Orchid Genetic Risk Score: Type 2 Diabetes

Orchid has developed advanced genetic risk scores (GRS) for a variety of diseases. Here we present our data on our GRS of type 2 diabetes.

Written by the Orchid Team

1. Type 2 Diabetes

Type 2 diabetes is a disease involving high levels of blood sugar, a relative deficiency of insulin, and high levels of insulin resistance [1]. It is a common disease, and the most common form of diabetes, accounting for as many as 90% of cases worldwide [2]. Type 2 diabetes can be caused by obesity or inactivity, but it is substantially influenced by genetics; the heritability of the disease was estimated to be between 61-78% in a study of 34,166 twin pairs from the DISCOTWIN consortium [3].

2. Clinical Impact and Prevalence

More than 30 million Americans (about 9% of the population) are estimated to have type 2 diabetes, and two out of five Americans will eventually develop the condition during their lifetime [4], [5]. Findings from the Swedish National Diabetes Registry suggest the average age of diagnosis to be around 57.5-62.5 years of age [6], but in the United States the typical age of onset has been shrinking dramatically in recent decades, and screening is recommended for Americans aged 45 and older [7], [8]. People with type 2 diabetes often experience no symptoms, but may suffer from increased thirst and urination, weight loss, numbness or tingling in the extremities, or blurred vision [1]. Poorly managed type 2 diabetes is a major risk factor for heart disease, stroke [9], and amputations, and is the leading cause of kidney failure. In some cases, type 2 diabetes may be managed through a strict low-calorie diet [10], but most patients will take one or more blood-sugar lowering medications, such as metformin or insulin [11]. Better control of blood sugar is associated with lower rates of complications [11].

3. Genetic risk score (GRS)

A genetic risk score quantifies the degree to which an individual's genetics increases their likelihood of developing a specific disease. The GRS for type 2 diabetes includes 1,105,354 variants and was developed based on the variants identified in a study that analyzed genomes of 159,208 individuals of European ancestry. The study included 26,676 cases (individuals with type 2 diabetes) and 132,532 controls [12]. The summary statistics from the meta-analysis were then adjusted for linkage disequilibrium using PRScs [13].

Number of variants in genetic risk score	1,105,354
Discovery GWAS(n=159,208)	Cases: 26,676
	Controls: 132,532

Table 1: **Discovery cohort statistics.** Variants in GRS and sample number used in the Type 2 Diabetes GWAS.

4. Performant Type 2 Diabetes risk stratification

4.1. Validated using a large cohort of adults with known type 2 diabetes status

Adults in the 99th percentile of genetic risk have a 20.3 percent absolute risk of type 2 diabetes, compared to a 7.47% risk for the baseline risk, within the UK Biobank cohort. Baseline risk is the prevalence of the disease in the entire reference population.

Importantly, this is different from the lifetime prevalence figure reported above, for two reasons: the UK Biobank cohort has a median age of 58 [14], so many people who will develop diabetes have not done so and the UK Biobank population is somewhat healthier than average [15]. Both of those will lower the diabetes prevalence of the UK Biobank cohort.

In the UK Biobank, cases were identified using self reported type 2 diabetes (UK Biobank field 20002, code 1223) relevant ICD-10 diagnosis and death codes. See the supplementary table for full details. In the validation, prevalence of the disease increased with GRS. We restricted our analysis to self-reported British whites whose genetic ancestry matched. With our phenotype definition there were 30,507 cases of type 2 diabetes and 378,013 controls.

Elevated Genetic Risk Definition	Prevalence	Odds ratio
Baseline Prevalence	7.5%	1
Top 5% of distribution	16.5%	2.6
Top 3% of distribution	17.6%	2.88
Top 1% of distribution	20.3%	3.15
Top 0.5% of distribution	21.4%	3.37

Table 2: **Prevalence and odds ratios for elevated genetic risk subgroups.** Adults at the tail end of GRS distribution were at an elevated risk for and had higher odds for the disease in comparison to the baseline rate of 7.5%.

4.2. Comparing baseline and elevated risk for type 2 diabetes

Adults in the 99th percentile of genetic risk develop type 2 diabetes at 2.71 times the baseline rate. Baseline rate is the
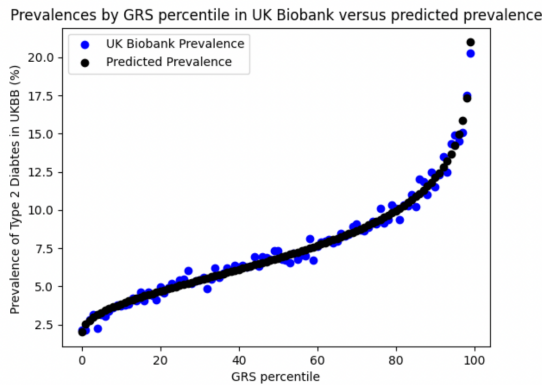
Figure 1: Risk gradient for type 2 diabetes. Each blue dot represents a percentile of Genetic Risk Score, with its percent prevalence in UK Biobank self-reported White British in the y-axis. The black line represents the predicted prevalence from a logistic regression derived from the data.

prevalence of the disease in the entire reference population.

The odds ratio represents the relative odds of an event in the presence or absence of a specific factor. In this case, it can be used to show the strength of association between a genetic risk score and a disease. The odds ratio for developing **type 2 diabetes** while being in the top 1% of the elevated risk population is 3.15.

5. Comparison to Published Benchmarks

Orchid's model achieves comparable stratification performance with an AUC of 0.704 compared to the benchmark of 0.698.

We compared the performance of our model as validated on the UK Biobank with the performance of the best model in Khera et al. To make a comparison of models, we restricted our validation sample to those in Phase II of the UK Biobank release, as in Khera et. al. In the first column, we give the results for our predictor with the phenotype as described above. In the second, we report the metrics for the best-performing predictor in Khera et. al using the same phenotype as ours.

Model	Orchid	Reference[1]
AUC of model with GRS, age, and PCs[2]	0.703	0.698[3]

Table 3: Accuracy metric comparison. Our model compared to reference.

References

[1] Orchid Team. Type 2 diabetes. Available at `https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/type-2-diabetes`, 2022. [cited 7 Jan 2022].

[2] Orchid Team. Type 2 diabetes. Available at `https://www.who.int/health-topics/diabetes`, 2022. [cited 7 Jan 2022].

[3] Gonneke Willemsen, Kirsten J Ward, Christopher G Bell, Kaare Christensen, Jocelyn Bowden, Christine Dalgård, Jennifer R Harris, Jaakko Kaprio, Robert Lyle, Patrik KE Magnusson, et al. The concordance and heritability of type 2 diabetes in 34,166 twin pairs from international twin registers: the discordant twin (discotwin) consortium. *Twin Research and Human Genetics*, 18(6):762–771, 2015.

[4] Orchid Team. Type 2 diabetes.24 nov 2021. Available at `https://www.cdc.gov/diabetes/basics/type2.html`, 2022. [cited 7 Jan 2022].

[5] Edward W Gregg, Xiaohui Zhuo, Yiling J Cheng, Ann L Albright, KM Venkat Narayan, and Theodore J Thompson. Trends in lifetime risk and years of life lost due to diabetes in the usa, 1985–2011: a modelling study. *The lancet Diabetes & endocrinology*, 2(11):867–874, 2014.

[6] Naveed Sattar, Araz Rawshani, Stefan Franzén, Aidin Rawshani, Ann-Marie Svensson, Annika Rosengren, Darren K McGuire, Björn Eliasson, and Soffia Gudbjörnsdottir. Age at diagnosis of type 2 diabetes mellitus and associations with cardiovascular and mortality risks: findings from the swedish national diabetes registry. *Circulation*, 139(19):2228–2237, 2019.

[7] Richelle J Koopman, Arch G Mainous, Vanessa A Diaz, and Mark E Geesey. Changes in age at diagnosis of type 2 diabetes mellitus in the united states, 1988 to 2000. *The Annals of Family Medicine*, 3(1):60–63, 2005.

[8] Karly Pippitt, Marlana Li, and Holly E Gurgle. Diabetes mellitus: screening and diagnosis. *American family physician*, 93(2):103–109, 2016.

[9] Orchid Team. American diabetes association. 10. cardiovascular disease and risk management: Standards of medical care in diabetes-2021. diabetes care., 2021.

[10] Shaminie J Athinarayanan, Rebecca N Adams, Sarah J Hallberg, Amy L McKenzie, Nasir H Bhanpuri, Wayne W Campbell, Jeff S Volek, Stephen D Phinney, and James P McCarter. Long-term effects of a novel continuous remote care intervention including nutritional ketosis for the management of type 2 diabetes: a 2-year non-randomized clinical trial. *Frontiers in endocrinology*, 10:348, 2019.

[11] Orchid Team. American diabetes association. 10. cardiovascular disease and risk management: Standards of medical care in diabetes-2021. diabetes care., 2021.

[12] Robert A Scott, Laura J Scott, Reedik Mägi, Letizia Marullo, Kyle J Gaulton, Marika Kaakinen, Natalia Pervjakova, Tune H Pers, Andrew D Johnson, John D Eicher, et al. An expanded genome-wide association study of type 2 diabetes in europeans. *Diabetes*, 66(11):2888–2902, 2017.

[13] Tian Ge, Chia-Yen Chen, Yang Ni, Yen-Chen Anne Feng, and Jordan W Smoller. Polygenic prediction via bayesian regression and continuous shrinkage priors. *Nature communications*, 10(1):1776, 2019.

[14] Orchid Team. Type 2 diabetes. Available at `https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id=21022`, 2022. [cited 22 Feb 2022].

[15] Anna Fry, Thomas J Littlejohns, Cathie Sudlow, Nicola Doherty, Ligia Adamska, Tim Sprosen, Rory Collins, and Naomi E Allen. Comparison of sociodemographic and health-related characteristics of uk biobank participants with those of the general population. *American journal of epidemiology*, 186(9):1026–1034, 2017.

[16] Amit V Khera, Mark Chaffin, Krishna G Aragam, Mary E Haas, Carolina Roselli, Seung Hoan Choi, Pradeep Natarajan, Eric S Lander, Steven A Lubitz, Patrick T Ellinor, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics*, 50(9):1219–1224, 2018.

6. Appendix

Disease case identification and number of cases in UK Biobank

Phenotype	ICD-10 Codes	Self-Report Codes	Cases in UK Biobank (white British)
Prostate cancer	C61, D075	1044	13,806
Type 2 diabetes	E11.1-9	1223	30,507
Coronary artery disease	I2104,I219,I220, I221,I228,I232 I233,I235,I236 I238,I249,I252	1075	22,451
Breast cancer	C5.0-9, D05.0, D059	1002	18,588
Inflammatory bowel disease	K51	1461,1462 ,1463	5,959
Atrial fibrillation	148.0-4,148.9	1471,1483	22,472
Schizophr enia	F20.0-9, F21, F23.0-3, F23.8	1289	1,376
Alzheimer's disease	F00.0-2, F00.9, G30.0-1,8-9.	1263	2,547
Celiac disease	K900	1456	3,253
Bipolar disease	F31	1291	1,855
Type 1 diabetes	*	*	421

Table 4: Supplementary Table: How each disease case is defined in evaluating genetic risk scores in the UK Biobank

*Type 1 diabetes was defined as a combination the following inclusion and exclusion criteria:

- Self-diagnosed diabetes (any type)

- No self-diagnosed Type 2 diabetes

- Age of diabetes onset between 0 and 20 years

- Started insulin within one year of diagnosis of diabetes