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

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.

<table>
<thead>
<tr>
<th>Celiac Disease GRS</th>
<th>Orchid</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.83</td>
<td>0.84</td>
</tr>
</tbody>
</table>

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

References


6. Appendix

Disease case identification and number of cases in UK Biobank

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>ICD-10 Codes</th>
<th>Self-Report Codes</th>
<th>Cases in UK Biobank (white British)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate cancer</td>
<td>C61, D075</td>
<td>1044</td>
<td>13,806</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>E11.1-9</td>
<td>1223</td>
<td>30,507</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>I2104,1219,1220, I221,1228,1232 I233,1235,1236 I238,1249,1252</td>
<td>1075</td>
<td>22,451</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>C5.0-9, D05.0, D059</td>
<td>1002</td>
<td>18,588</td>
</tr>
<tr>
<td>Inflammatory bowel disease</td>
<td>K51</td>
<td>1461,1462,1463</td>
<td>5,959</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>I48.0-4,148.9</td>
<td>1471,1483</td>
<td>22,472</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>F20.0-9, F21, F23.0-3, F23.8</td>
<td>1289</td>
<td>1,376</td>
</tr>
<tr>
<td>Alzheimer’s disease</td>
<td>F00.0-2, F00.9, G30.0-1,8-9.</td>
<td>1263</td>
<td>2,547</td>
</tr>
<tr>
<td>Celiac disease</td>
<td>K900</td>
<td>1456</td>
<td>3,253</td>
</tr>
<tr>
<td>Bipolar disease</td>
<td>F31</td>
<td>1291</td>
<td>1,855</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>*</td>
<td>*</td>
<td>421</td>
</tr>
</tbody>
</table>

Note: 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

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