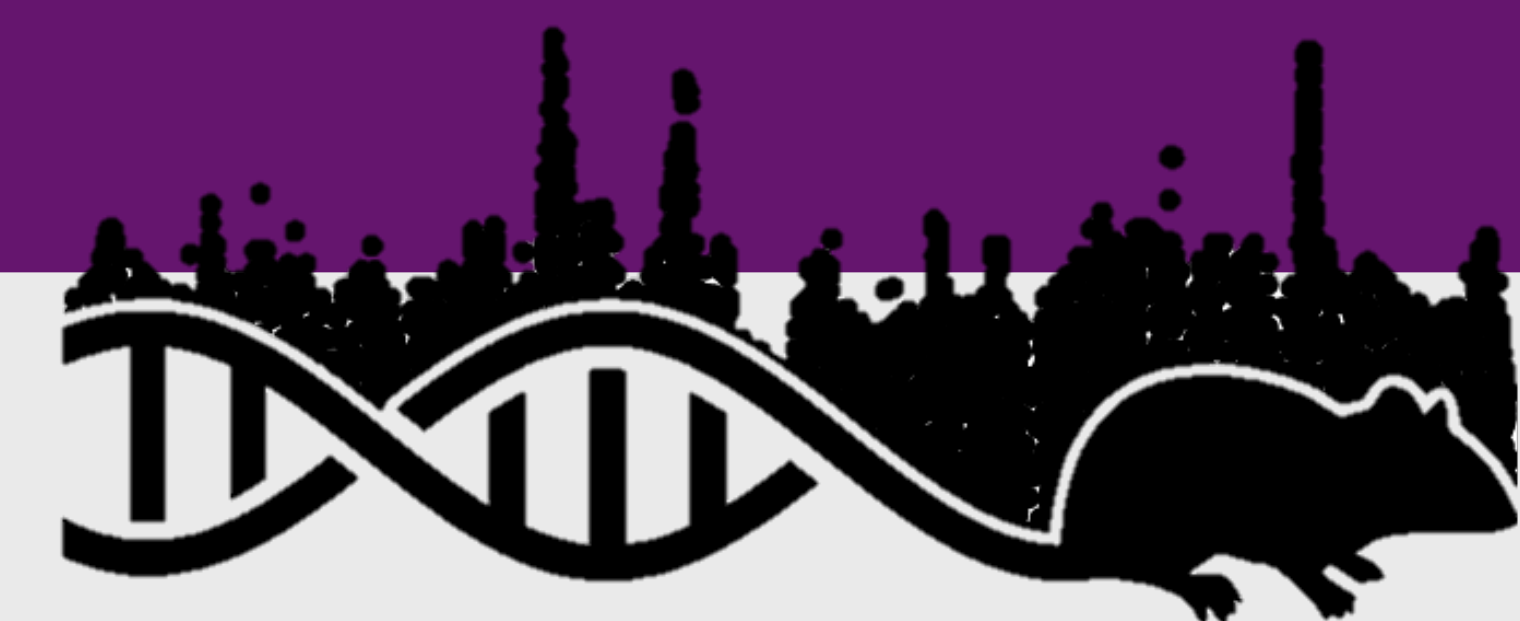


RATTACA:

Genetic predictions for genetic correlation and experimental design in outbred rats



NIDA Center for GWAS in Outbred Rats



Benjamin B. Johnson^{1*}, Robert M. Vogel¹, Thiago M. Sanches¹, Mika H. Okamoto^{1,2}, Khai-Min H. Nguyen¹, Clara A. Ortez¹, Oksana Polesskaya¹, Abraham A. Palmer^{1,3°}

THE PROBLEM

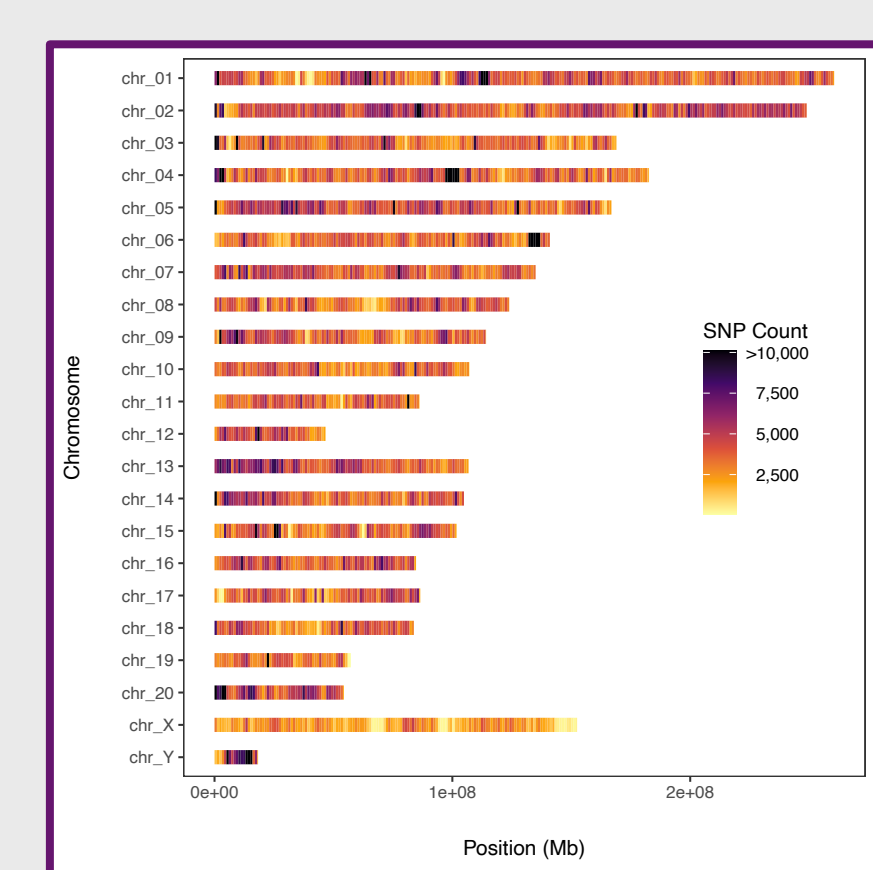
- Genetic correlations are foundational in identifying causal links between genotype and phenotype
- However, observed trait correlations regularly arise between populations without any shared genetic architecture
- Methods to avoid confounding between genetic and environmental contributions to phenotype (e.g., trait comparisons across large inbred panels) require intensive breeding programs that demand substantial funds, effort, and time

THE SOLUTION

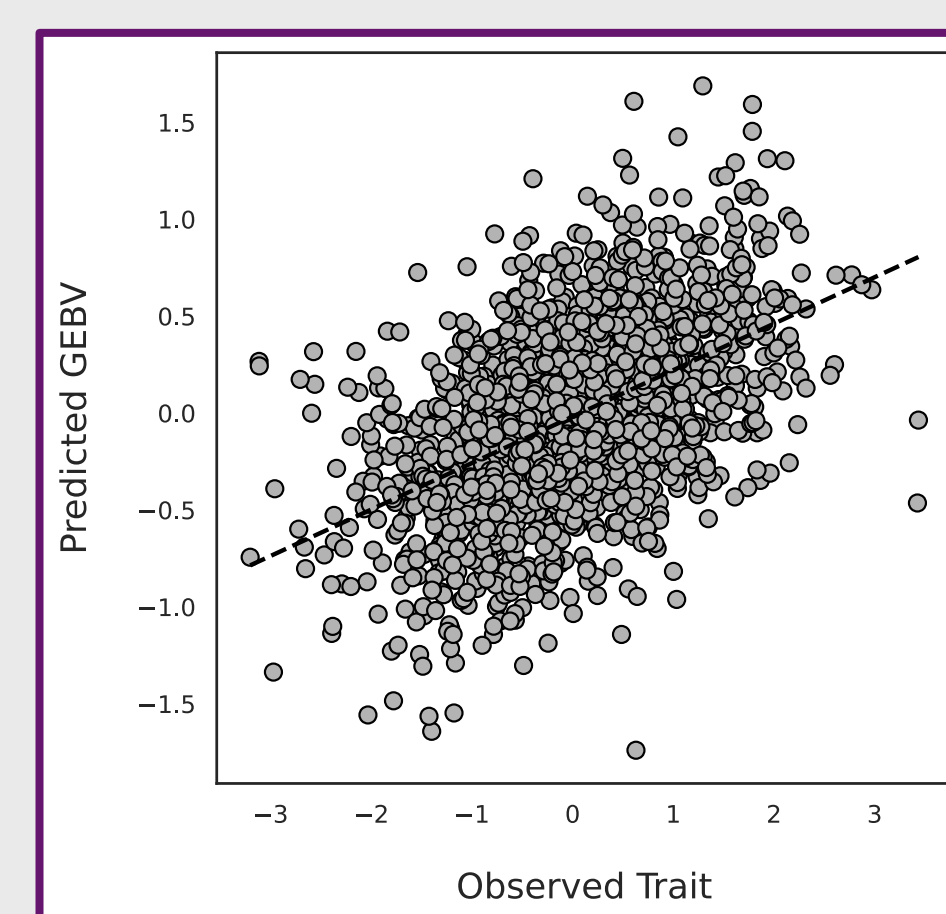
- Genetic prediction!
- Genotype & phenotype measurements from GWASs can be used to estimate per-variant genetic effect sizes on the phenotype
- These effect sizes can be used to predict trait values in genotyped individuals without the need for phenotyping
- Trait predictions allow researchers to design experimental samples expected to show desired phenotypic values
- Correlations between the predicted trait and a second measured trait suggest a genetic correlation between phenotypes

THE RATTACA PARADIGM

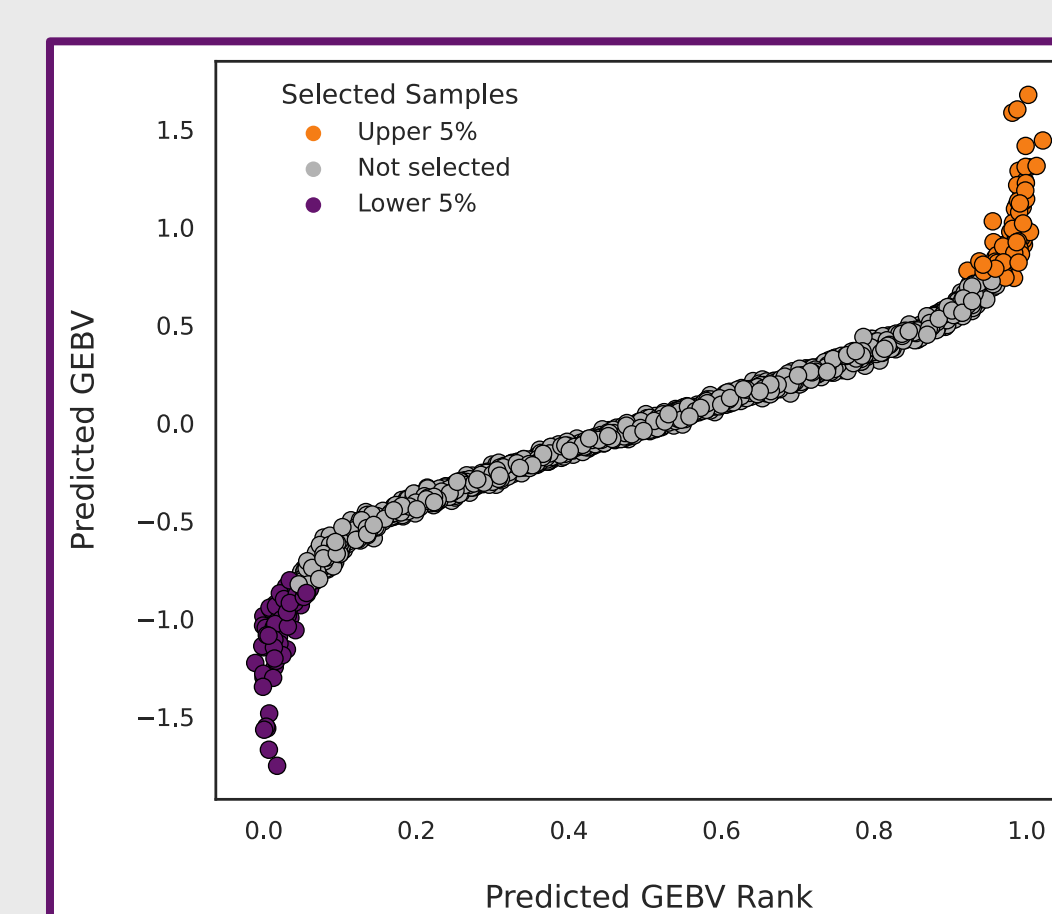
1 Genotyping + phenotyping



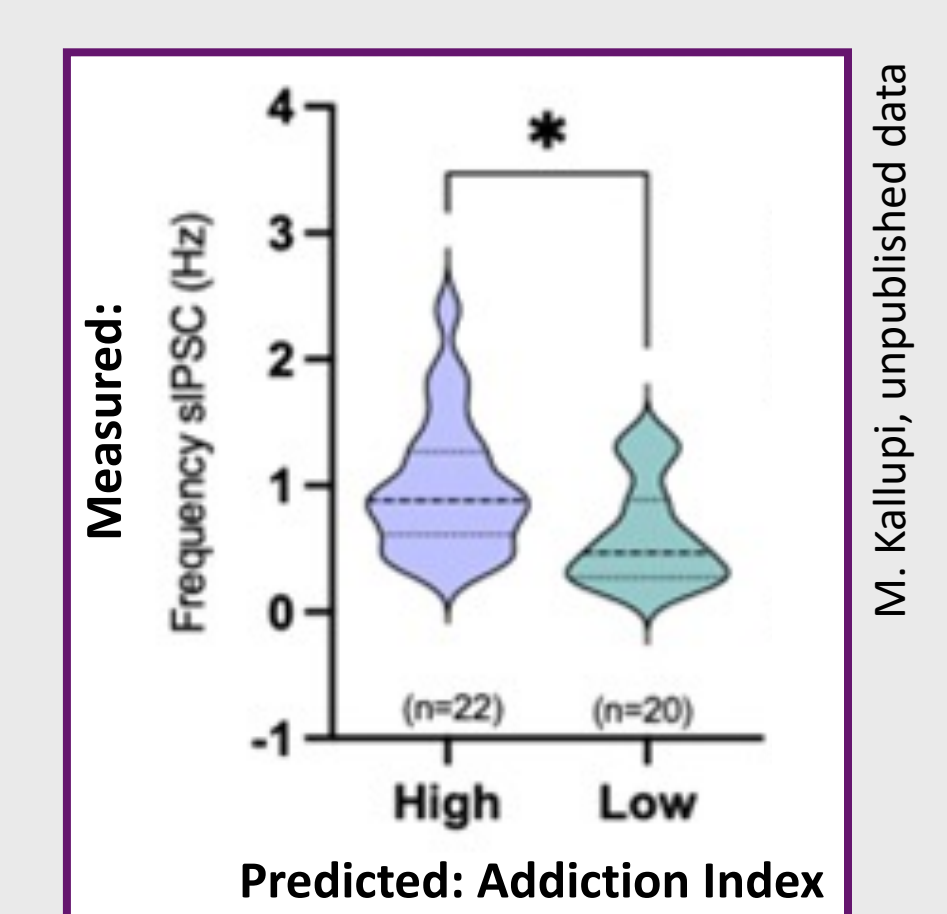
2 Genetic prediction



3 Sample selection

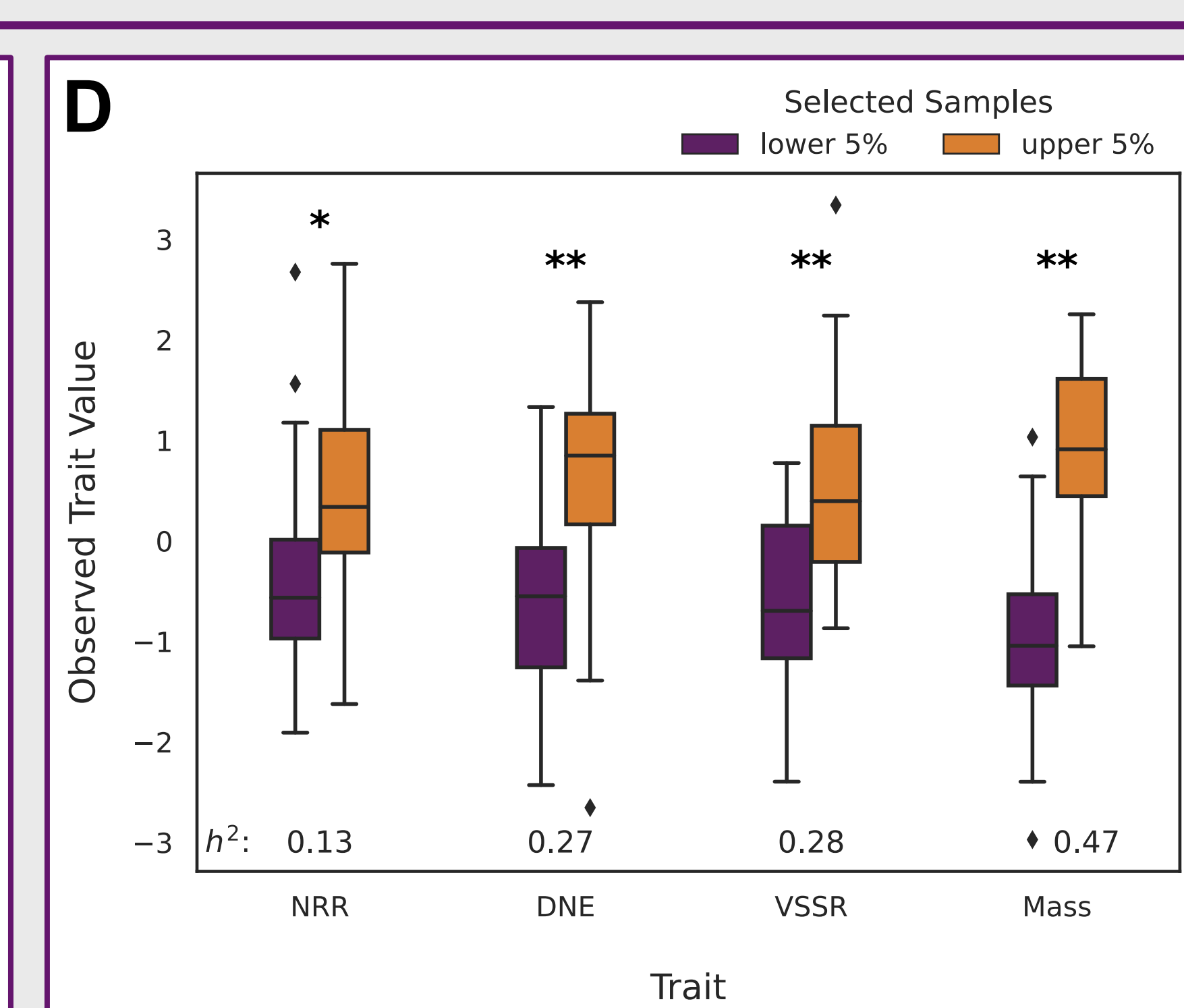
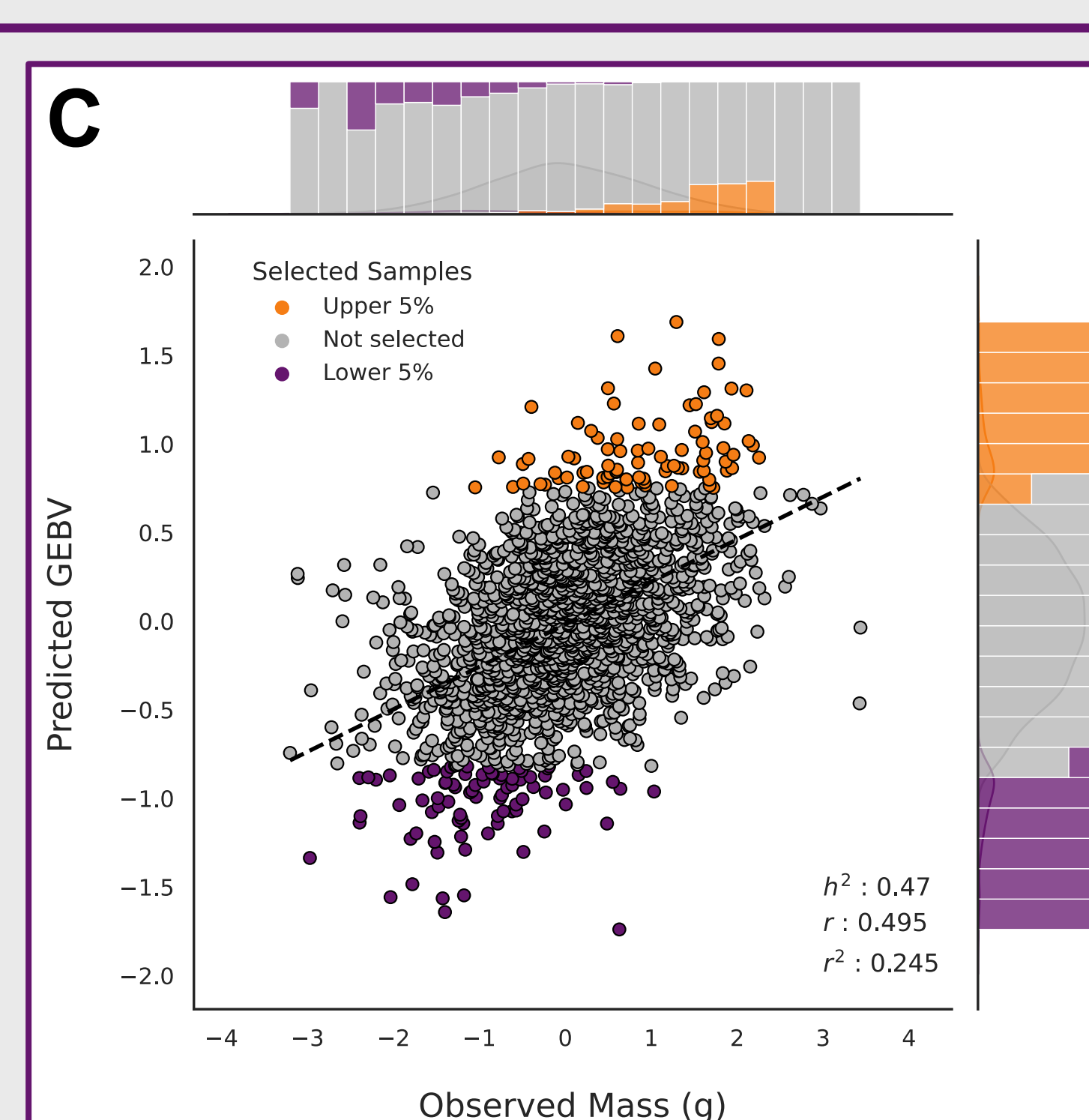
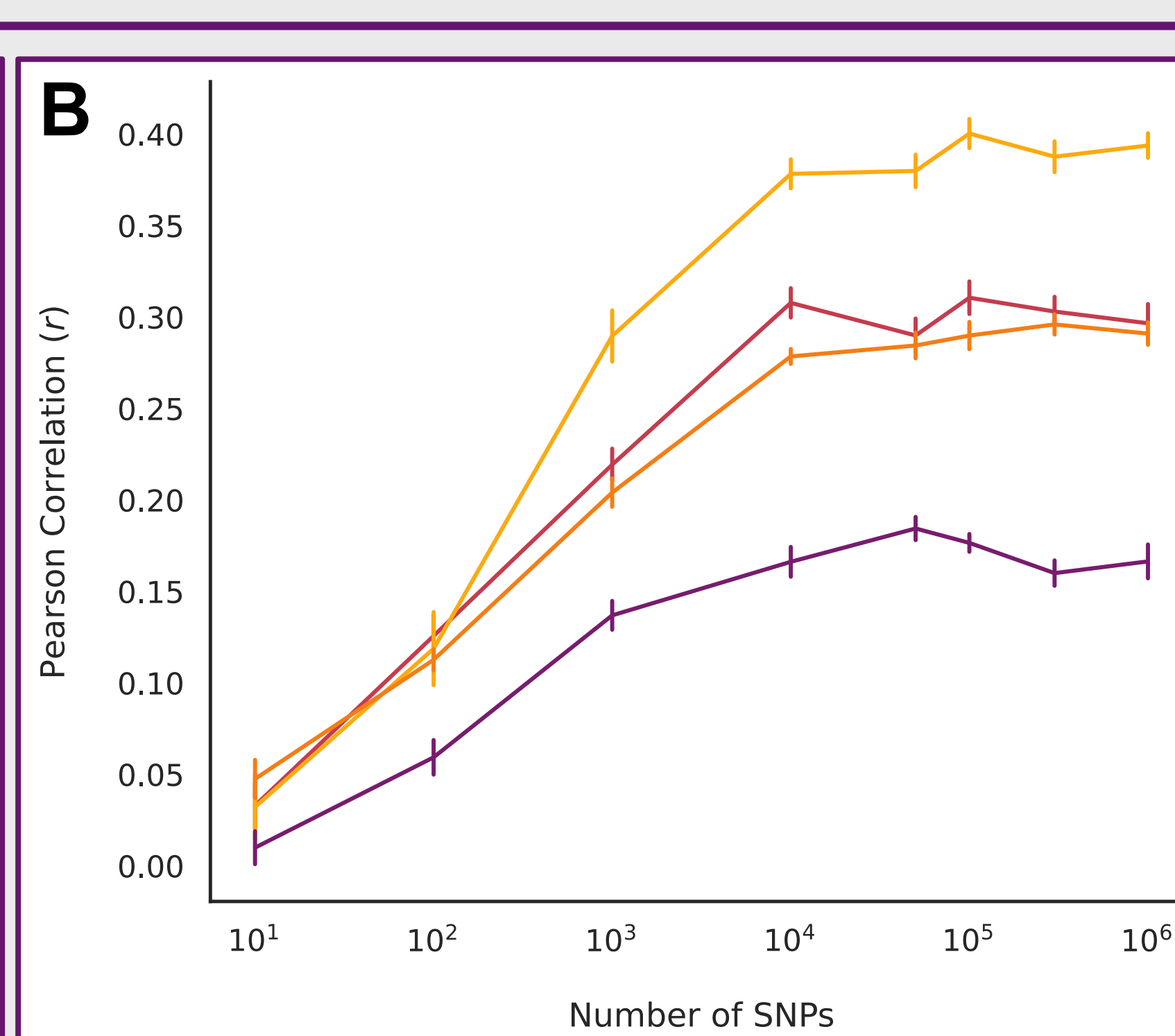
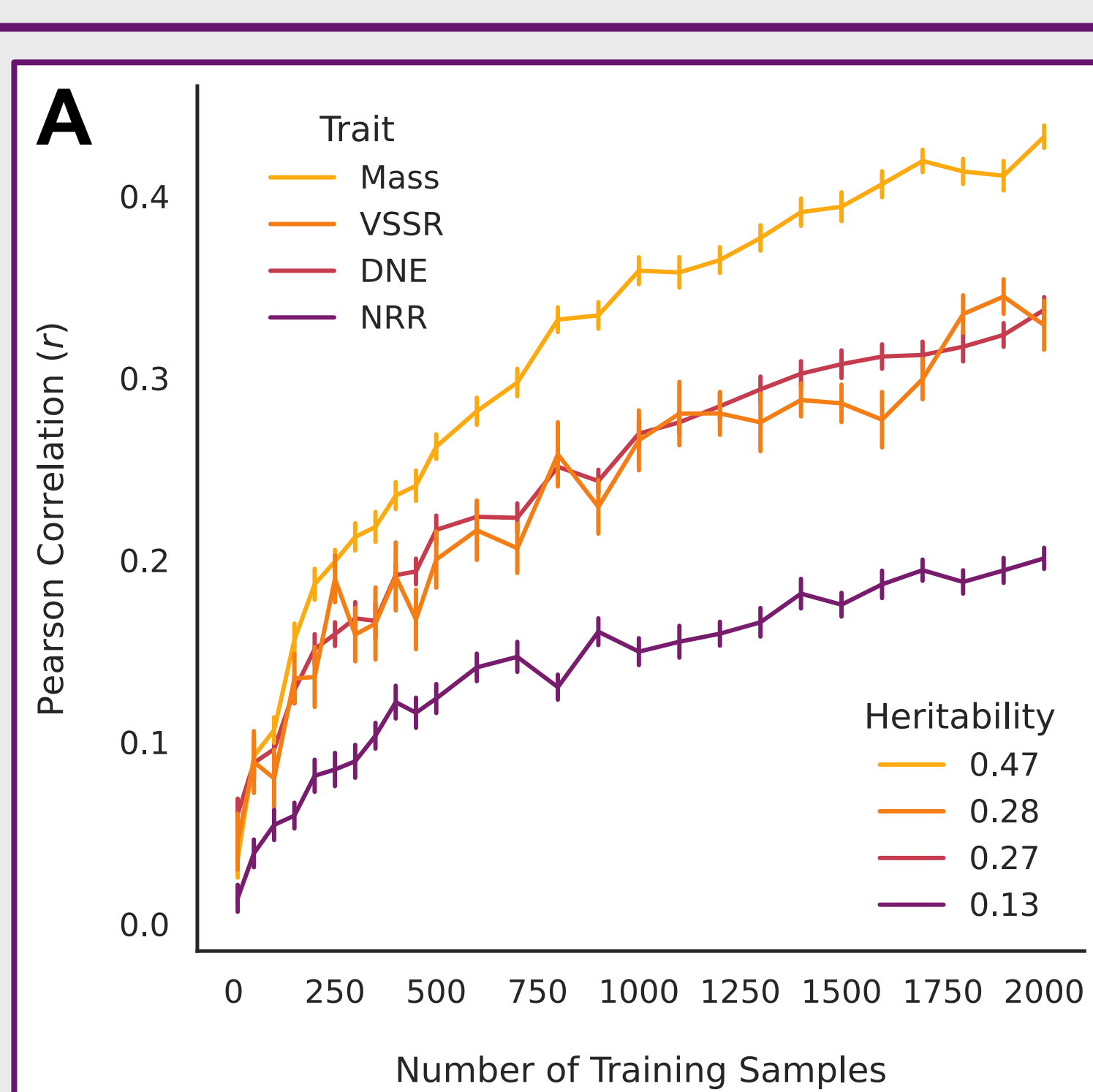


4 Genetic correlation



RAT Trait Ascertainment using Common Alleles: 1 The NIDA Center for GWAS in Outbred Rats maintains a database of experimental phenotype measurements and genome-wide SNP variation from >16,000 HS rats used in previous GWASs. Each new generation of rats bred at the Center is genotyped at weaning. 2 BLUP estimates of genetic effect sizes from the measured (training) sample are used to predict phenotypes for traits of interest on new genotypes. 3 Samples with predicted extreme trait values are selected from the new population for inclusion in later experiments. 4 Correlations between trait predictions and experimental measurements of a new trait offer preliminary evidence for a genetic correlation between the two traits.

THE DETAILS



Prediction performance depends heavily on trait heritability (h^2): the higher the h^2 , the more accurate our predictions (A+B). Training sample size also positively influences performance: larger phenotyped samples used to train models result in more accurate predictions (A). The number of SNPs used to train models is less impactful: Greater SNP samples increase prediction accuracy at low sample sizes, but performance plateaus above ~10K SNPs (B). (Accuracy = Pearson correlation r between measured and predicted traits)

Trait predictions successfully distinguish extreme samples with distinct mean trait values. BLUP predictions of genomic-estimated breeding values (GEBVs) correlate with experimentally observed phenotypes (C). Selecting the upper and lower 5% of predicted GEBVs produces selected samples that overlap in their distribution of trait values (top marginal histogram), but prediction accuracy is high enough that mean values are statistically distinct (D). The difference between groups increases with increasing trait heritability (h^2). T-test $P < 1 \times 10^{-3}$ for all traits.

¹Dept. Psychiatry, University of California San Diego School of Medicine, La Jolla, CA, USA

²College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

³Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA

*bbjohnson@health.ucsd.edu

°aapalmer@health.ucsd.edu