Genetic colocalisation

Chris Wallace 🐦 chr1swallace 🏠 chr1swallace.github.io

wellcome UNIVERSITY OF CAMBRIDGE MRC Biostatistics Unit

Colocalisation: are two traits under control of the same causal variant(s)?

Colocalisation doesn’t care about causality between traits

Literally - do causal variants for two traits share the same location

Study all variants in a single genetic region (~ 1000)

Anticipate correlation between these variants (linkage disequilibrium - LD)

Explicitly do not assume causal variant for either is known

Colocalisation doesn’t care about causality between traits

Literally - do causal variants for two traits share the same location

Study all variants in a single genetic region (~ 1000)

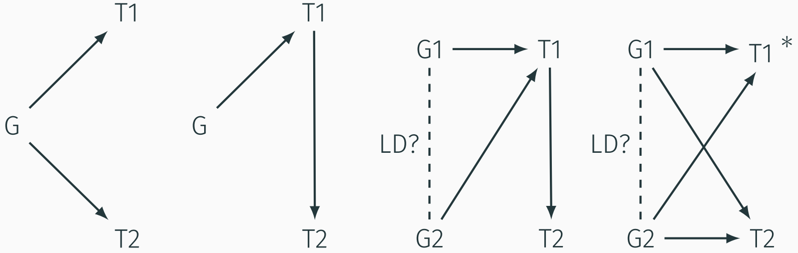
Anticipate correlation between these variants (linkage disequilibrium - LD)

Explicitly do not assume causal variant for either is known

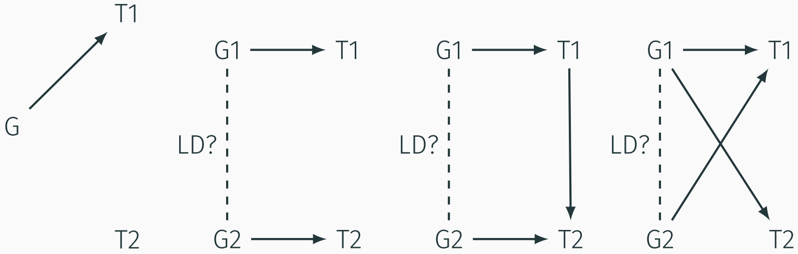
Nonetheless - often used to make causal inference if one trait is biologically likely to precede another

This ignores potential for single variant to have two independent effects

✓These signals colocalise



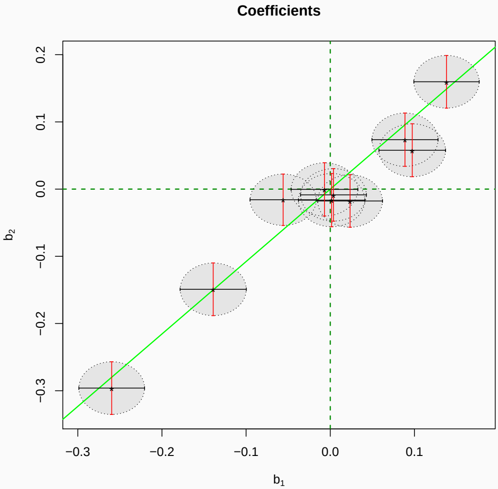
✗These signals don't



(1) Proportional colocalisation

Proportional colocalisation

If two traits share one or more causal variants, then regression coefficients for the traits against any set of variants *in the neighbourhood of those causal variants* should be proportional.



Testing proportionality

Let $\mathbf{b}_X$, $\mathbf{b}_Y$ be estimates of regression coefficients $\boldsymbol{\beta}_X$, $\boldsymbol{\beta}_Y$, with variance-covariance matrices $\mathbf{V}_X$ and $\mathbf{V}_Y$ respectively.

$$H_0 : \boldsymbol{\beta}_X = \frac{1}{\eta}\boldsymbol{\beta}_Y \text{ for some } \eta$$

If η known $X(\eta)^2 = \mathbf{u}^T\mathbf{V}^{-1}\mathbf{u} \sim \chi^2_q$, where

$$\mathbf{u} = \left(\mathbf{b}_X - \frac{1}{\eta}\mathbf{b}_Y\right), \qquad \mathbf{V} = \mathbf{V}_X + \frac{1}{\eta^2}\mathbf{V}_Y$$

Replace η by $\hat{\eta}$, then $X(\hat{\eta})^2 \sim \chi^2_{q-1}$

Plagnol et al., Biostatistics 2009

Testing proportionality

Note: η is a nuisance parameter (initially)

Null hypothesis is proportionality

Test is for departure from proportionality → failure to reject the null can be colocalisation *or* lack of power

Selection of variants

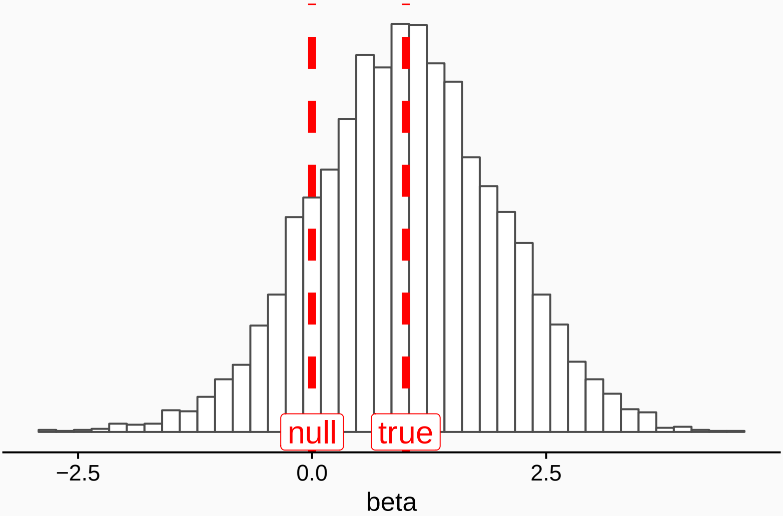
Too many variants → too many degrees of freedom

Weakly associated variants → loss of power

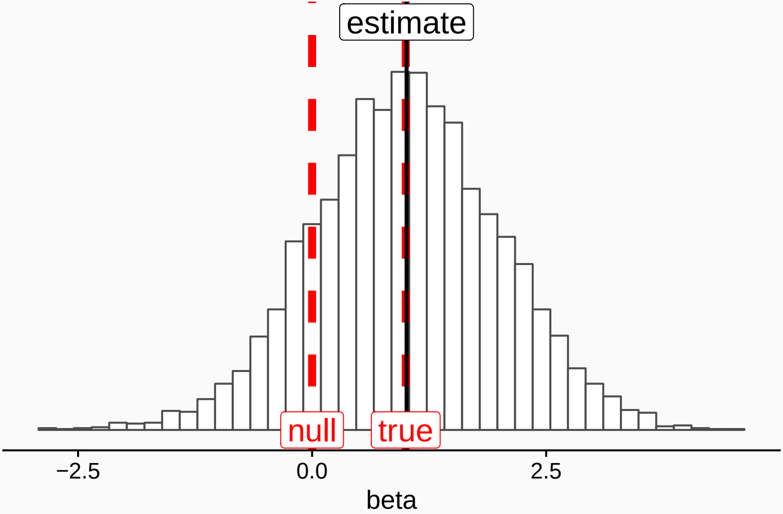
Selecting most significant → biased coefficients (winner's curse)

Wallace, Genet Epidemiol, 2014

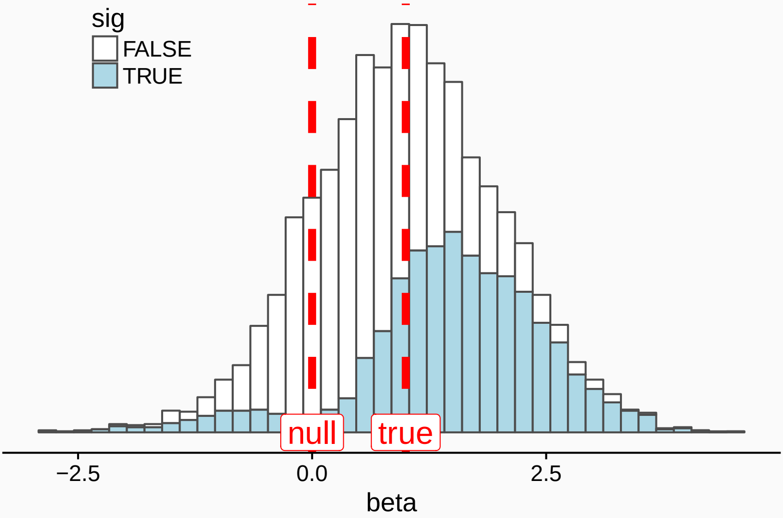
Winner’s curse



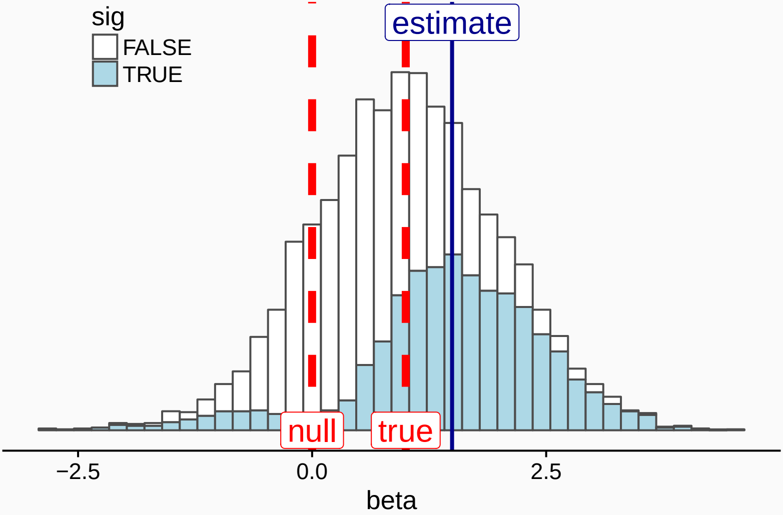
Winner’s curse



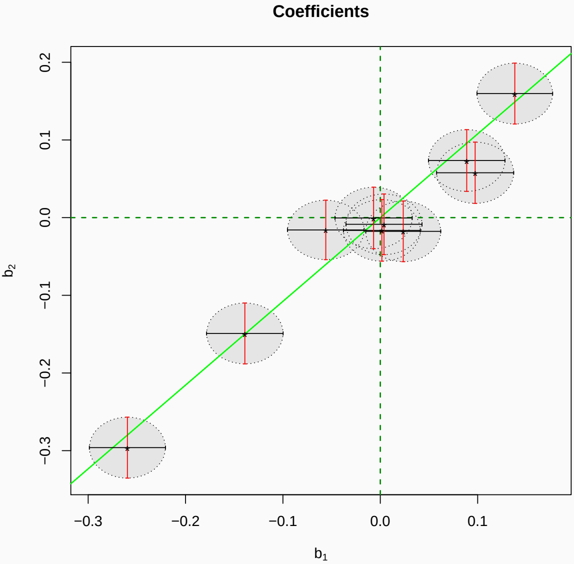
Winner's curse



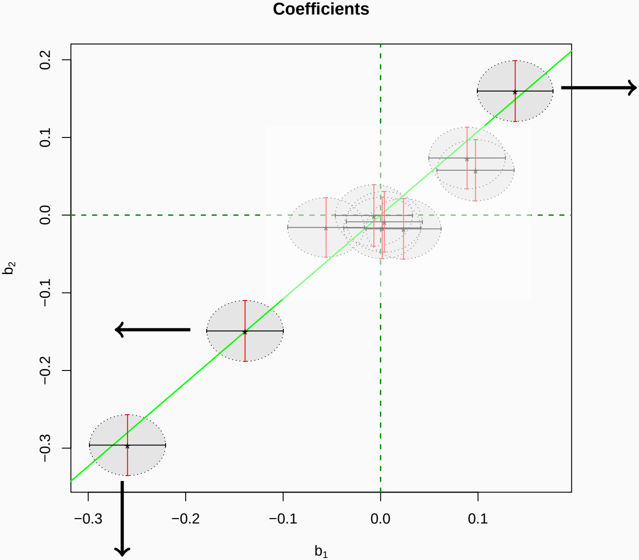
Winner's curse



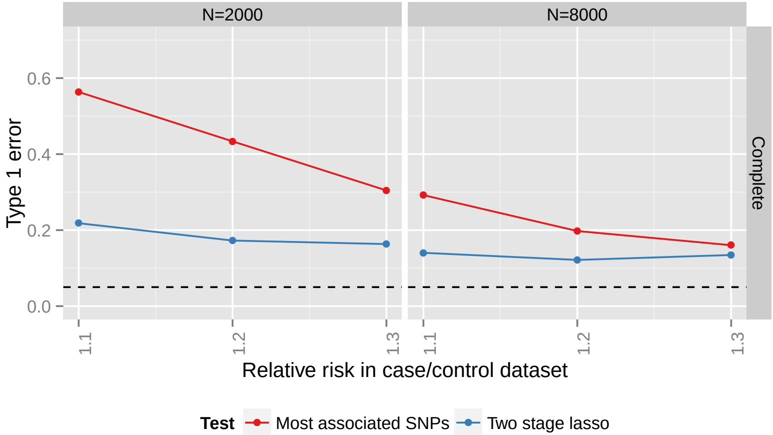
Proportionality test for selected “best” variants



Proportionality test for selected “best” variants



Using most associated variants inflates type 1 error



Two proposed solutions

Principal components

- Summarize genetic variation by principal components
- Do a (high degree of freedom) test for colocalisation based on the most important components
- *No obvious choice for optimal number of principal components.*

Wallace, Genet Epidemiol, *2013*

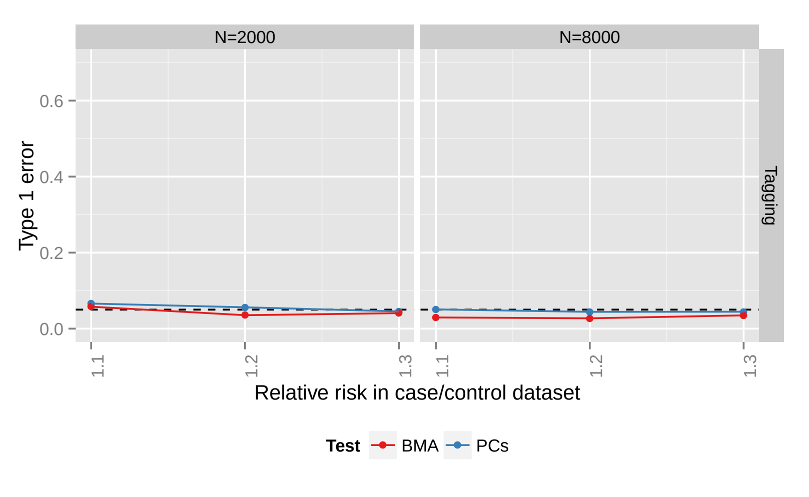
Two proposed solutions

Model Averaging

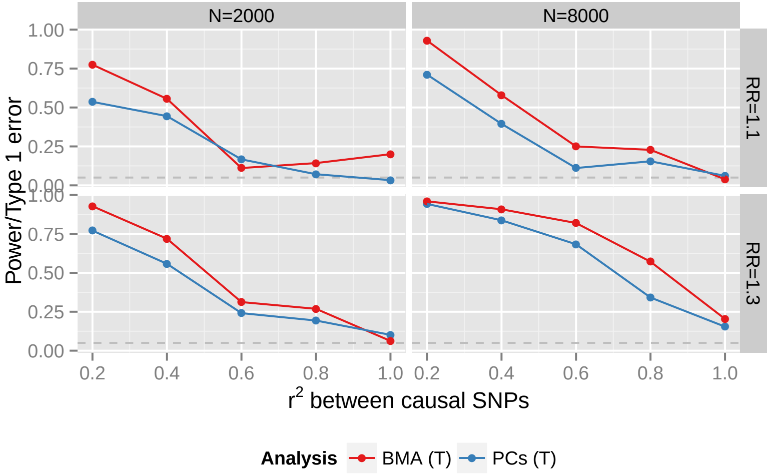
- Test colocalisation for all possible two SNP models
- Calculate model probabilities via approximate Bayes factors
- Average p values, weighted by model posterior probabilities
- Generate a posterior predictive p value.
- *Computationally slower (but tolerable: minutes not hours).*
- NB Posterior predictive p values are not p values!

Wallace, Genet Epidemiol, *2013*

Either maintains type 1 error...



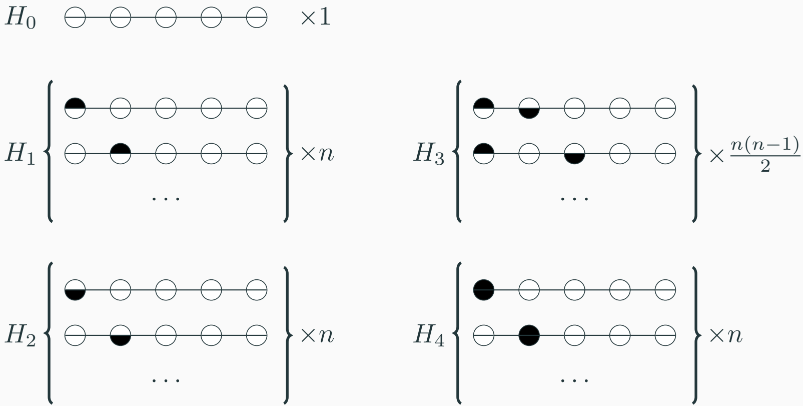
… but BMA tends to have better power



(2) Colocalisation analyis via enumeration

Alternative approach

- Assume at most one causal variant in a region
- Enumerate all possible configurations of association and partition into hypotheses



Giambartolomei et al., PLOS Genet 2014

Use Bayes factors to summarize evidence

Bayes factor for a configuration for SNP i, trait k is

$$BF_{i0}^{(k)} = \frac{P(D|\text{SNP } i \text{ causal})}{P(D|\text{no SNP causal})} = \frac{P(D|\text{SNP } i \text{ causal})}{P(D|\text{SNP } i \text{ not causal})}$$

Bayes factor for a set of configurations can be calculated for independent datasets as

$$BF_{10} = \sum_i BF_{i0}^{(1)} \qquad BF_{30} = \sum_{i \neq j} BF_{i0}^{(1)} BF_{j0}^{(2)}$$

$$BF_{20} = \sum_i BF_{i0}^{(2)} \qquad BF_{40} = \sum_i BF_{i0}^{(1)} BF_{i0}^{(2)}$$

Incorporates no prior knowledge about relative effect sizes

Giambartolomei et al., PLOS Genet 2014

Approximate Bayes factor calculations

Given estimated effect size for SNP i, trait k, $\hat{\beta}_i^{(k)}$ and standard error $\sqrt{V_i^{(k)}}$:

$$BF_{i0}^{(k)} = \frac{1}{\sqrt{1-r}} \exp\left(\frac{-{Z_i^{(k)}}^2}{2} r \right)$$

$$\hat{\beta}_i^{(k)} \sim N(0, W) \qquad Z_i^{(k)} = \frac{\hat{\beta}_i^{(k)}}{\sqrt{V_i^{(k)}}} \qquad r = \frac{W}{V_i^{(k)} + W}$$

- W chosen according to effect size considered unlikely *a priori*
- eg, if $V = 1$, W$= 0.15$ corresponds to an effect explaining 1% of trait variance
- only summary statistics required

Wakefield, Genet Epidemiol 2009

Priors

$$\pi(H_1) = 1 \times 10^{-4} \qquad \pi(H_3) = 1 \times 10^{-8}$$
$$\pi(H_2) = 1 \times 10^{-4} \qquad \pi(H_4) = 1 \times 10^{-5}$$

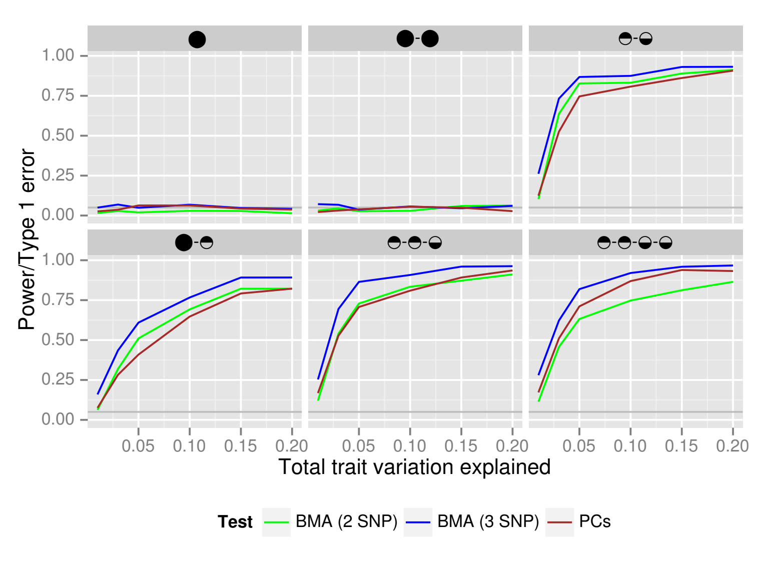
$\pi(H_1)$, $\pi(H_2)$ correspond to the proportion of genotyped SNPs thought to be associated with disease.

$\pi(H_4)/\pi(H_1)$ corresponds to the fraction of SNPs associated with trait 1 we think may be also be associated with trait 2.

By working with Bayes factors, can consider a range of priors without computational cost.

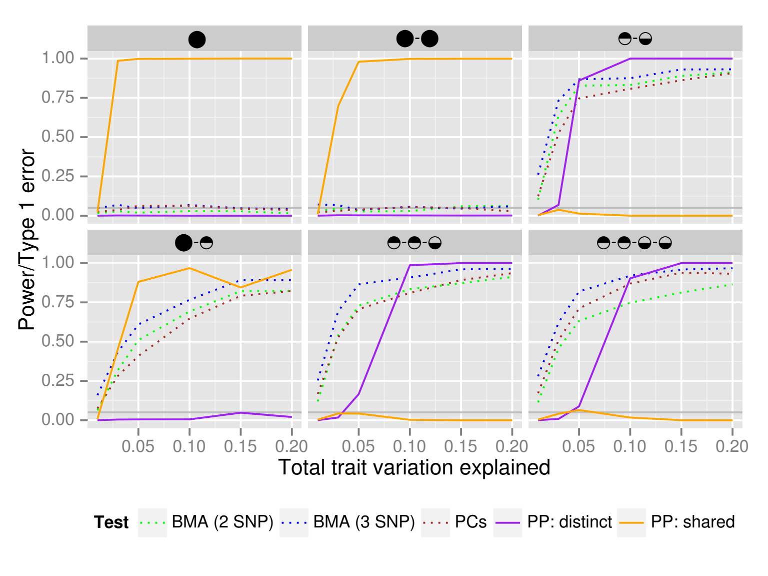
Comparision of proportional and enumeration approaches

Effect of multiple causal variants on colocalisation



Giambartolomei et al., PLOS Genet 2014

Effect of multiple causal variants on colocalisation



Giambartolomei et al., PLOS Genet 2014

Methods summary

Proportional testing

- Must be applied carefully (selection of variants)
- Requires access to full genotype data for both traits - can be difficult
- Prioritises any non-sharing

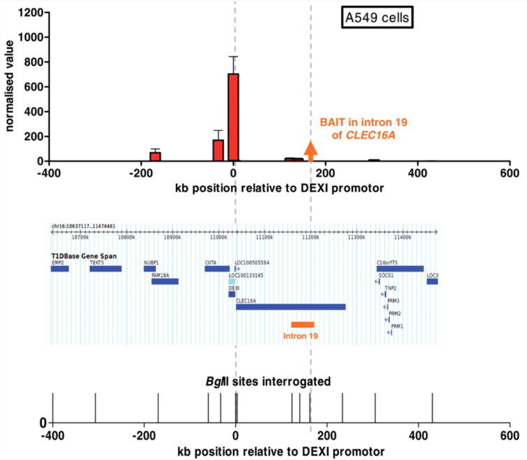
Enumeration

- Accepts summary statistics
- Assumes single causal variant
- Prioritises any sharing

Applications of colocalisation

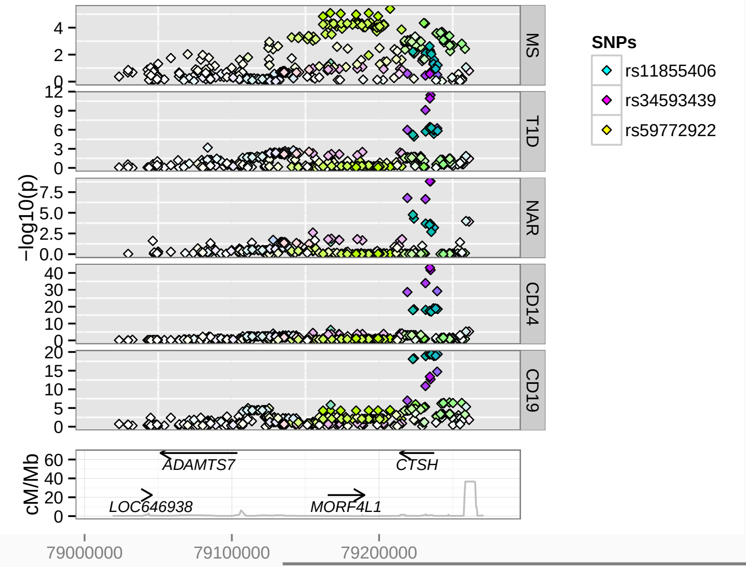
Colocalisation of T1D signals, monocyte eQTLs identified *DEXI* as a candidate causal T1D gene

Additional support from chromosome conformation capture



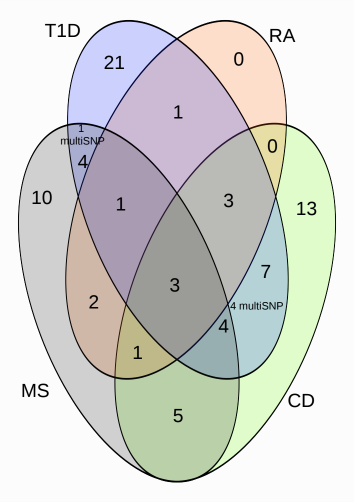
Wallace et al., Hum Mol Genet 2012, *Davison et al.*, Hum Mol Genet 2012

Six candidate causal autoimmune genes identified with confidence in monocytes and B cells



Guo et al., Hum Mol Genet, *2015*

Can be applied to pairs of diseases, allowing for shared controls



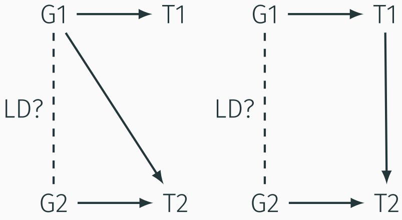
Fortune et al., Nat Genet, *in press*

Discussion

Colocalisation is **not** MR but if a trait is suitable for MR, then it should colocalise with target trait

Colocalisation explicitly tackles uncertainty in causal variants, winner's curse, shared subjects

Does interpretation of colocalisation for partial sharing depend on trait labels?



G1 T1
LD?
G2 T2

G1 T1
LD?
G2 T2

References

Proportional testing

- Plagnol et al. 2009. “Statistical Independence of the Colocalized Association Signals for Type 1 Diabetes and RPS26 Gene Expression on Chromosome 12q13.” *Biostatistics*
- Wallace 2013. “Statistical Testing of Shared Genetic Control for Potentially Related Traits.” *Genet Epidemiol*

Enumeration

- Wakefield 2009. “Bayes Factors for Genome-Wide Association Studies: Comparison with P -Values.” *Genet Epidemiol*
- Giambartolomei et al. 2014. “Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics.” *PLoS Genet*
- Fortune et al. 2015. “Statistical Colocalization of Genetic Risk Variants for Related Autoimmune Diseases in the Context of Common Controls.” *Nat Genet*

Examples

- Guo et al. 2015. “Integration of Disease Association and eQTL Data Using a Bayesian Colocalisation Approach Highlights Six Candidate Causal Genes in Immune-Mediated Diseases.” *Hum Mol Genet*
- Davison et al. 2012. “Long-Range DNA Looping and Gene Expression Analyses Identify *DEXI* as an Autoimmune Disease Candidate Gene.” *Hum Mol Genet*

Thanks to

Claudia Giambartolomei, Vincent Plagnol, Mary Fortune, Hui Guo

W wellcome

UNIVERSITY OF CAMBRIDGE

MRC Biostatistics Unit