Received: 17 December 2021 | Accepted: 3 January 2022

DOI: 10.1111/pedi.13310

PATHOPHYSIOLOGY AND PREVENTION

Improved genetic risk scoring algorithm for type 1 diabetes prediction

Hui-Qi Qu[1] | Jingchun Qu[1] | Joseph Glessner[1,2,3] | Yichuan Liu[1] | Frank Mentch[1] | Xiao Chang[1] | Michael March[1] | Jin Li[4] | Jeffrey D. Roizen[2] | John J. Connolly[1] | Patrick Sleiman[1,2,3] | Hakon Hakonarson[1,2,3,5]

[1]The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

[2]Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[3]Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

[4]Department of Cell Biology, Tianjin Medical University, Tianjin, China

[5]Division of Pulmonary Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

Correspondence
Hakon Hakonarson, Center for Applied Genomics, 3615 Civic Center Blvd, Abramson Building, Philadelphia, PA 19104, USA.
Email: hakonarson@email.chop.edu

Funding information
The study was supported by Institutional Development Funds from the Children's Hospital of Philadelphia to the Center for Applied Genomics and The Children's Hospital of Philadelphia Endowed Chair in Genomic Research to H.H. The eMERGE Network was initiated and funded by the NHGRI through the following grants for Phase 4: U01HG011175 (Children's Hospital of Philadelphia).

Abstract

Background: Precise risk prediction of type 1 diabetes (T1D) facilitates early intervention and identification of risk factors prior to irreversible beta-islet cell destruction, and can significantly improve T1D prevention and clinical care. Sharp et al. developed a genetic risk scoring (GRS) system for T1D (T1D-GRS2) capable of predicting T1D risk in children of European ancestry. The T1D-GRS2 was developed on the basis of causal genetic variants, thus may be applicable to minor populations, while a trans-ethnic GRS for T1D may avoid the exacerbation of health disparities due to the lack of genomic information in minorities.

Methods: Here, we describe a T1D-GRS2 calculator validated in two independent cohorts, including African American children and European American children. Participants were recruited by the Center for Applied Genomics at the Children's Hospital of Philadelphia.

Results: It demonstrates that GRS2 is applicable to the T1D risk prediction in the AA cohort, while population-specific thresholds are needed for different populations.

Conclusions: The study highlights the potential to further improve T1D-GRS2 performance with the inclusion of additional genetic markers.

KEYWORDS
eMERGE, genetic risk score, PRS, screening, type 1 diabetes

1 | INTRODUCTION

Type 1 diabetes (T1D), which is caused by autoimmune destruction of pancreatic β-cells, is most prevalent in individuals of European ancestry, but also presents a serious burden among individuals of African ancestry.[1] Once diagnosed, the disease progress is irreversible, and patients will require lifelong insulin therapy. Precise T1D risk prediction is required to support preventative studies where intervention in advance of pancreatic β-cells destruction has enormous therapeutic potential. The genetic risk scoring (GRS) system for T1D, developed by Sharp et al., T1D-GRS2, uses 67 SNPs from known autoimmune loci, and is demonstrably capable of predicting T1D in children of European ancestry.[2] To further explore the clinical potential of T1D-GRS2, we developed a computer code written in Python, which enables the calculation of T1D-GRS2 based on genotyping data generated by Illumina SNP arrays. By assessing performance in children of African as well as European ancestry, this study aims to increase the precision and broaden the scope of T1D-GRS2, both as a tool with immediate clinical application and driver of future translational studies.

2 | RESEARCH DESIGN AND METHODS

2.1 | Computer code for T1D-GRS2 scoring

The T1D-GRS2 scoring was developed by Sharp et al.[2] with 67 SNPs from known autoimmune loci, including 35 SNPs from the human leukocyte antigen (*HLA*) region, and 32 SNPs from 31 non-*HLA* T1D susceptibility loci (Table S1). Our code takes the input of the genetic information from a set of PLINK files containing the 67 GRS2 SNPs. The original genotyping data are from the Illumina Genotyping BeadChip with at least 550,000 SNPs genotyped. With the original genotyping data, genome-wide imputation is performed using the TOPMed Imputation Server (https://imputation.biodatacatalyst.nhlbi.nih.gov/#!) harboring the TOPMed (Version R2 on GRC38) Reference Panel. Of the 67 GRS2 SNPs, 65 can be imputed with the quality filter $R^2 > 0.3$ (average $R^2 = 0.885 \pm 0.152$). Two SNPs, rs144530872 corresponding to the *HLA-A*2902* allele, and rs540653847 corresponding to the *HLA_B_3906* allele, respectively, need to be imputed using the SNP2HLA software.[3] In addition, the user has the option to include four more SNPs to calculate GRS2′ (*HLA-DQ*: rs9273363; non-*HLA*: rs926169, rs10788599, and rs56380902) validated for T1D association in both European Americans (EAs) and African Americans (AAs), with their roles in T1D risk prediction in AAs demonstrated previously.[4] Genotyping information for these four SNPs is available in the TOPMed imputation results. The publicly available code for calculating the T1D GRS2 score was written in Bash and Python, and is available on GitHub (https://github.com/huiqi-qu/GRS2). The collected datasheet on the T1D GRS2 by Marc Vaudel (https://github.com/mvaudel/diabetesRiskScores/blob/master/resources/scores/T1D-GRS2) is referred to in our study with corrections.

2.2 | Assessment of the performance of T1D-GRS2 in EA and AA children

Subjects: Two population samples were investigated in this study, including (1) 168 T1D AA cases versus 1366 non-diabetes AA controls; (2) 361 EA T1D cases versus 1943 non-diabetes EA controls (Table 1). Both cohorts were recruited between 2006 and 2020 by the Center for Applied Genomics (CAG) at the Children's Hospital of Philadelphia (CHOP), which has established a large pediatric biobank coupled to comprehensive electronic medical record. Each individual was genotyped with an Illumina Genotyping BeadChip with at least 550,000 SNPs genotyped.

2.3 | Data analysis

The population ancestry of each individual was both self-reported and validated by principal component analysis with genome-wide SNP markers. The GRS2 and the GRS2′ (with four additional SNPs associated with T1D in African Population) were calculated for each subject. The GRS2 and the GRS2′ were compared between different groups with independent *t* test using IBM SPSS Statistics Version 23 software. The GRS scores were assessed for their predictive performance in each population by the area under the ROC curve (AUC).

3 | RESULTS

3.1 | Lower GRS2 and GRS2′ in the AA population

As shown in Tables 1 and 2, a significant difference was detected between T1D cases and controls for both GRS2 and GRS2′ in both AA and EA cohorts, which suggests the feasibility of T1D prediction by GRS2 and/or GRS2′. T1D AA cases had lower GRS2 and GRS2′ scores than the EA cases, and AA controls had lower GRS2 and GRS2′ scores than the EA controls. These findings suggest population-specific thresholds of GRS2 and GRS2′ are needed for AA and EA populations.

3.2 | ROC analysis of GRS2 and GRS2′ in AA and EA populations

Consequently, we performed ROC analysis in both AA and EA populations. The GRS2 had an AUC (95% CI) of 0.807 (0.779, 0.835) to predict T1D in the CAG AA cohort, compared to AUC (95% CI) of 0.823

TABLE 1 General information of the two population samples

Population		Cases	Controls	*p* value
AA	*N*	168	1366	
	Male	89 (53.0%)	662 (48.5%)	0.269
	Female	79 (47.0%)	704 (51.5%)	
	Age	13.6 ± 5.1	14.3 ± 3.2	0.012
	GRS2	8.13 ± 2.33	5.24 ± 2.32	1.23×10^{-48}
	GRS2′	9.77 ± 3.31	5.70 ± 2.71	2.25×10^{-65}
EA	*N*	361	1943	
	Male	184 (51.0%)	990 (51.0%)	0.995
	Female	177 (49.0%)	953 (49.0%)	
	Age	12.7 ± 4.5	13.5 ± 3.4	9.15×10^{-5}
	GRS2	10.52 ± 2.20	7.41 ± 2.53	2.37×10^{-96}
	GRS2′	12.70 ± 3.04	8.37 ± 3.10	1.17×10^{-117}

Abbreviations: AA, African American; EA, European American; GRS, genetic risk scoring.

TABLE 2 *p* values of comparisons between different groups

	AA Cases	AA Controls	EA Cases	EA Controls
GRS				
AA cases	-	1.23×10^{-48}	3.62×10^{-27}	-
AA controls	1.23×10^{-48}	-	-	8.86×10^{-127}
EA cases	3.62×10^{-27}	-	-	2.37×10^{-96}
EA controls	-	8.86×10^{-127}	2.37×10^{-96}	-
GRS2				
AA cases	-	2.25×10^{-65}	8.98×10^{-22}	-
AA controls	2.25×10^{-65}	-	-	8.69×10^{-133}
EA cases	8.98×10^{-22}	-	-	1.17×10^{-117}
EA controls	-	8.69×10^{-133}	1.17×10^{-117}	-

Abbreviations: AA, African American; EA, European American; GRS, genetic risk scoring.

(0.804, 0.842) in the CAG EA cohort. The prediction of T1D has a sensitivity of 0.613 and a specificity of 0.834 with the maximum Matthews correlation coefficient at the cutoff of GRS2 at 7.43 in AA (Table S2), compared to a sensitivity of 0.623 and a specificity of 0.833 at the cutoff of GRS2 = 9.75 in EA (Table S3). These results suggest that the T1D GRS2 is applicable to both AA and EA populations.

The GRS2 performance improved in both AA and EA by including the four additional SNPs of T1D association in both African and European populations (*HLA-DQ*: rs9273363; non-*HLA*: rs926169, rs10788599, and rs56380902). The GRS2′ AUC (95% CI) = 0.826 (0.800, 0.852) improved in the CAG AA cohort (Table S4), and the AUC (95% CI) = 0.839 (0.822, 0.857) also improved in the CAG EA cohort (Table S5). With a specificity of 0.834, the GRS2′ has a sensitivity of 0.643 at the cutoff of GRS2′ = 8.19 in AA; a sensitivity of 0.690 at the cutoff of GRS2′ = 11.31 in EA.

4 | DISCUSSION

This study demonstrated that GRS2 is applicable to the T1D risk prediction in the AA cohort, though with a lower AUC score. The AA-specific GRS has been developed by Onengut-Gumuscu et al. with demonstrated performance.[4] In the meantime, a trans-ethnic GRS for T1D may avoid the exacerbation of health disparities due to the lack of genomic information in minorities.[5] The T1D-GRS2 was developed on the basis of causal genetic variants, thus may be applicable to minor populations. Both the AA and EA individuals were recruited at the CHOP. As a limitation of this study, the sample size of recruited T1D individuals of other ethnicities was too limited for this study. Additionally, due to the required sample size and the long time period for patient recruitment, patients recruited in earlier time were mainly based on clinical diagnosis of T1D, without the results of the four T1D autoantibodies, that is, islet cell antibodies (ICA, against cytoplasmic proteins in the β-cell), antibodies to glutamic acid decarboxylase (GAD-65), insulin autoantibodies (IAA), and IA-2A, antibodies to protein tyrosine phosphatase. However, the potential of mixed non-autoimmune pediatric diabetes might decrease the power of this study and cause bias toward false-negative results, thus will not make our demonstration of the performance of T1D-GRS2 less convincing.

A lower GRS2 score is observed in the AA cohort, highlighting the importance of using population-specific reference values in the GRS2 for T1D risk prediction in AAs. The AA cohort has a lower prevalence of T1D than the EA cohort (0.57/1000 vs. 2.0/1000).[6] The lower GRS2 score observed in the AA cohort partially represents the lower genetic risk of T1D in the AA population.

As shown by this study, the performance of GRS2 in the AA population could be further improved by including four additional SNP markers associated with AA T1D. The marker rs926169 is from the *CTLA4* region. *CTLA4* encodes cytotoxic T-lymphocyte-associated protein 4, which transmits inhibitory signals to attenuate T-cell activation.[7] One SNP, rs3087243, has been included in the GRS2. However, a previous study has shown that more than one association signals have been seen in the *CTLA4* region.[8] The additional marker, rs926169, has a low linkage disequilibrium (LD) $r^2 = 0.162$, $D' = 0.940$ in the AA population, and $r^2 = 0.358$, $D' = 0.816$ in the EA population.

A previous study showed that the additional marker, rs9273363, at the *HLA DR-DQ* region maps to a potential enhancer region of *HLA-DQB1*, and is associated with T1D by tagging the *HLA DQB1*03:02* haplotype.[9] The GRS2 scoring system includes 35 SNPs from the *HLA* region. A SNP, rs9275490, is included in GRS2 to tag the *DQA1**03:0X-*DQB1**03:02 haplotype. However, the two SNPs have a low r^2 of 0.253, with $D' = 0.987$ in the AA cohort, and a low r^2 of 0.308, $D' = 1$ in the EA population.

The marker rs10788599 is from the renalase, FAD dependent amine oxidase gene (*RNLS*) region. *RNLS* may contribute to T1D genetic susceptibility by its role in JAK–STAT signaling in immune-mediated diseases by activating STAT3.[10] The SNP rs60888743 has been included in the GRS2. As shown by this data, the two SNPs have a low LD r^2 of 0.007, with $D' = 0.148$ in the AA cohort, and r^2 of 0.083, $D' = 0.387$ in the EA cohort.

The SNP rs56380902 maps to the gasdermin B gene (*GSDMB*) locus. This locus at chr17q21.1 was not covered in the GRS2 scoring system. *GSDMB* encodes a gasdermin-domain containing protein that is involved in T cell-mediated cytotoxicity by inducing pyroptosis.[11] The association of *GSDMB* and T1D has been validated in both European[12–14] and African populations.[4] Besides T1D, *GSDMB* has reported associations with other autoimmune diseases, such as

rheumatoid arthritis,[15] and autoinflammatory disease, such as asthma.[16] This locus has not been covered in the GRS2 system.

At this time, the T1D-GRS2 calculation code developed in this study takes input of the genotyping imputation results. Due to the automated TOPMED imputation requiring elevated privileges, the entirety of a GRS calculation pipeline from the original genotyping BeadChip data cannot be run in one go on a high performance computing cluster. For this purpose, the T1D-GRS2 calculation based on a cloud computing platform, for example, Amazon Web Services, to enable its clinical application, is under plan.

In conclusion, the results of this study are twofold. This study demonstrates the performance of the T1D-GRS2 calculator in both AA and EA cohorts, which implies that GRS2 may be applicable to individuals of other ethnicities or mixed ethnicity. On the other hand, the results of this study demonstrate that GRS2 improves in both AA and EA populations with the inclusion of four additional targeted SNPs, *HLA-DQ*: rs9273363; non-*HLA*: rs926169, rs10788599, and rs56380902 (noted by GRS2′). In addition, the lower GRS2 scores observed in AA individuals highlight population-specific reference values in GRS2 with consequences for T1D risk prediction in the AA population. The performance, as well as population-specific reference values, of the T1D-GRS2 in other ethnicities warrants for further study.

ACKNOWLEDGMENTS

Dr Hakon Hakonarson is the guarantor of this work. He 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.

AUTHOR CONTRIBUTIONS

H.Q.Q., J.C., J.G., Y.L., P.S., J.J.C. and H.H. performed the research. H.Q.Q., J.C. and H.H. designed the research study. F.M., X.C., M.M., J.L., and J.D.R. contributed essential reagents or tools. H.Q.Q., J.C. and J.G. analysed the data. H.Q.Q., J.C., J.J.C. and H.H. wrote the paper.

CONFLICT OF INTEREST

The authors declare no competing interests.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/pedi.13310.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

This study was approved by the Children's Hospital of Philadelphia (CHOP) Institutional Review Board (IRB). Informed consent was obtained from all subjects or, if subjects are under 18, from a parent and/or legal guardian with assent from the child if 7 years or older.

ORCID

Hui-Qi Qu https://orcid.org/0000-0001-9317-4488

REFERENCES

1. Mayer-Davis EJ, Beyer J, Bell RA, et al. Diabetes in African American youth: prevalence, incidence, and clinical characteristics: the SEARCH for Diabetes in Youth Study. *Diabetes Care*. 2009;32(suppl 2):S112-S122.
2. Sharp SA, Rich SS, Wood AR, et al. Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. *Diabetes Care*. 2019;42(2):200-207.
3. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One*. 2013;8(6):e64683.
4. Onengut-Gumuscu S, Chen W-M, Robertson CC, et al. Type 1 diabetes risk in African-ancestry participants and utility of an ancestry-specific genetic risk score. *Diabetes Care*. 2019;42(3):406-415.
5. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584-591.
6. Maahs DM, West NA, Lawrence JM, Mayer-Davis EJ. Epidemiology of type 1 diabetes. *Endocrinol Metab Clin*. 2010;39(3):481-497.
7. Ostrov DA, Shi W, Schwartz J-CD, Almo SC, Nathenson SG. Structure of murine CTLA-4 and its role in modulating T cell responsiveness. *Science*. 2000;290(5492):816-819.
8. Qu H, Bradfield J, Grant S, Hakonarson H, Polychronakos C. Remapping the type I diabetes association of the CTLA4 locus. *Genes Immun*. 2009;10(1):S27-S32.
9. Inshaw JRJ, Walker NM, Wallace C, Bottolo L, Todd JA. The chromosome 6q22.33 region is associated with age at diagnosis of type 1 diabetes and disease risk in those diagnosed under 5 years of age. *Diabetologia*. 2018;61(1):147-157.
10. Onengut-Gumuscu S, Vogler N, Faidas M, Pickin RR, Gersz E, Rich SS. Functional evaluation of RNLS, a gene harboring risk variants for type 1 diabetes in European and African Ancestry Subjects. *Am Diabetes Assoc*. 2018;67(Supplement 1):1702-P.
11. Zhou Z, He H, Wang K, et al. Granzyme A from cytotoxic lymphocytes cleaves GSDMB to trigger pyroptosis in target cells. *Science*. 2020;368(6494):eaaz7548.
12. 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(6):703-707.
13. Onengut-Gumuscu S, Chen W-M, Burren O, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet*. 2015;47(4):381-386.
14. Plagnol V, Howson JM, Smyth DJ, et al. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. *PLoS Genet*. 2011;7(8):e1002216.
15. Eyre S, Bowes J, Diogo D, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet*. 2012;44(12):1336-1340.
16. Moffatt MF, Kabesch M, Liang L, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*. 2007;448(7152):470-473.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Qu H-Q, Qu J, Glessner J, et al. Improved genetic risk scoring algorithm for type 1 diabetes prediction. *Pediatr Diabetes*. 2022;23(3):320-323. doi:10.1111/pedi.13310