Comparative Evaluation of Cross-Ancestry Polygenic Risk Scoring of Type 1 Diabetes in the All of Us Cohort

Mohammad Ahangari[1], Spencer Moore[1], Ivan Davidson[1], Jonathan Anomaly[1], Robert Maier[1], Jeremiah H. Li[1], Michael Christensen[1], David Stern[1]*, Tobias Wolfram[1]*

[1]Herasight Research, USA. * Contributed equally.

Abstract

Type 1 diabetes is a highly heritable autoimmune condition characterized by the destruction of pancreatic beta cells, resulting in insulin deficiency. Here, we developed a novel polygenic score we call the HLA-Augmented SBayesRC Framework (HLA-ARC). HLA-ARC integrates direct modeling of HLA haplotypes, with a Bayesian regression approach for the non-HLA component. SBayesRC leverages extensive functional genomic annotations and linkage disequilibrium patterns across approximately 7.4 million variants, substantially enhancing predictive accuracy. We systematically compared HLA-ARC to three existing T1D polygenic scores (Polygenic Risk Score extension for Diabetes Mellitus [PRSedm], Trans-Ancestry Polygenic Score for Diabetes [TA-PS], and Type 1 Diabetes Multi-Ancestry Polygenic Score [T1D-MAPS]) using data from the ancestrally-diverse All of Us cohort. Among the three existing methods, T1D-MAPS showed superior performance in all ancestry groups. However, HLA-ARC consistently outperformed the existing methods, achieving AUROC values exceeding 0.91 in European individuals and 0.89 in non-European groups. Our results demonstrate that integrating HLA haplotype modeling with genomic annotation and ancestry-informed linkage disequilibrium methods significantly improves polygenic risk prediction for autoimmune diseases characterized by major genetic risk loci.

Introduction

Type 1 diabetes (T1D) is a highly heritable autoimmune disease characterized by the destruction of pancreatic beta cells and absolute insulin deficiency, with substantial genetic risk conferred by variants in the human leukocyte antigen (HLA) region on chromosome 6. The HLA region encodes proteins essential for immune recognition and antigen presentation. Although the lifetime risk of T1D is approximately 0.4% in the general population, it increases significantly to around 6% for individuals with an affected dizygotic twin, and to 25–50% for those with an affected monozygotic twin[1].

The HLA region accounts for a large fraction of T1D heritability[2], with higher risk conferred by specific HLA class II haplotypes. The DR4-DQ8 and DR3-DQ2 haplotype combination demonstrates strong associations, with odds ratios reaching $\sim$15 for T1D susceptibility[3,4]. These haplotypes and their specific combinations, influence T1D risk by determining the peptide-binding

specificity of HLA class II molecules, which enhances presentation of beta cell autoantigens and facilitates the breakdown of immune tolerance.

Beyond the HLA region, genome-wide association studies (GWAS) have identified over 50 significant non-HLA loci that modify disease risk, each contributing relatively modest effects to T1D susceptibility[5]. This genetic architecture, dominated by a few high-effect HLA haplotypes with additional contribution from many small-effect variants, poses distinct challenges for constructing accurate polygenic scores (PGS).

Conventional genome-wide scoring methods typically treat all loci uniformly, failing to account for the outsized contribution of the HLA region in T1D. This limitation is compounded across diverse populations where HLA haplotype frequencies and LD patterns differ substantially. For instance, European-derived T1D polygenic scores have substantially reduced predictive accuracy when applied to non-European populations, particularly when HLA-specific risks are not explicitly modeled[6]. This discrepancy underscores the need for ancestry-aware modeling frameworks that properly handle both HLA and genome-wide variation across ancestries.

In this study, we introduce the HLA-Augmented SBayesRC Framework (HLA-ARC), an extension of T1D-MAPS[7] designed to further refine the non-HLA component of T1D PGS by integrating LD information and functional annotations for approximately 7.4 million genetic markers, using the Bayesian PGS construction method SBayesRC[8]. SBayesRC has demonstrated enhanced cross-ancestry predictive accuracy, achieving improvements of up to $\sim$34% compared to its predecessor, SBayesR[9]. HLA-ARC combines direct genotyping and phased scoring of high-risk HLA DRB1-DQA1-DQB1 haplotypes with the genome-wide non-HLA component derived from SBayesRC, resulting in a unified, ancestry-informed PGS (Figure 1). Notably, HLA-ARC can be extended to other diseases that require careful modeling of both HLA and non-HLA components, such as multiple sclerosis and other autoimmune conditions.

We find HLA-ARC attained meaningful risk stratification between T1D cases and controls across diverse ancestry groups using data from the All of Us (AoU) cohort[10], comprising over 400,000 individuals with whole-genome sequencing data. We then systematically benchmarked HLA-ARC against three published state-of-the-art PGS frameworks – PRSedm[11], TA-PS[12], and T1D-MAPS – and found that HLA-ARC outperformed all tested methods across all ancestries. HLA-ARC thus provides a more accurate and equitable framework for T1D genetic risk assessment across ancestry groups, surpassing current state-of-the-art methods in predictive performance.

Results

Study populations and modeling overview

All models were evaluated in the All of Us (AoU) Research Program using strict EHR-based T1D definitions. After quality control and phenotype curation, AoU provided 200 T1D cases and 21,802 controls (EUR $n = 78/7{,}716$; AFR $n = 54/4{,}029$; AMR $n = 51/6{,}462$; other ancestries were under-powered for ancestry-specific primary analyses). Classical HLA alleles (DRB1, DQA1, DQB1) were inferred and phased to haplotypes; the non-HLA component of the score was constructed using SBayesRC applied to $\sim$7.4M variants. Both components were then combined in a weighted sum to compute individual scores (see Methods).

HLA-ARC score distributions highlight T1D risk stratification and provide practical risk thresholds

Kernel-density plots of the HLA-ARC scores with continuous ancestry adjustment show clear case-control stratification within each ancestry group (Figure 2A). The right-shifted distributions in T1D cases relative to controls are most pronounced in EUR and remain distinct in AFR and AMR, as well as combined cohorts. Using the equal-density threshold (shown as a dashed line), we demonstrate robust discrimination, but also highlight slightly reduced separation in non-EUR groups, with sensitivity/specificity of 0.84/0.89 in EUR and 0.77/0.86 in non-EUR cohorts, respectively. Figure 2B displays AUROC curves from continuous ancestry-adjusted HLA-ARC scores (see Methods). The left-skewed, rapidly rising profiles show that sensitivity is achievable at low false-positive rates across all ancestries studied, further supporting the portability of these scores across ancestries.

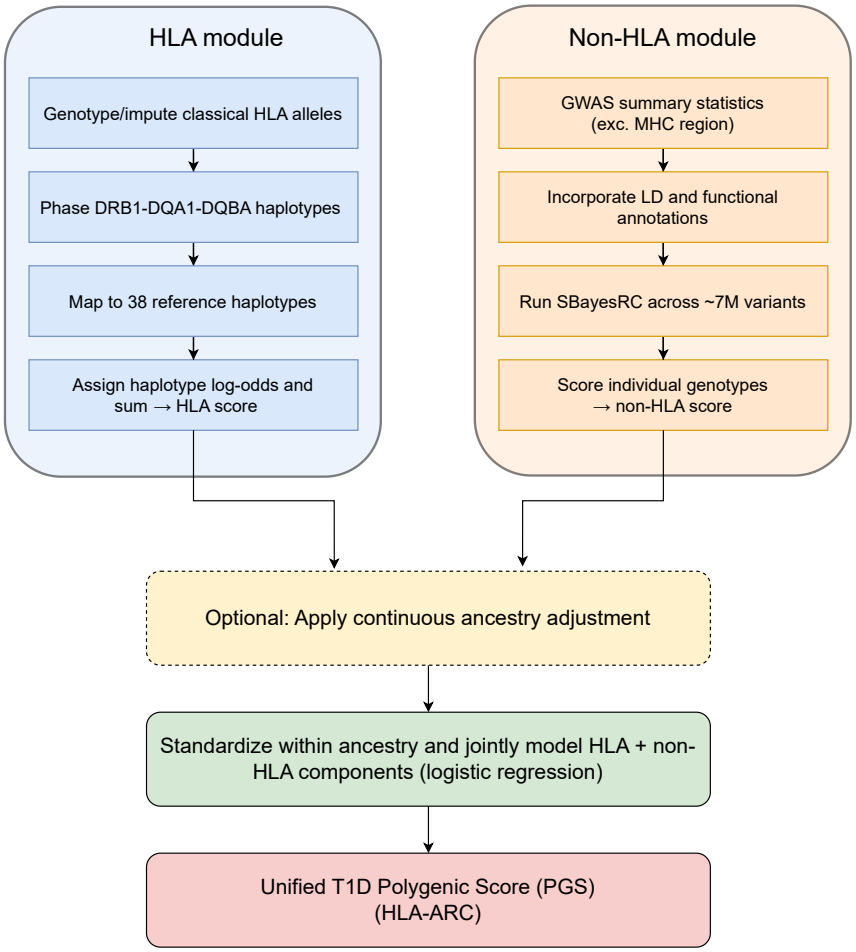

Figure 1: **HLA ARC workflow.**
HLA alleles are directly genotyped or imputed and scored as phased DRB1 DQA1 DQB1 haplotypes. The non-HLA variants are scored using SBayesRC posterior effects estimated across ∼7.4M variants with LD and functional annotations. The two standardized components are combined with a joint model to produce unified T1D PGS weights. The two components are then jointly modeled to generate the unified PGS.

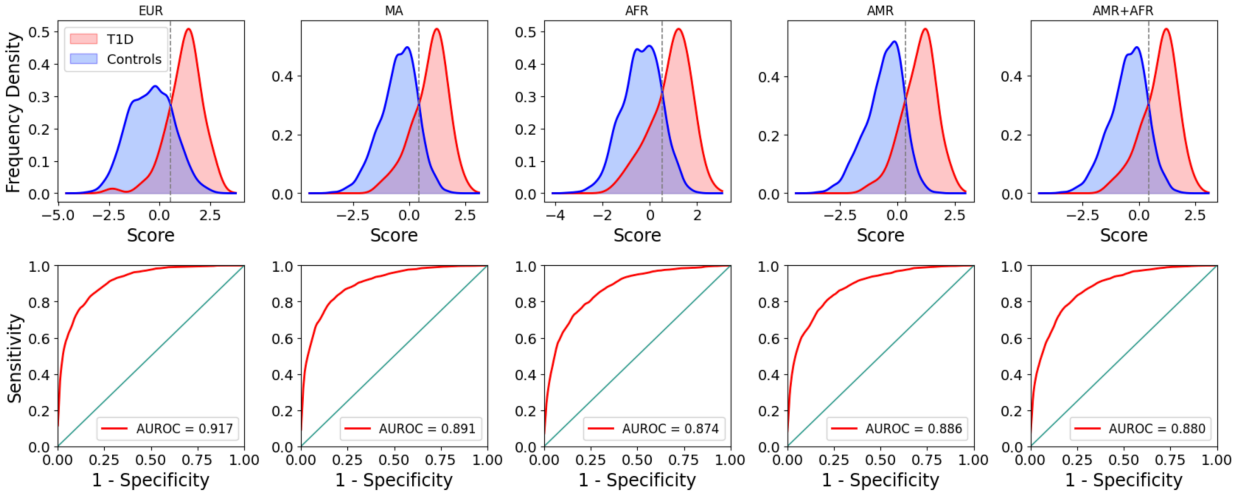

Figure 2: **Ancestry-stratified performance of the HLA-RC scores with continuous ancestry adjustment.**
(A) kernel-density estimates of cases (red) and controls (blue). The dashed line marks the equal-density threshold (intersection of KDEs). Reported operating points at this threshold were: EUR 0.84/0.89 (sensitivity/specificity), AFR 0.75/0.85, AMR 0.81/0.87, and AMR+AFR 0.77/0.86. (B) AUROC curves of the same scores, with diagonal lines indicating no discrimination null model and red lines showing the observed performance. Curves remain well above the diagonal across all ancestries, indicating robust discrimination.

Comparison with three published approaches

In analyses run alongside established methods (Table 1), HLA-ARC matched or exceeded their performance (Figure 3). The standard PRSedm model showed moderate discrimination for T1D across ancestries in the AoU cohort, consistent with prior literature[7]. In the EUR subset, PRSedm achieved an AUROC of ∼0.882, whereas performance was notably lower in the AFR subset (∼0.784) and intermediate in the AMR group (∼0.793). Correspondingly, variance explained on the liability scale by PRSedm was modest across all groups, with the lowest (0.189) observed in AFR individuals.

Incorporating genome-wide PGS alongside the HLA component improved predictions relative to PRSedm. TA-PS, which uses PRS-CS-derived weights[13], raised AUROC in AFR and AMR ancestries (e.g., from ∼0.80 to above 0.83 in AMR), with EUR performance remaining around ∼0.88. Overall, including genome-wide TA-PS data increased the fraction of T1D liability variance explained compared to PRSedm.

Table 1: **Summary of the PGS methods compared in this study, highlighting how each approach handles HLA and non-HLA components of the PGS.**

Method	HLA component	non-HLA component	Score integration approach
PRSedm	Fixed set of 67 established variants	Limited to 32 non-HLA markers	Unweighted linear sum
TA-PS	Fixed set of 67 established variants	1.2M non-HLA with PRS-CS	Unweighted linear sum
T1D-MAPS	Direct genotyping of DRB1, DQA1, and DQB1	1.2M non-HLA markers with PRS-CS	Weighted linear combination
HLA-ARC	Direct genotyping of DRB1, DQA1, and DQB1	7.4M non-HLA markers with SBayesRC	Weighted linear combination

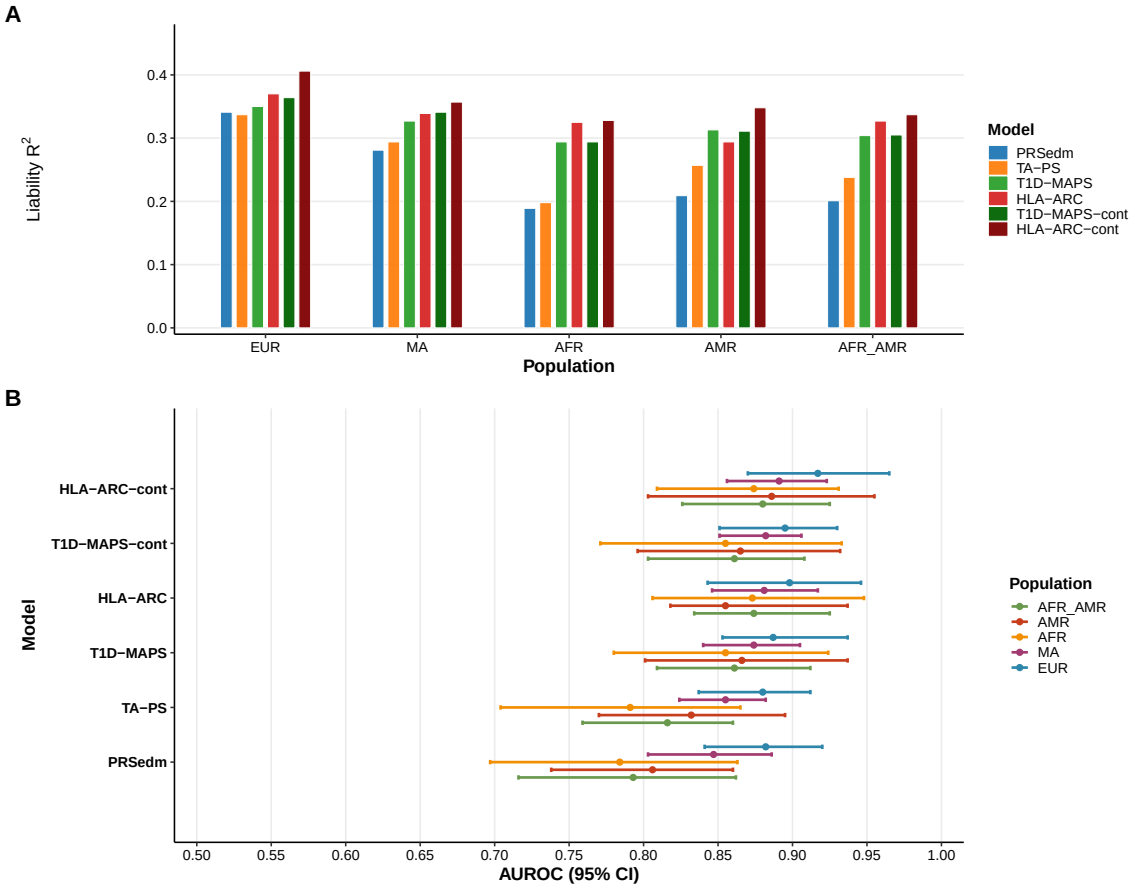

Figure 3: **Performance of polygenic risk score models for Type 1 Diabetes across diverse ancestries.**

(A) Liability-scale R^2 values for six T1D prediction models across five ancestry groups (EUR: European, MA: Middle Eastern, AFR: African, AMR: American, AFR_AMR: African American). (B) Area under the receiver operating characteristic curve (AUROC) with 95% confidence intervals for the same models.

T1D-MAPS had superior predictive accuracy compared to PRSedm and TA-PS across all ancestries (Figure 3), with AUROC increasing from 0.78 to 0.85 in T1D in AFR ancestry. Adopting the HLA-ARC approach with comprehensive SBayesRC weights further enhanced discrimination, achieving AUROC values around 0.90 in EUR and more than 0.87 in non-EUR individuals (Figure 3). Consistent with this high prediction accuracy and discrimination power, HLA-ARC explained the largest proportion of liability variance across ancestries.

The relative predictive power of the HLA and non-HLA component may differ with ancestry. Specifically, we have previously shown that our SBayesRC scoring approach has reduced performance in ancestries that scales with the genetic distance of the test set of individuals from the training set[14]. Thus, we applied a continuous ancestry adjustment to the SBayesRC component of the score, as the dispersed polygenic signal likely more closely obeys the ancestry attenuation patterns observed across polygenic traits, whereas the HLA alleles would not necessarily follow the same effect size attenuation pattern. As expected, we observed ancestry-dependent reduction in the effect sizes of both the HLA and non-HLA components when jointly regressed of T1D status that scaled proportionally to ancestral distance. Notably, the *relative* predictive importance of the associations of the HLA and non-HLA components increased with genetic ancestry. The ratio of the liability R^2 of the HLA to non-HLA components effect sizes increased from 2.02 in EUR to 3.95 in AMR and 4.73 in AFR, where notably the non-HLA component was not significantly associated with T1D status (Table 2), underscoring the importance of accurate HLA haplotyping in non-EUR individuals. Applying continuous ancestry adjustments to account for the change in relative importance of the HLA and non-HLA components provided marginal gains for both T1D-MAPS and HLA-ARC. For example, HLA-ARC EUR scores slightly improved to $\sim$0.91, and non-EUR (AFR+AMR) scores increased from 0.87 to 0.88. These results suggest accounting for gradations of ancestry can modestly enhance prediction accuracy, particularly in non-EUR cohorts.

Table 2: **HLA-ARC-cont odds ratios (OR) with 95% confidence intervals (CI) and liability-scale R^2 values for HLA and non-HLA components across ancestries. HLA consistently shows stronger effects and higher liability R^2, while non-HLA effects attenuate in non-EUR groups.**

Ancestry	HLA OR	HLA 95% CI	HLA liability R^2	non-HLA OR	non-HLA 95% CI	non-HLA liability R^2
EUR	3.2	2.656 – 3.852	0.265	2.3	1.900 – 2.769	0.131
MA	3.0	2.400 – 3.721	0.247	1.8	1.440 – 2.232	0.091
AMR	2.7	2.079 – 3.445	0.265	1.6	1.232 – 2.041	0.067
AFR	2.3	1.552 – 3.197	0.258	1.21	0.816 – 1.681	0.045
AFR_AMR	2.5	1.875 – 3.250	0.261	1.45	1.087 – 1.885	0.053

Discussion

Our study introduces HLA-ARC, a novel T1D genetic risk score that integrates HLA-haplotype calling with a functionally informed polygenic scoring with SBayesRC. Previous T1D PGS efforts have taken one of three approaches: not explicitly modeling HLA, relying on a fixed set of

known variants genome-wide, or using standard genome-wide PGS approaches such as PRS-CS for genome-wide data. Here we show that bridging these approaches can be advantageous: by modeling the HLA region in a biologically informed manner via direct genotyping of important HLA alleles, and simultaneously applying annotation-aware PGS approaches such as SBayesRC (which includes a much larger set of genome-wide markers for the non-HLA scoring), one can consistently achieve better overall predictive performance compared to published methods. Our analysis highlights the importance of separately modeling HLA and non-HLA components of polygenic risk in T1D, and underscores how different modeling strategies can impact performance across ancestries.

T1D-MAPS's flexible framework for direct genotyping of HLA haplotypes, combined with precise LD mapping and functional annotation-aware genome-wide scoring of non-HLA components of the risk from SBayesRC, provides a more complete assessment of the genetic architecture of T1D, and hence a more robust estimate of the genetic risk across different ancestries. For example, in the EUR cohort, the HLA-ARC model reached an AUROC as high as $\sim$0.91, slightly above the $\sim$0.88 achieved using PRS-CS weights, with similar trends in the non-EUR cohorts. Although these performance gains appear modest in magnitude, they were consistently observed, indicating a tangible improvement from PRS-CS to SBayesRC. Additionally, our results align with the broader observation that SBayesRC often outperforms other PGS algorithms by better fine-mapping causal signals[8]. When predicting rare diseases like T1D, even a small increment in AUROC or variance explained by the PGS can provide meaningful improvements for disease stratification, potentially translating to better identification of high-risk individuals.

While HLA-ARC showed the best performance overall, it is important to note that TA-PS also demonstrated robust cross-ancestry performance and clear improvement over the baseline PRSedm approach. This suggests that non-HLA loci continue to have a meaningful impact on the genetic risk of T1D, and comprehensive inclusion of them into the overall T1D PGS is crucial. However, despite clear improvements in the non-HLA portion of the score, TA-PS still relies on PRSedm to derive the HLA risk component, which means that TA-PS's HLA scores are not as finely tuned as T1D-MAPS's direct haplotype-based genotyping and scoring of the HLA region.

In contrast to the two methods that use genome-wide data for non-HLA portions of the risk, PRSedm provides the simplest and quickest approach for T1D genetic risk stratification by scoring only 67 risk variants known to have high effect on T1D. Despite the relatively good performance of PRSedm, lack of explicit ancestry tailoring for the HLA region and not using genome-wide data for non-HLA regions, results in an observable performance lag for PRSedm compared to both TA-PS and T1D-MAPS, especially in non-EUR samples. This observation is further reflected in the larger decline in predictive accuracy of T1D PGS scores for AFR and AMR individuals with PRSedm relative to both TA-PS and T1D-MAPS. For instance, PRSedm saw a pronounced drop in AUROC from EUR to AFR individuals, whereas HLA-ARC maintained high discrimination even in non-EUR individuals (0.89 in non-EUR vs 0.91 in EUR), effectively closing most of the cross-ancestry performance gap.

We have demonstrated improvements in T1D PGS prediction, but there are several limitations to our study. First, we were unable to perform within-family validation of the proposed HLA-ARC approach. Such within-family validations are important to confirm that PGS performance is not inflated by population stratification or other confounding factors. However, the AoU dataset was underpowered for this purpose due to its limited sample size and the scarcity of multiplex T1D families. Nevertheless, the genetic architecture of T1D suggests that attenuation of PGS

performance from family structure is likely minimal. HLA alleles are the strongest risk factors for T1D, and they exert large effects even within families. For example, siblings who share both high-risk HLA haplotypes have a significantly higher chance of developing autoimmune disease than siblings without them[15]. This supports the view that our PGS is capturing direct genetic effects rather than potential confounding. Still, as larger cohorts or family-based datasets become available, future work should formally confirm PGS performance within families to further validate this approach.

Second, sample size imbalances across ancestry groups, particularly the limited number of T1D cases of non-EUR ancestry, may have affected the stability of performance estimates. The wide confidence intervals and overlapping performance metrics across methods suggest that some comparisons, such as the gain of HLA-ARC over T1D-MAPS, may not consistently reach statistical significance with current sample sizes. This underscores the need for larger cohorts to definitively distinguish performance differences.

Third, our case-control definitions relied on established EHR-based criteria, which carry a risk of phenotype misclassification, undetected inclusion of atypical diabetes cases, or misclassified controls, all of which could attenuate the apparent predictive power of the PGS. Future work should therefore prioritize external validation in other large biobank cohorts and refine case definitions, ideally leveraging direct clinician-confirmed diagnoses, when available.

Fourth, although our evaluation included EUR, AFR, MA, and AMR cohorts, not enough Asian samples with T1D were available in AoU for a proper analysis. Nonetheless, our framework is broadly generalizable: the major genetic determinants of T1D, including HLA-DR/DQ alleles, and key non-HLA loci, are relevant across ancestral backgrounds, even if the specific risk variants or allele frequencies differ[1]. For instance, DR3-DQ2 and DR4-DQ8 drive much of the risk in EUR populations, whereas East Asian populations carry distinct high-risk alleles such as HLA-DR9 in Japan[16]. If our current score were applied directly to East Asian cohorts without re-weighting, performance would likely be attenuated because many EUR-trained variants have different allele frequencies or weaker effect sizes in East Asians, consistent with cross-ancestry portability challenges seen in other complex diseases. However, our proposed framework can recover a substantial amount of accuracy by incorporating population-specific HLA signals through direct genotyping and by recalibrating genome-wide weights with SBayesRC using East Asian reference data. Based on prior PGS transferability studies for complex diseases, we would expect performance close to what we achieve in AMR, and possibly higher than AFR populations[17].

Fifth, we did not perform direct HLA allele-calling; instead, we imputed classical DRB1–DQA1–DQB1 alleles using SNP2HLA[18]. AoU's current large-scale genomics release is based on short-read WGS, which does not by itself resolve long-range phase across the MHC. While HLA imputation can be quite accurate for common European haplotypes (e.g., ∼95–97% two-field with the T1DGC panel), it degrades for rarer alleles and across ancestries, introducing non-differential misclassification that will attenuate HLA effect sizes and, by extension, our performance estimates[18]. In head-to-head studies, short-read WES/WGS allele-calling (e.g., HLA-HD/HLA*LA) is more accurate than imputation and increases discovery power for autoimmune traits, underscoring that even a short-read caller would likely improve our HLA component[19]. Nevertheless, long-read sequencing (especially PacBio HiFi) enables full-length, fully phased HLA haplotypes with ∼99.8% per-base accuracy and has been shown to resolve class II ambiguities, which we expect would yield the highest-fidelity DR/DQ haplotypes and sharpen risk stratification in future iterations of our pipeline[20]. Given the relatively greater importance of the HLA to non-HLA

component of HLA-ARC, accurate HLA allele-calling is crucial particularly for maximizing risk prediction accuracy in underrepresented ancestries.

The success of our approach in the multi-ancestry AoU cohort demonstrated the feasibility of developing a robust cross-population T1D PGS. However, continued efforts to expand portability of PGS to diverse ancestries and refine modeling strategies will be critical to delivering clinically meaningful risk assessments for T1D worldwide.

Methods

Study Cohort

This study used genetic and phenotypic data from the All of Us (AoU) Research Program's Controlled Tier Dataset Curated Data Repository version 8, available to authorized users on the Researcher Workbench. T1D case and control definitions were based on the validated eMERGE phenotyping algorithm[21]. The criteria integrate diagnosis codes, medication records, and age of onset to maximize specificity for autoimmune, early-onset T1D while minimizing misclassification from other forms of diabetes. Individuals were classified as T1D cases for the analysis if they satisfied all of the following criteria to ensure high specificity for T1D while minimizing inclusion of secondary atypical diabetes: early-onset T1D while minimizing inclusion of secondary or atypical diabetes:

- At least two recorded T1D diagnosis codes to reduce the risk of miscoding common in EHR data.
- Age at first T1D diagnosis under 40 to enrich for autoimmune T1D and exclude late-onset presentations more consistent with type 2 or atypical diabetes.
- Initiation of outpatient insulin therapy within 365 days of the first T1D diagnosis to capture insulin dependence from disease onset.
- No recorded use of non-insulin diabetes medications prior to insulin initiation, ensuring that treatment history was consistent with T1D rather than type 2 or atypical diabetes.
- No diagnostic codes indicative of cystic fibrosis, cancer, drug-induced diabetes, or other secondary causes of diabetes, to exclude confounding conditions.

Controls were defined using a conservative eMERGE recommended phenotyping algorithm to minimize false negatives and exclude individuals with any potential diabetes-related conditions or medications. Individuals used as controls were required to have:

- At least two least of longitudinal EHR data between the first and last recorded condition, ensuring adequate opportunity to capture any diabetes diagnosis, if present.
- No diagnosis codes for type-1 or type-2 diabetes.
- No prescription for either insulin or non-insulin diabetes medications.
- No diagnosis codes for secondary forms of diabetes or malignant neoplasms.

This strict case-control framework prioritizes specificity over sensitivity, providing a high-confidence T1D phenotype for genetic analyses while minimizing contamination from type 2 or atypical diabetes.

The final cohort included 200 cases and 21,802 controls, with the ancestry distribution detailed in Table 3. For ancestry-specific analyses, only participants of EUR, AFR, and AMR ancestry were retained due to power limitations in other groups.

Table 3: **Distribution of cases and controls by genetic ancestry in the study cohort. MA includes AFR, AMR, and any other ancestry group that did not reach a threshold required for the analysis.**

Cohort	N_Cases	N_Controls
EUR	78	7,716
AFR	54	4,029
AMR	51	6,462
MA	122	14,086

For the genetic data, we used AoU Allele Count Allele Frequency (ACAF) whole-genome sequencing (WGS) dataset for all subsequent analyses. This dataset includes variant calls from WGS, encompassing both SNPs and INDELs, with a population-specific allele frequency filter of over 1% and an allele count exceeding 100 in any of the computed subpopulations. It comprises approximately 58 million variant sites across $\sim$415,000 samples and is optimized for efficient analysis of common genetic variation in AoU by minimizing rare-variant noise and reducing computational burden.

PGS models

PRSedm. The PRSedm approach is an improved version of the GRS2x score, which integrates 67 variants, including those in HLA class I, HLA class II, and non-HLA genome-wide data, partitioned into separate component scores. For each category, a partitioned polygenic score is calculated as the weighted sum of risk alleles using published effect sizes.

More specifically, HLA scores are derived from 36 variants marking HLA haplotypes. Due to imperfect LD between SNPs and true HLA alleles, the method uses imputation and frequency based filtering to assign the most likely haplotype pair, and then converts these into a polygenic distribution. The non-HLA scores are based on the remaining risk variants outside the HLA region. Each allele effect is estimated from the GWAS sources on T1D, with missing variants imputed using reference frequencies under Hardy–Weinberg equilibrium expectations. The combined scores are then obtained by summing the three partitions into a total standardized score. To ensure comparability across studies, PRSedm total score is standardized to a fixed theoretical range using minimum to maximum relative risk normalization. This approach allows separate inspection of HLA and non-HLA contributions, while also providing a single unified score (called GRS2x in PRSedm) for the overall impact of genetic risk on T1D.

TA-PS. The TA-PS approach is a simple extension of PRSedm that combines an optimized non-HLA polygenic score by replacing the non-HLA component of PRSedm (which includes

only a limited number of markers) with genome-wide weights derived from PRS-CS. The two components of the score are then combined into a unified PGS using a simple linear approach.

T1D-MAPS. T1D-MAPS is a recently developed approach that explicitly models HLA and non-HLA components of the T1D risk by directly genotyping HLA alleles and combining them with genome-wide non-HLA scores. For the HLA component, classic HLA alleles are imputed using SNP2HLA approach with a multi-ancestry WGS reference panel. Phased multi-locus haplotypes spanning HLA-DRB1, HLA-DQA1, and HLA-DQB1 are then compared to a reference set of 38 haplotypes derived from a multi ancestry T1D GWAS, with each haplotype assigned a score equal to its log odds of T1D association. In case of ambiguous or unresolved haplotypes, a neutral score of zero is assigned to that haplotype. For the non-HLA component, genome-wide variants summary statistics data from outside the MHC region were used to generate polygenic weights using the PRS-CS method by applying continuous shrinkage prior to approximately 1.3M HapMap3 variants using the EUR 1000 Genomes reference panel.

To construct the overall score, a logistic regression model was fit in the MGB Biobank[22] to estimate weights for the HLA and non-HLA components and the resulting coefficients (β-HLA and β-non-HLA) were then applied to combine the two partitions into a single linear prediction.

HLA-ARC. Our proposed approach extends T1D-MAPS by retaining the original partitioned design and explicit modeling of HLA haplotypes, but replacing the PRS-CS based genome-wide component with weights derived from SBayesRC. As in T1D-MAPS, classical alleles at HLA-DRB1, HLA-DQA1, and HLA-DQB1 were imputed using SNP2HLA, then phased into multi-locus haplotypes in the cohort. For the non-HLA region, we jointly modeled GWAS summary statistics data with LD from the 1000 Genomes panel and functional annotations (BaselineLD) to shrink SNP effects toward biologically informed priors that account for functional effect of each variant and its LD pattern. Posterior mean effect sizes from SBayesRC were calculated across 7.4 million variants, and individual non-HLA scores were obtained by summing allele dosages weighted by these posterior estimates using the scoring function in plink2. Finally, the combined score was derived by using the same formula used in the original T1D-MAPS publication to yield the integrated predictor. This formulation preserved the strong disease-defining contribution of HLA haplotypes while leveraging the greater modeling capacity of SBayesRC for the genome-wide component of the score.

Continuous ancestry adjustments for T1D-MAPS and HLA-ARC. To account for continuous genetic ancestry, we used the approach described in T1D-MAPS2, which adjusts an individual's PGS by weighting the proportion of EUR ancestry using the UK Biobank's pan-ancestry framework. Individuals with higher EUR ancestry receive greater weight from the PRSedm score, which is optimized for EUR ancestry, while the direct genotyping–based score receives more weight as ancestry diverges from EUR. This formulation allows the score to adapt smoothly across the ancestry spectrum rather than relying on a discrete EUR versus non-EUR classification. By modeling ancestry as a continuous probability instead of a categorical label, the approach can be applied consistently across populations without requiring self-reported or PCA-derived ancestry assignment, thereby reducing bias and improving score portability.

Evaluation metrics

Model performances were assessed using the area under the receiver operating characteristic curve (AUROC) and the liability-scaled R^2.

AUROC quantifies the discriminative ability of the scores by measuring the probability that a randomly chosen case has a higher score than a randomly chosen control, equivalent to the rank-based separation of cases and controls across thresholds.

The liability-scaled R^2 estimates the proportion of variance in the unobserved liability to disease explained by the score, adjusting for assumed disease prevalence in the population (Table 4).

Table 4: **Population prevalence of T1D across different ancestries.**

Cohort	Population prevalence
EUR	0.0068
AFR	0.0063
AMR	0.0061

Observed-scale R^2 was first derived using Cohen's d as:

$$R_o^2 = \frac{d^2 p(1-p)}{1 + d^2 p(1-p)}$$

where p is the sample prevalence and d is the standardized mean difference between cases and controls. This was then transformed to the liability scale following Lee et al[23]:

$$R_l^2 = \frac{C R_o^2}{1 + C\theta R_o^2}$$

$$t = \Phi^{-1}(1-k), \quad z = \varphi(t), \quad i = \frac{z}{k}, \quad C = \frac{(k(1-k))^2}{z^2 p(1-p)}, \quad \theta = i\left(\frac{p-k}{1-k}\right)\left(i\left(\frac{p-k}{1-k}\right) - t\right)$$

where k denotes lifetime disease risk. Standard errors were obtained by bootstrapping with 2,000 replicates.

Data access statement

This study used genetic and phenotypic data from the All of Us Research Program's Controlled Tier Dataset Curated Data Repository version 8, available to authorized users on the Researcher Workbench.

Acknowledgements

We gratefully acknowledge All of Us participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's All of Us Research Program for making available the participant data examined in this study.

References

[1] Jahromi MM, Eisenbarth GS. Genetic determinants of type 1 diabetes across populations *Ann N Y Acad Sci.* 2006;1079:289–299.

[2] Dorman JS, Steenkiste AR, O'Leary LA, McCarthy BJ, Lorenzen T, Foley TP. Type 1 diabetes in offspring of parents with type 1 diabetes: the tip of an autoimmune iceberg? *Pediatr Diabetes.* 2000;1:17–22.

[3] Tisch R, McDevitt H. Insulin-dependent diabetes mellitus *Cell.* 1996;85:291–297.

[4] Pang H, Lin J, Luo S, et al. The missing heritability in type 1 diabetes *Diabetes Obes Metab.* 2022;24:1901–1911.

[5] Michalek DA, Tern C, Zhou W, et al. A multi-ancestry genome-wide association study in type 1 diabetes *Hum Mol Genet.* 2024;33:958–968.

[6] Redondo MJ, Gignoux CR, Dabelea D, et al. Type 1 diabetes in diverse ancestries and the use of genetic risk scores *Lancet Diabetes Endocrinol.* 2022;10:597–608.

[7] Deutsch AJ, Bell AS, Michalek DA, et al. Development and validation of a type 1 diabetes multi-ancestry polygenic score medRxiv 2025.

[8] Zheng Z, Liu S, Sidorenko J, et al. Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries *Nat Genet.* 2024;56:767–777.

[9] Lloyd-Jones LR, Zeng J, Sidorenko J, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics *Nat Commun.* 2019;10:5086.

[10] All of Us Research Program Genomics Investigators . Genomic data in the All of Us Research Program *Nature.* 2024;627:340–346.

[11] Luckett AM, Oram RA, Deutsch AJ, et al. Standardized measurement of type 1 diabetes polygenic risk across multiancestry population cohorts *Diabetes Care.* 2025;48:e81–e83.

[12] Jumentier B, Qu H, Lu T, et al. Development and validation of a trans-ancestry polygenic risk score for type 1 diabetes medRxiv 2025.

[13] Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors *Nat Commun.* 2019;10:1776.

[14] Moore S, Davidson I, Anomaly J, et al. Development and validation of polygenic scores for within-family prediction of disease risks medRxiv 2025.

[15] Valdes AM, Thomson G, Erlich HA, Noble JA. Association between type 1 diabetes age of onset and HLA among sibling pairs *Diabetes.* 1999;48:1658–1661.

[16] Kumagai T, Yamanaka T, Wataya Y, et al. A strong association between HLA-DR9 and gelatin allergy in the Japanese population *Vaccine.* 2001;19:3273–3276.

[17] Wang Y, Guo J, Ni G, Yang J, Visscher PM, Yengo L. Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations *Nat Commun.* 2020;11:3865.

[18] Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens *PLoS One.* 2013;8:e64683.

[19] Butler-Laporte G, Farjoun J, Nakanishi T, et al. HLA allele-calling using multi-ancestry whole-exome sequencing from the UK Biobank identifies 129 novel associations in 11 autoimmune diseases *Commun Biol.* 2023;6:1113.

[20] Wenger AM, Peluso P, Rowell WJ, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome *Nat Biotechnol.* 2019;37:1155–1162.

[21] Szczerbinski L, Mandla R, Schroeder P, et al. Algorithms for the identification of prevalent diabetes in the All of Us Research Program validated using polygenic scores *Sci Rep.* 2024;14:26895.

[22] Boutin NT, Schecter SB, Perez EF, et al. The evolution of a large biobank at Mass General Brigham *J Pers Med.* 2022;12:1323.

[23] Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies *Am J Hum Genet.* 2011;88:294–305.