THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

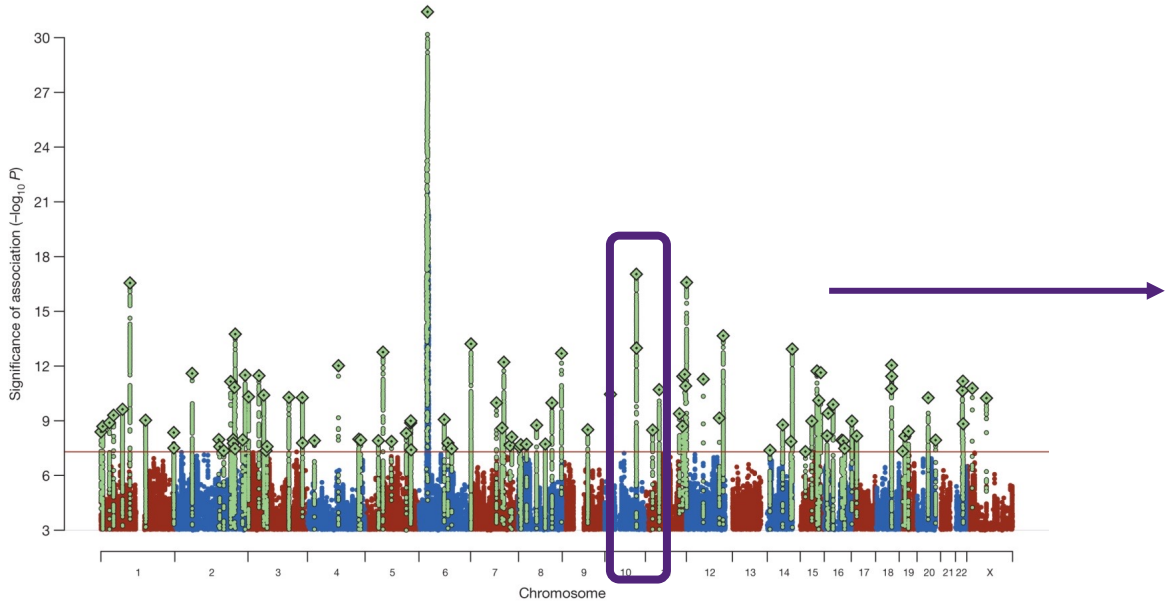
Lecture 11: Fine Mapping

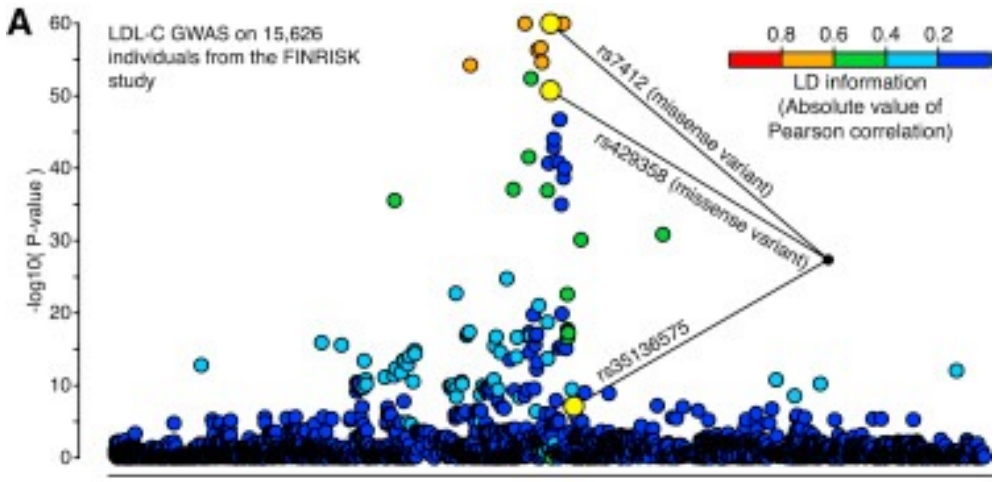
Genetics & Genomics Winter School
Module 1

Alesha Hatton

What is fine-mapping?

An approach to identify and prioritise SNPs driving GWAS association signals





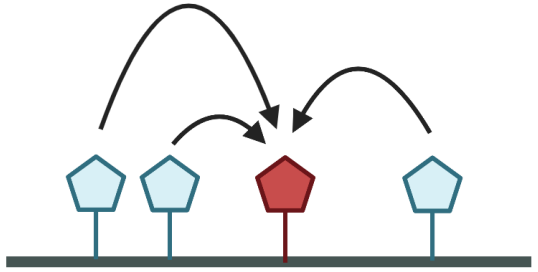
S Ripke et al. *Nature* (2014)

Benner et al. *AJHG*, 2017

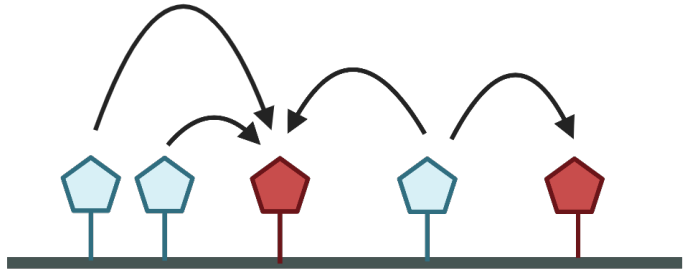
Why don't we take the top associated SNP?

Red – causal variant

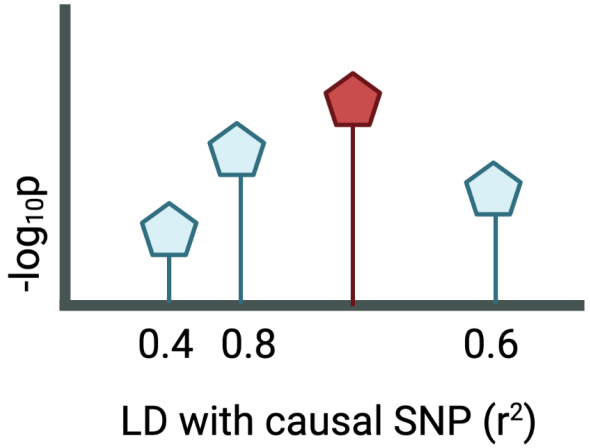
Single causal variant



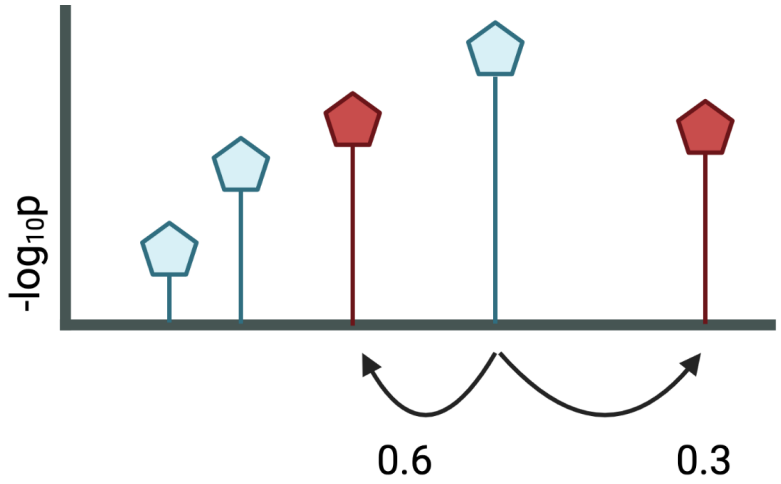
What about when there is ≥ 1 causal variant



GWAS result

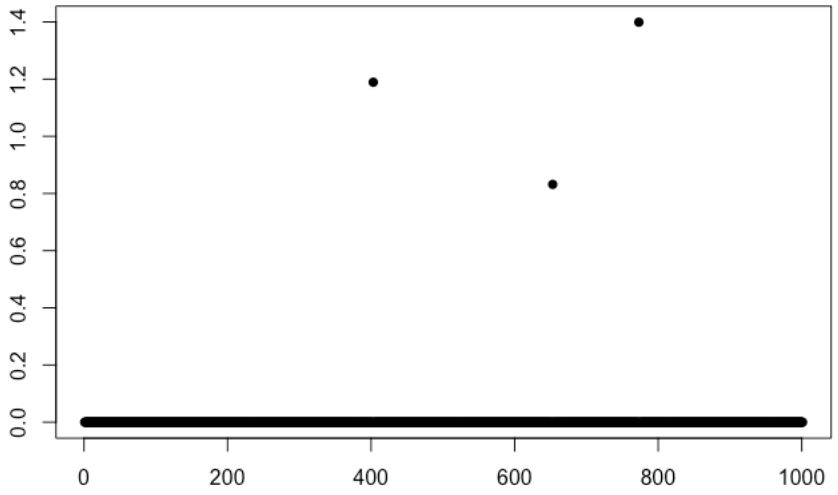


GWAS result

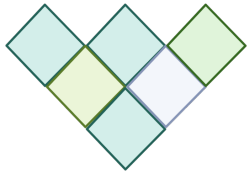


Simplistic fine-mapping example

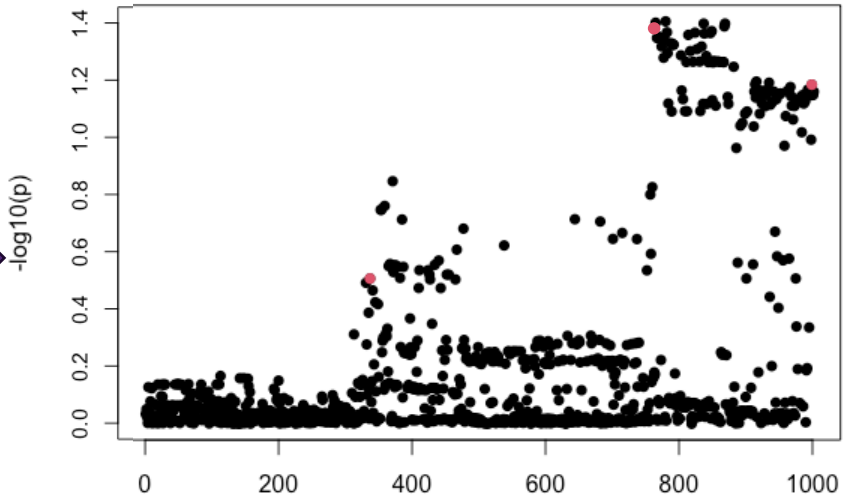
True effects



+ LD



Estimated effect



Fine mapping attempts to do the reverse..

https://stephenslab.github.io/susieR/articles/index.html

Many fine-mapping methods

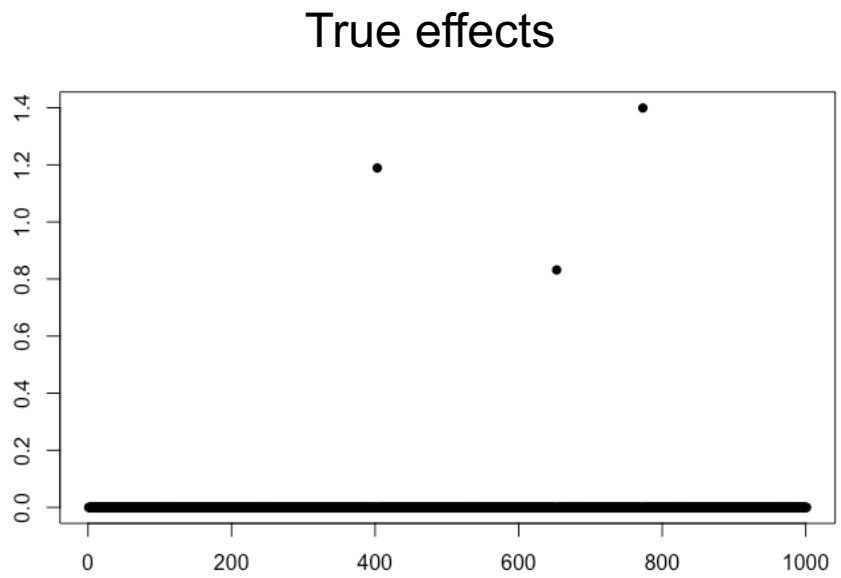
Majority of methods follow a Bayesian framework

- prior × data = posterior

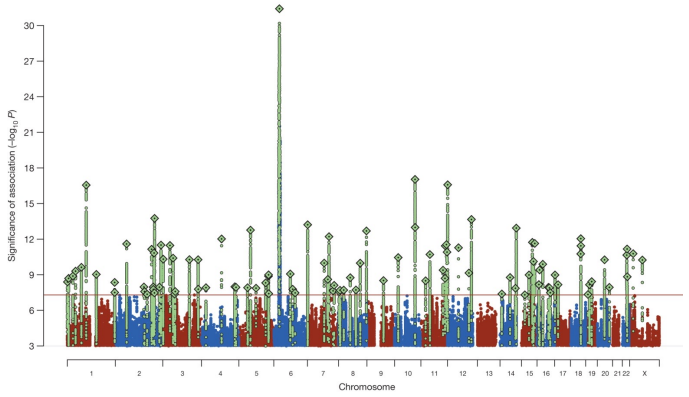
E.g. Prior knowledge of distribution of true SNP effects

Methods

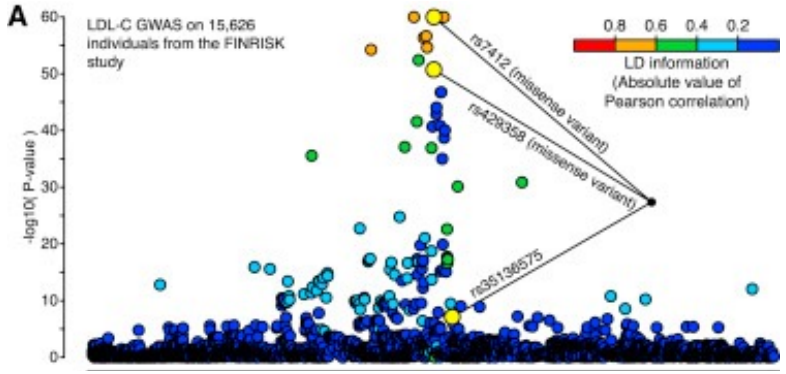
- BIMBAM (Servin and Stephens, 2007)
- CAVIAR (Hormozdiari, 2014)
- CAVIARBF (Chen, 2015)
- FINEMAP (Benner, 2016)
- PAINTOR (Kichaev, 2014)
- SuSiE (Wang 2020)

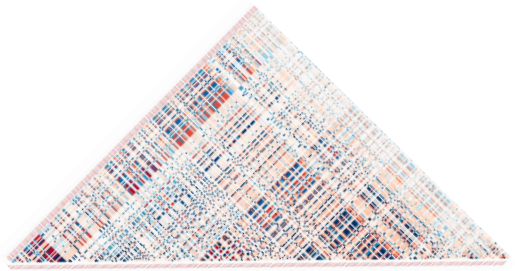


How does this work?



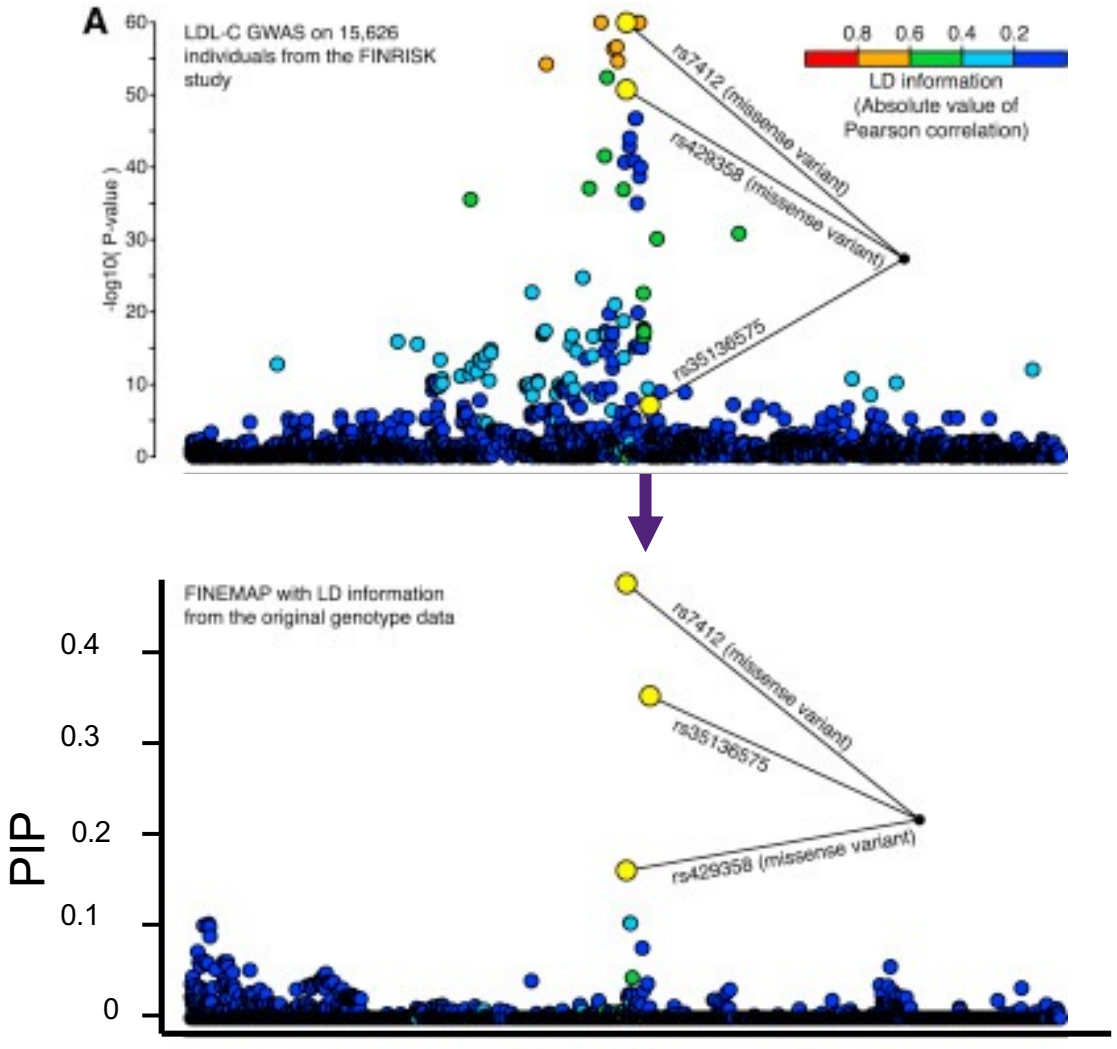
Each SNP has equal probability of being causal
Joint SNP effects are sparse





Probability of causality for each SNP

Posterior inclusion probability (PIP)



Probability (according to the model) the variant is casual

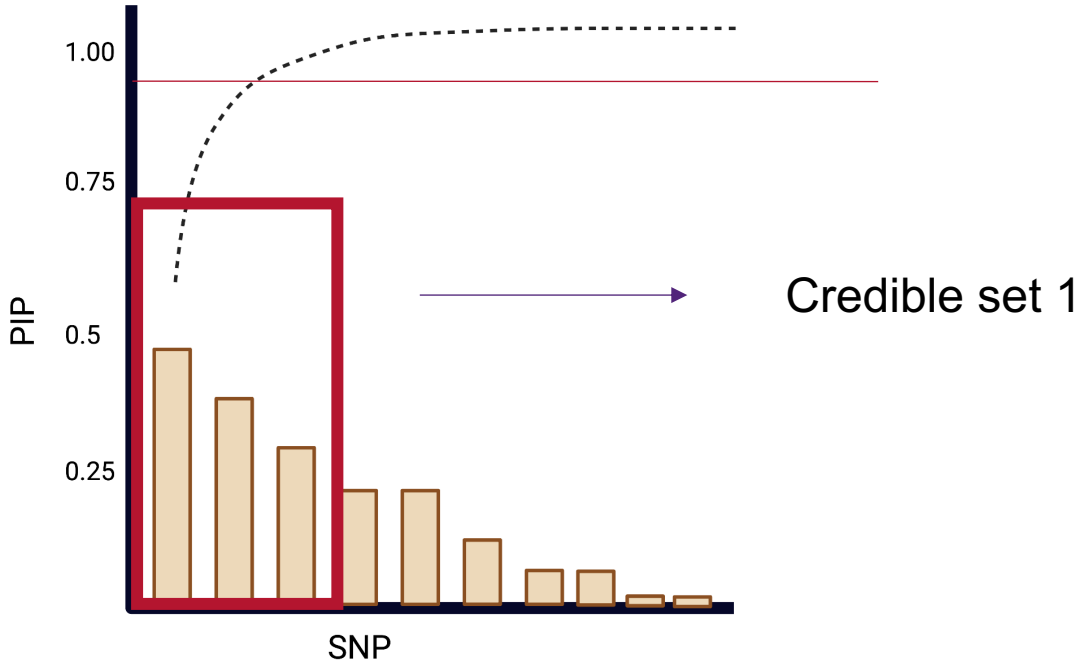
$$\mathrm{PIP_i} = \Pr\left(b_j \neq 0 \,\middle|\, X, Y\right)$$

⇧ PIP = more confidence

⇩ PIP = less likely to be driving signal

Maller et al. Nature Genetics (2012)

Credible Sets (CS)



Threshold	0.95
Size	0.97

A set of putative causal variants for further investigation.

- Sorting PIP for each SNPs in descending order
- Cumulatively sum until reach the threshold

Formal Definition

Smallest set of SNPs with >95% probability of containing a single causal variant

Multiple causal variants

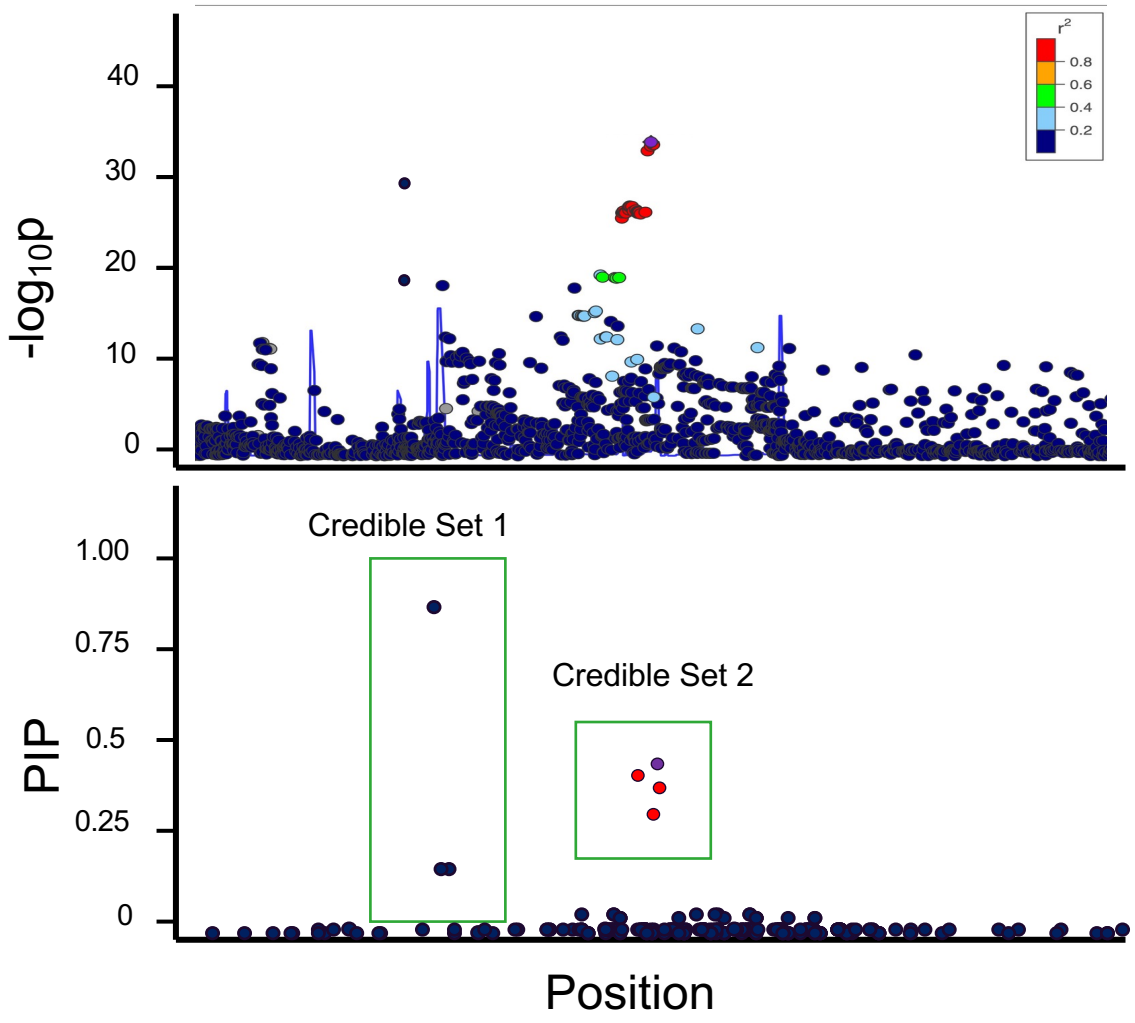
Most newer methods allow for the assumption of multiple causal variants

Multiple credible sets

- Smallest set of SNPs with >95% probability of containing a single causal variant

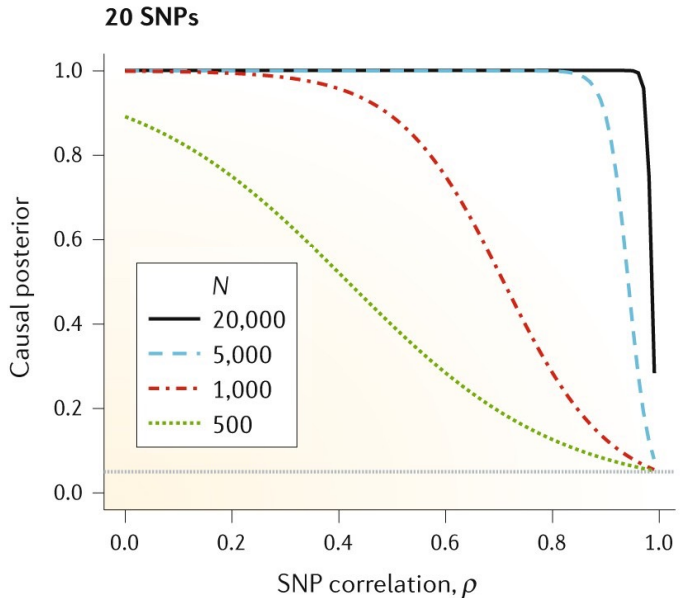
In general, the more independent signals, the less statistical power to detect credible sets

Specify minimum correlation between SNPs allowed within a credible set



Factors influence fine-mapping performance

- The local LD structure
- Sample size
- Number of causal SNPs in a region
- LD reference matches the data
- Whether the causal variants are measured



Example:

- 20 SNPs
- All SNPs have equal LD (x-axis)
- One causal SNP (R^2 1%)
- PIP of causal SNP on yaxis

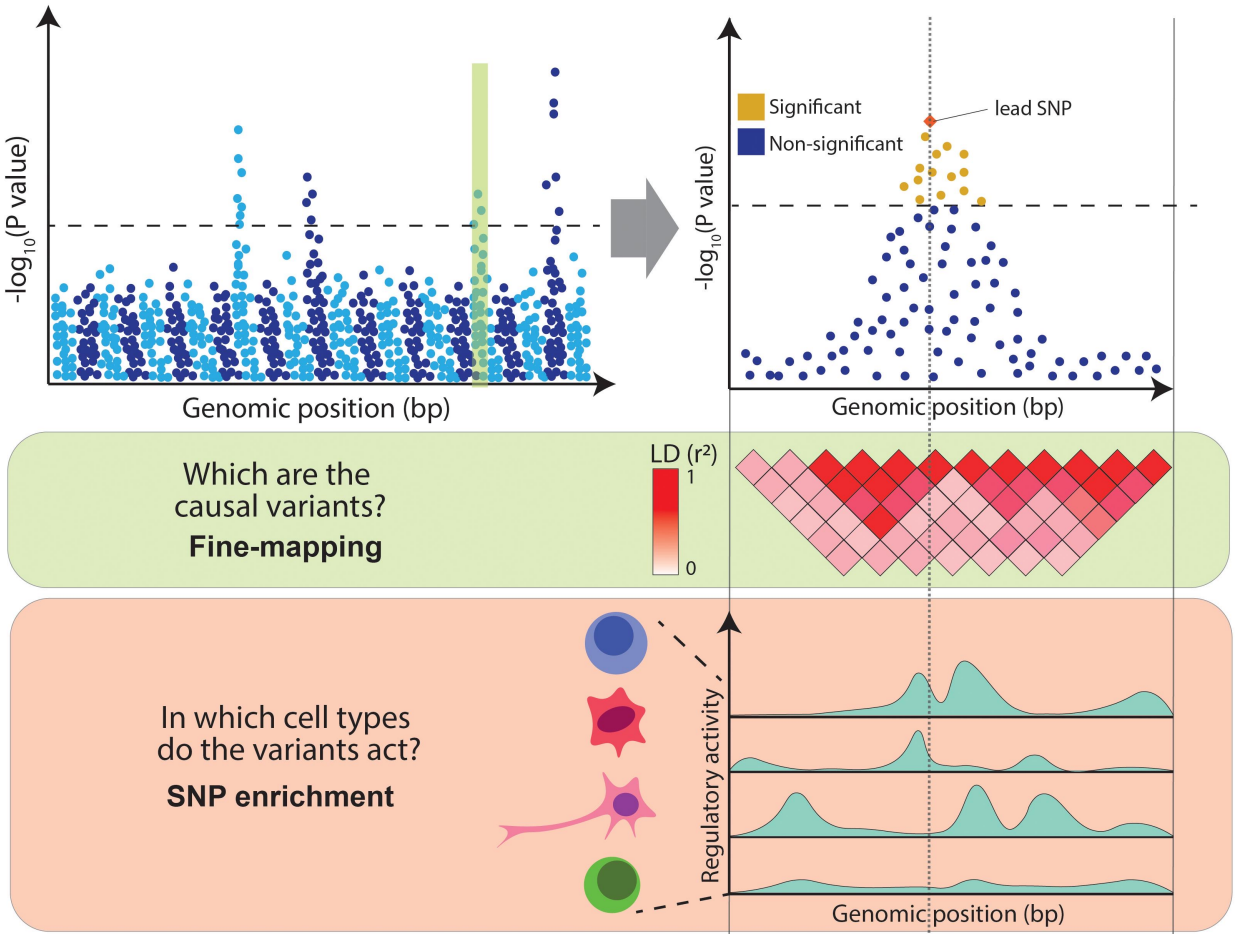
Schaid, 2018 Nature Reviews Genetics

Integrate functional annotation

Use our “prior” knowledge about the SNPs

1. Estimates prior causal probabilities for all SNPs using functional annotations
2. Perform fine-mapping using these prior causal probabilities.

Methods: PolyFun, PAINTOR, fastPAINTOR, CAVIARBF

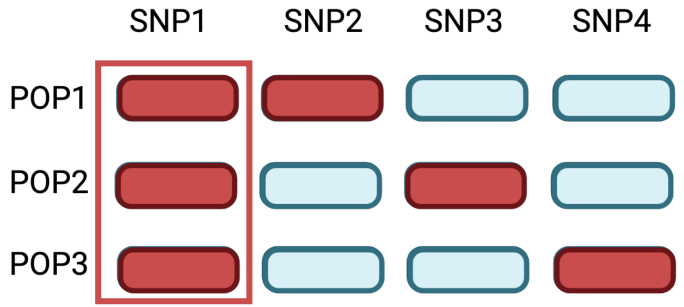


Cano-Gamez et al. Frontiers in Genetics (2020)

Cross-ancestry fine-mapping

- Utilise populations of different ancestries to prioritise SNPs
- Relies on the assumption that causal variants are shared between populations (generally supported in literature).
- Leveraging differences in LD between populations
- Methods: SuSiEx, MS-Caviar

Example



Biobank Japan (BBJ)
79 complex traits

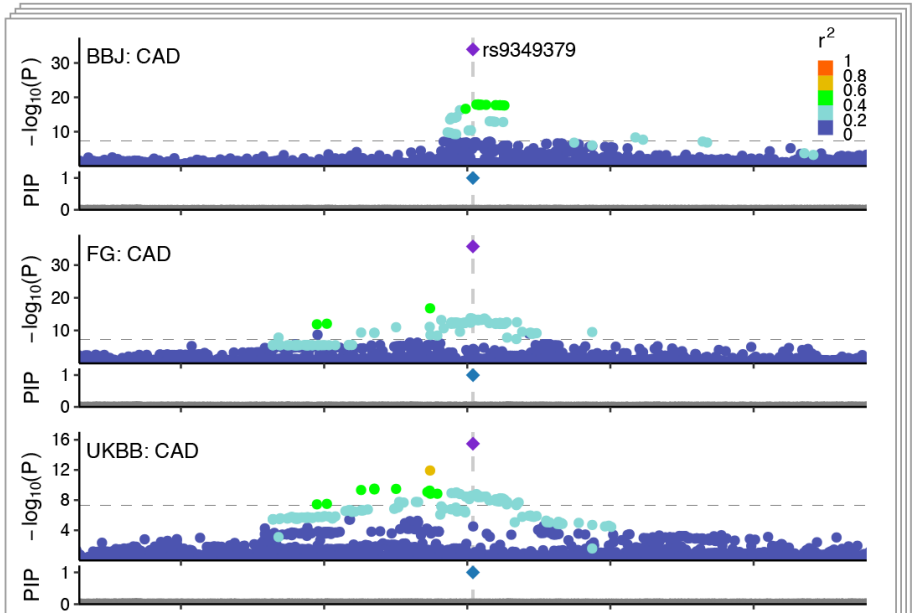
Japanese
(*n* = 178,726)

FinnGen (FG) release 6
67 complex traits

Finnish
(*n* = 271,341)

UK Biobank (UKBB)
119 complex traits

White British
(*n* = 361,194)



Kanai et al, medRxiv (2021)

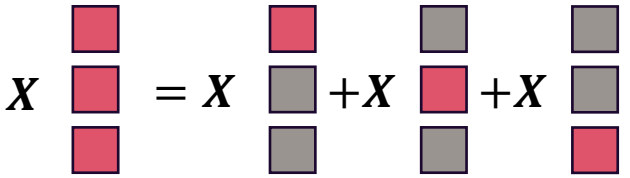
Sum of Single Effects (SuSiE)

Methods for fine-mapping multiple causal variants sets

For each causal variant $\boldsymbol{b_i} = (b_{l1}, \ldots, b_{lJ})$ single effect vector

For multiple causal variants, sums over multiple vectors of single effects

$\boldsymbol{b} = \sum \boldsymbol{b_i}$



Iterative Bayesian stepwise selection

- Can quantify uncertainty in variables selected

Outputs 95% credible sets with PIPs for each SNP

Wang 2020
Zou 2022

Genetic Mapping Session 11 - Fine-mapping

Alesha Hatton

UQWS 2023

Objective

The goal of fine-mapping is to prioritise SNPs that are most likely to be causal (or in LD with causal variants).

In this practical session will apply the Sum of Single Effects (SuSiE) fine-mapping method to investigate the genetic effects underlying height.

We will use summary statistics from Wood et al (2014). This is a GWAS meta-analysis of adult height from 79 studies consisting of 253,288 individuals of European ancestry. This study identified 697 SNPs that reached genome-wide significance ($P < 5 \times 10^{-8}$) that were approximately conditionally independent. We will fine-map some of the regions surrounding these loci.