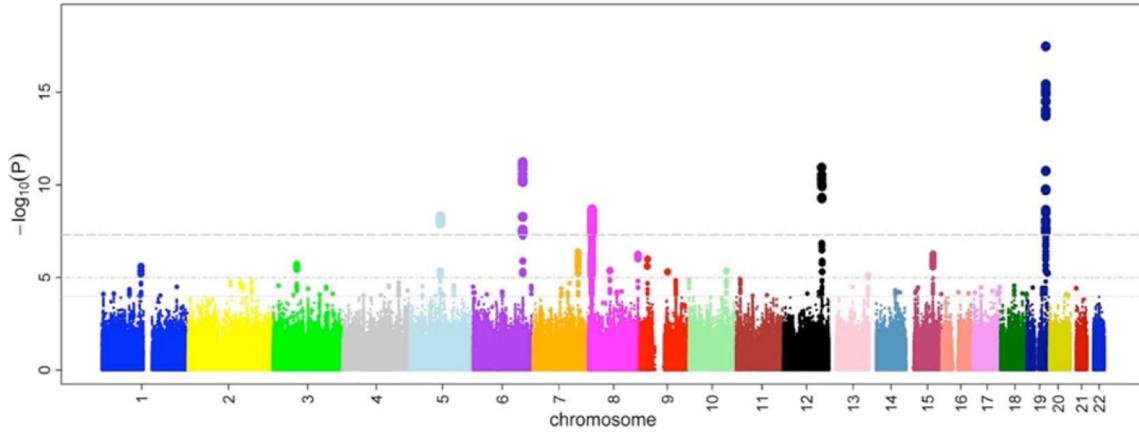


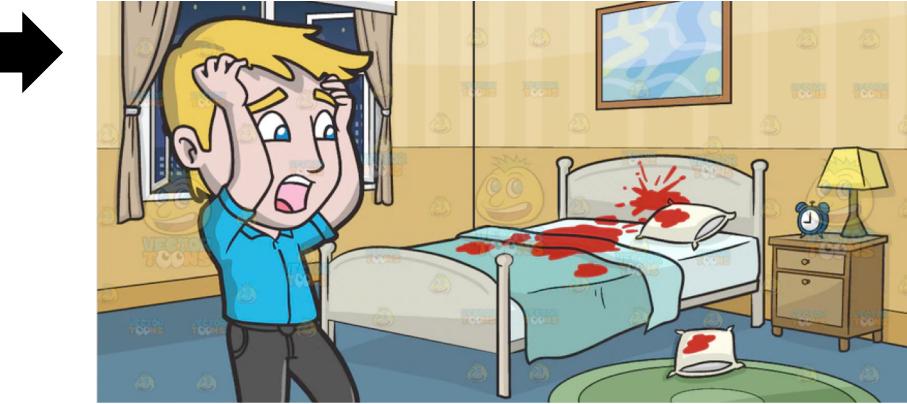
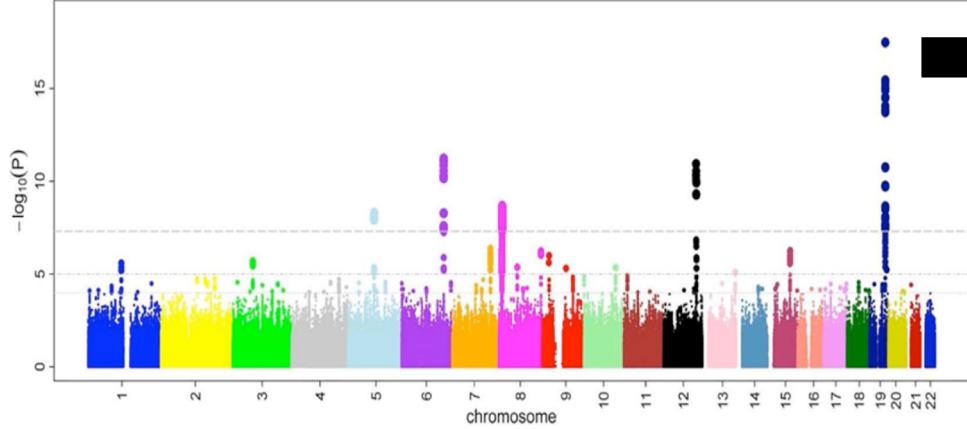
Introduction to Fine-Mapping

Ananyo.Choudhury@wits.ac.za

Associations



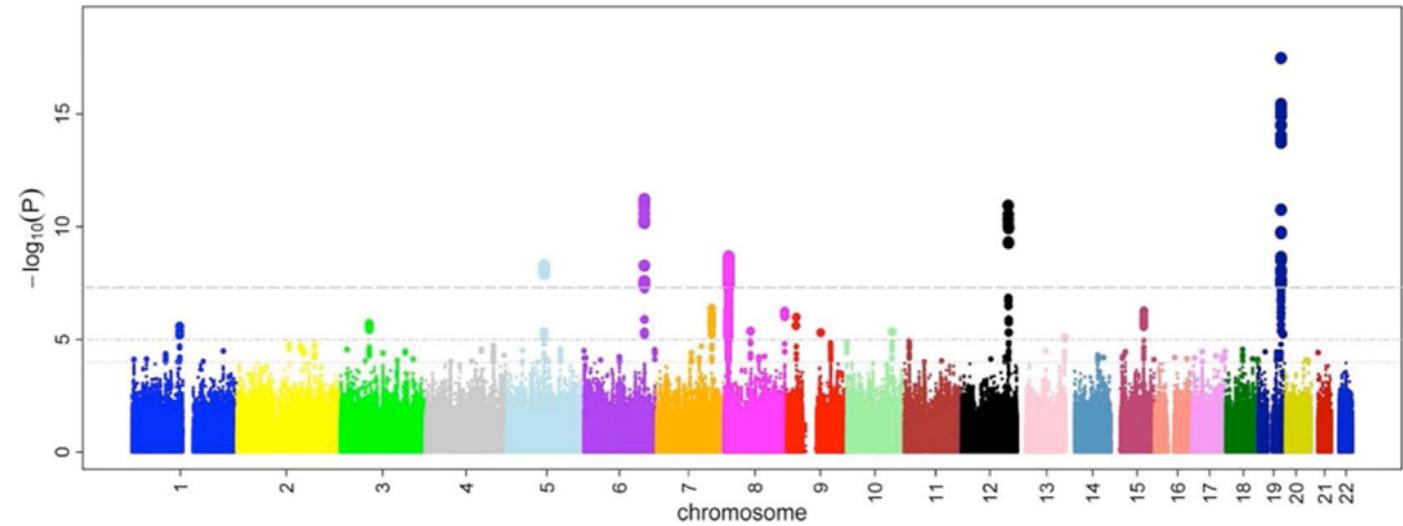
Analogy



Why so?



Lead SNP is Causal SNP?
(1,000 cases and 1,000 controls)
79% chance at OR=1.5 ; RAF= 0.50
2.4% chance at OR=1.1 and RAF=0.05

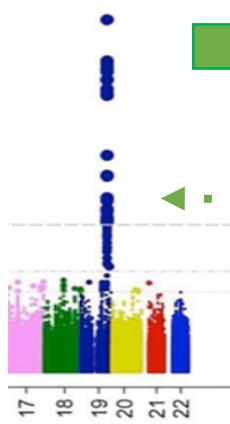


- The top association is often not “causal”.
- Not all SNPs are genotyped (or imputed)
- Causal associations for complex traits – **modest effect size** and therefore often not the lead

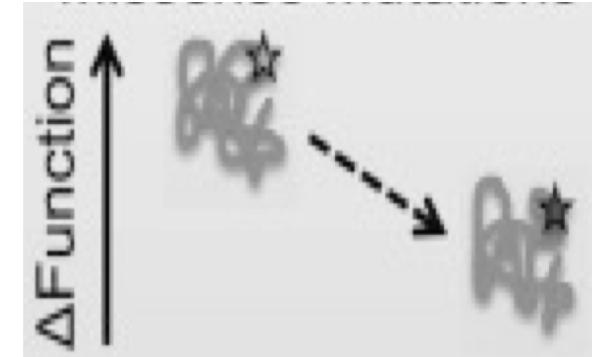
Analogy - continued



From Crime Scene to Criminal



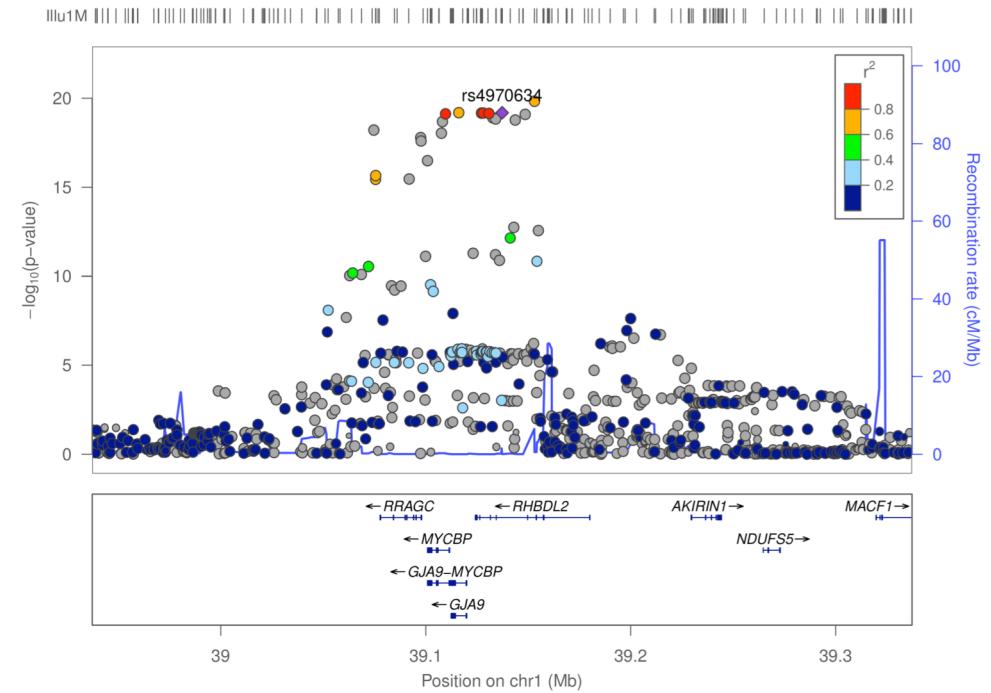
From Index SNP(s) to causal variant(s)



Understanding the scene (analogy continues)



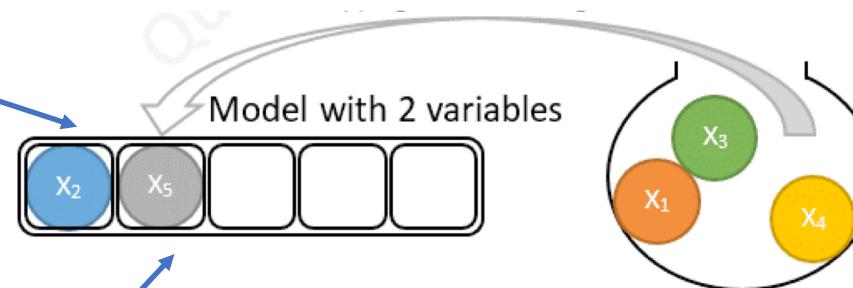
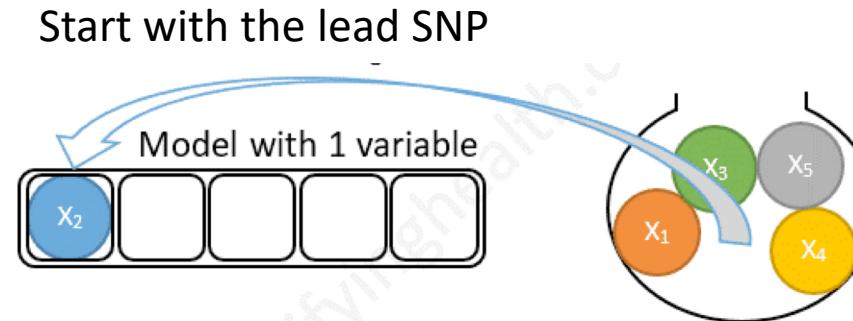
How many independent crimes?



- Multiple SNPs in a region are often associated with a trait due to correlation of non-causal SNPs with a single causal SNP.
- The LD make each SNP appear to be associated with a trait when analysing one SNP at a time.

Partition to independent regions

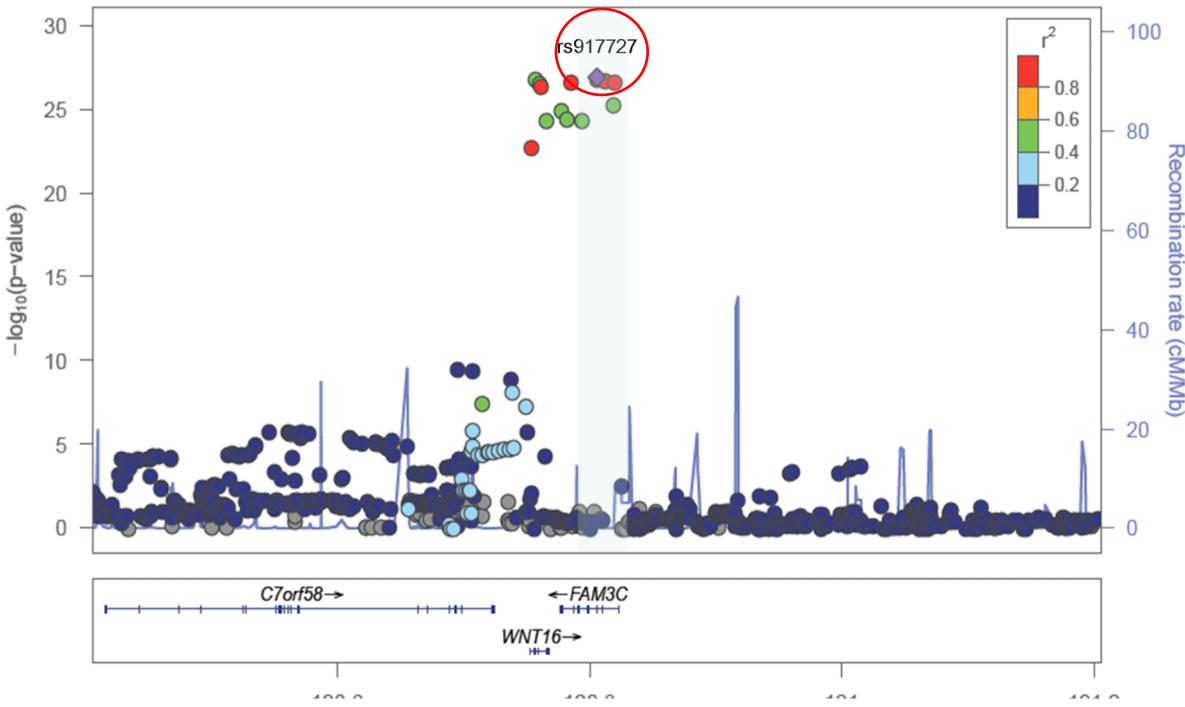
Conditioning on the lead SNP from a GWAS by treating it as an a covariate in a regression model and testing the remaining SNPs in the region of interest



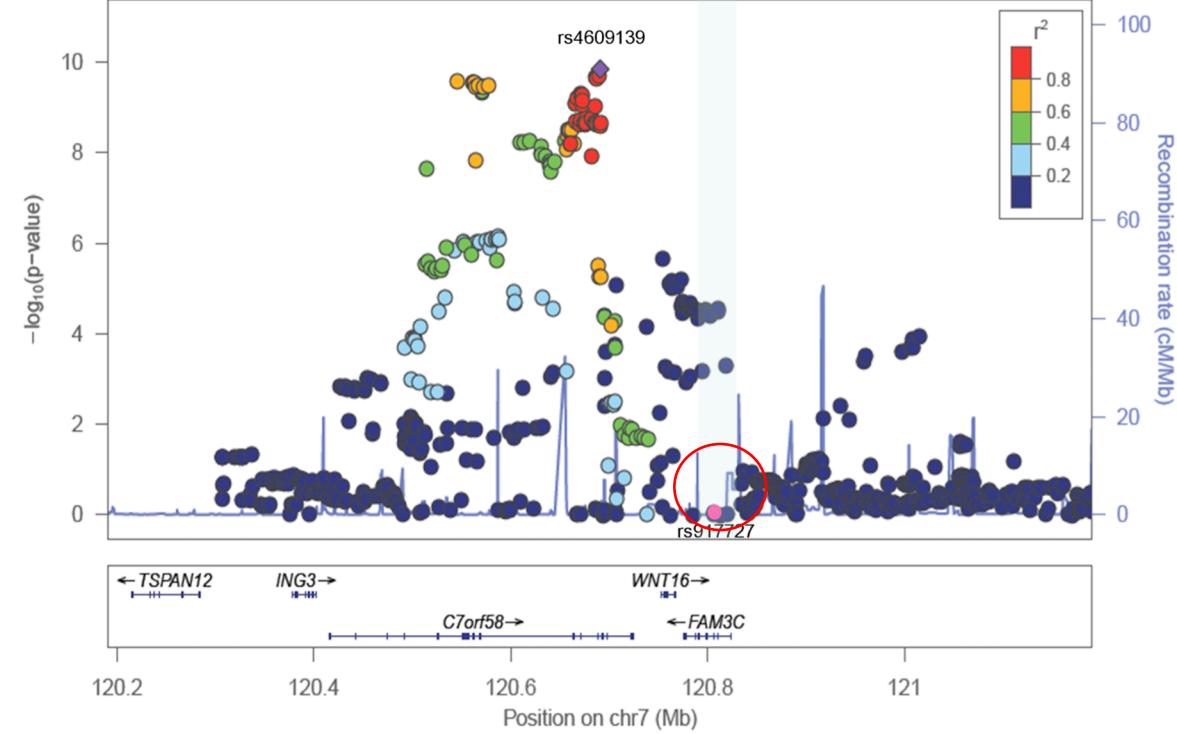
If a second SNP is found significant after conditioning – also add it as covariate and look for a third association and continue..
If not just report one independent signal.

Example

A



B

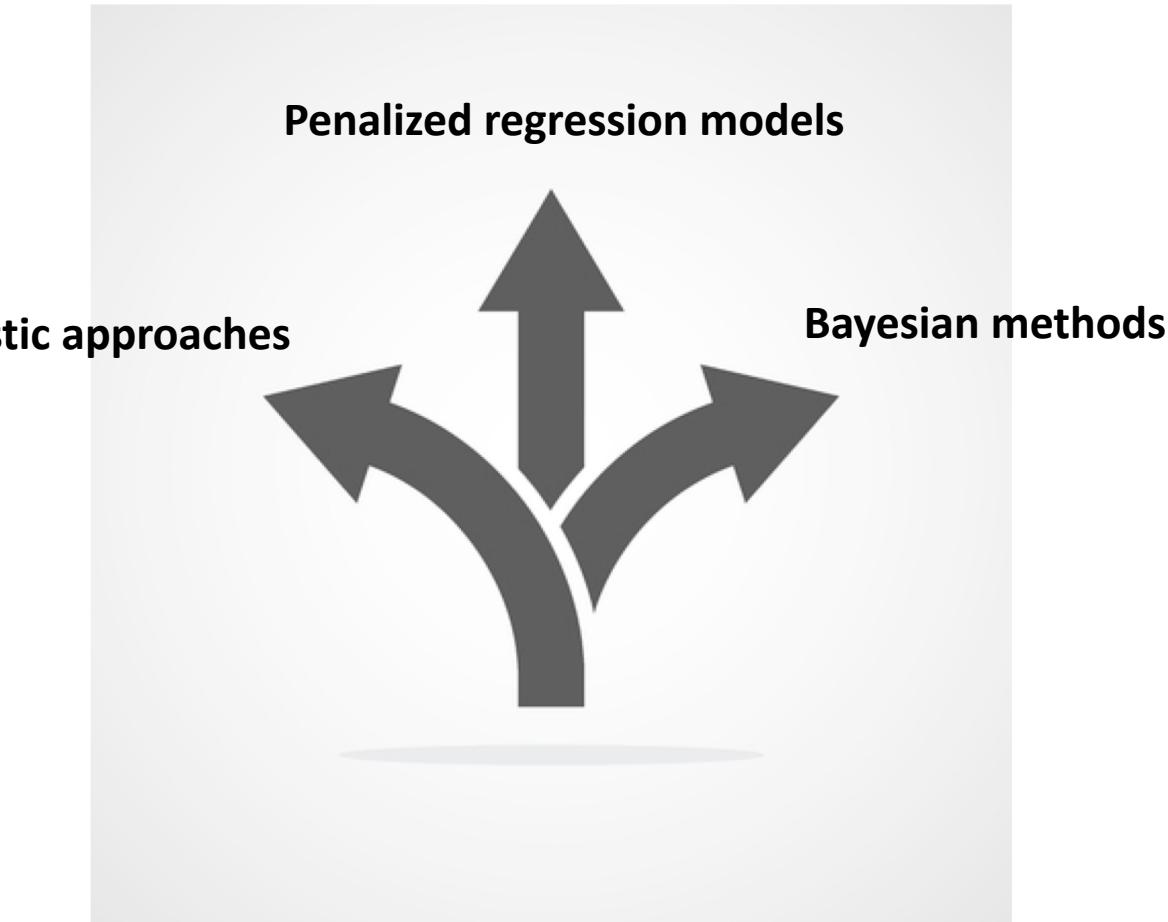


Conditional Analysis- Cautions and Challenges

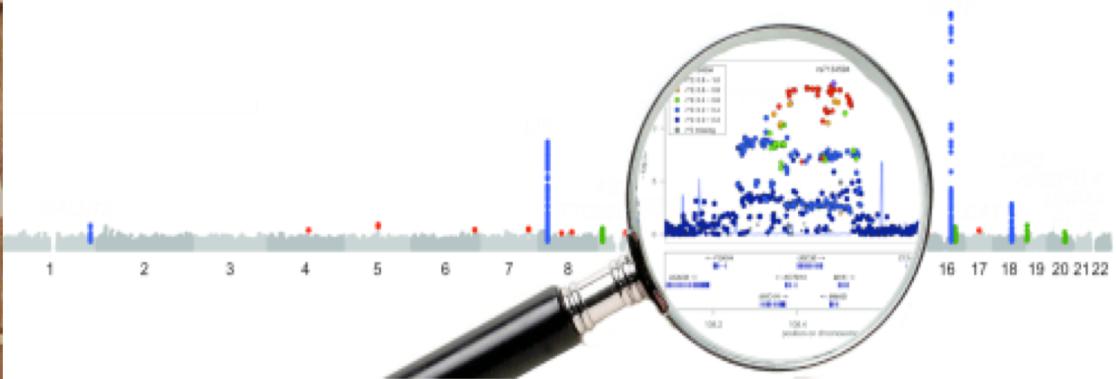
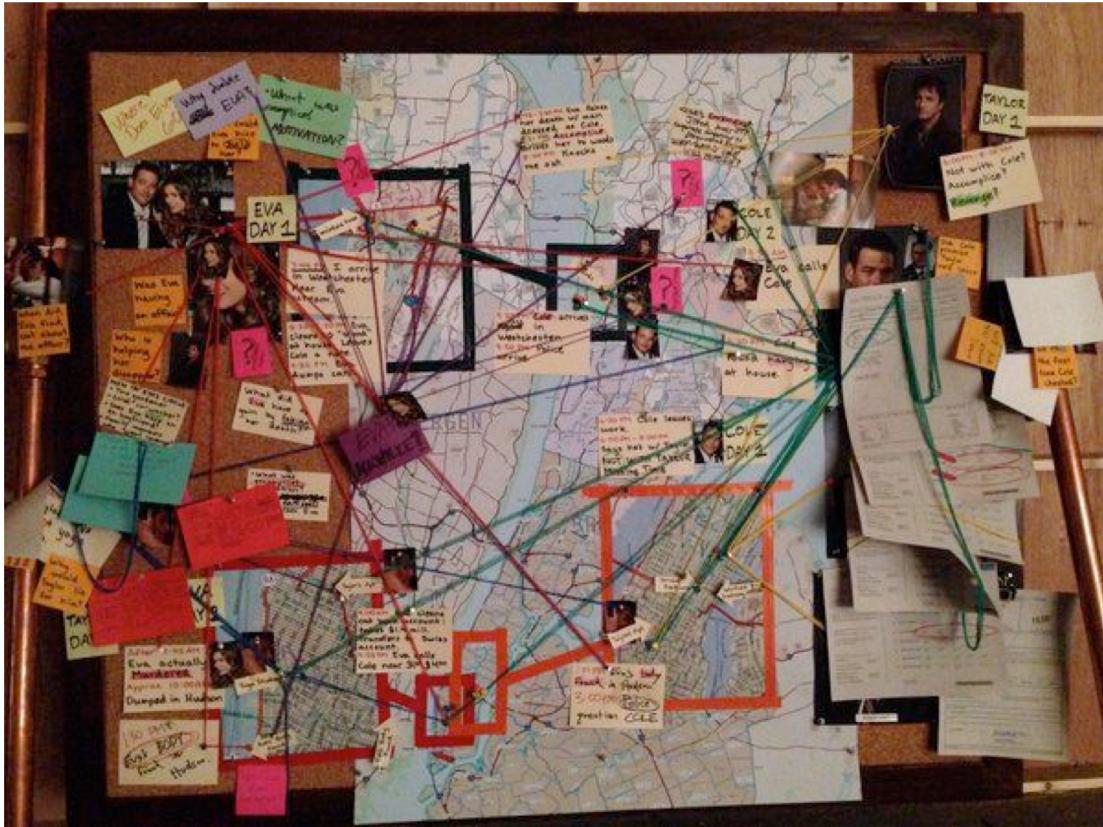
- Higher the **number of steps**, more statistical tests greater the **chance of a false positive result**.
- Power **diminishes** as the correlation of a primary SNP and a secondary SNP increases, and **when the effect size** of a secondary SNP is weaker than that for a primary SNP.
- Low power can cause **missed secondary associations**
- informative but **fails to provide probabilistic measures of causality** for individual variants.



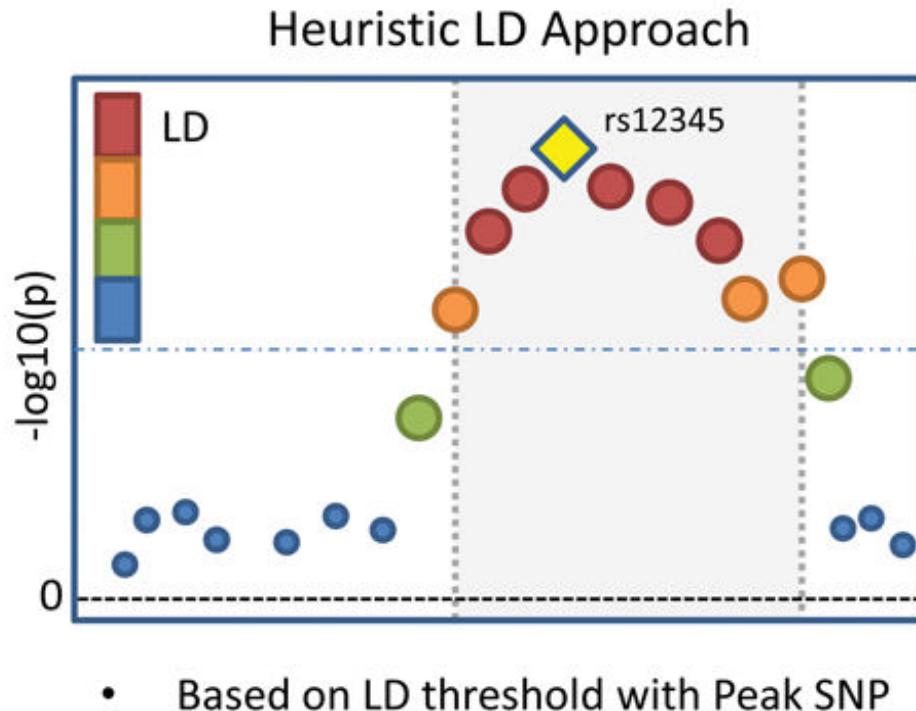
Fine-mapping



Heuristic fine-mapping -Using LD



Heuristic fine-mapping approaches



- Filter SNPs according to (r^2) with the lead SNP and retain as potentially causal only those SNPs with an r^2 above a threshold.
- Hierarchical clustering of all SNPs in a region based on their pairwise r^2 to create clusters .
- Based on pairwise LD among SNPs infer haplotypes, using software such as Haploview

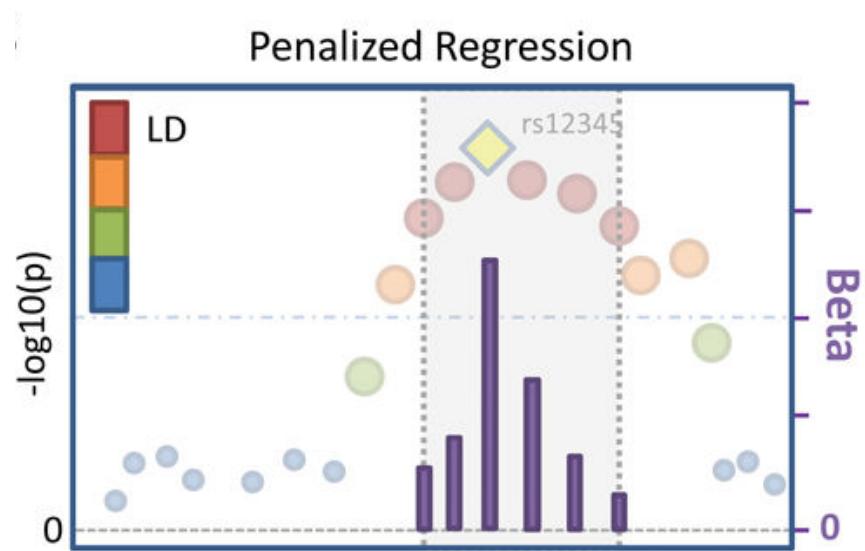


Challenges with LD based approaches

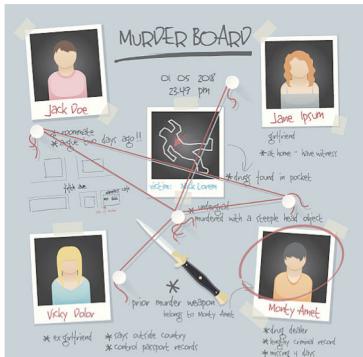
- Evaluating one SNP at a time can be misleading due to the complex patterns of LD
- Do not account for the joint effects of the SNPs on the trait
- Factors beyond recombination influence LD
- No objective measure of the confidence that a SNP is causal
- Arbitrary thresholds and subjective interpretations



Penalized regression models



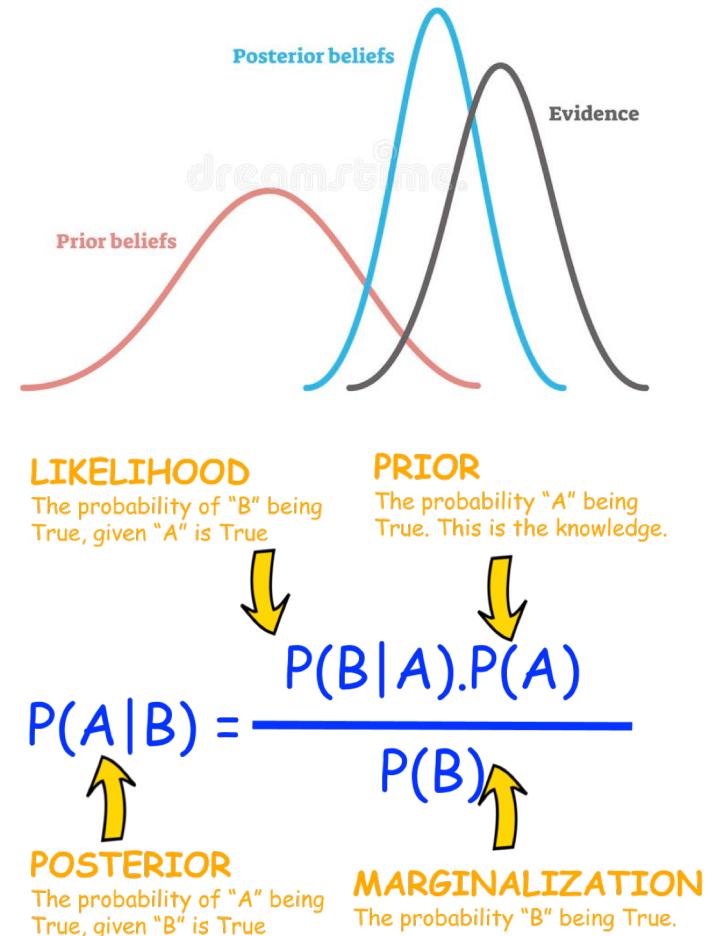
- Based on all SNPs with non-zero betas



- Perform estimation of SNP effect sizes and SNP selection simultaneously into a model by shrinking small effect estimates toward zero.
- Tuning parameters to select SNPs into a model
- Produces sparse models
- Includes non-causal SNPs and excludes a causal SNP when they are highly correlated.

Bayesian methods

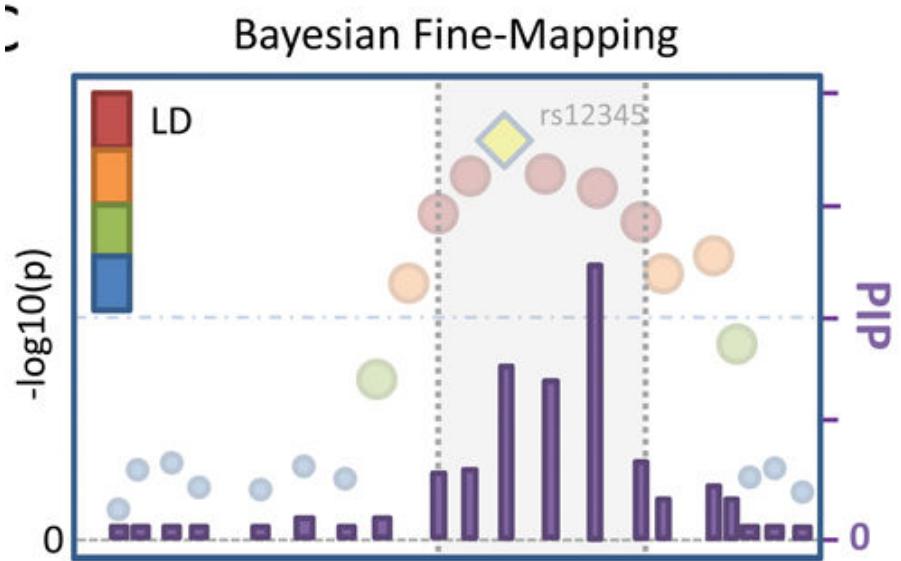
- Given m significant SNPs in a GWAS – which SNPs are *causal*?
- Select the set of SNPs that we *hypothesise* are causal (eg SNP 3, 33, 34, 37, 421 are causal)
- each SNP, can have two values - 1 for causal and 0 for not.
- Given m SNPs, there are 2^m possible models
- $P(M|D)$ – probability of the model given the data we see
- Find the M with greatest probability



PIP and Credible Set

- The PIP for a SNP is the probability of including a SNP as causal in any of the models. For SNP i , the PIP is computed by the sum of the posteriors over all models that include SNP i as causal.

$$PIP_i = \sum \{P(M|D) : \text{SNP } i \text{ causal in } M\}$$

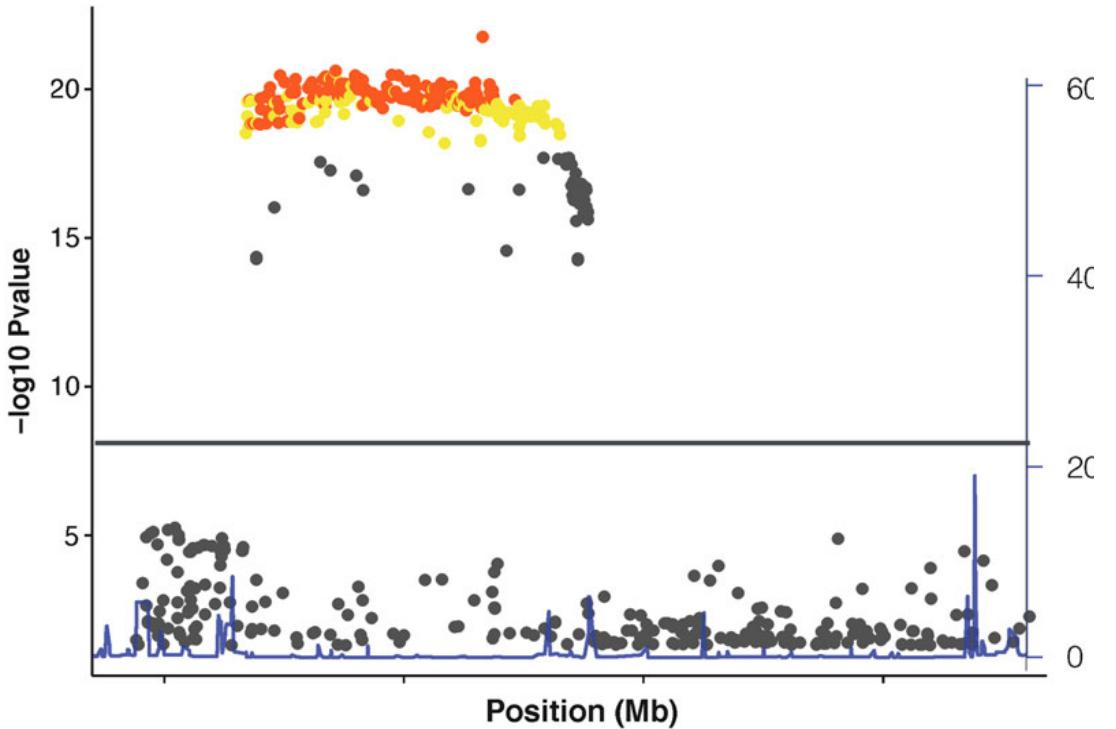


- Bayesian methods can also be used to determine the α credible set, the minimum set of SNPs that contains all causal SNPs with probability α .

- Credible set based on SNP PIPs
- Test and find associated SNPs
- Rank SNPs by PIP (descending)
- Pick the smallest number so that sum of PIPs = α

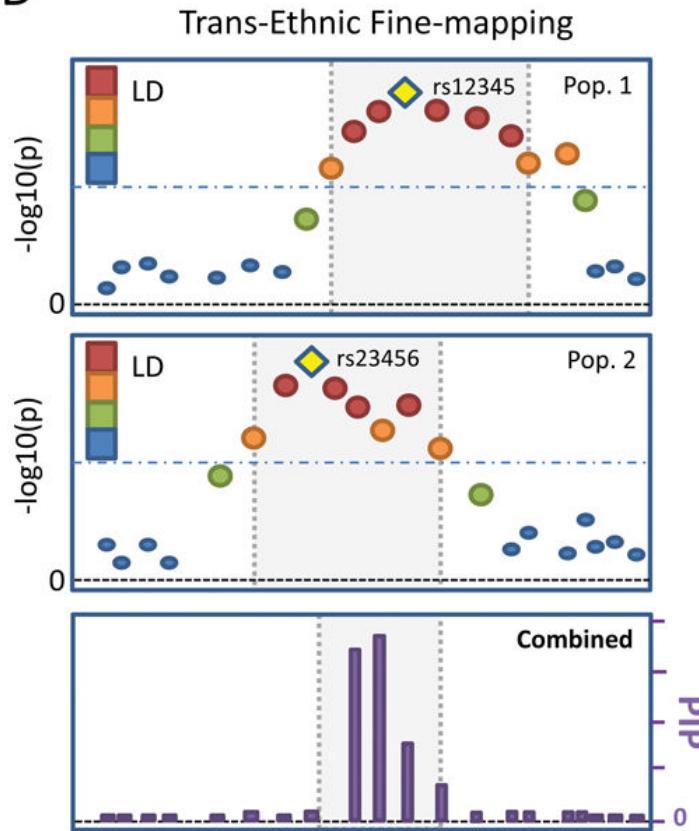
Advantages

- Allows for multiple causal SNPs in a region.
- Posterior probabilities for SNPs can be directly compared.
- Select fewer SNPs as potentially causative.
- Control for SNPs with large effects.

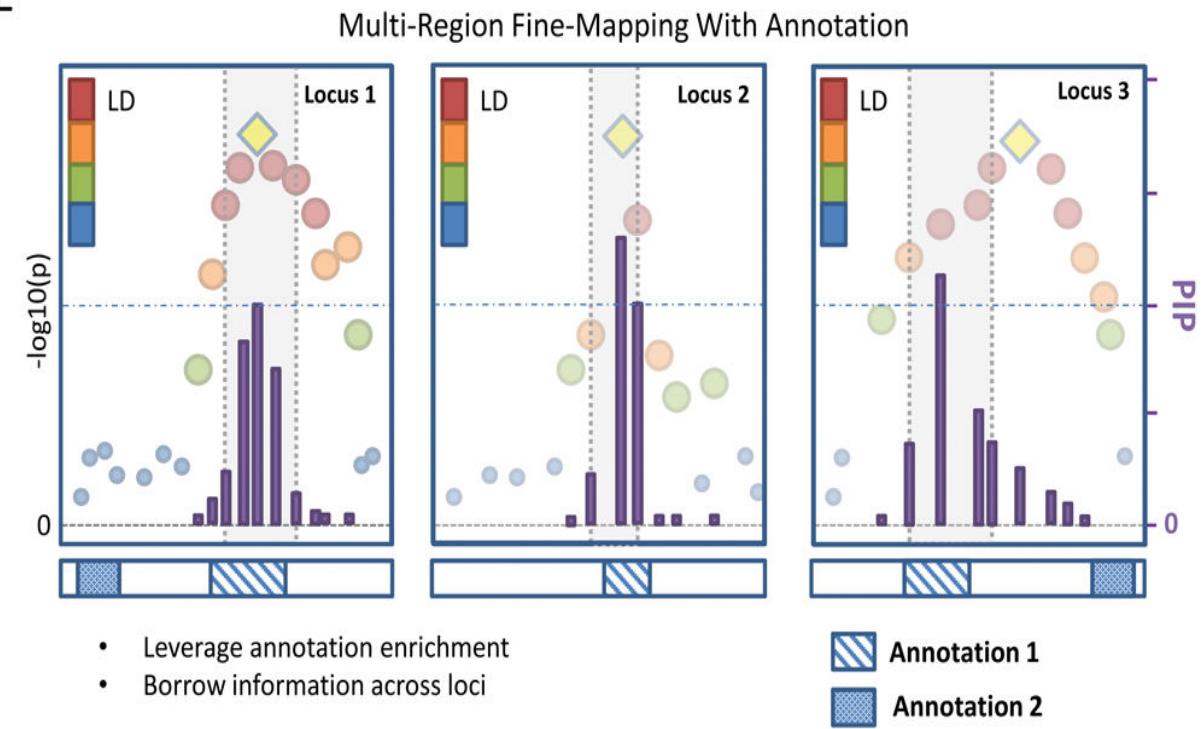


Other approaches to enhance fine mapping

D



E



- Leverage Ethnic Differences in LD at a given locus

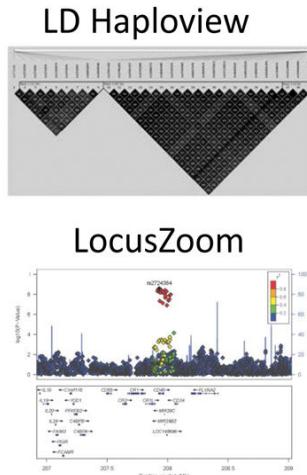
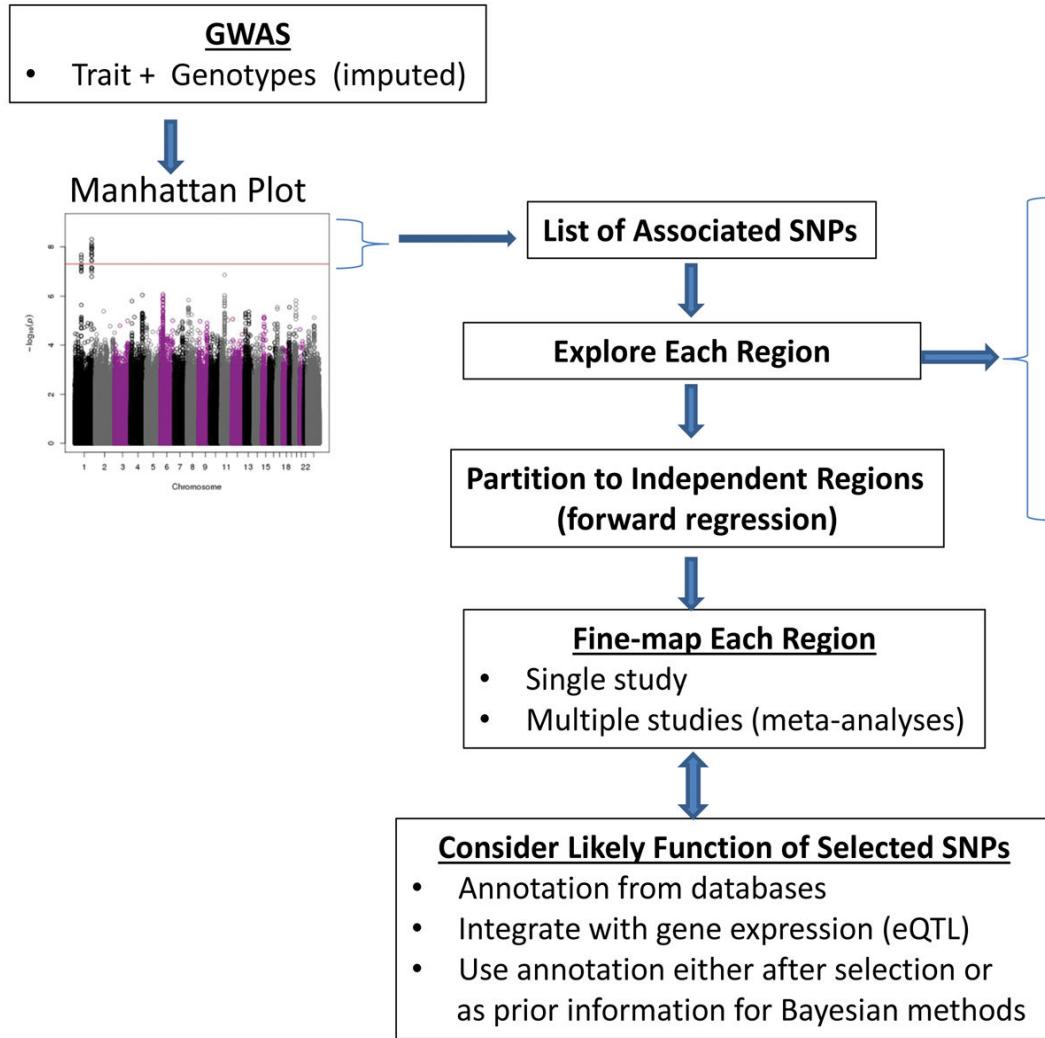
Software	Trait type ^a	Input covariates ^b	Uses summary statistics?	Maximum number of causal variants ^c	Input annotation?	Causal search
BIMBAM v1.0	qt/binary	No	No	Fixed	No	Exhaustive
mvBIMBAM v1.0.0 ^{115,116}	mqt	No	Yes	1	No	Exhaustive
SNPTTEST v2.5.4-beta3	qt/binary/mqt/multinomial	No	No	1	No	Exhaustive
piMASS v0.9	qt/binary	No	No	Computed	No	MCMC
BVS v4.12.1 ^{117,118,119}	binary	Yes	No	Computed	Yes	MCMC
FM-QTL	qt	No	No	Computed	Yes	MCMC
DAP v1.0.0	qt	Yes	Yes	1/Fixed/Computed	Yes	Exhaustive
Fine-mapping	multinomial	Yes	No	Computed	No	Greedy
Trinculo	multinomial	Yes	No	Computed	No	Greedy
BayesFM	binary	Yes	No	20	No	MCMC
ABF	qt/binary ^d	Yes	Yes	1	No	Exhaustive
fgwas v0.3.6	qt/binary ^d	No	Yes	1	Yes	Exhaustive
CAVIAR/eCAVIAR	qt/binary ^d	No	Yes	Fixed	No	Exhaustive
PAINTOR v3.0	qt/binary ^d /mqt	No	Yes	Fixed/Computed	Yes	Exhaustive/MCMC
CAVIARBF v0.2.1	qt/binary ^d	No	Yes	Fixed	Yes	Exhaustive
FINEMAP v1.1	qt/binary ^d	No	Yes	Fixed	No	Shotgun stochastic search
JAM in R2BGLiMS v0.1	qt/binary ^d	No	Yes	Fixed/Computed	No	Exhaustive/MCMC

Toolkit for Fine-Mapping



shutterstock.com · 570144070

Summary



Schaid et al. 2018. *Nature review genetics*
Broekema et al. 2019 *Open Biology*
Spain and Barrett 2015, *Human Molecular Genetics*

