



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

CREATE CHANGE

Lecture 10 : Estimation of Non-Additive Genetic Variance and Genetic Correlation

Genetics & Genomics Winter School

Valentin Hivert

contact: v.hivert@imb.uq.edu.au

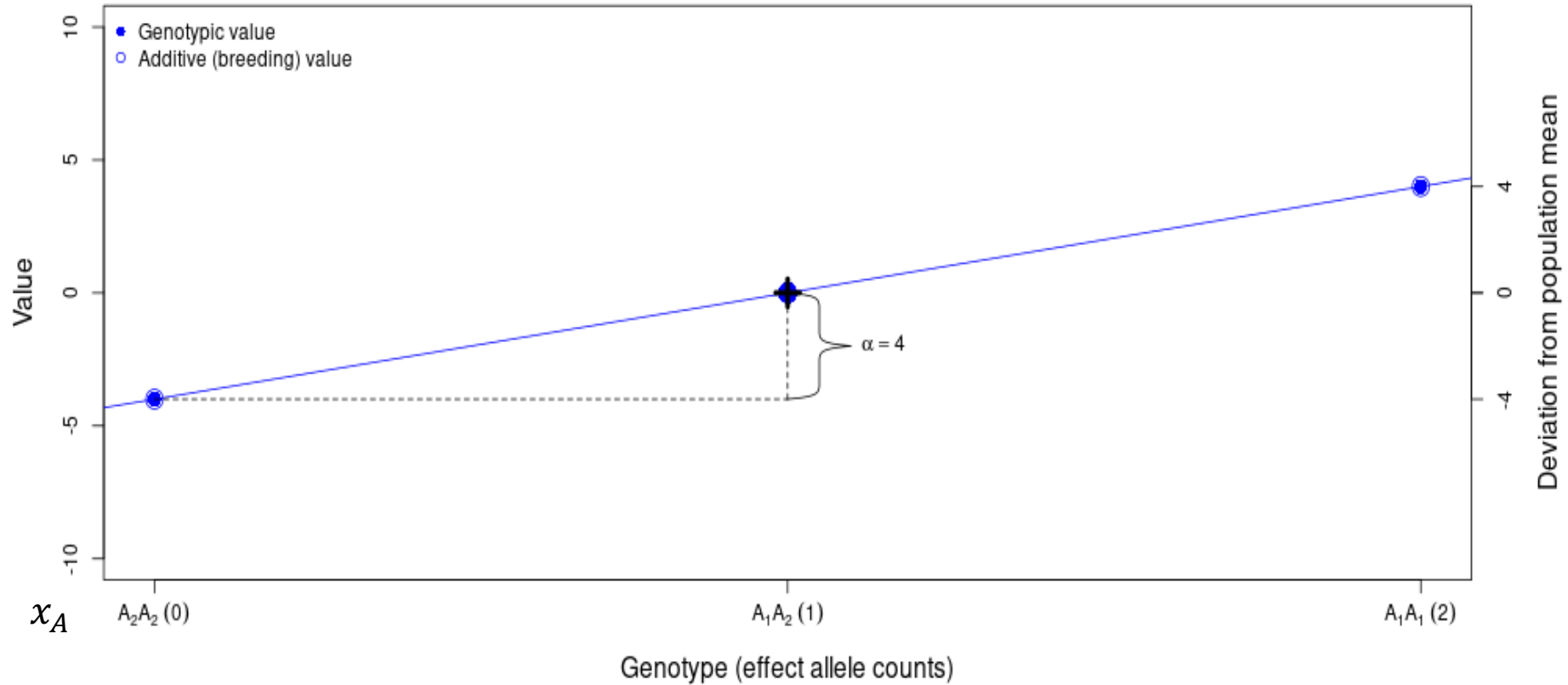
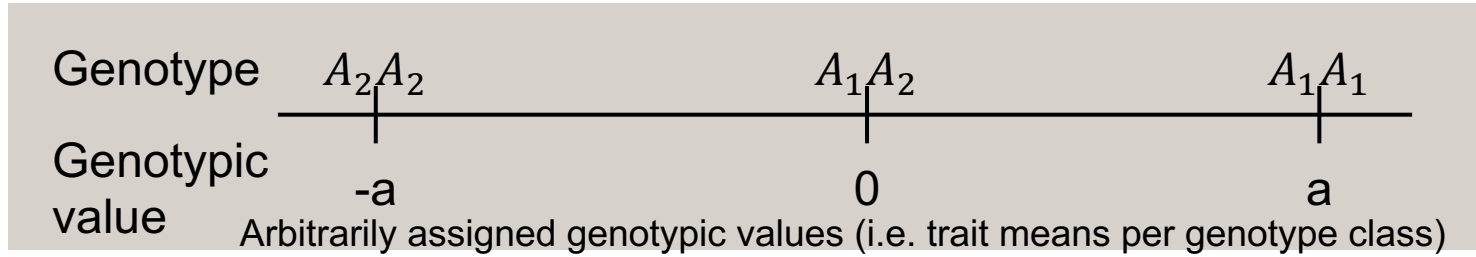
Institute for Molecular Bioscience



Program in Complex
Trait Genomics

Part I: Estimation of non-additive genetic variance

$$Y = G + E$$



$$\alpha = a$$

$$\sigma_A^2 = 2p(1-p)\alpha^2$$

$$Y = G_A + E$$

Assuming an additive model, we have:

$$Y = \mu + \sum_j w_{A(j)} \alpha_j + e$$

$$\text{With } w_{A(j)} = \frac{x_{A(j)} - 2p_j}{\sqrt{2p_j(1-p_j)}}$$

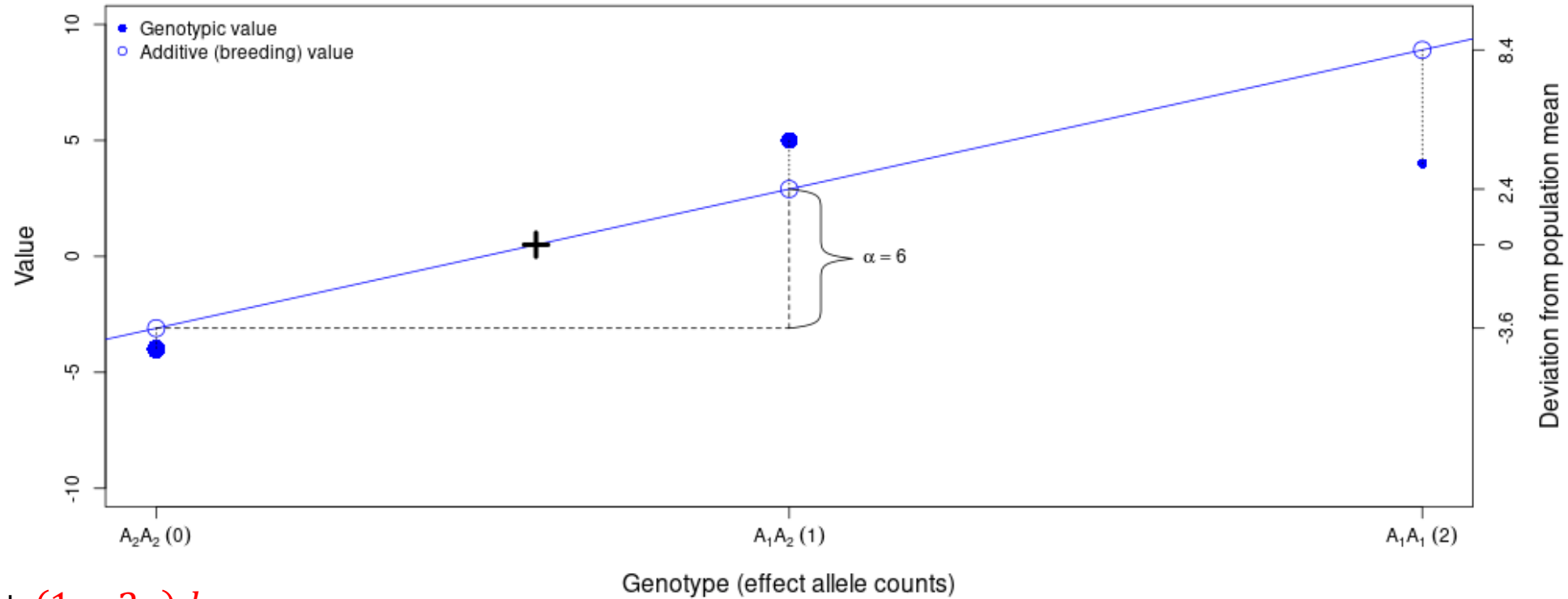
$$\text{Var}(Y) = \Theta_A \sigma_A^2 + I \sigma_e^2$$

With Θ_A the additive Genomic Relationship Matrix (GRM).

We can estimate $h_{SNP}^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2}$ using GREML or HE regression in GCTA.

The total genetic variance of a trait can be partitioned into additive, dominance and epistatic variance:

$$\text{Var}(Y) = \Theta_A \sigma_A^2 + \Theta_D \sigma_D^2 + \Theta_{AA} \sigma_{AA}^2 + \dots + I \sigma_e^2$$



$$\alpha = a + (1 - 2p)d$$

$$\sigma_A^2 = 2p(1 - p)\alpha^2$$

$$\sigma_D^2 = [2p(1 - p)d]^2$$

Falconer Shiny App: <https://shiny.cnsgenomics.com/Falconer2/>

The total genetic variance of a trait can be partitioned into additive and dominance variance:

$$\text{Var}(Y) = \Theta_A \sigma_A^2 + \Theta_D \sigma_D^2 + I \sigma_e^2$$

- Many effort to characterize non-additive genetic effect from twin studies, **BUT confounding effect with shared environmental factors**
- Zhu *et al.* (2015) proposed to estimate dominance variance from a sample of **unrelated individuals**

The total genetic variance of a trait can be partitioned into additive and dominance variance:

$$\text{Var}(Y) = \Theta_A \sigma_A^2 + \Theta_D \sigma_D^2 + I \sigma_e^2$$

- $\Theta_A = G$ is the additive Genomic Relationship Matrix (GRM)
- Θ_D is the dominance GRM where the dominance genetic relatedness between two individuals j and k is computed as:

$$\pi_{D(jk)} = \frac{1}{m} \sum_i \frac{(x'_{D(ij)} - 2p_i^2)(x'_{D(ik)} - 2p_i^2)}{4p_i^2(1-p_i)^2}, \text{ with } x'_D = 0, 2p, 4p - 2 \leftarrow \text{ensure the orthogonality of the model}$$

```
gcta --bfile input_file --make-grm-d --out dominance_grm
```

The total genetic variance of a trait can be partitioned into additive, dominance and epistatic variance:

$$\text{Var}(Y) = \Theta_A \sigma_A^2 + \Theta_D \sigma_D^2 + \Theta_{AA} \sigma_{AA}^2 + I \sigma_e^2$$

We aim at estimating :

- $h_{SNP}^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_D^2 + \sigma_e^2}$
- $\delta_{SNP}^2 = \frac{\sigma_D^2}{\sigma_A^2 + \sigma_D^2 + \sigma_e^2}$

**Haseman-Elston regression (OLS) and
REML analysis using GCTA software**

```
gcta --mgrm multi-grm.txt --reml --out test
```

Assume a sample of unrelated individuals in an outbred population:

$$G_{ij} \sim \mathcal{N}(0, \text{Var}(G_{ij}))$$

Visscher et al. (2014) showed that under OLS:

$$\text{Var}(\hat{h}_{SNP}^2 | OLS) \simeq \frac{2}{N^2 \text{Var}(G_{ij})}$$

Visscher & Goddard (2015) showed that for REML estimation:

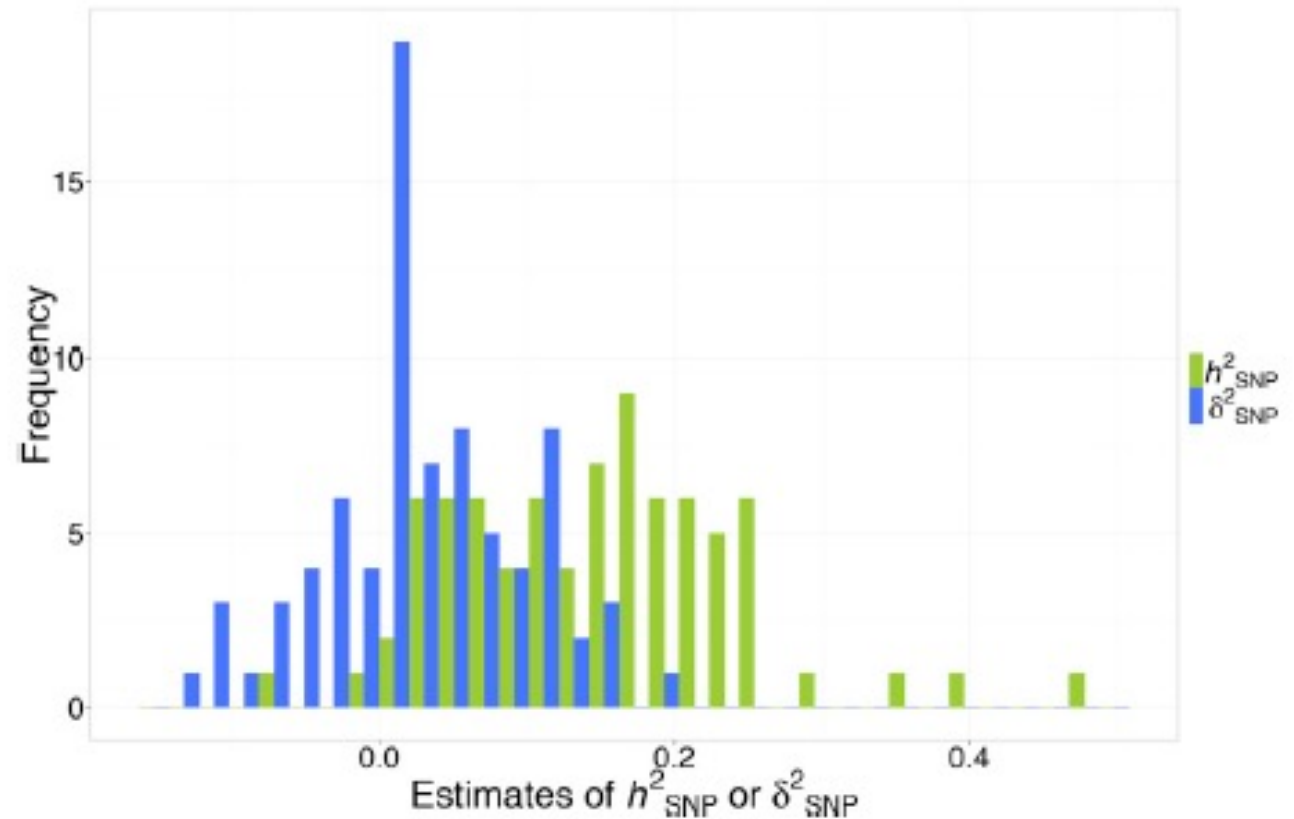
$$\text{Var}(\hat{h}_{SNP}^2 | REML) \simeq \frac{2}{N \text{Var}(\lambda_G)} \simeq \frac{2}{N^2 \text{Var}(G_{ij})}$$

Applications

```
gcta --mgrm multi-grm.txt --reml --reml-no-constrain --out test
```

```
Summary result of REML analysis:
Source  Variance      SE
V(G1)   0.423959      0.086312
V(G2)   -0.062331     0.107661
V(e)    0.348104      0.134472
Vp      0.709732      0.018560
V(G1)/Vp 0.597351      0.118408
V(G2)/Vp -0.087823     0.151827
Sum of V(G)/Vp 0.509528     0.190842
```

$$Var(\hat{\delta}_{SNP}^2) \approx \frac{2}{N^2 Var(\Theta_{Dij})}$$

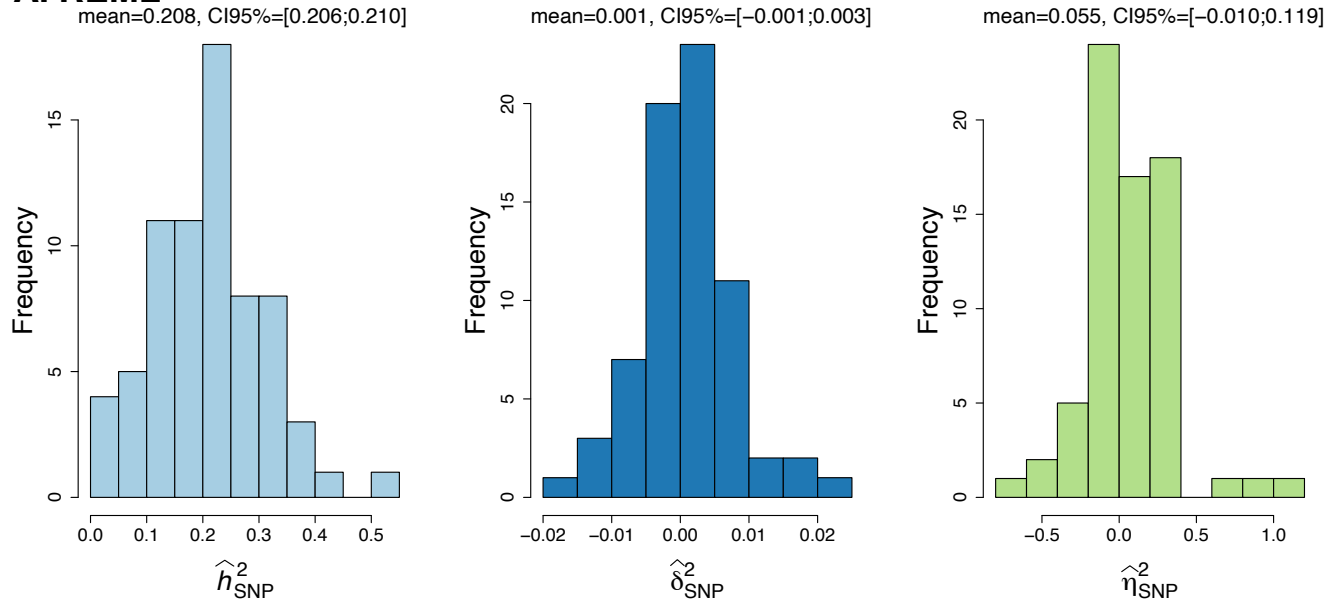


Zhu et al. (2015)

Applications

Humans

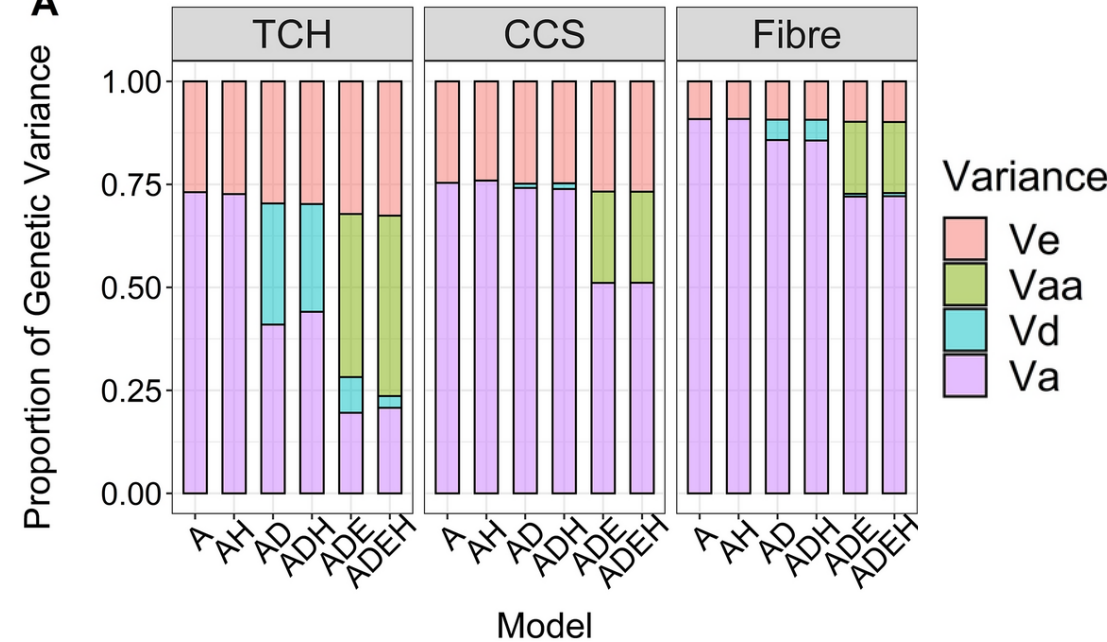
A. REML



Hivert et al. (2021)

Sugarcane

A



Yadav et al. (2021)

Part II: Genetic Correlation between two traits

- What is the proportion of variance that two traits share due to genetic causes?

Bivariate GREML analysis to estimate genetic correlation in unrelated individuals

```
gcta64 --reml-bivar --grm test --pheno test.phen --out test
```

$$\mathbf{y}_1 = \mathbf{X}_1\mathbf{b}_1 + \mathbf{g}_1 + \mathbf{e}_1$$

$$\mathbf{y}_2 = \mathbf{X}_2\mathbf{b}_2 + \mathbf{g}_2 + \mathbf{e}_2$$

$$\mathbf{V} = \text{var} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1\sigma_{g1}^2 + \mathbf{I}\sigma_{e1}^2 & \mathbf{G}_{12}\sigma_{g1g2} + \mathbf{I}\sigma_{e1e2} \\ \mathbf{G}_{12}\sigma_{g1g2} + \mathbf{I}\sigma_{e1e2} & \mathbf{G}_2\sigma_{g2}^2 + \mathbf{I}\sigma_{e2}^2 \end{bmatrix}$$

For traits measures on different samples

$$\mathbf{V} = \text{var} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1\sigma_{g1}^2 + \mathbf{I}\sigma_{e1}^2 & \mathbf{G}_{12}\sigma_{g1g2} \\ \mathbf{G}_{12}\sigma_{g1g2} & \mathbf{G}_2\sigma_{g2}^2 + \mathbf{I}\sigma_{e2}^2 \end{bmatrix}$$

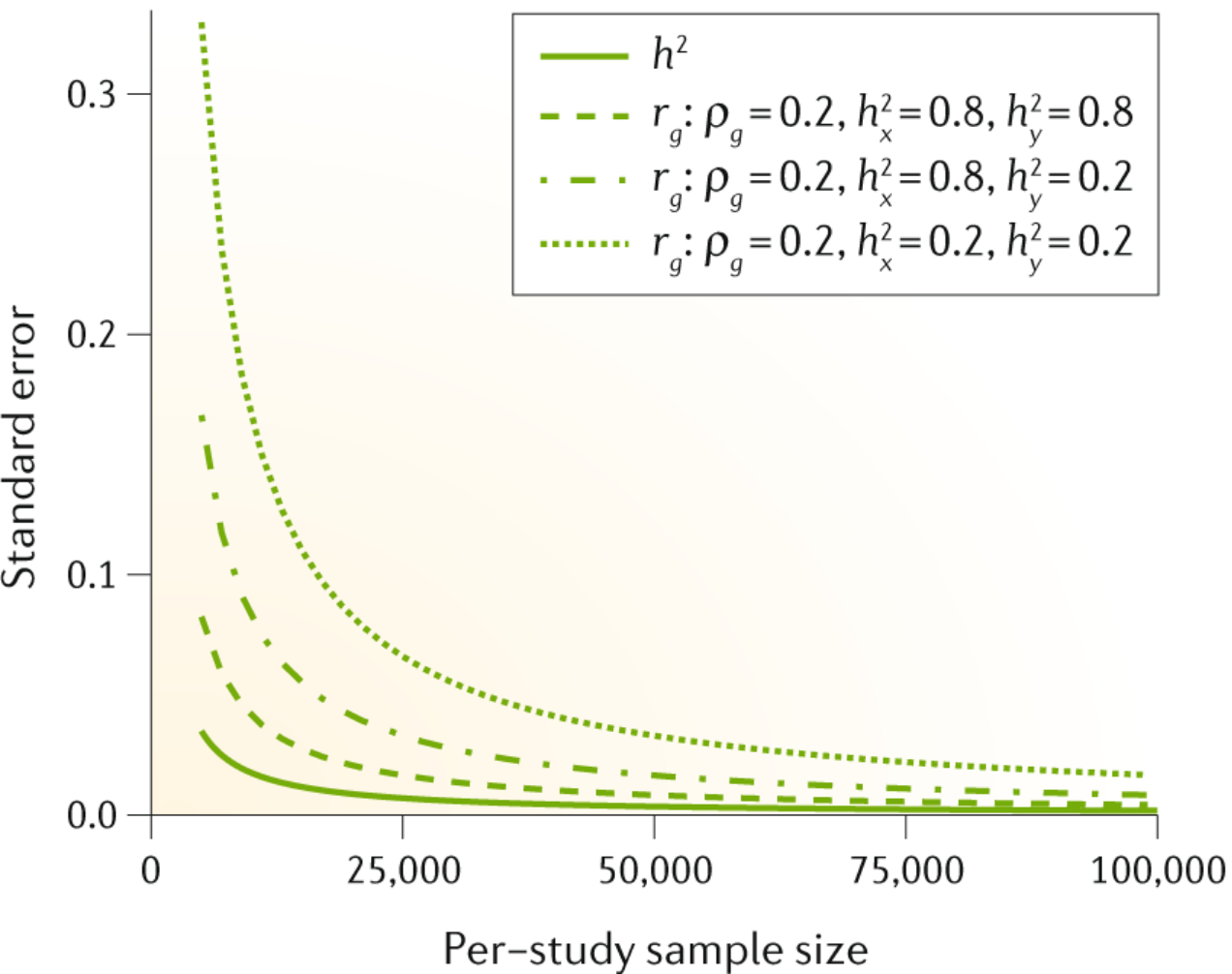
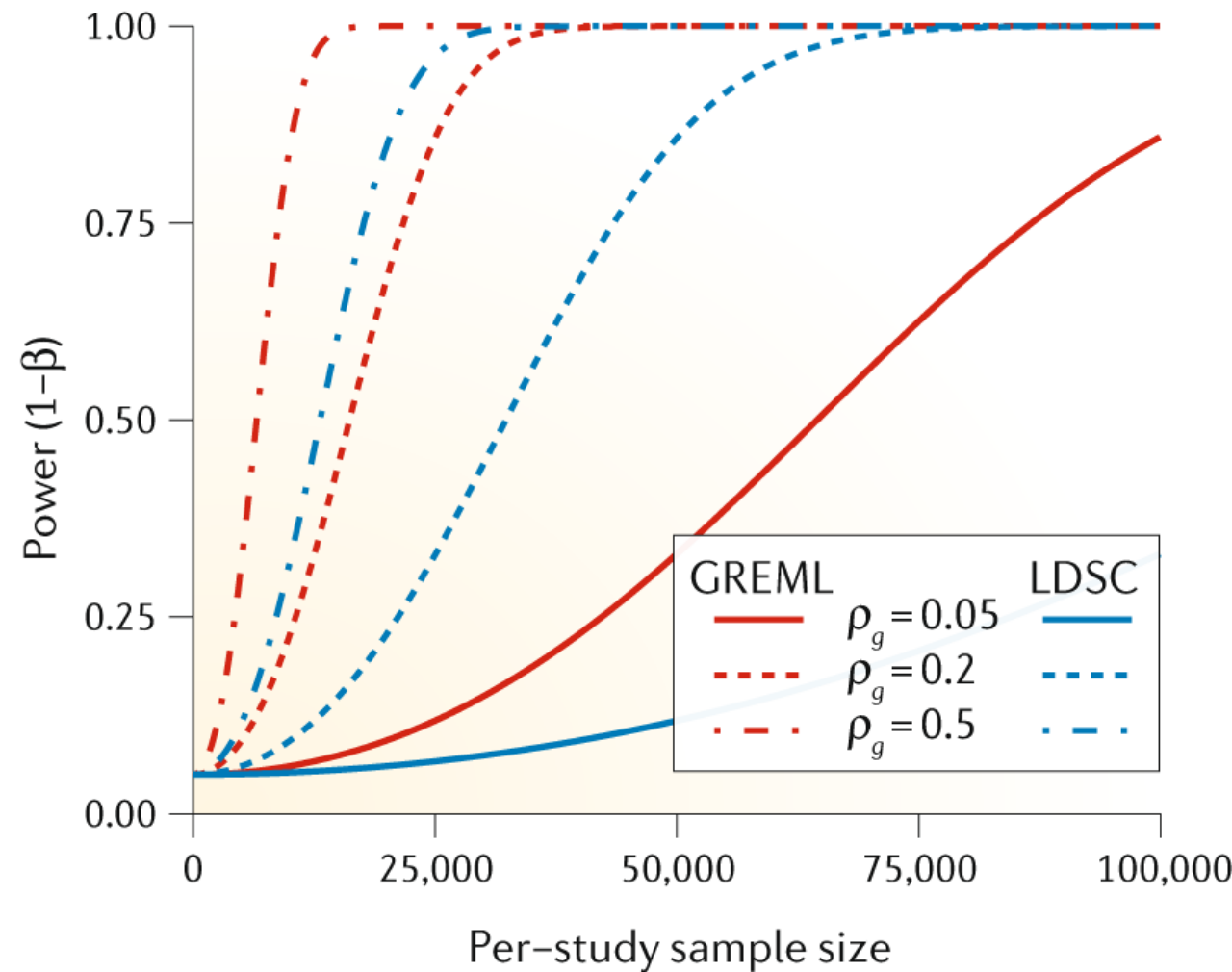
Bivariate GREML analysis to estimate genetic correlation in unrelated individuals

Sampling variance of \hat{r}_g for traits measured on the same sample:

$$\text{Var}(\hat{r}_g) \approx \frac{(1 - r_g r_p)^2 + (r_g - r_p)^2}{h_{G1}^2 h_{G2}^2 N^2 \text{Var}(G_{ij})}$$

Sampling variance of \hat{r}_g for traits measured on different samples:

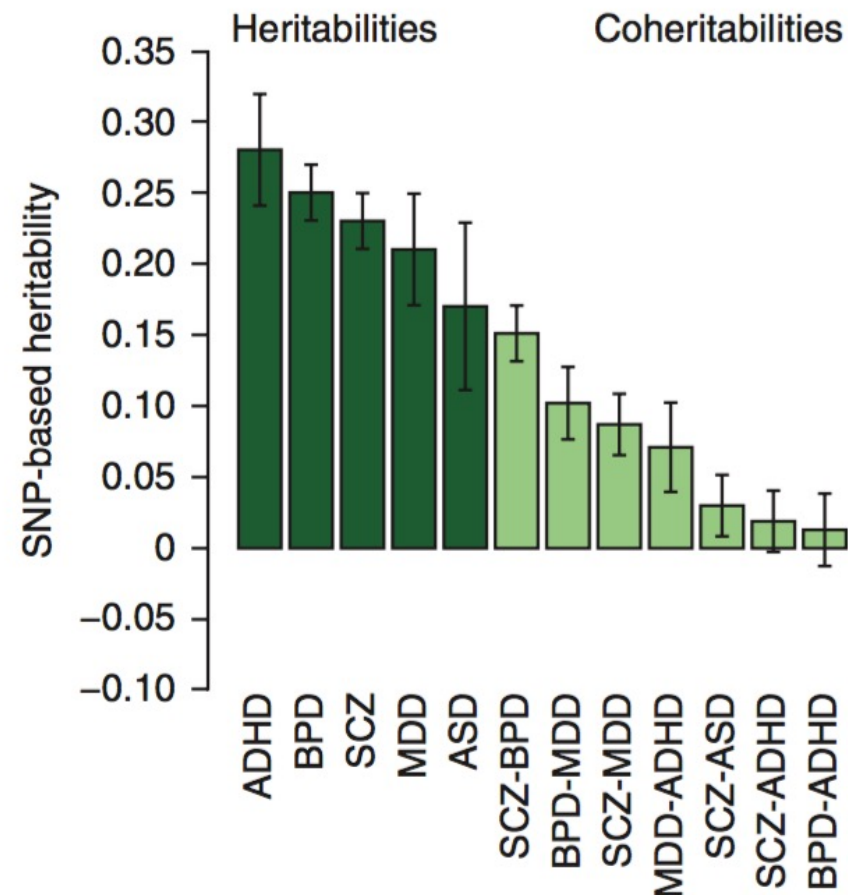
$$\text{Var}(\hat{r}_g) \approx \frac{r_g^2 (N_1^2 h_{G1}^4 + N_2^2 h_{G2}^4) + 2h_{G1}^2 h_{G2}^2 N_1 N_2}{2h_{G1}^4 h_{G2}^4 N_1^2 N_2^2 \text{Var}(G_{ij})}$$

a Precision of genetic correlation estimates

b Power to detect non-null genetic correlations


Estimating genetic correlation between traits measured on different samples

```
gcta64 --reml-bivar --grm test --pheno test.phen --out test
```

Source	Variance	SE
V(G)_tr1	8.704961	1.440557
V(G)_tr2	31.130274	5.619958
C(G)_tr12	0.050401	2.073481
V(e)_tr1	14.838636	1.404844
V(e)_tr2	61.047385	5.512983
C(e)_tr12	12.059304	2.036266
Vp_tr1	23.543596	0.433530
Vp_tr2	92.177659	1.695391
V(G)/Vp_tr1	0.369738	0.059975
V(G)/Vp_tr2	0.337720	0.059952
rG	0.003062	0.125778
logL	-28744.199	
n	11980	



Estimating genetic correlation between traits measured in different populations (e.g. ancestries, breeds)

Brown et al. (2016) introduced the genetic impact correlation, the correlation between the standardized effect sizes in two populations.

$$\rho_{gi} = \text{Cor}(\sigma_1\beta_1, \sigma_2\beta_2).$$

With $\sigma_i = \sqrt{2p_i(1 - p_i)}$ with p_i the allele frequencies in population i .

→ Implemented in the Python package Popcorn.



Practical 10: Use GCTA GREML to perform
1) the joint estimation of additive and dominance variance in a trait, and 2)
the estimation of genetic correlation between two traits