



MODULE 1 | GENETIC MAPPING

Session 9. Fine Mapping

July 2026

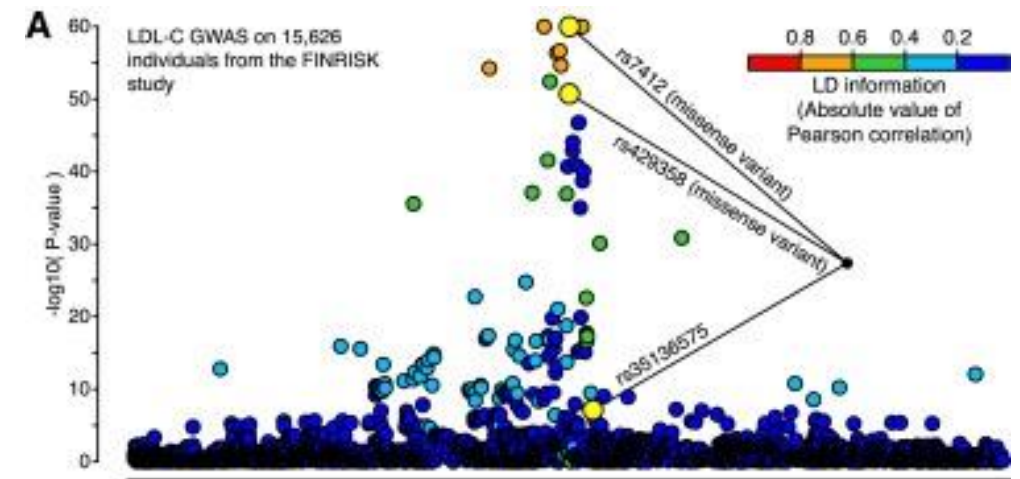
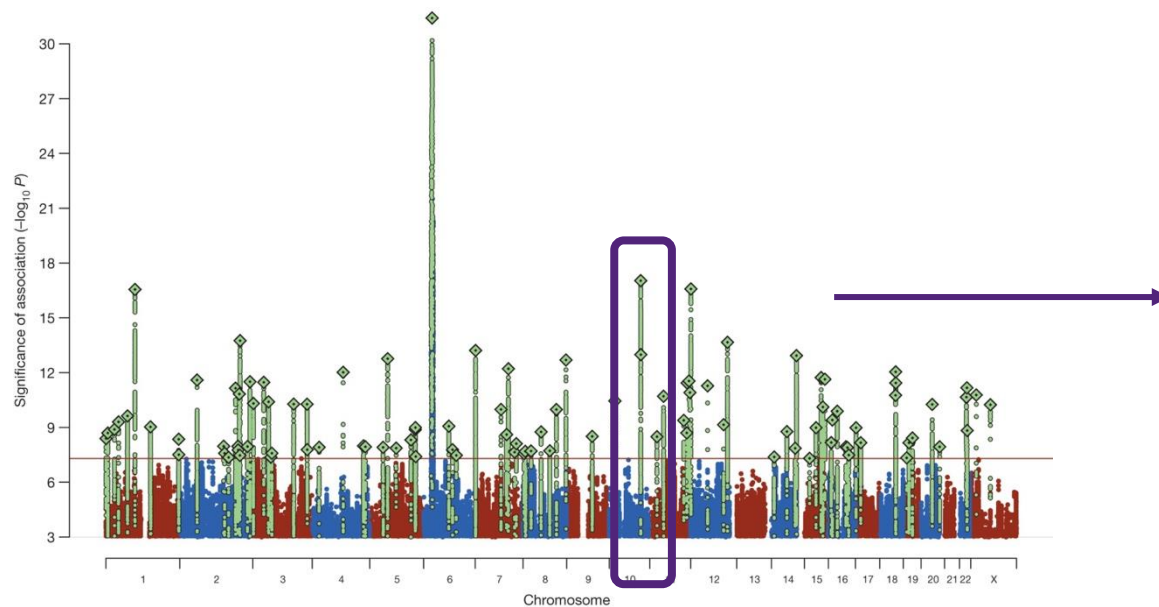
Slides can be downloaded from the website:

<https://cnsgenomics.com/data/teaching/GNGWS26/module1/>

What is fine-mapping?

An approach to identify and prioritise SNPs driving GWAS association signals

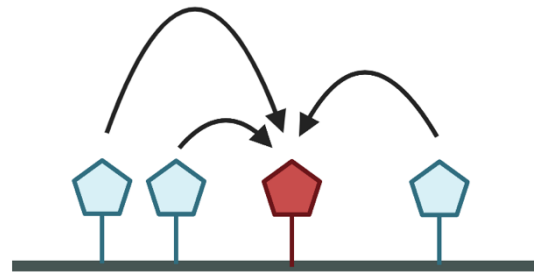
We want to understand the mechanisms by which genetic variants impact associated traits



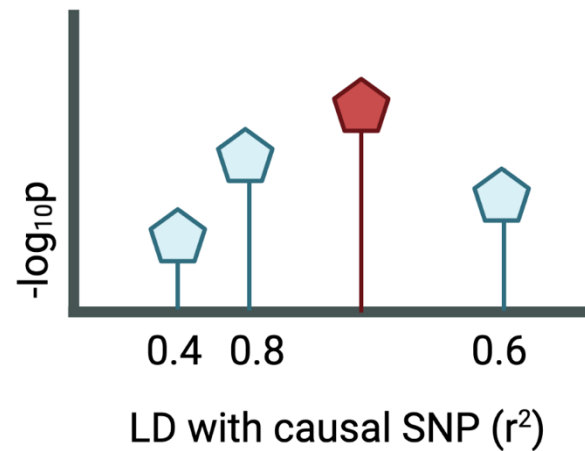
Why don't we take the top associated SNP?

Single causal variant

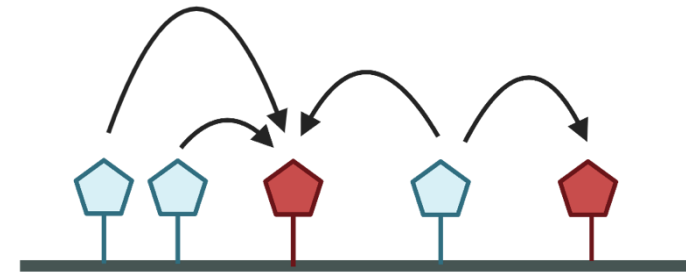
Red – causal variant



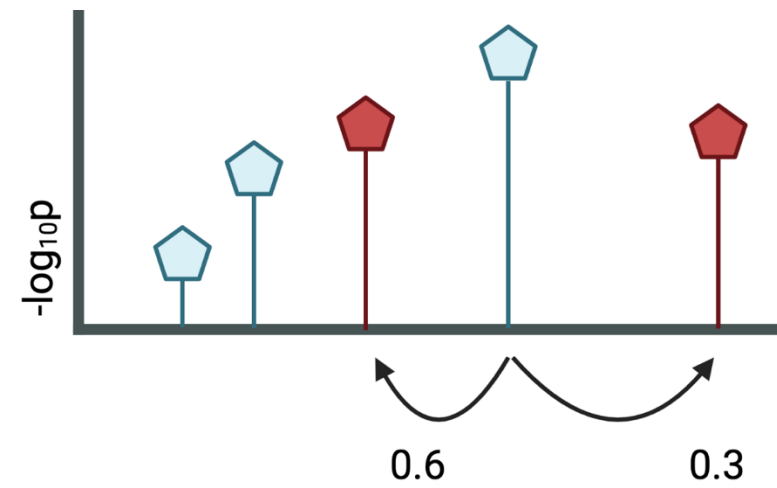
GWAS result



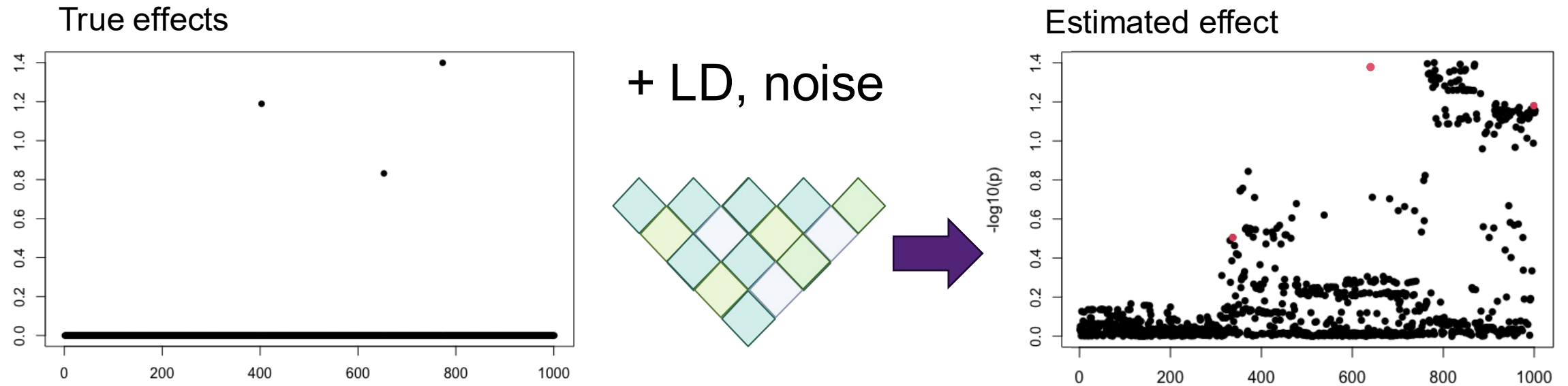
What about when there is ≥ 1 causal variant



GWAS result



Simplistic fine-mapping example



Fine mapping attempts to do the reverse..



Fine mapping assumes the causal variant has been measured!

- This is a big assumption!
- For example, FinnGen provides publicly available fine-mapping results
 - For ALS they report two credible sets.
 - The top variant in one of the sets had a PIP of 1.00 – that is equivalent to the fine mapping method reporting 100% certainty this is the causal variant.
 - However, it is unlikely causal - there is a C9orf72 repeat expansion that is in strong LD with the SNP

Motor neuron disease

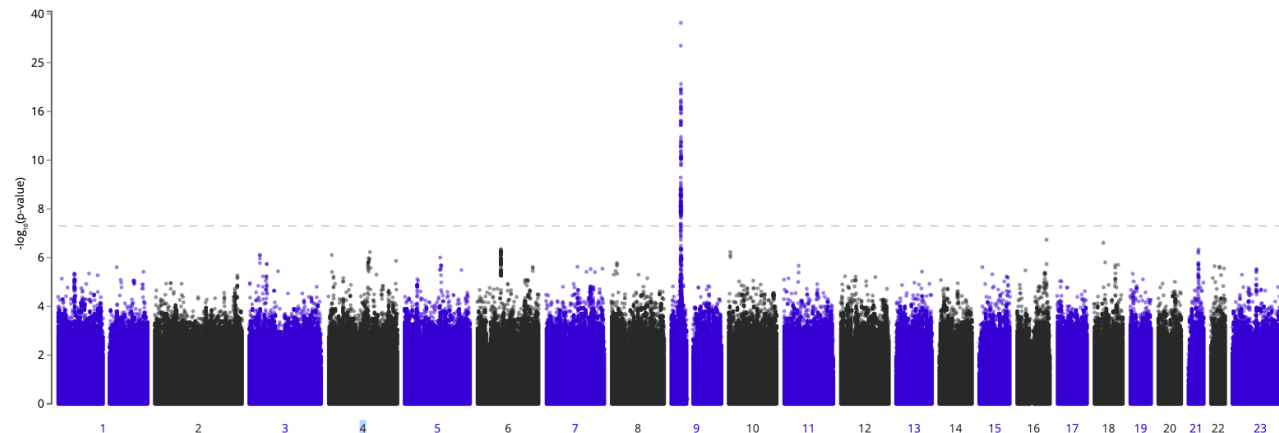
Diseases of the nervous system (G6_)

RISKEYS

431 cases

159840 controls

Phenotype not found in UKBB results



https://r8.finnngen.fi/pheno/G6_ALS

Many fine-mapping methods

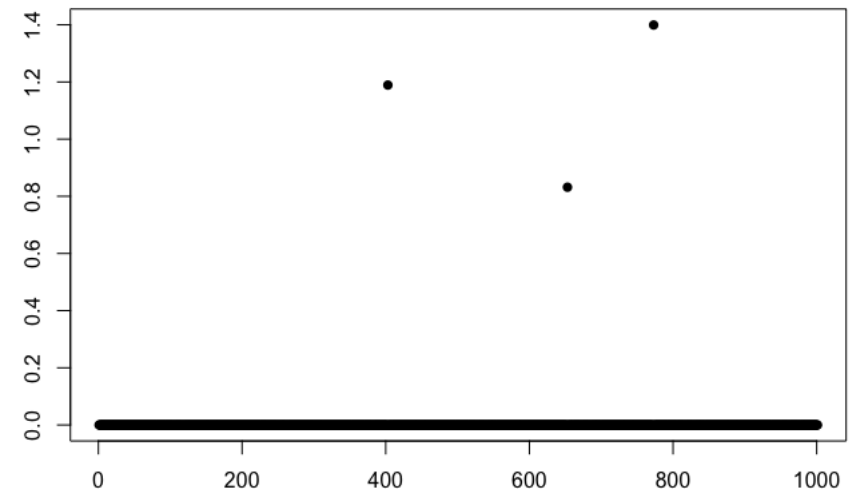
Majority of methods follow a Bayesian framework

- prior \times data = posterior

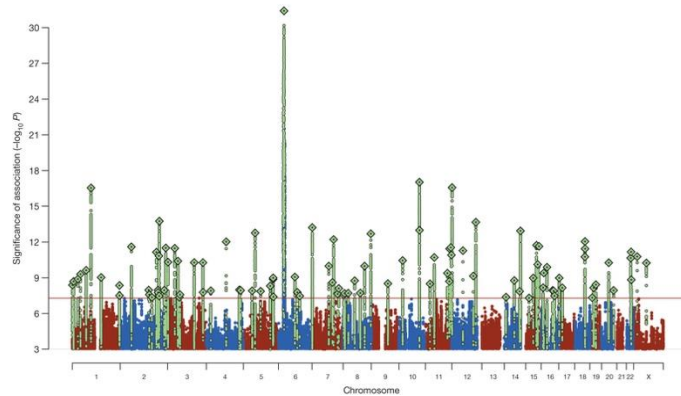
E.g. Prior knowledge of distribution of true SNP effects

Methods

- BIMBAM (Servin and Stephens, 2007)
- CAVIAR (Hormozdiari, 2014)
- PAINTOR (Kichaev, 2014)
- CAVIARBF (Chen, 2015)
- FINEMAP (Benner, 2016)
- SuSiE (Wang, 2020)
- Susie-inf & FINEMAP-inf (Cui, 2024)
- SBayesRC (Wu, 2025 medRxiv)

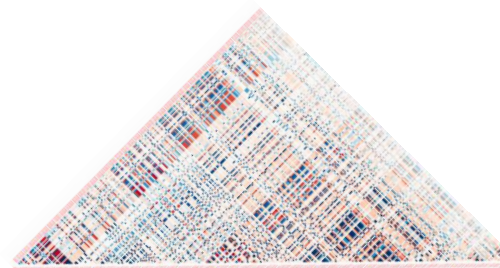
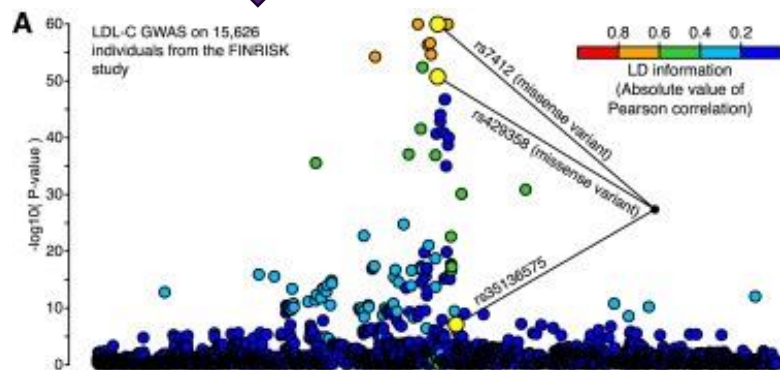


How does this work?



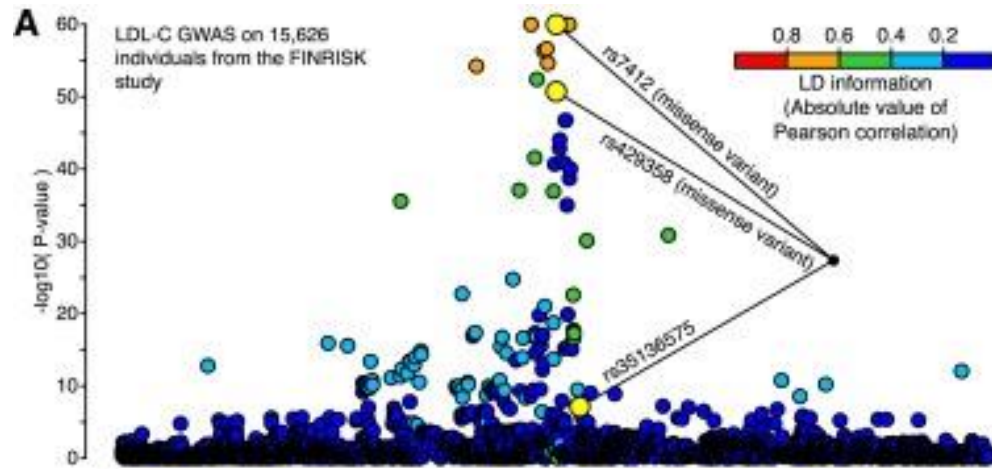
Prior

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



Probability of causality for each SNP

Posterior inclusion probability (PIP)

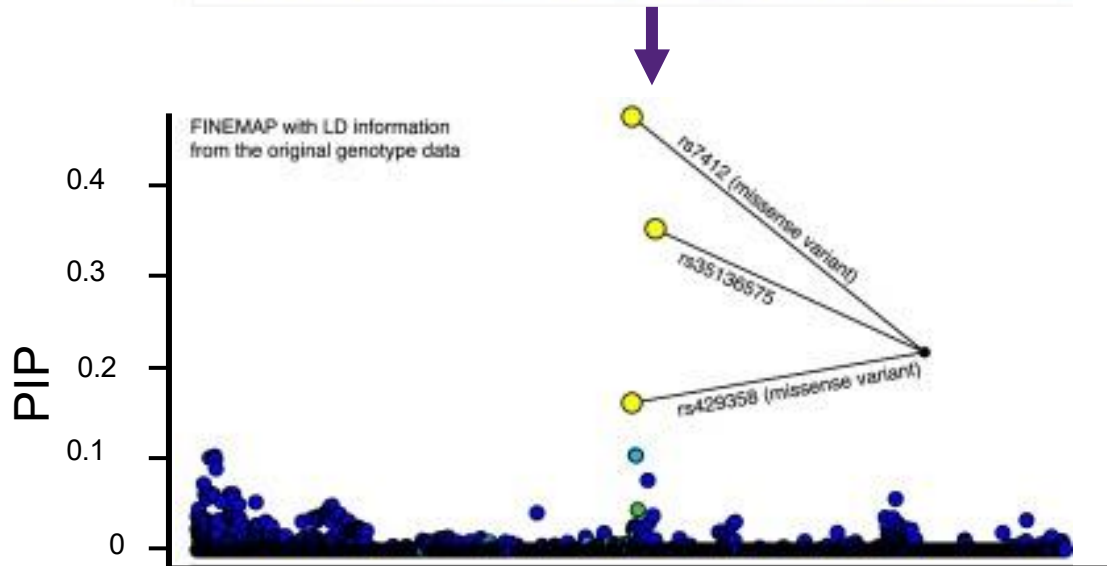


Probability (according to the model) the variant is casual

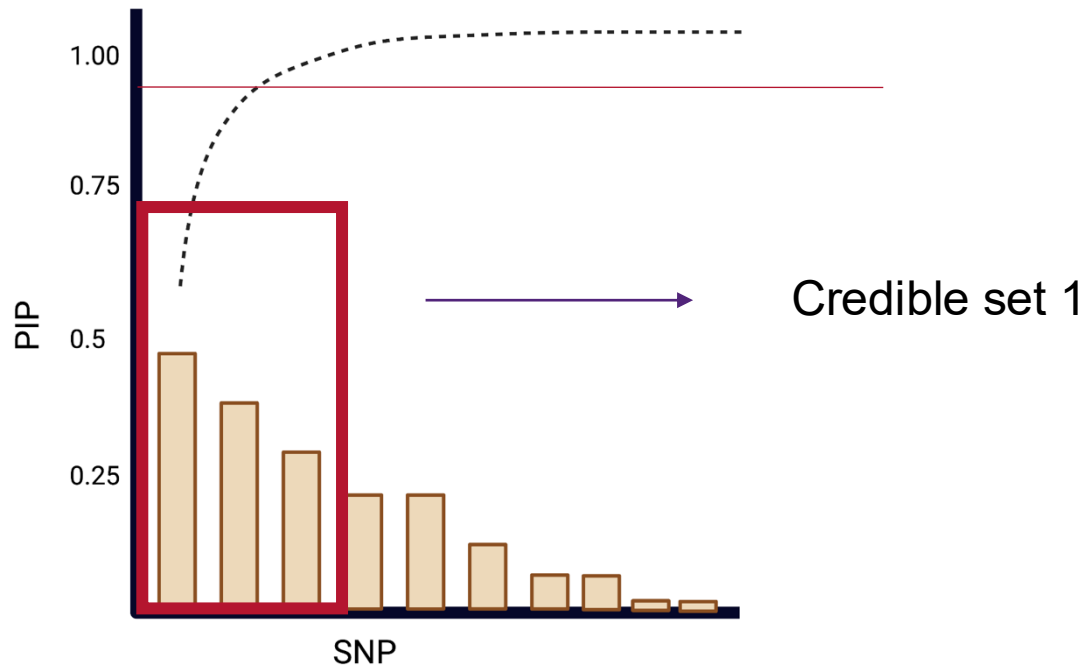
$$PIP_i = \Pr(b_j \neq 0 | X, Y)$$

↑ PIP = more confidence

↓ PIP = less likely to be driving signal



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

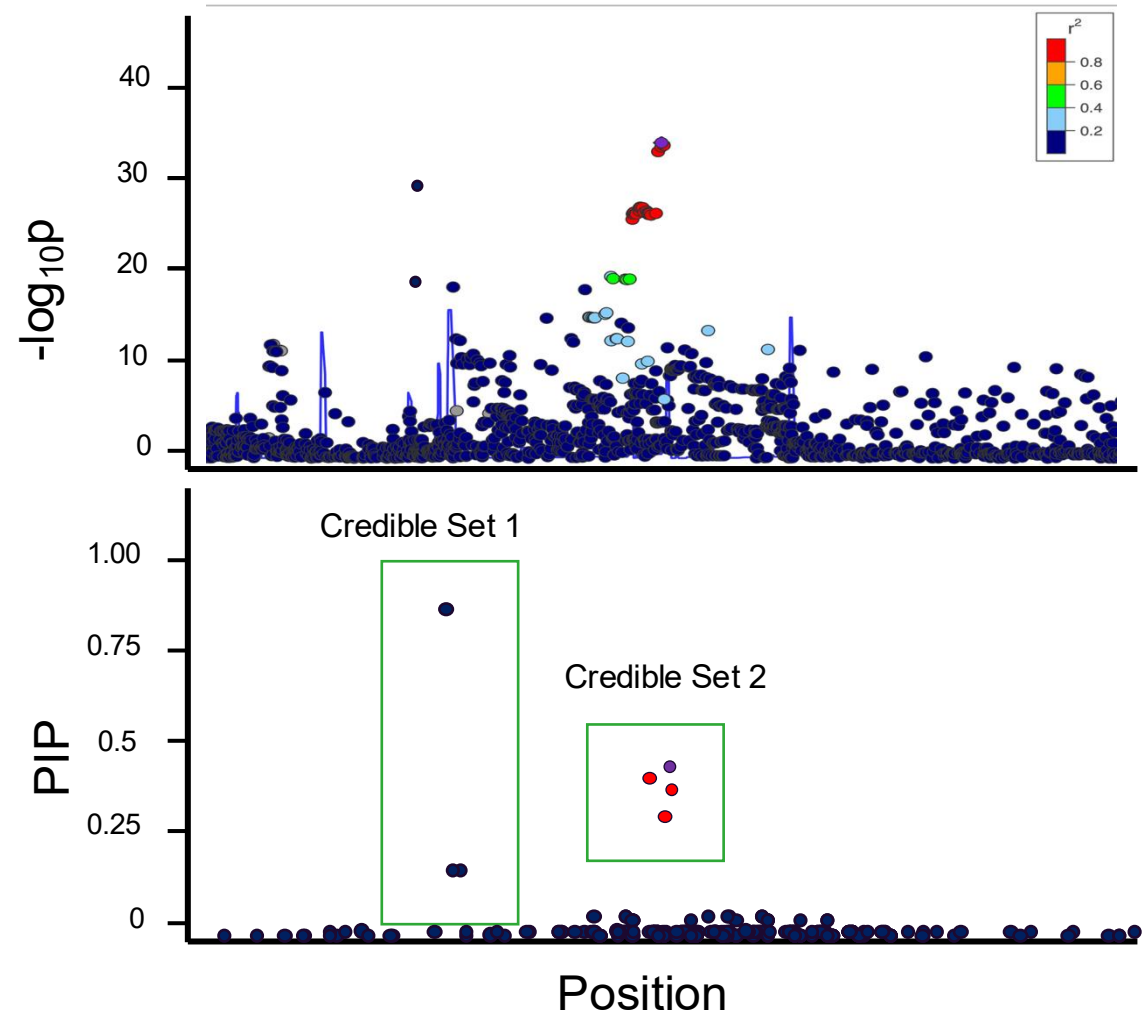
Some assuming **infinitesimal model** - a small number of larger effects, large number of infinitesimal effects (e.g. Susie-INF)

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



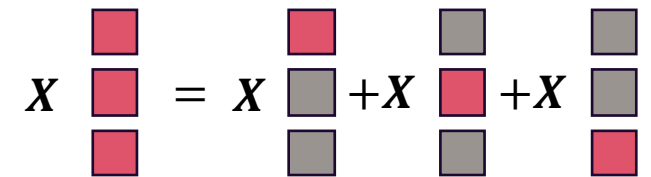
Sum of Single Effects (SuSiE)

Method for fine-mapping multiple causal variants

For each causal variant $\mathbf{b}_i = (b_{i1}, \dots, b_{iJ})$ **single effect vector**

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

$$\mathbf{b} = \sum \mathbf{b}_i$$



$$X \begin{matrix} \blacksquare \\ \blacksquare \\ \blacksquare \end{matrix} = X \begin{matrix} \blacksquare \\ \blacksquare \\ \blacksquare \end{matrix} + X \begin{matrix} \blacksquare \\ \blacksquare \\ \blacksquare \end{matrix} + X \begin{matrix} \blacksquare \\ \blacksquare \\ \blacksquare \end{matrix}$$

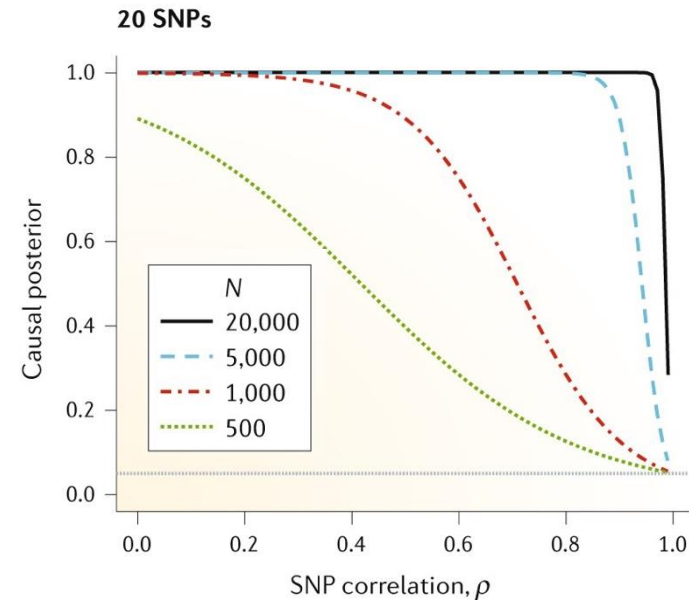
Iterative Bayesian stepwise selection

- Can quantify uncertainty in variables selected

Outputs 95% credible sets with PIPs for each SNP

Factors influence fine-mapping performance

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



Example:

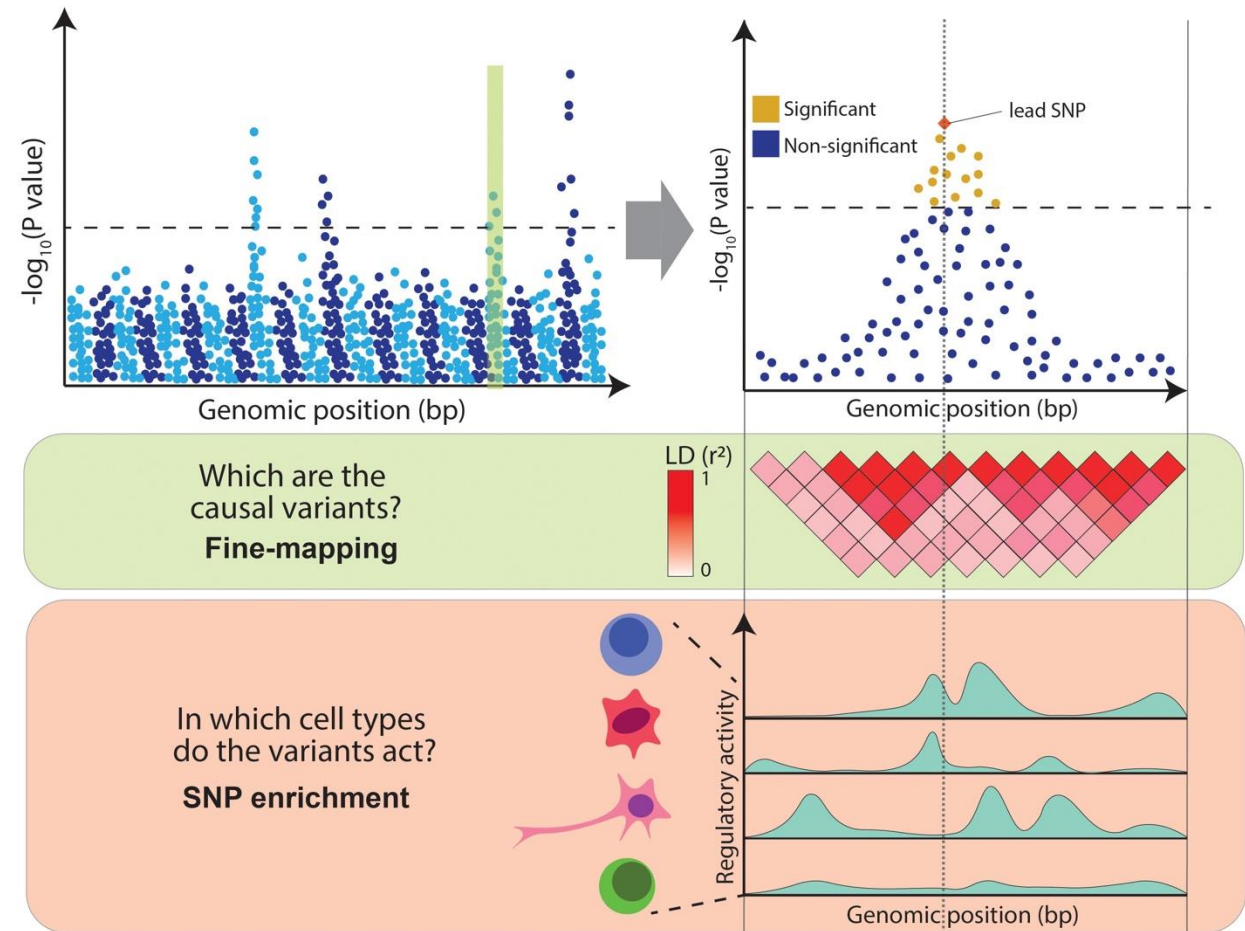
- 20 SNPs
- All SNPs have equal LD (x-axis)
- One causal SNP (R^2 1%)
- PIP of causal SNP on yaxis

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: SBayesRC, PolyFun, PAINTOR, fastPAINTOR, CAVIARBF



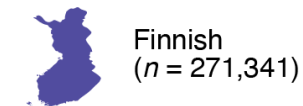
Cross-ancestry fine-mapping

- 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

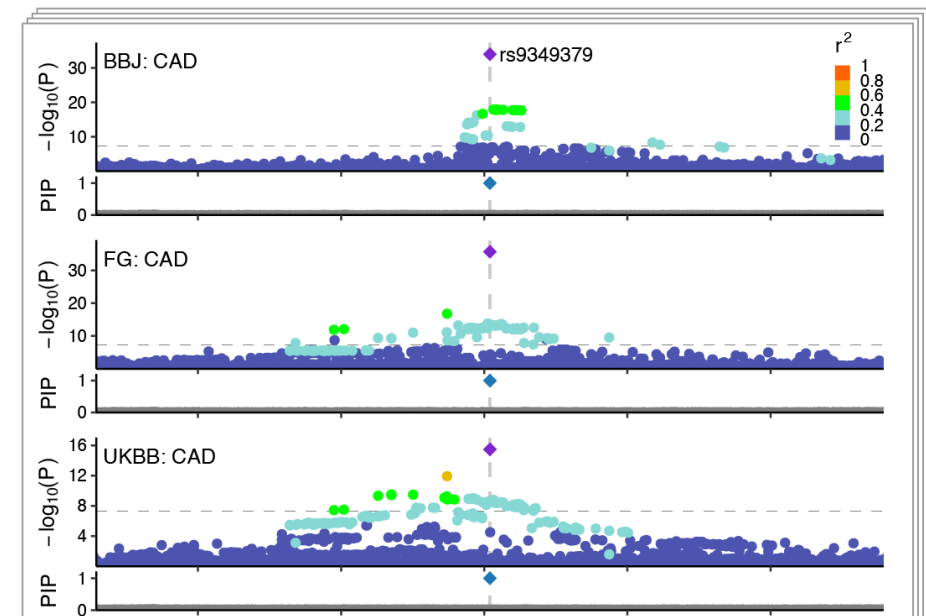
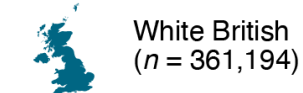
Biobank Japan (BBJ)
79 complex traits



FinnGen (FG) release 6
67 complex traits



UK Biobank (UKBB)
119 complex traits



Use caution when applying fine-mapping

- Beware of reference LD, follow [Weissbrod et al. 2020 NG guidelines](#).
 - Recommend using in-sample LD from the GWAS target sample
- Meta-analysis fine-mapping is tricky, see [Kanai et al. 2022 Cell Genomics](#).
- Don't forget to use covariate-adjusted LD when the cohort has more complex population structure, e.g. admixture. See [Pan-UKBB LD documents](#).
- Keep in mind that model misspecifications and missing causal variants exist in real data applications. Use caution when interpreting fine-mapping results.
- Run different methods if you can.

Fine-mapping resources

Published fine-mapping results by large studies:

FinnGen: <https://finngen.gitbook.io/documentation/methods/finemapping>

UK Biobank: <https://www.finucanelab.org/data>

PGC Schizophrenia study: [Supplementary Table 11](#)

LD resources:

PolyFun published UK Biobank LD matrices.

Pan-UKBB published multi ancestry LD matrices.

Fine-mapping pipelines:

FinnGen pipeline: <https://github.com/FINNGEN/finemapping-pipeline>

UK Biobank pipeline: <https://github.com/mkanai/finemapping-pipeline>

Practical

- Use SuSiE to investigate the genetic effects underlying height.
- We will use summary statistics from Wood et al (2014) and perform fine-mapping in the regions surrounding some of the GWAS 'hits' to identify SNPs that are most likely to be causal (or in LD with causal variants).
- The practical html can be found on the website
https://cnsgenomics.com/data/teaching/GNGWS26/module1/9_FineMappingPrac.html
- The data can be found on the cluster
[/data/module1/9_FineMappingPrac](#)