June 30, 2020

Transcriptome-wide association study for COVID19-HGI Freeze 3 phenotypes using S-PrediXcan

Gita A Pathak, Frank R Wendt, Renato Polimanti
gita.pathak@yale.edu, frank.wendt@yale.edu, renato.polimanti@yale.edu
ORCID: GAP (0000-0003-3943-0895); FRW (0000-0002-2108-6822); RP (0000-0003-0745-6046)

Department of Psychiatry, Yale School of Medicine, Yale University, New Haven, CT, USA
Veteran Affairs Connecticut Healthcare System, West Haven, CT, USA

SUMMARY: Variation in the human genome is expected to determine part of the individual susceptibility to COVID-19 severity and outcomes. The differences in the frequency of the genetic variants with respect to the phenotype-of-interest have small yet cumulative effect on the human transcriptome. In a transcriptome-wide association study (TWAS), the effects of genetic variants are aggregated for gene expression and tested for association with the phenotype-of-interest, reducing the multiple-testing burden with respect to a single-variant association analysis. Here, we performed TWAS using S-PrediXcan with GTEx v8 MASHR (multivariate adaptive shrinkage in R) model for lung and whole blood tissue. The MASHR models are provided at the developers host site (https://github.com/hakyimlab/MetaXcan/blob/master/README.md). The GTEx contains genotype-gene-expression data for 54 tissues from 838 donors (https://www.gtexportal.org/home/).

DESCRIPTION: We performed TWAS based on the genome-wide association data for each of the six phenotypes released for COVID19-HGI Freeze3 on June 28, 2020 (https://www.covid19hg.org/results/). The S-PrediXcan (alternatively known as MetaXcan) framework consists of two prediction models for GTEx v8; we used the MASHR-based model for deriving eQTL values. The MASHR model is biologically informed, with fine mapped variables (1). The variants and LD was used for EUR ancestry. The Bonferroni-corrected significance threshold accounting for the number of phenotypes, genes, and tissues tested is p-value = 2.77e-7. Significant associations were observed for *BET1L* (whole blood; phenotype-B1_V2 p-value = 2.33e-7) and *CCR9* (whole blood; phenotype-B2_V2 p-value = 1.00e-14 | phenotype-C2_V2 p-value=9.10e-9). The detailed report is available in the attached xlsx file.

ATTACHED FILE CONTENTS:

Filename: COVID-freeze3_MetaXcan_PathakG.xlsx

Phenotype Legend: The Excel workbook has 12 sheets, i.e. two tissues for each phenotype. Each sheet is named as -phenotype_model_tissue name. The sheets with significant or suggestive associations are colored in orange, and rows with significant observations are highlighted in yellow. The phenotypes are as follows:

- A2_V2: very severe respiratory confirmed covid vs. population
- B1_V2: hospitalized covid vs. not hospitalized covid
- B2_V2: hospitalized covid vs. population
- C1_V2: covid vs. lab/self-reported negative
- C2_V2: covid vs. population
- D1_V2: predicted covid from self-reported symptoms vs. predicted or self-reported non-covid

Table Legend:

- gene: a gene's id: as listed in the Tissue Transcriptome model. Ensemble Id is provided for most gene model releases. May also indicate an intron id for splicing model releases.
- gene_name: gene name as listed by the Transcriptome Model, typically HUGO. It can also be an intron id.
- zscore: S-PrediXcan association result for the indicated gene, typically HUGO.
- effect_size: S-PrediXcan association effect size for the indicated gene. Can only be computed when beta from the GWAS is used.
- pvalue: P-value of the effect size statistic.
- n_snps_used: number of SNPs from GWAS used in S-PrediXcan analysis
- n_snps_in_cov: number of SNPs in the covariance matrix
- n_snps_in_model: number of SNPs in the model
- var_g: variance of gene expression, calculated as W' * G * W (where W is the vector of SNP weights in a gene model, W' is its transpose, and G is the covariance matrix)
- best_gwas_p: the highest p-value from GWAS SNPs used in this model
- largest_weight: the largest (absolute value) weight in this model

Bibliography

1. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nature Communications. 2018. p. 1825.