Colocalization theory

Benjamin Neale, PhD

Boulder Workshop 2026

Early conceptualization

JOURNAL ARTICLE

Statistical independence of the colocalized association signals for type 1 diabetes and *RPS26* gene expression on chromosome 12q13

Vincent Plagnol, Deborah J. Smyth, John A. Todd, David G. Clayton

Biostatistics, Volume 10, Issue 2, April 2009, Pages 327–334,
https://doi.org/10.1093/biostatistics/kxn039
Published: 27 November 2008 **Article history**

“Following the recent success of genome-wide association studies in uncovering disease-associated genetic variants, the next challenge is to understand how these variants affect downstream pathways. The most proximal trait to a disease-associated variant, most commonly a single nucleotide polymorphism (SNP), is differential gene expression due to the *cis* effect of SNP alleles on transcription, translation, and/or splicing gene expression quantitative trait loci (eQTL). ”

Formalizing colocalization as a test of the null that the same SNPs drive the signal

Consider 2 SNP effect estimates - this test assess whether:
$\beta_1 \propto \beta_2$, i.e. $\beta_1 = (1/\eta)\beta_2 = \beta$

Human Molecular Genetics, 2012, Vol. 21, No. 12 **2815–2824**
doi:10.1093/hmg/dds098
Advance Access published on March 8, 2012

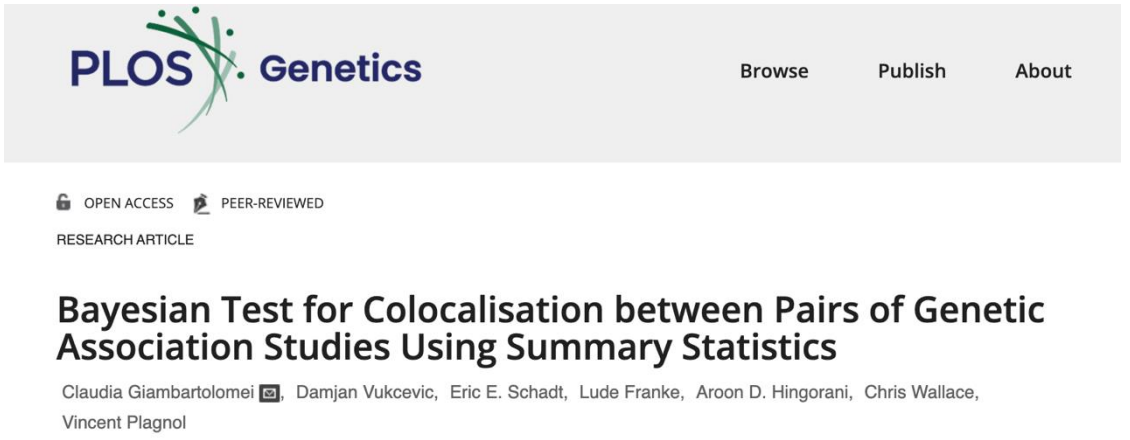
Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes

Chris Wallace[1,*,†], Maxime Rotival[2,†], Jason D. Cooper[1], Catherine M. Rice[3], Jennie H.M. Yang[1], Mhairi McNeill[1], Deborah J. Smyth[1], David Niblett[3], François Cambien[2], The Cardiogenics Consortium, Laurence Tiret[2], John A. Todd[1], David G. Clayton[1] and Stefan Blankenberg[4,*]

Shifting to a Bayesian approach and formalizing outcomes

Contemplates 5 hypotheses for the patterns of association:

- H_0: No association with either trait
- H_1: Association with trait 1, not with trait 2
- H_2: Association with trait 2, not with trait 1
- H_3: Association with trait 1 and trait 2, two independent SNPs
- H_4: Association with trait 1 and trait 2, one shared SNP



Extending to multiple potential causal variants in the locus

Estimates the colocalization posterior probability (CLPP) that is the chance that each variant is driving the association signal in both studies

$$\phi_i = \frac{\sum_{C^{*(\mathrm{p})}} \mathrm{P}\left(S^{(\mathrm{p})} \mid C^{(\mathrm{p})} = C^{*(\mathrm{p})}\right)\mathrm{P}\left(C^{*(\mathrm{p})}\right)\mathbb{I}\left(c_i^{*(\mathrm{p})} = 1\right)}{\sum_{C^{*(\mathrm{p})}} \mathrm{P}\left(S^{(\mathrm{p})} \mid C^{(\mathrm{p})} = C^{*(\mathrm{p})}\right)\mathrm{P}\left(C^{*(\mathrm{p})}\right)} \times \frac{\sum_{C^{*(\mathrm{e})}} \mathrm{P}\left(S^{(\mathrm{e})} \mid C^{(\mathrm{e})} = C^{*(\mathrm{e})}\right)\mathrm{P}\left(C^{*(\mathrm{e})}\right)\mathbb{I}\left(c_i^{*(\mathrm{e})} = 1\right)}{\sum_{C^{*(\mathrm{e})}} \mathrm{P}\left(S^{(\mathrm{e})} \mid C^{(\mathrm{e})} = C^{*(\mathrm{e})}\right)\mathrm{P}\left(C^{*(\mathrm{e})}\right)}.$$

(Equation 8)

ARTICLE

Colocalization of GWAS and eQTL Signals Detects Target Genes

Farhad Hormozdiari,[1] Martijn van de Bunt,[2,3] Ayellet V. Segrè,[4] Xiao Li,[4] Jong Wha J. Joo,[1] Michael Bilow,[1] Jae Hoon Sul,[5,6] Sriram Sankararaman,[1,8] Bogdan Pasaniuc,[7,8] and Eleazar Eskin[1,8,*]

Conclusions and considerations

- Functional variation is widespread and so methods that allow for multiple causal variants are essential
- Data quality issues including accuracy of genotype calling and imputation are critical potential confounds that may bias the results