Module 14: Advanced Heritability Analysis

1. Estimating SNP Heritability
2. Genome Partitioning
3. Intensity of Heritability
4. Testing Different Models
5. Using Other Datatypes
6. Bivariate Analysis
7. Mixed Model Analysis
8. Gene-Based Association Testing
9. Adaptive MultiBLUP
10. Applying to Animal and Plant Data

1. Estimating SNP Heritability
2. Genome Partitioning
3. Intensity of Heritability
4. Testing Different Models
5. Using Other Datatypes
6. Bivariate Analysis
7. Mixed Model Analysis
8. Gene-Based Association Testing
9. Adaptive MultiBLUP
10. Applying to Animal and Plant Data

Recap of Module 9

We looked at estimating heritability from SNP data. To do this in a mixed model framework, it is necessary to construct a kinship matrix **K**, which estimates relatedness / genetic similarity between all pairs of individuals

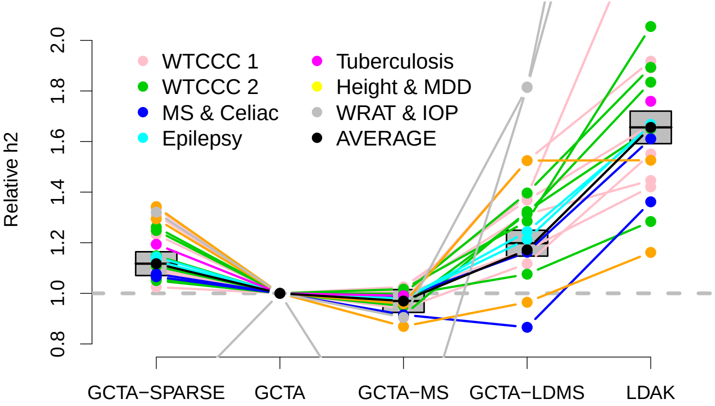
Most commonly, **K** takes the form of allelic correlations $\mathbf{K} = \mathbf{X}\mathbf{X}^T$, in which case, the mixed model is equivalent to a random effect regression model

When individuals are related, we will obtain an estimate of narrow-sense heritability, h^2

When individuals are unrelated, (and there is no inflation due to population structure or genotyping error) we will obtain an estimate of SNP heritability, h^2_{SNP}

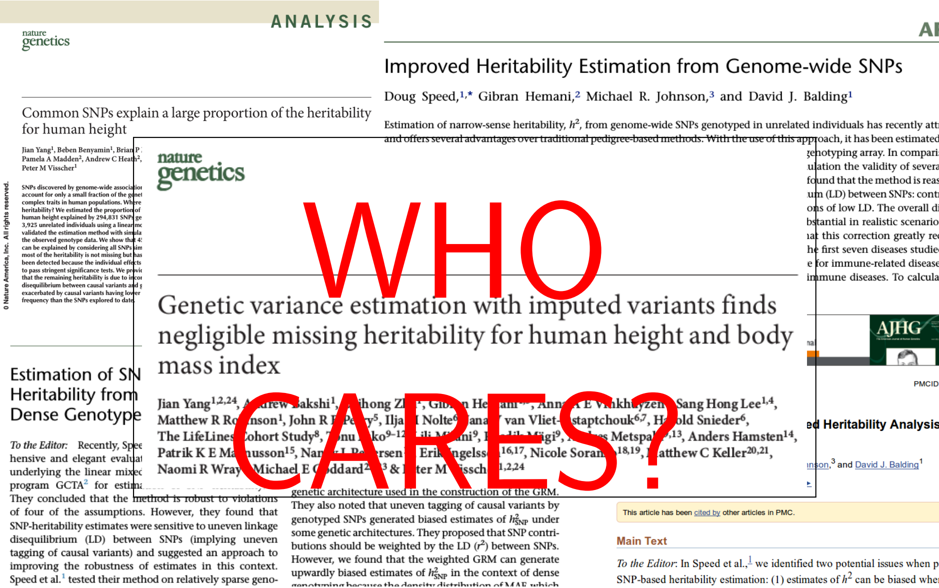
Recap of Module 9

The standard **K** used when estimating SNP heritability, corresponds to a (very) specific set of model assumptions



Changing **K** changes estimates of h^2_{SNP}; sometimes by a large amount

Different Methods Give Different Estimates



ANALYSIS

nature genetics

Common SNPs explain a large proportion of the heritability for human height

Jian Yang[1], Beben Benyamin[1], Brian P
Pamela A Madden[2], Andrew C Heath[2],
Peter M Visscher[1]

SNPs discovered by genome-wide associatio
account for only a small fraction of the genet
complex traits in human populations. Where
heritability? We estimated the proportion of
human height explained by 294,831 SNPs ge
3,925 unrelated individuals using a linear mo
validated the estimation method with simula
the observed genotype data. We show that 45
can be explained by considering all SNPs sim
most of the heritability is not missing but has
been detected because the individual effects
to pass stringent significance tests. We provi
that the remaining heritability is due to incor
disequilibrium between causal variants and g
exacerbated by causal variants having lower
frequency than the SNPs explored to date.

© Nature America, Inc. All rights reserved.

Improved Heritability Estimation from Genome-wide SNPs

Doug Speed,[1,*] Gibran Hemani,[2] Michael R. Johnson,[3] and David J. Balding[1]

Estimation of narrow-sense heritability, h^2, from genome-wide SNPs genotyped in unrelated individuals has recently att
and offers several advantages over traditional pedigree-based methods. With the use of this approach, it has been estimated
genotyping array. In compari
ulation the validity of severa
found that the method is reas
um (LD) between SNPs: cont
ons of low LD. The overall di
bstantial in realistic scenario
at this correction greatly re
he first seven diseases studie
e for immune-related disease
immune diseases. To calcula

nature genetics

WHO

Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index

Jian Yang[1,2,24], Andrew Bakshi[1], Zhihong Zhu[1], Gibran Hemani[2,3], Anna A E Vinkhuyzen[1], Sang Hong Lee[1,4],
Matthew R Robinson[1], John R B Perry[5], Ilja M Nolte[6], Jana V van Vliet-Ostaptchouk[6,7], Harold Snieder[6],
The LifeLines Cohort Study[8], Tonu Esko[9–12], Lili Milani[9], Reedik Mägi[9], Andres Metspalu[9,13], Anders Hamsten[14],
Patrik K E Magnusson[15], Nancy L Pedersen[15], Erik Ingelsson[16,17], Nicole Soranzo[18,19], Matthew C Keller[20,21],
Naomi R Wray[1], Michael E Goddard[22,23] & Peter M Visscher[1,2,24]

CARES?

AJHG

PMCID

ed Heritability Analysis

nson,[3] and David J. Balding[1]

This article has been cited by other articles in PMC.

Main Text

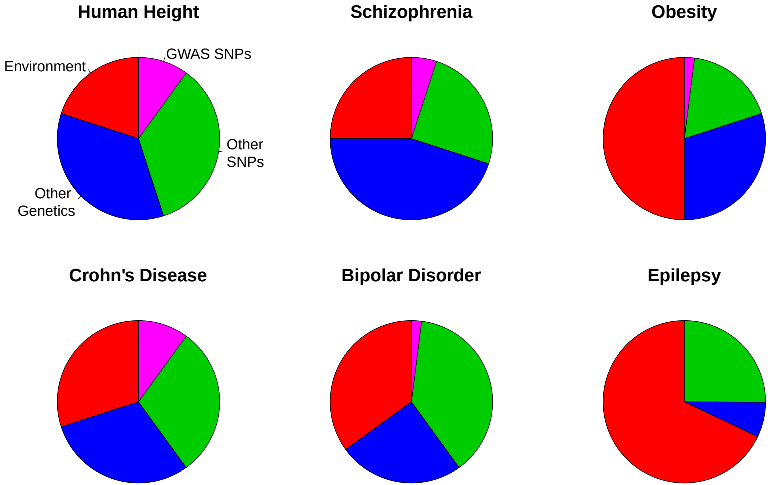
To the Editor: In Speed et al.,[1] we identified two potential issues when p
SNP-based heritability estimation: (1) estimates of h^2 can be biased whe

Estimation of SNP Heritability from Dense Genotype

To the Editor: Recently, Speed et al. published a comprehensive and elegant evaluation of the assumptions underlying the linear mixed model implemented in the program GCTA[2] for estimating SNP heritability. They concluded that the method is robust to violations of four of the assumptions. However, they found that SNP-heritability estimates were sensitive to uneven linkage disequilibrium (LD) between SNPs (implying uneven tagging of causal variants) and suggested an approach to improving the robustness of estimates in this context. Speed et al.[1] tested their method on relatively sparse geno-

genetic architecture used in the construction of the GRM. They also noted that uneven tagging of causal variants by genotyped SNPs generated biased estimates of h^2_{SNP} under some genetic architectures. They proposed that SNP contributions should be weighted by the LD (r^2) between SNPs. However, we found that the weighted GRM can generate upwardly biased estimates of h^2_{SNP} in the context of dense genotyping because the density distribution of MAF which

The Missing Heritability Problem is Solved



Regardless of the exact numbers, SNPs explain considerable heritability

Purpose of a GWAS

The three main objectives are:

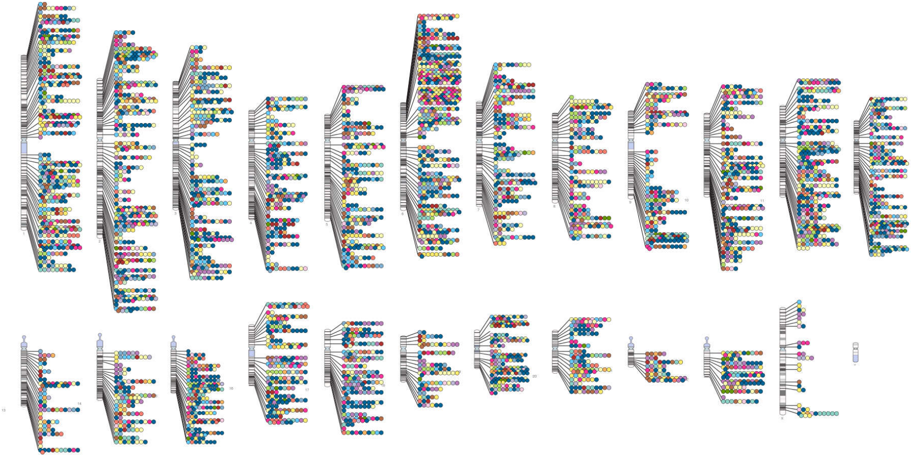
1 - Identify the causal variants

2 - Create prediction models

3 - Understand the genetic architecture of the trait

The most popular analysis is single-SNP testing

1 - Identify the Causal Variants



2 - Creating Prediction Models

Successful for traits with very large causal variants (e.g., BRCA1 for breast cancer, HLA for celiac)

But limited success for most common traits

3 - Understanding Genetic Architecture

Would like to know how much causal variation is due to:

- common variation
- rare variation
- additivity
- dominance / recessiveness / epistasis
- coding regions
- copy number changes, frameshifts, splicings, methlation, etc.

Single-SNP GWAS analysis mainly tells us that there are few common variants of strong effect

SNP-Based Heritability Analysis to the Rescue

1 Estimating SNP Heritability
2 Genome Partitioning
3 Intensity of Heritability
4 Testing Different Models
5 Using Other Datatypes
6 Bivariate Analysis
7 Mixed Model Analysis
8 Gene-Based Association Testing
9 Adaptive MultiBLUP
10 Applying to Animal and Plant Data

Genome Partitioning

The basic (single-kinship) model assumes

$$\begin{aligned}Y = {} & \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 \\ & + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} \\ & + \beta_{15} X_{15} + \beta_{16} X_{16} + \beta_{17} X_{17} + \beta_{18} X_{18} + \beta_{19} X_{19} + \beta_{20} X_{20} + \beta_{21} X_{21} \\ & + \beta_{22} X_{22} + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{25} X_{25} + \beta_{26} X_{26} + \beta_{27} X_{27} + \beta_{28} X_{28} \\ & + \ldots + \beta_{500\,000} X_{500\,000} \\ & + e.\end{aligned}$$

Assume $\beta_j \sim \mathbb{N}(0, \sigma_g^2/N)$ and $e \sim \mathbb{N}(0, \sigma_e^2)$.

Then $Y \sim \mathbb{N}(0, K\sigma_g^2 + I\sigma_e^2)$, where $K = \frac{XX^T}{N}$

Genome Partitioning

We can extend this to

$$\begin{aligned} Y = {} & \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7 \\ & + \beta_8X_8 + \beta_9X_9 + \beta_{10}X_{10} + \beta_{11}X_{11} + \beta_{12}X_{12} + \beta_{13}X_{13} + \beta_{14}X_{14} \\ & + \beta_{15}X_{15} + \beta_{16}X_{16} + \beta_{17}X_{17} + \beta_{18}X_{18} + \beta_{19}X_{19} + \beta_{20}X_{20} + \beta_{21}X_{21} \\ & + \beta_{22}X_{22} + \beta_{23}X_{23} + \beta_{24}X_{24} + \beta_{25}X_{25} + \beta_{26}X_{26} + \beta_{27}X_{27} + \beta_{28}X_{28} \\ & + \ldots + \beta_{500\,000}X_{500\,000} \\ & + e. \end{aligned}$$

Assume $\beta_j \sim \mathbb{N}(0, \sigma^2_{g1}/N_1)$ and $\beta_k \sim \mathbb{N}(0, \sigma^2_{g2}/N_2)$.

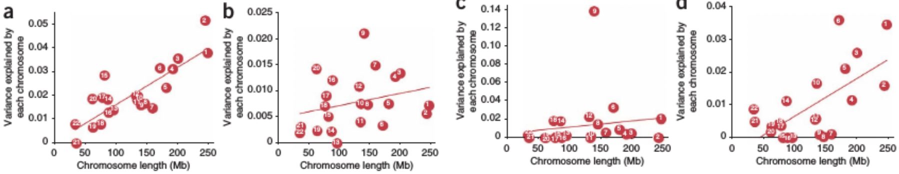
Then $Y \sim \mathbb{N}(0, K_1\sigma^2_{g1} + K_2\sigma^2_{g2} + I\sigma^2_e)$, where $K_1 = \frac{X_1X_1^T}{N_1}$ and $K_2 = \frac{X_2X_2^T}{N_2}$

Instead of estimating only total h^2_{SNP}, we can now estimate h^2 of RED SNPs and h^2 of BLUE SNPs

Genome Partitioning

Genome partitioning of genetic variation for complex traits using common SNPs

Jian Yang[1*], Teri A Manolio[2], Louis R Pasquale[3], Eric Boerwinkle[4], Neil Caporaso[5], Julie M Cunningham[6], Mariza de Andrade[7], Bjarke Feenstra[8], Eleanor Feingold[9], M Geoffrey Hayes[10], William G Hill[11], Maria Teresa Landi[12], Alvaro Alonso[13], Guillaume Lettre[14], Peng Lin[15], Hua Ling[16], William Lowe[17], Rasika A Mathias[18], Mads Melbye[8], Elizabeth Pugh[16], Marilyn C Cornelis[19], Bruce S Weir[20], Michael E Goddard[21,22] & Peter M Visscher[1]



Height BMI vWF QTi

1. Estimating SNP Heritability
2. Genome Partitioning
3. Intensity of Heritability
4. Testing Different Models
5. Using Other Datatypes
6. Bivariate Analysis
7. Mixed Model Analysis
8. Gene-Based Association Testing
9. Adaptive MultiBLUP
10. Applying to Animal and Plant Data

Intensity of Heritability

Suppose you have used genome partitioning to estimate contributions of genic and inter-genic regions

h^2_{SNP}=60%

h^2_{GENIC}=40%

$h^2_{\text{INTER-GENIC}}$=20%

Genes

Non-Genes

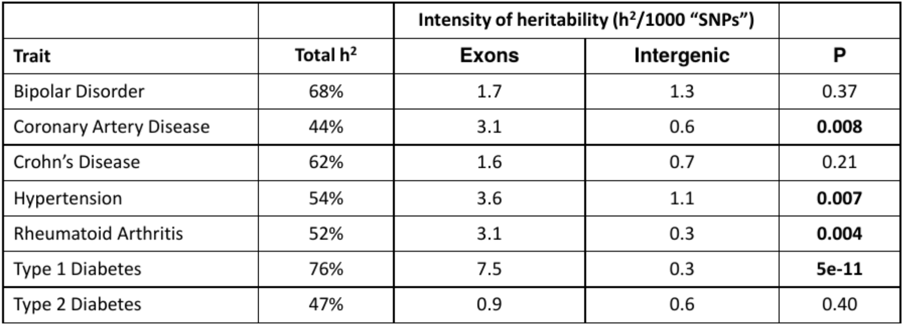
Want to decide whether a partitioning is significant

Define a region's intensity of heritability, I, as its h^2 divided by how much variation it captures (its sum of SNP weightings)

Can then test whether two (or more) partitions have significantly different intensity of heritability

The Role of Genes

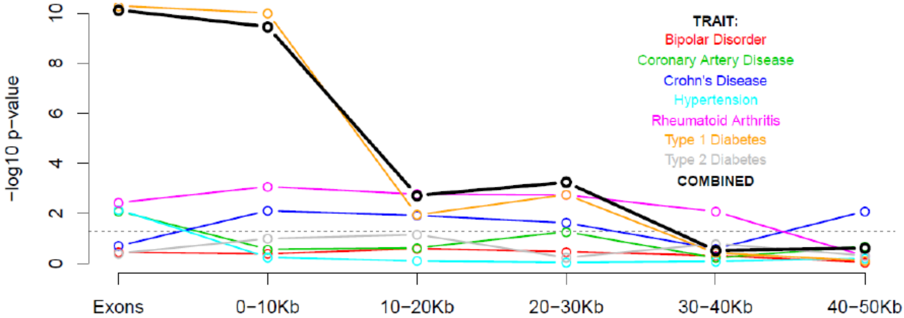
Is intensity of heritability higher for exonic than inter-genic SNPs?



		Intensity of heritability (h^2/1000 "SNPs")		
Trait	**Total h^2**	**Exons**	**Intergenic**	**P**
Bipolar Disorder	68%	1.7	1.3	0.37
Coronary Artery Disease	44%	3.1	0.6	**0.008**
Crohn's Disease	62%	1.6	0.7	0.21
Hypertension	54%	3.6	1.1	**0.007**
Rheumatoid Arthritis	52%	3.1	0.3	**0.004**
Type 1 Diabetes	76%	7.5	0.3	**5e-11**
Type 2 Diabetes	47%	0.9	0.6	0.40

Inter-genic defined as $>$100kb from a coding region.

The Role of Genes

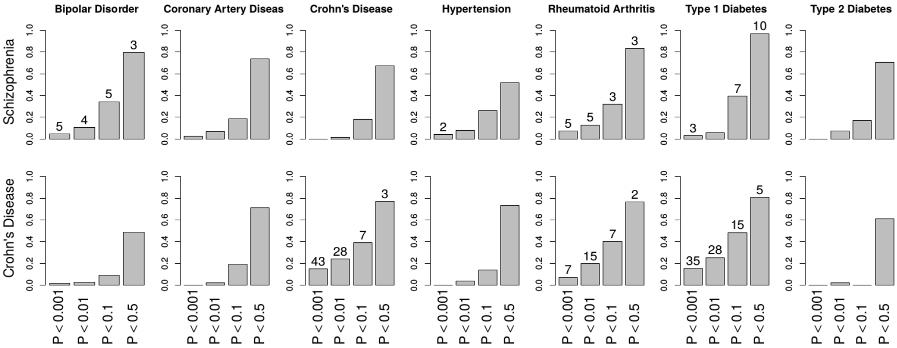


Can investigate what happens as we move away from exons.

Find intensity of heritability significantly high until >30kb.

Concordance Between Traits

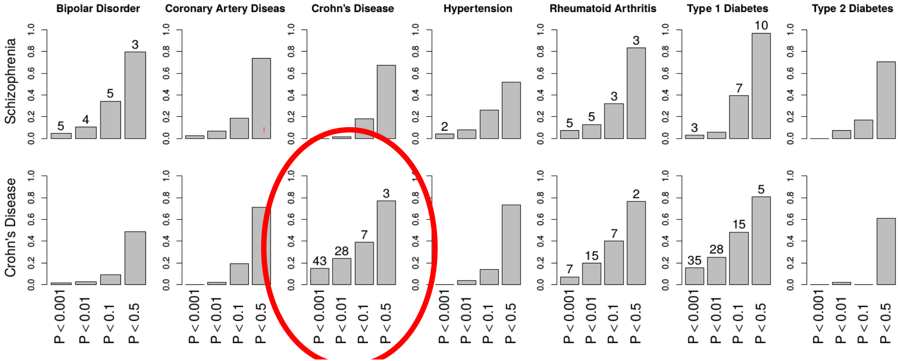
Are SNPs associated with one trait more important for others.



p-values for Schizophrenia and Crohn's obtained from independent studies.

Concordance Between Traits

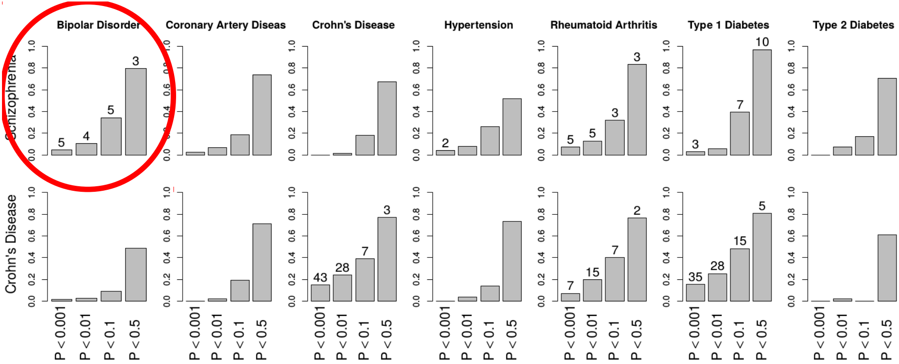
Are SNPs associated with one trait more important for others.



SNPs associated with Crohn's are more important for Crohn's. (Good!)

Concordance Between Traits

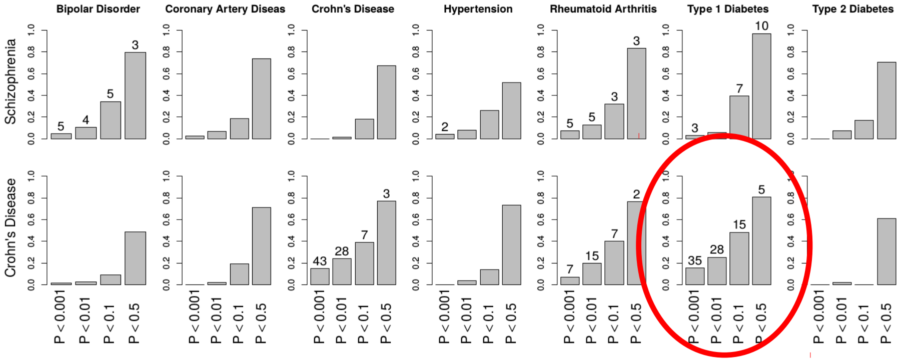
Are SNPs associated with one trait more important for others.



SNPs associated with Schizophrenia are more important for Bipolar.

Concordance Between Traits

Are SNPs associated with one trait more important for others.



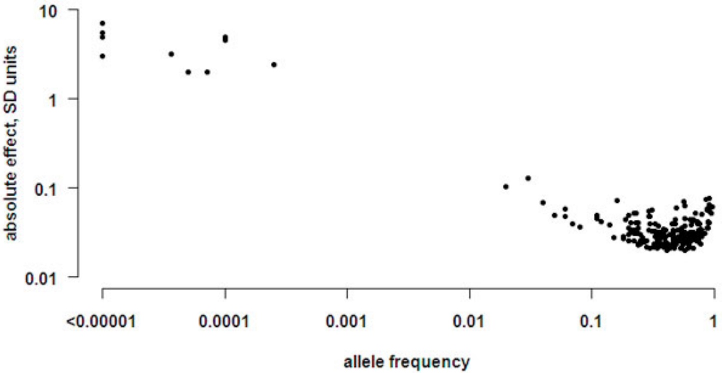
Find concordance between Crohn's and Type 1 Diabetes.

1. Estimating SNP Heritability
2. Genome Partitioning
3. Intensity of Heritability
4. **Testing Different Models**
5. Using Other Datatypes
6. Bivariate Analysis
7. Mixed Model Analysis
8. Gene-Based Association Testing
9. Adaptive MultiBLUP
10. Applying to Animal and Plant Data

Examining the Relationship between MAF and Effect Size

The default assumption is that all SNPs (GCTA) or genetic variations (LDAK) contribute equal heritability

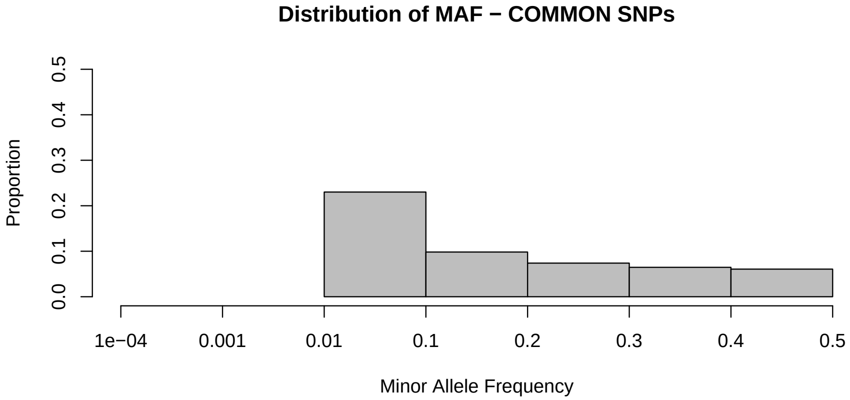
This seems to be supported by empirical evidence for some traits



Genetic architecture of body size in mammals, Kemper et al (2012)

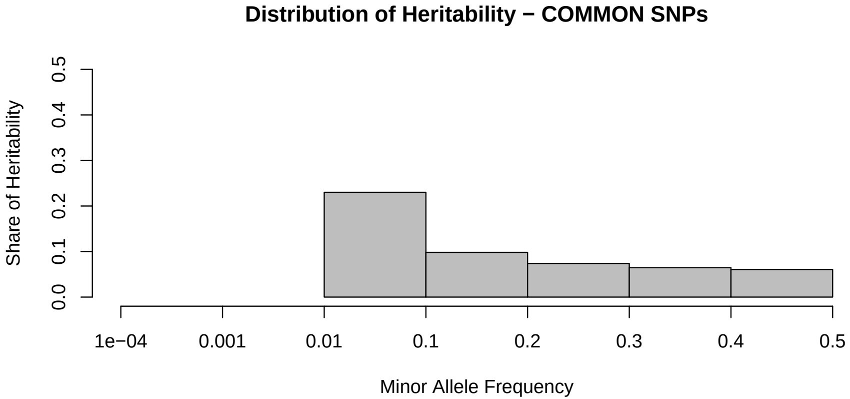
Distribution of Heritability across the MAF Spectrum

(If true), this gives us an idea of how heritability varies with MAF



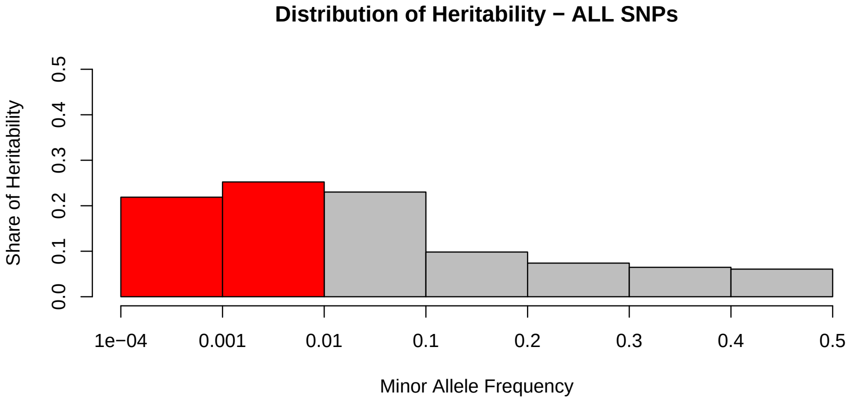
Distribution of Heritability across the MAF Spectrum

(If true), this gives us an idea of how heritability varies with MAF



Distribution of Heritability across the MAF Spectrum

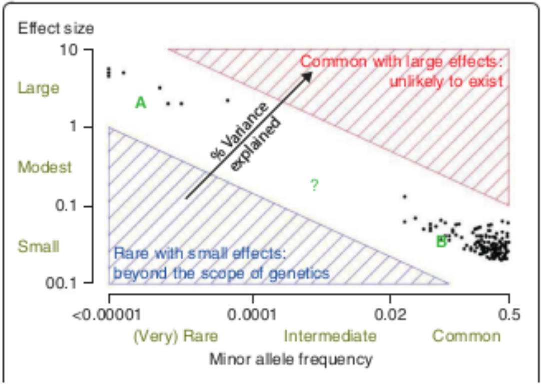
(If true), this gives us an idea of how heritability varies with MAF



In particular, it suggests the relative contribution of rare (red) and common (grey) variants

Is this Assumed Distribution Accurate?

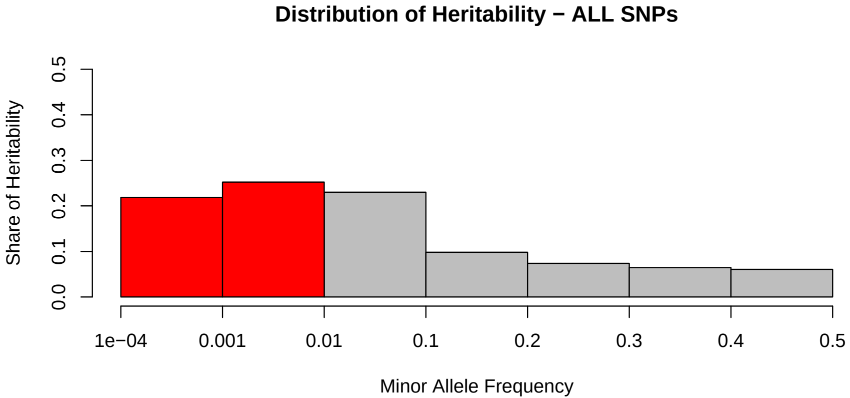
Must recognise there is an implicit bias, because we tend to only find the highest heritability variants



Understanding complex traits: from farmers to pharmas, Speed and Balding et al (2012)

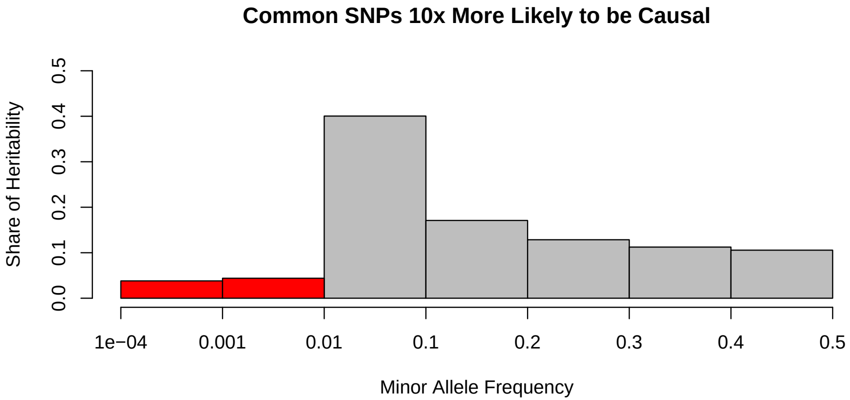
Distribution of Heritability across the MAF Spectrum

Also, this assumes rare SNPs are equally likely to be causal as common SNPs. Is this a fair assumption?



Distribution of Heritability across the MAF Spectrum

Things would change if the probability of a variant being causal depended on MAF



Testing How Heritability Varies with MAF

We can test this by performing heritability analysis with different kinship matrices, each corresponding to a different assumed relationship, and seeing which fare best

$$K_{ij} = \frac{1}{\sum_{w_j}} \sum w_j (X_{ij} - mean(X_j))(X_{ij} - mean(X_j)) \times [Var(X_j)]^{\alpha}$$

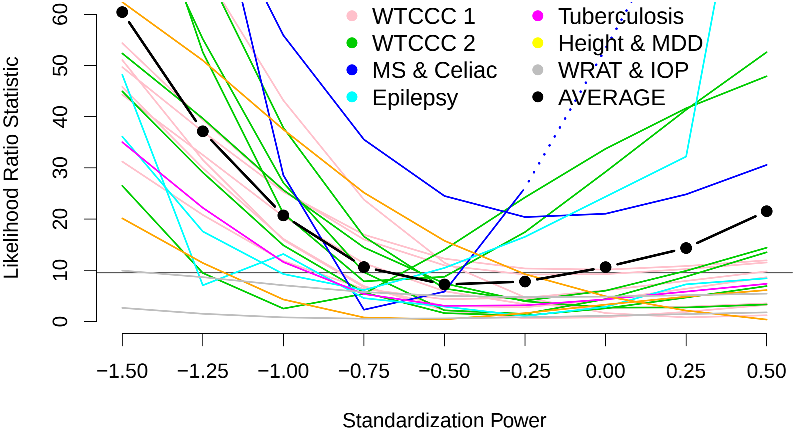
Smaller α means that rare variants contribute more h^2

The default in humans is $\alpha = -1$ (all variants contribute equal h^2)

The default in animals is $\alpha = 0$ (more common $\Rightarrow$ more h^2)

We can now analyse our data using multiple α and see which fits best

Results for 22 GWAS Traits



Smaller LRT (y-axis) $\Rightarrow$ more support for corresponding α

We deduce that $\alpha = -.5$ bests fit for these 22 traits

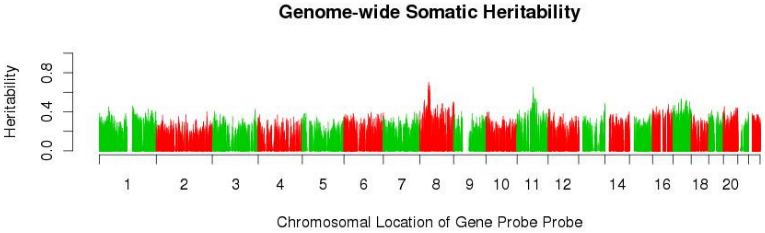
1 Estimating SNP Heritability
2 Genome Partitioning
3 Intensity of Heritability
4 Testing Different Models
5 Using Other Datatypes
6 Bivariate Analysis
7 Mixed Model Analysis
8 Gene-Based Association Testing
9 Adaptive MultiBLUP
10 Applying to Animal and Plant Data

Variance Explained by Copy Number Variants

We are not restricted to considering SNPs

the methodology can be applied to any data type

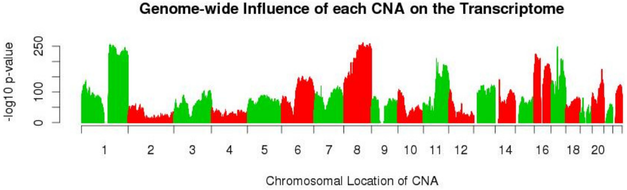
For breast cancer data, we constructed kinship matrices based on copy number variants (so now, each row of **X** represents the number of copies of a variant possessed by each individual)

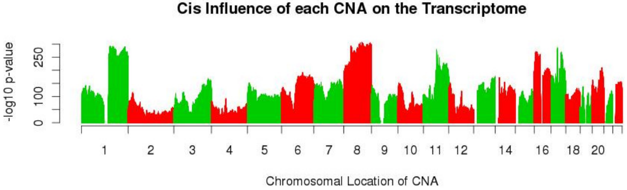


Instead of variance explained by SNPs, we estimate variance explained by copy number (where the phenotype is gene expression).

Variance Explained by Gene Expression

Alternatively, you can construct kinship matrices based on gene expressions (each row of **X** now represents intensity of a particular gene probe)


Genome-wide Influence of each CNA on the Transcriptome


Cis Influence of each CNA on the Transcriptome

We are now estimating variance explained by gene expression (where the phenotype is copy number change at a particular locus).

1 Estimating SNP Heritability

2 Genome Partitioning

3 Intensity of Heritability

4 Testing Different Models

5 Using Other Datatypes

6 Bivariate Analysis

7 Mixed Model Analysis

8 Gene-Based Association Testing

9 Adaptive MultiBLUP

10 Applying to Animal and Plant Data

Bivariate Analysis

$$\begin{aligned}\text{Trait 1: } Y_1 &= Z\alpha_1 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{500\,000} X_{500\,000} + e_1 \\ &= Z\alpha_1 + g_1 + e_1\end{aligned}$$

$$\begin{aligned}\text{Trait 2: } Y_2 &= Z\alpha_2 + \gamma_1 X_1 + \gamma_2 X_2 + \ldots + \gamma_{500\,000} X_{500\,000} + e_2 \\ &= Z\alpha_2 + g_2 + e_2\end{aligned}$$

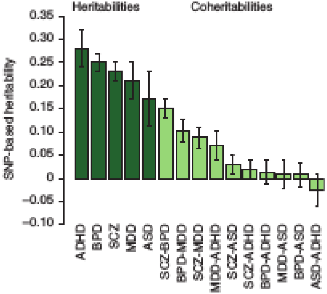
Now interested in the correlation between genetic effects:
$\rho = cor(g_1, g_2)$.

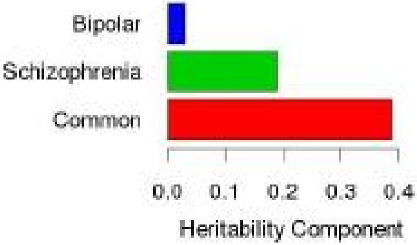
Or equivalently can think of the average correlation between effect sizes:
$\rho = cor(\beta_j, \gamma_j)$

Examining Concordance Between Traits

Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs

Cross-Disorder Group of the Psychiatric Genomics Consortium*





Hong will explain more in Module 17

1. Estimating SNP Heritability
2. Genome Partitioning
3. Intensity of Heritability
4. Testing Different Models
5. Using Other Datatypes
6. Bivariate Analysis
7. Mixed Model Analysis
8. Gene-Based Association Testing
9. Adaptive MultiBLUP
10. Applying to Animal and Plant Data

Mixed Model Association Analysis

Introduced in Module 5

Suppose we want to test SNP j for association with Y

Standard Single-SNP Analysis:

$$Y = \beta_j X_j + e$$

Mixed Model Association Analysis:

$$Y = \beta_j X_j + g + e \quad g \sim \mathbb{N}(0, K\sigma_g^2)$$

Including the background (polygenic) random effect g effectively allows for confounding due to familiar relatedness and population structure

1. Estimating SNP Heritability
2. Genome Partitioning
3. Intensity of Heritability
4. Testing Different Models
5. Using Other Datatypes
6. Bivariate Analysis
7. Mixed Model Analysis
8. **Gene-Based Association Testing**
9. Adaptive MultiBLUP
10. Applying to Animal and Plant Data

GBAT: Gene-Based Association Testing

Standard GWAS analysis examines each SNP individually.

There are many reasons we might prefer a gene-based analysis.

Reduces total number of tests
Biologically plausible
Can accumulate evidence across SNPs

GBAT: Gene-Based Association Testing

Standard GWAS analysis examines each SNP individually.

There are many reasons we might prefer a gene-based analysis.

- Reduces total number of tests
- Biologically plausible
- Can accumulate evidence across SNPs

Our software GBAT performs set-based tests of association.

- Fast and Powerful.
- Bayesian version accommodate prior information.
- Suitable for meta-analysis.

Gene-Based Association Testing

To estimate the total contribution of all SNPs, we use the model:

$$\begin{aligned} Y = {} & \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 \\ & + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} \\ & + \beta_{15} X_{15} + \beta_{16} X_{16} + \beta_{17} X_{17} + \beta_{18} X_{18} + \beta_{19} X_{19} + \beta_{20} X_{20} + \beta_{21} X_{21} \\ & + \beta_{22} X_{22} + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{25} X_{25} + \beta_{26} X_{26} + \beta_{27} X_{27} + \beta_{28} X_{28} \\ & + \ldots + \beta_{500\,000} X_{500\,000} \\ & + e. \end{aligned}$$

Gene-Based Association Testing

To test a set of SNPs S, can reduce it to:

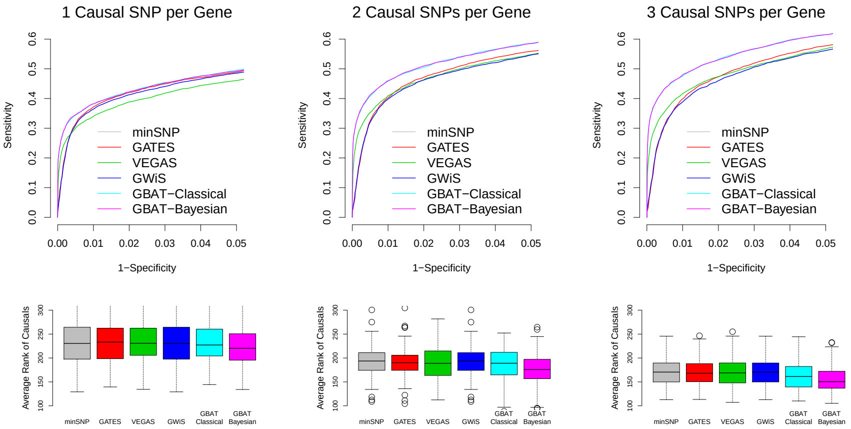
$$\begin{aligned} Y = {} & \cancel{\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7} \\ & + \cancel{\beta_8 X_8 + \beta_9 X_9 +} \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} \\ & + \beta_{15} X_{15} + \beta_{16} X_{16} + \beta_{17} X_{17} + \beta_{18} X_{18} \cancel{+ \beta_{19} X_{19} + \beta_{20} X_{20} + \beta_{21} X_{21}} \\ & + \cancel{\beta_{22} X_{22} + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{25} X_{25} + \beta_{26} X_{26} + \beta_{27} X_{27} + \beta_{28} X_{28}} \\ & + \cancel{\ldots + \beta_{500\,000} X_{500\,000}} \\ & + e. \end{aligned}$$

i.e., $Y = \sum_{j \in S} X_j \beta_j + e$ with $\beta_j \sim \mathbb{N}(0, \sigma_S^2 / N_S)$.

Perform a likelihood ratio test for $\mathbb{P}(\sigma_S^2 > 0)$.

Simulation Study

Generate phenotypes where 50 out of 1000 genes contribute heritability.


1 Causal SNP per Gene

2 Causal SNPs per Gene

3 Causal SNPs per Gene

GBAT most powerful and fastest (does not require permutations).

Accommodating Prior Information

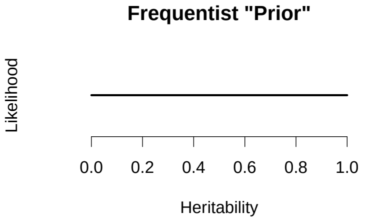
Frequentist methods base inferences only on evidence from data.

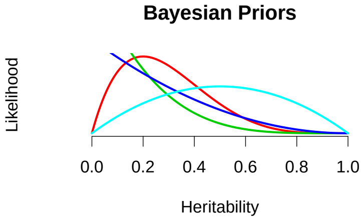
Bayesian methods combine evidence from data with prior beliefs.

Accommodating Prior Information

Frequentist methods base inferences only on evidence from data.

Bayesian methods combine evidence from data with prior beliefs.



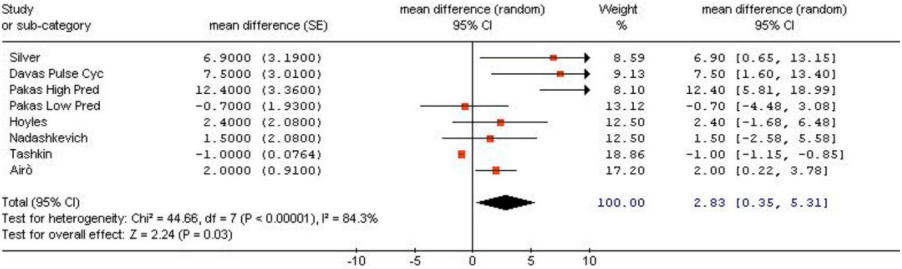


Frequentist analysis assumes a gene is as likely to explain 1% or 99% of heritability.

Bayesian GBAT lets the user specify a prior distribution for heritability.

Single-SNP Meta-Analysis

Meta-analysis allows us to combine evidence for SNP associations across cohorts.



Study or sub-category	mean difference (SE)	Weight %	mean difference (random) 95% CI
Silver	6.9000 (3.1900)	8.59	6.90 [0.65, 13.15]
Davas Pulse Cyc	7.5000 (3.0100)	9.13	7.50 [1.60, 13.40]
Pakas High Pred	12.4000 (3.3600)	8.10	12.40 [5.81, 18.99]
Pakas Low Pred	-0.7000 (1.9300)	13.12	-0.70 [-4.48, 3.08]
Hoyles	2.4000 (2.0800)	12.50	2.40 [-1.68, 6.48]
Nadashkevich	1.5000 (2.0800)	12.50	1.50 [-2.58, 5.58]
Tashkin	-1.0000 (0.0764)	18.86	-1.00 [-1.15, -0.85]
Airò	2.0000 (0.9100)	17.20	2.00 [0.22, 3.78]
Total (95% CI)		100.00	2.83 [0.35, 5.31]

Test for heterogeneity: Chi² = 44.66, df = 7 (P < 0.00001), I² = 84.3%

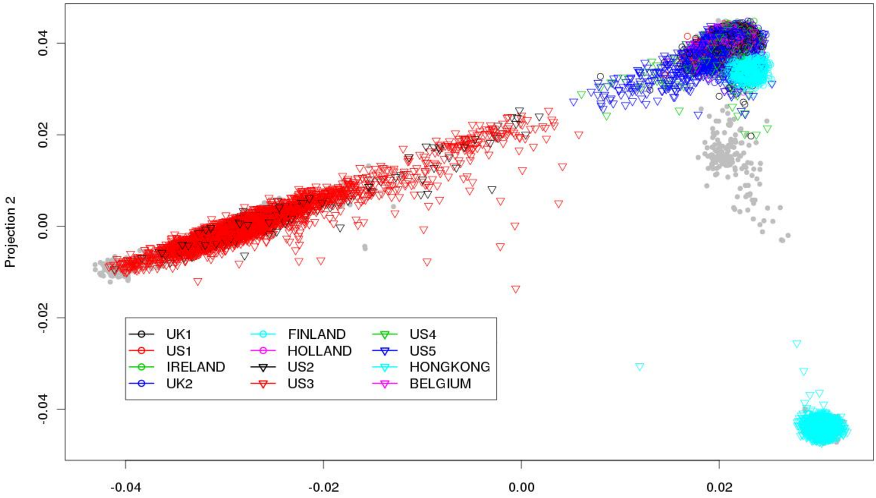
Test for overall effect: Z = 2.24 (P = 0.03)

Existing gene tests are not well-suited for meta-analysis

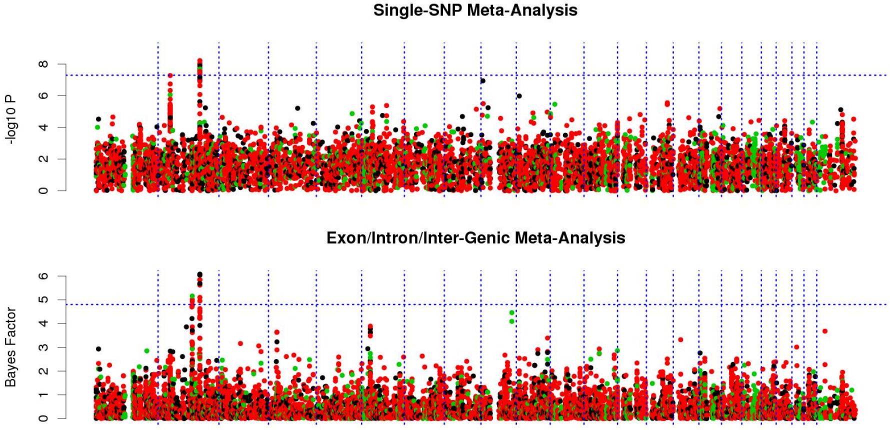
But when we score genes based on heritability, meta-analysis becomes straightforward

12-Cohort Gene-Based Meta-Analysis for Epilepsy

Work for the International League Against Epilepsy (ILAE)

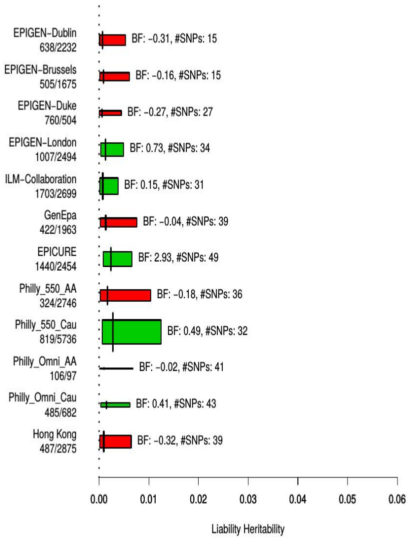


12-Cohort Gene-Based Meta-Analysis for Epilepsy

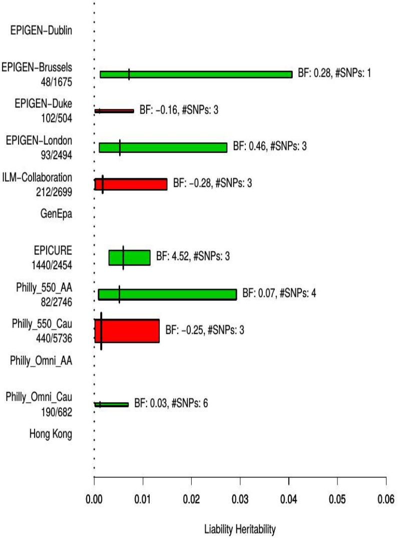


BLACK: EXONS — RED: INTRONS — GREEN: INTER-GENIC

Two Hits on Chromosome 2

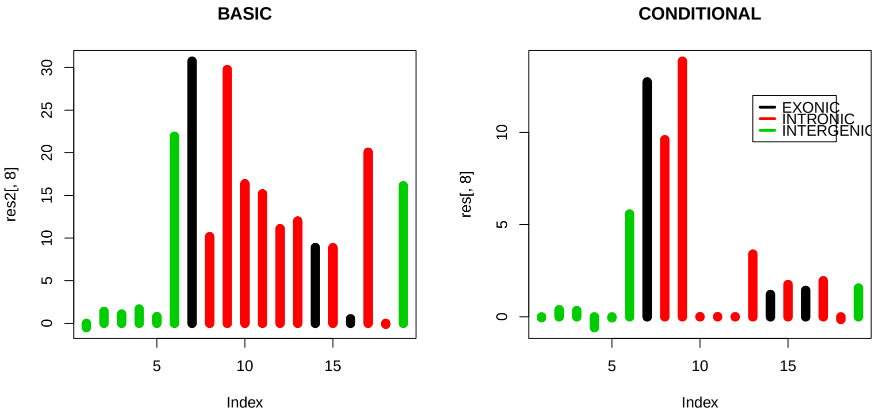


SCN1A



C17orf76-AS1

Fine Mapping of SCN1A



For each region of interest, can regress phenotype on each exon/intron/inter-genic chunk conditional on all other SNPs

1 Estimating SNP Heritability

2 Genome Partitioning

3 Intensity of Heritability

4 Testing Different Models

5 Using Other Datatypes

6 Bivariate Analysis

7 Mixed Model Analysis

8 Gene-Based Association Testing

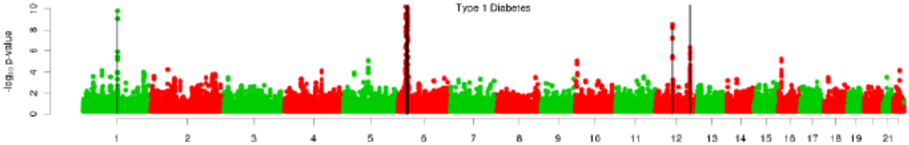
9 Adaptive MultiBLUP

10 Applying to Animal and Plant Data

Adaptive MultiBLUP - See Module 11

Step 1: Divide genome into (say) 75kbp overlapping chunks.

Step 2: Test each chunk for association (using GBAT).



Step 3: Identify all significant chunks (say $P < 10^{-5}$).
(Merge these chunks with neighbouring chunks with $P < 0.01$.)

E.g., for Type 1 Diabetes, obtain 4 local regions.

Step 4: Run MultiBLUP with five random effects.

1 Estimating SNP Heritability

2 Genome Partitioning

3 Intensity of Heritability

4 Testing Different Models

5 Using Other Datatypes

6 Bivariate Analysis

7 Mixed Model Analysis

8 Gene-Based Association Testing

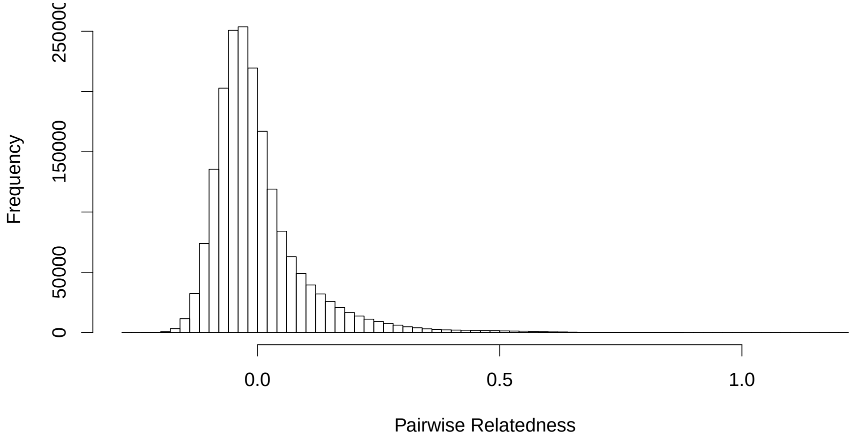
9 Adaptive MultiBLUP

10 Applying to Animal and Plant Data

Applying these Methods to Animal Data

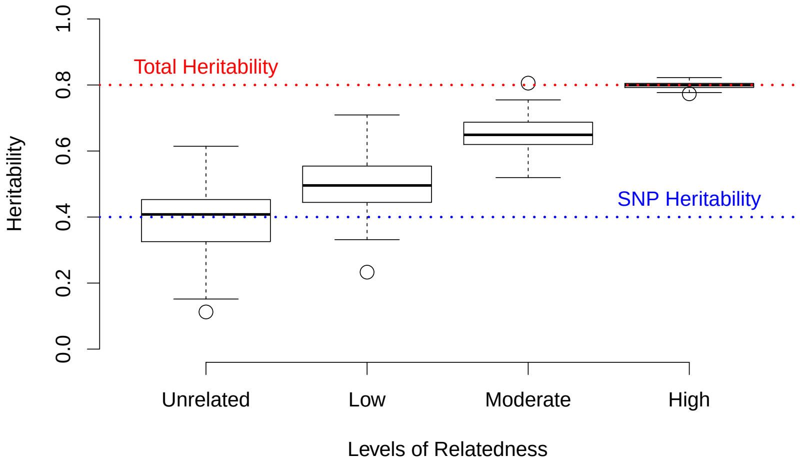
Human Genetics	**Animal Genetics**
Low Relatedness	High Relatedness
10 000s of Individuals	100/1000s of Individuals
Millions of SNPs	10/100 000s of SNPs
Binary Traits	Quantitative Traits
Loadsa Money	Nice Waterproof Clothing

Wellcome Trust Hetergeneous Stock Mice



1940 mice, descended from 8 founders. 10 091 SNPs.

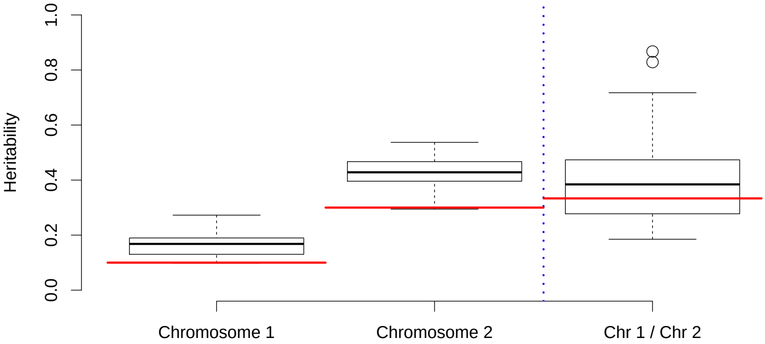
Estimating SNP Heritability



WILL NOT WORK - Requires individuals to be unrelated

Genome Partitioning

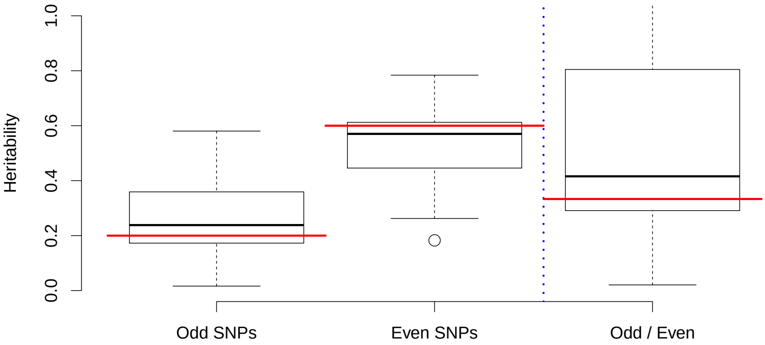
Simulated 80% heritability, of which 10% on Chr 1, 30% on Chr 2.



SEMI WORKS - Absolute values inflated, but relative values reasonable

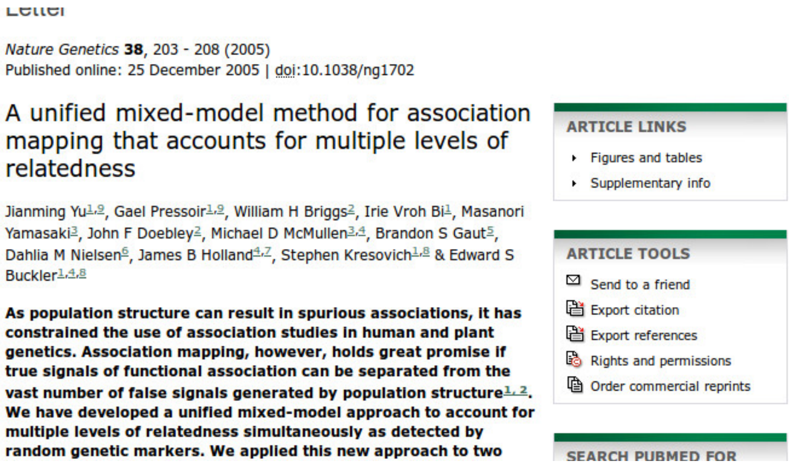
Intensity of Heritability

SNPs 1, 3, 5, ..., 10091 explain 20% of variance
SNPs 2, 4, 6, ..., 10090 explain 60% of variance



WORKS - Although precision low even with sparse SNP density

Mixed Model Association Analysis



Letter

Nature Genetics **38**, 203 - 208 (2005)
Published online: 25 December 2005 | doi:10.1038/ng1702

A unified mixed-model method for association mapping that accounts for multiple levels of relatedness

Jianming Yu[1,9], Gael Pressoir[1,9], William H Briggs[2], Irie Vroh Bi[1], Masanori Yamasaki[3], John F Doebley[2], Michael D McMullen[3,4], Brandon S Gaut[5], Dahlia M Nielsen[6], James B Holland[4,7], Stephen Kresovich[1,8] & Edward S Buckler[1,4,8]

As population structure can result in spurious associations, it has constrained the use of association studies in human and plant genetics. Association mapping, however, holds great promise if true signals of functional association can be separated from the vast number of false signals generated by population structure[1, 2]. We have developed a unified mixed-model approach to account for multiple levels of relatedness simultaneously as detected by random genetic markers. We applied this new approach to two

ARTICLE LINKS
Figures and tables
Supplementary info
ARTICLE TOOLS
Send to a friend
Export citation
Export references
Rights and permissions
Order commercial reprints
SEARCH PUBMED FOR

WORKS - this is what it was designed for!

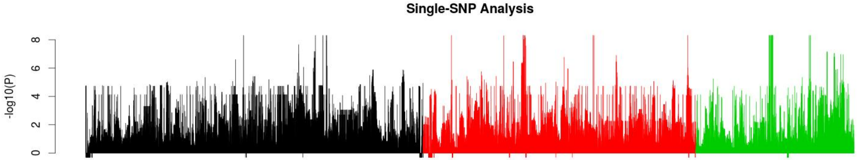
Analysis of Schizosaccharomyces pombe

Data from Dan Jeffares

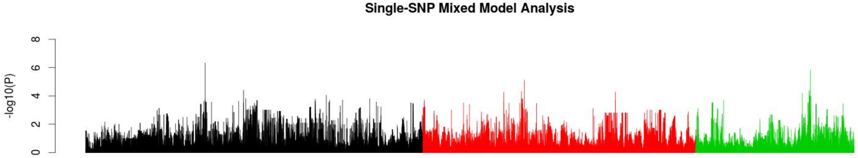
Collection of 161 natural isolates, phenotyped for 53 measurements of cell shape and growth rates.

Data highly structured; while genotyped for over 1 000 000 SNPS, the effective size of the genome is closer to 1000.

Analysis of Schizosaccharomyces pombe



Standard Single-SNP analysis leads to high inflation



This is controlled by mixed model association analysis, but power is low

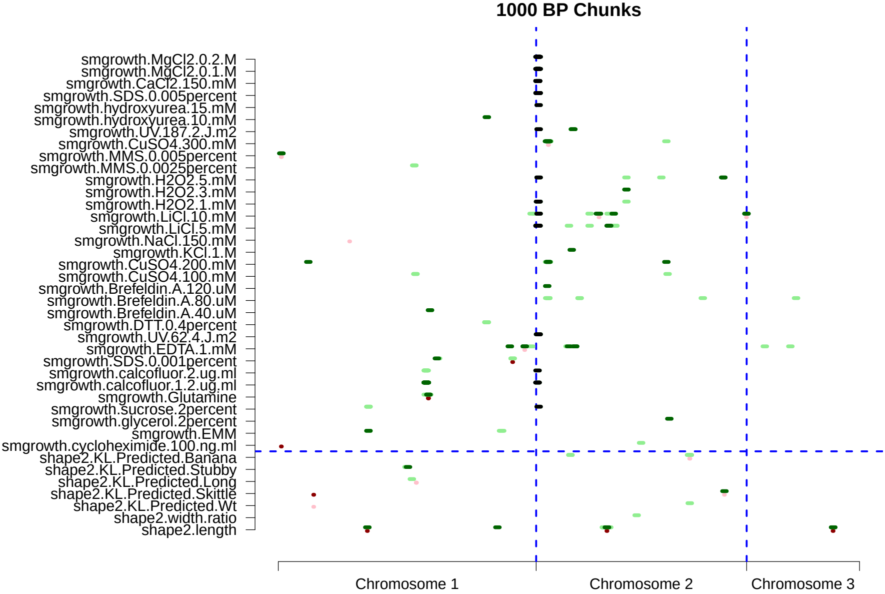
Analysis of Schizosaccharomyces pombe

On next slide: across 58 traits, green are loci significant through chunk-based testing, red are those significant through single-SNP test

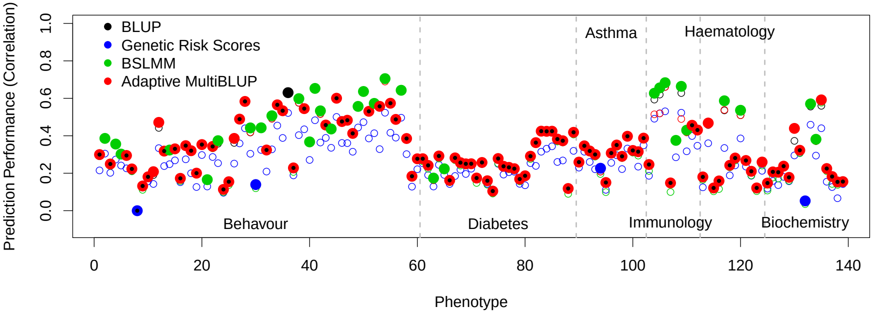
There are many more green (chunk-significant) hits than red (single-SNP-significant) hits

A gene (chunk) based test can have much higher power, because it considers groups of variants (typically in high LD) together and can afford a far lower significance threshold

Analysis of Schizosaccharomyces pombe



Adaptive MultiBLUP



WORKS - across 143 mice traits find modest improvement over BLUP

Summary

Yang, Visscher *et al.* showed that by applying mixed-models to GWAS data for unrelated individuals, it is possible to estimate h^2_{SNP}

While this was a major breakthrough in the missing heritability discussion, in my view, this is just the tip of the iceberg

By realising that different **K** correspond to different underlying models, we can now use (extensions of) SNP-based heritability analysis to investigate traits in fantastic detail

e.g., genome partitioning, intensity of heritability, distribution of h^2, gene-based association analysis, MultiBLUP prediction

While estimating h^2_{SNP} require unrelated individuals

MANY OF THE EXTENSIONS DO NOT