Functionally informed fine-mapping and polygenic localization of complex trait heritability

@oweissb

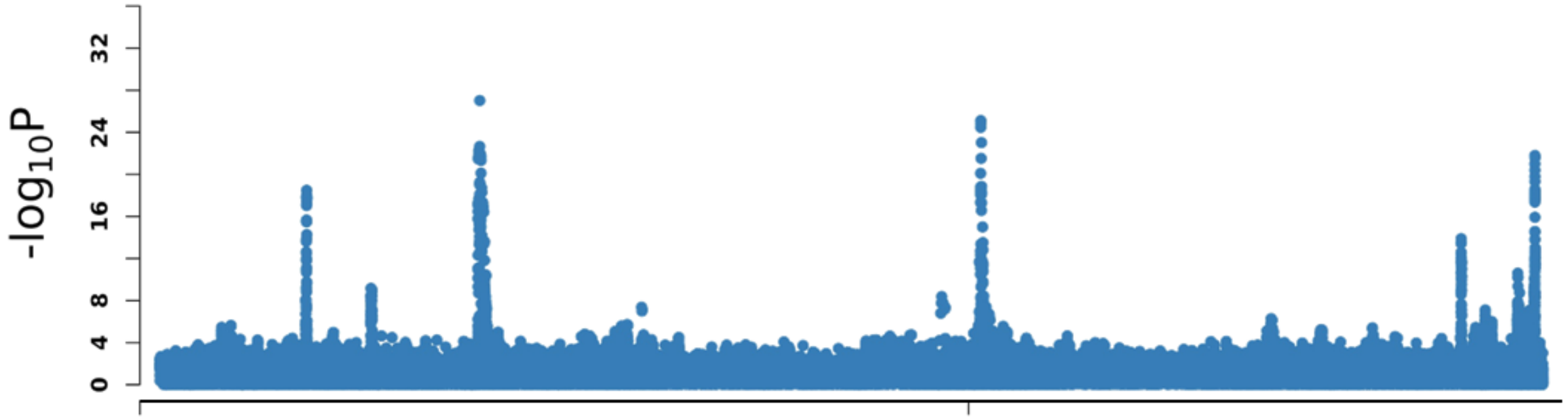
Omer Weissbrod (presented by Farhad Hormozdiari)
Alkes Price Group
Harvard School of Public Health
10.17.2019

Weissbrod *et al.* 2019 bioRxiv

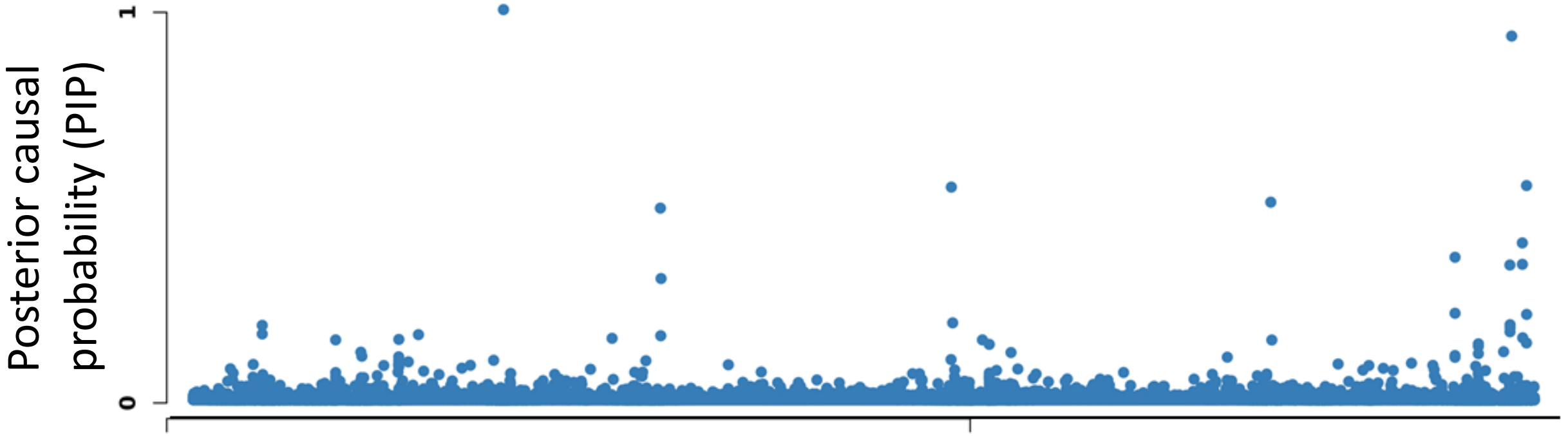
Outline

- **Motivation**
- Methods
- Results
- Polygenic localization

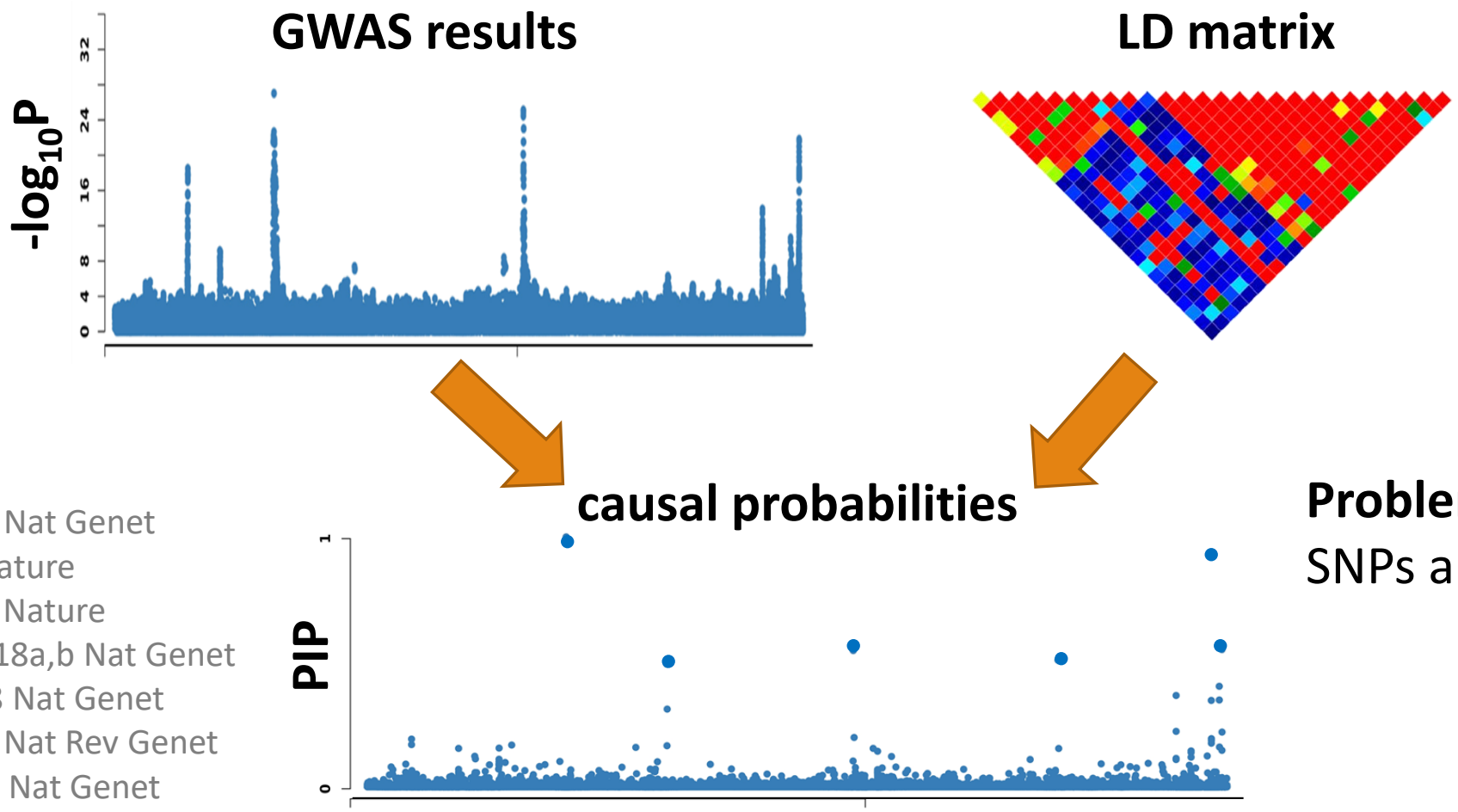
GWAS identify associations, not causality



We want causality, not associations



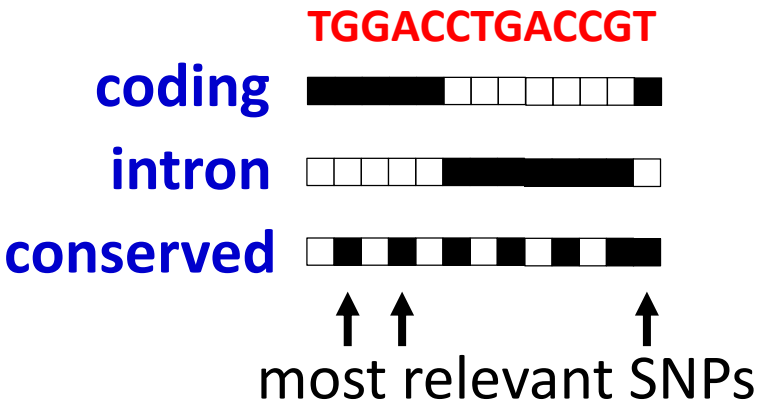
Fine-mapping identifies causal SNPs

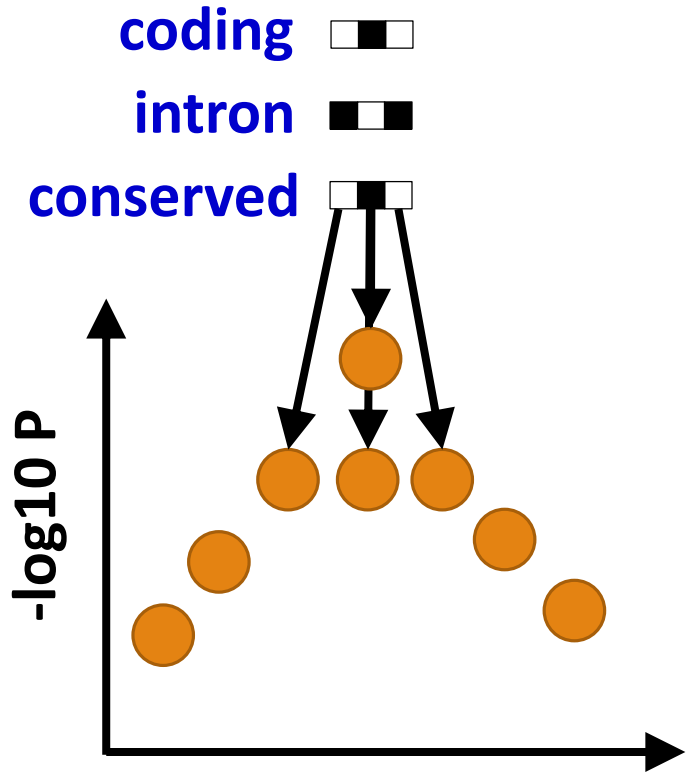


Maller *et al.* 2012 Nat Genet
Farh *et al*. 2015 Nature
Huang *et al.* 2017 Nature
Mahajan *et al.* 2018a,b Nat Genet
Westra *et al*. 2018 Nat Genet
Schaid *et al.* 2018 Nat Rev Genet
Ulirsch *et al.* 2019 Nat Genet

Problem:
SNPs are in strong LD!

Functional annotations tease apart SNPs in strong LD





Kichaev *et al.* 2014 PLOS Genet
Chen *et al.* 2016 Genetics
Wen *et al.* 2016 AJHG
Mahajan *et al.* 2018b Nat Genet

PolyFun: fine-mapping with polygenic functional priors

Problem: previous functionally-informed fine-mapping methods can either:

- ✘ Analyze information from only a few loci (<20)
- ✘ Use only a few functional annotations (<20)

PolyFun leverages **modern fine-mapping methods** and **stratified LD-score regression** to:

- ✓ Analyze genome-wide information
- ✓ Use hundreds of functional annotations

Kichaev *et al.* 2014 PLOS Genet
Chen *et al.* 2016 Genetics
Wen *et al.* 2016 AJHG
Mahajan *et al.* 2018b Nat Genet

FINEMAP: Benner *et al.* 2016 Bioinformatics, 2018 bioRxiv
SuSiE: Wang *et al.* 2018 bioRxiv
S-LDSC: Finucane *et al.* 2015 Nat Genet
PolyFun: Weissbrod *et al.* 2019 bioRxiv

Outline

- Motivation
- **Methods**
- Results
- Polygenic localization

PolyFun leverages the fine-mapping model and the stratified LD-score regression (S-LDSC) baseline-LF model

Fine-mapping model: SNP effect β_i is either zero (null SNP) or normally distributed (causal SNP)

S-LDSC model:

$$\underbrace{\sum_{c} \tau^c \cdot a_i^c}_{\text{S-LDSC model}} = \underbrace{\text{var}[\beta_i | \boldsymbol{a}_i]}_{\text{per-SNP } h^2} = \underbrace{\text{P}(\beta_i \neq 0 \mid a_i)}_{\text{prior causal probability}} \cdot \underbrace{\text{var}[\beta_i \mid \beta_i \neq 0]}_{\text{causal variance}}$$

annotation coefficient

annotations

Finucane *et al.* 2015 Nat Genet
Gazal *et al.* 2017,2018,2019 Nat Genet

PolyFun is robust to modeling misspecification

PolyFun procedure:

1. Estimate per-SNP heritabilities on even (resp. odd) chromosomes using L2-regularized stratified LD score regression

2. Partition SNPs into bins of similar per-SNP heritability

3. Re-estimate per-SNP heritabilities in each bin using odd (resp. even) chromosomes

4. Compute prior causal probabilities

Finucane *et al.* 2015 Nat Genet

Outline

- Motivation
- Methods
- **Results**
- Polygenic localization

Data analysis details

Data:

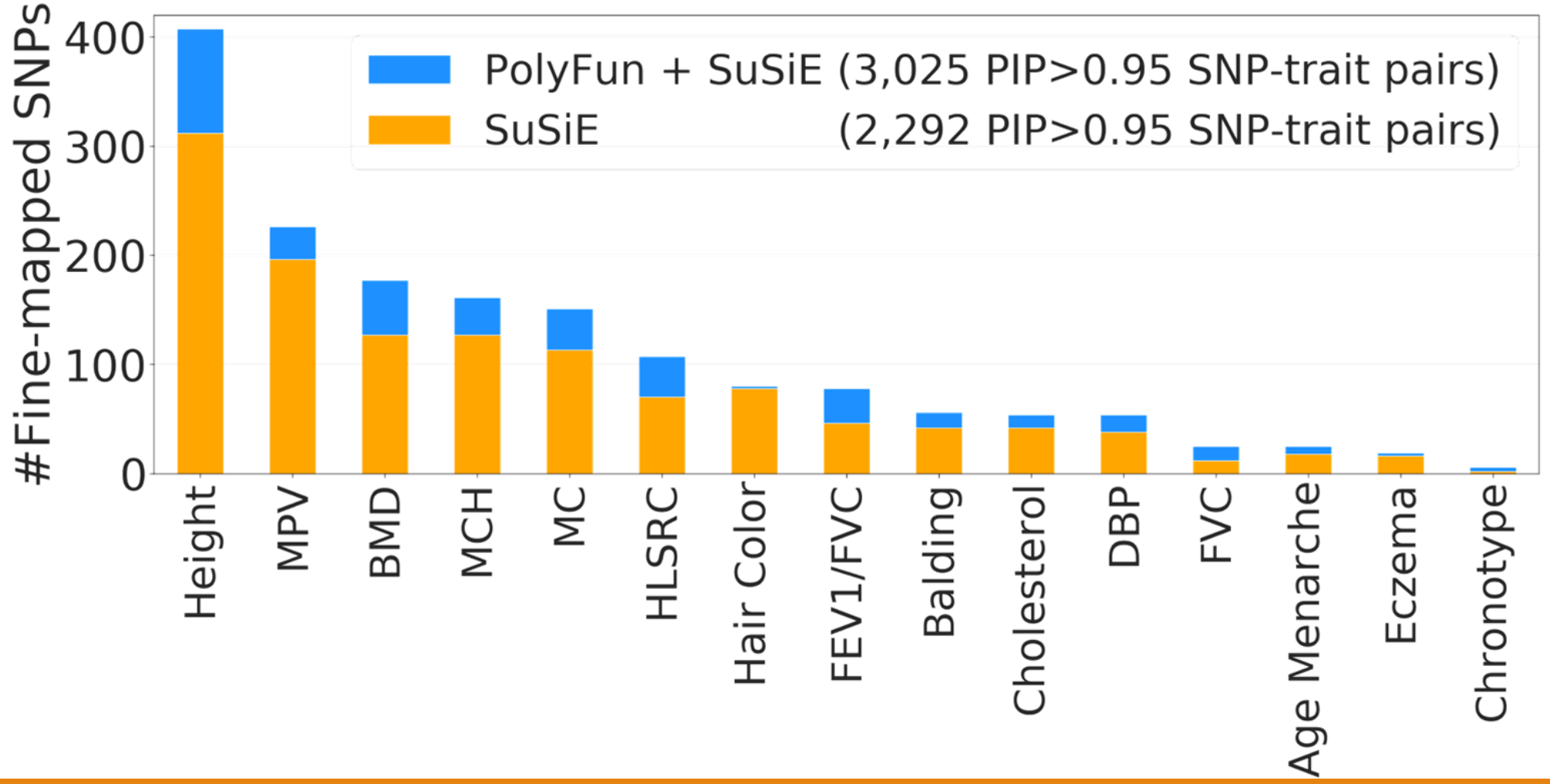
- 47 UK Biobank traits (average N=317K)
- 19M SNPs with MAF≥0.1% (excluding MHC)

Annotations:

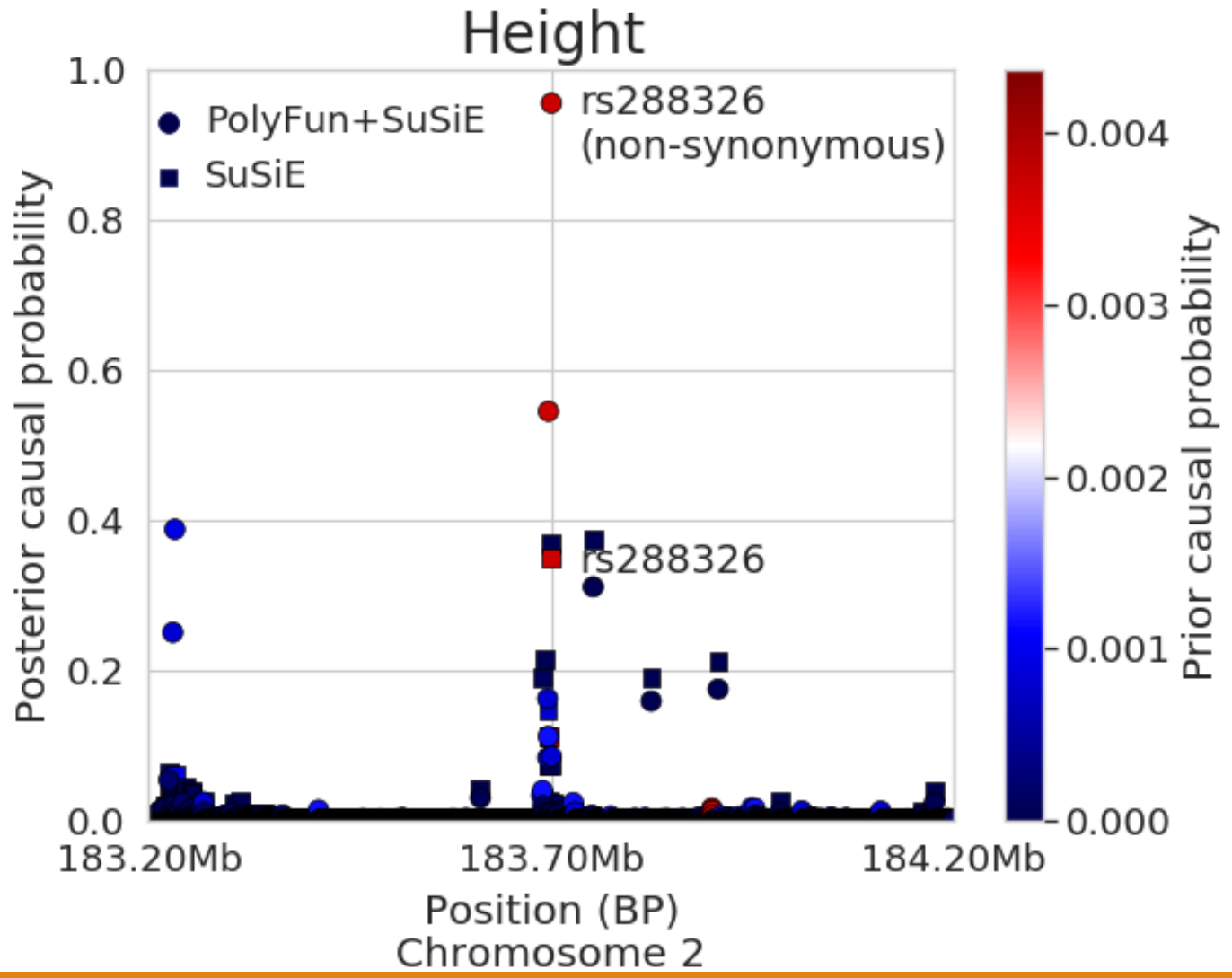
- 187 functional annotations from baseline-LF model
  (a broad set of coding, conserved, regulatory, MAF and LD-related annotations)

Gazal *et al.* 2017,2018,2019 Nat Genet

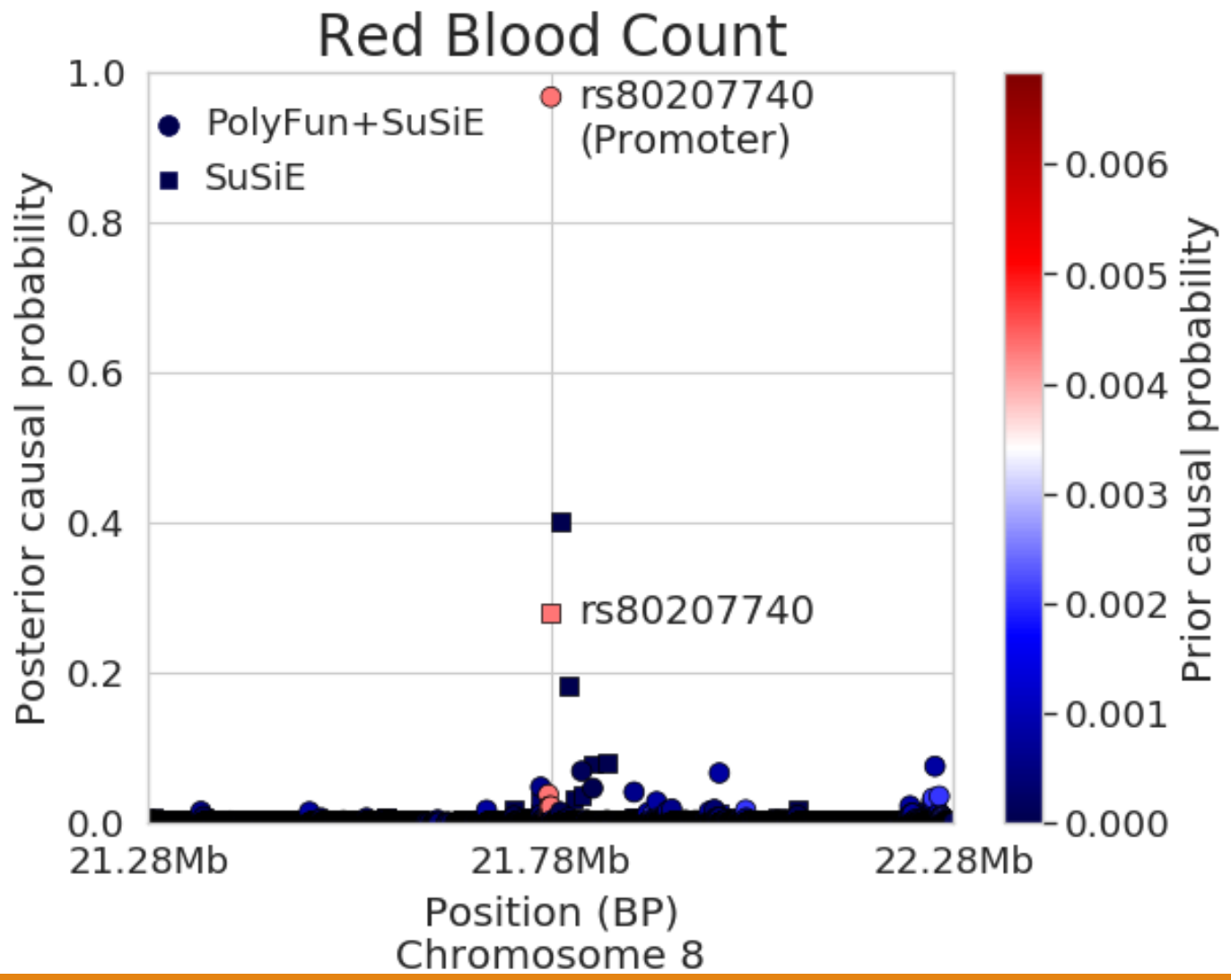
PolyFun finds 32% more fine-mapped SNPs (PIP>0.95) than non-functionally informed fine-mapping



Functional annotations improve fine-mapping resolution

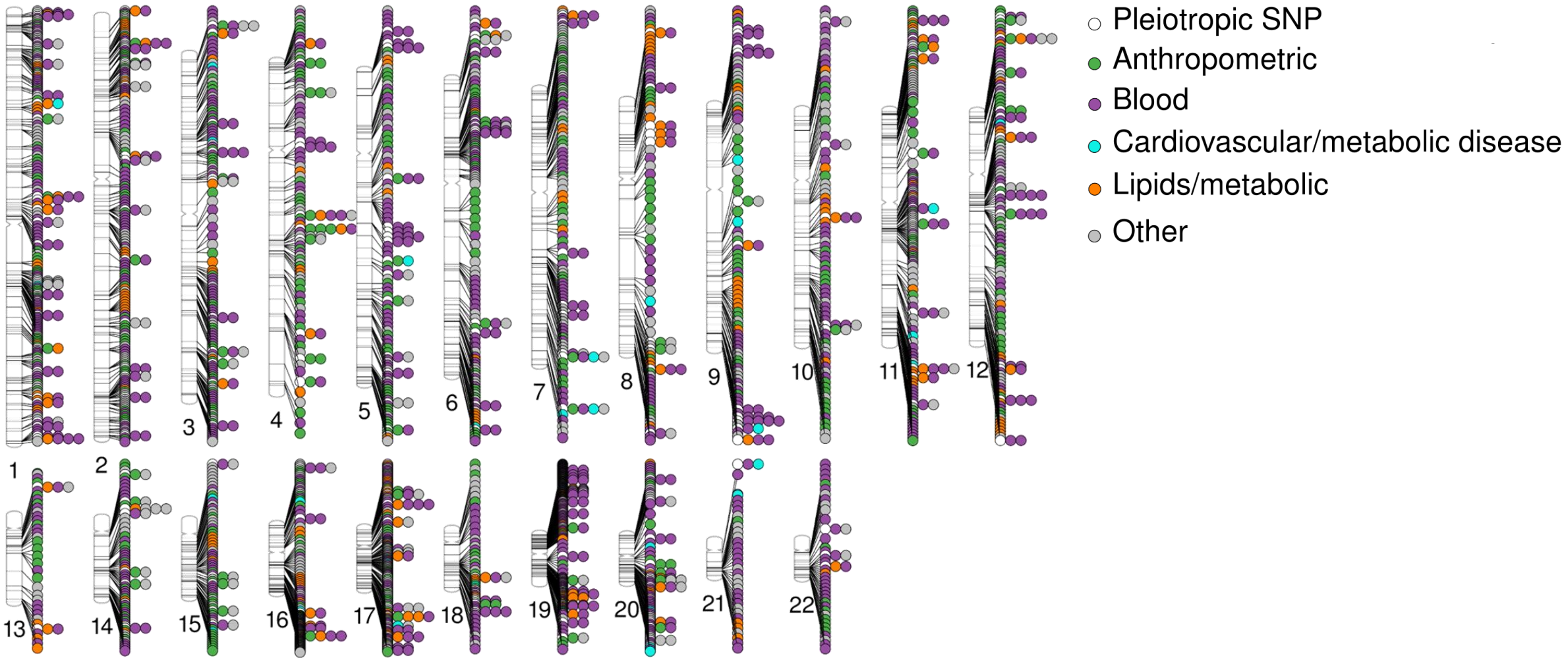


Functional annotations improve fine-mapping resolution



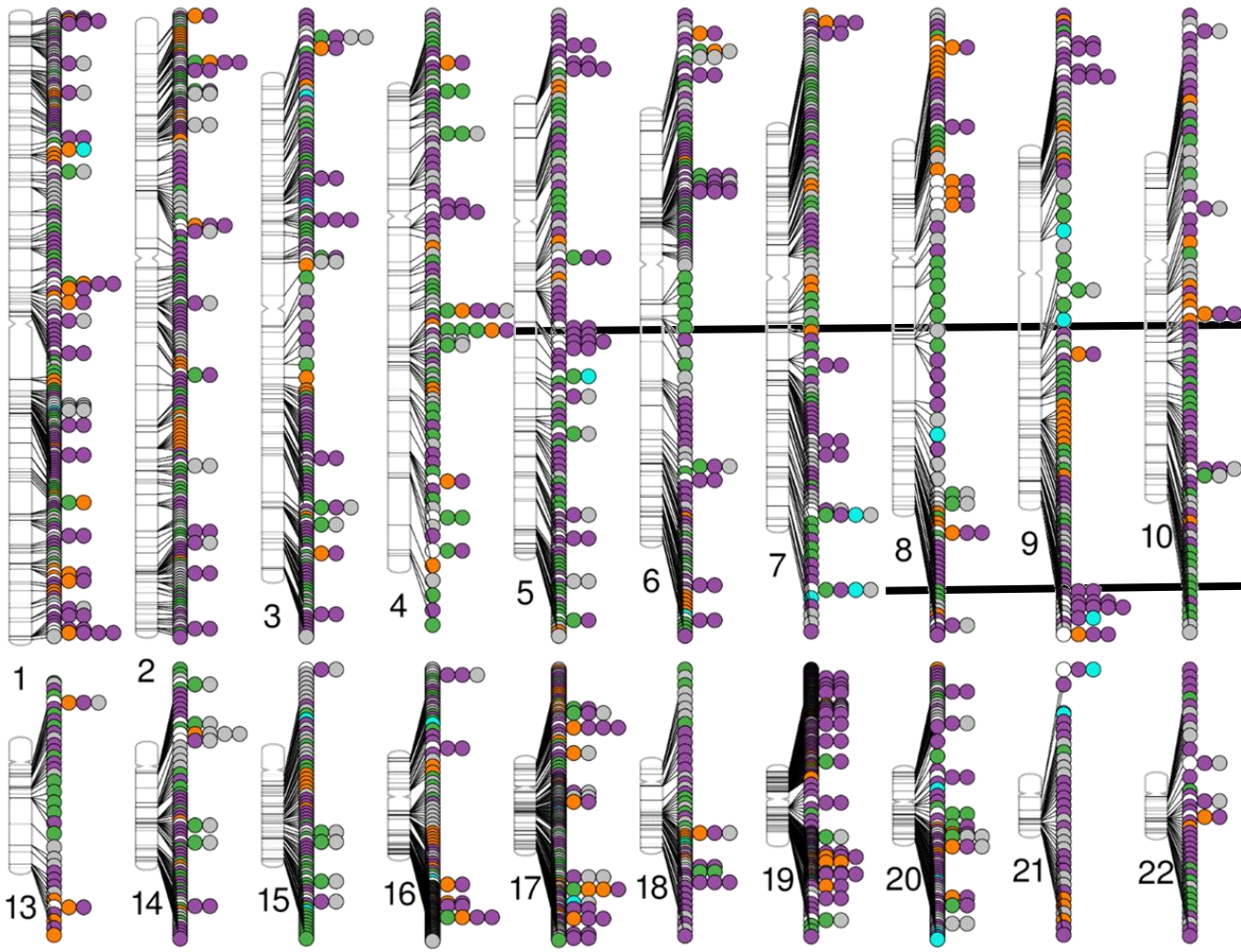
PolyFun identifies over 3,000 fine-mapped (PIP>0.95) SNP-trait pairs across anthropometric, blood, disease and lipid traits in the UK Biobank

Fine-mapping results



PolyFun finds 223 pleiotropic fine-mapped SNPs (PIP>0.95) across genetically uncorrelated traits

Fine-mapping results



rs76895963
Height, triglycerides, RBC, BMD, SBP

rs13107325
Height, BMI, WHR-adj-BMI, RBC, total cholesterol, DBP, FVC, balding

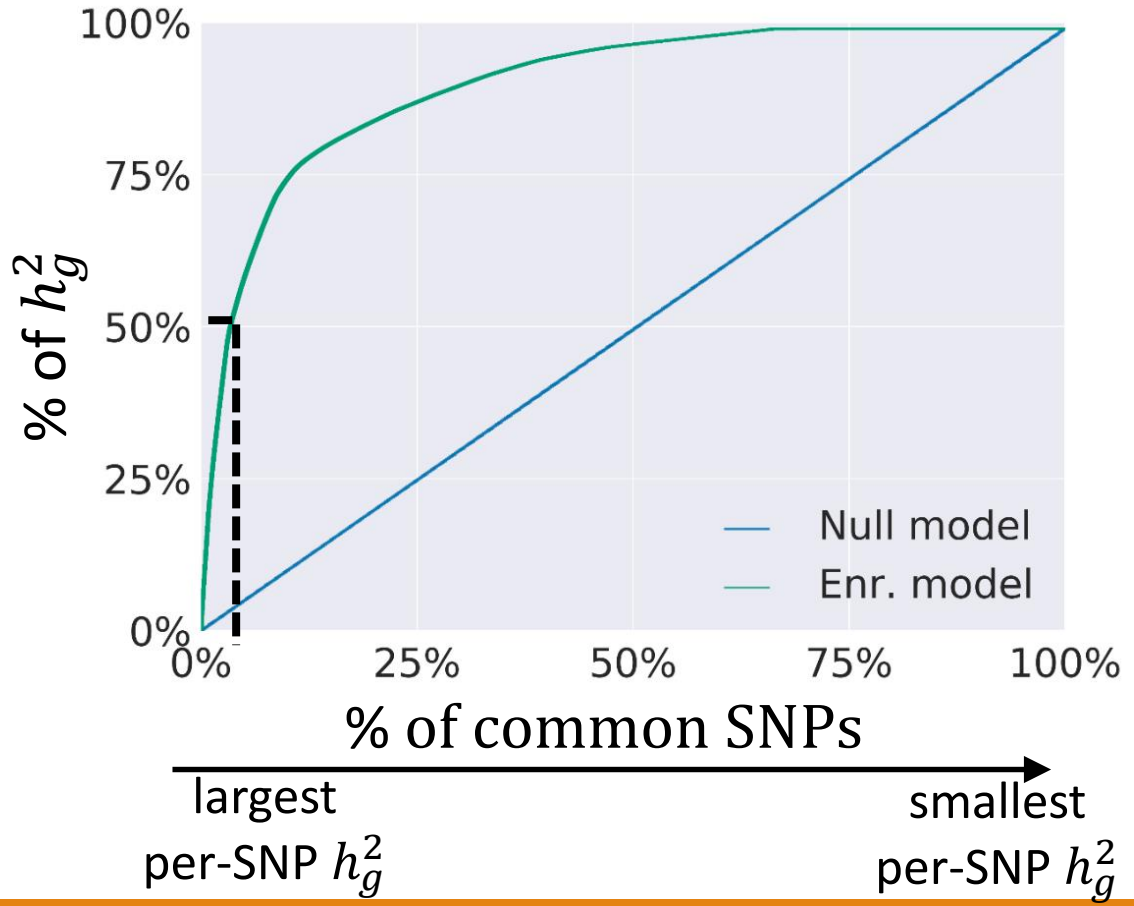
rs11556924
Height, cardiovascular disease, platelet count, age of menarche

Outline

- Motivation
- Methods
- Results
- **Polygenic localization**

Polygenic localization: localizing heritability

- **Motivation**: PIP>0.95 SNPs causally explain a small $\%h_g^2$. Where is the rest?
- **Definition**: Identify a **minimal** set of SNPs causally explaining (e.g.) 50% of h_g^2



Polygenic localization: method

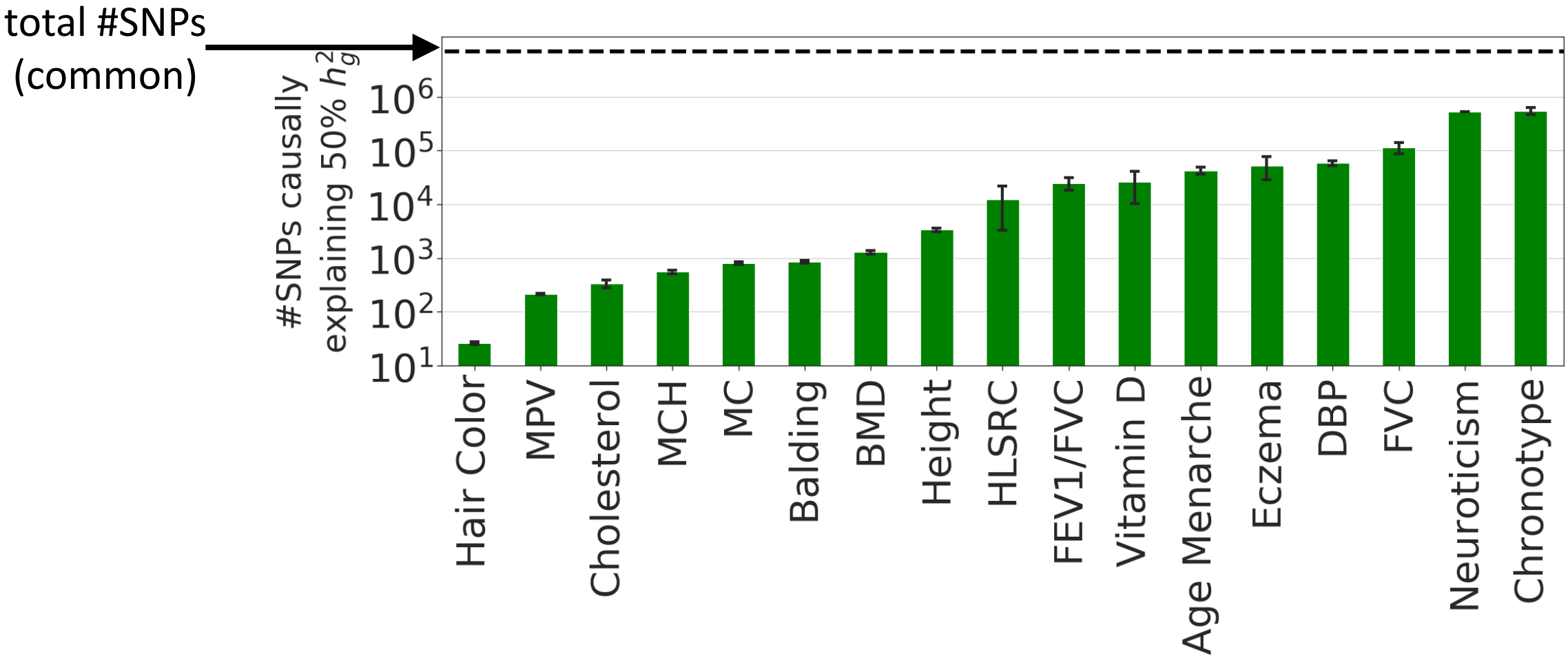
Cannot reuse PolyFun estimates of per-SNP heritability

Beware of winner`s curse...

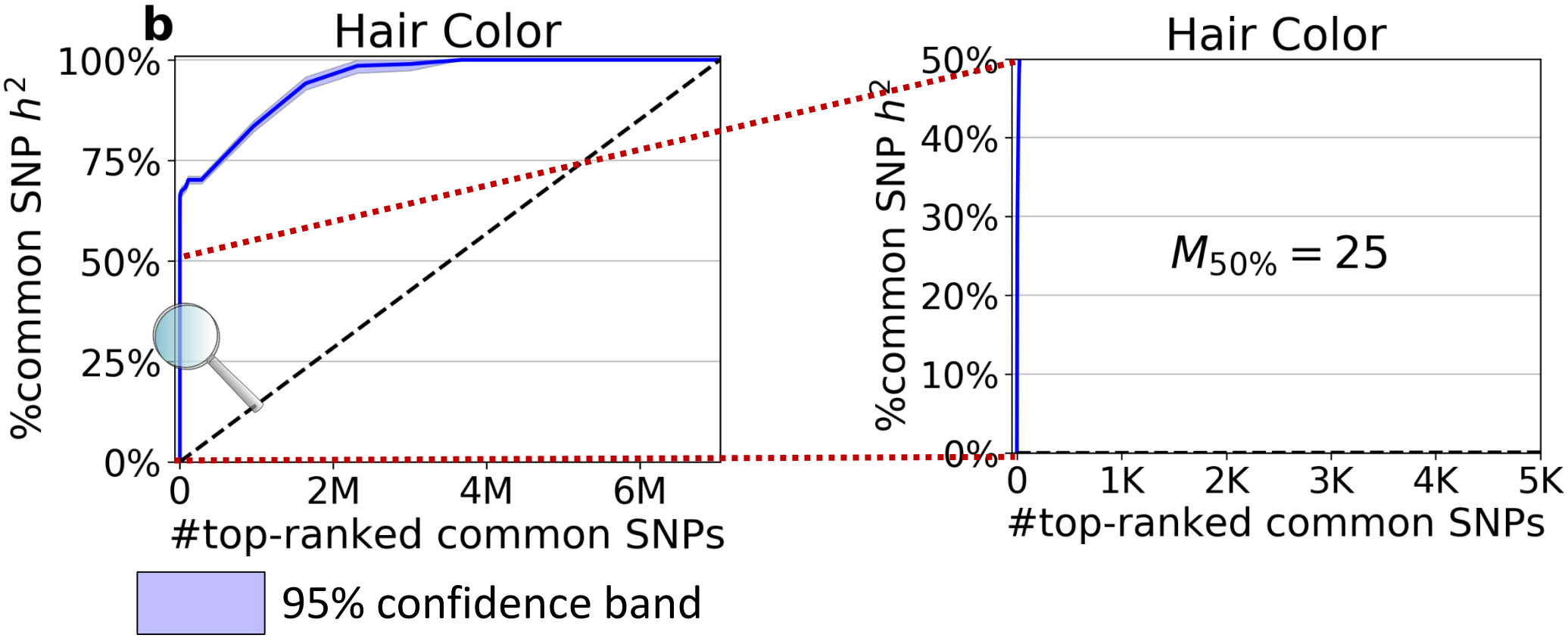
Instead:

- Estimate per-SNP heritabilities with PolyFun
- Partition SNPs into bins of similar per-SNP heritability
- Re-estimate heritability in each bin with S-LDSC, using different data (N=122K UK Biobank individuals not in the N=337K PolyFun dataset)

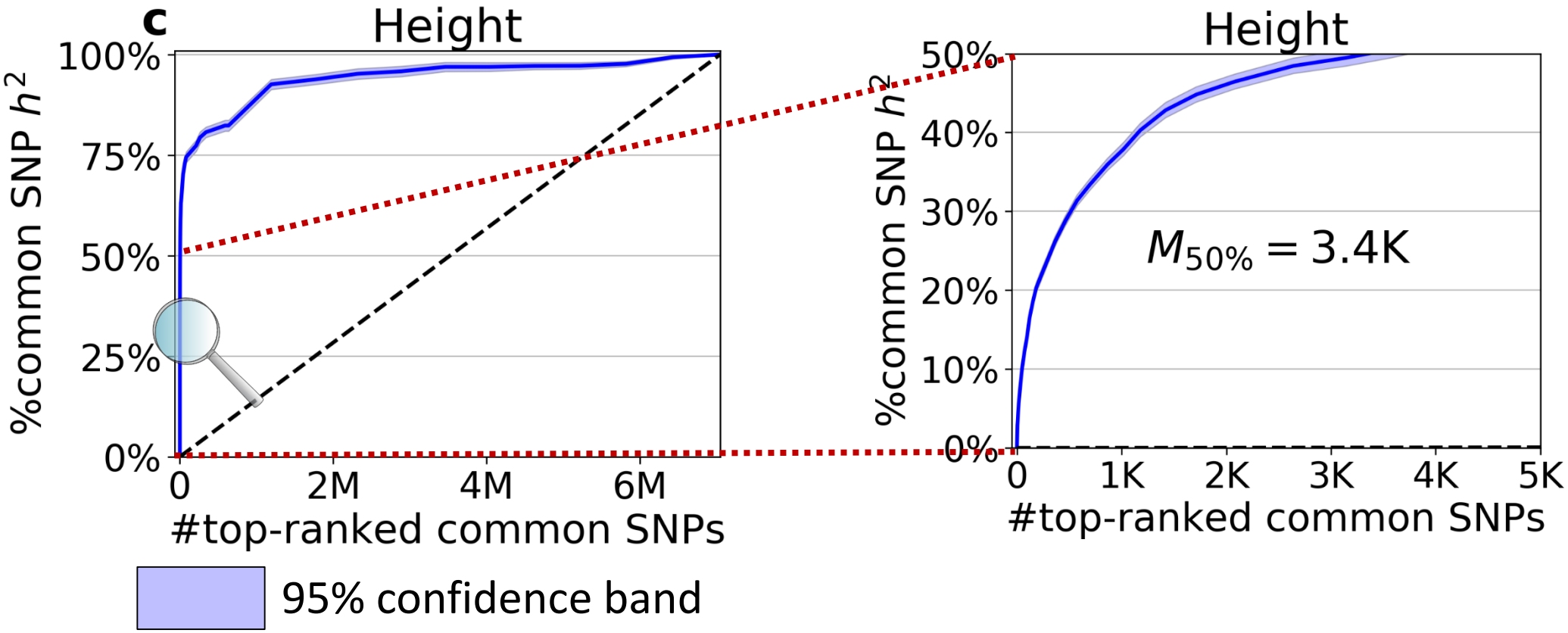
Minimal SNP sets causally explaining 50% h_g^2 vary in size across orders of magnitude



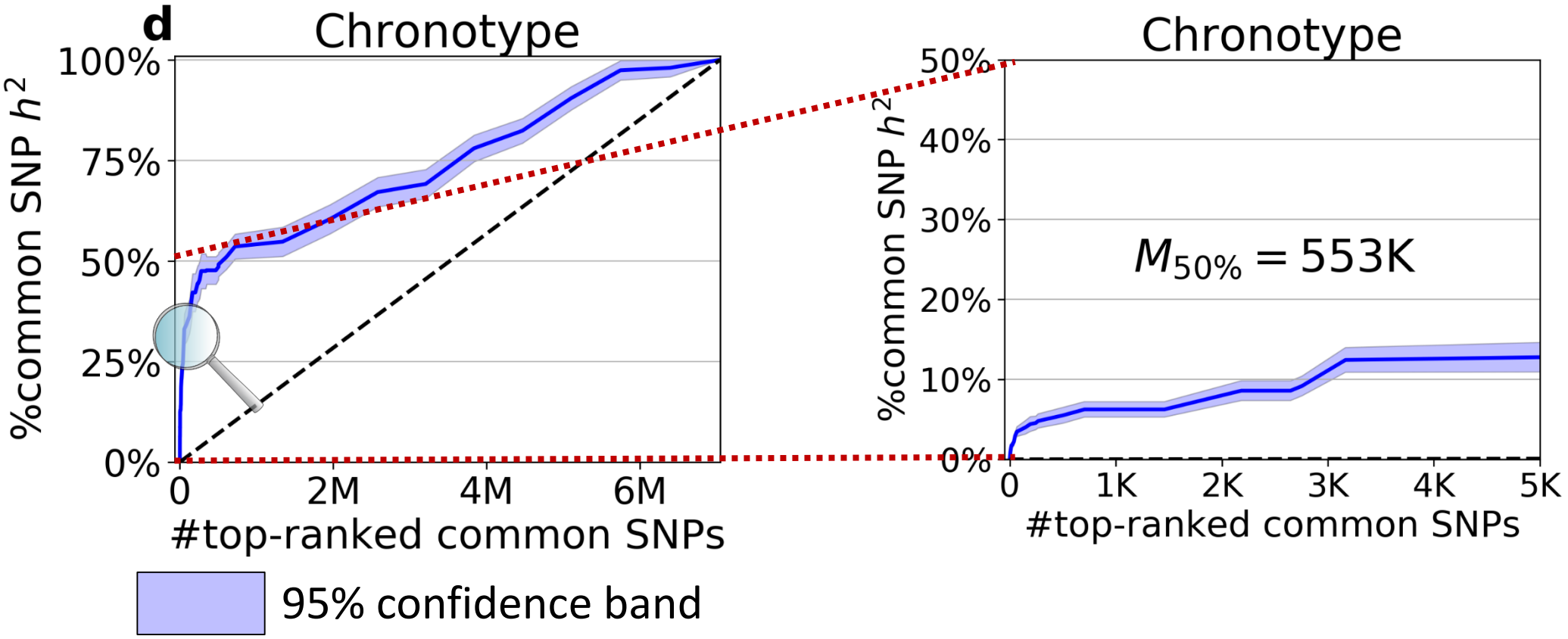
Hair color: strong polygenic localization



Height: intermediate polygenic localization



Chronotype (morning person): weak polygenic localization



Conclusions

We propose:

- **PolyFun**: Functionally-informed fine-mapping with polygenic priors
- **Polygenic localization**: Find minimal SNP sets causally explaining a given proportion of SNP heritability

Results:

- PolyFun + SuSiE finds >3,000 fine-mapped SNP-trait pairs
- Many SNPs are pleiotropic for multiple traits
- 50% of SNP heritability is causally explained by 25-550,000 SNPs

Acknowledgements

Alkes Price
Farhad Hormozdiari
Steven Gazal
Armin Schoech
Price Group

Hilary Finucane
Ran Cui
Jacob Ulirsch
Luke O'Connor

Christian Benner

Weissbrod *et al.* 2019 bioRxiv