Orchid Genetic Risk Score: Prostate Cancer

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

Written by the Orchid Team

---

1. Prostate Cancer

Prostate cancer is a form of cancer that affects the prostate, a gland in the male reproductive system [1]. Like other types of cancer, it is caused by cells which begin to divide abnormally [2]. Genetics are implicated in the development of prostate cancer, the heritability of which is estimated to be between 52% to 63% based on an analysis of almost 50,000 twin pairs drawn from the Nordic Twin Study of Cancer [3].

2. Clinical Impact and Prevalence

Prostate cancer affects roughly 3.2 million American men, and 12.5% will be diagnosed at some point during their lifetime [4], making it the most common non-skin cancer in men [5]. The average age of diagnosis is 66 [6]. Typical symptoms of prostate cancer include frequent and difficult urination, pain in and around the groin, and pain in adjacent parts of the body such as the pelvis and hips [7]. Prostate cancer is very common and about 80% of men diagnosed with prostate cancer will live at least five years after their diagnosis [8]. Many cases of prostate cancer do not require active treatment, but some may require medical treatment by an oncologist, surgery, and/or radiotherapy [7].

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 prostate cancer includes 1,106,020 variants and was developed by selecting the genome-wide significant SNPs identified in a study that analyzed genomes of about 74,849 individuals of European ancestry [9]. The study included 46,939 cases (individuals with prostate cancer) and 27,910 healthy controls [9]. The summary statistics from the meta-analysis were then adjusted for linkage disequilibrium using PRScs.

Number of variants in genetic risk score	1,106,020
Discovery GWAS(n=74,849)	Cases: 46,939
	Controls: 27,910

Table 1: **Discovery cohort statistics**. Variants in GRS and sample number used in the prostate cancer GWAS.

4. Performant prostate cancer risk stratification

4.1. Validated using a large cohort of real world men with known prostate cancer status

Within the UK Biobank, men in the 99th percentile of genetic risk have a 30.5 percent absolute risk of prostate cancer, compared to a baseline prevalence of 7.4 percent. This is different from the lifetime prevalence figure reported above mainly because the UK Biobank cohort has a median age of 58 [10], so the majority of men who will develop prostate cancer have not yet done so, or it has not been detected yet. This will lower the prevalence of prostate cancer in the UK Biobank cohort.
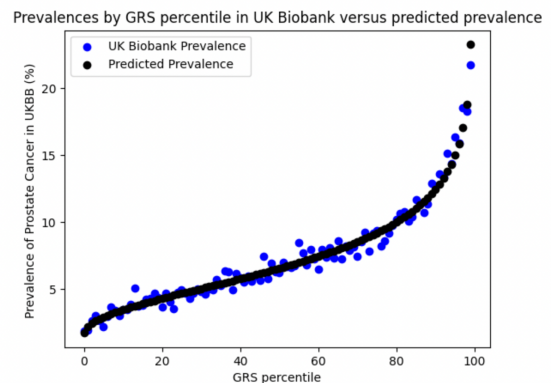

Figure 1: Risk gradient for prostate cancer. Each blue dot represents a percentile of Genetic Risk Score, with its percent prevalence in UK Biobank self-reported White British in the y-axis. The black line represents the predicted prevalence from a logistic regression derived from the data.

In the UK Biobank, cases were identified using self-reported prostate cancer (UK Biobank field 20002) relevant ICD-10 diagnosis and death codes. See supplementary table for more details. In the validation, prevalence of the disease increased with GRS. Using our phenotype definition, in our sample of self-reported white British men, there were 13,806 cases of prostate cancer and 173,859 controls (prevalence of 7.36%).

People at the tail end of GRS distribution were at an elevated risk for developing the disease in comparison to the baseline rate. For each GRS threshold, we also computed the odds ratio (with the baseline rate as a comparison group). Baseline rate is the prevalence of the disease in the entire reference population.

Elevated Genetic Risk Definition	Prevalence	Odds ratio
Baseline Prevalence	7.36%	1
Top 5% of distribution	22.49%	3.65
Top 3% of distribution	24.85%	4.16
Top 1% of distribution	30.53%	5.53
Top 0.5% of distribution	33.12%	6.24

Table 2: **Average prevalence and odds ratio for elevated genetic risk subgroups.** Men at the tail end of GRS distribution were at an elevated risk for developing the disease in comparison to men overall, who had a prevalence of 7.36%.

4.2. Comparing baseline and elevated risk for prostate cancer

Men in the 99th percentile of genetic risk develop prostate cancer at 4.15 times the baseline rate, with an odds ratio of 5.53. Baseline rate is the prevalence of the disease in the entire reference population.

5. Comparison to Published Benchmarks

Orchid's model achieves comparable stratification performance with an AUC of 0.683 compared to the benchmark at 0.662.

We compared the performance of our model as validated on the UK Biobank with the performance of the best model in Jia et al [11]. To make a comparison of models, we restricted our validation sample to those in Phase II of the UK Biobank release, as in Jia 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 Jia et al. from their paper.

Model	Orchid	Reference[1]
AUC, GRS alone[2]	0.683	0.662

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

References

[1] Orchid Team. What is prostate cancer? 23 aug 2021, 2022. [cited 4 Jan 2022].

[2] Orchid Team. What causes prostate cancer?, 2022. [cited 4 Jan 2022].

[3] Jacob B Hjelmborg, Thomas Scheike, Klaus Holst, Axel Skytthe, Kathryn L Penney, Rebecca E Graff, Eero Pukkala, Kaare Christensen, Hans-Olov Adami, Niels V Holm, et al. The heritability of prostate cancer in the nordic twin study of cancer. *Cancer Epidemiology, Biomarkers & Prevention*, 23(11):2303–2310, 2014.

[4] Orchid Team. National cancer institute. cancer of the prostate - cancer stat factsnational cancer institute. cancer of the prostate - cancer stat facts, 2022. [cited 4 Jan 2022].

[5] Vera Nelen. Epidemiology of prostate cancer. *Prostate cancer*, pages 1–8, 2007.

[6] Orchid Team. American cancer society. key statistics for prostate cancer., 2022. [cited 4 Jan 2022].

[7] Orchid Team. How is prostate cancer treated? 13 sep 2021, 2022. [cited 4 Jan 2022].

[8] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2018. *CA: a cancer journal for clinicians*, 68(1):7–30, 2018.

[9] Fredrick R Schumacher, Ali Amin Al Olama, Sonja I Berndt, Sara Benlloch, Mahbubl Ahmed, Edward J Saunders, Tokhir Dadaev, Daniel Leongamornlert, Ezequiel Anokian, Clara Cieza-Borrella, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nature genetics*, 50(7):928–936, 2018.

[10] Orchid Team. Data-field 21022, 2022. [cited 22 Feb 2022].

[11] Lu Y Jia G, Long J Wen W, and et al. Liu Y, Tao R. Evaluating the utility of polygenic risk scores in identifying high-risk individuals for eight common cancers. jnci cancer spectr. *Cancer Epidemiology, Biomarkers & Prevention*, 2020.

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
Schizophrenia	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