

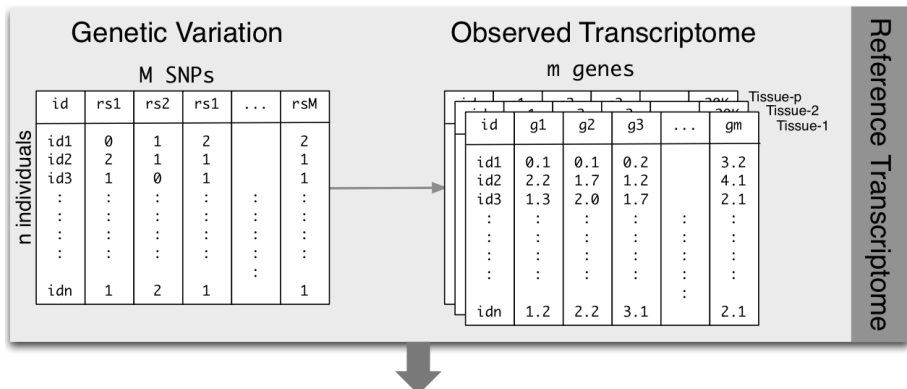
Understanding the genetic architecture of gene expression

Heather E. Wheeler, PhD

The University of Chicago
hwheeler@bsd.uchicago.edu

February 18, 2015

PrediXcan Step 1: Build and Test Predictors



PrediXcan Step 2: Build database of Best Predictors

PredictDB: Database
of Prediction Models

M SNPs

	rs1	rs2	rs3	...	rsM
g1	w11	w12	w13		w1M
g2	w21	w22	w23		w2M
g3	w31	w32	w33		w3M
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
gm	wm1	wm2	wm3		wmM

m genes

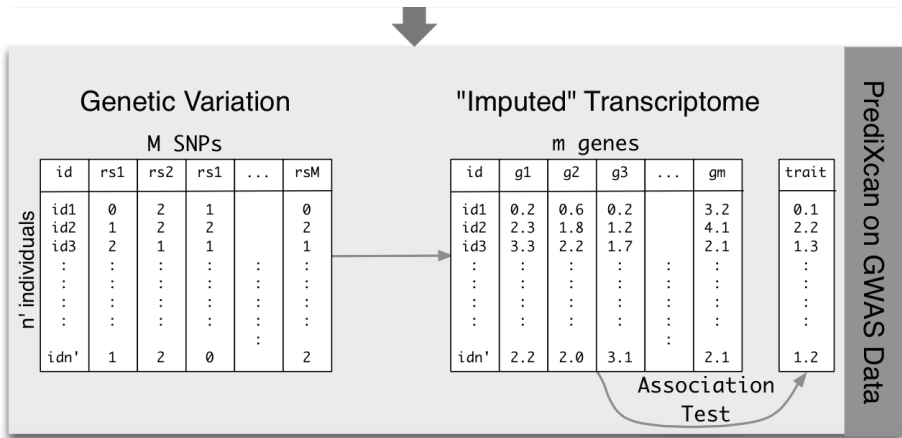
Tissue-p
Tissue-2
Tissue-1

Additive model of gene
expression trait trained in
reference transcriptome
datasets

$$T = \underbrace{\sum_k w_k X_k}_{GReX} + \epsilon$$

Weights stored in PredictDB

PrediXcan Step 3: Impute gene expression and test for association with phenotype



Explore the Genetic Architecture of Transcriptome Regulation

Optimizing predictors for PrediXcan also tells us about the underlying genetic architecture of gene expression.

We can ask what proportion of genes have:

- *cis* vs. *trans* effects
- sparse vs. polygenic effects
- cross-tissue vs. tissue-specific effects

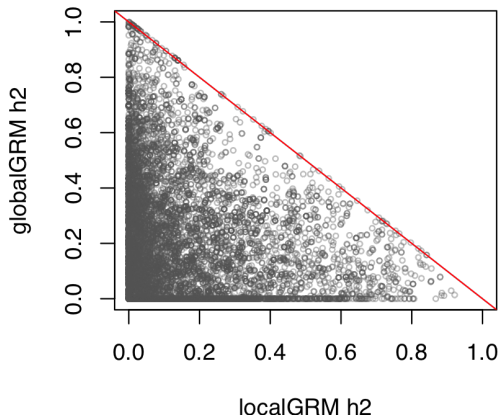
Primary cohort: DGN

- Battle et al. “Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals.” *Genome Research* 2014, 24(1):14-24
- Whole blood from Depression Genes and Networks study
- $n = 922$
- RNA-seq: “normalized gene-level expression data used for trans-eQTL analysis. The data was normalized using HCP (Hidden Covariates with Prior) where the parameters were optimized for detecting ‘trans’ trends”
- 600K genotypes: I have imputed to 1000 Genomes, but some earlier analyses were genotyped data only.

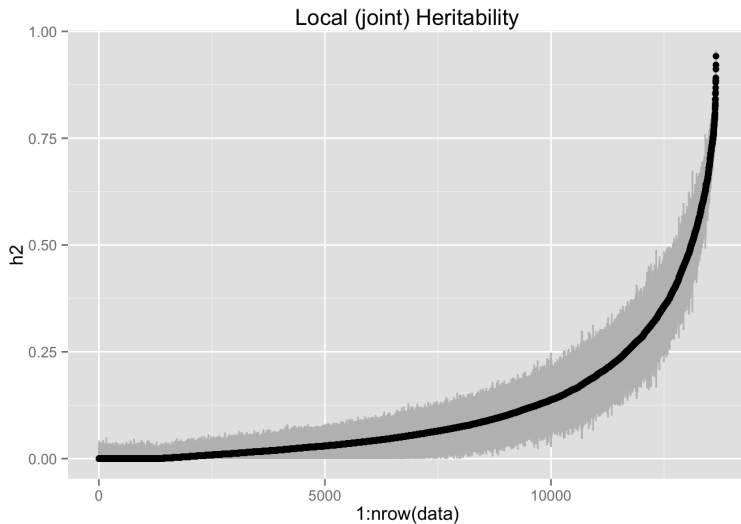
cis vs. *trans* effects

Estimate the heritability of gene expression in a joint analysis: localGRM (SNPs w/in 1Mb) + globalGRM (all SNPs)

DGN-WB GCTA

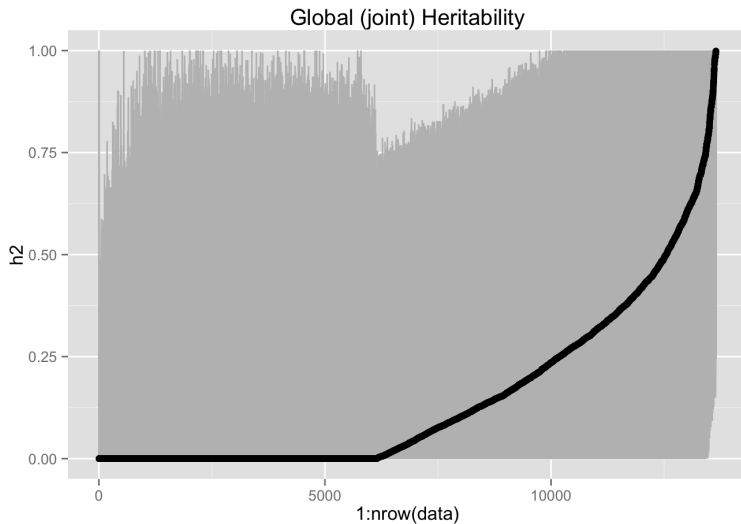


Local (joint) sorted h^2 estimates with 95% CI from GCTA



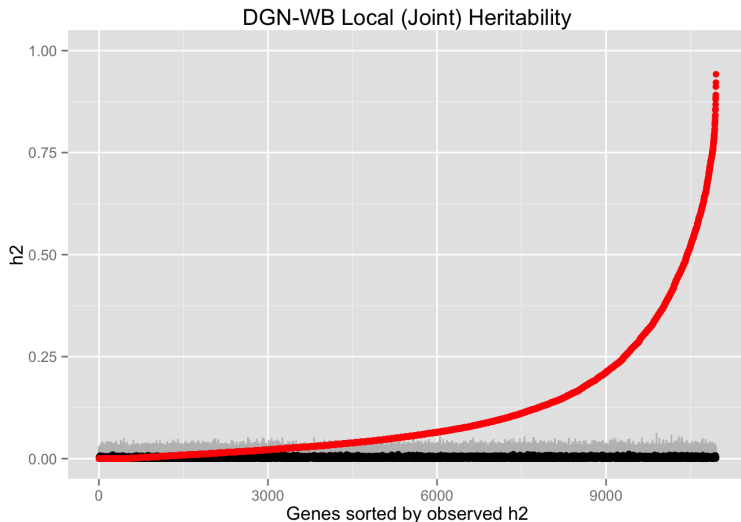
https://github.com/hwheeler01/cross-tissue/blob/master/analysis/sources/heritab_analysis.html

Global (joint) sorted h^2 estimates with 95% CI from GCTA

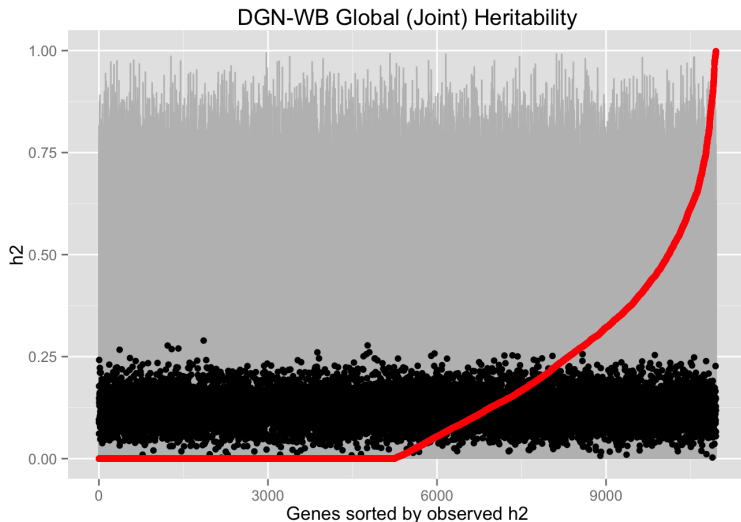


https://github.com/hwheeler01/cross-tissue/blob/master/analysis/sources/heritab_analysis.html

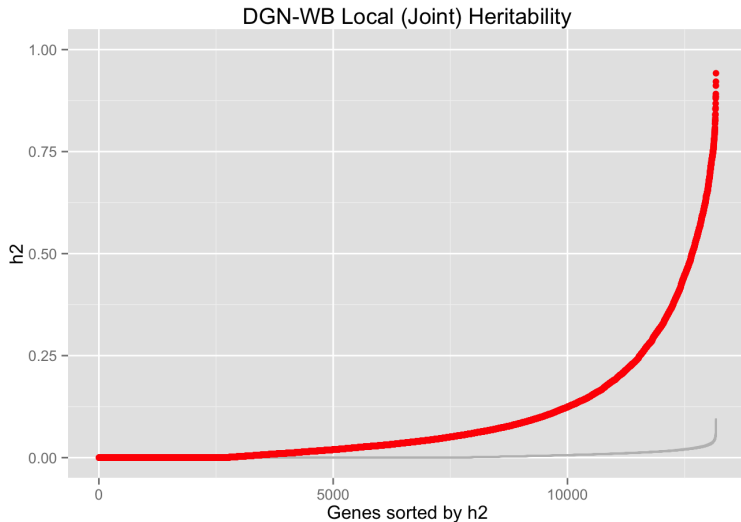
100 permutations to determine expected distribution of h^2 estimates



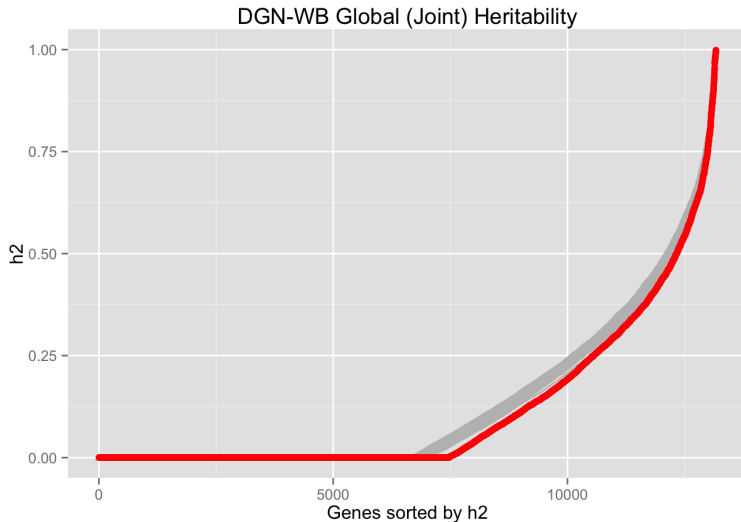
100 permutations to determine expected distribution of h^2 estimates



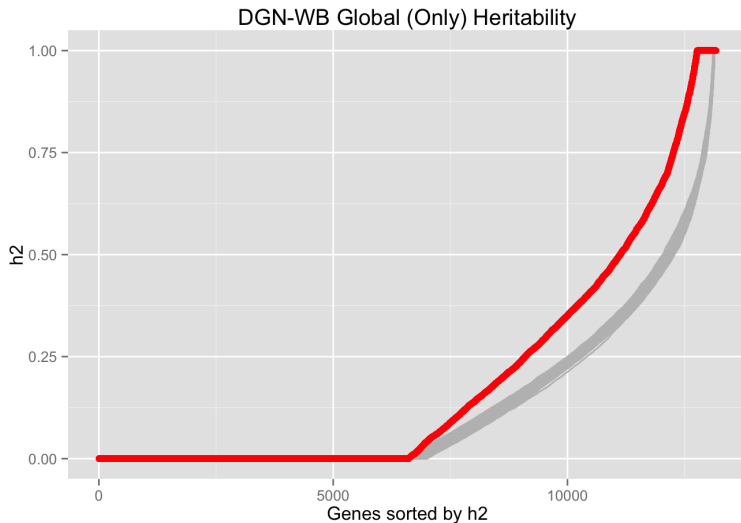
Sort the h^2 from each permutation



Sort the h^2 from each permutation



Sort the h^2 from each permutation



Try a larger sample to better capture *trans* effects

Framingham Heart Study

- $n = 5257$
- exon expression array and genotype array

sparse vs. polygenic effects

glmnet solves the following problem

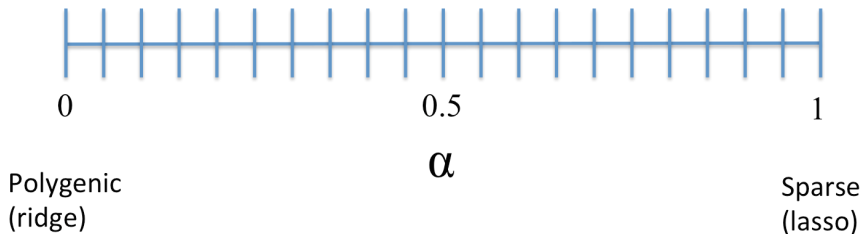
$$\min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N w_i l(y_i, \beta_0 + \beta^T x_i) + \lambda \left[(1 - \alpha) \|\beta\|_2^2 / 2 + \alpha \|\beta\|_1 \right],$$

over a grid of values of λ covering the entire range.

The elastic-net penalty is controlled by α , and bridges the gap between lasso ($\alpha = 1$, the default) and ridge ($\alpha = 0$). The tuning parameter λ controls the overall strength of the penalty.

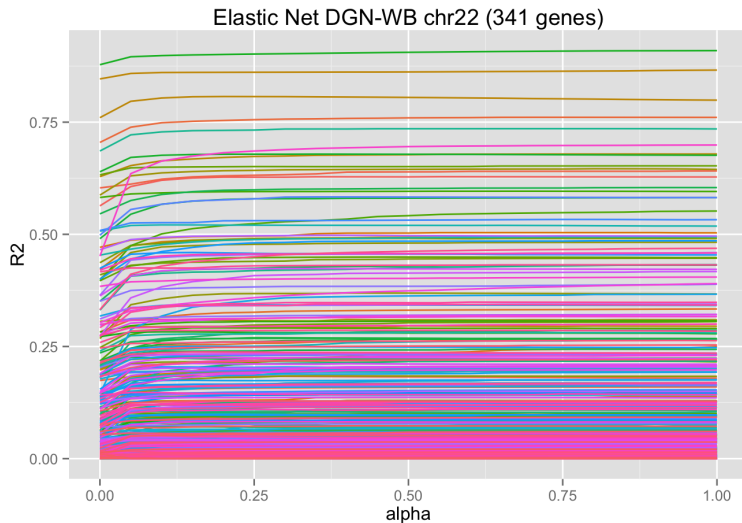
http://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html

sparse vs. polygenic effects

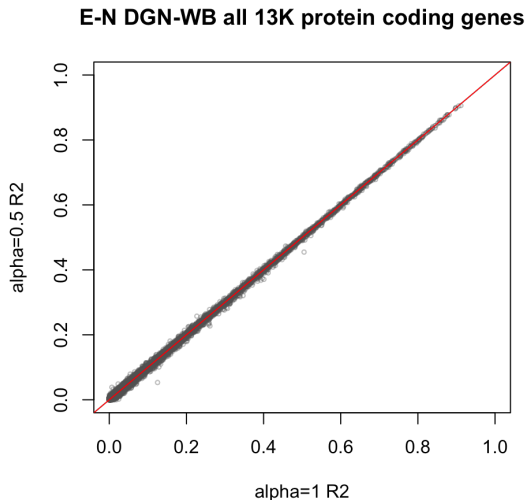


For each gene, determine α with best 10-fold CV predictive performance using *cis* SNPs.

Predictive performance consistent across most alphas



Predictive performance consistent between $\alpha=0.5$ and $\alpha=1$



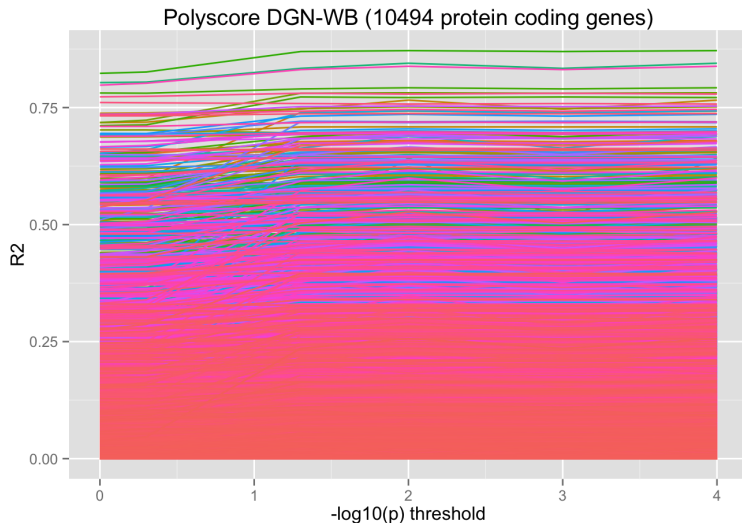
Also tested Polyscore predictive performance using 10-fold CV

$$expression = \sum \hat{w} * gt$$

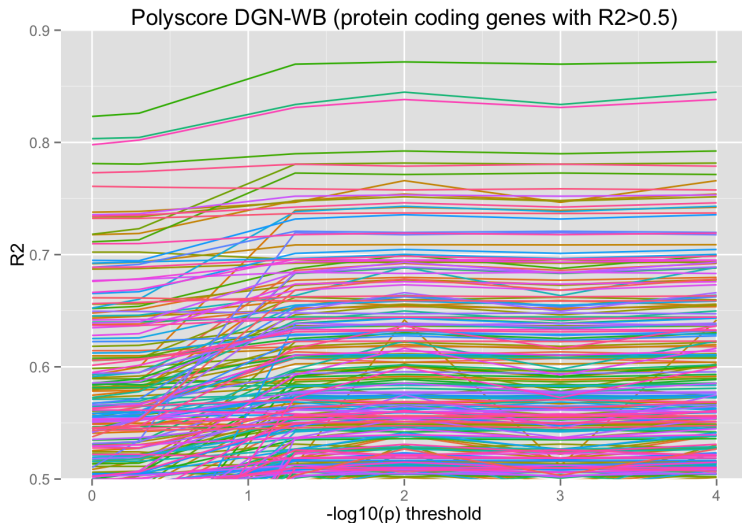
Single variant linear regression coefficients (w) at several P-value thresholds included in the additive model:

- $P < 0.0001$
- $P < 0.001$
- $P < 0.01$
- $P < 0.05$
- $P < 0.5$
- $P < 1$

Polyscore (*cis* SNPs only) predictive performance

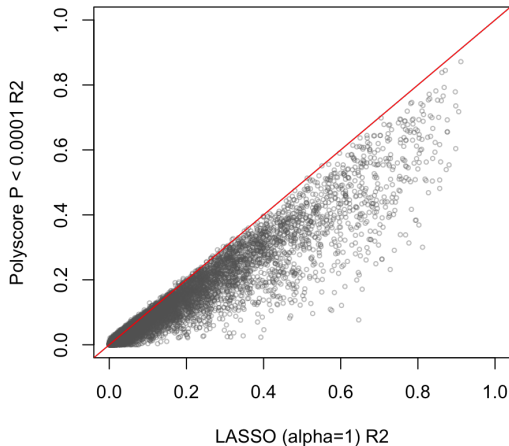


Polyscore (*cis* SNPs only) predictive performance

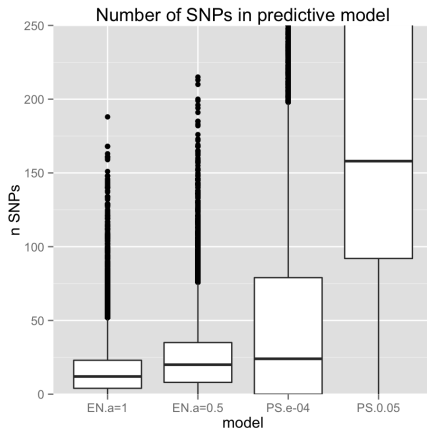
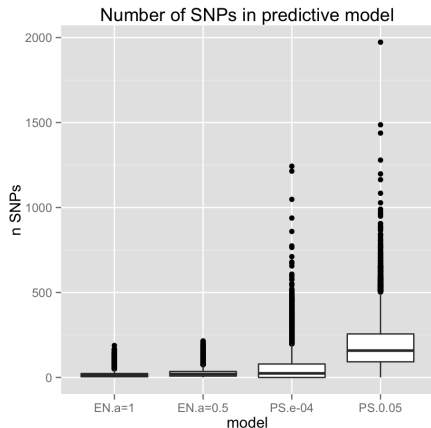


LASSO predicts gene expression better than Polyscore

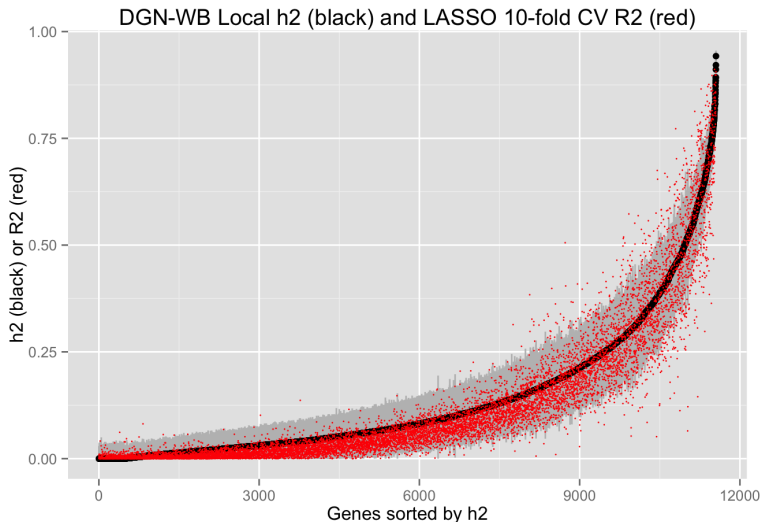
DGN-WB predictive performance



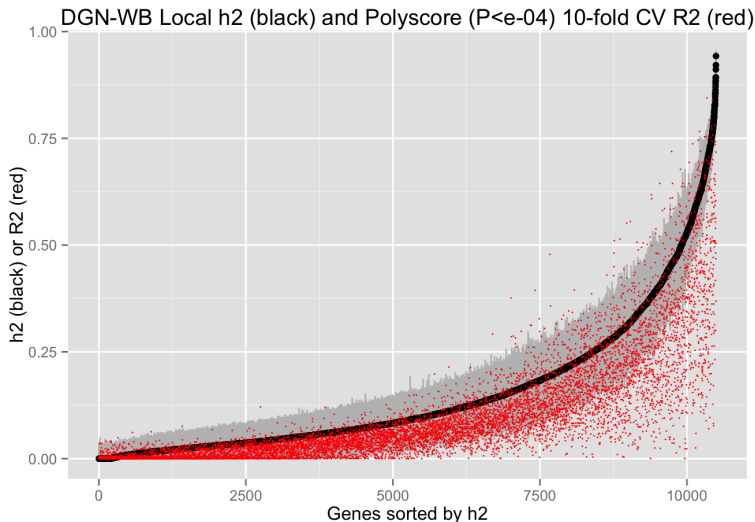
For robustness, consider EN ($\alpha=0.5$) for PrediXcan



LASSO predictive performance reaches (or exceeds?) local h^2 of most genes

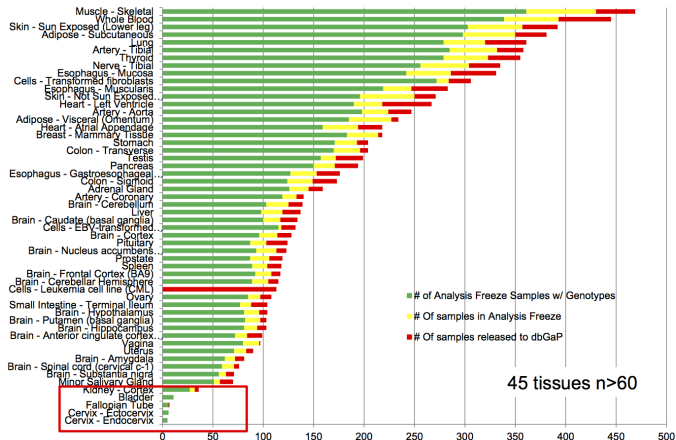


Polyscore predictive performance does not reach local h^2 of many genes



cross-tissue vs. tissue-specific effects with GTEx

RNA Seq Samples per tissue



Modeling cross-tissue expression

Linear mixed effect model using GTEx_Data_2014-06-13 release

- 8555 tissues across 544 subjects
- limited to ~17K protein coding genes

```
library(lme4)

fit <- lmer(expression ~ (1|SUBJID) + TISSUE
+ GENDER + PEERs)

#cross-tissue expression
fitranef <- ranef(fit)

#tissue-specific expression
fitresid <- resid(fit)
```

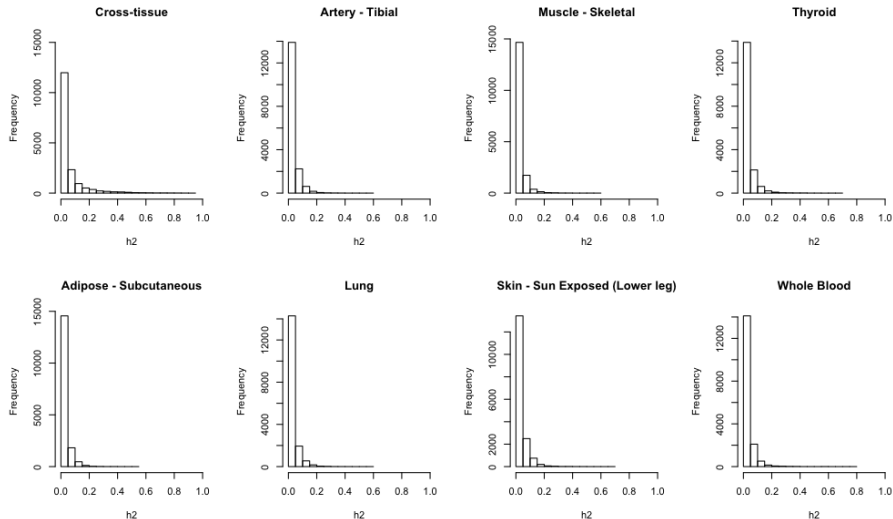
Estimating heritability with GCTA

Tested two genetic relationship matrix (GRM) models for each expressed gene

- localGRM (SNPs within 1 Mb of gene)
- localGRM + globalGRM (all SNPs)

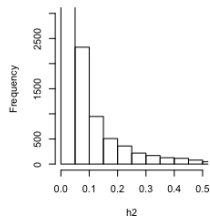
First pass: estimated h^2 of cross-tissue expression and tissue-specific expression in the 7 tissues with the most samples

GCTA heritability: localGRM h^2

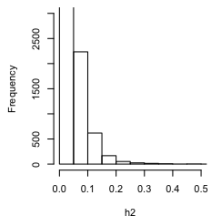


GCTA heritability: localGRM h^2 ZOOM

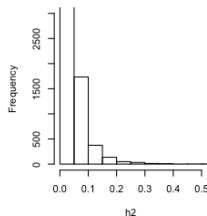
Cross-tissue



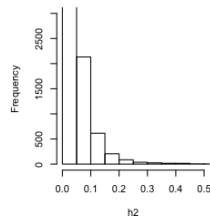
Artery - Tibial



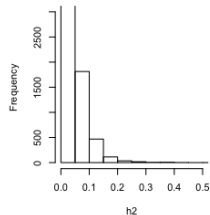
Muscle - Skeletal



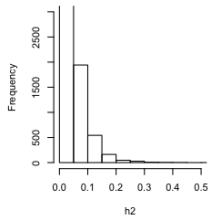
Thyroid



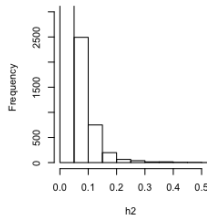
Adipose - Subcutaneous



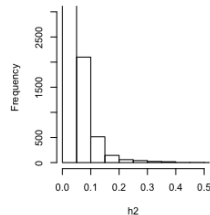
Lung



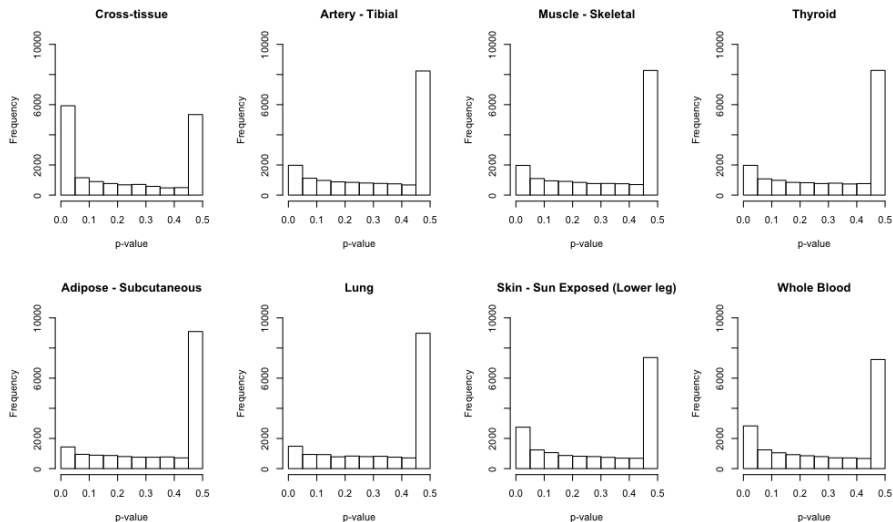
Skin - Sun Exposed (Lower leg)



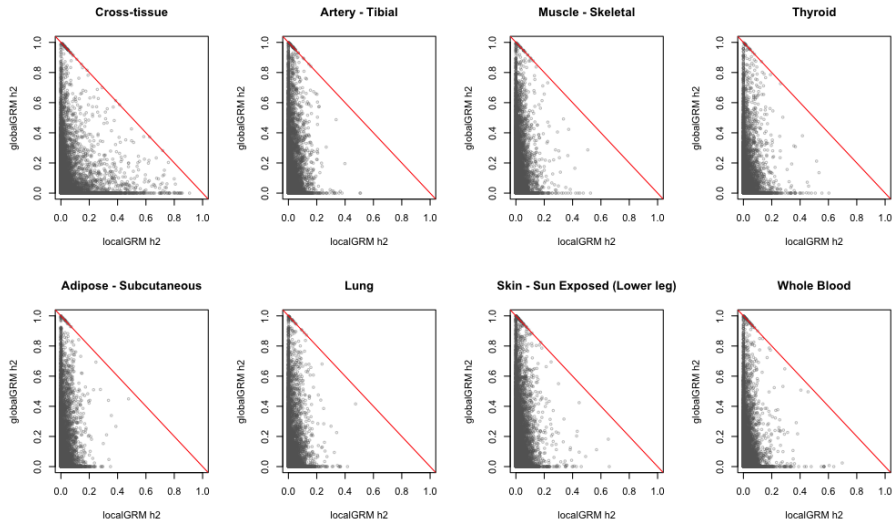
Whole Blood



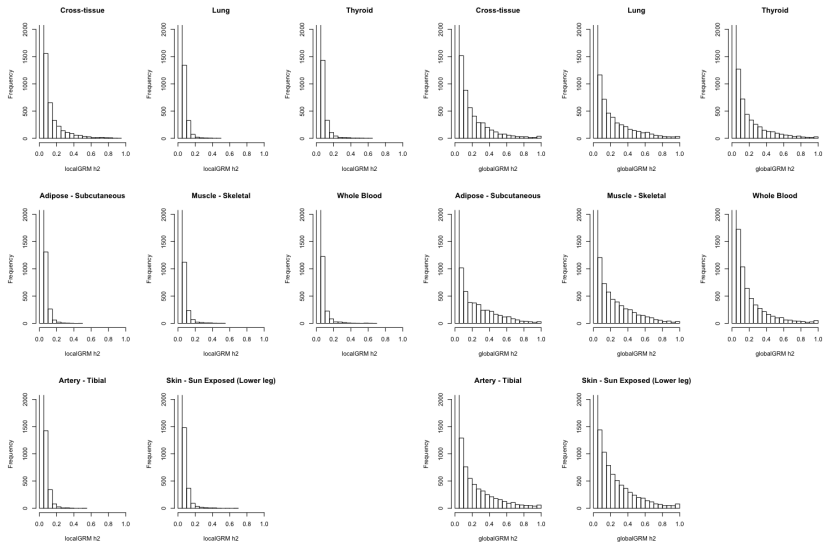
GCTA heritability: localGRM p-values



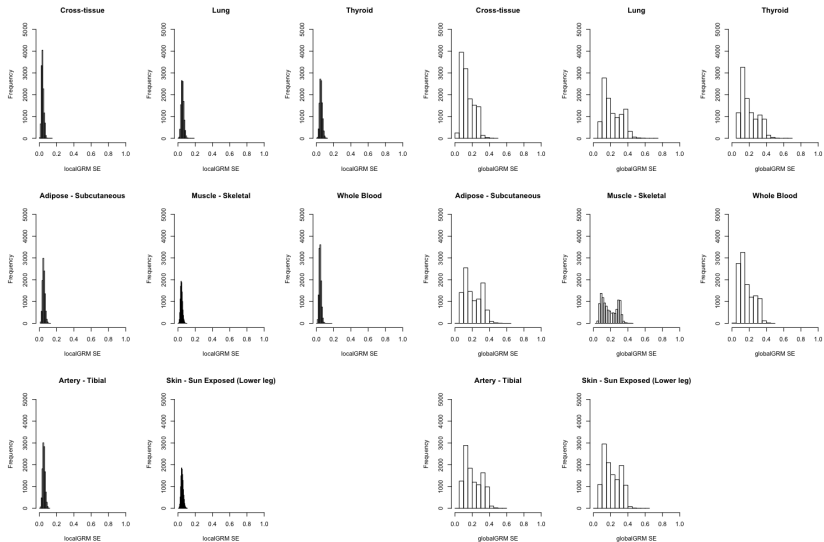
GCTA joint heritability: localGRM + globalGRM h^2



GCTA joint heritability: localGRM + globalGRM h^2



GCTA joint heritability: localGRM + globalGRM SE



***cis* vs. *trans* effects**

- *cis* effects are easily detected: 37% of DGN-WB genes have local $h^2 > 0.1$, 81% have local $h^2 > 0.01$
- good estimates of *trans* h^2 will require larger n

sparse vs. polygenic effects

- *cis* effects are mostly sparse (LASSO beats ridge and polyscore)

cross-tissue vs. tissue-specific effects

- GTEx cross-tissue local h^2 is greater than tissue-specific local h^2