**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 into 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 ~22,000 non-European individuals, publicly available summary statistics used for our discovery population.

**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.
- Top Med 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.

**Results**

- 944 Bonferroni significant, colocated, and replicated protein-trait pairs.
- In non-European PAGE data: more significant results when using the AFA (dark blue) and HHS (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**
- 944 total associations across all training models
- 27 unique protein-trait pairs.

**Tables**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRP</td>
<td>CRP</td>
<td>C-reactive</td>
</tr>
<tr>
<td>Apo E</td>
<td>LDL cholesterol</td>
<td>Total cholesterol</td>
</tr>
<tr>
<td>Apo E2</td>
<td>LDL cholesterol</td>
<td>C-reactive</td>
</tr>
<tr>
<td>Apo E3</td>
<td>LDL cholesterol</td>
<td>Total cholesterol</td>
</tr>
<tr>
<td>Apo E4</td>
<td>LDL cholesterol</td>
<td>C-reactive</td>
</tr>
<tr>
<td>HP</td>
<td>Mixed-type haptoglobin</td>
<td>Total cholesterol</td>
</tr>
<tr>
<td>CD36</td>
<td>CD36 antigen</td>
<td>C-reactive</td>
</tr>
<tr>
<td>CSF3</td>
<td>G-CSF</td>
<td>WBC count</td>
</tr>
<tr>
<td>IL6R</td>
<td>IL-6 sRa</td>
<td>C-reactive</td>
</tr>
<tr>
<td>IL1RN</td>
<td>IL-10Ra</td>
<td>C-reactive</td>
</tr>
<tr>
<td>FRZB</td>
<td>sFRP-3</td>
<td>Height</td>
</tr>
</tbody>
</table>

**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 R35HG009569. We would also like to thank the Loyola Undergraduate Research Opportunities Program and the Mulchey Scholars Program who made this work possible.

**References**