COVID-19 GWAB
Functional Analysis of COVID-19 GWAS Data with Network-based Boosting

Seungbyn Baek[1], Sunmo Yang[1], and Insuk Lee[1]*

[1] Department of Biotechnology, College of Life Science & Biotechnology, Yonsei University, Seoul 03722, Korea

I. Introduction

- Large scale Genome-wide association studies (GWAS) with COVID-19 patients have been published by COVID-19 Host Genetics Initiative (HGI) and other research groups.
- While GWAS data alone can provide valuable information about disease-related genotypes, it is often difficult to use for translational research or to make phenotypic connections to diseases.
- GWAS data can be enhanced by network-based boosting methods, such as GWAB.
- Updated version of GWAB specialized in analyzing COVID-19 GWAS data is in development progress.

II. GWAB
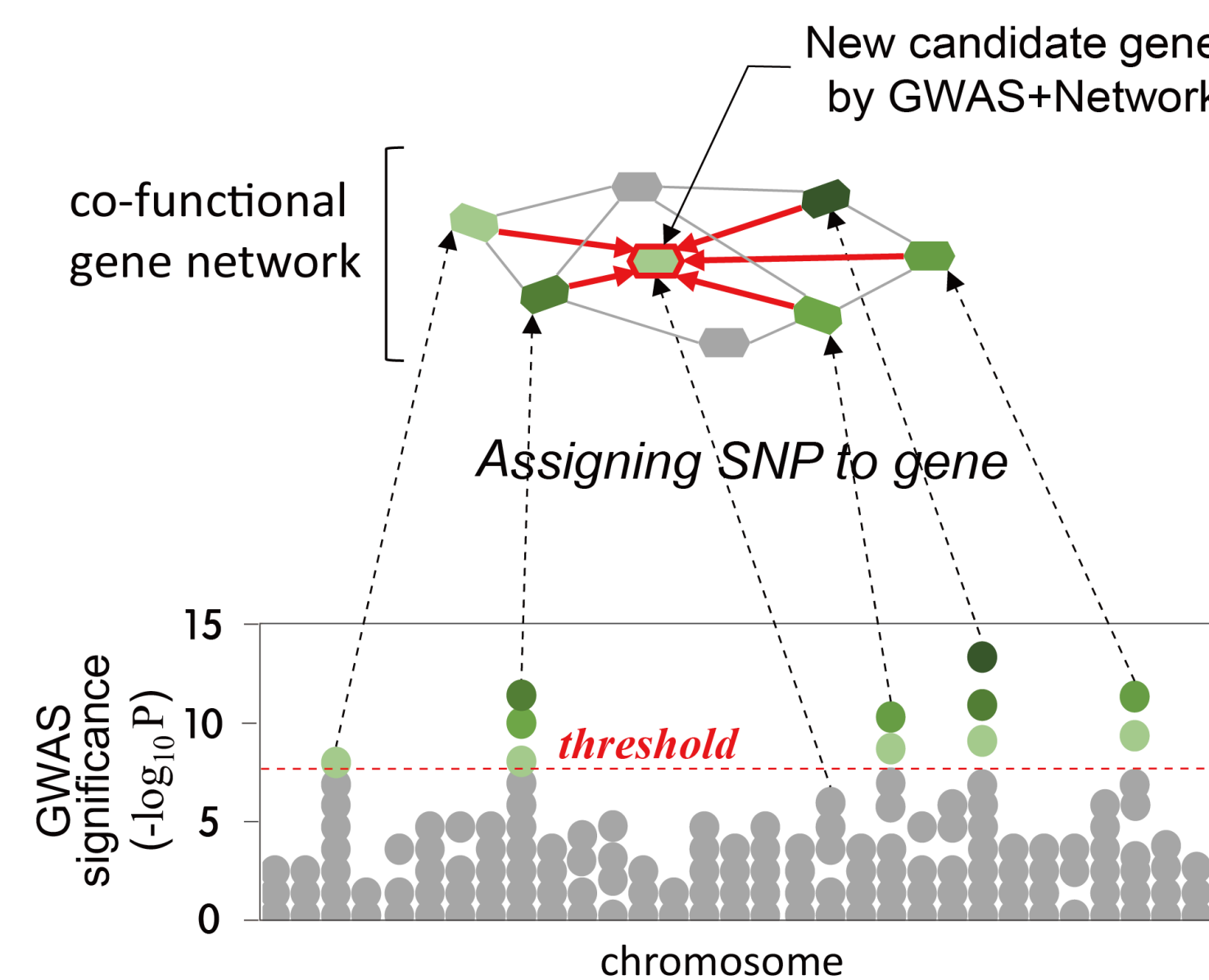
- Genome-Wide Association Boosting (GWAB) is a web server for the network based boosting of human GWAS data[1].
- GWAB uses co-functional gene network called HumanNet[2].
- GWAB enhances GWAS results by suggesting new candidate genes connected to neighbors genes with optimal p-value threshold within networks

## III. COVID GWAB Pipeline (development in progress)
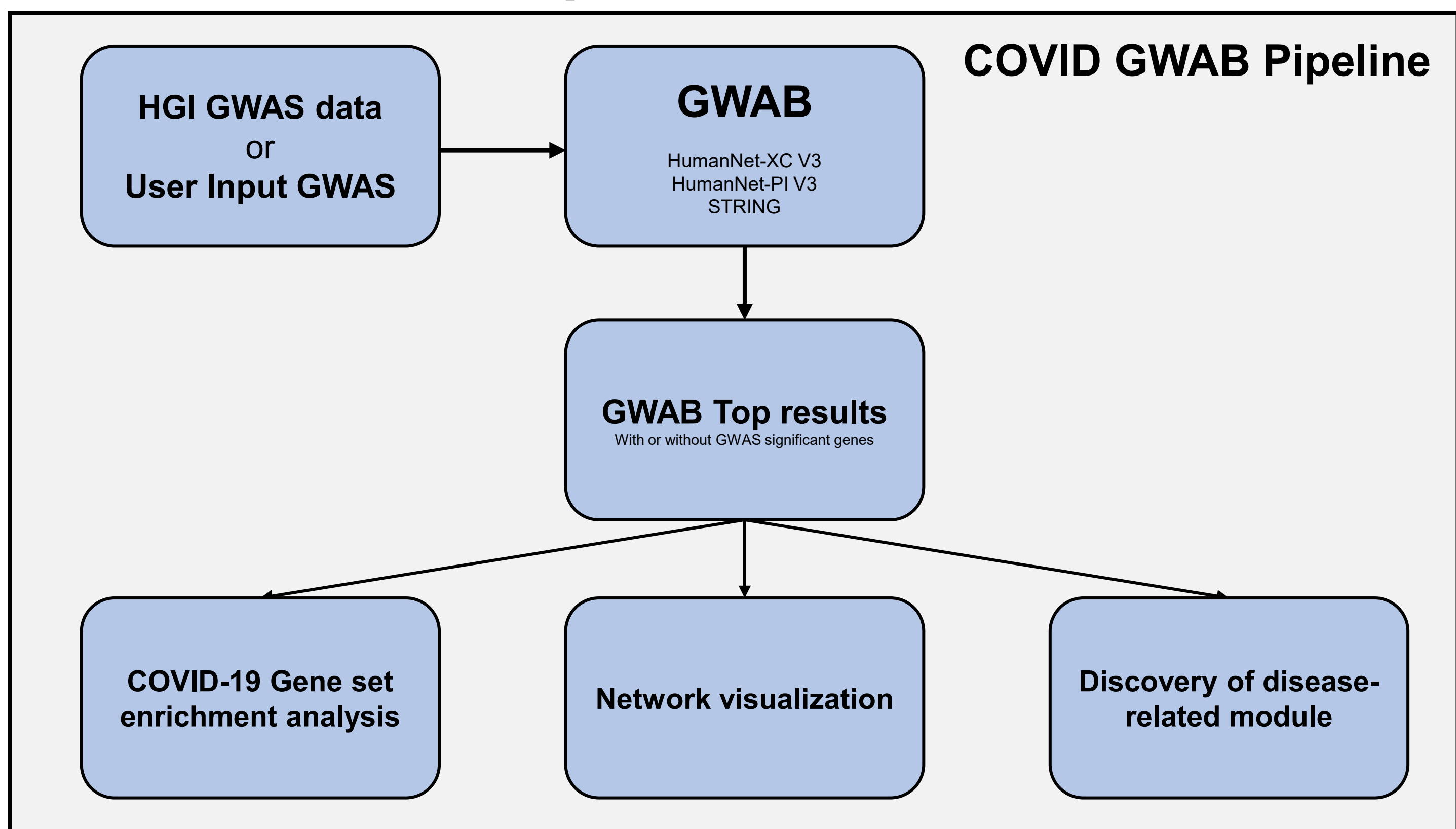
- Either COVID-19 HGI GWAS data or user input COVID-19 GWAS data can be used as input for COVID GWAB
- Networks available for GWAB include updated version of HumanNet (version 3, publication in progress) and STRING
- Top candidate genes after GWAB boosting can be additionally filtered based on several categories (GWAS p-value, gene families, etc.)
- COVID-19 gene sets from single-cell and bulk RNA-seq data are available as validation sets for gene set enrichment analysis
- Network visualization and sub-network module analysis are available
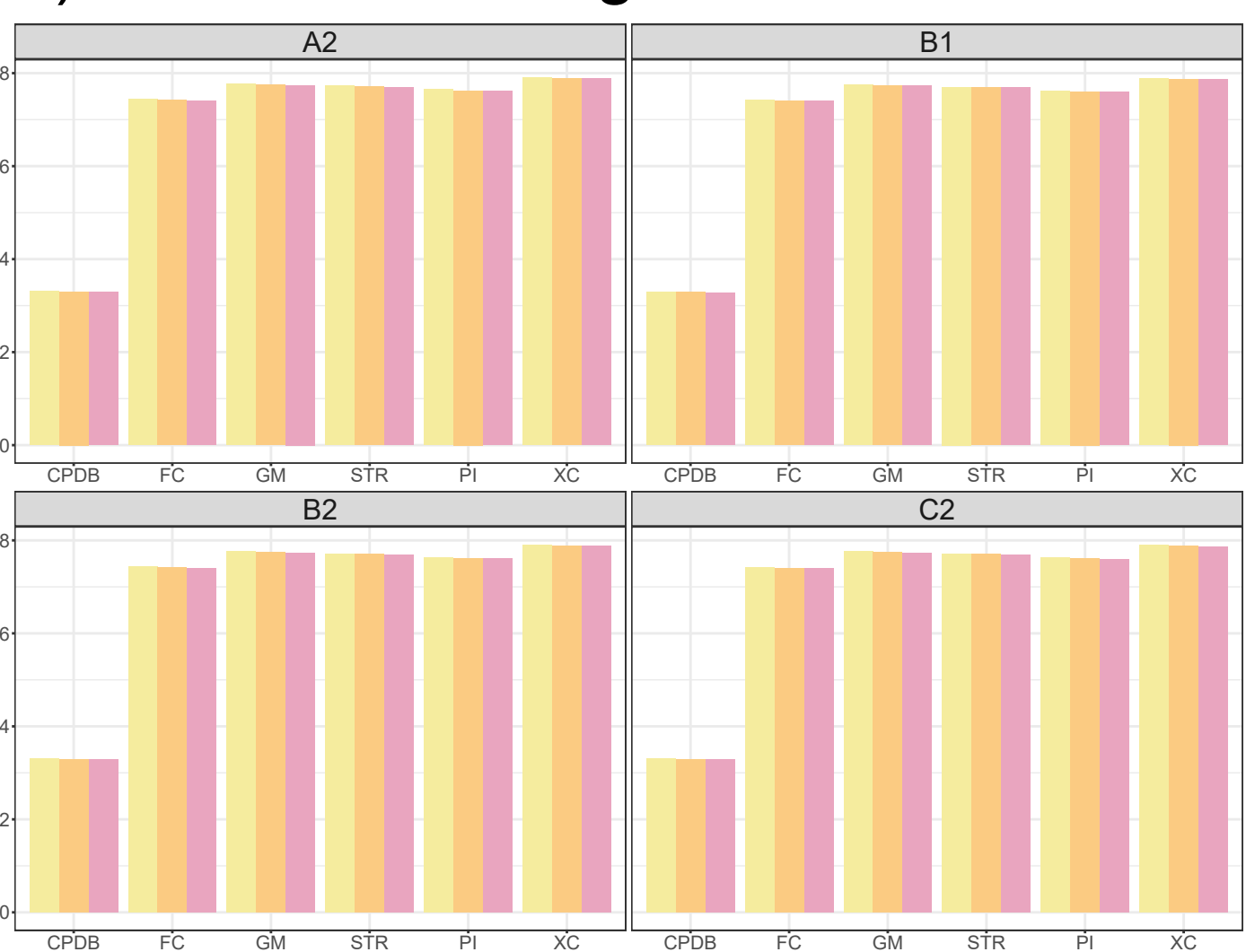
IV. GWAB with various networks
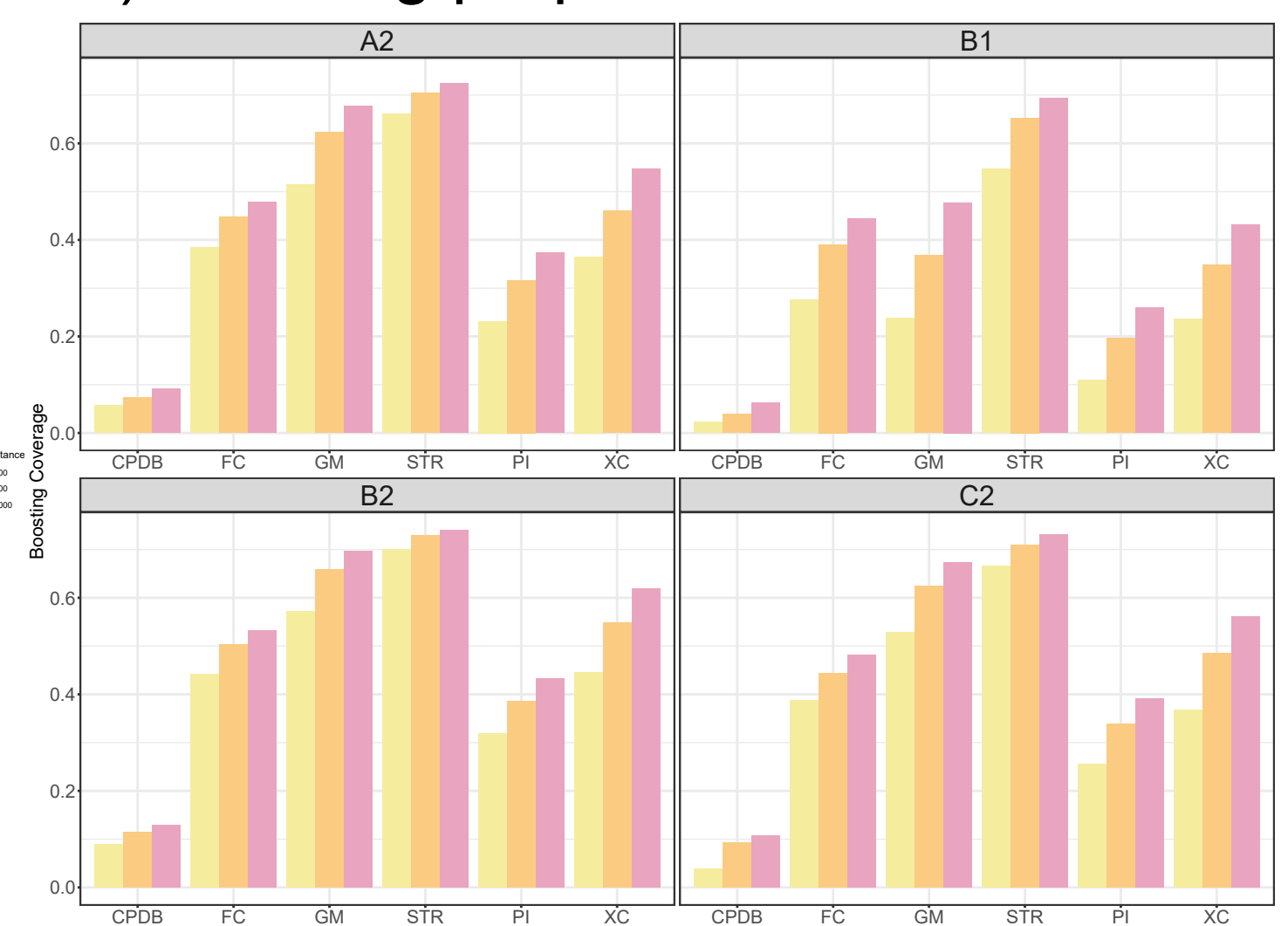
A) Network weight distribution
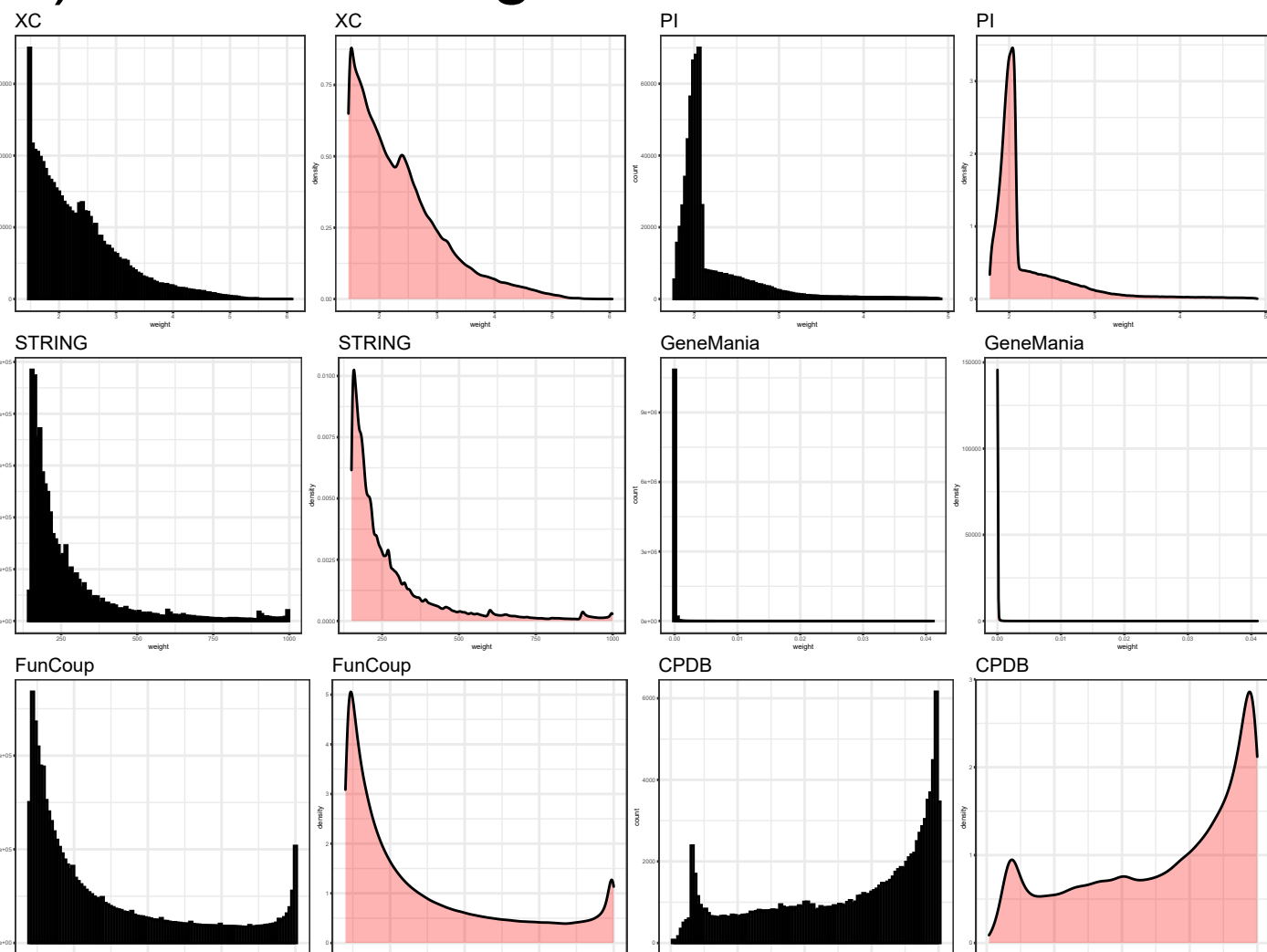
### B) Significant GWAS genes in top results
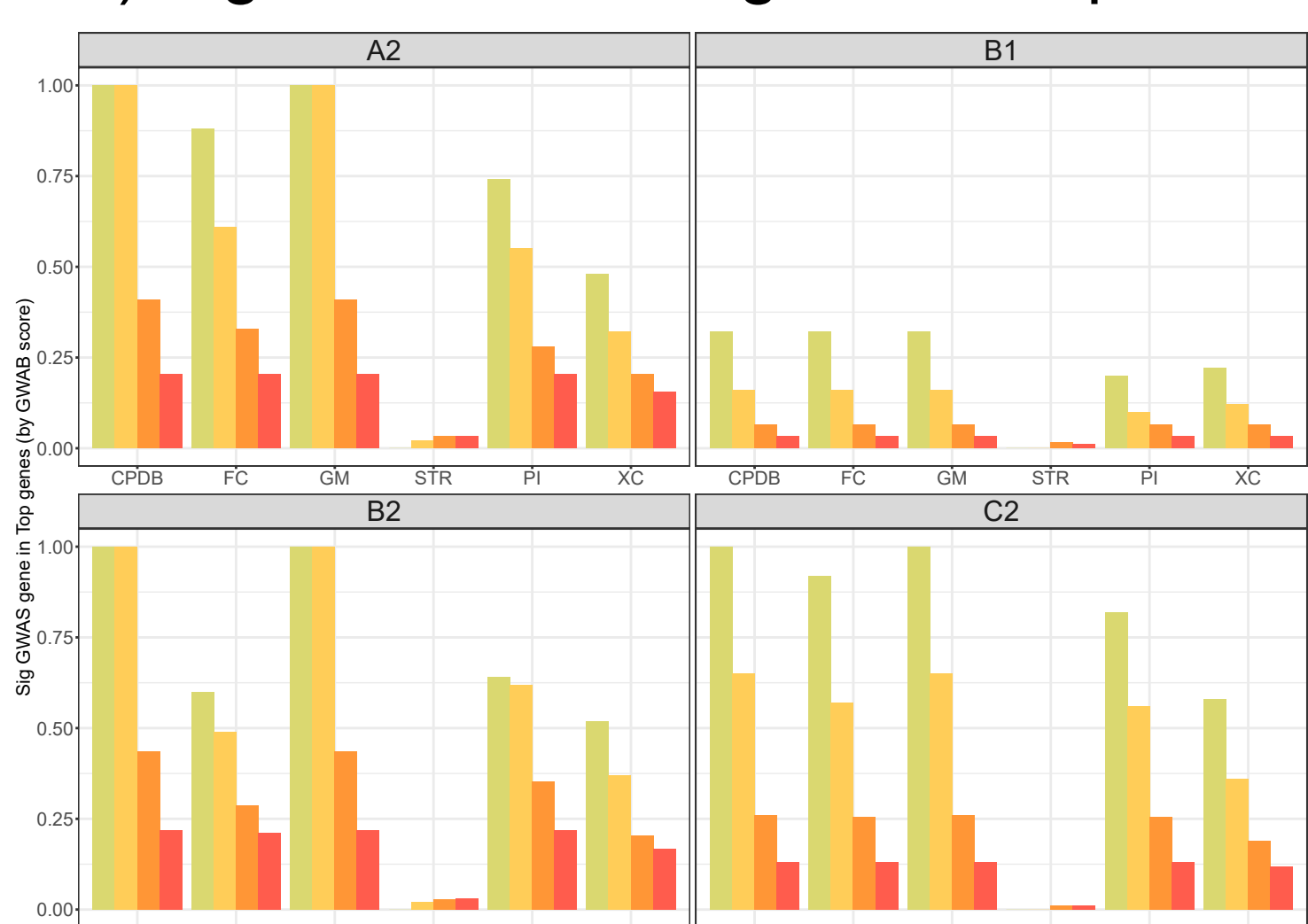

C) Network coverage

D) Boosting proportion

- A) GWAB was orginally built for HumanNet which utilizes likelihood score for network weight. Since other networks often use different approaches for network weight, distributions of the weights were checked for similarity with HumanNet. Networks with similar distributions tend to result in similar boosting patterns as HumanNet
- B) Proportion of GWAS significant genes (5×10^{-5} p-value) in top 50/100/250/500 GWAB genes. Too many GWAS significant genes often mean under-boosting results and too few mean possible false positives.
- C) Proportion of GWAB genes in each network's nodes.
- D) Proportion of GWAB genes that are boosted by each network.

V. COVID-19 Geneset overlap

A) COVID vs Healthy (COVID genes)
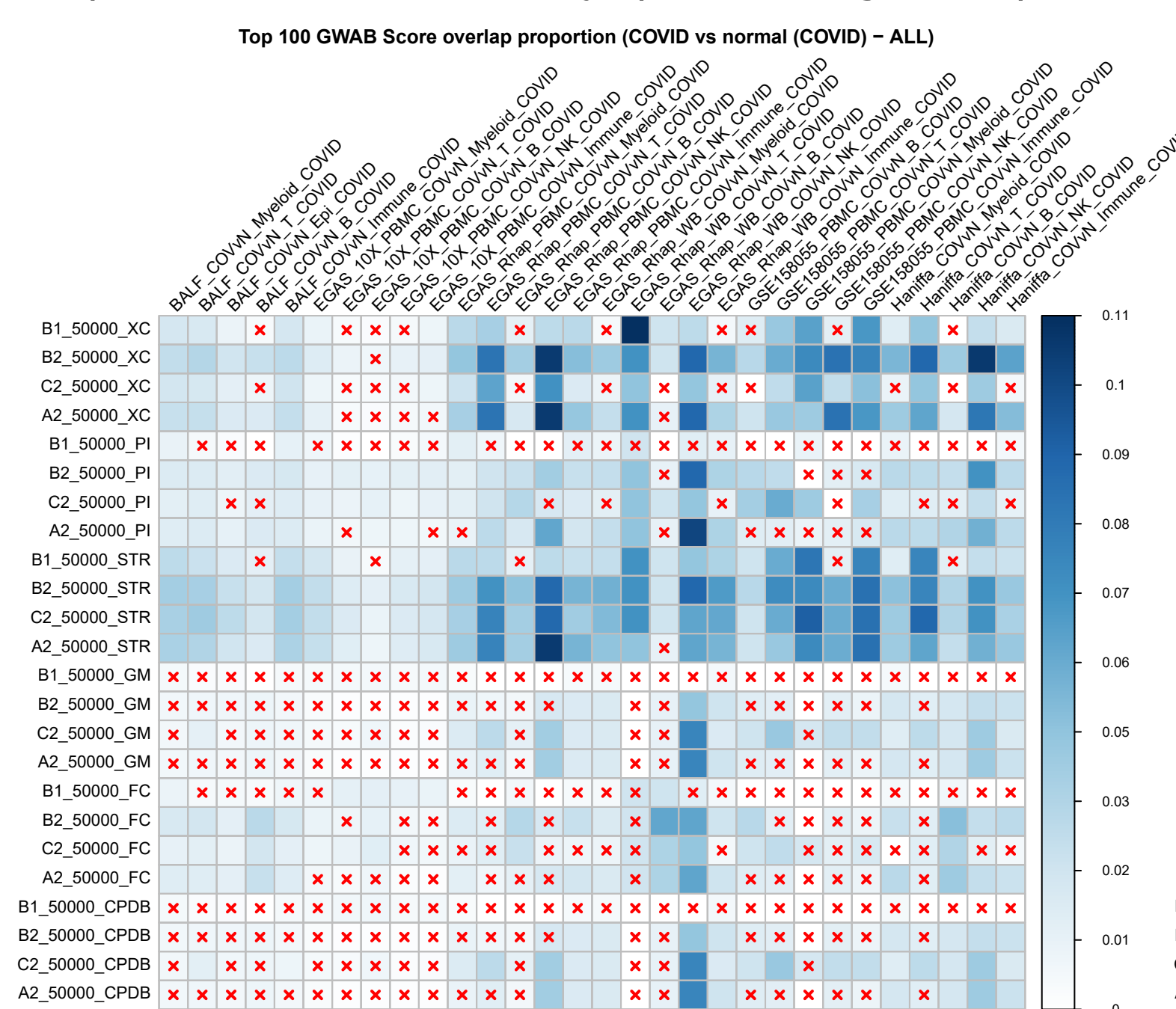
B) COVID vs Healthy (Healthy genes)
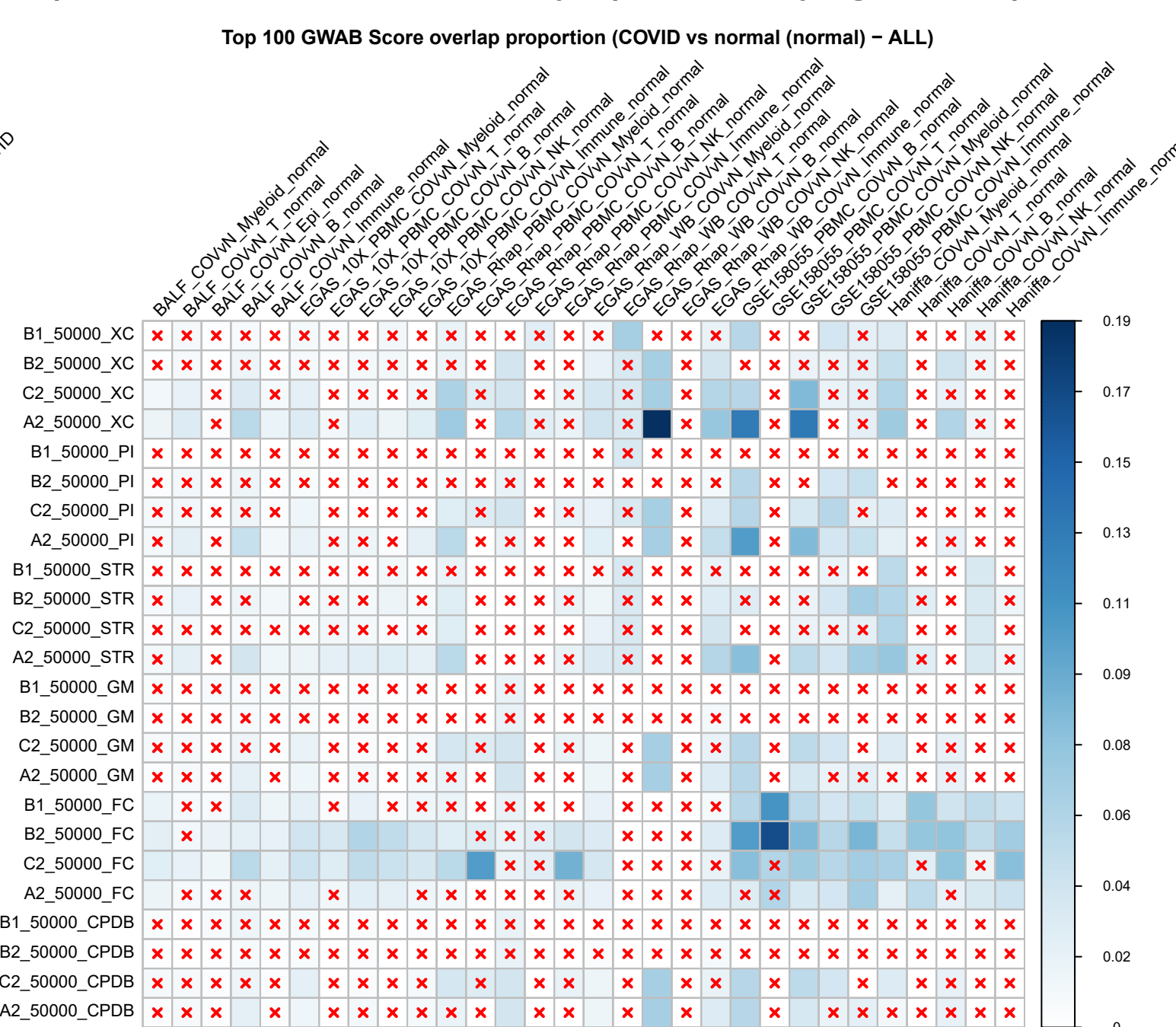
- Single-cell RNA-seq data for COVID-19 for BALF[3,4] and PBMC[4,5,6] were collected for three phenotypic comparisons: severe and mild, severe and healthy, COVID (severe and mild) and healthy.
- Major celltypes (B, NK, T, myeloid, epithelial) are provided
- GWAB results with HumanNet_XC and STRING showed the most overlaps with single-cell data for COVID-specific genes and not much overlaps for healthy-specific genes
- Other networks (GeneMania, FunCoup, CPDB) did not show as much overlaps as HumanNet or STRING

VI. GWAB networks
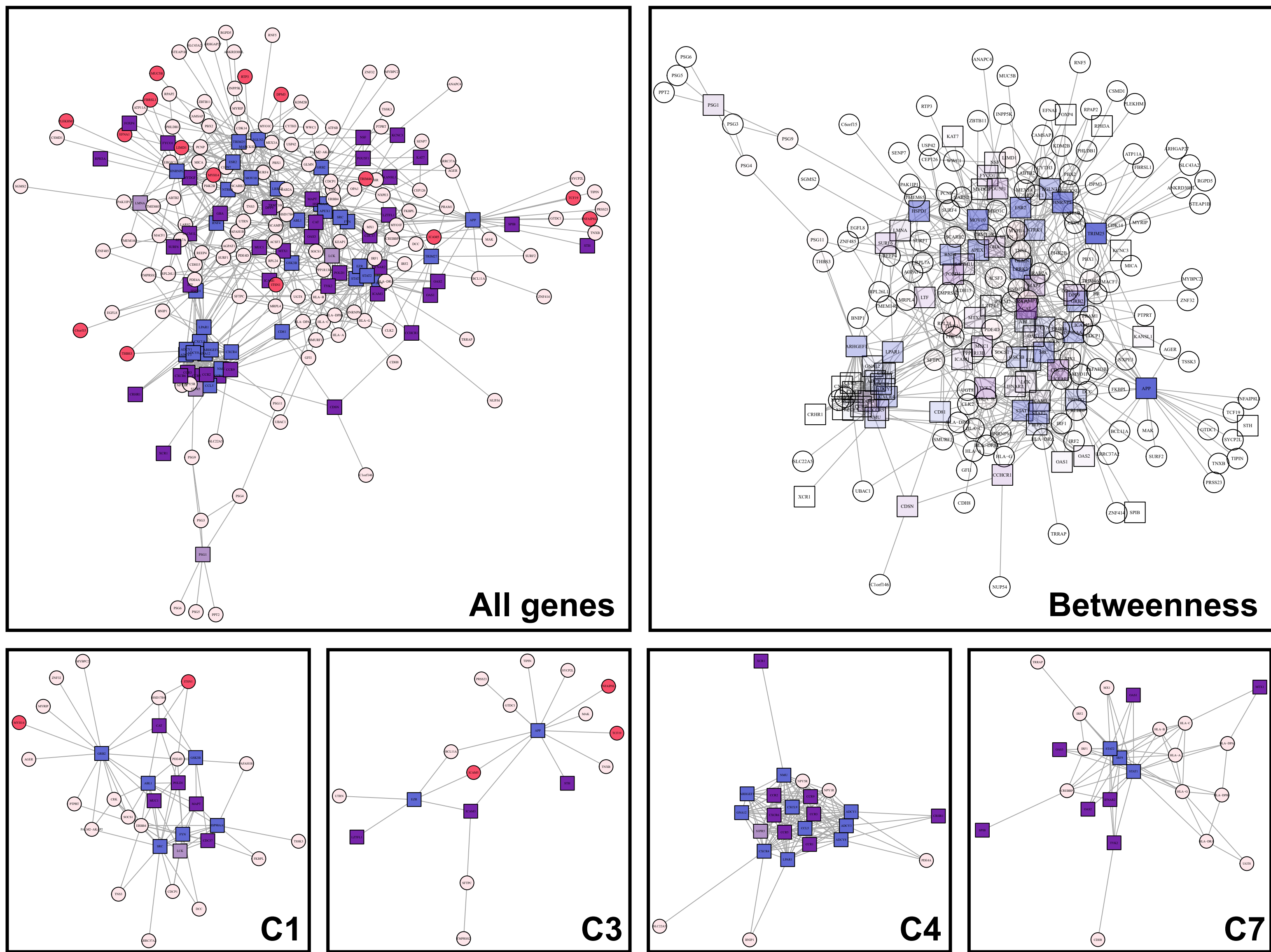
- The example network is constructed with top 100 GWAB results with COVID-19 HGI B2 (Hospitalized covid vs. population) and HumanNet-PI.
- The GWAB top results are indicated as squares while boosting genes (over GWAB p-value threshold) are indicated as circles.
- Blue nodes are new candidates genes (below GWAS p-value threshold) and purple nodes are genes in top GWAB results and GWAS-significant.
- Betweenness centrality was visualized with transparency with more solid colors indicating more betweenness
- Sub-network modules can be analyzed with network clustering. C1,3,4,7 are results of louvain clustering methods and each cluster can be interpreted as disease-related modules.
- Genes in C1 cluster are mostly related to kinase activities while genes in C4 are related to various chemokines.

Reference
1. Shim, Jung Eun, et al. "GWAB: a web server for the network-based boosting of human genome-wide association data." Nucleic acids research 45.W1 (2017): W154-W161.
2. Hwang, Sohyun, et al. "HumanNet v2: human gene networks for disease research." Nucleic acids research 47.D1 (2019): D573-D580.
3. Liao, Mingfeng, et al. "Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19." Nature medicine 26.6 (2020): 842-844.
4. Ren, Xianwen, et al. "COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas." Cell 184.7 (2021): 1895-1913.
5. Schulte-Schrepping, Jonas, et al. "Severe COVID-19 is marked by a dysregulated myeloid cell compartment." Cell 182.6 (2020): 1419-1440.
6. Stephenson, Emily, et al. "Single-cell multi-omics analysis of the immune response in COVID-19." Nature medicine 27.5 (2021): 904-916.