HERITABILITY OF GENE EXPRESSION

Joseph Powell

SISG- 2018

CONTENTS

- What is heritability
- How do we estimate it?
- What are the results?

OBSERVATION

The expression levels of many (most) transcripts vary across individuals.

OBSERVATION

The expression levels of many (most) transcripts vary across individuals.

Why do they vary?

OBSERVATION

The expression levels of many (most) transcripts vary across individuals.

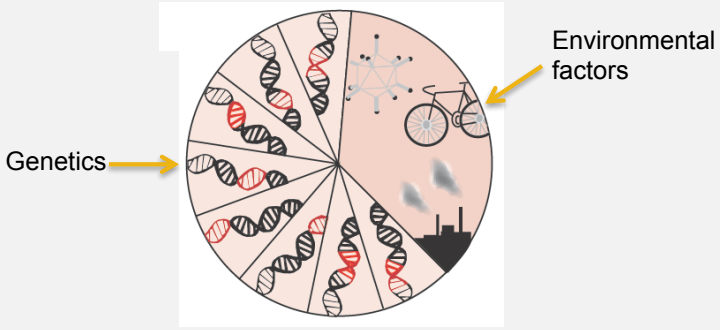
Why do they vary?

(1) Differences in the environment between individuals

(2) Technical variation in the sample collection, preparation and sequencing

(3) Stochastic variation

(4) Genetic variation between individuals



HERITABILITY

- Heritability: is a statistic that provides an estimate of how much variation in a trait in a population is due to genetic variation among individuals.
- Other causes of variation in a trait are characterized as environmental factors.

HERITABILITY

- Heritability: is a statistic that provides an estimate of how much variation in a trait in a population is due to genetic variation among individuals.
- Other causes of variation in a trait are characterized as environmental factors.

$$\text{Phenotype (P)} = \text{Genotypes (G)} + \text{Environment (E)}$$

$$\text{Var(P)} = \text{Var(G)} + \text{Var(E)} + 2(\text{Cov(G,E)})$$

HERITABILITY

- Heritability: is a statistic that provides an estimate of how much variation in a trait in a population is due to genetic variation among individuals.
- Other causes of variation in a trait are characterized as environmental factors.

$$\text{Phenotype (P)} = \text{Genotypes (G)} + \text{Environment (E)}$$

$$\text{Var(P)} = \text{Var(G)} + \text{Var(E)} + 2(\text{Cov(G,E)})$$

$$\text{Var}(P) = \text{Var}(G) + \text{Var}(E) + 2\ \text{Cov}(G,E).$$

In a planned experiment Cov(G,E) can be controlled and held at 0. Thus,

HERITABILITY

$H^2=Var(G)/Var(P)$

There are different forms of Var(G). H^2 is the broad-sense heritability.

Var(G) includes all genetic contributions to phenotypic variance including additive (A), dominant, epistatic, as well as maternal and paternal effects.

Additive only heritability is calculated,

$h^2=Var(A)/Var(P)$

ESTIMATING HERITABILITY FOR GENE EXPRESSION

- What data do we need?
  - Genotypes
  - Normalized gene expression levels
  - Covariates

- What else do we need to consider?
  - Are the sample matched? (Look up MixUpMapper)
  - Population stratification

ESTIMATING HERITABILITY FOR GENE EXPRESSION

Twin model

Heritability estimates in humans are commonly made using the resemblance between monozygotic (MZ) and dizygotic (DZ) twins.

MZ twins are genetically identical whereas DZ twins, on average, have 50% of their alleles identical by descent (IBD).

The correlation of mRNA transcript levels in MZ (rMZ) and DZ (rDZ) can be used to estimate the additive genetic contribution (VA) to phenotypic variance by;

VA = 2(rMZ − rDZ).

The contribution of environmental variance (VE) can be estimated by subtracting rMZ from1 as inVE = 1 − rMZ

and the contribution of common environmental effects (VC) byVC = rMZ − VA.

ESTIMATING HERITABILITY FOR GENE EXPRESSION

Parent-offspring

- Parents share 50% of their alleles IBD with their offspring. If we regress the transcript levels of an mRNA transcript measured in offspring again the levels measured in their parents then the slope of the regression (β) is equal to
- $\beta = cov(P,O)/Var(P) = ½\ h^2$.
- In other words, the heritability can be estimated as $2 * \beta$. This method assumes no common environmental effects.

ESTIMATING HERITABILITY FOR GENE EXPRESSION

Population based

- For example, the GREML method (Yang *et al*. NG 2010, Powell *et al*. NRG, 2010) uses a linear mixed-effects model:
- $y = g + e$, where y is a $nx1$ vector of normalized gene expression levels for a transcript; g is n∗1 vector of random polygenic effects with $g \sim N(0, Var(G)\boldsymbol{A})$, with $\boldsymbol{A}$ the genetic relationship matrix (GRM) estimated from common SNPs; and is a $nx1$ vector of residuals with
- $e \sim N(0, Var(E)\mathbf{I})$, with **I** as the incidence matrix.

SOFTWARE

Related and unrelated individuals in the study design

PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses

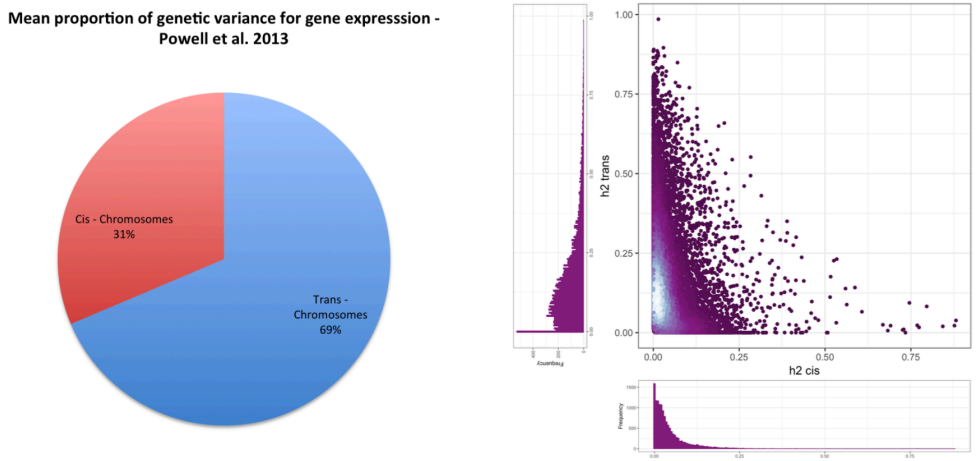
Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I. W. de Bakker, Mark J. Daly, and Pak C. Sham

plink - *(pngu.mgh.harvard.edu/ purcell/plink/)*

Unrelated individuals in the study design

GCTA - emphhttp://cnsgenomics.com/software/gcta/

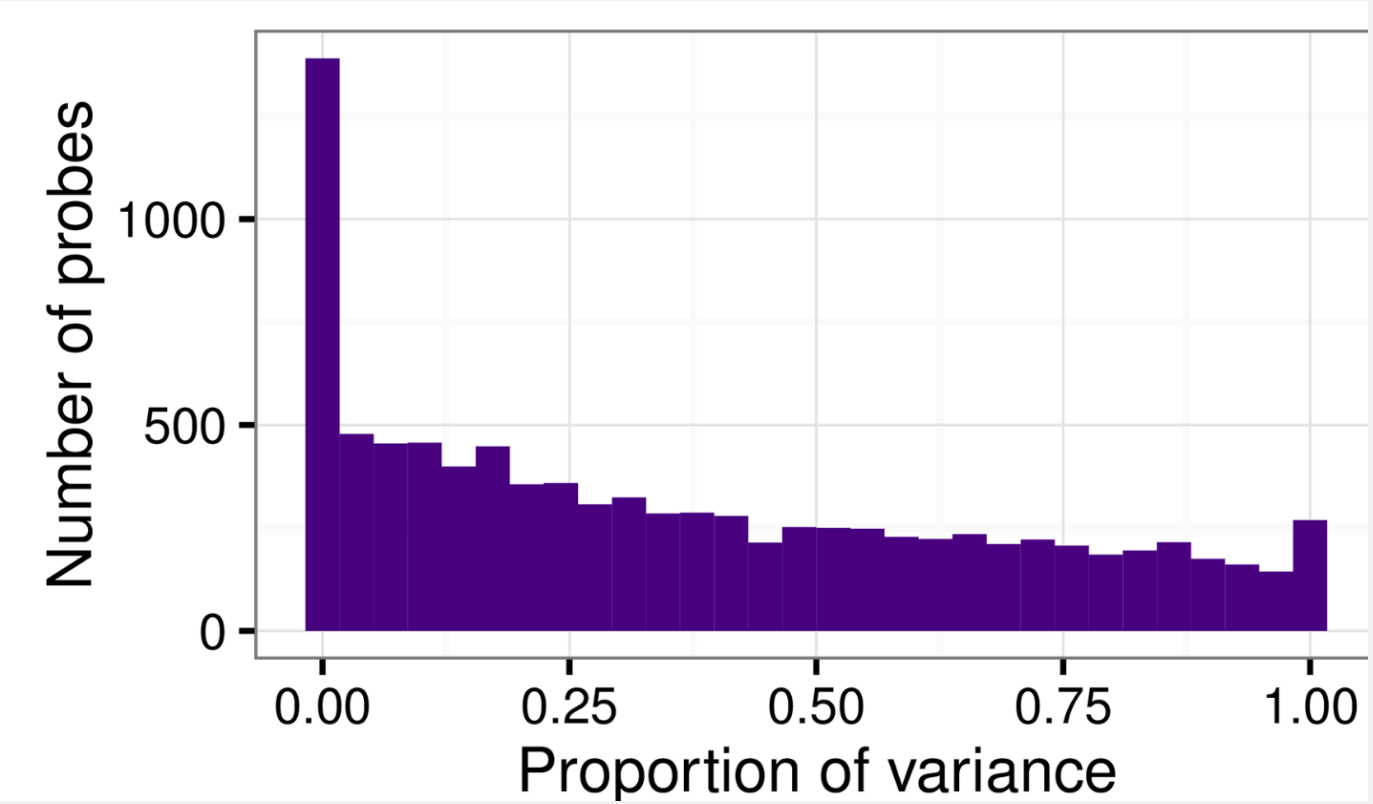
TYPICAL RESULTS



The majority of heritability is located on *trans* chromosomes

What about other forms of genomic architecture?

NOT ALL GENES ARE THE SAME



WHAT ABOUT DIFFERENT TISSUES?

Table: Summary of the estimates of heritability for gene expression levels from large-scale studies.

Study	Tissue	N Transcripts or Genes	Mean h^2	Method	Sample Size
Dixon *et al.*	LCLs	20,599	0.23	Sib pairs	400
Price *et al.*	Peripheral blood	18,735	0.16	Population IBD	687
Price *et al.*	Adipose tissue	19,099	0.24	Population IBD	496
Wright *et al.*	Peripheral blood	18,392	0.14	Twin model	2,752
Powell *et al.*	LCLs	9,555	0.38	Twin model	100
Powell *et al.*	Peripheral blood	9,555	0.32	Twin model	100
Powell *et al.*	Peripheral blood	17,994	0.24	Complex family	862
Grundberg *et al.*	Adipose tissue	23,596	0.26	Twin model	714
Grundberg *et al.*	Skin	23,596	0.16	Twin model	540
Grundberg *et al.*	LCLs	23,596	0.21	Twin model	718
Lloyd-Jones *et al.*	Peripheral blood	36,778	0.192	Population	2813

THANK YOU

- Email me: j.powell@garvan.org.au
- Twitter: @JP_Garvan