Orchid Genetic Risk Score: Celiac Disease

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

Written by the Orchid Team

1. Celiac Disease

Celiac disease is an autoimmune condition triggered by gluten ingestion [1]. This leads to characteristic symptoms in affected individuals, such as diarrhea, abdominal bloating, and in some cases, nutrient deficiencies from malabsorption. The heritability of celiac disease is around 75% based on analysis of 513 twins with a diagnosis of celiac disease [2]. It is also more common among 1st-degree relatives of celiac patients, by up to 4-5 times [3].

2. Clinical Impact and Prevalence

Conservative estimates put the prevalence of celiac disease in the U.S. at about 1 in 133 or around 0.75% of the population [4]. Celiac disease can develop later in life, but a recent study found that most cases develop in childhood, with the majority occurring before age 10 [5]. The mainstay of managing celiac disease is consuming a gluten-free diet [6]. In addition, since celiac disease can cause micronutrient deficiencies, patients with celiac disease may want to seek advice from a dietician with experience in managing celiac disease.

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 celiac disease includes **42** variants and was developed based on the variants identified in a study that analyzed genomes of about **24,269** individuals of European ancestry. The study included **12,041** cases (individuals with celiac disease) and **12,228** healthy controls [7].

Our celiac disease GRS has some special characteristics relative to our usual GRS. For this disease, one specific locus, HLA DQ, confers a disproportionate share of genetic risk. That is, celiac disease risk follows an *oligogenic* model, not a classic *polygenic* model. For that reason, the resulting GRS is not normally distributed.

4. Performant celiac disease risk stratification

4.1. Validated using a large cohort of individuals with known disease status

Within the UK Biobank cohort, adults in the 99th percentile of genetic risk have a 10.11% absolute risk of celiac disease,

Number of variants in genetic risk score	42
Discovery GWAS(n=24,269)	Cases: 12,041 Controls: 12,228

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

compared to 0.80% for the baseline rate. Baseline rate is the prevalence of the disease in the entire reference population.
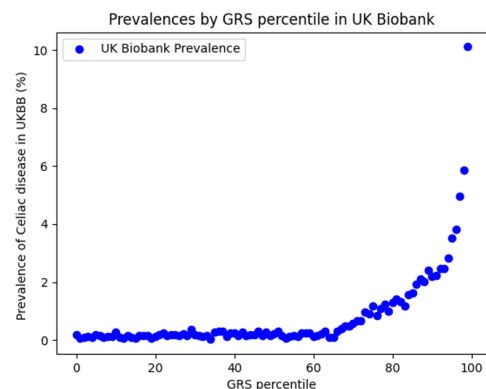

Figure 1: Risk gradient for celiac disease. Each blue dot represents a percentile of Genetic Risk Score, with its percent prevalence in UK Biobank self-reported White British in the y-axis.

Validation in UK Biobank. In the UK Biobank, cases were identified using self-reported celiac disease and relevant ICD-9/ICD-10 diagnosis. See the supplementary table in the appendix for more details. In the validation, prevalence of celiac disease increased with GRS. We restricted our analysis to self-reported British individuals whose genetic ancestry matched their self-identification. With our phenotype definition there were 3,253 cases of celiac disease and 408,099 controls.

4.2. Identification of adults at 14 times the baseline risk of celiac disease

Adults in the 99th percentile of genetic risk have a 10.11% risk of developing celiac disease. The odds ratio for adults in the 99th percentile of genetic risk was 14.01.

Elevated Genetic Risk Definition	Prevalence	Odd ratio
Baseline Prevalence	0.80%	1.0
Top 5% of distribution	5.66%	7.47
Top 3% of distribution	6.98%	9.34
Top 1% of distribution	10.11%	14.01
Top 0.5% of distribution	11.26%	15.81

Table 2: **Prevalence and odds ratios of celiac disease in 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 0.80%.

5. Comparison to Published Benchmarks

Orchid's model achieves comparable stratification performance with an AUC of 0.83 compared to the benchmark at 0.84 reported in Sharp et al [7]. 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 Sharp et. al using the numbers reported in their paper.

Celiac Disease GRS	Orchid	Reference[1]
AUC	0.83	0.84

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

References

[1] Giacomo Caio, Umberto Volta, Anna Sapone, Daniel A Leffler, Roberto De Giorgio, Carlo Catassi, and Alessio Fasano. Celiac disease: a comprehensive current review. *BMC medicine*, 17:1–20, 2019.

[2] Ralf Kuja-Halkola, Benjamin Lebwohl, Jonas Halfvarson, Cisca Wijmenga, Patrik KE Magnusson, and Jonas F Ludvigsson. Heritability of non-hla genetics in coeliac disease: a population-based study in 107 000 twins. *Gut*, 65(11):1793–1798, 2016.

[3] Yasar Dogan, Serap Yldrmaz, and Ibrahim Hanifi Özercan. Prevalence of celiac disease among first-degree relatives of patients with celiac disease. *Journal of pediatric gastroenterology and nutrition*, 55(2):205–208, 2012.

[4] Orchid Team. Beyond celiac–celiac disease: Fast facts. Available at `https://www.beyondceliac.org/celiac-disease/facts-and-figures/`, 2022. [cited 7 Jan 2022].

[5] Dong F Liu E and et al Barón AE. High incidence of celiac disease in a long-term study of adolescents with susceptibility genotypes. Available at `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5533620/`, 2017. [cited 4 Jan 2022].

[6] Alberto Rubio-Tapia, Ivor D Hill, Ciarán P Kelly, Audrey H Calderwood, and Joseph A Murray. American college of gastroenterology clinical guideline: diagnosis and management of celiac disease. *The American journal of gastroenterology*, 108(5):656, 2013.

[7] Seth A Sharp, Samuel E Jones, Robert A Kimmitt, Michael N Weedon, Anne M Halpin, Andrew R Wood, Robin N Beaumont, Seema King, David A van Heel, Patricia M Campbell, et al. A single nucleotide polymorphism genetic risk score to aid diagnosis of coeliac disease: a pilot study in clinical care. *Alimentary pharmacology & therapeutics*, 52(7):1165–1173, 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
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