

# Validating Orchid's Type 2 Diabetes Genetic Risk Score

Written by Orchid Bioinformatics Staff

## Introduction

Type 2 diabetes (T2D) is a chronic metabolic disease in which the body becomes resistant to insulin and gradually fails to maintain normal blood glucose levels. Common symptoms include increased thirst and urination, fatigue, blurred vision, and slow-healing sores, and long-term uncontrolled disease can lead to serious complications involving the heart, kidneys, nerves, and eyes.<sup>1</sup> Major risk factors include obesity, diet, physical inactivity, and age, with disease driven by insulin resistance and impaired insulin secretion.<sup>2</sup>

T2D affects hundreds of millions of adults worldwide<sup>2</sup> with an estimated lifetime risk of 35.65% in the US.<sup>3</sup> This risk is strongly modified by obesity, diet, physical inactivity, and socioeconomic change.<sup>2</sup> Treatment of T2D focuses on lowering blood sugar and reducing long-term complications through a combination of lifestyle changes and medications. Weight loss, improved diet, and increased physical activity are foundational, while medications are often required to improve insulin action or reduce excess glucose. Multiple drugs are commonly used together as the disease progresses. Ongoing management also targets related risk factors such as blood pressure and cholesterol to reduce the risk of heart disease, stroke, and other complications.<sup>2</sup>

## Genetic Risk Score

T2D is shaped by both environmental and genetic factors. Monogenic testing is not available because no single gene causes the condition. Genetic risk scores (GRS), which combine the small effects of many variants into a single score, are currently the only way to estimate genetic risk. Although not diagnostic, a GRS can indicate how likely an individual is to develop the disease.

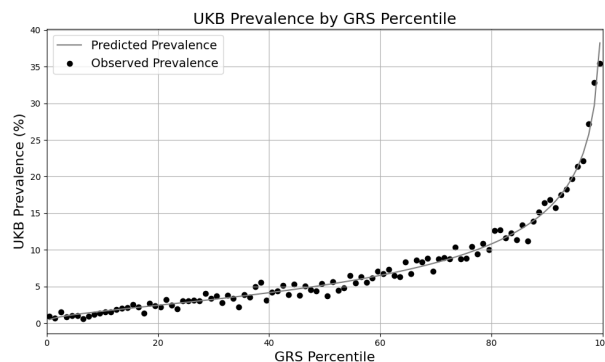
The type 2 diabetes GRS was developed using data from multiple large studies, including the DIAGRAM Consortium, the UK Biobank (UKB), FinnGen, the Million Veteran Program (MVP), and All of Us, consisting of individuals of European ancestry.<sup>4-8</sup> For the UKB, a subset of individuals was held out to use in the validation study. Summary statistics were generated for the UKB and All of Us using the Regenie method.<sup>9</sup> Summary statistics across all studies were then meta-analyzed using METAL.<sup>10</sup> The final GRS was trained using SBayesRC and includes over 7 million variants.<sup>11</sup>

Risk predictions are adjusted to each individual's ancestry, with predictive power decaying as genetic distance from the predominately European training data increases.<sup>12</sup> Orchid considers a GRS meaningfully predictive if individuals

at roughly the 97.7th percentile have an odds ratio (OR) of at least 2. The T2D GRS meets this criterion for all common ancestry groups.

## Evaluation on UK Biobank Data

We evaluated the predictive accuracy of Orchid's type 2 diabetes GRS using the UK Biobank (UKB), a research database of roughly 500,000 genotyped individuals from the United Kingdom.<sup>5</sup> Because some data from the UKB was used to train the GRS, our validation study included only the set of samples that were held out for testing and not included in the GRS training. We also restricted the analysis to individuals of British ancestry and defined type 2 diabetes using the E11.x ICD-10 code, yielding 6,590 cases and 86,074 controls (7.1% 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.



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

Table 1 shows the type 2 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 2 diabetes, supporting the predictive accuracy of the GRS to identify individuals with elevated risk.

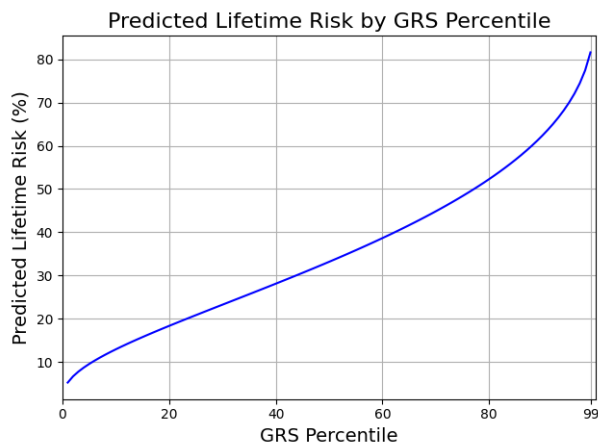
GRS Group	Observed UKB Prevalence	Odds Ratio
Baseline (50th percentile)	4.53%	1.00
Top 10%	22.05%	5.96
Top 5%	27.21%	7.88
Top 1%	35.29%	11.49

**Table 1. Observed prevalence of type 2 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 2 diabetes.

## Estimating Lifetime Risk

The average observed prevalence of type 2 diabetes in the UKB was 7.1%. This is considerably lower than the lifetime prevalence in the US general population, which has been estimated to be approximately 35.65%.<sup>3</sup> 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.<sup>13</sup> Additionally, the observed prevalence in the UKB includes people still living who could develop the disease when they are older, and so does not capture the full lifetime risk of the disease.

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).<sup>14</sup> 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 35.65% estimate.<sup>3</sup>

GRS Percentile	Predicted Lifetime Risk	Relative Risk
50th (baseline)	33.19%	1.00x
95th	70.00%	2.11x
97th	74.44%	2.24x
99th	81.58%	2.46x

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

## Conclusion

In this study, we evaluated our type 2 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 2 diabetes GRS model is available to individuals of all ancestry groups.

## Acknowledgments

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

## References

- Centers for Disease Control and Prevention. Diabetes: Signs and Symptoms. <https://www.cdc.gov/diabetes/signs-symptoms/index.html>. Accessed 2025-12-05.
- R. DeFronzo, E. Ferrannini, L. Groop, et al. Type 2 diabetes mellitus. *Nat Rev Dis Primers*, 1:15019, 2015. doi:10.1038/nrdp.2015.19.
- K. M. V. Narayan, J. P. Boyle, T. J. Thompson, et al. Lifetime risk for diabetes mellitus in the United States. *JAMA*, 290(14):1884–1890, 2003. doi:10.1001/jama.290.14.1884.
- Anubha Mahajan, Daniel Taliun, Matthias Thurner, Neil R Robertson, Jason M Torres, Nigel W Rayner, Andrew J Payne, Valgerdur Steinthorsdottir, Laura J Scott, Niels Grarup, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics*, 50(11):1505–1513, 2018.
- 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.
- FinnGen Consortium. FinnGen documentation of R12 release. 2023. URL: <https://finngen.gitbook.io/documentation/>.

7. J.M. Gaziano, J. Concato, M. Brophy, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *Journal of Clinical Epidemiology*, 70:214–223, 2016. doi:10.1016/j.jclinepi.2015.09.016.
8. All of Us Research Program Investigators. The “All of Us” Research Program. *New England Journal of Medicine*, 381(7):668–676, 2019. doi:10.1056/NEJMsrl809937.
9. J. Mbatchou, L. Barnard, J. Backman, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics*, 53:1097–1103, 2021. doi:10.1038/s41588-021-00870-7.
10. C.J. Willer, Y. Li, and G.R. Abecasis. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17):2190–2191, 2010. doi:10.1093/bioinformatics/btq340.
11. Z. Zheng et al. Leveraging functional genomic annotations and LD structure to improve polygenic prediction. *Nature Communications*, 13:1–12, 2022. doi:10.1038/s41467-022-29849-5.
12. Florian Privé 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.
13. 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. *Am J Epidemiol*, 186:1026–1034, 2017. doi:10.1093/aje/kwx246.
14. N. Chatterjee, J. Shi, M. García-Closas, et al. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*, 17:392–406, 2016. doi:10.1038/nrg.2016.27.

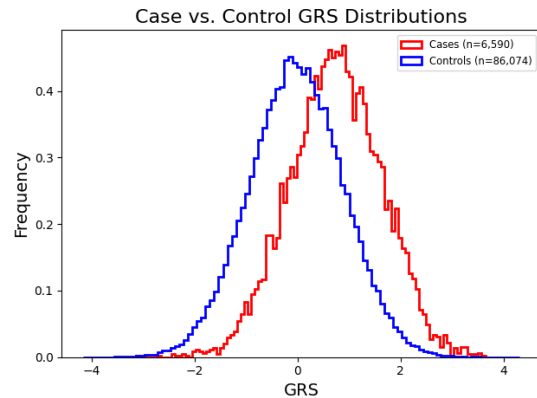
### Supplementary Information

Baseline Risk	OR per SD	OR per 2 SD
33.19%	2.56	6.56

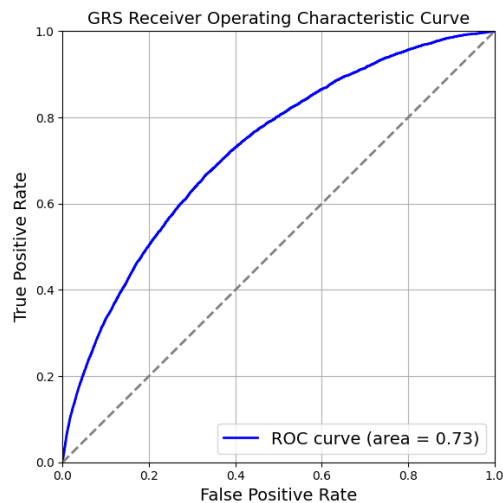
**Table 3. OR per SD.** The baseline risk for an individual with a median GRS, and the predicted OR at one and two SDs, respectively. A GRS must have a predicted OR >2 at 2 SD to be included in Orchid’s clinical reports.

UKB Prevalence	Population Prevalence	Liability R <sup>2</sup>
7.1%	35.65%	23.45%

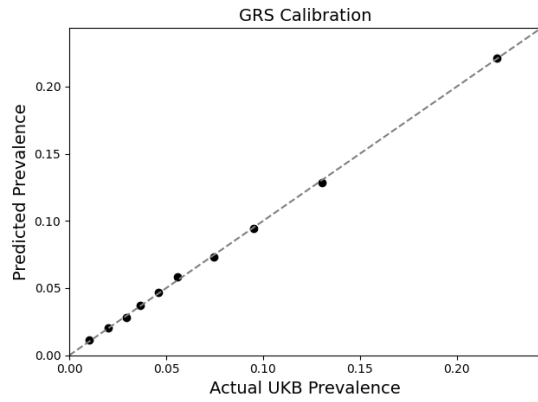
**Table 4. Liability R<sup>2</sup>.** The estimated liability R<sup>2</sup> using a population prevalence of 35.65%.



**Figure 3. GRS histograms.** GRS distributions for cases and controls. Both are approximately normal, with the case distribution shifted noticeably higher compared to the controls.



**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.



**Figure 5. Calibration Curve.** Calibration plot showing observed disease prevalence versus predicted risk across GRS deciles.