# Polygenic Risk Scores for SSNS, SRNS

Alejandro Ochoa

StatGen, Biostatistics & Bioinformatics — Duke University

2024-02-07 — NS U01 working group

# Overview

- Completed PRS testing on a new setup
- Redid PRS testing on old setup
- Under the hood
  - Streamlined code
  - Added PC conditioning, which slightly increased correlations

# How PRS works

Score is generally a linear model:

$$\text{PRS}_j = \sum_i \beta_i x_{ij}.$$

- $i$: variant index
- $j$: individual index
- $\beta_i$: coefficient of variant $i$
- $x_{ij}$: genotype (0,1,2) at variant $i$, individual $j$

Challenge is about picking $\beta_i$:

- Not all variants are in all datasets
- If starting from GWAS, need to decorrelate (LD or clumping), shrink (p-value threshold or fancier models)

# Basics of PRS construction and evaluation

- PRS construction and validation requires 3 disjoint datasets:
  - Base set: Used to fit "GWAS summary statistics": variant coefficients (betas), standard errors, p-values
  - Training set: Used to fit PRS parameters: p-value threshold, or heritability and sparsity
    - Modifies betas, usually by shrinking them to zero and reducing correlation due to LD
  - Testing set: Data where nothing was trained, reveals true performance (correlation to trait)

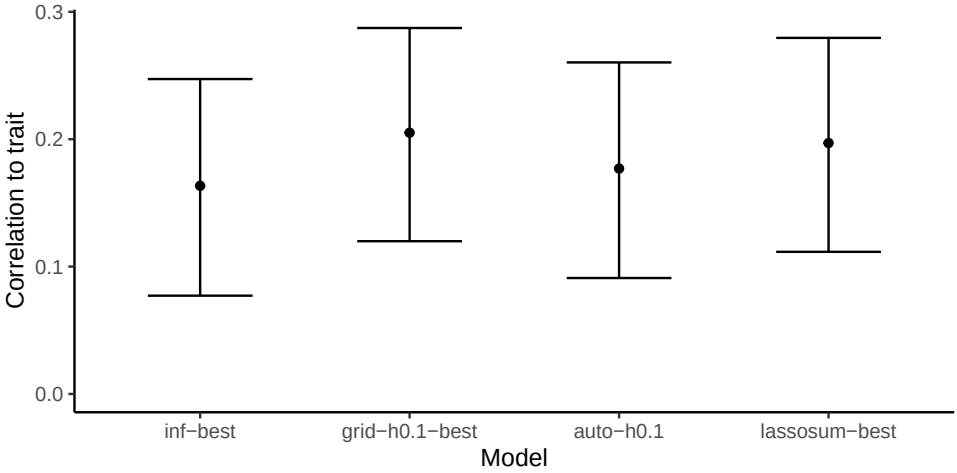
# Testing setups

| Name | Base | Train | Test |
|---|---|---|---|
| Old SSNS-SRNS | Discov SSNS-SRNS (725/193) | Bristol SSNS-SRNS 70% (252/103) | Bristol SSNS-SRNS 30% (113/46) |
| Old SSNS-Ctrl | Discov SSNS-Ctrl (725/3553) | Bristol SSNS-SRNS 70% (252/103) | Bristol SSNS-SRNS 30% (113/46) |
| New SSNS-Ctrl | Discov SSNS-Ctrl (532/3553) | Discov SSNS-SRNS (193/193) | Bristol SSNS-SRNS 100% (365/149) |

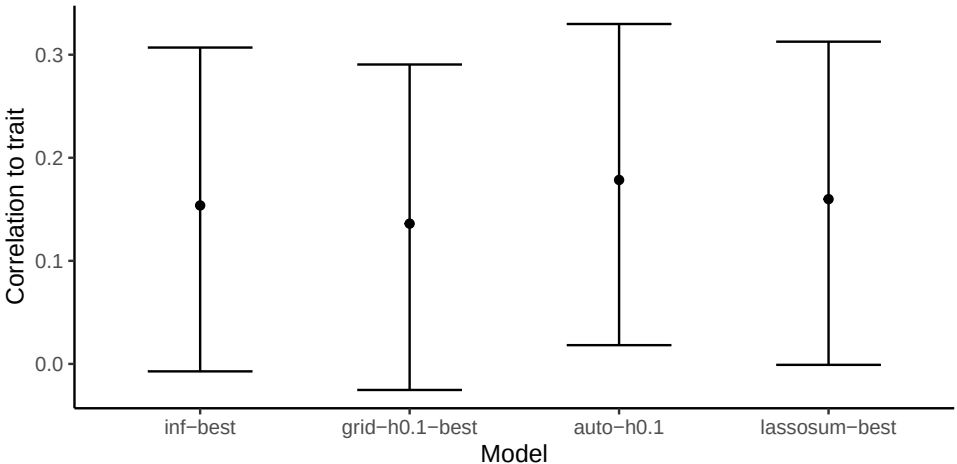
To incorporate soon, CureGN version SSNS-SRNS (250/170), based on these rules:

- SSNS: MCD and age $\leq 21$
- SRNS: FSGS and age $\leq 21$

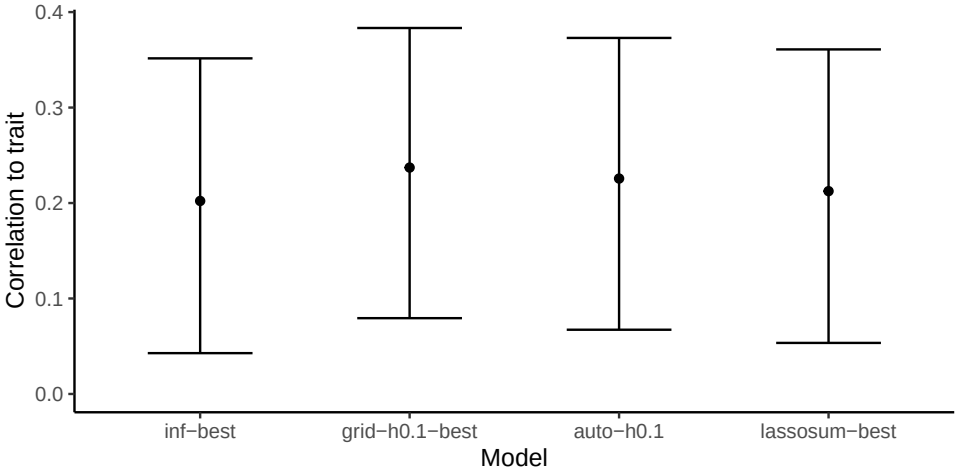
# Test results: New SSNS-Ctrl



# Test results: Old SSNS-Ctrl



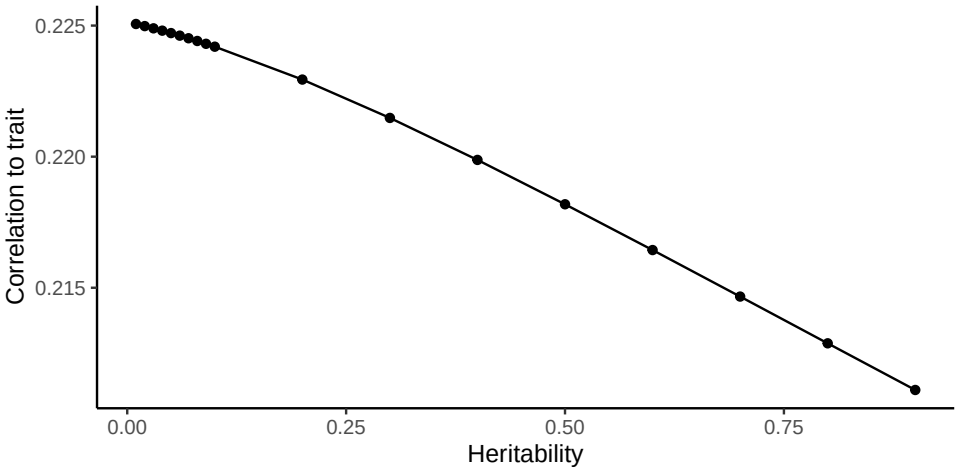
# Test results: Old SSNS-SRNS



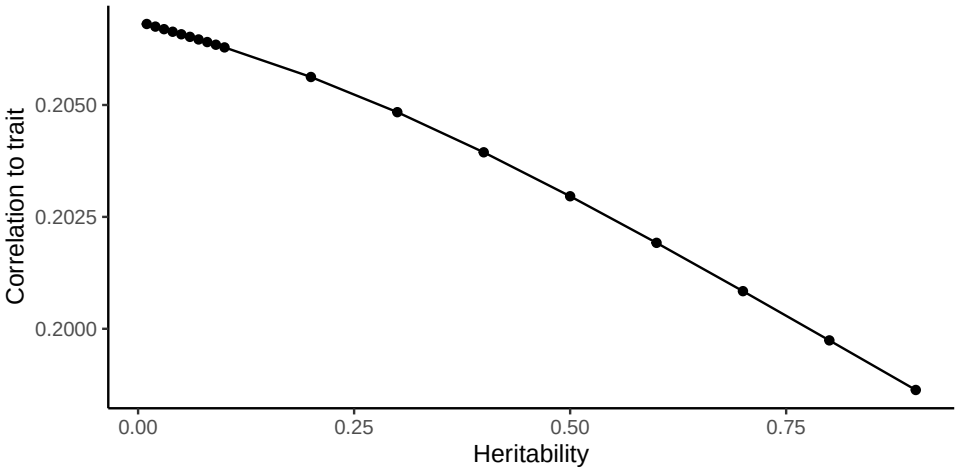
# Next steps

- Use LD of base data instead of training data
- Include clump and threshold method
- Include CureGN's SSNS/SRNS (allows base data to be for both SSNS-SRNS and SSNS-Ctrl, as with old setup)
- Use HLA haplotypes!
- Vary SNP set filters
  - Due to LD runtime, only using array SNPs right now
  - Could enrich for more significant ones or higher severity variants
- Try Barry et al., 2023 base data

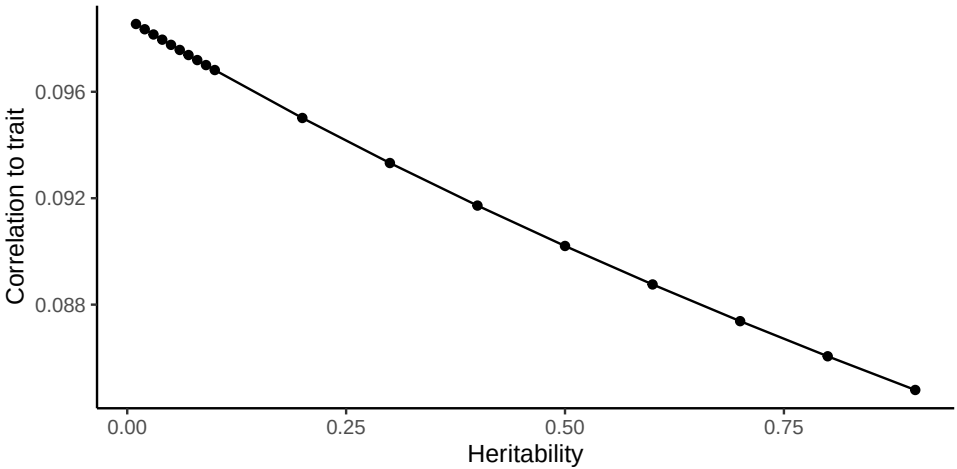
# Train results: New ldpred-inf



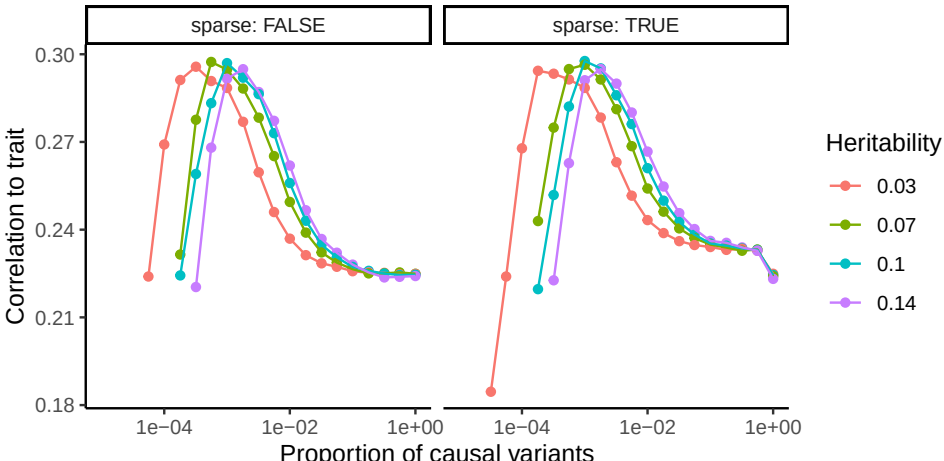
# Train results: Old SSNS-Ctrl ldpred-inf



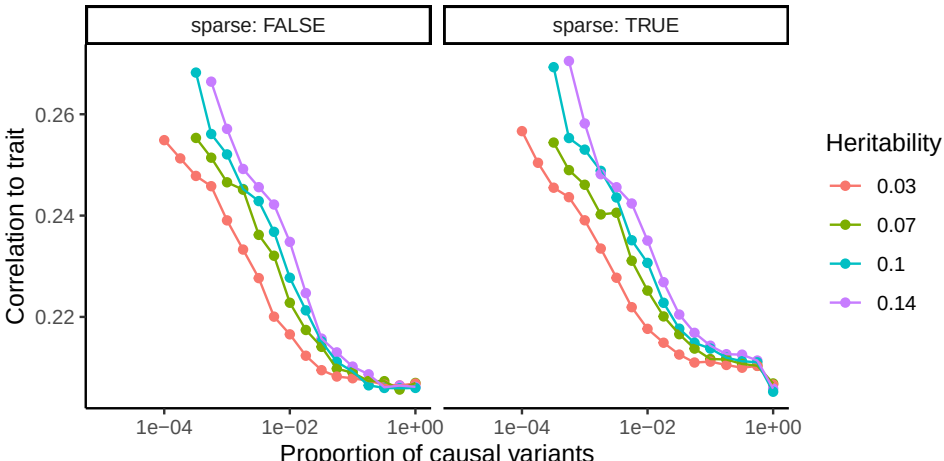
# Train results: Old SSNS-SRNS ldpred-inf



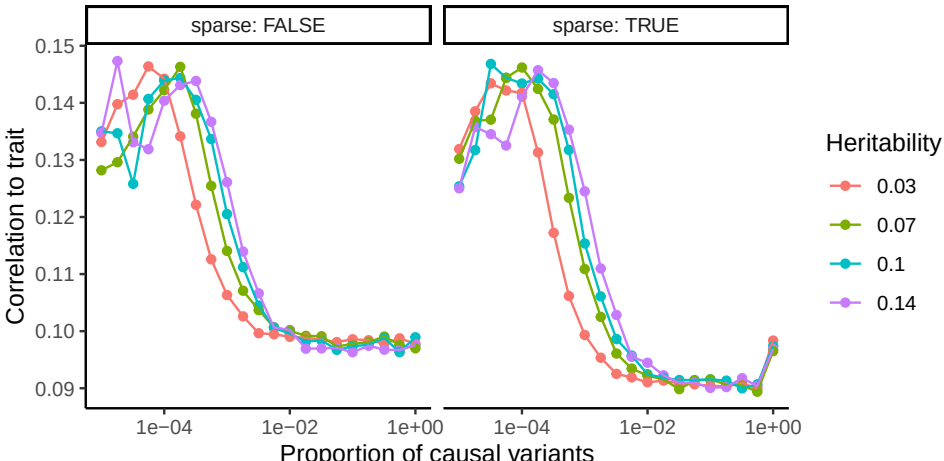
# Train results: New ldpred-grid-h0.1



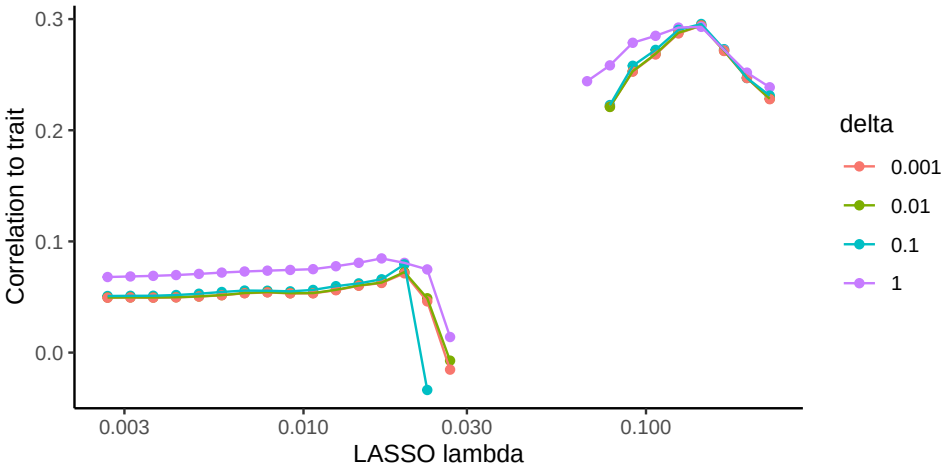
# Train results: Old SSNS-Ctrl ldpred-grid-h0.1



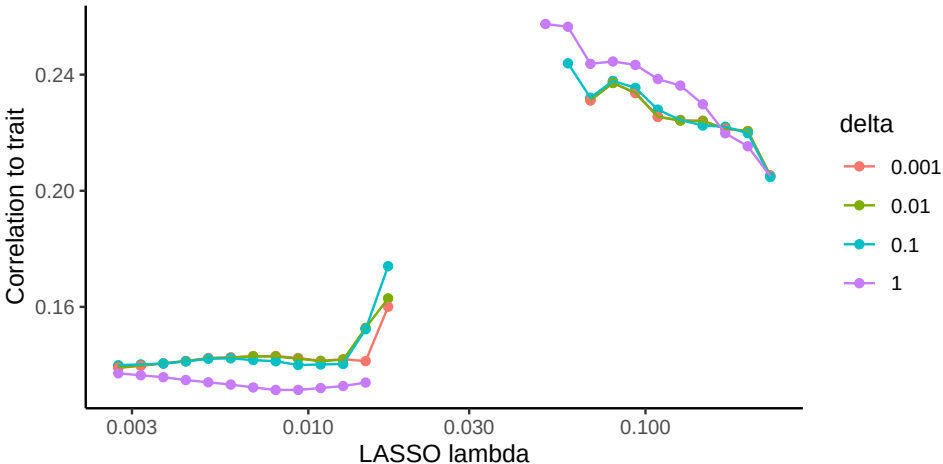
# Train results: Old SSNS-SRNS ldpred-grid-h0.1



# Train results: New ldpred-lassosum



# Train results: Old SSNS-Ctrl ldpred-lassosum



# Train results: Old SSNS-SRNS ldpred-lassosum