Cross-phenotypic Association Testing Using Biobank PheWAS Results

Yongwen Zhuang
PhD student, University of Michigan

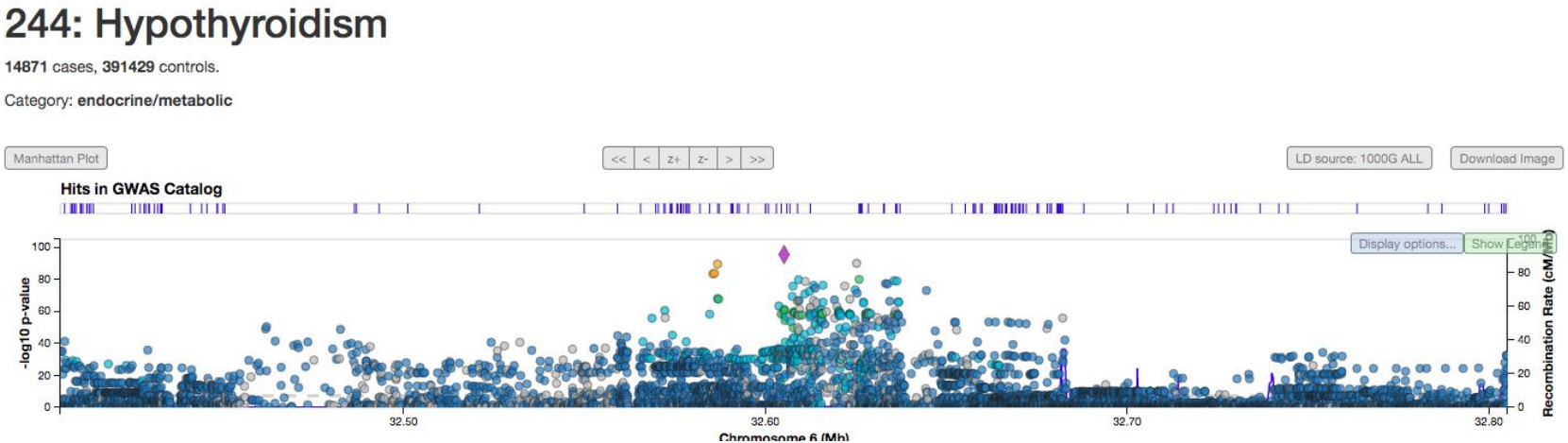
Motivation

- Mapping the relationship across phenotypic traits is an essential way to understand the **genetic mechanisms underlying common human diseases**
- **Phenome-wide association studies (PheWAS)** identify genetic variants that are significantly associated with multiple phenotypes, yet **do not provide information on the significance** of association among phenotypes

1. **Can we quantitatively describe the relatedness between phenotype pairs?**
2. **How to efficiently perform the test at a large scale?**

Data

1. **Phenotypes**
   A total of 1,403 binary traits in 408,961 UK Biobank samples from white British participants
2. **PheWAS test results**



Method

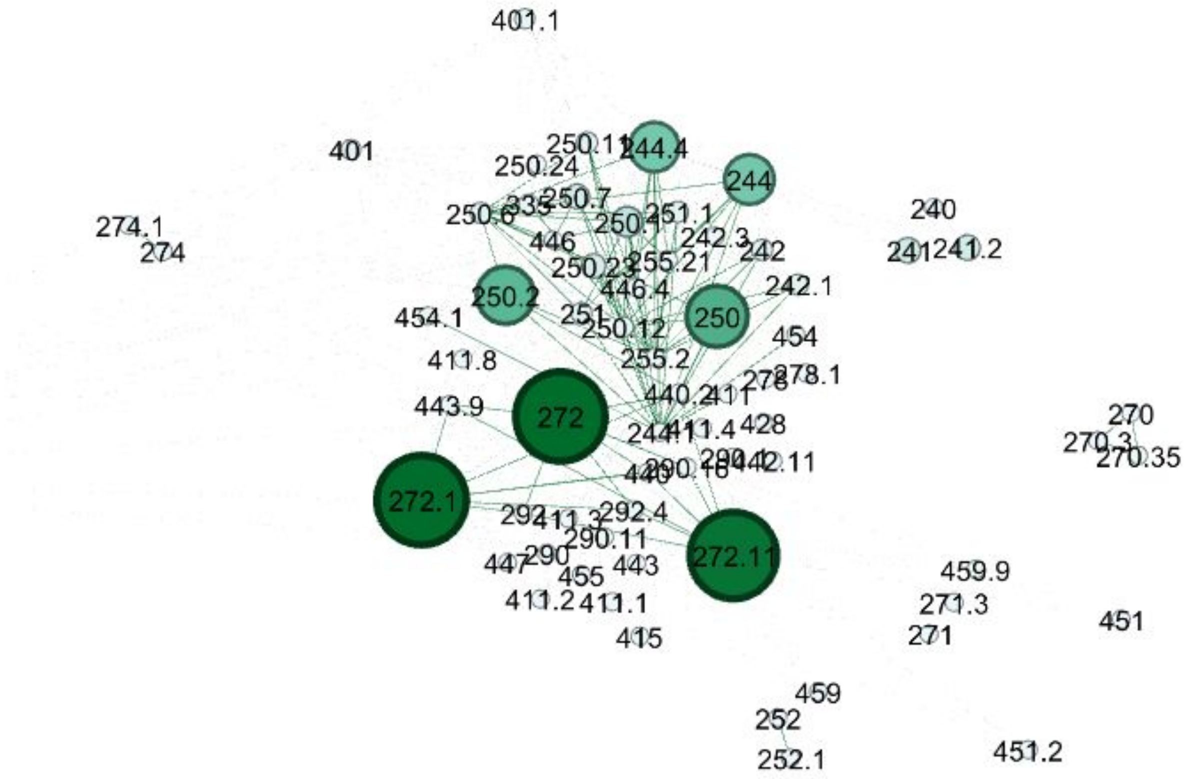
1. **Common clump identification**
   Clump significant SNPs in PheWAS results within 250kb genetic distance that are in linkage disequilibrium (correlated) with the index SNP
2. **Association test**

   Given N SNPs, K are associated with Pheno 1, n are associated with Pheno 2,

   $$p := \Pr(k \text{ SNPs are associated with both phenotypes}) = \frac{\binom{n-k}{N-K}^{k}_{K}}{\binom{n}{N}}$$

3. **Parallelization**
   For each pair of phenotypes, the above process was executed simultaneously on a multiprocessor cluster;
   Results were aggregated into a single file

Result



401.1
401
250.11
244.4
250.24
244
250.6
335
250.7
250.1
251.1
274.1
274
446
242.3
242
240
241
241.2
250.23
255.21
242.1
250.2
446.4
454.1
251
250
250.12
454
411.8
255.2
278
278.1
440.2
411
443.9
272
244.1
411.4
428
270
270.3
270.35
290.1
442.11
440
290.16
272.1
292
411.3
292.4
290.11
272.11
447
290
443
455
459.9
411.2
411.1
271.3
451
415
271
459
252
252.1
451.2