



Improving the Prediction of Type 1 Diabetes Across Ancestries

Diabetes Care 2022;45:e48–e50 | <https://doi.org/10.2337/dc21-1254>

John S. Kaddis,¹ Daniel J. Perry,²
 Anh Nguyet Vu,¹ Stephen S. Rich,³
 Mark A. Atkinson,^{2,4}
 Desmond A. Schatz,⁴
 Bart O. Roep,⁵ and Todd M. Brusko^{2,4}

Prediction of type 1 diabetes (T1D) is essential for disease prevention and early interventional therapies. Genetic risk score (GRS) models developed to predict T1D have largely been determined using Eurocentric populations and applied to all individuals, regardless of race, ethnicity, or ancestry (1). Recent T1D incidence data suggest significant increases of new cases in Black, Hispanic, and Asian/Pacific Islander populations, emphasizing the need for personalized assessment strategies (2). Here, we demonstrate that a frequently used GRS model fails to recognize T1D in individuals of African descent (AFR); however, these same individuals can be identified using an AFR-derived model. We further illustrate the feasibility of determining genetic ancestry and show that it conflicts with self-reported ethnicity.

To demonstrate the application of ancestry-inclusive GRS models, we examined a cohort ($n = 377$) of unrelated individuals from the Network for Pancreatic Organ Donors with Diabetes (nPOD) program; this is a widely studied biorepository established to further our understanding of the pathological features of T1D. Our strategy consisted of three phases. First, we determined the genetic admixture of each individual, using the 1000 Genomes phase 3 data set to train our model, and compared it to self-reported ethnicity; here, race and ethnicity are defined as social

constructs of human variability, whereas ancestry is based on genetic DNA data. Second, single-nucleotide polymorphism (SNP)-based GRSs were calculated using European (EUR) (30 SNPs) and AFR (7 SNPs) cohort-derived models, as described in their original articles, and termed EUR GRS and AFR GRS, respectively (1,3). Of the 377 examined, those with a single ancestry proportion of ≥ 0.88 that clustered together were considered, with only ancestry groups having both control individuals without diabetes and case individuals with T1D included in the GRS analysis. This resulted in 207 individuals with AFR (11 control, 6 case), admixed American (AMR) (13 control, 6 case), and EUR (84 control, 87 case) ancestry. Finally, we investigated the differences in GRS models. Both of these models attempt to estimate the contribution of genetic factors in discriminating and/or predicting T1D. Detailed methods, including statistical analysis, data files, and code, for all three phases of this approach are available on Github through Zenodo (<https://doi.org/10.5281/zenodo.4944879>).

Disease risk models that do not account for the diversity in populations hinder the comparability between studies and may even increase health disparities (4). However, determining the genetic ancestry of individuals is feasible irrespective of unreliable, incomplete, or

missing self-reported race or ethnicity data. To illustrate this, the proportions of AFR, EUR, AMR, East Asian (EAS), and South Asian (SAS) were determined in the nPOD cohort, grouped based on the five major categories of the 1000 Genomes project. Each person was represented by a single ancestry (proportion ≥ 0.88), as an admixture, or had approximately equal contributions from two or more (denoted MIX) ancestries (Fig. 1A and B). Overall, genetic ancestries were 80% concordant with self-stated ethnicities, whereas 14% were partially concordant and 6% discordant (Fig. 1C). Concordance between genetic ancestry and self-reported ethnicity was highest in EAS ($n = 5/5$, 100%) and EUR ($n = 234/261$, 90%), low in AFR ($n = 30/59$, 51%), and absent from MIX ($n = 0/6$, 0%) groups. These results point to the characterization of genetic ancestry as a more reliable measure than self-reported ethnicity when examining potential genetic contributions to disease, particularly in heterogeneous populations.

This analysis revealed that only the AFR GRS was able to distinguish between individuals without diabetes and individuals with T1D in all three populations examined (Fig. 1D and E). Specifically, there was a statistically significant difference in the AFR GRS individuals without diabetes and individuals with T1D within the EUR ($\Delta = 3.177$; 95% confidence limits [CL] 2.525, 3.402; $P <$

¹Department of Diabetes and Cancer Discovery Science, Arthur Riggs Diabetes and Metabolism Research Institute, Beckman Research Institute, City of Hope, Duarte, CA

²Department of Pathology, Immunology and Laboratory Medicine, College of Medicine, Diabetes Institute, University of Florida, Gainesville, FL

³Center for Public Health Genomics, University of Virginia, Charlottesville, VA

⁴Department of Pediatrics, College of Medicine, Diabetes Institute, University of Florida, Gainesville, FL

⁵Department of Diabetes Immunology, Arthur Riggs Diabetes and Metabolism Research Institute, Beckman Research Institute, City of Hope, Duarte, CA

Corresponding author: John S. Kaddis, jkaddis@coh.org, or Todd M. Brusko, tbrusko@ufl.edu

Received 16 June 2021 and accepted 9 December 2021

© 2022 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/journals/pages/license>.

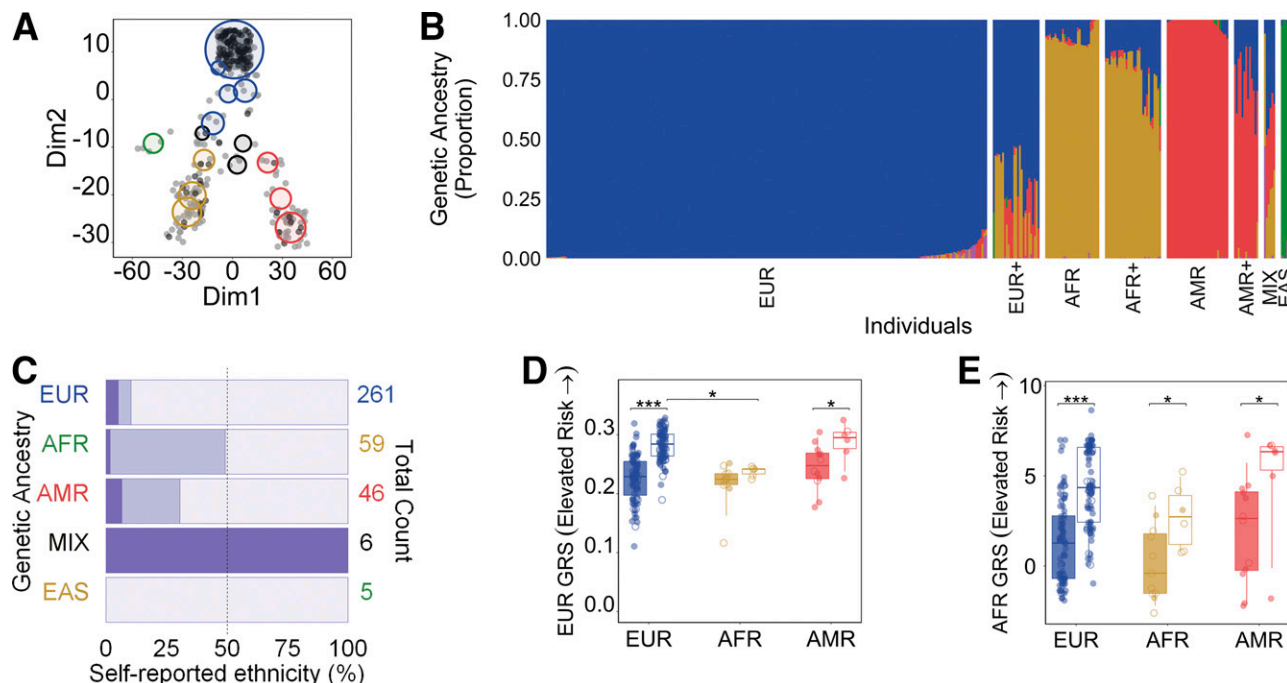


Figure 1—Analysis of T1D GRSs by ancestry. Colors denote individuals of EUR (blue), AFR (brown), AMR (red), EAS (green), and SAS (purple) genetic ancestry. **A:** Individuals ($n = 377$) are plotted as light (no T1D) and dark (T1D) gray circles according to their proportions of EUR, AFR, AMR, EAS, and SAS ancestry. Each non-gray-colored circle represents the center of a cluster of individuals; the size of the circle is proportional to the number of people in a cluster. Each cluster contains individuals of a single ancestry with a proportion of ≥ 0.88 (4 circles; blue circle at the top, red and yellow circles at the bottom, and the green circle), multiple ancestries in equal proportions (3 circles; termed MIX, black), or a single ancestry with contributions from others (8 remaining circles). **B:** Ancestry composition of all 377 individuals. A plus sign (+) denotes individuals within a group with increasing proportions from other ancestries. **C:** The percentage of agreement between the calculated genetic ancestry (**B**) and self-reported ethnicity taken from hospital medical charts (Caucasian, African American, Hispanic/Latino, Asian, American Indian/Alaska Native, and >1 group). Transparent purple represents concordance (e.g., an individual determined to be of AMR ancestry and self-reported as Hispanic/Latino), light purple is partial agreement (e.g., AMR+ ancestry and Hispanic/Latino), and dark purple denotes discordance (e.g., AMR or AMR+ and Caucasian). **D** and **E:** GRS values were compared within each ancestry for those with (box with no color fill) and without (solid color) T1D using EUR (**D**) and AFR (**E**) models. Data points with no fill color represent individuals with admixture ranging from a proportion of >0.01 to 0.11. * $P \leq 0.05$; *** $P \leq 0.0001$.

0.0001), AFR ($\Delta = 2.537$; 95% CL 0.264, 4.961; $P = 0.03$), and AMR ($\Delta = 2.928$; 95% CL -0.141 , 7.057; $P = 0.05$) populations. However, there also was a statistically significant difference in the EUR GRS between individuals without diabetes and individuals with T1D within EUR ($\Delta = 0.054$; 95% CL 0.044, 0.065; $P < 0.0001$) and AMR ($\Delta = 0.040$; 95% CL 0.001, 0.084; $P = 0.04$) but not AFR ($\Delta = 0.015$; 95% CL -0.005 , 0.028; $P = 0.15$) populations. Interestingly, the GRS was significantly different in individuals with T1D of AFR versus EUR ancestry using only the EUR ($\Delta = 0.046$; 95% CL 0.027, 0.063; $P^* \leq 0.0002$ [P^* values are significant only if less than a multiple-comparison-corrected nominal α of 0.025]) but not the AFR ($\Delta = 1.688$; 95% CL -0.247 , 3.580; $P^* = 0.08$) GRS model. The clinical utility of the AFR GRS to distinguish case and control individuals across ancestries will need to be evaluated in future studies. While the

sample size is small, limiting the genetic diversity seen, these results suggest that the development and application of GRS models tailored to all ancestries are warranted.

Very few studies have examined the issue of race, ethnicity, or ancestry in T1D risk prediction (3,5). We conclude that genetic contributions to T1D should be examined using ancestry-inclusive GRS models to more broadly determine whether certain features or disease endotypes are unique within or shared across populations. To avoid bias and potential false expectations, the biomedical research community needs to specify to which ancestries their study applies and take into account any admixture that may exist. Indeed, the progress we make will be measured by the value it brings to the lives of the patients we serve.

Acknowledgments. The authors thank Dave Ko (Arthur Riggs Diabetes and Metabolism

Research Institute, Beckman Research Institute, City of Hope, Duarte, CA) for developing a web-based toolkit to implement the containerized admixture pipeline used in this analysis, Srikar Chamala (College of Medicine, Diabetes Institute, University of Florida, Gainesville, FL) for contributions to genotype data processing, Amanda L. Posgai (College of Medicine, Diabetes Institute, University of Florida, Gainesville, FL) for technical assistance in editing and formatting this work, and Catherine S. Kaddis (Glendora, CA) for helpful discussions about the manuscript. The authors express apologies to those colleagues whose work was not cited, or cited and not discussed in detail.

Funding. Research reported in this publication was supported by nPOD (RRID:SCR_014541), a collaborative T1D research project sponsored by the JDRF (5-SRA-2018-557-Q-R, 25-2013-268, 25-2012-380, and 25-2007-874 to M.A.A., including a subcontract to J.S.K.), and The Leona M. & Harry B. Helmsley Charitable Trust (grant no. 2018PG-T1D053). Organ procurement organizations partnering with nPOD to provide research resources are listed at <https://www.jdrfnpod.org/for-partners/npod-partners/>. This research was performed using resources and/or funding provided by the National Institute of Diabetes and Digestive and Kidney Diseases—supported

Human Islet Research Network (HIRN) (RRID: SCR_014393; <https://hirnetwork.org>, U24DK104162 to J.S.K.) and a National Institute of Allergy and Infectious Diseases program project grant (P01 AI042288 to M.A.A. and T.M.B.).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or JDRF. The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data or in the writing or revision of the manuscript.

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

Author Contributions. J.S.K. and B.O.R. conceived and designed the work, drafted and revised the manuscript, and interpreted the data. J.S.K., D.J.P., and A.N.V. acquired and analyzed the data. M.A.A. and T.M.B. acquired the data. S.S.R. is part of the Population Architecture using Genomics and

Epidemiology (PAGE) Consortium, working to develop better GRS models by expanding non-European ancestry populations and increased methodologic work. All authors contributed to discussions about and revisions of the manuscript, assume personal accountability for their own contributions, and approved the final version of this work. J.S.K. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation. Parts of this study were presented in abstract form at the virtual 81st Scientific Sessions of the American Diabetes Association, 25–29 June 2021.

References

1. Patel KA, Oram RA, Flanagan SE, et al. Type 1 diabetes genetic risk score: a novel tool to dis-

criminate monogenic and type 1 diabetes. *Diabetes* 2016;65:2094–2099

2. Divers J, Mayer-Davis EJ, Lawrence JM, et al. Trends in incidence of type 1 and type 2 diabetes among youths—selected counties and Indian reservations, United States, 2002–2015. *MMWR Morb Mortal Wkly Rep* 2020; 69:161–165

3. Onengut-Gumuscu S, Chen W-M, Robertson CC, et al.; SEARCH for Diabetes in Youth; Type 1 Diabetes Genetics Consortium. Type 1 diabetes risk in African-ancestry participants and utility of an ancestry-specific genetic risk score. *Diabetes Care* 2019;42:406–415

4. 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;51:584–591

5. Perry DJ, Wasserfall CH, Oram RA, et al. Application of a genetic risk score to racially diverse type 1 diabetes populations demonstrates the need for diversity in risk-modeling. *Sci Rep* 2018;8:4529