

Statistical Fine-mapping

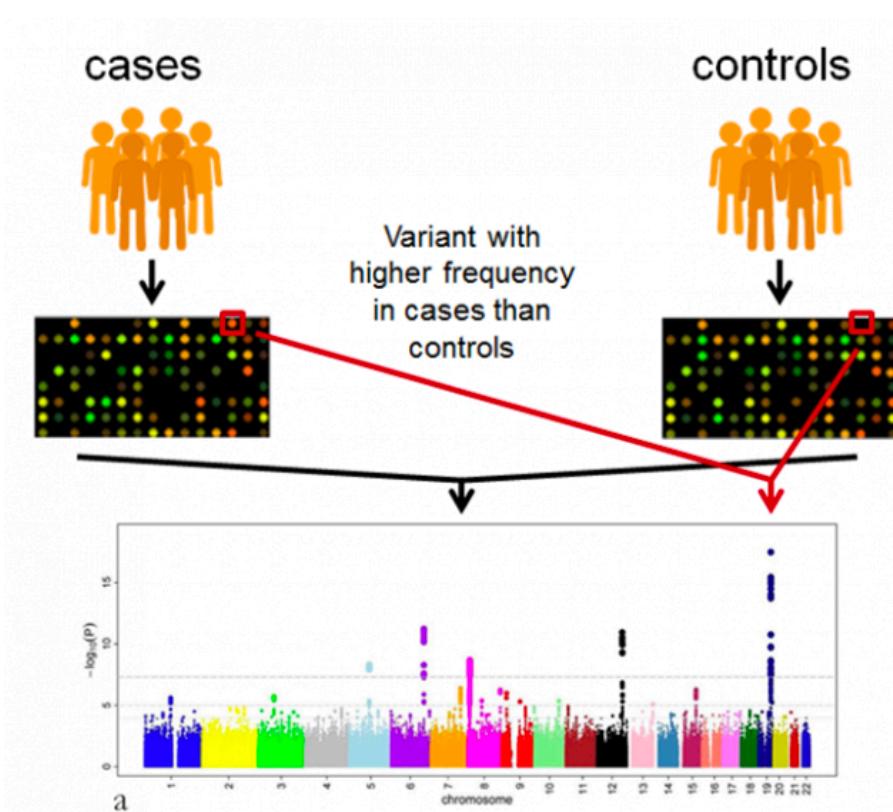
From genome-wide associations to candidate causal variants

GMS statistics course wrap-up session

Jiayuan Zhang

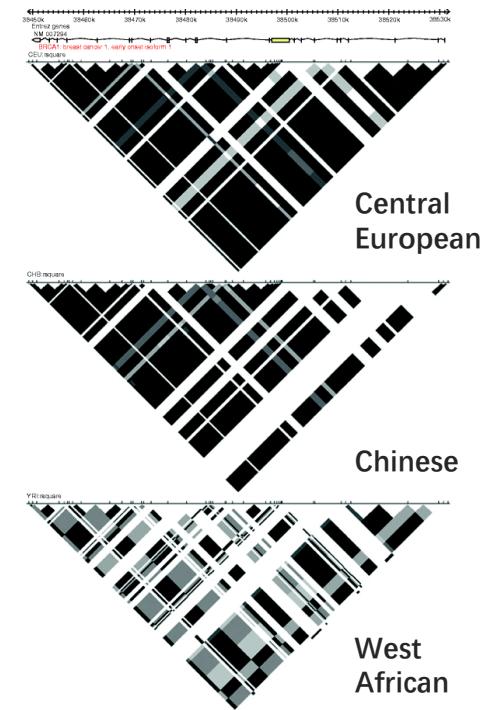
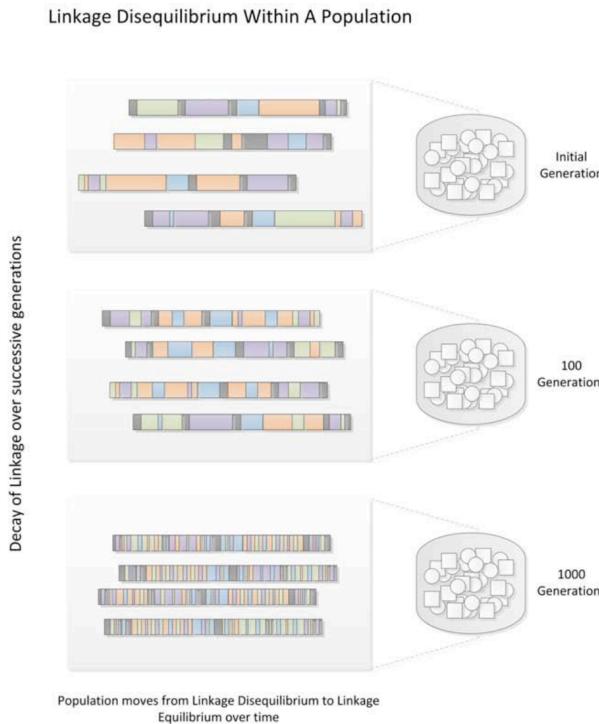
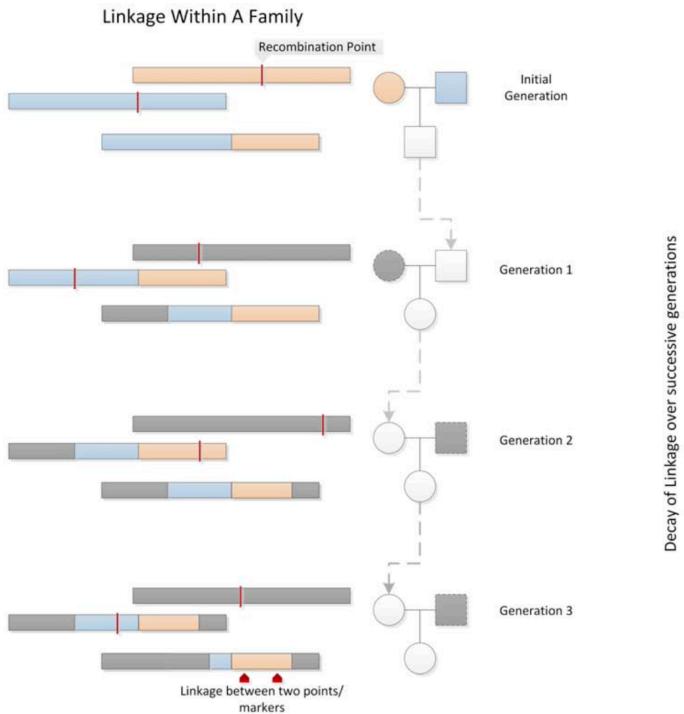
6 Dec 2019

Overview of GWAS



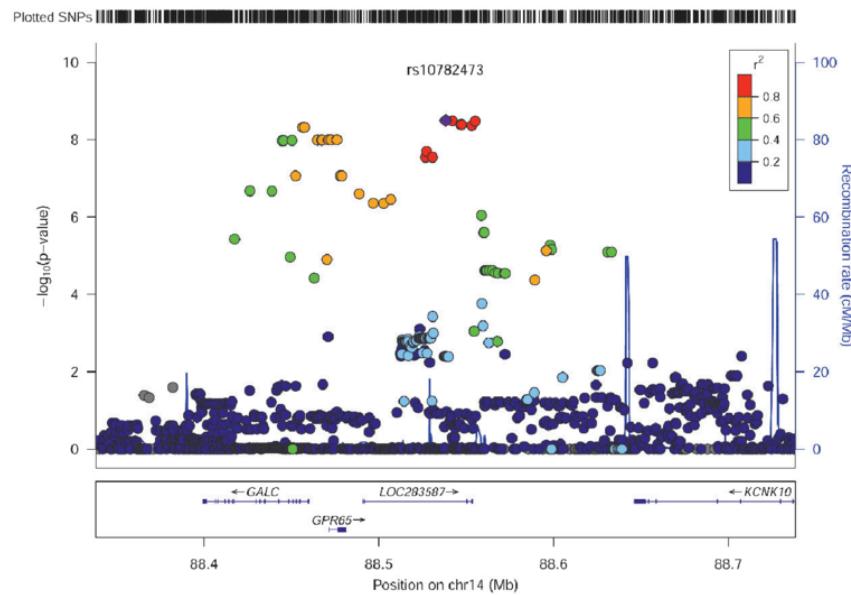
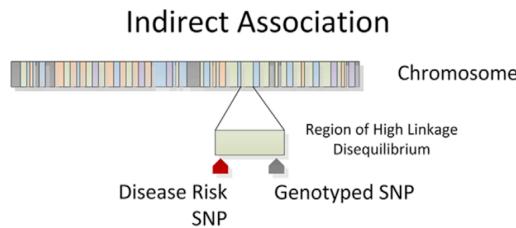
Only a subset of SNPs can be **genotyped** in genome-wide association studies. **Imputation** methods can infer the alleles of hidden variants and use those inferences to test the hidden variants for association.

Linkage disequilibrium



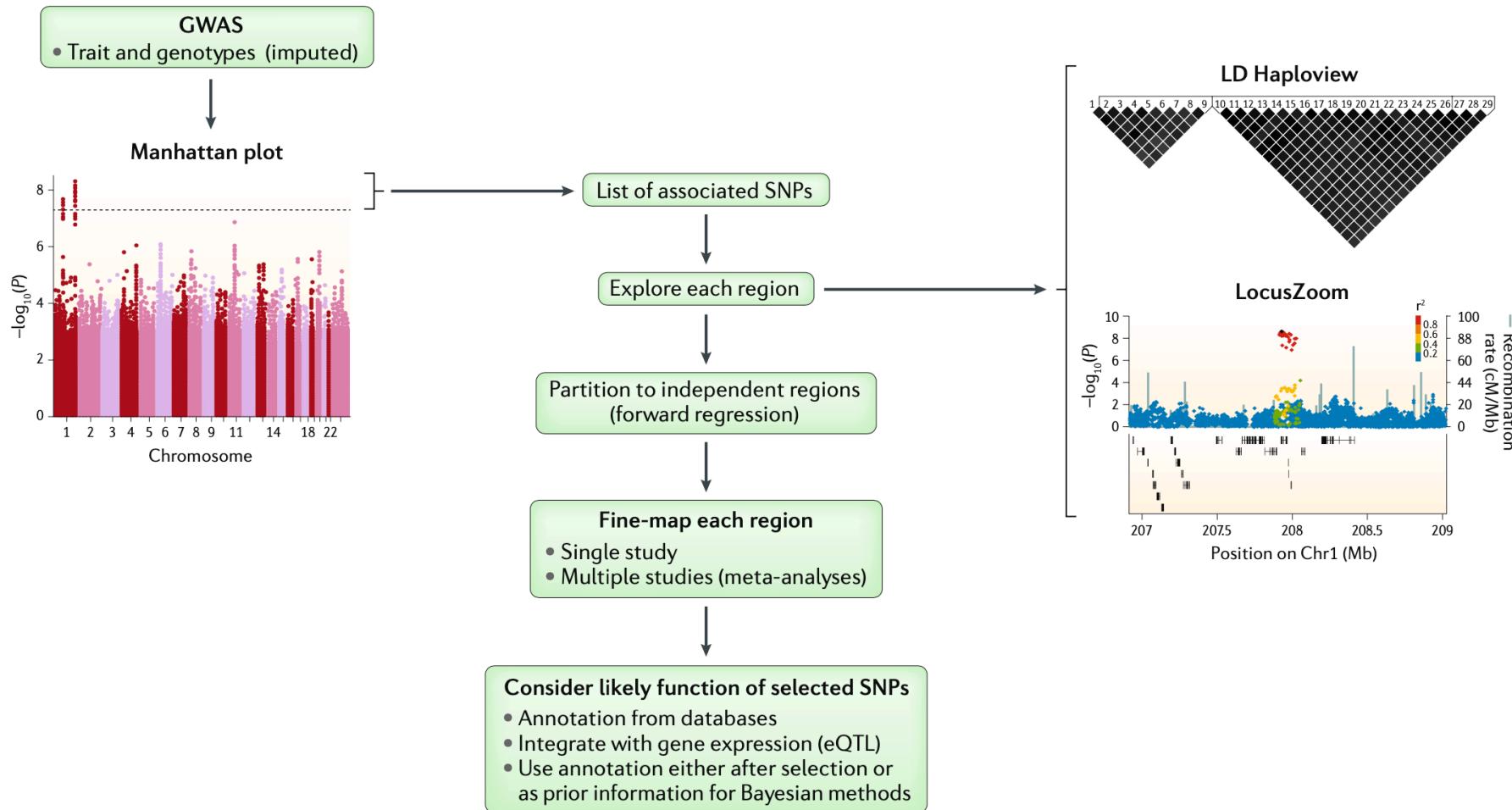
Imputation is possible due to the correlation between SNPs, i.e. **linkage disequilibrium (LD)**.

Indirect association



- It is highly likely that the **influential SNP** is not directly typed, but instead a **tag SNP** in high LD with the influential SNP is genotyped and statistically associated to the phenotype (**indirect association**).
- **Fine-mapping**: refining the genomic localization of **causal variants** by the use of statistical, bioinformatic or functional methods.

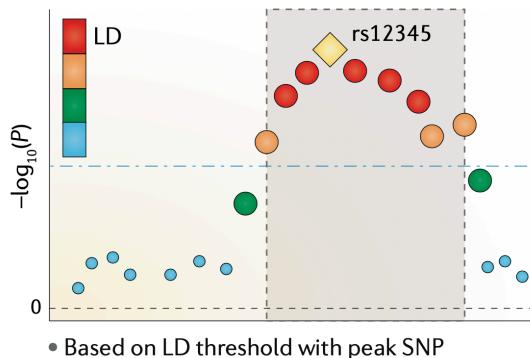
Fine-mapping workflow



Schaid, D. J., Chen, W., & Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics*, 19(8), 491–504. <https://doi.org/10.1038/s41576-018-0016-z>

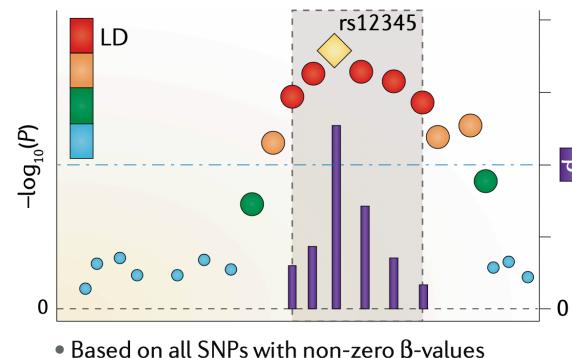
Fine-mapping strategies

A Heuristic LD approach



- **Joint effects** of SNPs on a trait are not modelled
- LD thresholds can be **arbitrary**

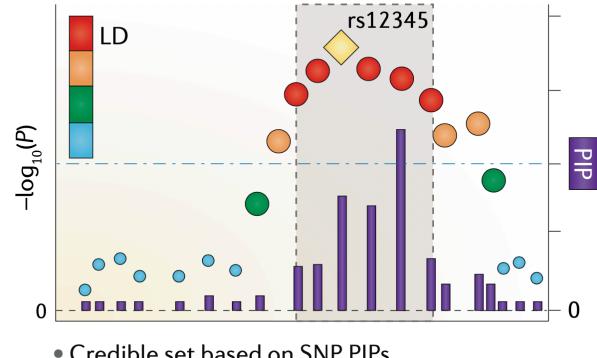
B Penalized regression



$$\sum_{i=1}^M \left(y_i - \sum_{j=0}^p w_j \times x_{ij} \right)^2 + \lambda \sum_{j=0}^p |w_j|$$

- Using **penalized regression models** like lasso and elastic net
- Tuning parameters are often estimated by **cross-validation**

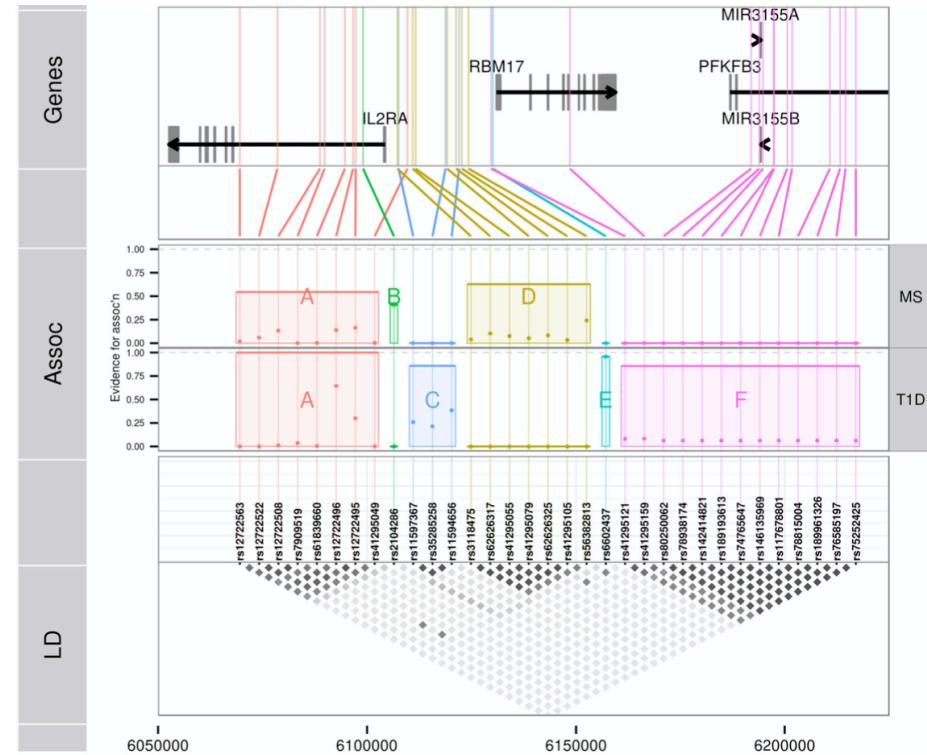
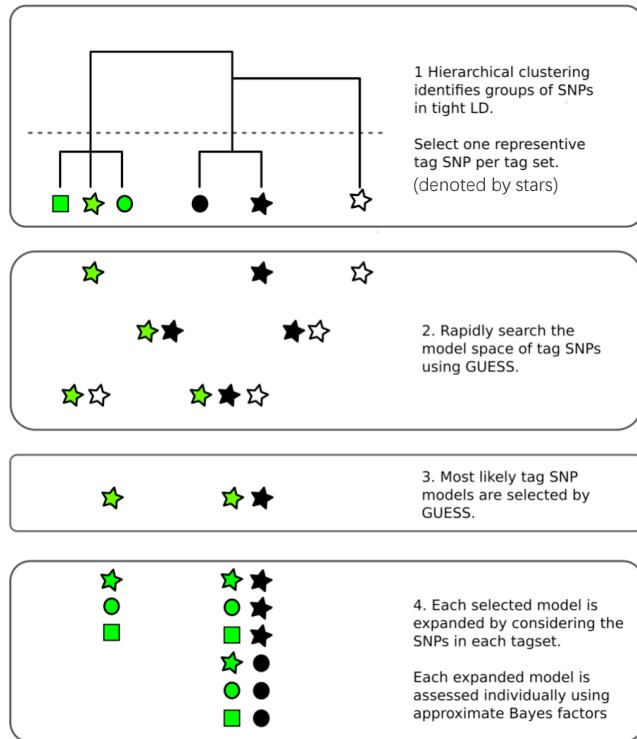
C Bayesian fine-mapping



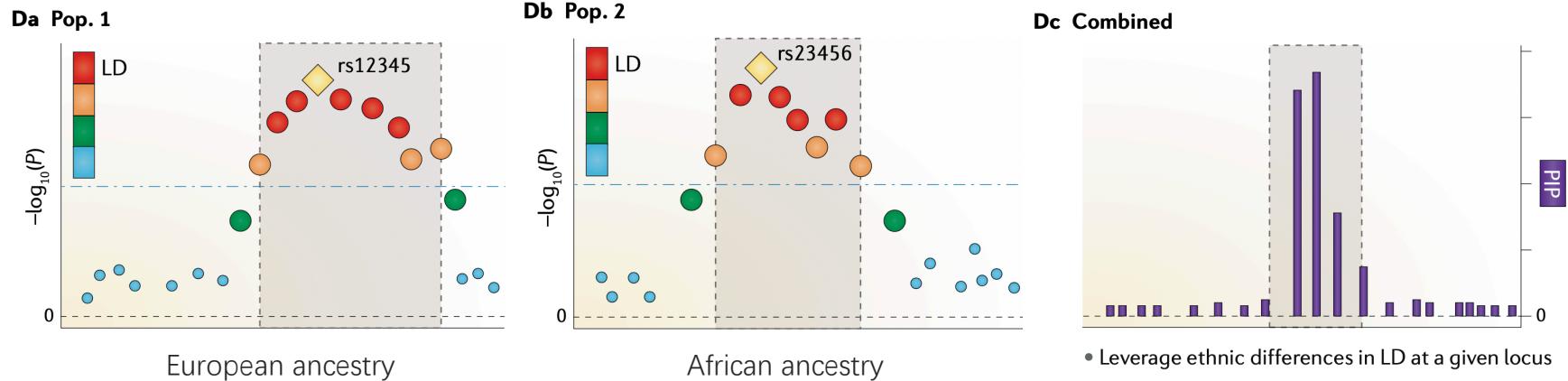
$$PIP_j = P(c_j = 1 | D) = \sum_{M, c_j=1} P(M | D)$$

- **Stochastic searching** is generally applied over a wide array of possible models.

Bayesian fine-mapping example: GUESSFM

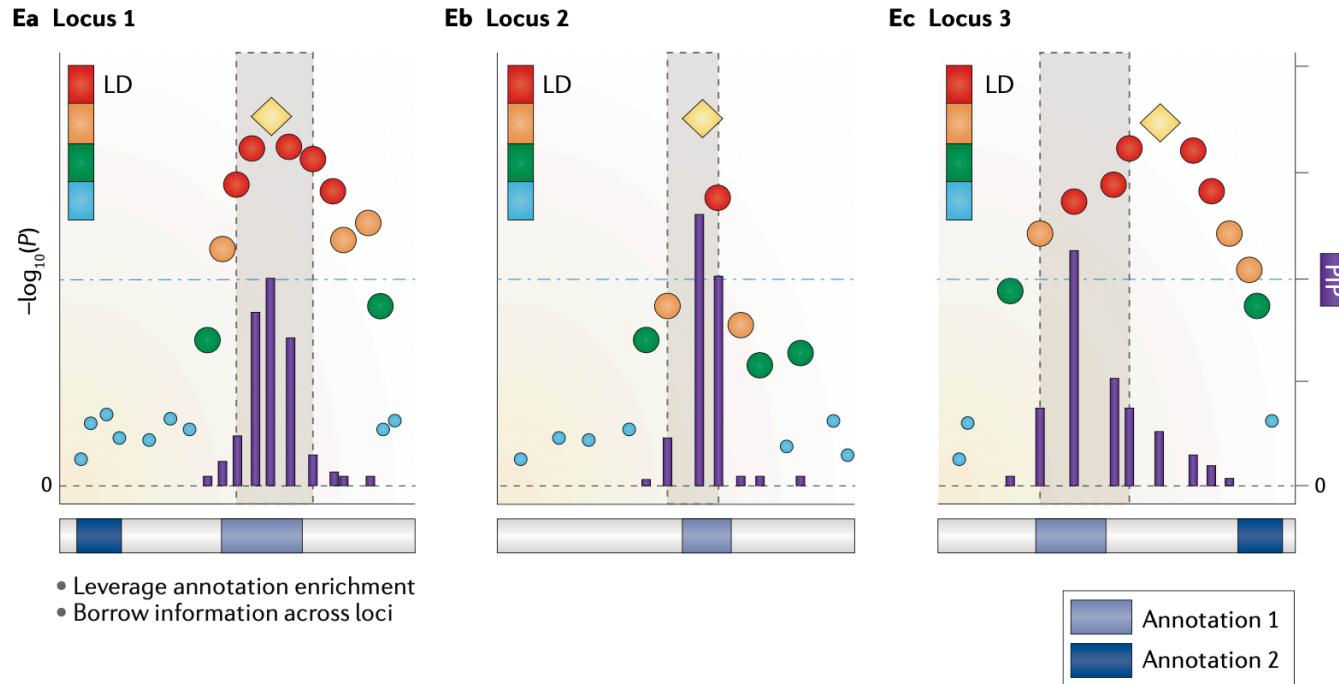


Trans-ethnic fine-mapping



- A substantial reduction in the size of Bayesian fine-mapping **credible sets** can be attained by including subjects of **African ancestry** because they have **much narrower LD** due to the accumulation of more recombination events.
- **Random-effects models** are often used, recognizing that a SNP can have different effect sizes across different ethnic groups.

Multi-region fine-mapping with annotation



- For loci 1 and 2, the peak SNPs overlap with Annotation 1, indicating **enrichment**. This enrichment results in the SNP with the highest PIP in locus 3 to be different from the peak SNP in locus 3.
- Prior probability** that a SNP is causal is adjusted depending on **annotation**.
- Challenge:** current understanding of broad genomic function may be too limiting to accurately improve prior probabilities of causation.