

Orchid Genetic Risk Score: Breast Cancer

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

Written by the Orchid Team

1. Breast Cancer

Breast cancer is a cancer that forms in the cells of the breasts [1]. Like other types of cancer, it is caused by cells which begin to divide abnormally. There is a substantial genetic component to breast cancer; individuals with a first degree female relative (mother, sister, or daughter) with the disease are twice as likely to be diagnosed than individuals without a family history [2]. The genetics of breast cancer illustrate both monogenic and polygenic components: there are rarer monogenic variants (such as BRCA1/2) which substantially increase breast cancer risk [3, 4] but there is a polygenic component to breast cancer risk as well, which can sometimes confer risk comparable to monogenic variants [5]. These polygenic scores are still useful for further determining risk among BRCA1/2 carriers [6]. The heritability of breast cancer is approximately 31%, based on an analysis of 200,000 twin pairs drawn from the population registries of Denmark, Norway, Sweden, and Finland [7].

2. Clinical Impact and Prevalence

Breast cancer affects over 3.8 million Americans, and 12% of women in the US will be diagnosed in their lifetime [1], [8]. It is usually diagnosed between 55 and 64, with the median age of diagnosis being 62 [9] [10]. The first noticeable symptoms of breast cancer are changes to the breast including a breast lump or thickening that feels different from the surrounding tissue, change in the size, shape or appearance of a breast, and peeling, scaling, or flaking of the pigmented area of skin surrounding the nipple (areola) or breast skin [1]. The prognosis of breast cancer depends on the stage the disease has progressed to when it is diagnosed. Mammography screening can help with earlier detection [11], with different professional associations recommending different start dates for screening. Several treatments such as medications (including chemotherapy), surgery, and radiotherapy may be prescribed by an oncologist [1].

3. Genetic risk score (GRS)

Orchid's genetic risk score quantifies the degree to which an individual's genetics increases their likelihood of developing a specific disease. The GRS for breast cancer includes 1,106,006 variants and was developed based on the variants identified in

a study that analyzed genomes of 228,951 individuals of European ancestry [12]. The study included 122,977 cases (individuals with breast cancer) and 105,974 healthy controls [8]. The summary statistics from the meta-analysis were then adjusted for linkage disequilibrium using PRSCs [13, 14].

3.1. Limitations of the GRS

5% to 10% of all breast cancer cases are linked to rare genetic variants in the BRCA1 and BRCA2 genes. Because most of these variants are individually very rare, they are not included in genome-wide association studies for breast cancer and thus they are not captured by our genetic risk score. A low breast cancer GRS is therefore no guarantee for not having an elevated genetic risk for breast cancer. However, more than 90% of all breast cancer cases are influenced by genetic risk factors outside of the BRCA1 and BRCA2 genes, which are captured by our breast cancer GRS. Any individual with a family history of early-onset breast cancer, multiple relatives with breast cancer, or has other risk factors [15] should seek guidance from a genetic counselor with expertise in cancer genetics, since they may carry a rare BRCA1/2 variant not captured in Orchid's GRS.

Number of variants in genetic risk score	1,106,006
Discovery GWAS(n=228,951)	Cases: 122,977 Controls: 105,974

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

4. Performant breast cancer risk stratification

4.1. Validated using a large cohort of real world women with known breast cancer status

Women in the 99th percentile of genetic risk have a 15.05 percent prevalence of breast cancer, compared to the baseline prevalence of 8.42%. This is lower than the lifetime prevalence of breast cancer reported for US women, as a result of the median age of the UK Biobank (58), which means that many women who will eventually develop breast cancer have not yet done so. This biases the prevalence of breast cancer within the UK Biobank downwards.

Prevalences by GRS percentile in UK Biobank versus predicted prevalence

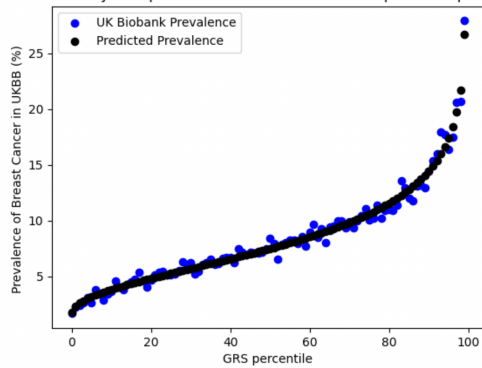


Figure 1: Risk gradient for breast cancer. Each blue dot represents a percentile of Genetic Risk Score, with its percent prevalence in UK Biobank self-reported White British in the y-axis. The black line represents the predicted prevalence from a logistic regression derived from the data.

4.2. Validation in the UK Biobank

Validation in UK Biobank. In the UK Biobank, cases were identified using self-reported breast cancer (UK Biobank field 20002, code 1002), relevant ICD-9/ICD-10 diagnosis, cancer and death codes. See Supplementary Table for more details. In the validation, prevalence of breast cancer increased with GRS. We restricted our analysis to self-reported British white women whose genetic ancestry matched their self-identification. With our phenotype definition there were 18,588 cases of breast cancer and 202,267 controls.

Elevated Genetic Risk Definition	Prevalence	Odd ratio
Baseline Prevalence	8.42%	1
Top 5% of distribution	20.65%	2.83
Top 3% of distribution	23.11%	3.27
Top 1% of distribution	27.93%	4.22
Top 0.5% of distribution	31.13%	4.92

Table 2: Disease prevalence in elevated genetic risk subgroups for White British individuals.

4.3. Identification of women at 3 times the baseline risk of breast cancer

Women in the 97th percentile of genetic risk develop breast cancer at 3.32 times the baseline rate. The odds ratio for women in the 99th percentile was 4.22. Baseline rate is the prevalence of the disease in the entire reference population.

5. Comparison to Published Benchmarks

Orchid’s model achieves greater stratification performance with an AUC of 0.670 compared to the benchmark of 0.644.

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	0.670	0.644

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

References

- [1] Orchid Team. Breast cancer., 2022. [cited 4 Jan 2022].
- [2] Orchid Team. What are the risk factors for breast cancer? 21 sep 2021, 2022. [cited 4 Jan 2022].
- [3] James D Fackenthal and Olufunmilayo I Olopade. Breast cancer risk associated with brca1 and brca2 in diverse populations. *Nature Reviews Cancer*, 7(12):937–948, 2007.
- [4] Karoline B Kuchenbaecker, John L Hopper, Daniel R Barnes, Kelly-Anne Phillips, Thea M Mooij, Marie-José Roos-Blom, Sarah Jervis, Flora E Van Leeuwen, Roger L Milne, Nadine Andrieu, et al. Risks of breast, ovarian, and contralateral breast cancer for brca1 and brca2 mutation carriers. *Jama*, 317(23):2402–2416, 2017.
- [5] Nasim Mavaddat, Kyriaki Michailidou, Joe Dennis, Michael Lush, Laura Fachal, Andrew Lee, Jonathan P Tyrer, Ting-Huei Chen, Qin Wang, Manjeet K Bolla, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *The American Journal of Human Genetics*, 104(1):21–34, 2019.
- [6] Daniel R Barnes, Matti A Rookus, Lesley McGuffog, Goska Leslie, Thea M Mooij, Joe Dennis, Nasim Mavaddat, Julian Adlard, Munaza Ahmed, Kristiina Aittomäki, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of brca1 and brca2 pathogenic variants. *Genetics in Medicine*, 22(10):1653–1666, 2020.
- [7] Lorelei A Mucci, Jacob B Hjelmberg, Jennifer R Harris, Kamila Czene, David J Havelick, Thomas Scheike, Rebecca E Graff, Klaus Holst, Sören Möller, Robert H Unger, et al. Familial risk and heritability of cancer among twins in nordic countries. *Jama*, 315(1):68–76, 2016.
- [8] Orchid Team. Breastcancer.org. u.s. breast cancer statistics., 2021. [cited 4 Jan 2022].
- [9] Orchid Team. National cancer institute. cancer of the breast (female) - cancer stat facts., 2022. [cited 4 Jan 2022].
- [10] Orchid Team. American cancer society. breast cancer statistics., 2022. [cited 4 Jan 2022].
- [11] Kirsten Bibbins-Domingo, David C Grossman, Susan J Curry, Karina W Davidson, John W Epling, Francisco AR García, Matthew W Gillman, Diane M Harper, Alex R Kemper, Alex H Krist, et al. Screening for colorectal cancer: Us preventive services task force recommendation statement. *Jama*, 315(23):2564–2575, 2016.
- [12] 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.
- [13] Orchid Team. Faucon a, samaroo j and ge t, davis lk and tao r, cox nj, et al., 2021.
- [14] 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.
- [15] Douglas K Owens, Karina W Davidson, Alex H Krist, Michael J Barry, Michael Cabana, Aaron B Caughey, Chyke A Doubeni, John W Epling, Martha Kubik, C Seth Landefeld, et al. Risk assessment, genetic counseling, and genetic testing for brca-related cancer: Us preventive services task force recommendation statement. *Jama*, 322(7):652–665, 2019.

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	I48.0-4, I48.9	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