Validating Orchid's Type 1 Diabetes Genetic Risk Score

ORCHID

Written by Orchid Bioinformatics Staff

Introduction

Type 1 diabetes is a chronic autoimmune disease in which the body's immune system destroys the insulin-producing beta cells of the pancreas, resulting in little or no insulin production and chronically elevated blood glucose levels. Common symptoms include increased thirst and urination, unintended weight loss, fatigue, blurred vision, and, if untreated, life-threatening complications.[1]

The prevalence of type 1 diabetes in the US general population is estimated to be approximately 0.3%.[2] Treatment of type 1 diabetes requires lifelong insulin therapy along with regular blood glucose monitoring and lifestyle management. Although there is no cure, advances in treatment have greatly improved prognosis, allowing many individuals to live long lives, though the disease remains associated with an increased risk of complications if blood glucose levels are poorly controlled.[1]

Genetic Risk Score

A person's risk of developing type 1 diabetes is shaped substantially by genetics.[3] Genetic risk scores (GRS) allow us to estimate this disease risk based on the DNA of a person or embryo.[4] Although not diagnostic, a GRS can indicate how likely an individual is to develop the disease compared to the population baseline risk.

The type 1 diabetes GRS used in Orchid's reports includes 63 variants and was developed based on variants identified in a study that included 6,481 cases (individuals with type 1 diabetes) and 9,247 healthy controls.[5] The final GRS score gives special weight to specific variants within the immune-related HLA region of chromosome 6, which account for a disproportionately large share of predicted genetic risk.[6] The specific HLA-DQ alleles considered by the model are listed in Supplementary Table 3.

Genetic risk scores that are trained primarily on data from individuals of European ancestry tend to be less accurate when applied to individuals of non-European ancestry.[7] This problem is especially challenging in the case of type 1 diabetes. Therefore, we currently offer type 1 diabetes model predictions only to individuals of European ancestry, although research efforts are underway to improve the predictive accuracy of this model for other ancestry groups.

Evaluation on UK Biobank Data

We evaluated the predictive accuracy of Orchid's type 1 diabetes GRS using the UK Biobank (UKB), a research database of roughly 500,000 genotyped individuals from the United Kingdom.[8] We restricted the analysis to people of British ancestry and defined T1D cases using strict criteria:

- Clinical diagnosis of diabetes from ICD-10 E10 codes in hospital inpatient records (fields 41204/41202), doctor-diagnosed diabetes reported by participants (field 2443), or self-reported type 1 diabetes (field 20002 code 1222).
- Absence of type 2 diabetes (excluding individuals with self-reported T2D code 1223 or ICD-10 E11 codes).
- Age of diabetes onset from 0 to 20 years old (field 2976).

This yielded 303 cases and 386,833 controls (0.08% prevalence). We then grouped individuals by GRS percentile and compared the observed disease prevalence within each group to our model's predictions (Figure 1). For additional technical details, see the Supplementary Information.

Table 1 shows the type 1 diabetes observed prevalence for individuals in the UKB grouped by GRS percentile range (top 10%, 5%, and 1%), as well as how their risk compares to the baseline risk at the 50th GRS percentile. Those with higher GRS relative to the population baseline also had substantially higher observed prevalence of type 1 diabetes, supporting the predictive accuracy of the GRS to identify individuals with elevated risk.

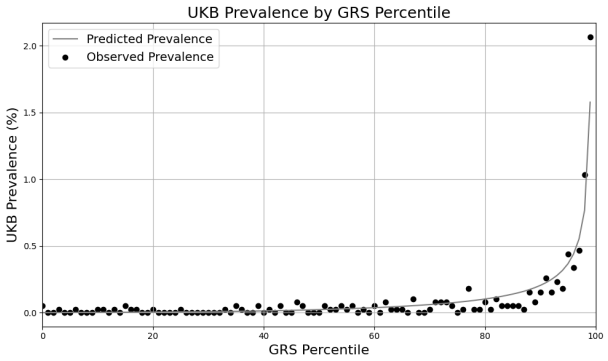

Figure 1. Risk Stratification. Predicted and observed prevalence in the UKB for individuals grouped by GRS percentile.

GRS Group	Observed UKB Prevalence	Odds Ratio
Baseline (50th percentile)	0.03%	1.00
Top 10%	0.53%	18.82
Top 5%	0.87%	30.80
Top 1%	2.07%	74.22

Table 1. Observed prevalence of type 1 diabetes in the UKB by GRS percentile range. Those with higher GRS relative to the population baseline also had substantially higher observed prevalence of type 1 diabetes.

Estimating Lifetime Risk

The average observed prevalence of type 1 diabetes in the UKB was 0.08%. This is considerably lower than the lifetime prevalence in the US general population, which has been estimated to be approximately 0.3%.[2] This is likely due in part to the fact that UKB participants tend to be healthier than the general population, which leads to lower observed disease prevalence.[9] Additionally, we used a strict filter to define type 1 diabetes cases in the UKB, so as to rule out type 2 diabetes cases that may have been misclassified, so this filter may have also eliminated some true type 1 diabetes cases.

Orchid's clinical reports include predicted lifetime disease risk, which we calculate by first estimating how disease risk varies across GRS in the UKB and then rescaling that pattern so the average matches the known lifetime population risk (Figure 2).[4] People at the high end of the GRS distribution are predicted to have an elevated lifetime risk of the disease relative to the population (Table 2).

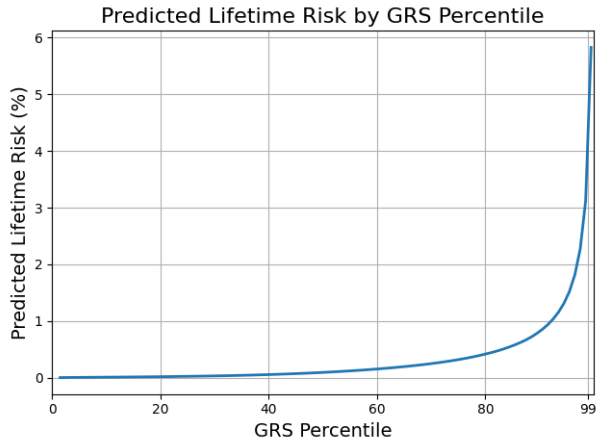

Figure 2. Adjusted Risk Stratification. Predicted risk estimates adjusted so that overall prevalence matches the approximately 0.3% estimate.[2]

GRS Percentile	Predicted Lifetime Risk	Relative Risk
50th (baseline)	0.09%	1.00x
95th	1.41%	15.07x
97th	2.01%	21.50x
99th	3.87%	41.39x

Table 2. Predicted lifetime prevalence of type 1 diabetes at different GRS percentiles. Individuals with the highest GRS percentiles are predicted to have an increased risk of type 1 diabetes relative to those at the 50th percentile.

Conclusion

In this study, we evaluated our type 1 diabetes GRS on data from the UKB. We found that it performed well, particularly for identifying individuals with elevated risk of the disease relative to the population. In our embryo and couple reports, we adjust the model to predict lifetime risk, which is generally higher than observed prevalence in the UKB. The type 1 diabetes GRS model is currently only available to individuals of European ancestry, due to concerns about loss of accuracy in other ancestry groups, although research efforts are underway to address this issue.

Acknowledgements

This research was conducted using the UK Biobank Resource under Application Number 80545.

References

1. Mayo Clinic Staff. Type 1 diabetes. https://www.mayoclinic.org/diseases-conditions/type-1-diabetes/symptoms-causes/syc-20353011, 2024. Accessed 2026.
2. A Menke, TJ Orchard, G Imperatore, KM Bullard, E Mayer-Davis, and CC Cowie. The prevalence of type 1 diabetes in the united states. *Epidemiology*, 24(5):773–774, 2013. doi:10.1097/EDE.0b013e31829ef01a.
3. CF Kuo, IJ Chou, MJ Grainge, et al. Familial aggregation and heritability of type 1 diabetes mellitus and coaggregation of chronic diseases in affected families. *Clinical Epidemiology*, 10:1447–1455, 2018. doi:10.2147/CLEP.S172207.
4. N Chatterjee, J Shi, M García-Closas, et al. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature Reviews Genetics*, 17:392–406, 2016. doi:10.1038/nrg.2016.27.
5. SA Sharp, SS Rich, AR Wood, et al. Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. *Diabetes Care*, 42(2):200–207, 2019. doi:10.2337/dc18-1785.

6. J Sticht, M Álvaro-Benito, and S Konigorski. Type 1 diabetes and the hla region: Genetic association besides classical hla class ii genes. *Frontiers in Genetics*, 12:683946, 2021. doi:10.3389/fgene.2021.683946.

7. Florian Privé, Hugues Aschard, Shai Carmi, et al. Portability of 245 polygenic scores when derived from the uk biobank and applied to 9 ancestry groups from the same cohort. *American Journal of Human Genetics*, 109(1):12–23, 2022. doi:10.1016/j.ajhg.2021.11.008.

8. C. Sudlow, J. Gallacher, N. Allen, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3):e1001779, 2015. doi:10.1371/journal.pmed.1001779.

9. A. Fry, T.J. Littlejohns, C. Sudlow, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American Journal of Epidemiology*, 186:1026–1034, 2017. doi:10.1093/aje/kwx246.

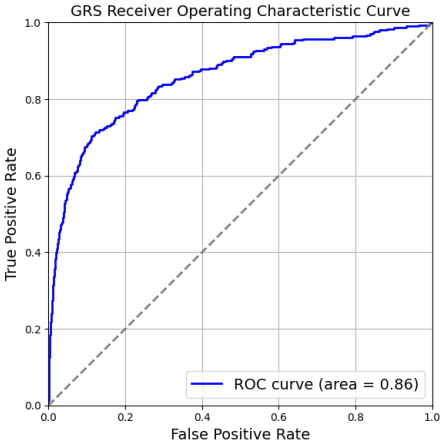

Figure 4. The receiver operating characteristic (ROC) used to compute the ROC area under the curve (AUC). The ROC curve is a graphical representation of a binary classifier's performance, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across different decision thresholds. A curve closer to the top-left indicates a better model, while a diagonal line (AUC = 0.5) represents random guessing.

Supplementary Information

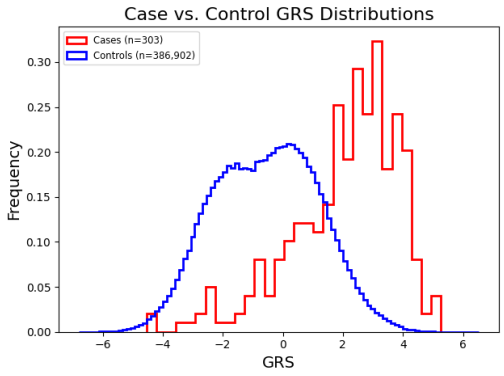

Figure 3. GRS histograms. GRS distributions for cases and controls. The case distribution is shifted noticeably higher compared to the controls.

HLA-DQ Allele
HLA-DQ2.2
HLA-DQ2.5
HLA-DQ4.2
HLA-DQ5.1
HLA-DQ5.3
HLA-DQ6.1
HLA-DQ6.2
HLA-DQ6.3
HLA-DQ6.9
HLA-DQ7.3
HLA-DQ7.5
HLA-DQ8.1
HLA-DQ9.2
HLA-DQ9.3

Table 3. HLA-DQ alleles incorporated into the type 1 diabetes GRS. The type 1 diabetes model accounts for the listed HLA-DQ haplotype alleles in the HLA region of chromosome 6, which account for a disproportionately large share of predicted genetic risk.