

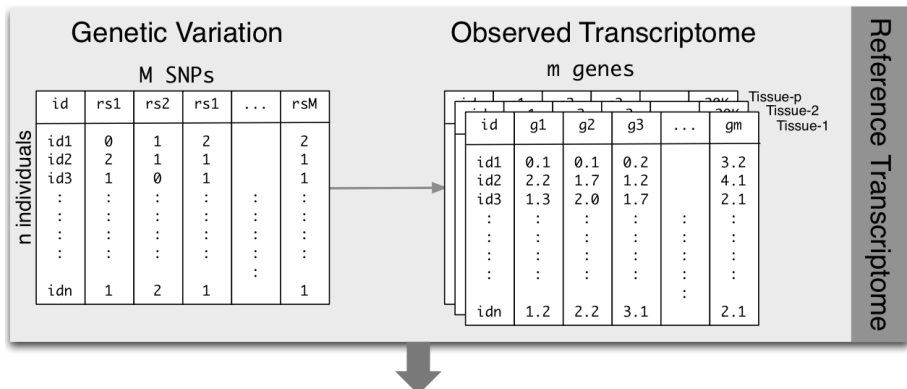
Understanding the genetic architecture of gene expression

Heather E. Wheeler, PhD

The University of Chicago
hwheeler@bsd.uchicago.edu

February 13, 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

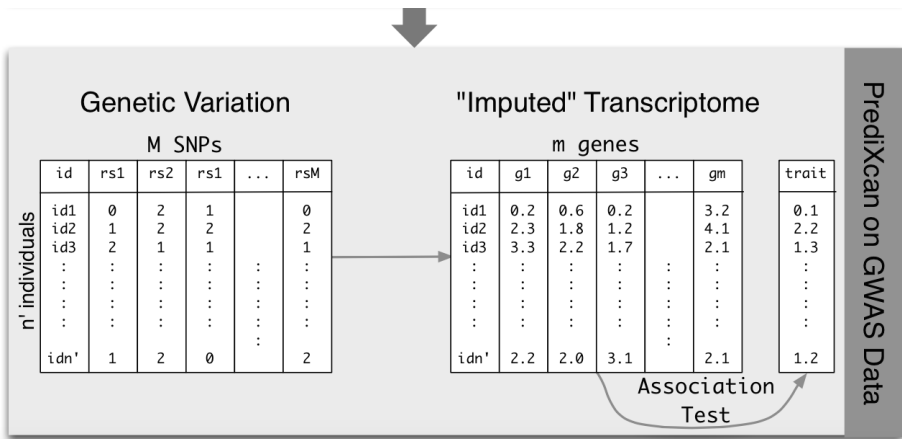
Tissue-p
Tissue-2
Tissue-1

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

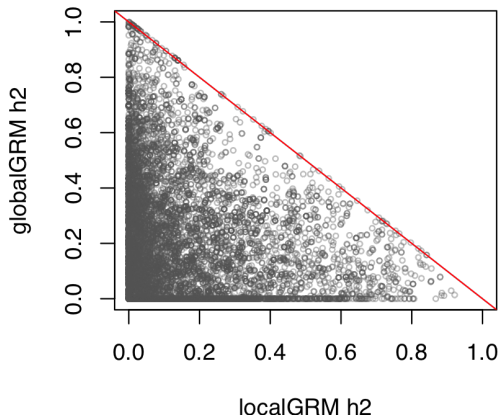
Primary cohort: DGN

- 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.

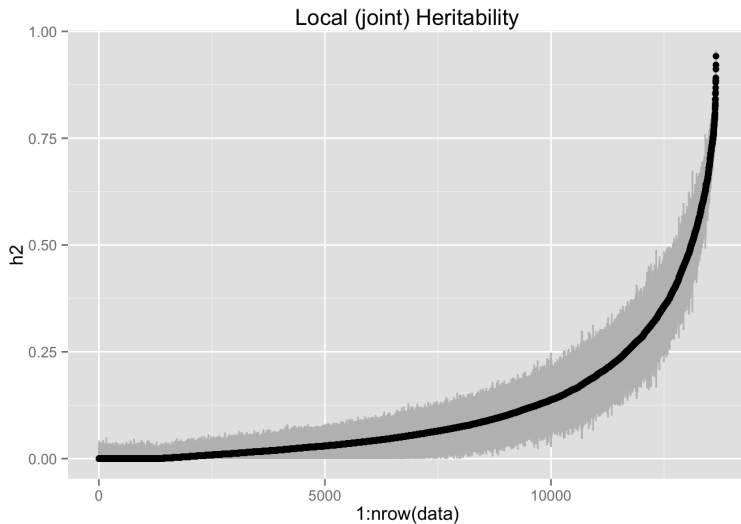
cis vs. *trans* effects

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

DGN-WB GCTA

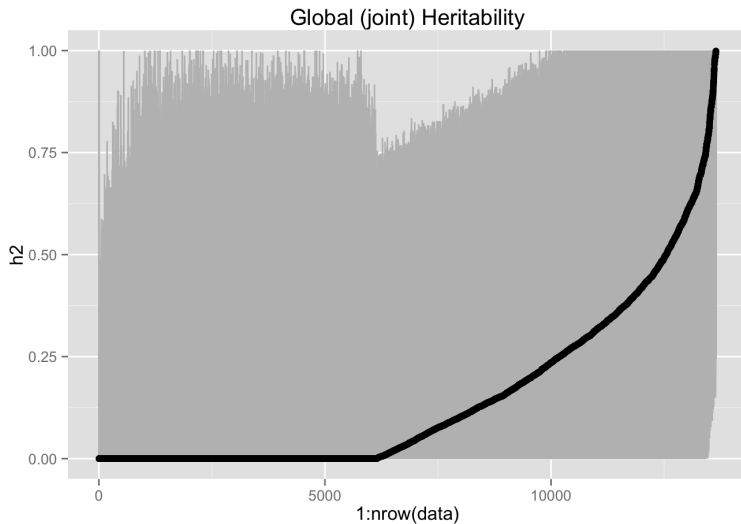


Local (joint) sorted h^2 estimates with 95% CI from GCTA



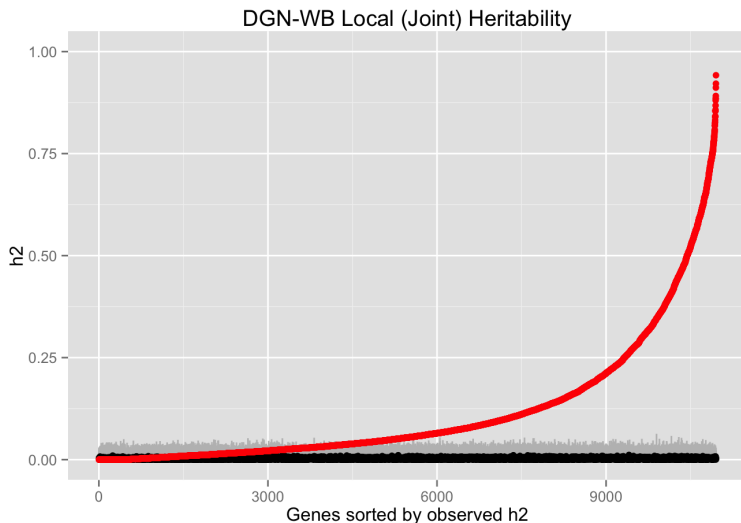
https://github.com/hwheeler01/cross-tissue/blob/master/analysis/sources/heritab_analysis.html

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

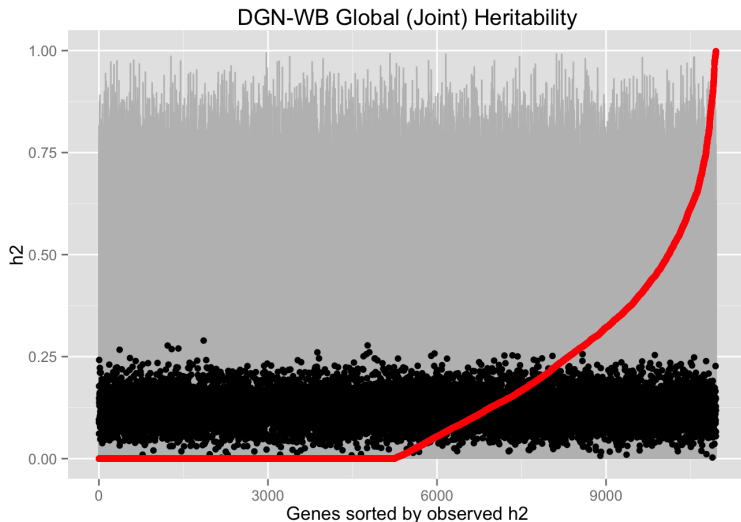


https://github.com/hwheeler01/cross-tissue/blob/master/analysis/sources/heritab_analysis.html

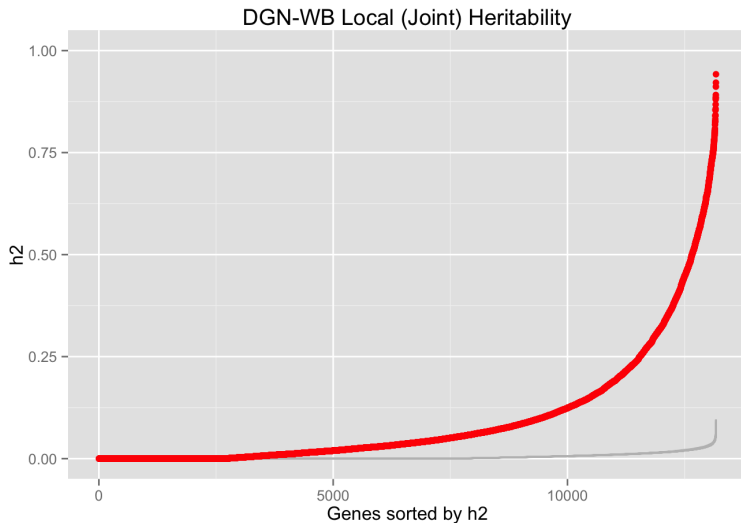
100 permutations to determine expected distribution of h^2 estimates



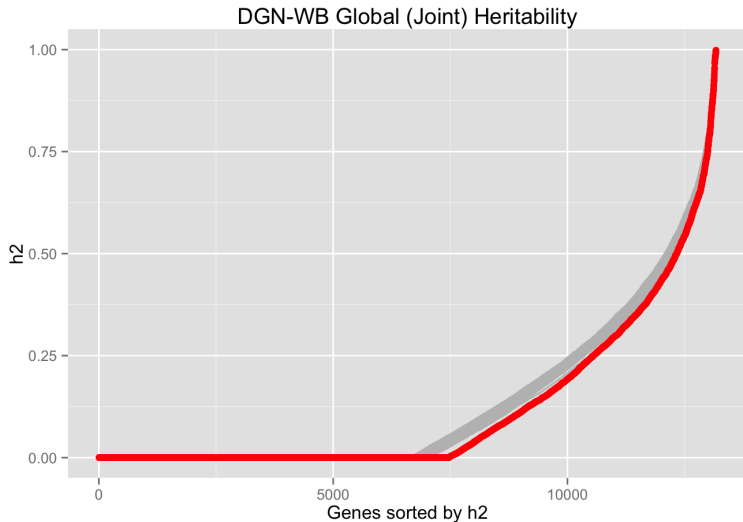
100 permutations to determine expected distribution of h^2 estimates



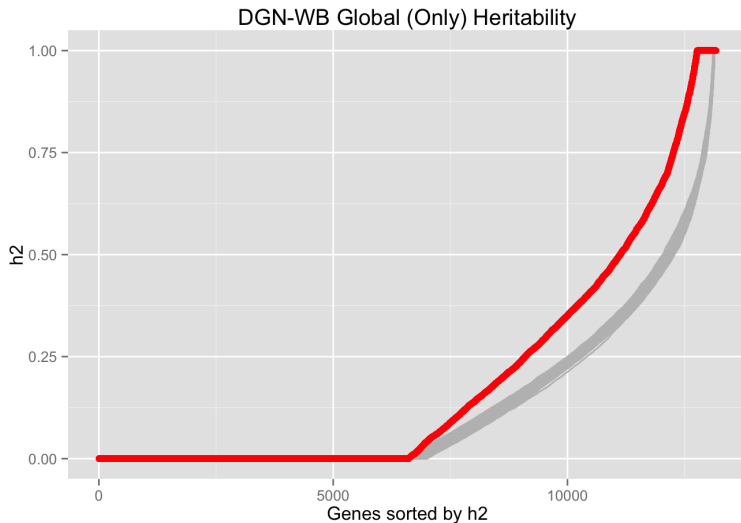
Sort the h^2 from each permutation



Sort the h^2 from each permutation



Sort the h^2 from each permutation



Modeling cross-tissue expression

Linear mixed effect model

```
library(lme4)

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

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

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