Inferring Gene-Trait Associations with Diverse Population Gene Expression Prediction Models

Elyse Geoffroy, Isabelle Gregga, Dr. Heather Wheeler
Departments of Biology and Computer Science, Loyola University Chicago, Chicago, IL

Abstract

Recent genome-wide association studies (GWAS) primarily focus on European individuals; however, these results cannot be accurately applied to non-European populations due to differences in genetic architecture. Using previously published Multi-Ethnic Study of Atherosclerosis (MESA) gene-expression prediction models, we perform S-PrediXcan, a transcriptome wide association method, with previously published GWAS summary statistics to identify gene-trait associations. Wojcik et al. 2019 includes single nucleotide polymorphism-level associations for twenty-seven phenotypes in approximately 50,000 non-European individuals. From this S-PrediXcan output we can identify genes associated with various complex traits for non-European cohorts.

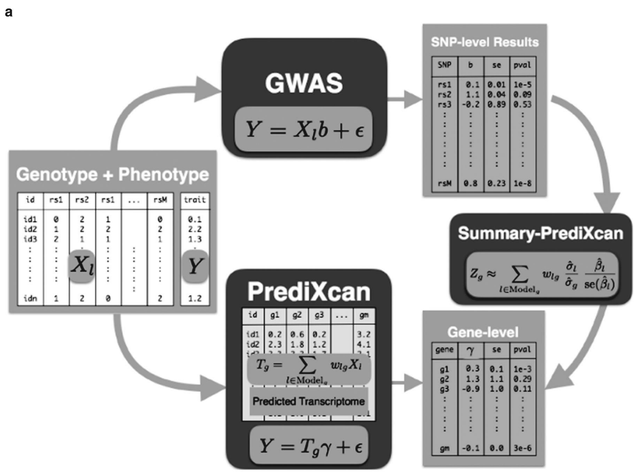
MESA Models

MESA Gene Expression Prediction Models Populations

Population	Number of Individuals	Model
African American	233	AFA
Hispanic	578	HIS
European	352	CAU
African American and Hispanic	585	AFHI
All Populations	1,163	ALL

- Multi-Ethnic Study of Atherosclerosis
- Previously published gene expression prediction models
- Data comes from the monocytes of > 1000 individuals

S-PrediXcan



Shows how much each gene contributed to the gene expression of a certain phenotypes

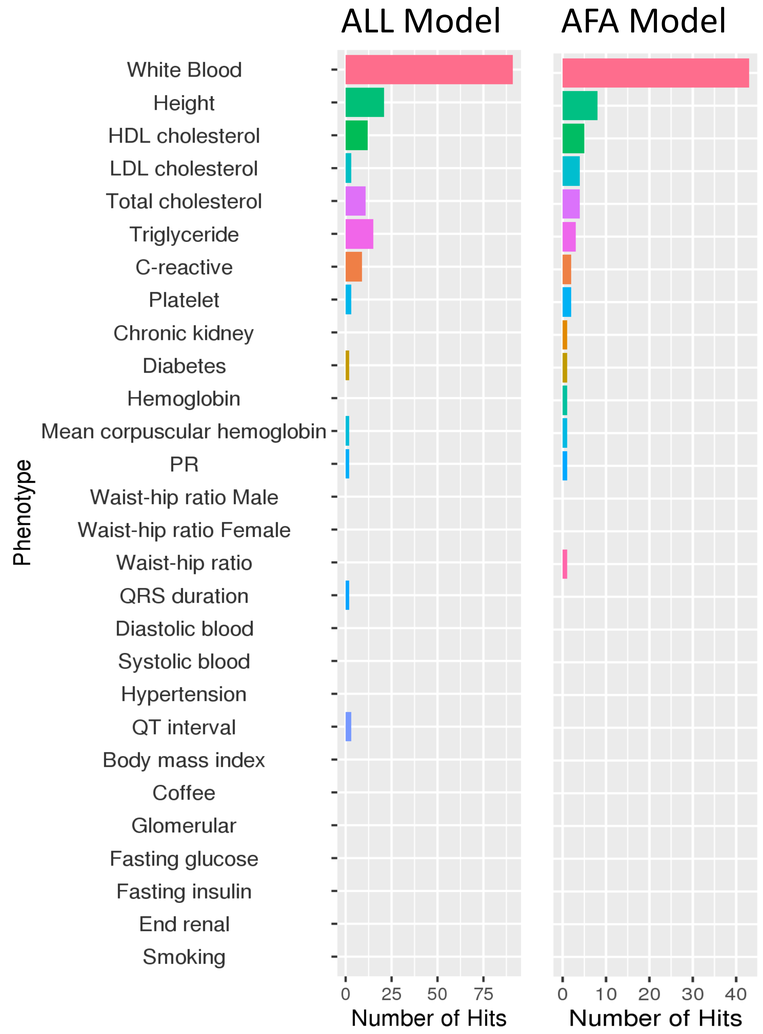
Gene predictors that are created can be used with phenotype data for different disease states

We will be using test-statistics from GWAS Summary Statistics with S-PrediXcan to find trait-associations

Wojcik et al 2019

Chronic Kidney Disease	Waist-hip ratio Female	Waist-hip ratio Male	Waist-hip ratio
LDL Cholesterol	HDL Cholesterol	Total Cholesterol	Triglyceride Levels
Mean corpuscular Hemoglobin	Body Mass Index	Systolic Blood Pressure	Diastolic Blood Pressure
Coffee Consumption	Hemoglobin Levels	Height	QRS Interval
C-reactive protein levels	White blood cell count	Fasting Glucose	Fasting Insulin
Cigarettes per day	Platelet Count	Type II Diabetes	End Stage Renal Failure
Hypertension	PR	QT interval	

Results

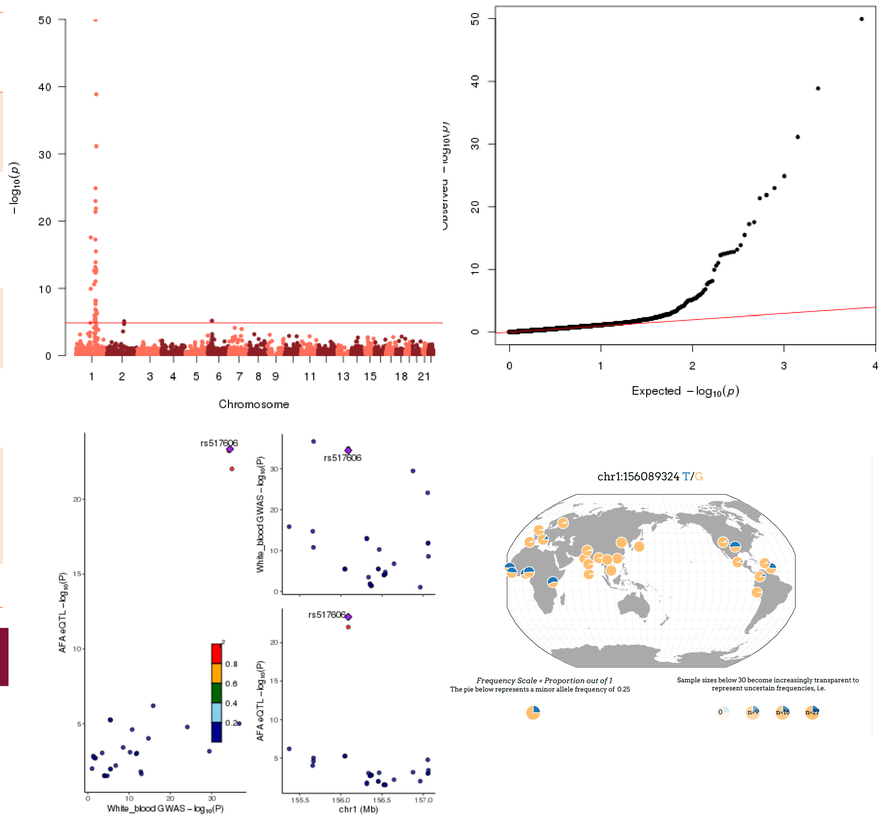


The number of significant genes found in S-PrediXcan between the ALL and AFA MESA models and the Wojcik phenotypes. Significance was determined by Bonferroni threshold.

Top Five Significant Gene Associations for White Blood Cell Count in African American Model

Gene	P-value	BETA
LMNA	2.42E-42	-0.063
UBAPL2	3.16E-31	-0.225
CD1B	8.61E-23	-0.099
DCAF8	8.55E-21	-0.223
B4GALT3	1.37E-19	-0.226

White Blood Cell Count Gene Associations in AFA Model



LMNA—chromosome 1

LMNA and *B4GALT3* found to be contributing to white blood cell count in PhenomeXcan analysis of Astle et al 2016.

Future Goals

We plan to run our model with more GWAS Summary Statistics data from other non-European studies. We also plan on studying other phenotypes and completing a similar analysis with the GTEx Consortium tissue models.

Acknowledgments

We would like to thank Dr. Wheeler for her mentorship and guidance. We would like to thank our lab members in the Wheeler lab for their support and collaboration. Finally, we would like to thank the Loyola Undergraduate Research Opportunities Program for supporting this project through the Mulcahy Scholars Program.

References

- Mogil LS, Andaleon A, Badalamenti A, Dickinson SP, Guo X, Rotter JI, et al. Genetic architecture of gene expression traits across diverse populations. PLoS Genetics. 2018;14(8):e1007586 10.1371/journal.pgen.1007586
- Gurdasani, D. *et al.* Genomics of disease risk in globally diverse populations. *Nat Rev Genet* (2019)
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., … Abecasis, G. R (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74.
- Lucia A. Hindorff, Vence L. Bonham, Lawrence C. Brody, et al. (2017). Prioritizing diversity in human genomics research. *Nature Reviews Genetic, 19* (3), 175-185.
- Stranger, Barbara E., et al. (2012) "Patterns of Cis Regulatory Variation in Diverse Human Populations." PLoS *Genetics PLoS, 8*(4).
- Li, Runze, and Hengjian Cui. (2014). Variable Selection via Regularization. *Wiley StatsRef: Statistic Reference Online.* doi:10.1002/9781118445112.stat07568
- Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V, Aquino-Michaels, K., Carroll, R. J., … Im, H. K. (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics*, *47*(9), 1091–1098
- Wojcik GL, Graff M, Nishimura KK, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019;570(7762):514-8.
- Astle, W. J. et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429 e19 (2016).
- Wheeler HE, Consortium G, et al. Survey of the Heritability and Sparse Architecture of Gene Expression Traits across Human Tissues. PLoS Genetics. 2016;**12**(11).
- Barbeira AN, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nature Communications. 2018;**9**(1).