Proteome Association Studies in Populations of Diverse Ancestries

Isabelle Gregga, Elyse Geoffroy, Ryan Schubert, Heather Wheeler, the TOPMed Consortium
Department of Biology, Program of Bioinformatics, Loyola University Chicago, Chicago, IL

Introduction

Most genome-wide association studies (GWAS) have been conducted in populations of European ancestries, but these results do not reflect the global population or replicate well in non-European populations. Additionally, investigating traits at the proteome level may provide more insight to biological mechanisms than at the genome level. Using data from the Trans-Omics for Precision Medicine (TOPMed) consortium, we have built protein models to perform proteome-wide association studies (PWAS) using S-PrediXcan in published multiethnic GWAS data from the Population Architecture using Genomics and Epidemiology (PAGE) study (Wojcik et al 2019). This output reveals significant associations between genes and a variety of complex traits in non-European populations.

Methods

- S-PrediXcan: statistical analysis software, takes GWAS summary statistics, protein level models, and phenotype data to find associations between proteome and traits
- Bonferroni significance threshold used to find the most significant associations
- The PAGE study: the most diverse GWAS to date, collecting data in 28 phenotypes in a sample size of ~50,000 non-European individuals; publicly available summary statistics used for our discovery population

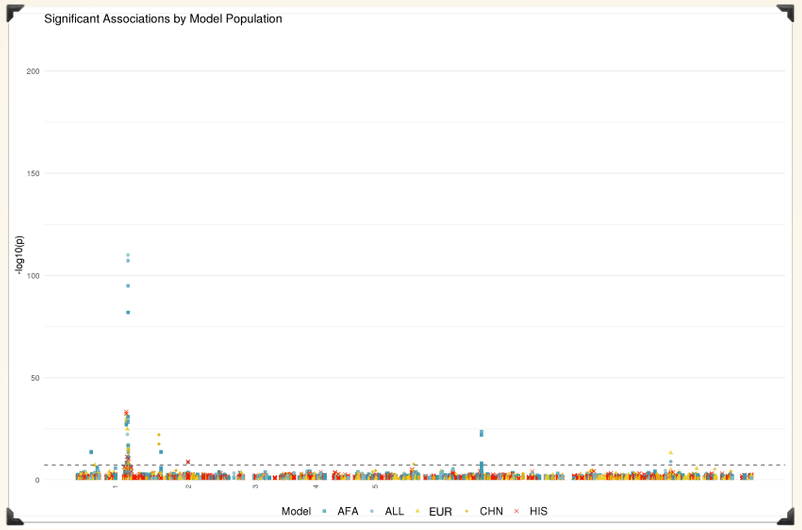
- Body mass index
- Chronic kidney disease
- Cigarettes per day (smoking)
- Coffee consumption
- C-reactive protein levels
- Diabetes (type II)
- Diastolic blood pressure
- End-stage renal failure
- Fasting glucose levels
- Fasting insulin levels
- Glomerular protein levels
- Height
- Hemoglobin levels
- HDL cholesterol levels
- Hypertension
- LDL cholesterol levels
- Mean corpuscular hemoglobin
- PR interval
- Platelet count
- QRS interval
- QT interval
- Systolic blood pressure
- Total cholesterol
- Triglyceride levels
- Waist-hip ratio
- Waist-hip ratio (female)
- Waist-hip ratio (male)
- White blood cell count

- TOPMed models: made with relatively diverse dataset from the TOPMed project
- PWAS is a new method compared to more common TWAS (transcriptome); we are still refining these protein models

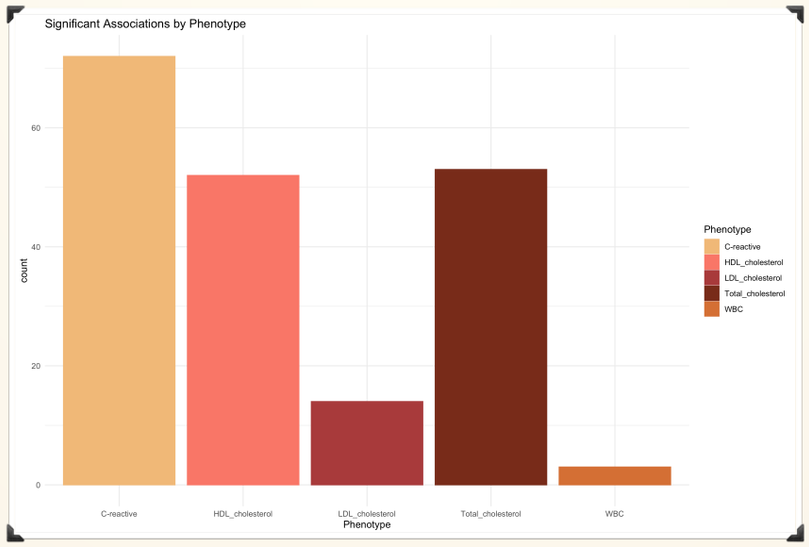
- AFA: African American, n=183
- CHN: Chinese, n=77
- EUR: European, n=414
- HIS: Hispanic, n=301
- ALL: all groups, n=975

- Coloc: software tool used to determine the colocalization of genes in a GWAS, provides insight to which of the significant S-PrediXcan results may be causal
- Replication datasets: taken from published GWAS summary statistics from the UK Biobank, other studies publicly available in the GWAS Catalog

Results



- 194 Bonferroni significant, colocalized, and replicated protein-trait pairs.
- In non-European PAGE data: more significant results when using the AFA (dark blue) and HIS (red) training models than the EUR (yellow) models, significance threshold shown as the dotted line



- Significant protein-trait associations found for 5 phenotypes:
  - C-reactive protein levels
  - HDL cholesterol levels
  - LDL cholesterol levels
  - Total cholesterol
  - White blood cell count

Discussion

- 194 total associations across all training models
  - 27 unique protein-trait pairs

Gene	Protein	Phenotype
CRP	CRP	C-reactive
APOE	Apo E	HDL cholesterol LDL cholesterol Total cholesterol C-reactive
	Apo E2	HDL cholesterol LDL cholesterol Total cholesterol C-reactive
	Apo E3	HDL cholesterol LDL cholesterol Total cholesterol C-reactive
	Apo E4	HDL cholesterol LDL cholesterol Total cholesterol C-reactive Triglycerides
HP	Mixed-type haptoglobin	LDL cholesterol Total cholesterol
CD36	CD36 antigen	C-reactive HDL cholesterol Platelet count
CSF3	G-CSF	WBC count
IL6R	IL-6 sRa	C-reactive
IL1RN	IL-1Ra	C-reactive
FRZB	sFRP-3	Height

- Published literature supports these associations and provides evidence to the validity of our proteome models. Only one pair, CD36 antigen (CD36 gene) and C-reactive protein level, is a novel finding among published GWAS. We plan to continue refining our training models and conducting more PAS in the future.
- Finding more significant associations in non-European populations with non-European models supports previous work that shows prediction improves with population-matched data. This further emphasizes the need for more diverse GWAS in the future.

Acknowledgements

We would like to thank Dr. Wheeler for her guidance and aid on this project, the members of the Wheeler Lab for their support, and the TOPMed Consortium and NIH grant R15HG009569. We would also like to thank the Loyola Undergraduate Research Opportunities Program and the Mulcahy Scholars Program who made this work possible.

References

- Mills, M.C., Rahal, C. A scientometric review of genome-wide association studies. Commun Biol 2, 9 (2019).
- Mogil LS, Andaleon A, Badalamenti A, Dickinson SP, Guo X, Rotter JI, Johnson WC, Im HK, Liu Y, Wheeler HE. (2018) Genetic architecture of gene expression traits across diverse populations. PLOS Genetics 14(8):e1007586.
- Wojcik GL, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019 Jun;570(7762):514-518.
- Barbeira, A.N., Dickinson, S.P., Bonazzola, R. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun 9, 1825 (2018).
- Allena, Naomi; Sudlowa, Cathie; Downey, Paul; Peakman, Tim; Danesh, John; Elliott, Paul; Gallacher, John; Green, Jane; Matthews, Paul; Pell, Jill; Sprosen, Tim; Collins, Rory. UK Biobank: Current status and what it means for epidemiology. Health Policy and Technology 1, 3 (2012).
-Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. (2014) Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet 10(5): e1004383.