S68. Leveraging population genetics to inform diverse cohorts and biobanks

Location: Conv Ctr/West Hall A/West Building

Session Time: Friday, October 28, 2022, 2:00 pm - 3:30 pm

ProgNbr 447. Multiadaptive shrinkage improves cross-population transcriptome prediction for transcriptome-wide association studies in underrepresented populations

Authors:

D. Araújo[1], C. Nguyen[1], X. Hu[2], S. S. Rich[2], J. I. Rotter[3], H. Im[4], A. W. Manichaikul[2], H. E. Wheeler[1], NHLBI TOPMed Consortium; [1]Loyola Univ. Chicago, Chicago, IL, [2]Univ. of Virginia, Charlottesville, VA, [3]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [4]Univ. of Chicago, Chicago, IL

Abstract:

The majority of genome- and transcriptome-wide association studies (GWAS/TWAS) are conducted in European ancestry populations, and consequently may not reflect genetic variants and linkage disequilibrium patterns found within non-European populations. Population-matched TWAS have been shown to increase the number of true discoveries; however, building accurate transcriptome prediction models for underrepresented populations is not always possible due to smaller samples sizes or data availability. Thus, we sought to build new transcriptome prediction models with better cross-population performance.
We used RNA-Seq data from the Trans-omics for Precision Medicine (TOPMed) Multi-Ethnic Study of Atherosclerosis (MESA) from up to 1287 African American (AFA), Chinese (CHN), European (EUR) and Hispanic/Latino (HIS) individuals of 3 blood cell types: peripheral blood mononuclear cells, CD16+ monocytes, and CD4+ T cells. Multivariate adaptive shrinkage in R (MASHR) was developed to estimate effects in genomic studies with multiple conditions and adapts to patterns present in the data, allowing for both shared and condition-specific effects. Here, we first used Matrix eQTL to perform cis-eQTL mapping in each population-tissue pair to generate single population effect sizes and standard errors, which were used as the input for MASHR. We then carried the most significant SNP in each population forward to prediction models, using the MASHR-generated effect sizes. We compared prediction performance in Geuvadis of MASHR models to unadjusted top SNP (Matrix eQTL) and elastic net (EN) approaches. MASHR models performed better than Matrix eQTL in all cases, whereas in comparison to EN, MASHR either performed the same or better. For example, AFA MASHR and EN models performed the same in Geuvadis Yoruba (YRI, p=0.21), but AFA MASHR models performed better than AFA EN in Geuvadis British (GBR, p=2.0x10-24). With the EUR models, MASHR outperformed EN in YRI (p=6.0x10-34) and in GBR (p=0.00052).
Next, we performed TWAS using GWAS summary statistics for 28 complex traits from the Population Architecture using Genomics and Epidemiology study. Across all population models and phenotypes tested, MASHR identified the most significant gene-trait pairs (MASHR=310, Matrix eQTL=295, EN=200), but the least unique significant pairs (MASHR=108, Matrix eQTL=136, EN=132), which could indicate that MASHR models give more consistent results across populations. We expect that by improving cross-population transcriptome prediction, we will identify new, more consistent gene-trait associations to better understand the underlying mechanisms of complex traits.

