

Uncovering predisposition to COPD with predicted gene expression (eQTL)



Angeline Yasodhara¹, M. Aziz Mezlini^{1,2}, Michael H. Cho³, Craig P. Hersh³, Anna Goldenberg^{1,2}



¹University of Toronto; ²Genetics & Genome Biology, SickKids Research Institute; ³Brigham and Women's Hospital, Harvard Medical School, Boston

Introduction

Facts & Figures:

1. Chronic Obstructive Pulmonary Disease (COPD): lung conditions with increased breathlessness, e.g. emphysema, chronic bronchitis, refractory (non-reversible) asthma, and some bronchiectasis.

2. In 2015, ~3 million deaths accounted by COPD (~ 5% of all deaths in that year)¹

3. Causes of COPD:

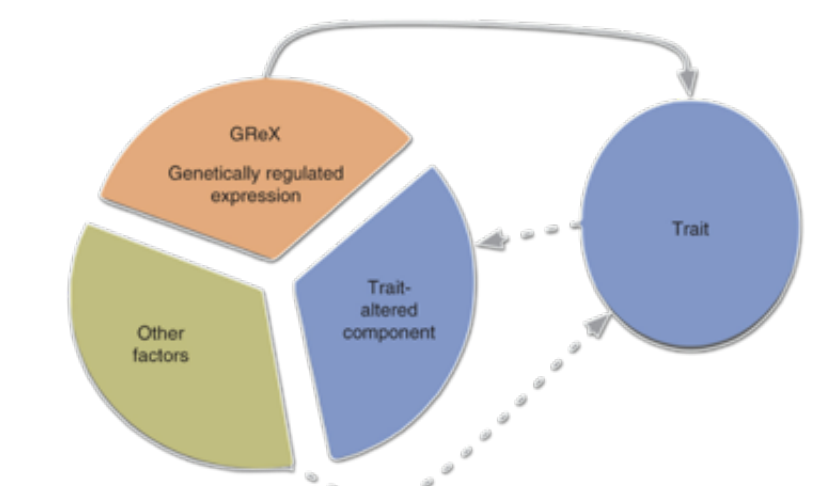
- environmental factors (smoking, exposure to chemical pollutants)
- genetic factors (37.7% rate of heritability)²

4. **What we know so far:** 532 genes (868 loci; GWAS/meta-analysis)^{3,4}

5. **What we don't know:** the mechanisms through which these genes affect the condition

To give us more insight, we can analyze transcriptome data. The challenges:

- Small number of individuals in the observed transcriptome compared to genotype.
- Transcriptome level is affected by multiple factors.

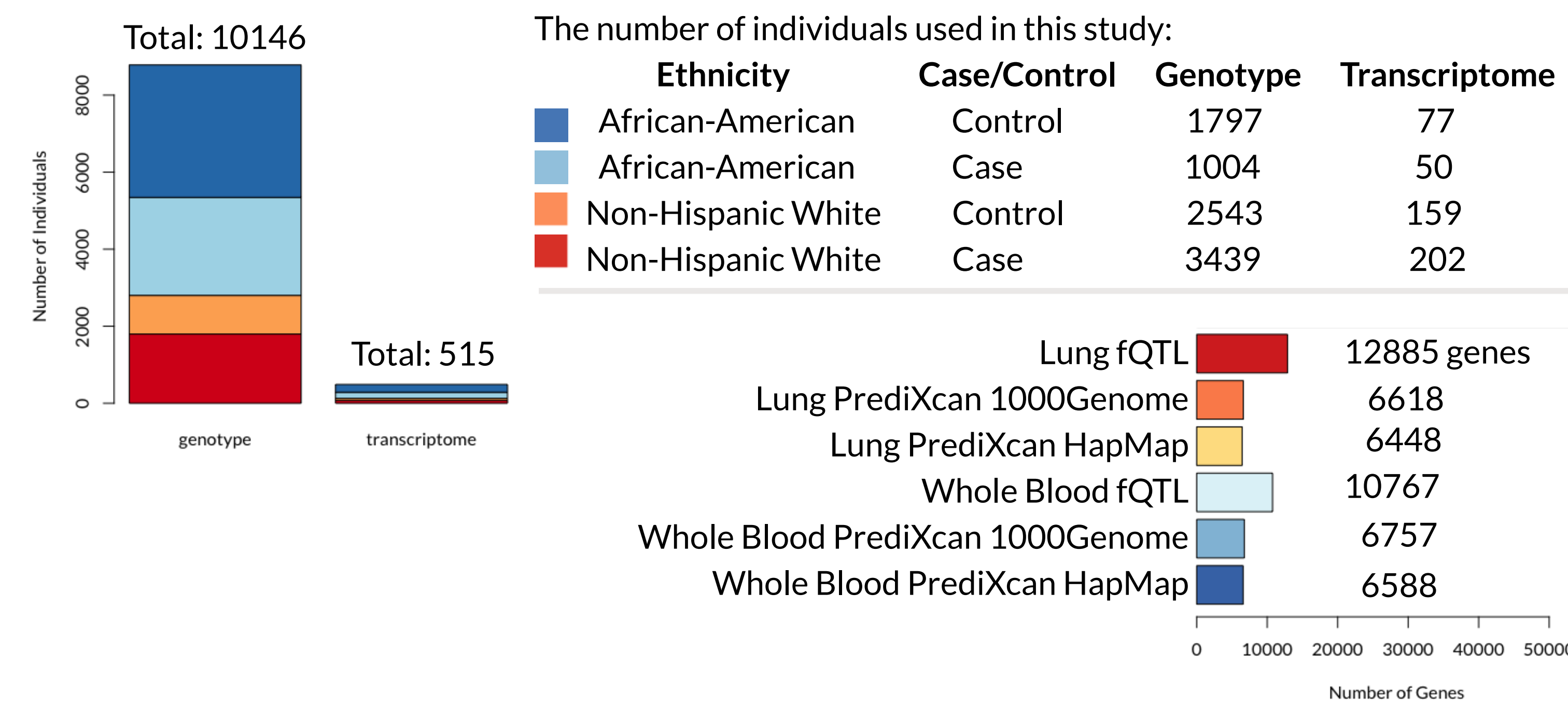


Gene Expression Decomposition⁶

Study Aim

To identify COPD-associated genes by evaluating proportion of gene expression explained by eQTL (displayed with the colour orange in the graph above). This value can be computed using fQTL⁵ and PrediXcan⁶.

Data



Methods

- Predict gene expression with fQTL and PrediXcan to for each individual in different tissues.
- Find association using 2 hypothesis:

Hypothesis 1: Genes have different proportions of genetically regulated component in their expression.

- Find associated genes on observed transcriptome
- Find associated genes on predicted expression of top 500 highly-associated genes from above

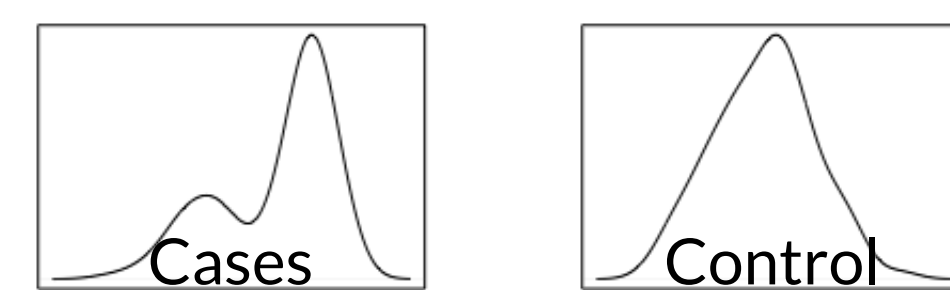
Hypothesis 2: Genes with higher genetic regulation plays a larger role in heritability.

- Compute correlation of predicted value to observed transcriptome.
- Find associated genes on predicted expression of genes with correlation ≥ 0.2 .

Statistical tests used:

- T-test
- Logistic regression on case/control status
- Logistic regression on case/control status and smoking duration
- A new statistical method developed by Mezlini, M. A.

Mezlini developed a statistical test to analyze if two populations are different by looking for enrichment close to the ends of the distribution as depicted above.

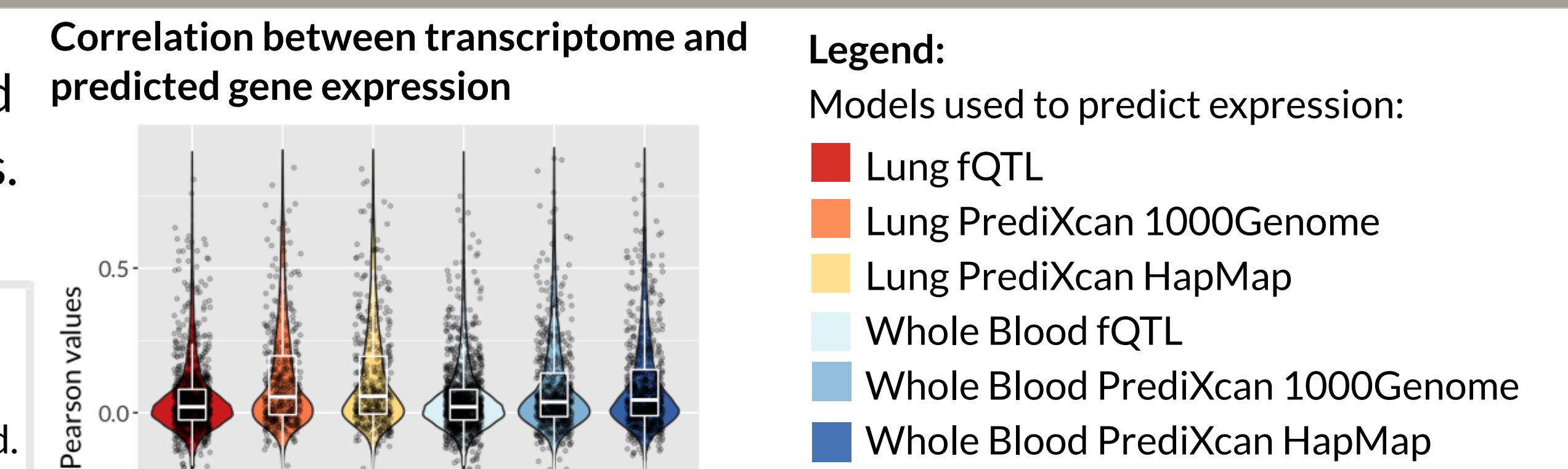


Multiple testing correction: p-value adjustment with false discovery rate (FDR) method with threshold 0.05.

References:
 1. Chronic obstructive pulmonary disease (COPD). (2017). World Health Organization. Retrieved 14 August 2017, from <http://www.who.int/mediacentre/factsheets/fs315/en/>
 2. Zhou, J., Cho, M., Castaldi, P., Hersh, C., Silverman, E., & Laird, N. (2013). Heritability of Chronic Obstructive Pulmonary Disease and Related Phenotypes in Smokers. doi: 10.1164/rccm.201302-0263oc
 3. Probert, K., Miller, S., Kheirallah, A. K., & Hall, I. P. (2015). Developmental genetics of the COPD lung. doi: 10.1186/s40749-015-0014-x
 4. Burdett, T., Hall, P. N., Hastings, E., Hindorf, L. A., Jenkins, H. A., Klemm, A. K., ... Welter, D. (2017). The NHGRI-EBI Catalog of published genome-wide association studies. Retrieved 14 August 2017, from www.ebi.ac.uk/gwas
 5. Park, Y., Sarkar, A., Bhutani, K., & Kellis, M. (2017). Multi-tissue polygenic models for transcriptome-wide association studies. doi: 10.1101/107623
 6. Gamazon, E., Wheeler, H., Shah, K., Mozaffari, S., Aquino-Michaels, K., Carroll, R. J., ... Kyung, H. (2015). A gene-based association method for mapping traits using reference transcriptome data. doi: 10.1038/ng.3367

Results

Figure 1. Violin plot of Pearson correlation values between observed transcriptome and predicted gene expression using different models.



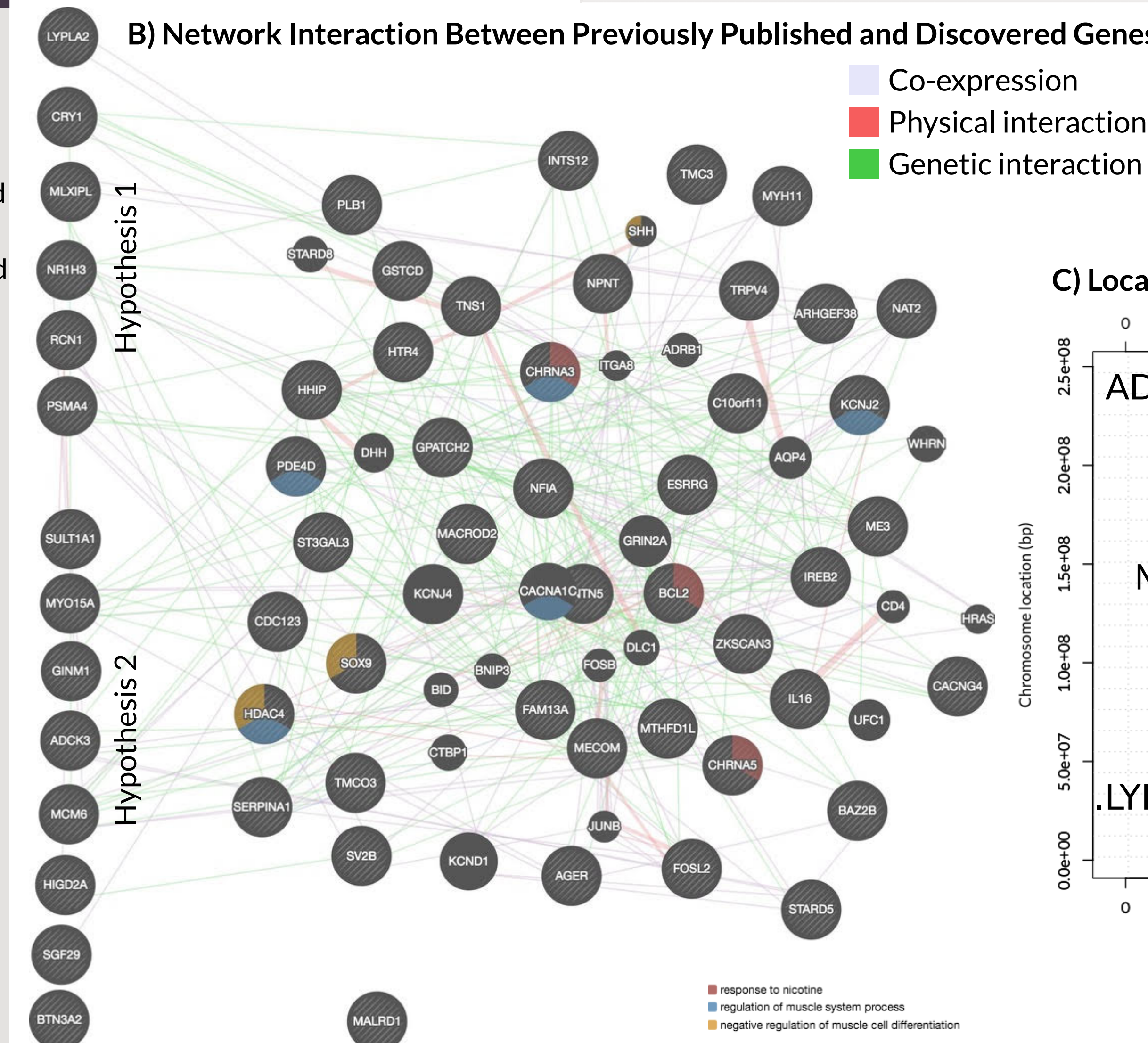
Legend:
 Models used to predict expression:
 Lung fQTL
 Lung PrediXcan 1000Genome
 Lung PrediXcan HapMap
 Whole Blood fQTL
 Whole Blood PrediXcan 1000Genome
 Whole Blood PrediXcan HapMap

Statistical tests used for association:
 A new statistical method developed by Mezlini, M. A.
 Logistic regression on case/control status
 Logistic regression on case/control + smoking duration
 T-test

Figure 2. Discovered COPD associated genes.

- A list of genes with its functions and ethnicity of population where association was found.
- Network interaction between discovered genes and a few genes previously published³.
- Locations of discovered genes and published SNPs^{3,4}.

A) Gene	Function	Ethnicity
LYPLA2	membrane enzyme	NHW
CRY1	circadian regulation	NHW
MLXIPL	triglyceride synthesis	Combined
NR1H3	lipid homeostasis	Combined
RCN1	Ca ²⁺ binding, ER	AA
PSMA4	proteasome	NHW
SULT1A1	sulfate conjugation	NHW
MYO15A	actin organization	NHW
GINM1	glycoprotein	NHW
ADCK3	electron transfer	NHW
MCM6	DNA replication	NHW
HIGD2A	mitochondrial respiration	NHW
SGF29	transcriptional regulation	NHW
BTN3A2	immune response	NHW



C) Locations of Previously Published SNPs and Newly Discovered Genes

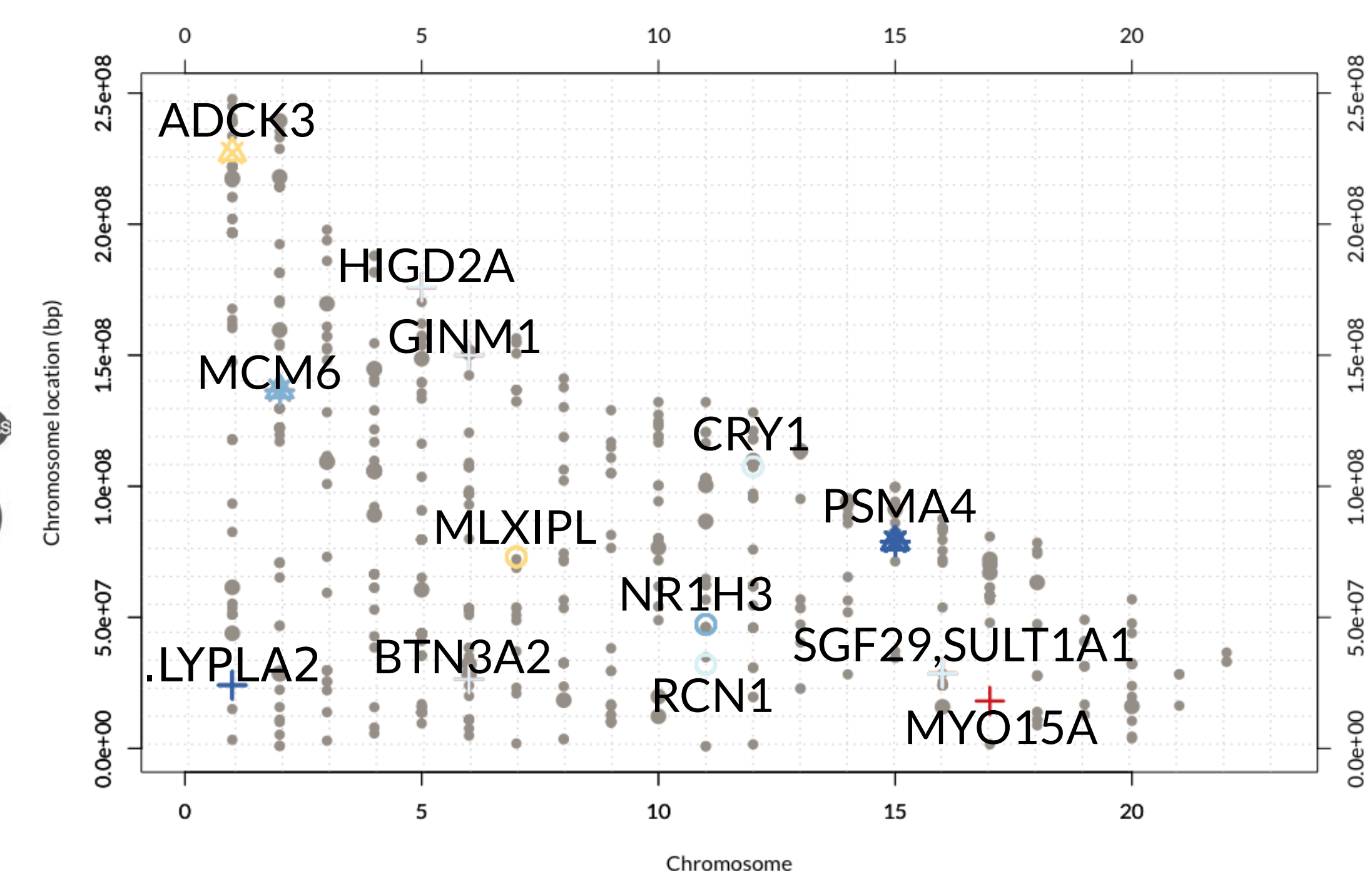
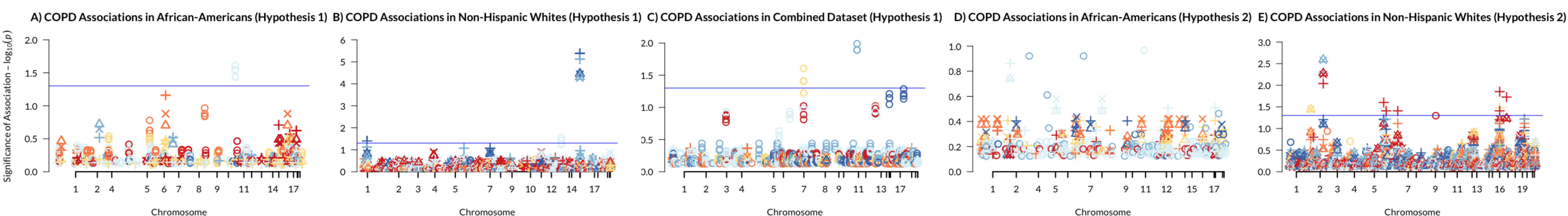


Figure 3. Manhattan plot of the discovery of genes associated with COPD using predicted gene expression with different models and statistical tests. FDR-corrected p-values for A-C) top 500 highly-associated genes observed in transcriptome; D,E) genes with expression correlation ≥ 0.2 . Note: only associations with p-value less than 0.75 is plotted.



Conclusions

Differences between statistical tests:

The logistic regressions and t-test found similar results with each other. Mezlini's statistical test is able to pick up associations that were not detected by the other methods.

Discovered 14 different genes that are statistically associated with COPD:

- 7 close to genes previously discovered by GWAS
- Mostly found in Non-Hispanic White dataset
- Predisposition to COPD in different ethnicities vary

fQTL vs PrediXcan:

fQTL models predict for more genes than PrediXcan models. However, expression predicted using PrediXcan models has a higher correlation with the observed transcriptome.

Acknowledgements

Thank you to Dr. Anna Goldenberg and Aziz Mezlini for the continuous guidance and support!

Shoutout to all the lab members for all the help, discussions, and brainstorming sessions: Ben, Lauren, Daniel H., Ladislav, Shems, Mingjie, Kingsley, Daniel C., Lebo, Jaryd, Mohamed, and Walter!

www.goldenberglab.ca
angelineyasodhara@gmail.com