Proportional Variance Explained by QTL and Statistical Power

Partitioning the Genetic Variance

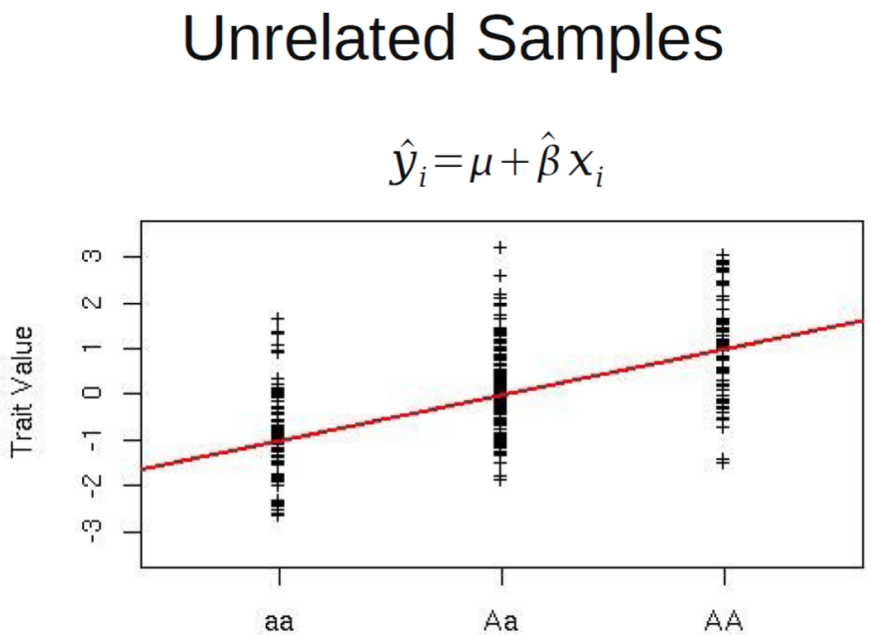
- We previously focused on obtaining variance components of a quantitative trait to determine the proportion of the variance of the trait that can be attributed to both genetic (additive and dominance) and environment (shared and unique) factors
- We demonstrated that resemblance of trait values among relatives we can be used to obtain estimates of the variance components of a quantitative trait without using genotype data.
- For quantitative traits, there generally is no (apparent) simple Mendelian basis for variation in the trait

Partitioning the Genetic Variance

- May be a single gene strongly influenced by environmental factors
- May be the result of a number of genes of equal (or differing) effect
- Most likely, a combination of both multiple genes and environmental factors
- Examples: Blood pressure, cholesterol levels, IQ, height, etc.

GWAS and Linear Regression

- Genome-wide association studies (GWAS) are commonly used for the identification of QTL
- Single SNP association testing with linear regression models are often used



Partition of Phenotypic Values

- Today we will focus on
  - Contribution of a QTL to the variance of a quantitative trait
  - Statistical power for detecting QTL in GWAS
- Consider once again the classical quantitative genetics model of $Y = G + E$ where Y is the phenotype value, G is the genotypic value, and E is the environmental deviation that is assumed to have a mean of 0 such that $E(Y) = E(G)$

Representation of Genotypic Values

- For a single locus with alleles A_1 and A_2, the genotypic values for the three genotypes can be represented as follows

$$\text{Genotype Value} = \begin{cases} -a & \text{if genotype is } A_2A_2 \\ d & \text{if genotype is } A_1A_2 \\ a & \text{if genotype is } A_1A_1 \end{cases}$$

- If p and q are the allele frequencies of the A_1 and A_2 alleles, respectively in the population, we previously showed that

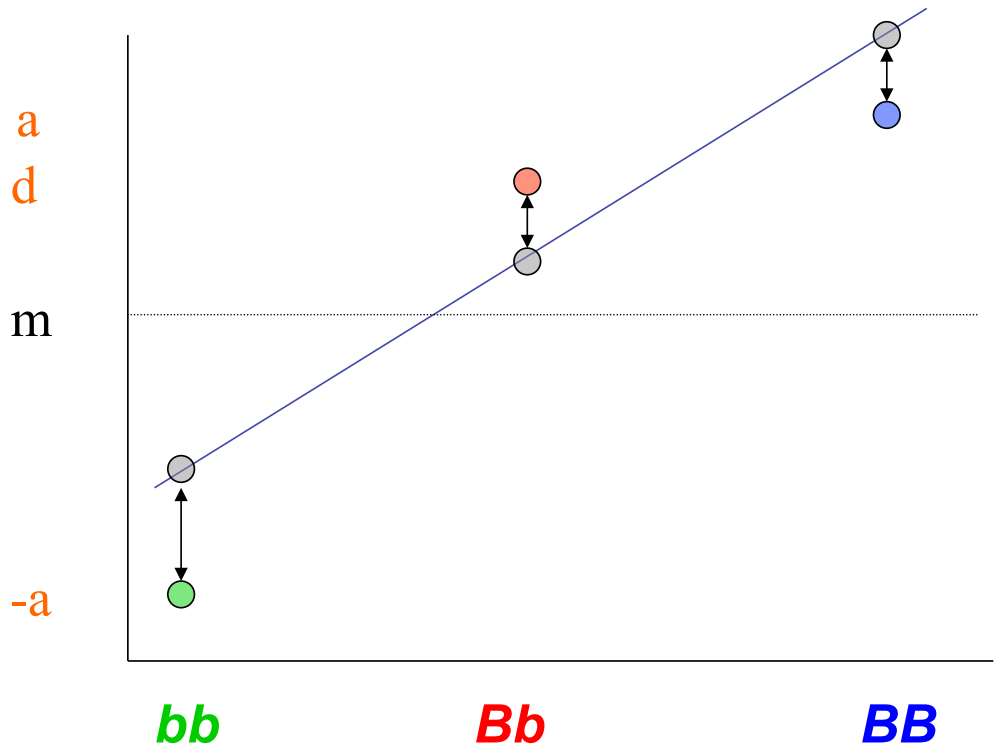
$$\mu_G = a(p-q) + d(2pq)$$

and that the genotypic value at a locus can be decomposed into additive effects and dominance deviations:

$$G_{ij} = G_{ij}^A + \delta_{ij} = \mu_G + \alpha_i + \alpha_j + \delta_{ij}$$

Linear Regression Figure for Genetic Values

Falconer model for single biallelic QTL



Var (X) = Regression Variance + Residual Variance
= Additive Variance + Dominance Variance

Decomposition of Genotypic Values

- The model can be given in terms of a linear regression of genotypic values on the number of copies of the A_1 allele such that:

$$G_{ij} = \beta_0 + \beta_1 X_1^{ij} + \delta_{ij}$$

  where X_1^{ij} is the number of copies of the type A_1 allele in genotype G_{ij}, and with $\beta_0 = \mu_G + 2\alpha_2$ and $\beta_1 = \alpha_1 - \alpha_2 = \alpha$, the average effect of allele substitution.
- Recall that $\alpha = a + d(q - p)$ and that $\alpha_1 = q\alpha$ and $\alpha_2 = -p\alpha$

Additive Genetic Model

- The following additive model is commonly used association studies with quantitative traits

$$Y = \beta_0 + \beta_1 X + \epsilon$$

where X is the number of copies of the reference allele (A_1) and individual has

- For this a single locus trait, how would you interpret ϵ for this particular model?

Statistical Power for Detectng QTL

$$Y = \beta_0 + \beta_1 X + \epsilon$$

- Assume, without loss of generality, that Y is a standardized trait with $\sigma_Y^2 = 1$
- Test statistics for $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$

$$T = \hat{\beta}_1 / \sigma(\hat{\beta}) \sim \mathbf{t}_{N-2} \approx N(0,1) \text{ for large } N$$

$$T^2 = \hat{\beta}_1^2 / var(\hat{\beta}) \sim \mathbf{F}_{1,N-2} \approx \chi_1^2 \text{ for large } N$$

- And we have that

$$var(\hat{\beta}) = \frac{\sigma_\epsilon^2}{S_{XX}} \approx \frac{\sigma_\epsilon^2}{N\sigma_X^2} = \frac{\sigma_\epsilon^2}{2Np(1-p)}$$

where S_{XX} is the corrected sum of squares for the X_i's

Statistical Power for Detecting QTL

- $Y = \beta_0 + \beta_1 X + \epsilon$, so we have that

$$\sigma_Y^2 = \beta_1^2 \sigma_X^2 + \sigma_\epsilon^2 = \beta_1^2 2p(1-p) + \sigma_\epsilon^2$$

- Recall that $\beta_1^2 = \alpha^2 = [a + d(q-p)]^2$, so

$$\sigma_Y^2 = 2p(1-p)[a + d(q-p)]^2 + \sigma_\epsilon^2 = h_s^2 + \sigma_\epsilon^2$$

  where $h_s^2 = 2p(1-p)[a + d(q-p)]^2$.
- Interpret h_s^2 (note that we assume that trait is standardized such that $\sigma_Y^2 = 1$)

Statistical Power for Detecting QTL

- Also note that $\sigma_\epsilon^2 = 1 - h_s^2$, so we can write $Var(\hat{\beta}_1)$ as the following:

$$var(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{2Np(1-p)} = \frac{1 - h_s^2}{2Np(1-p)}$$

- To calculate power of the test statistic T^2 for a given sample size N, we need to first obtain the expected value of the non-centrality parameter λ of the chi-squared distribution which is the expected value of the test statistic T squared:

$$\lambda = [E(T)]^2 \approx \frac{\beta_1^2}{var(\hat{\beta}_1)} = \frac{2Np(1-p)[a + d(q-p)]^2}{1 - h_s^2} = \frac{Nh_s^2}{1 - h_s^2}$$

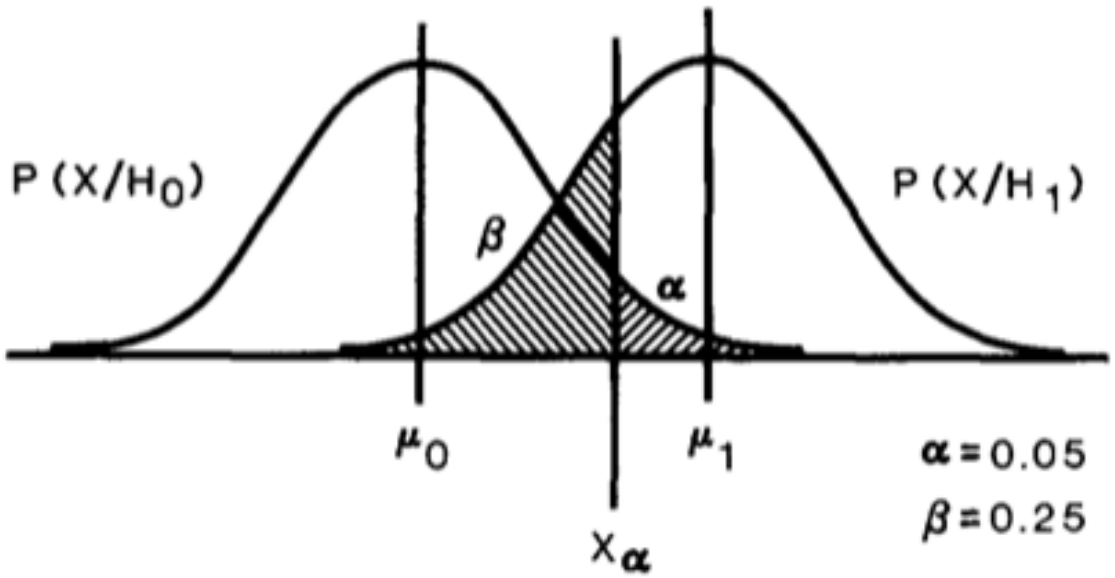
Power

- Can also obtain the required sample size given type-I error α and power $1-\beta$, where the type–II error is β :

$$N = \frac{1-h_s^2}{h_s^2}\left(z_{(1-\alpha/2)} + z_{(1-\beta)}\right)^2$$

where $z_{(1-\alpha/2)}$ and $z_{(1-\beta)}$ are the $(1-\alpha/2)$th and $(1-\beta)$th quantiles, respectively, for the standard normal distribution.

Statistical Power for Detecting QTL



Genetic Power Calculator (PGC)
http://pngu.mgh.harvard.edu/~purcell/gpc/

Genetic Power Calculator

Genetic Power Calculator

S. Purcell & P. Sham, 2001-2009

This site provides automated power analysis for variance components (VC) quantitative trait locus (QTL) linkage and association tests in sibships, and other common tests. Suggestions, comments, etc to *Shaun Purcell*.

If you use this site, please reference the following Bioinformatics article:

```
Purcell S, Cherny SS, Sham PC. (2003) Genetic Power Calculator:
design of linkage and association genetic mapping studies of complex
traits. Bioinformatics, 19(1):149-150.
```

Modules

Case-control for discrete traits	Notes
Case-control for threshold-selected quantitative traits	Notes
QTL association for sibships and singletons	Notes
TDT for discrete traits	Notes
TDT and parenTDT with ascertainment	Notes
TDT for threshold-selected quantitative traits	Notes
Epistasis power calculator	Notes
QTL linkage for sibships	Notes
Probability Function Calculator	Notes

Genetic Power Calculator

QTL Association for Sibships

```
Total QTL variance                 :          (0 - 1)
Dominance : additive QTL effects   :          (0 - 1)  No dominance (* see below)
QTL increaser allele frequency     :          (0 - 1)
Marker M1 allele frequency         :          (0 - 1)
Linkage disequilibrium (D-prime)   :          (0 - 1)
Sibling correlation                :          (0 - 1)  (* see below)

Sample Size                        :          (0 - 10000000)  (N=families, not individuals)
Sibship Size                       :  Pairs   Both parents genotyped

User-defined type I error rate     :  0.05    (0.00000001 - 0.5)
User-defined power: determine N    :  0.80    (0 - 1)
(1 - type II error rate)
```

Process Reset

Genetic Power Calculator (Shaun Purcell)

http://pngu.mgh.harvard.edu/~purcell/gpc/

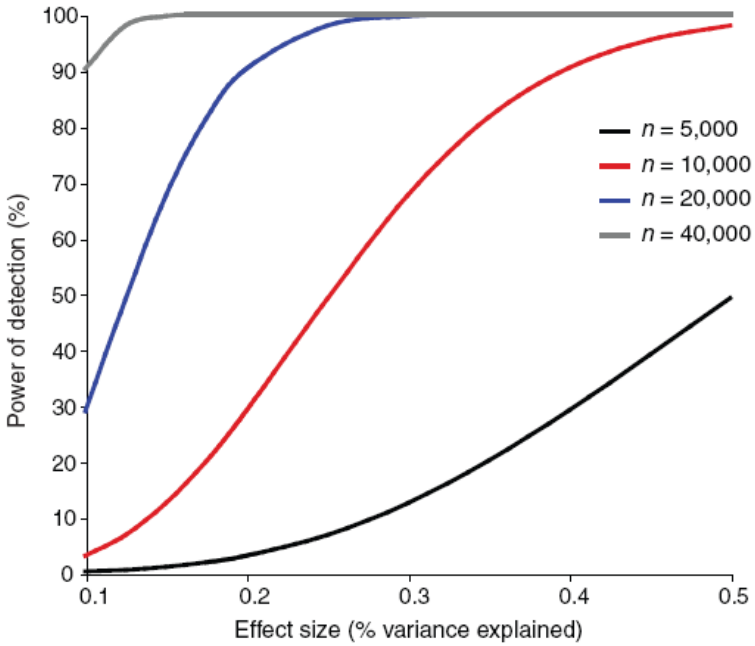
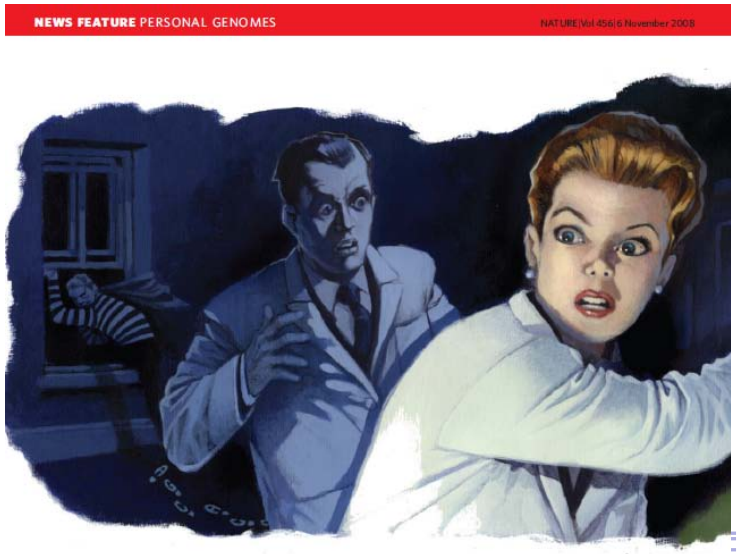


Figure 1 Statistical power of detection in GWAS for variants that explain 0.1–0.5% of the variation at a type I error rate of 5×10^{-7} (calculated using the Genetic Power Calculator[15]). Shown is the power to detect a variant with a given effect size, assuming this type I error rate, which is typical for a GWAS with a sample size of $n = 5{,}000$–$40{,}000$.

Missing Heritability

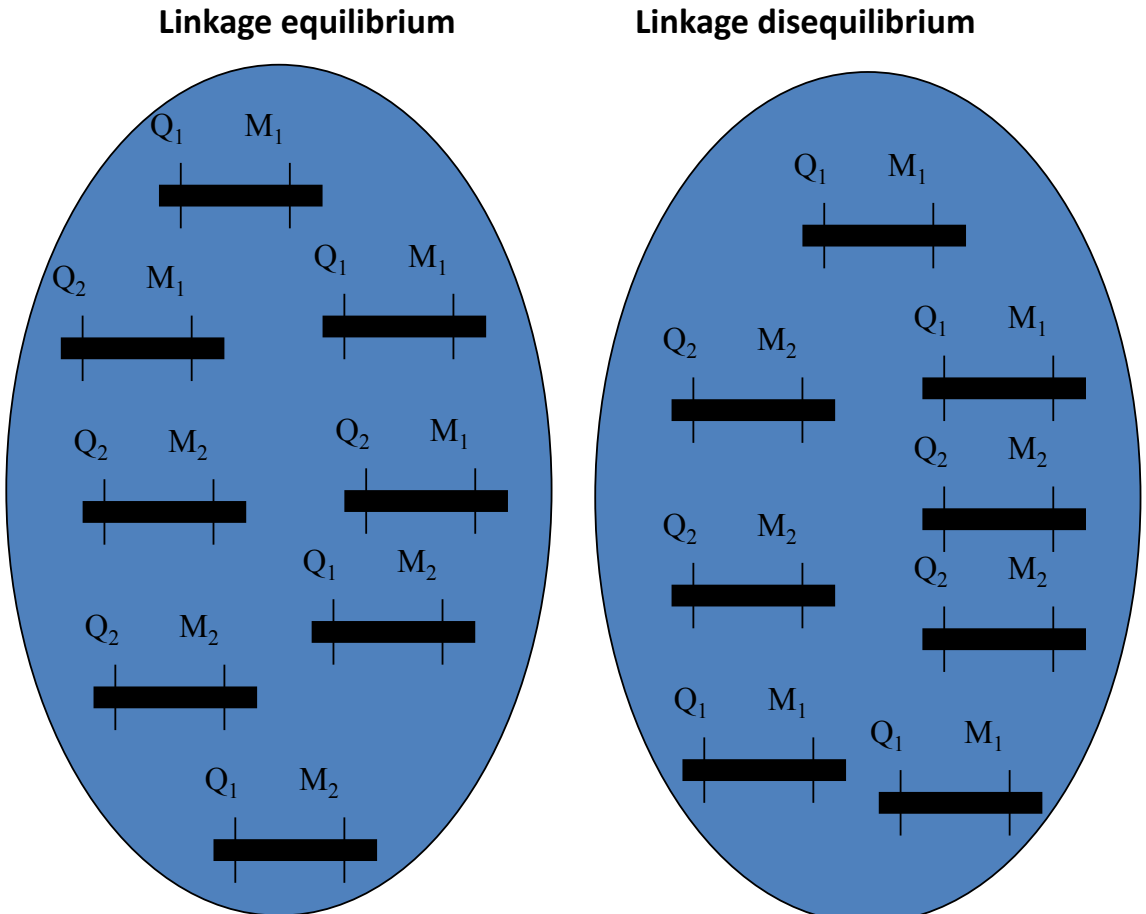
Disease	Number of loci	Percent of Heritability Measure Explained	Heritability Measure
Age-related macular degeneration	5	50%	Sibling recurrence risk
Crohn's disease	32	20%	Genetic risk (liability)
Systemic lupus erythematosus	6	15%	Sibling recurrence risk
Type 2 diabetes	18	6%	Sibling recurrence risk
HDL cholesterol	7	5.2%	Phenotypic variance
Height	40	5%	Phenotypic variance
Early onset myocardial infarction	9	2.8%	Phenotypic variance
Fasting glucose	4	1.5%	Phenotypic variance

- GWAS works
- Effect sizes are typically small
  - Disease: OR ~1.1 to ~1.3
  - Quantitative traits: % var explained <<1%



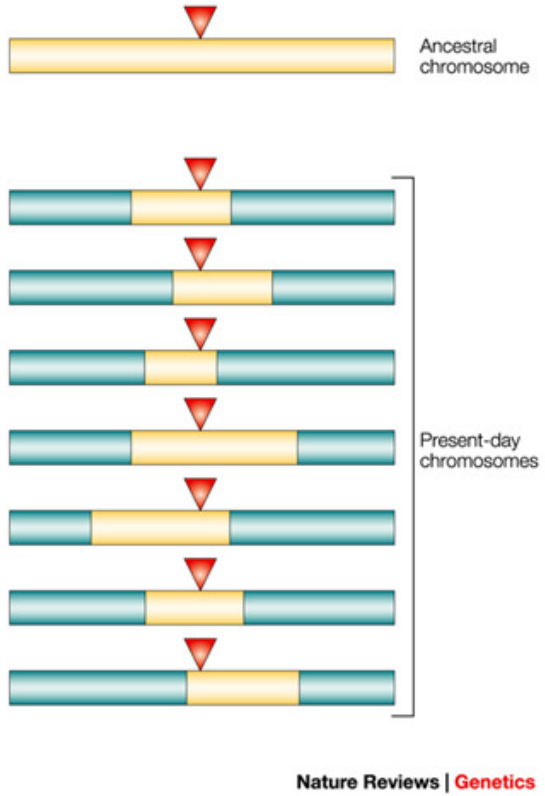
LD Mapping of QTL

- For GWAS, the QTL will generally not be genotyped in a study



LD

Linkage disequilibrium around an ancestral mutation



[Ardlie et al. 2002]

LD Mapping of QTL

- r^2 = LD correlation between QTL and genotyped SNP
- Proportion of variance of the trait explained at a SNP $\approx r^2 h_s^2$
- Required sample size for detection is

$$N \approx \frac{1 - r^2 h_s^2}{r^2 h_s^2} \left(z_{(1-\alpha/2)} + z_{(1-\gamma)} \right)^2$$

- Power of LD mapping depends on the experimental sample size, variance explained by the causal variant and LD with a genotyped SNP