

# SBayesRC: Polygenic Prediction Incorporating functional annotations

Jian Zeng

[j.zeng@uq.edu.au](mailto:j.zeng@uq.edu.au)

# Functional genomic annotations

Functional genomic annotations provide orthogonal information useful for polygenic prediction.

- Chromatin states
- Biological functions
- Molecular quantitative trait loci (xQTL)
- .....

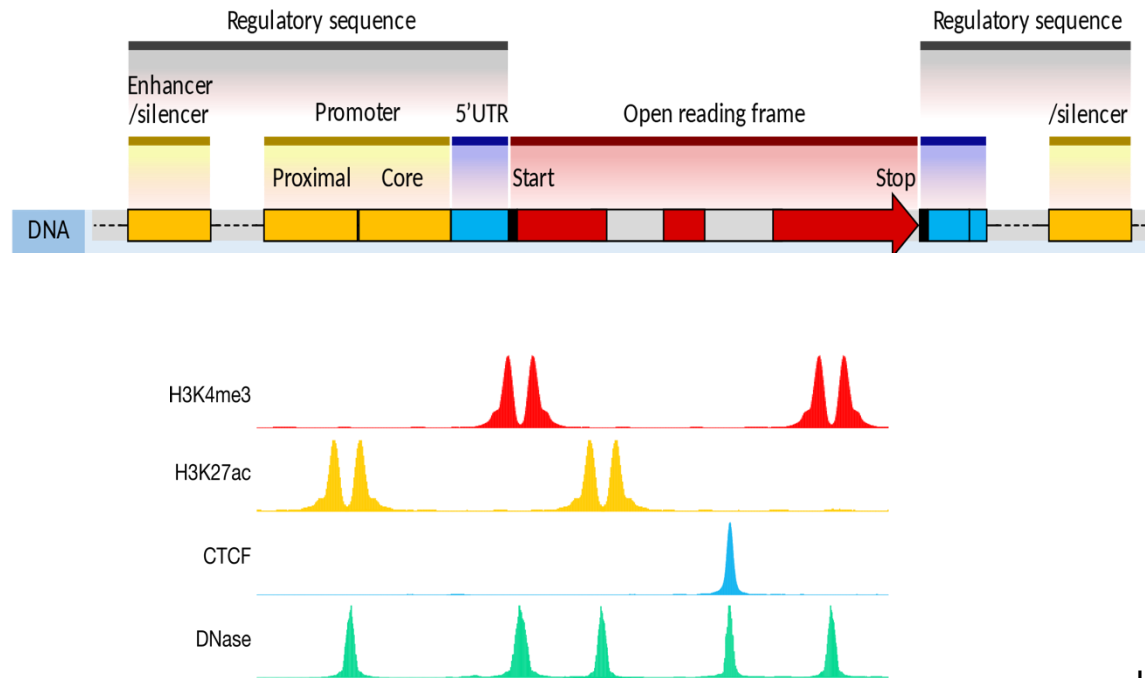


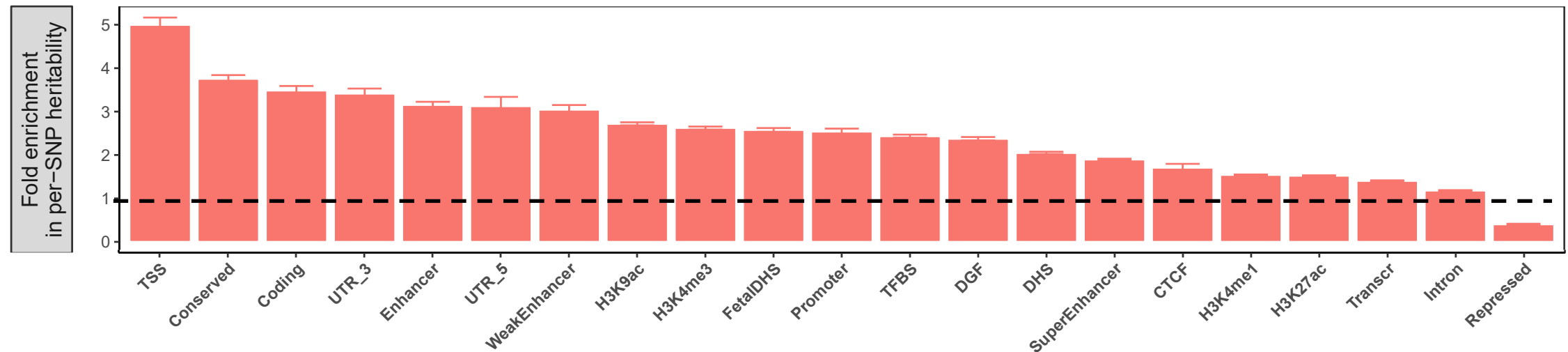
Image from ENCODE

# Functional genomic annotations

Functional genomic annotations provide orthogonal information useful for polygenic prediction.

- Chromatin states
- Biological functions
- Molecular quantitative trait loci (xQTL)
- .....

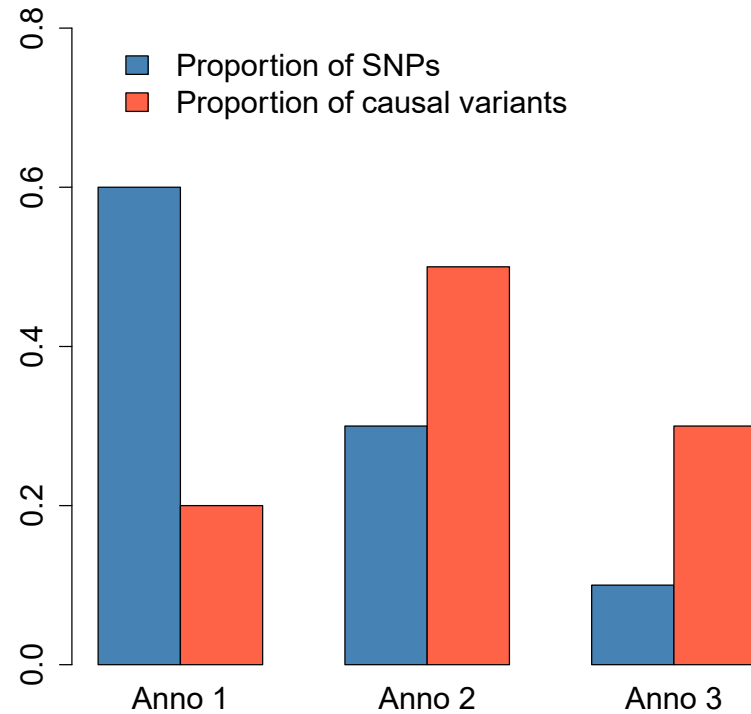
Zeng et al 2021 Nature Communications



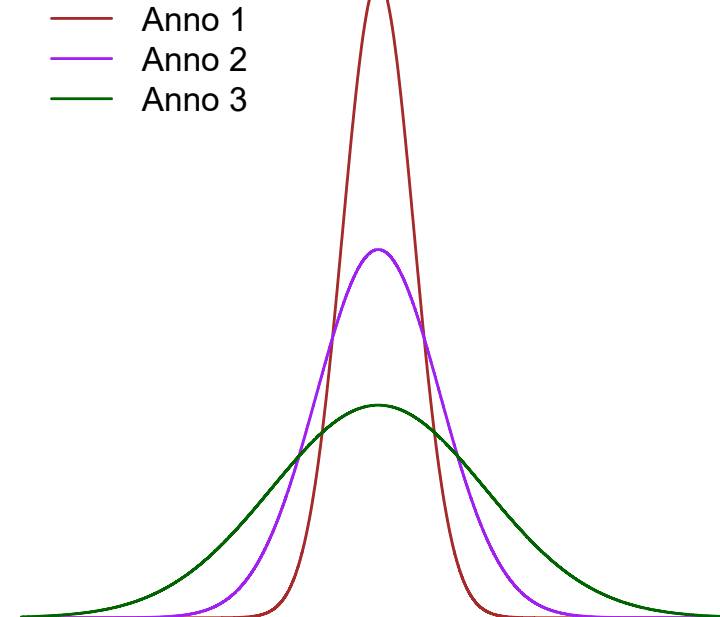
# Two dimensions of information

**Functional annotations are informative on both the presence of causal variants and the distribution of causal effect sizes.**

Differences in proportion of causal variants

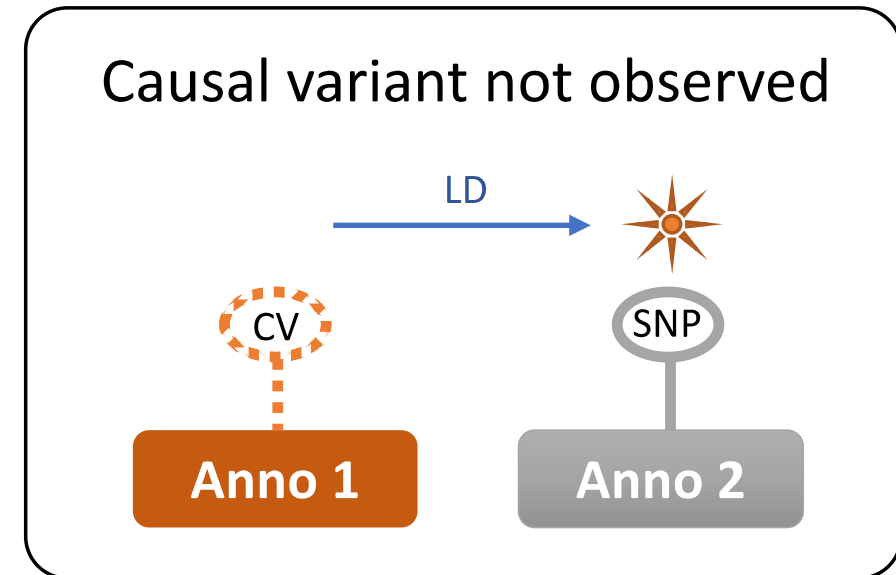
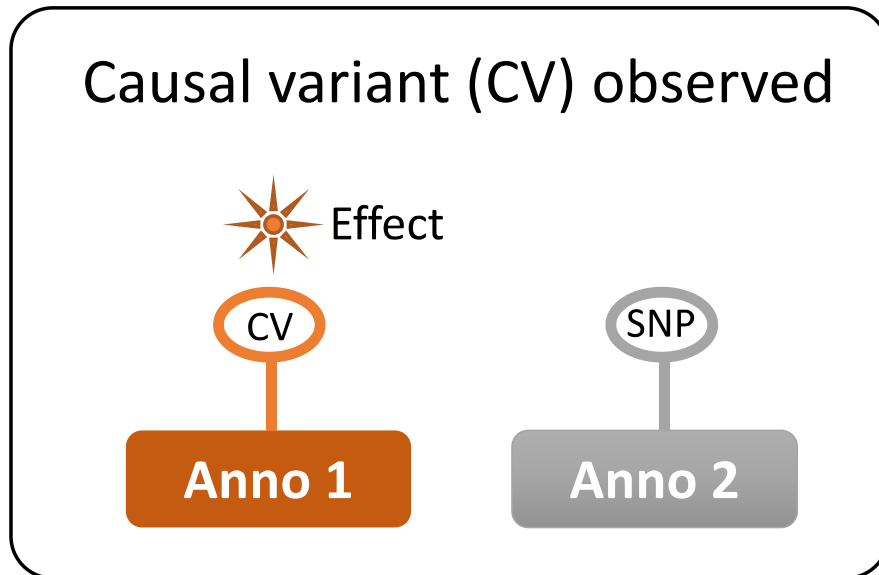


Differences in distribution of causal effects



# Opportunities/challenges

When causal variants are not observed, SNP markers can tag the causal variant by LD but may not tag by annotation.



**It's best to model all SNPs simultaneously with their annotations!**

## nature communications

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature communications](#) > [articles](#) > [article](#)

Article | [Open Access](#) | [Published: 18 October 2021](#)

### Incorporating functional priors improves polygenic prediction accuracy in UK Biobank and 23andMe data sets

[Carla Márquez-Luna](#) , [Steven Gazal](#), [Po-Ru Loh](#), [Samuel S. Kim](#), [Nicholas Furlotte](#), [Adam Auton](#), [23andMe Research Team](#) & [Alkes L. Price](#) 

### LDpred-funct

### Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits

[I. M. MacLeod](#) , [P. J. Bowman](#), [C. J. Vander Jagt](#), [M. Haile-Mariam](#), [K. E. Kemper](#), [A. J. Chamberlain](#), [C. Schrooten](#), [B. J. Hayes](#) & [M. E. Goddard](#)

*BMC Genomics* **17**, Article number: 144 (2016) | [Cite this article](#)

**6209** Accesses | **146** Citations | **9** Altmetric | [Metrics](#)

### BayesRC

## PLOS COMPUTATIONAL BIOLOGY

 OPEN ACCESS  PEER-REVIEWED



RESEARCH ARTICLE

### Leveraging functional annotations in genetic risk prediction for human complex diseases

[Yiming Hu](#) , [Qiongshi Lu](#) , [Ryan Powles](#), [Xinwei Yao](#), [Can Yang](#), [Fang Fang](#), [Xinran Xu](#), [Hongyu Zhao](#) 

### AnnoPred

### Winner's Curse Correction and Variable Thresholding Improve Performance of Polygenic Risk Modeling Based on Genome-Wide Association Study Summary-Level Data

[Jianxin Shi](#) , [Ju-Hyun Park](#), [Jubao Duan](#), [Sonja T. Berndt](#), [Winton Moy](#), [Kai Yu](#), [Lei Song](#), [William Wheeler](#), [Xing Hua](#), [Debra Silverman](#), [Montserrat Garcia-Closas](#), [Chao Agnes Hsiung](#), [Jonine D. Figueroa](#), [ ... ], [Nilanjan Chatterjee](#)  [ view all ]

### P+T-funct-LASSO

## nature genetics

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature genetics](#) > [articles](#) > [article](#)

Article | [Published: 07 April 2022](#)

### Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores

[Omer Weissbrod](#) , [Masahiro Kanai](#), [Huwenbo Shi](#), [Steven Gazal](#), [Wouter J. Peyrot](#), [Amit V. Khera](#), [Yukinori Okada](#), [The Biobank Japan Project](#), [Alicia R. Martin](#), [Hilary K. Finucane](#) & [Alkes L. Price](#) 

*Nature Genetics* **54**, 450–458 (2022) | [Cite this article](#)

### PolyPred

Need new method that can

- simultaneously fit all SNPs and annotation data in a unified model
- account for variations in both causal variant proportion and causal effect distribution

Leveraging functional annotations for cross-ancestry prediction

nature genetics



Article

<https://doi.org/10.1038/s41588-024-01704-y>

## Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries

Received: 1 October 2022

Accepted: 5 March 2024

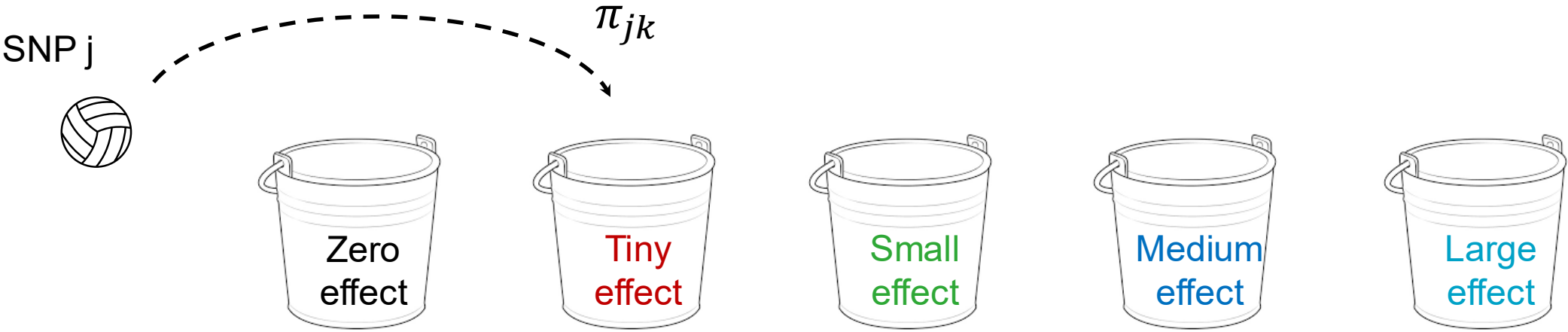
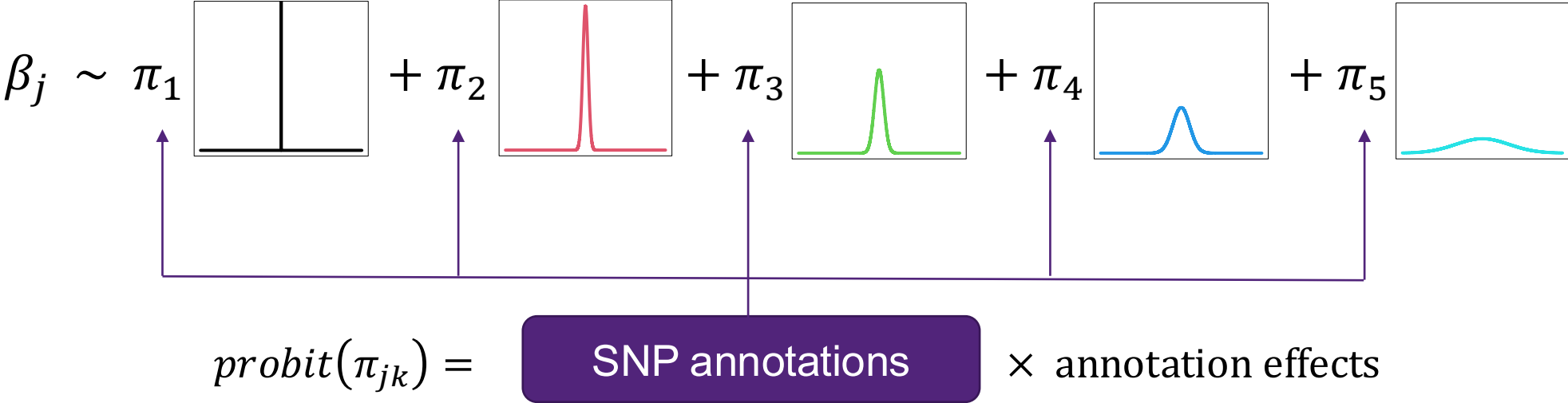
Published online: 30 April 2024

Check for updates

Zhili Zheng<sup>1,2,3</sup>✉, Shouye Liu<sup>1</sup>, Julia Sidorenko<sup>1</sup>, Ying Wang<sup>1</sup>, Tian Lin<sup>1</sup>, Loic Yengo<sup>1</sup>, Patrick Turley<sup>4,5</sup>, Alireza Ani<sup>6,7</sup>, Rujia Wang<sup>6</sup>, Ilja M. Nolte<sup>6</sup>, Harold Snieder<sup>6</sup>, LifeLines Cohort Study<sup>8</sup>, Jian Yang<sup>6,9</sup>, Naomi R. Wray<sup>1,10</sup>, Michael E. Goddard<sup>11,12</sup>, Peter M. Visscher<sup>1,13</sup> & Jian Zeng<sup>1</sup>✉

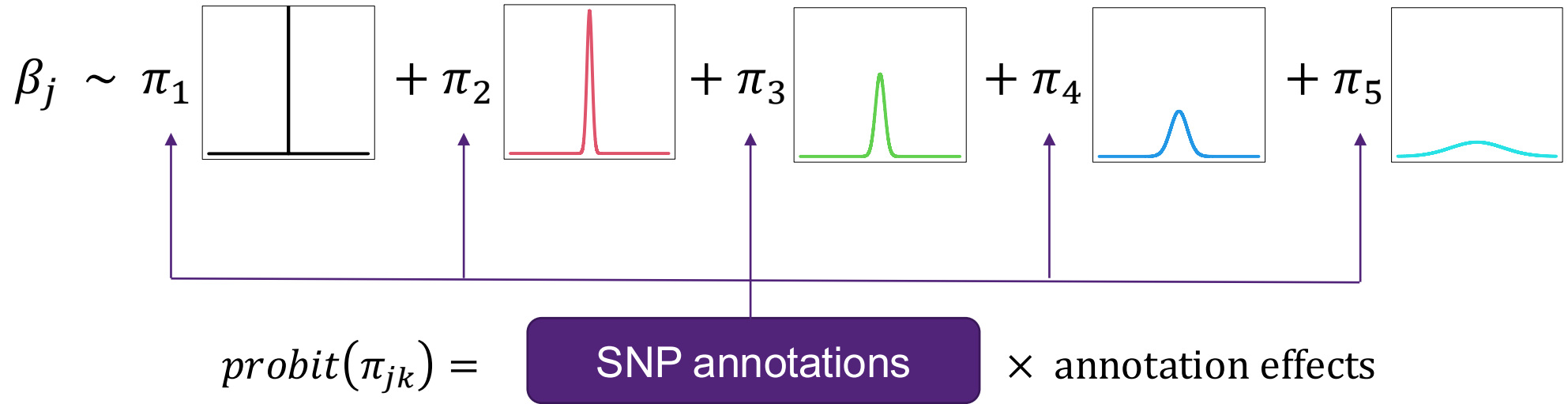
# SBayesRC

Incorporate functional annotations through a hierarchical prior:



# SBayesRC

Incorporate functional annotations through a hierarchical prior:



## Assumption

- Annotation effects are additive at the GLM scale.

## Pros

- Estimation of conditional effects.
- Allow annotation overlap.
- Interpretation.

## Cons

- # annotation effect parameters x 5.
- $\pi_{j_1} + \pi_{j_2} + \pi_{j_3} + \pi_{j_4} + \pi_{j_5} = 1$ .

# Reparameterisation of annotation effects

Suppose 4 components for simplicity

- A set of 2-component independent models:

- For all SNPs

$$\beta_j \sim (1 - p_2) \left[ \begin{array}{|c|c|} \hline & \\ \hline \end{array} \right] + p_2 \left[ \begin{array}{|c|c|c|} \hline \text{red peak} & \text{green peak} & \text{blue peak} \\ \hline \end{array} \right]$$

- For SNPs with nonzero effects (conditional on non-null SNPs)

$$\beta_j \sim (1 - p_3) \left[ \begin{array}{|c|} \hline \text{red peak} \\ \hline \end{array} \right] + p_3 \left[ \begin{array}{|c|c|} \hline \text{green peak} & \text{blue peak} \\ \hline \end{array} \right]$$

- For SNPs with at least medium effects (conditional on non-small-effect SNPs)

$$\beta_j \sim (1 - p_4) \left[ \begin{array}{|c|} \hline \text{green peak} \\ \hline \end{array} \right] + p_4 \left[ \begin{array}{|c|} \hline \text{blue peak} \\ \hline \end{array} \right]$$

*p<sub>2</sub>, p<sub>3</sub>, p<sub>4</sub> are independent!*

# Reparameterisation of annotation effects

- Probit link function:

$$\Phi^{-1}(p) = \sum \text{SNP annotation} \times \text{annotation effect}$$

where  $\Phi$  is the CDF of the standard normal distribution.

- It is straightforward to compute  $p = \Phi(\cdot)$

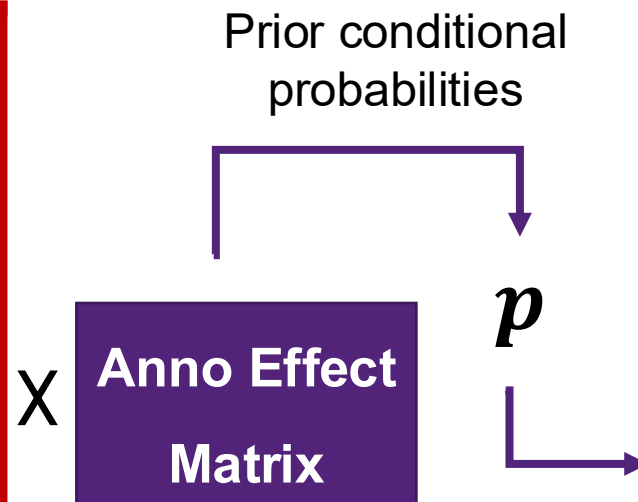
and  $\pi_1 = 1 - p_2$ ;  $\pi_2 = (1 - p_3)p_2$ ;  $\pi_3 = (1 - p_4)p_3p_2$ ;  $\pi_4 = p_2p_3p_4$

- Assume a normal prior distribution for each annotation effect.
- Gibbs sampling for all parameters.

# Toy example: inputs

	Genome	Region 1	Region 2	Region 3
SNP 1	1	1	0	0
SNP 2	1	0	1	0
SNP 3	1	1	1	0
SNP 4	1	0	0	1
SNP 5	1	1	0	0

Input data



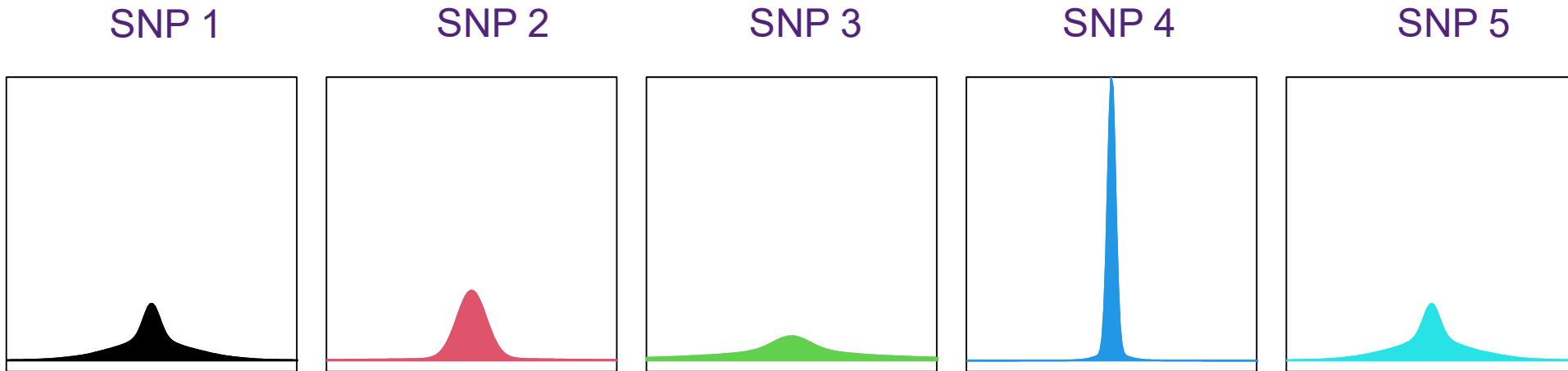
	$\pi_1$	$\pi_2$	$\pi_3$	$\pi_4$
SNP 1	0.2	0.1	0.6	0.1
SNP 2	0.8	0.02	0.02	0.16
SNP 3	0.2	0.0	0.2	<b>0.6</b>
SNP 4	0.9	0.08	0.01	0.01
SNP 5	0.2	0.1	0.6	0.1

prior mixing probabilities

# Toy example: consequences

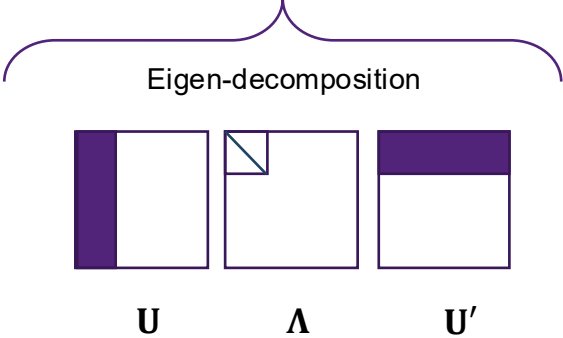
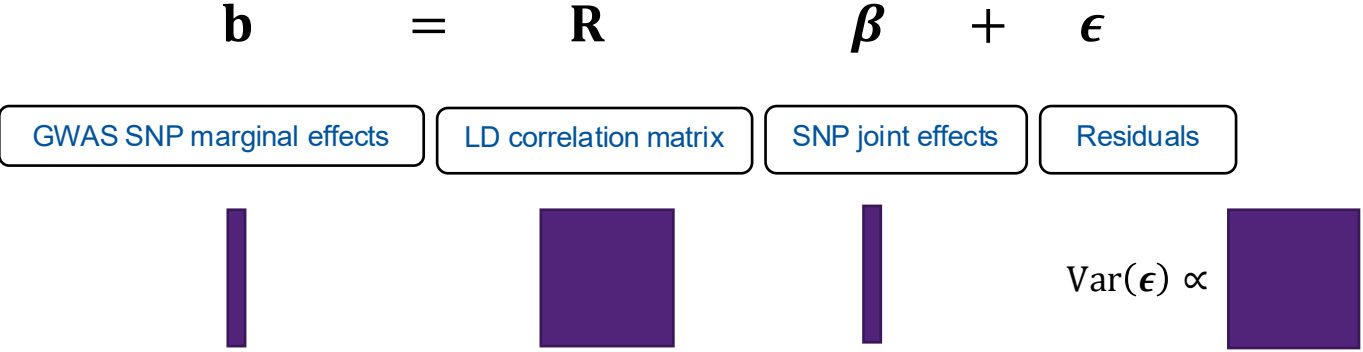
Prior distribution of SNP effect is annotation dependent.

	$\pi_1$	$\pi_2$	$\pi_3$	$\pi_4$
SNP 1	0.2	0.1	0.6	0.1
SNP 2	0.8	0.02	0.02	0.16
SNP 3	0.2	0.0	0.2	<b>0.6</b>
SNP 4	0.9	0.08	0.01	0.01
SNP 5	0.2	0.1	0.6	0.1



# Low-rank model (fits ultra high-density SNPs)

In each quasi-independent LD block:



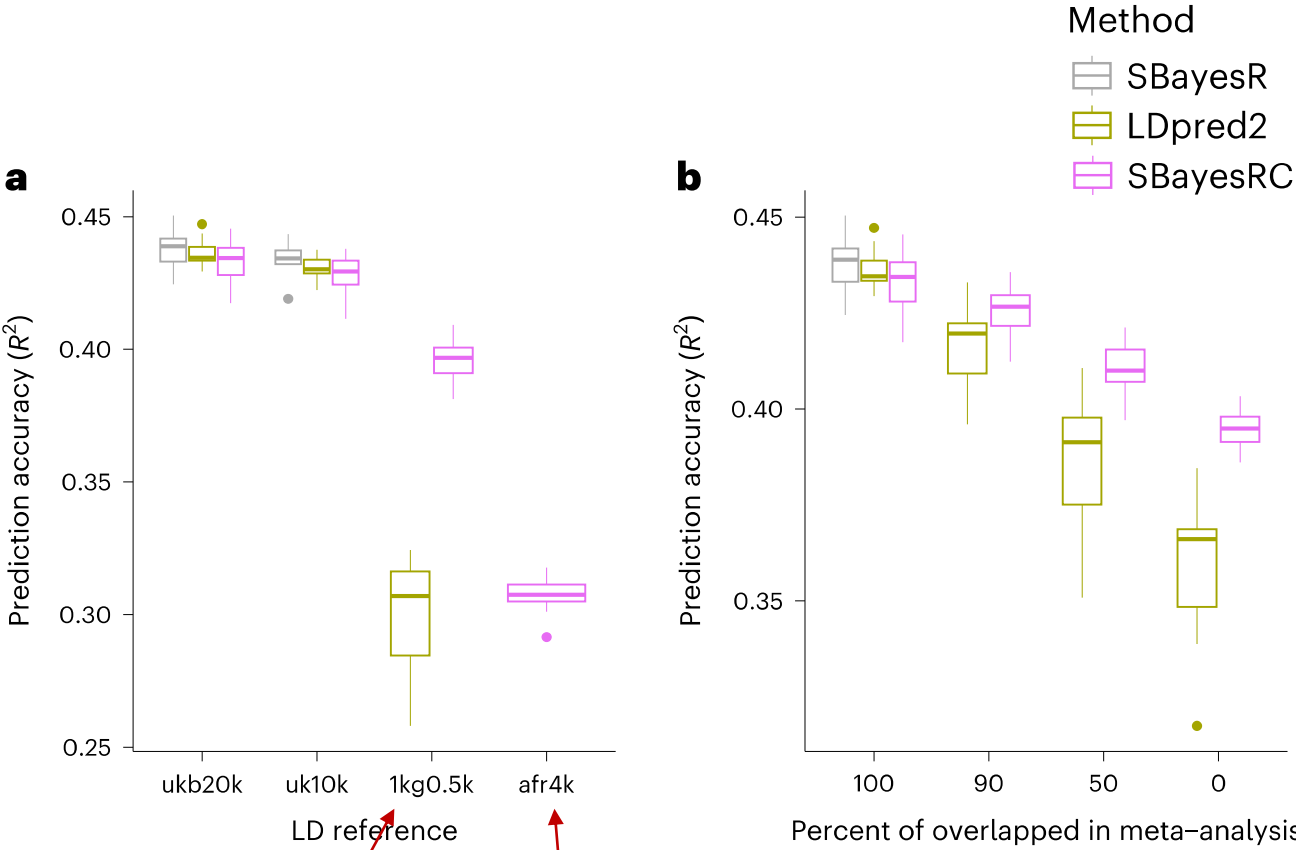
$$\Lambda^{-\frac{1}{2}} \mathbf{U}' \mathbf{b} = \Lambda^{\frac{1}{2}} \mathbf{U}' \boldsymbol{\beta} + \Lambda^{-\frac{1}{2}} \mathbf{U}' \boldsymbol{\epsilon}$$



*It only requires the top 20% PCs to explain 99.5% of the variance in LD!*

# Low-rank model

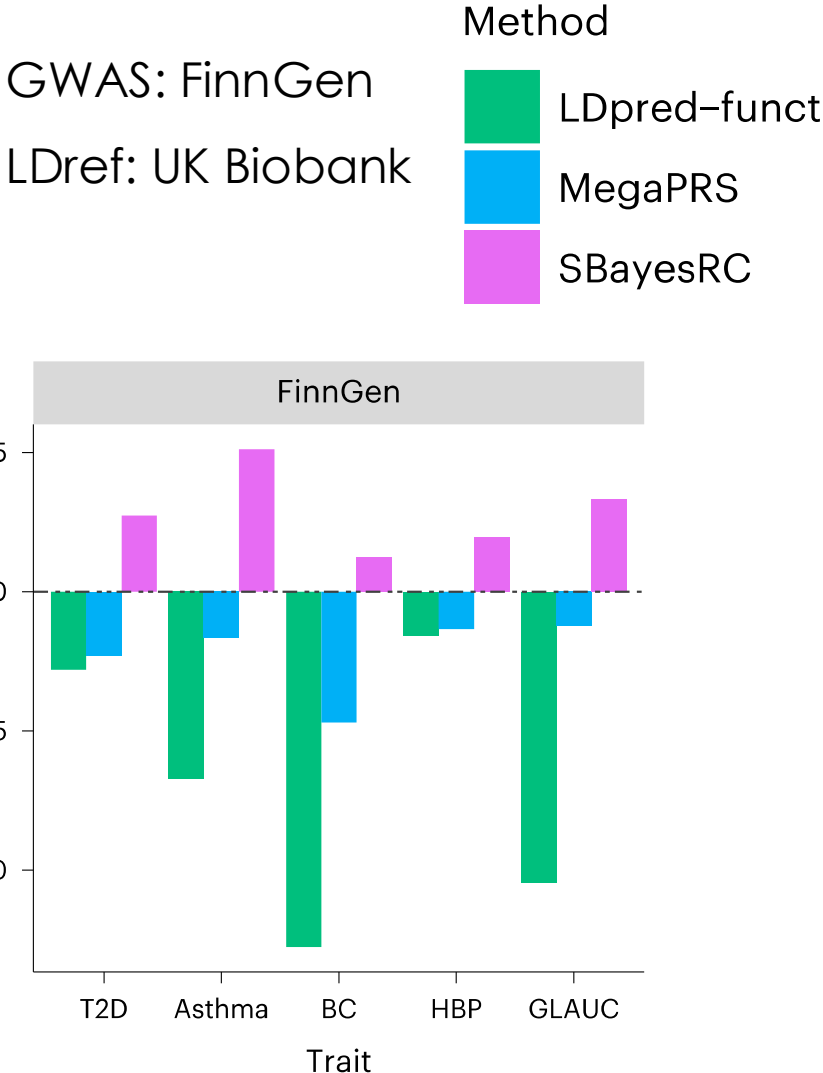
## Improved robustness



Very small LD ref sample size

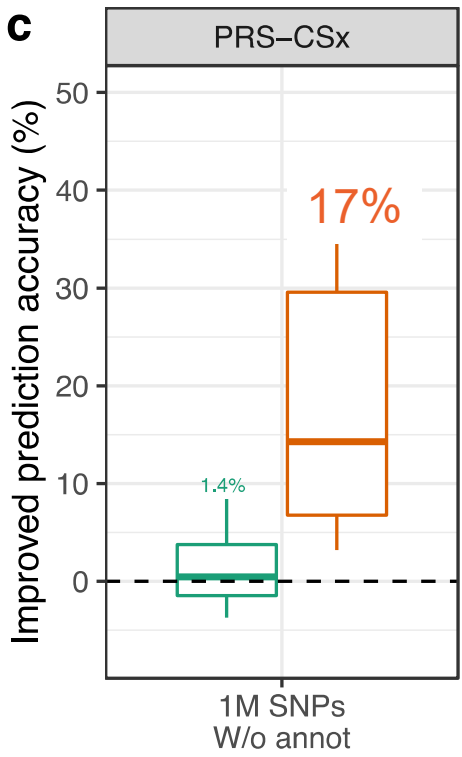
Different ancestry

Missing SNPs in sub cohorts



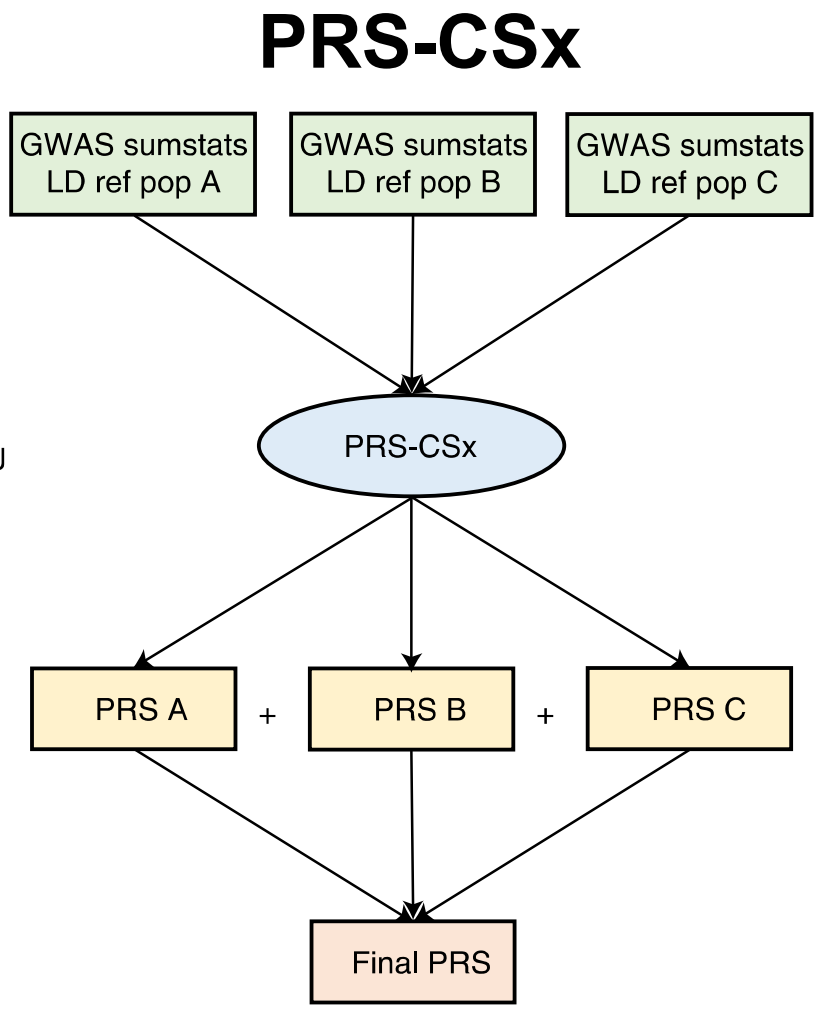
# Trans-ancestry prediction

Use GWAS data from UKB EUR and BBJ EAS to predict UKB EAS



Train dataset

- UKB-EUR
- UKB-EUR + BBJ



**nature genetics**

Explore content ▾ About the journal ▾ Publish with us ▾ Subscribe

[nature](#) > [nature genetics](#) > [articles](#) > [article](#)

Article | Published: 05 May 2022

### Improving polygenic prediction in ancestrally diverse populations

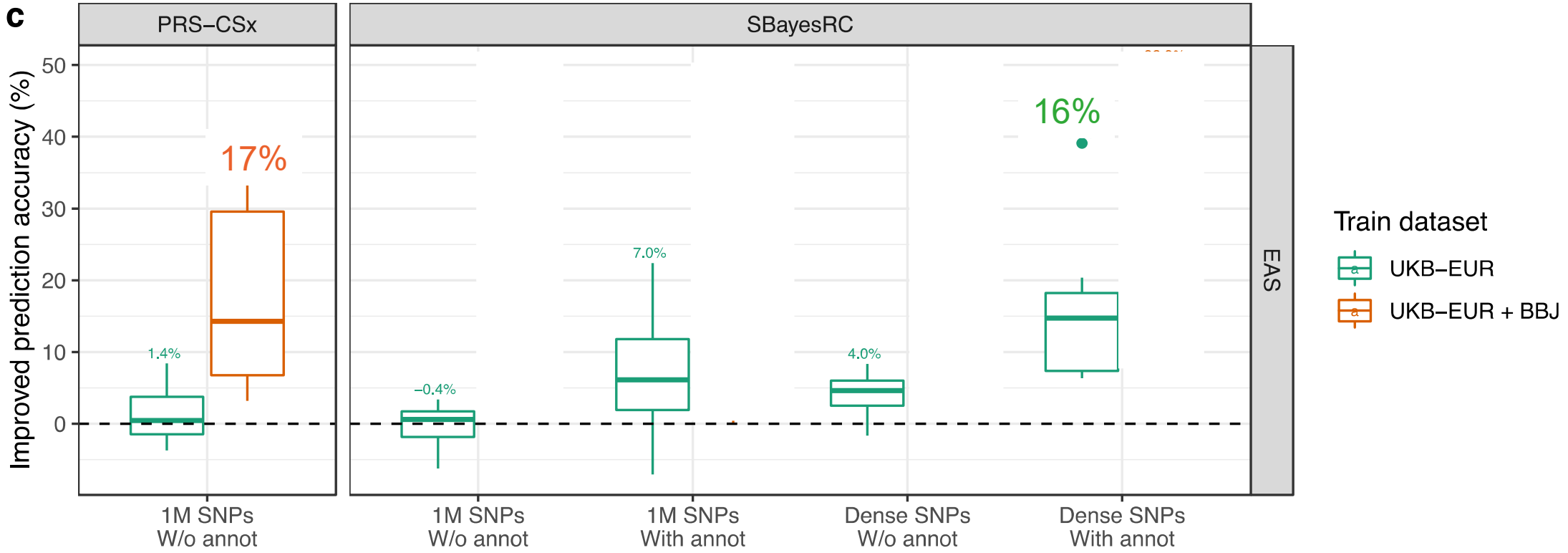
[Yunfeng Ruan](#), [Yen-Feng Lin](#), [Yen-Chen Anne Feng](#), [Chia-Yen Chen](#), [Max Lam](#), [Zhenglin Guo](#), [Stanley Global Asia Initiatives](#), [Lin He](#), [Akira Sawa](#), [Alicia R. Martin](#), [Shengying Qin](#) ✉, [Hailiang Huang](#) ✉ & [Tian Ge](#) ✉

*Nature Genetics* 54, 573–580 (2022) | [Cite this article](#)

How important is functional annotation data compare to another GWAS dataset from the target ancestry?

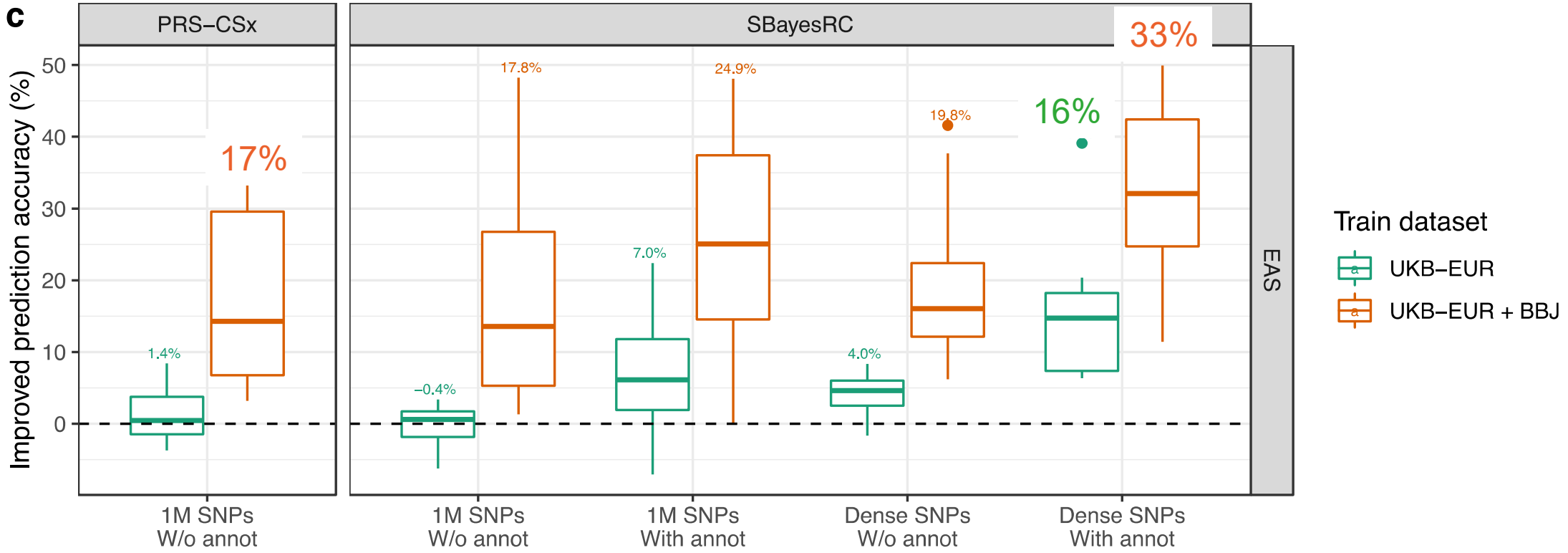
# Trans-ancestry prediction

Use GWAS data from UKB EUR and BBJ EAS to predict UKB EAS



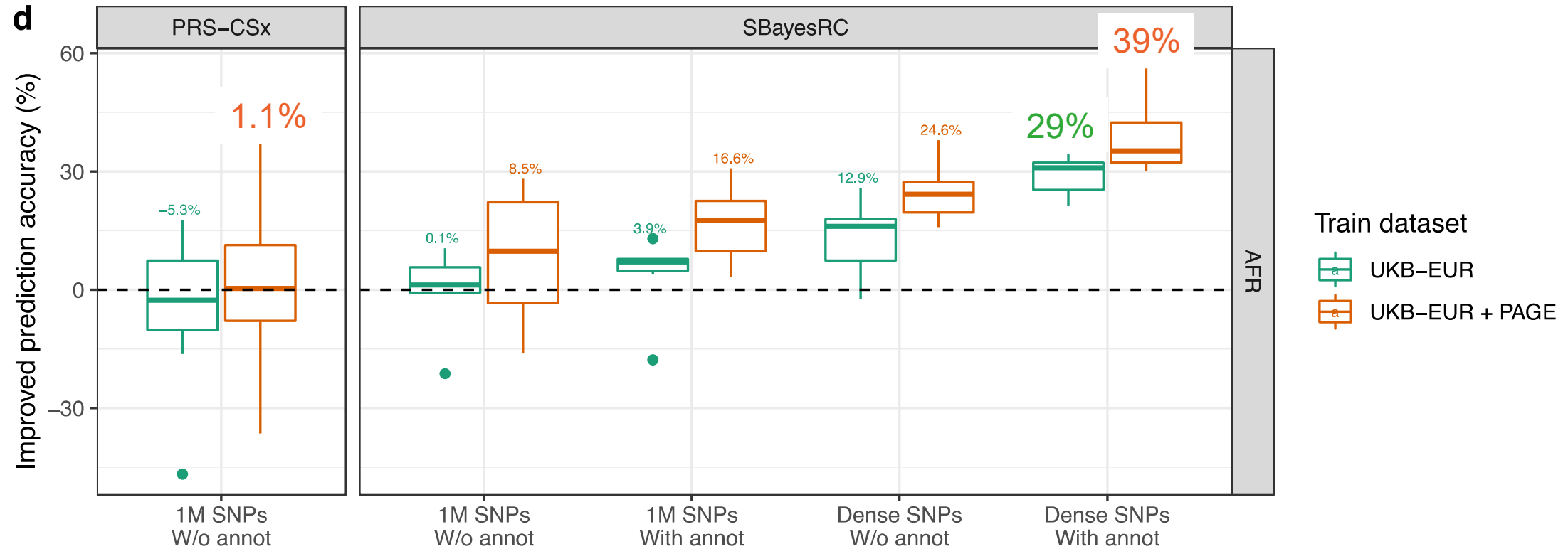
# Trans-ancestry prediction

Use GWAS data from UKB EUR and BBJ EAS to predict UKB EAS

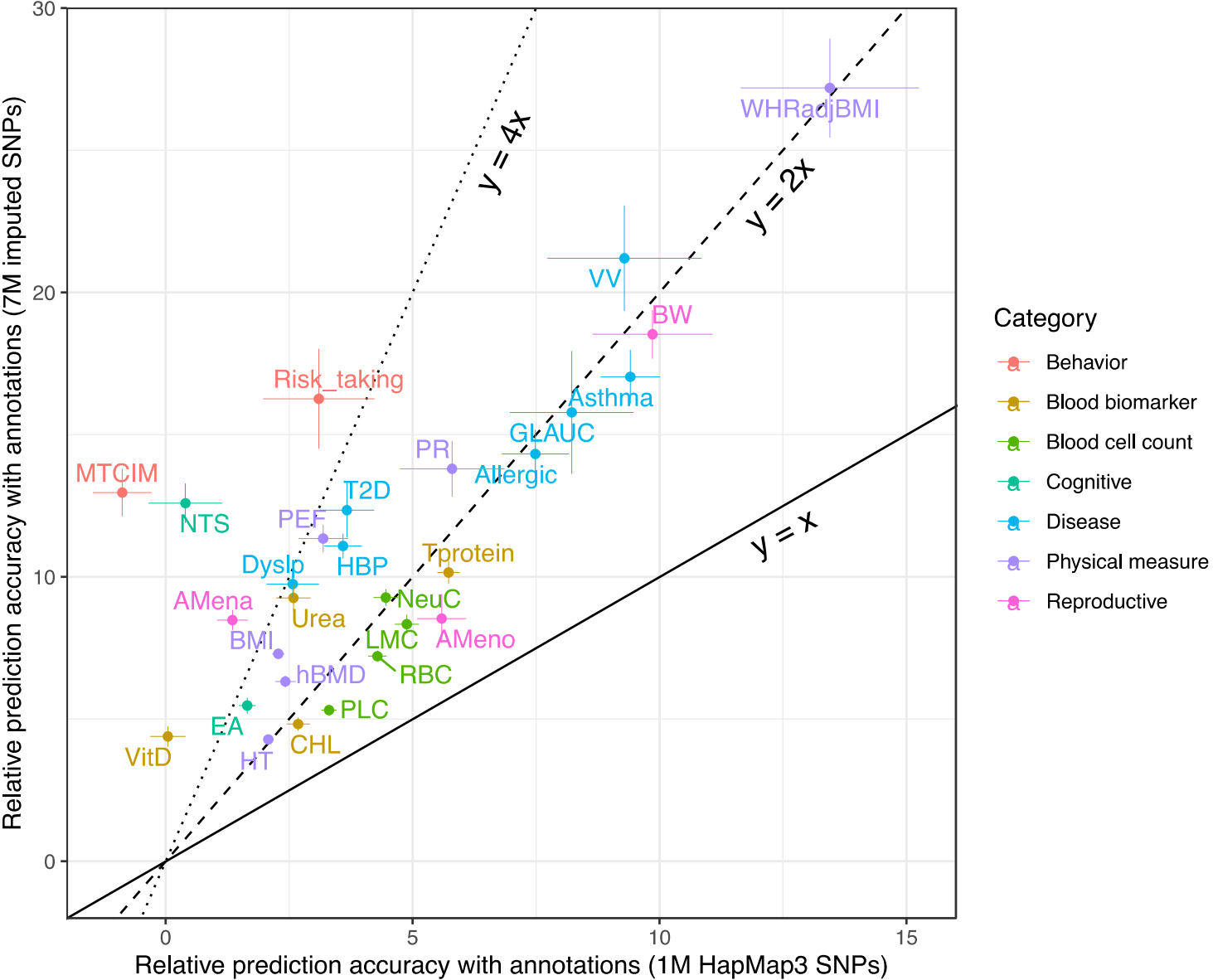


# Trans-ancestry prediction

Use GWAS data from UKB EUR and PAGE (mixed) AFR to predict UKB AFR



# Interaction between SNP density and annotation information



Improvement (%) in prediction accuracy with vs. without annotations:

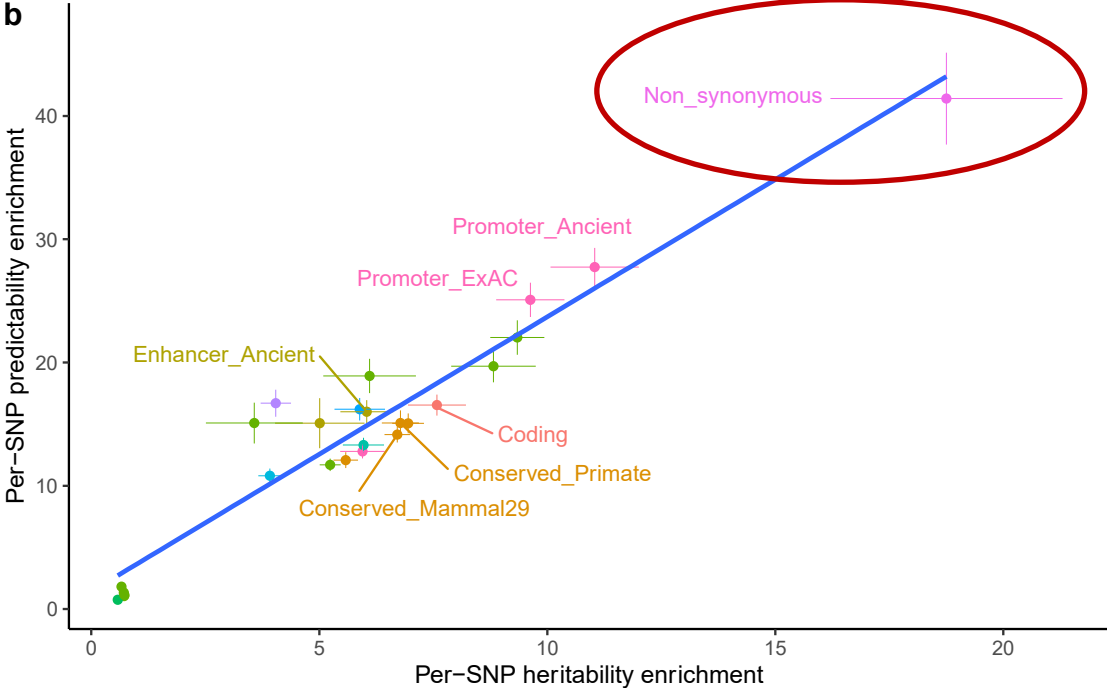
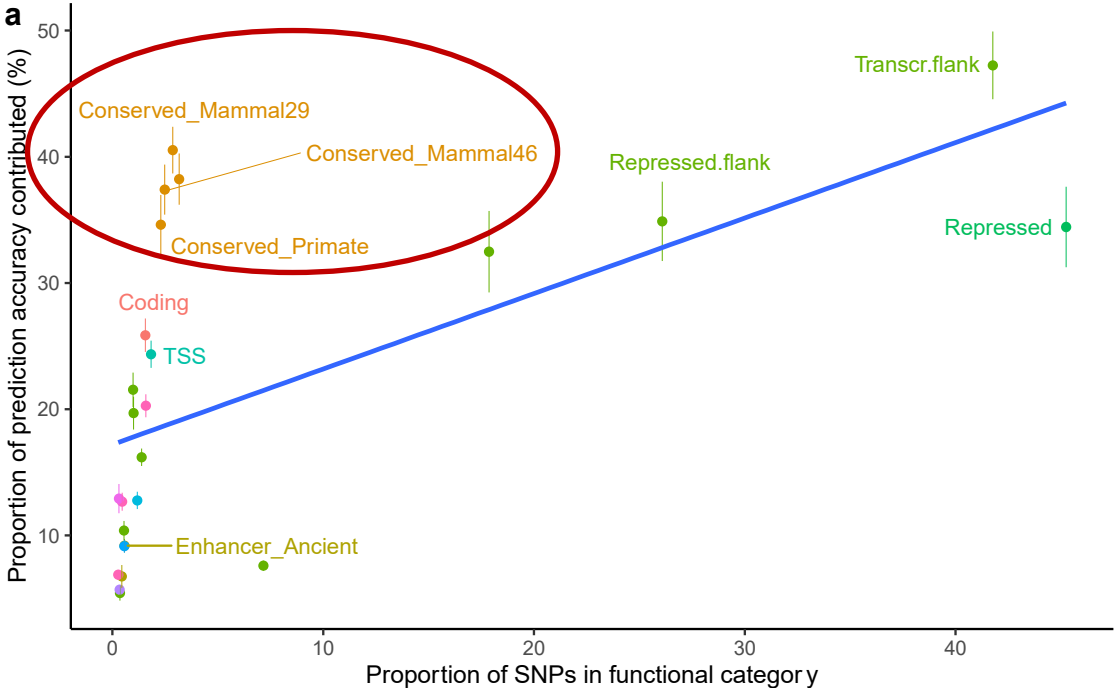
$$\frac{R^2_{\text{annot}} - R^2_{\text{wo}}}{R^2_{\text{wo}}}$$

using 7M imputed SNPs (y-axis) or 1M HapMap3 SNPs (x-axis).

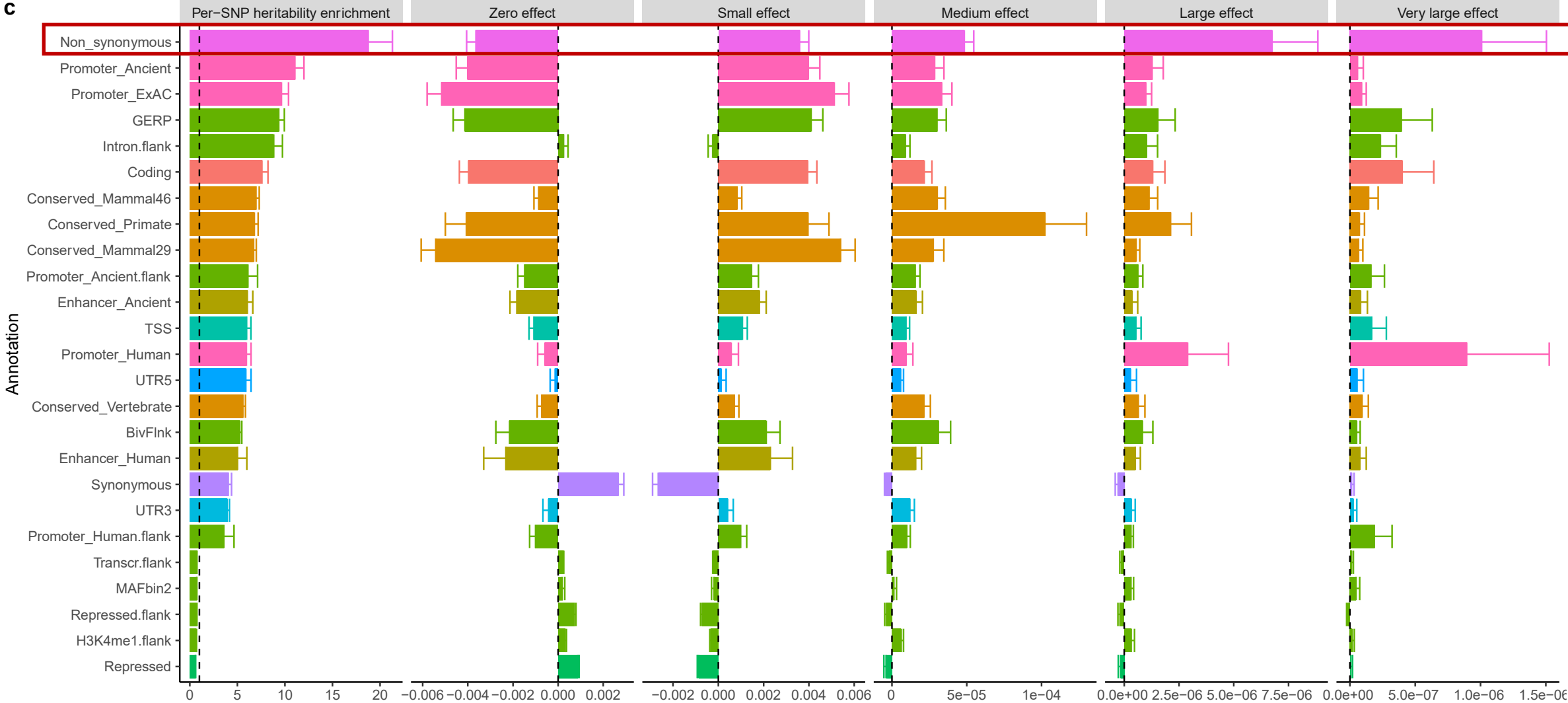
**Annotations help more with increased SNP density**

# Contributions of functional categories to prediction accuracy

Regions conserved across 29 mammals covers 3% genome but contributed 41% prediction accuracy!

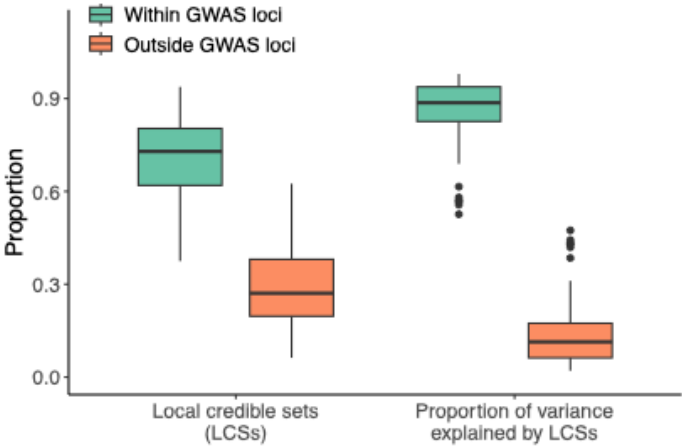
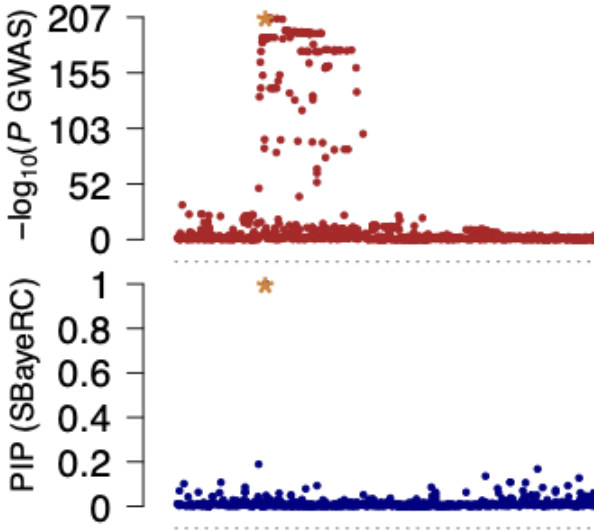


# Functional genetic architecture



# Incorporate annotations to improve fine-mapping

## Genome-wide fine-mapping



nature genetics



Article

<https://doi.org/10.1038/s41588-026-02549-3>

## Genome-wide fine-mapping improves identification of causal variants

Received: 20 July 2024

Accepted: 16 February 2026

Published online: 30 March 2026

Yang Wu<sup>1,2</sup>, Zhili Zheng<sup>2,3,4</sup>, Loic Thibaut<sup>2</sup>, Tian Lin<sup>2</sup>, Qian Feng<sup>2</sup>, Hao Cheng<sup>5</sup>, Loic Yengo<sup>2</sup>, Michael E. Goddard<sup>6,7</sup>, Naomi R. Wray<sup>2,8</sup>, Peter M. Visscher<sup>2,9</sup> & Jian Zeng<sup>2</sup>

### GCTB

A tool for Genome-wide Complex Trait Bayesian analysis

[GCTA](#) [SMR](#) [GSMR](#) [OSCA](#) **[GCTB](#)** [Program in CTG](#) [CTG forum](#)

- [Overview](#)
- [Download](#)
- [Basic options](#)
- [Bayesian alphabet](#)
- [Summary Bayesian Alphabet](#)
- [Tutorial](#)
  - [SBayesR Tutorial](#)
  - [SBayesRC Tutorial](#)
  - [Genome-wide Fine-mapping analysis](#)
- [FAQ](#)

#### Genome-wide Fine-mapping analysis

The Genome-wide Bayesian Mixture Model (GBMM) implemented in GCTB (e.g., SBayesRC) can perform genome-wide fine-mapping analysis. These methods require summary-level data from genome-wide association studies (GWAS) and linkage disequilibrium (LD) data from a reference sample. Our manuscript is currently under review and available at here (link to manuscript).

We outline below on how to perform the genome-wide fine-mapping (GWFM) analysis and calculate the credible set using GCTB.

#### Run genome-wide fine-mapping analysis

```
gctb --gwfm RC --ldm-eigen ldm --gwas-summary test.ma --annot annot.txt --gene-map gene_map.txt --thread 32 --out test
```

# Summary

## Methodology

- Incorporate functional annotations to better capture causal effects (**improved accuracy**).
- Develop a low-rank method that fits all SNPs to better model LD (**more robust & efficient**).
- Can apply to individual-level data and be used as a fine-mapping method.

## Science

- For trans-ancestry prediction, functional annotations provide **comparable and additive information** to the use of additional GWAS dataset of target ancestry.
- Significant **interaction** between SNP density and annotation information, suggesting whole-genome sequence variants with annotations may further improve prediction.
- Functional partitioning highlights a major contribution of **evolutionary constrained regions** to prediction accuracy and the largest per-SNP contribution from non-synonymous SNPs.

# Recommended reading

1. Zheng Z, *et al.* Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. *Nat Genet* 56, 767–777 (2024). (**SBayesRC for prediction**)
2. Wu Y, *et al.* Genome-wide fine-mapping improves identification of causal variants. *Nat Genet* 58, 940–951 (2026). (**SBayesRC for fine-mapping**)
3. Zeng J, Visscher PM. Harnessing functional annotation to improve the accuracy and transferability of polygenic scores. *Nat Rev Genet* 26, 805–806 (2025). (**Comment and vision**)