ROC AUC). Model refits on the DAISY GP and FDR group were performed using the above-mentioned multivariable logistic

regression model. Improvement in prediction by additional markers was quantified using the integrated discrimination improvement, IDI (22). Kaplan-Meier survival curves were obtained based on the resulting risk score and time from birth to development of type 1 diabetes. Individuals lost to follow-up were treated as censored data. Differences in survival curves were assessed using log-rank tests. All analyses were performed using R version 3.2.0 (23).

RESULTS

Using the 10-factor genetic risk model, risk scores were calculated for the 1,941 individuals from the DAISY cohort with SNP data available. The genetic risk score distributions from DAISY participants, both GP and FDR, who were T1D cases versus those who did not develop T1D are shown in Figure 1. For both GP and FDR, the children who developed T1D had a risk score distribution that was shifted to higher risk scores compared to those who did not progress to T1D. The FDR participants showed a much broader distribution of risk scores than the GP individuals, whose risk scores were more tightly clustered. Interestingly, amongst the FDR children, both T1D cases and controls showed lower risk score distributions relative to the GP group, which can be explained by the HLA selection of the GP group.

In order to examine the discriminative power of the 10-factor model in DAISY GP and FDR subjects, these two subgroups were examined separately using the receiver operating characteristic area under the curve (ROC AUC). Given the importance of HLA genotype, the ROC AUC was first calculated using HLA genotype as a sole predictor. A minimal risk model of 3 factors (HLA plus the top two weighted SNPs, *PTPN22* and *INS*) and the full 10-factor model were also examined for discrimination between T1D and non-T1D outcomes (Figure 2).

For both the GP and FDR groups, the probability of T1D outcome was calculated using HLA alone, the 3-factor model or the 10-factor model. In all cases, the mean risk score was higher those who did indeed develop T1D than in those who did not. Thus, the discrimination slope, an estimation of the difference in probabilities of outcome, was positive for all three risk score calculations in both GP and FDR (Figure 3). The improvement in prediction by new or additional information can be quantified using the integrated discrimination improvement, IDI, an estimation in the difference in discrimination slopes (22). Examination of the performance of the 10-factor model and the simpler 3-factor model relative to HLA risk group prediction, as well as to each other, was calculated in the DAISY GP and FDR groups. For the DAISY GP participants, both the 10-factor ($p=0.03$) and the 3-factor ($p=0.03$) models showed improvement over HLA alone (Figure 3A). Of note, in the GP group, comparison of the 10-factor model to the 3-factor model showed no significant improvement ($p=0.22$), indicating that the adding seven more SNPs provided no additional information compared to the more minimal 3-factor model. In the DAISY FDR group, the 10-factor model showed improvement over both HLA alone ($p=0.01$) and the 3-factor model ($p=0.02$). However, there was no significant difference between the 3-factor model and HLA alone ($p=0.12$, Figure 3B).

In order to demonstrate the impact of the various models on prediction, the three models were used to classify the GP (Table 1) and FDR (Table 2) study participants using a fixed sensitivity cutoff of 0.5. For the GP individuals, the 3-factor model was able to identify 12 true positives, compared to the HLA only model which did not identify any true positives at this cutoff. The 10-factor model showed no improvement in categorization over the 3-factor model, paralleling the findings from the reclassification index analysis. For the FDR individuals, there were no changes from HLA alone to the 3-factor model; however, the 10-factor model identified 54 additional true negatives by reducing the number of false positives, illustrating the superior performance of the 10-factor model demonstrated in the reclassification analysis.

The performance of the 10-factor risk model for predicting time to development of T1D was examined in the DAISY GP and FDR subgroups (Figure 4). Children from each group were stratified by risk score into the upper quintile, middle three quintiles and lower quintile. Kaplan-Meier analysis was used to determine cumulative risk of development of T1D over follow up from birth. The risk model showed significant discrimination for risk of development of T1D between the highest and lowest quintiles by risk score for both the GP (Figure 4a, $p=0.00006$) and FDR (Figure 4b, $p=0.00022$) groups. Of note, amongst the GP group, those with risk scores in the bottom quintile showed no incidence of T1D over the first 15 years (Figure 4B) while in the highest quintile, T1D-free survival probability at 15 years was 0.88 [95% CI: 0.93, 0.81]. For the DAISY FDR group, disease-free survival at 15 years in the highest vs. lowest quintiles by risk score was 0.87 [95% CI: 0.92, 0.80] vs. 0.98 [95% CI: 0.99, 0.92].

DISCUSSION

While we (10,19,24,25) and others (26–28) have previously examined the role of minor genetic susceptibility genes to improve prediction of autoimmunity and progression to T1D, this is the first study to evaluate the performance of a risk model derived from a group of first-degree relatives with T1D in a prospective cohort of individuals recruited from the general population as well as first-degree relatives. We have shown that the Winkler *et al.* 10-factor model (10) was able to effectively predict T1D outcome and showed improvement in classification over HLA class alone in both GP and FDR children.

A more minimalist 3-factor model using HLA type plus *PTPN22* and *INS* SNPs was able to improve outcome classification over HLA alone in GP, but did not reach significance for FDR children. The differences between the 3-factor model's performance in the FDR and GP subgroups may stem partly from the enrichment of the GP group for high-risk HLA types (Supplemental Figure 1). Amongst the enriched population, the role of *PTPN22* and *INS* in prediction of outcome becomes more pronounced. In contrast, in the more diversely distributed HLA background of the FDR group, HLA plays a more prominent role in prediction of outcome and therefore the 3-factor model does not provide enough new discriminating information to show improvement over HLA alone, but the 10 factor model does.

While the 10-factor model was significantly better at prediction of outcome than either HLA alone or the minimal 3-factor model in the FDR group, it showed no improvement over the 3-factor in the GP group. These findings may imply that children from the general population without a close family history of T1D may have a different profile of predictive minor genetic susceptibility genes from those with a T1D FDR. The 10-factor model was developed in the T1DGC dataset, a cohort composed of families with multiple members diagnosed with T1D. Of the 10 factors in the risk model, most of the SNPs with smaller effect sizes would not have been selected had the risk model been fit on either DAISY group, based on their performance in the model refitting (Supplemental Figure 2). Indeed, when features are ranked in DAISY GP and FDR separately and added in a stepwise fashion, the feature ranking algorithm selects a different subset of features for GP (*IL2RA* not selected) and FDR groups (*IL27* and *ORMDL3* not selected) (Supplemental Figure 4A and B). This could potentially reflect differences in populations. The DAISY study is composed of subjects from one geographic region. The T1DGC dataset, in contrast, was derived from multiple countries. Another potential effect is the role of age. The T1DGC dataset includes older adults with T1D among the parents of the family units, therefore including individuals who could have been diagnosed well into adulthood, while the DAISY population is younger. Only some of the risk SNPs identified by Winkler et al.'s 10-factor model have been identified in previous multivariate analysis of genetic risk markers in the DAISY cohort. Specifically, both *PTPN22* and *UBASH3A* (rs11203203 AA) were identified as being significantly associated with development of IA, while *GLIS3* and *IL2RA* showed borderline association (19). Additionally, *INS* and *UBASH3A* (rs11203203 AA) were significantly associated with progression from IA to T1D, while *PTPN22* showed borderline association (19).

We have shown that the 10-factor model derived by Winkler *et al.* (10) is also able to effectively discriminate between children who are likely to progress to T1D over the next 15 years and those who are not by survival analysis of longitudinal data. Of children from the DAISY cohort recruited from the general population, none of the children with a low risk score progressed to T1D over the follow up time, indicating that the stratification by risk score is useful in GP children as well as FDR children, as previously described (10). It is interesting to note that the GP children in the lowest risk quintile were largely from the lowest-risk HLA group (DRx/DRx, where x represents the non-DR3 and non-DR4-DQ8 alleles) (Supplemental Figure 3), underlining the importance of HLA genotype for prediction of T1D risk.

One limitation of this study is the difference between the *IL2RA* SNP used for the DAISY cohort from the SNP identified in the risk modeling from the T1DGC cohort. While these SNPs are associated by linkage disequilibrium, the difference may be enough to affect the performance of the risk model. While the performance of the risk model in GP versus FDR groups is intriguing, the enrichment of the GP group for moderate and high-risk HLA types may limit the generalizability to the population as a whole. Another limitation of this study is the characteristics of the non-T1D group as it represents a group already selected for T1D risk. Overall, the ROC AUCs in the DAISY cohort were lower than the values in the T1DGC data set and the German validation set described previously (0.87 and 0.84, respectively) (10), although ROC AUC would be expected to be lower in a validation set. Also, both of

the previously studied data sets contain a large group of control subjects who were from the general population and had only background risk for T1D. In contrast, the DAISY “control” groups consist of study subjects who were recruited due to their high-risk status as either a FDR or as having a high-risk HLA type and may still eventually progress to T1D in later life.

Although the majority of new-onset patients with T1D have no family history (1), the prevalence of T1D is relatively low in the general population, about 0.5% by age 20, use of these prediction models for general population screening would inevitably result in many false positives. As with the selection of the DAISY study population, an initial selection for higher-risk HLA types or first-degree relative status would improve the performance. Another strategy would be to apply the risk model with high sensitivity and then follow with a secondary screening. To date, this has meant serial islet autoantibody measurements. Future research may identify other layered screening strategies, for instance other prognostic markers in the blood, urine or microbiome. Alternately, identification of a constellation of pathogenic environmental exposures could yield an exposure risk assessment or score. A recent analysis found that the use of HLA with antibody measurement every 6 months until age 5 was not cost-effective if the outcome was prevention of DKA. However, less frequent autoantibody testing and/or advancements in laboratory technology enabling cost reduction in HLA and autoantibody testing to less than \$1 and \$0.03, respectively would render this strategy cost-effective (29).

In conclusion, we have shown that a 10-factor risk model, previously validated in first-degree relatives of T1D patients, is also effective in prediction of T1D outcome in a cohort of children including both those with and without family history of T1D. While the 10-factor model is superior to a more minimal 3-factor model including only HLA, *PTPN22* and *INS* in FDR children, its performance does not differ from the 3-factor model in GP children. Children from the general population with low risk scores, as determined by the 10-factor model, did not progress to T1D over up to 20 years of follow-up. This may indicate the utility of a genetic risk model for selection of children for future prevention trials or population screening programs. Further, the differences between model performance in the GP relative to the FDR groups indicate that there may be important genetic risk factors to be discovered by differentiating these populations in future analysis. The identification of subgroup differences in the performance of risk modeling may lead to increased insight into genetic factors which play a role in epidemiologic variation across regions and ethnic groups as well as first-degree relatives compared to those without family history of T1D.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium (T1DGC), a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and the Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. The UK case series collection was

additionally funded by the JDRF and Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Centre, at the Cambridge Institute for Medical Research, UK (CIMR), which is in receipt of a Wellcome Trust Strategic Award (079895). The SNP data from the T1DGC study were supplied by the NIDDK Central Repositories. This manuscript was not prepared in collaboration with Investigators of the T1DGC study and does not necessarily reflect the opinions or views of the T1DGC study, the NIDDK Central Repositories or the study sponsors.

This work was supported by JDRF grants 17-2013-535, 11-2010-206, 2-SRA-2015-13-Q-R and the National Institutes of Health grants R01 DK32083, DK32493, DK049654, and 5K12DK094712 and Diabetes Endocrinology Research Center (P30 DK57516). Also supported by Helmsley Charitable Trust Grant 2015PG-T1D072, Helmholtz HIRG-0018, and German BMBF-funded iMed Helmholtz Alliance.

Abbreviations

ROC AUC	Receiver operating characteristic area under the curve
SNP	Single-nucleotide polymorphism
T1DGC	Type 1 Diabetes Genetics Consortium
T1D	Type 1 Diabetes
IA	islet autoimmunity
FDR	first-degree relative
GP	general population

References

1. Michels A, Zhang L, Khadra A, Kushner JA, Redondo MJ, Pietropaolo M. Prediction and Prevention of Type 1 Diabetes: Update on Success of Prediction and Struggles at Prevention. *Pediatr Diabetes*. 2015; 16:465–84. [PubMed: 26202050]
2. Insel RA, Dunne JL, Ziegler A-G. General population screening for type 1 diabetes: has its time come? *Curr Opin Endocrinol Diabetes Obes*. 2015; 22:270–6. [PubMed: 26087338]
3. Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, Erlich HA. The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am J Hum Genet*. 1996; 59:1134–48. [PubMed: 8900244]
4. Ilonen J, Kiviniemi M, Lempainen J, et al. Genetic susceptibility to type 1 diabetes in childhood – estimation of HLA class II associated disease risk and class II effect in various phases of islet autoimmunity. *Pediatr Diabetes*. 2016; 17:8–16. [PubMed: 27411431]
5. Concannon P, Erlich HA, Julier C, et al. Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. *Diabetes*. 2005; 54:2995–3001. [PubMed: 16186404]
6. Cooper JD, Smyth DJ, Smiles AM, et al. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat Genet*. 2008; 40:1399–401. [PubMed: 18978792]
7. 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. [PubMed: 19430480]
8. Concannon P, Chen W-M, Julier C, et al. Genome-wide scan for linkage to type 1 diabetes in 2,496 multiplex families from the Type 1 Diabetes Genetics Consortium. *Diabetes*. 2009; 58:1018–22. [PubMed: 19136655]
9. Burren OS, Adlem EC, Achuthan P, Christensen M, Coulson RMR, Todd JA. T1DBase: update 2011, organization and presentation of large-scale data sets for type 1 diabetes research. *Nucleic Acids Res*. 2011; 39:D997–1001. [PubMed: 20937630]
10. Winkler C, Krumsiek J, Buettner F, et al. Feature ranking of type 1 diabetes susceptibility genes improves prediction of type 1 diabetes. *Diabetologia*. 2014; 57:2521–9. [PubMed: 25186292]

11. Thümer L, Adler K, Bonifacio E, et al. German new onset diabetes in the young incident cohort study: DiMelli study design and first-year results. *Rev Diabet Stud.* 2010; 7:202–8. [PubMed: 21409312]
12. Ziegler AG, Hummel M, Schenker M, Bonifacio E. Autoantibody appearance and risk for development of childhood diabetes in offspring of parents with type 1 diabetes: the 2-year analysis of the German BABYDIAB Study. *Diabetes.* 1999 Mar 1;48:460–8. [PubMed: 10078544]
13. Rewers M, Norris JM, Eisenbarth GS, et al. Beta-Cell Autoantibodies in Infants and Toddlers without IDDM Relatives: Diabetes Autoimmunity Study in the Young (DAISY). *Journal of Autoimmunity.* 1996; 9:405–10. [PubMed: 8816978]
14. Rewers M, Bugawan TL, Norris JM, et al. Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). *Diabetologia.* 1996; 39:807–12. [PubMed: 8817105]
15. Yu L, Rewers M, Gianani R, et al. Antiislet autoantibodies usually develop sequentially rather than simultaneously. *J Clin Endocrinol Metab.* 1996; 81:4264–7. [PubMed: 8954025]
16. Bonifacio E, Yu L, Williams AK, et al. Harmonization of glutamic acid decarboxylase and islet antigen-2 autoantibody assays for national institute of diabetes and digestive and kidney diseases consortia. *J Clin Endocrinol Metab.* 2010; 95:3360–7. [PubMed: 20444913]
17. Wenzlau JM, Juhl K, Yu L, et al. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc Natl Acad Sci U S A.* 2007; 104:17040–5. [PubMed: 17942684]
18. Wenzlau JM, Moua O, Sarkar SA, et al. SLc30A8 Is a Major Target of Humoral Autoimmunity in Type 1 Diabetes and a Predictive Marker in Prediabetes. *Annals of the New York Academy of Sciences.* 2008; 1150:256–9. [PubMed: 19120307]
19. Steck AK, Dong F, Wong R, et al. Improving prediction of type 1 diabetes by testing non-HLA genetic variants in addition to HLA markers. *Pediatr Diabetes.* 2014; 15:355–62. [PubMed: 25075402]
20. Ziegler A-G, Bonifacio E, Group the B-BS. Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. *Diabetologia.* 2012; 55:1937–43. [PubMed: 22289814]
21. SNAP Pairwise LD [Internet]. [cited 2015 Oct 30]. Available from: <https://www.broadinstitute.org/mpg/snap/ldsearchpw.php>
22. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *StatMed.* 2008; 30:157–72.
23. Core Team R. A language and environment for statistical computing. Vienna, Austria 2014 [Internet] 2015 [cited 2016 Apr 7]. Available from: <http://scholar.google.com/scholar?cluster=7833233836880142876&hl=en&oi=scholar>
24. Winkler C, Krumsiek J, Lempainen J, et al. A strategy for combining minor genetic susceptibility genes to improve prediction of disease in type 1 diabetes. *Genes Immun.* 2012; 13:549–55. [PubMed: 22932816]
25. Achenbach P, Hummel M, Thümer L, Boerschmann H, Höfelmann D, Ziegler AG. Characteristics of rapid vs slow progression to type 1 diabetes in multiple islet autoantibody-positive children. *Diabetologia.* 2013; 56:1615–22. [PubMed: 23539116]
26. Lempainen J, Härkönen T, Laine A, Knip M, Ilonen J, Finnish Pediatric Diabetes Register. Associations of polymorphisms in non-HLA loci with autoantibodies at the diagnosis of type 1 diabetes: INS and IKZF4 associate with insulin autoantibodies. *Pediatr Diabetes.* 2013; 14:490–6. [PubMed: 23721563]
27. Lempainen J, Laine A-P, Hammais A, et al. Non-HLA gene effects on the disease process of type 1 diabetes: From HLA susceptibility to overt disease. *J Autoimmun.* 2015; 61:45–53. [PubMed: 26074154]
28. Lempainen J, Hermann R, Veijola R, Simell O, Knip M, Ilonen J. Effect of the PTPN22 and INS risk genotypes on the progression to clinical type 1 diabetes after the initiation of β -cell autoimmunity. *Diabetes.* 2012; 61:963–6. [PubMed: 22357962]
29. Meehan C, Fout B, Ashcraft J, Schatz DA, Haller MJ. Screening for T1D risk to reduce DKA is not economically viable. *Pediatr Diabetes.* 2015; 16:565–72. [PubMed: 26392298]

Figure 1. SNP Risk Score Distribution for T1D vs. non-T1D controls in DAISY subjects recruited from first-degree relatives of T1D patients (FDR) or the general population (GP)

The risk score (x-axis) is calculated using the 10-factor weighted risk model. A shift to the right indicates a higher risk score. Density (y-axis) represents the portion of the subjects with a risk score at each level. Red lines represent children who have a first-degree relative with T1D (FDR). Black lines represented children recruited from the general population (GP). In each case, the risk score distribution of the T1D children is represented by dashed lines, while the risk score distribution of the non-T1D children is represented by solid lines.

Figure 2.

Receiver operator curve (ROC) for prediction of T1D outcome using HLA only, 3-factor model (HLA plus *PTPN22* and *INS* SNPs) and 10-factor model for (a) DAISY first degree relatives (FDR), (b) DAISY subjects from the general population (GP).

Figure 3. Comparison of classification performance using HLA prediction alone versus 3-factor or 10-factor models in DAISY subjects recruited from (a) general population (GP) or (b) first-degree relatives of T1D patients (FDR)

Discrimination slope measures (difference in mean probability of T1D outcome) are given for risk score calculated by HLA, 3-factor or 10-factor model. Reclassification index (IDI) measures the improvement in classification performance for one model relative to another. * $p < 0.05$.

Figure 4. Disease free survival without T1D for DAISY GP and FDR children, separated by highest risk score quintile (red), middle three risk quintiles (gray) and lowest risk quintile (blue) as determined using the 10-factor risk model

Time measured from birth to event. (a) General population subjects in DAISY cohort (highest vs lowest quintile, p-value = 0.00006) (b) First-degree relatives in DAISY cohort, (highest vs. lowest quintile, p-value = 0.00022).

Reclassification in DAISY general population screened participants for a sensitivity cutoff of 0.5. ND, nondiabetic; T1D, type 1 diabetes. (n=1097).

Table 1

HLA Only				3-Factor Model				10-Factor Model			
Model Prediction		True Outcome		Model Prediction		True Outcome		Model Prediction		True Outcome	
		ND	T1D			ND	T1D			ND	T1D
ND	1075	22		ND	932	10		ND	940	11	
T1D	0	0		T1D	143	12		T1D	135	11	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2
Reclassification in DAISY first-degree relatives for a sensitivity cutoff of 0.5. ND, nondiabetic; T1D, type 1 diabetes. (n=844).

HLA Only			3-Factor Model			10-Factor Model		
Model Prediction	True Outcome		Model Prediction	True Outcome		Model Prediction	True Outcome	
	ND	T1D		ND	T1D		ND	T1D
ND	674	18	ND	677	19	ND	731	18
T1D	134	18	T1D	131	17	T1D	77	18