THE UNIVERSITY OF QUEENSLAND AUSTRALIA | CREATE CHANGE

# SBayesRC: Polygenic Prediction Incorporating functional annotations

Jian Zeng

j.zeng@uq.edu.au

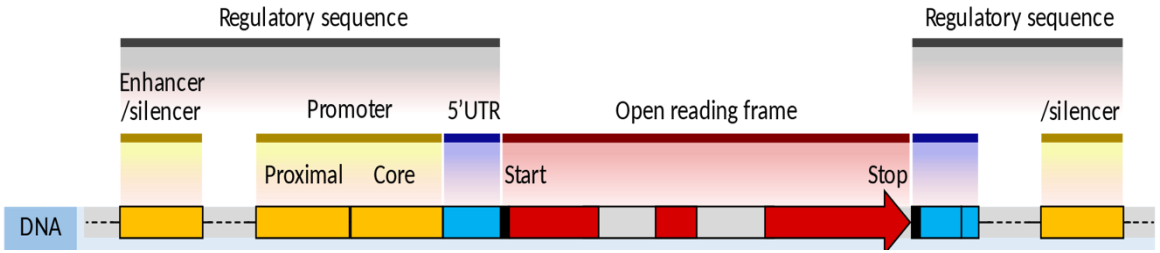
THE UNIVERSITY OF QUEENSLAND AUSTRALIA | Institute for Molecular Bioscience

Program in Complex Trait Genomics

# Functional genomic annotations

Functional genomic annotations provide orthogonal information useful for polygenic prediction.

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



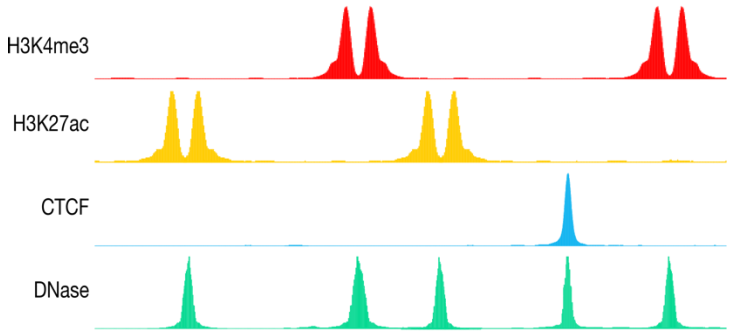



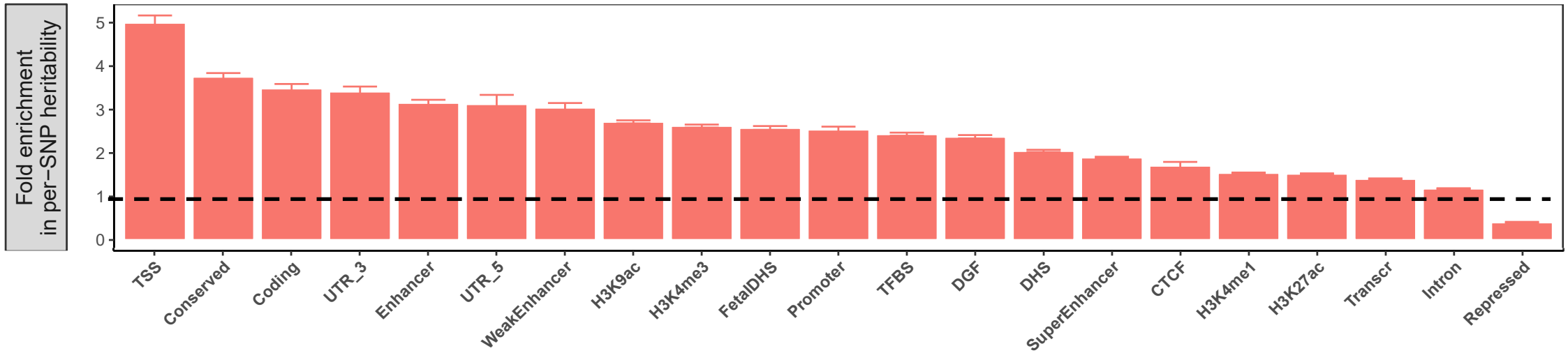

Image from ENCODE

# Functional genomic annotations

Functional genomic annotations provide orthogonal information useful for polygenic prediction.

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

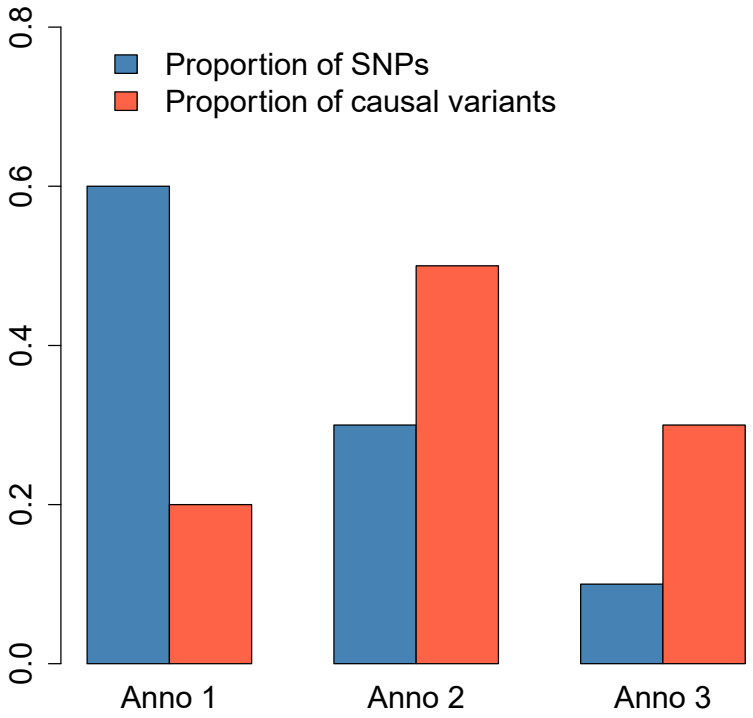
Zeng et al 2021 Nature Communications



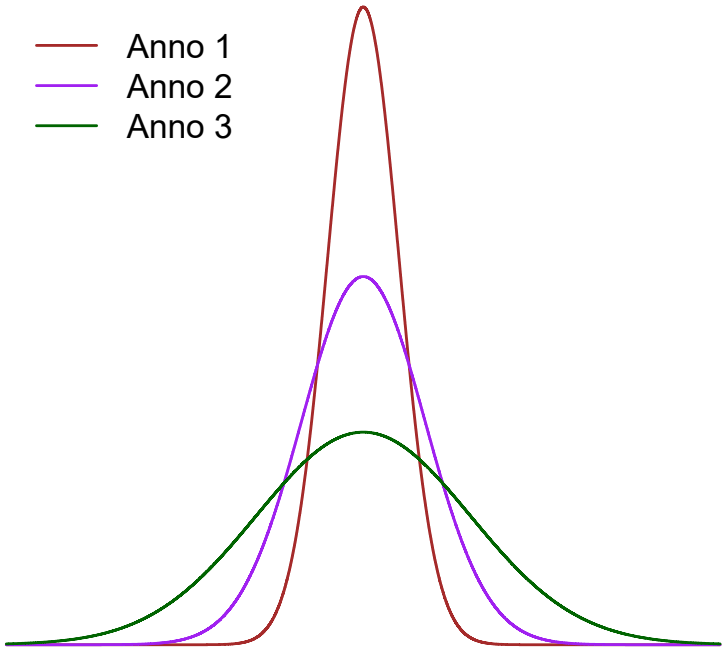
# Opportunities/challenges

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

Differences in proportion of causal variants

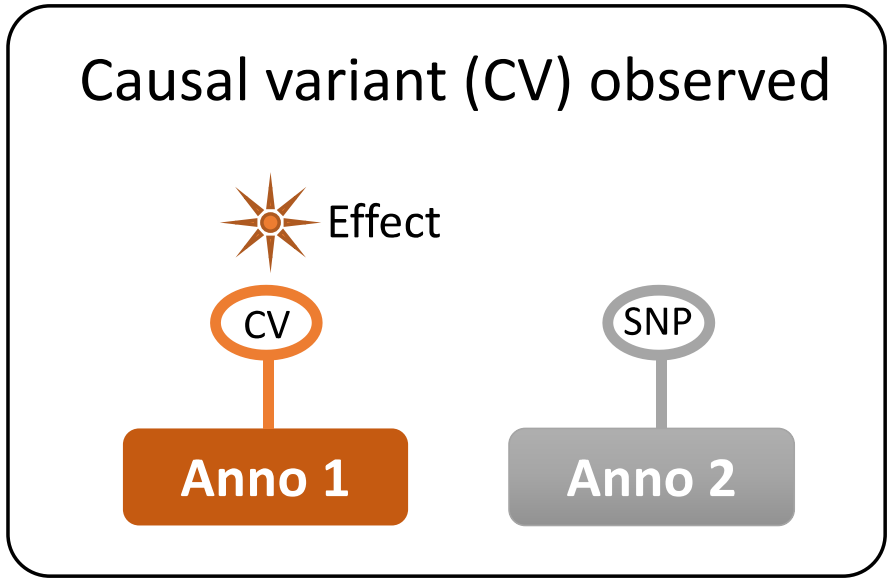


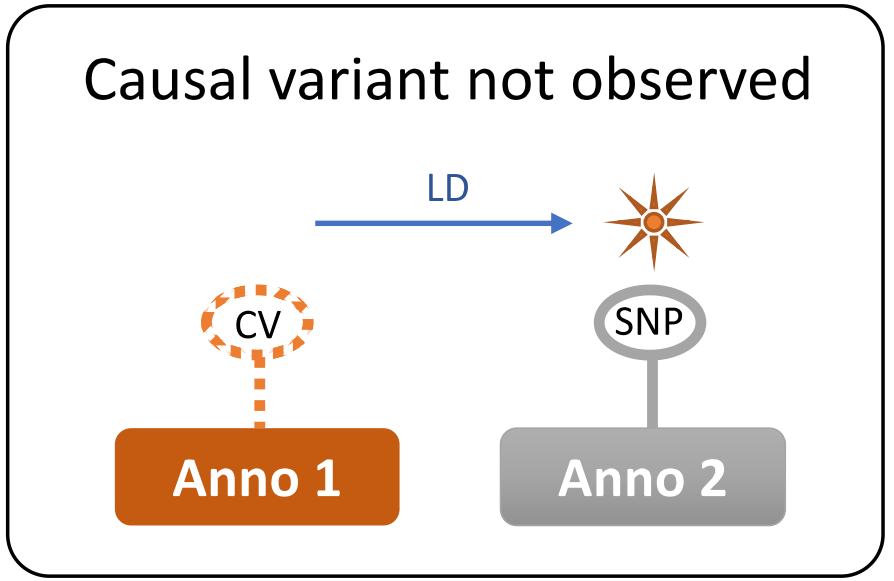

Differences in distribution of causal effects



# 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!**

# Literature

The University Of Queensland AUSTRALIA

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

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

## 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

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

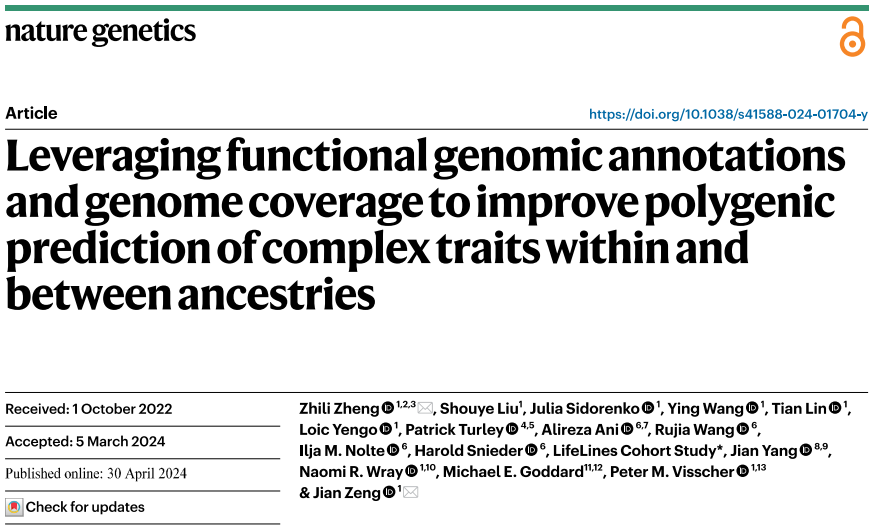
PolyPred

# Gaps

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

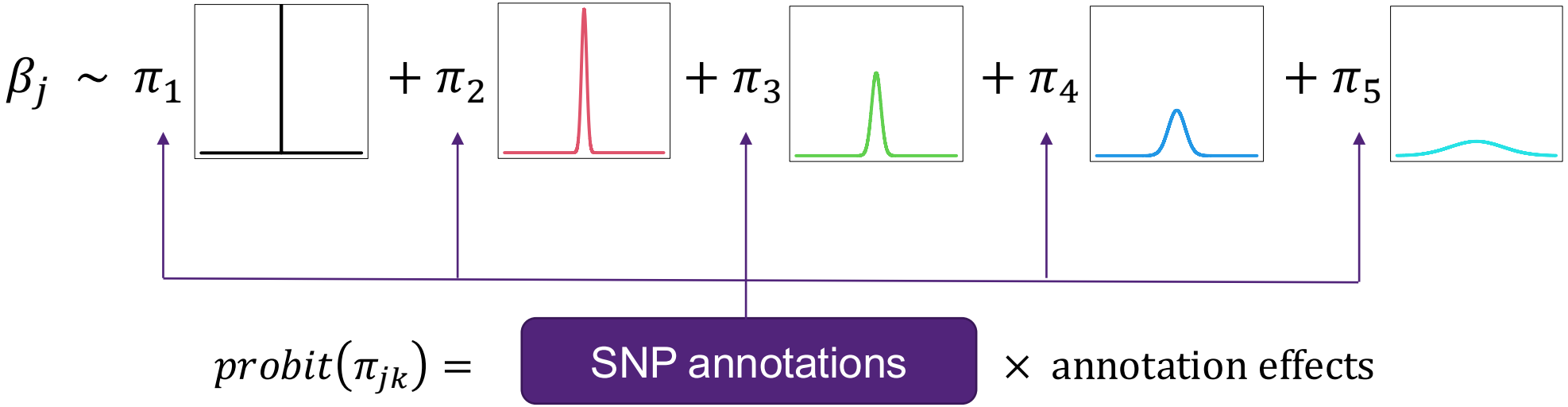
Published online: 30 April 2024

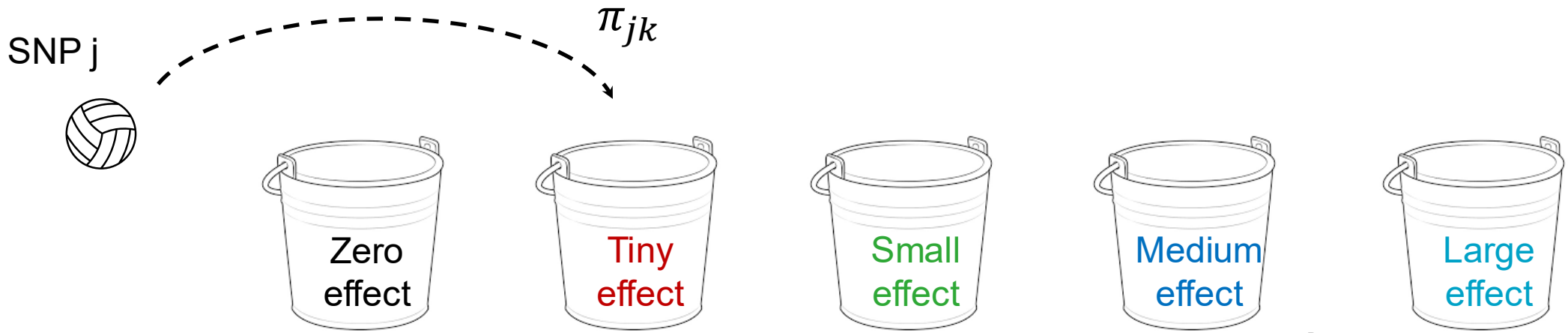
Check for updates

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

# SBayesRC

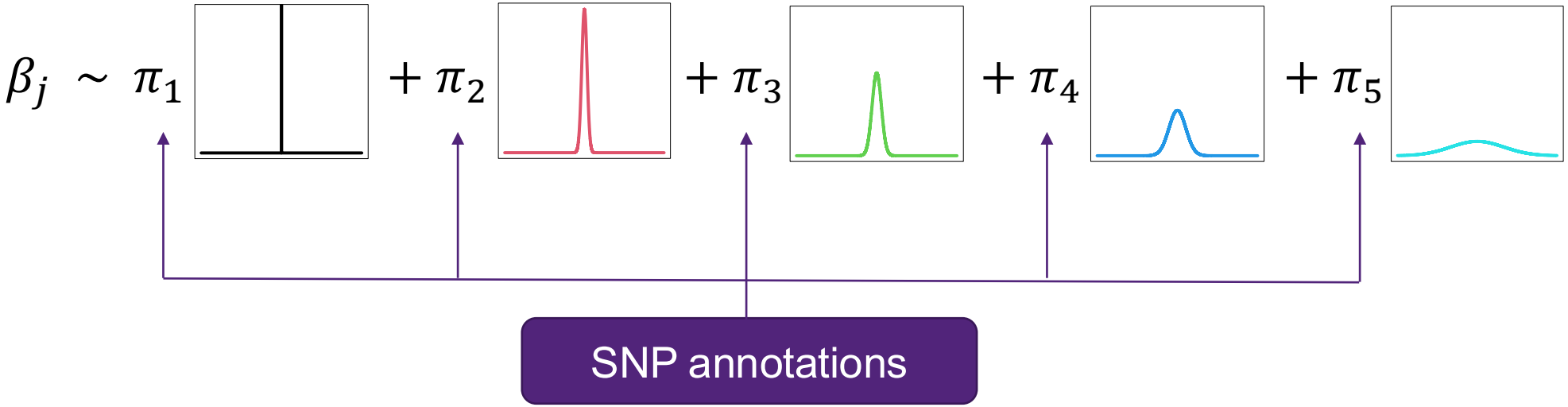
Incorporate functional annotations through a hierarchical prior:



# SBayesRC

Incorporate functional annotations through a hierarchical prior:

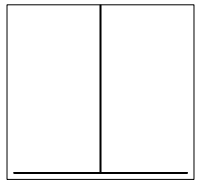
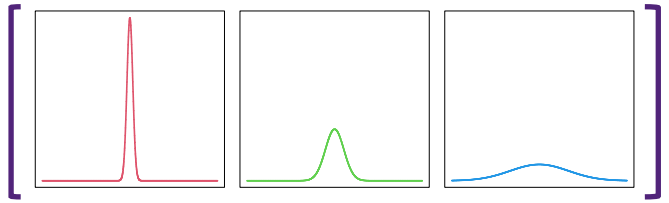



| Assumption | Pros | Cons |
|---|---|---|
| • Annotation effects are additive at the GLM scale. | • Estimation of conditional effects.<br>• Allow annotation overlap.<br>• Interpretation. | |

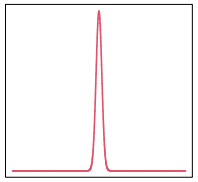
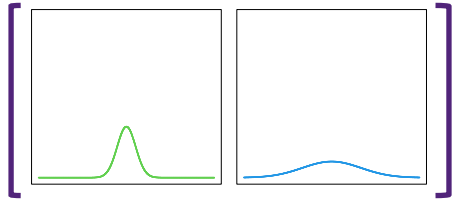
# Reparameterisation of annotation effects

## Suppose 4 components for simplicity

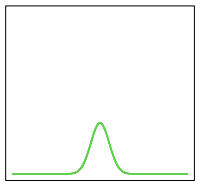
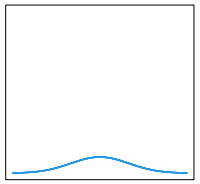
- A set of 2-component independent models:
- For all SNPs

$$\beta_j \sim (1-p_2)$$  $$+\ p_2$$ 

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

$$\beta_j \sim (1-p_3)$$  $$+\ p_3$$ 

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

$$\beta_j \sim (1-p_4)$$  $$+\ p_4$$ 

$p_2, p_3, p_4$ 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

| | 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

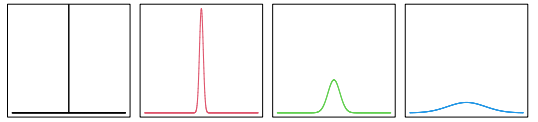
X

**Anno Effect Matrix**

Estimate from the data

Prior conditional probabilities

$\boldsymbol{p}$



| | $\pi_1$ | $\pi_2$ | $\pi_3$ | $\pi_4$ |
|---|---|---|---|---|
| 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 | **0.6** |
| 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

Prior distribution of SNP effect is annotation dependent.

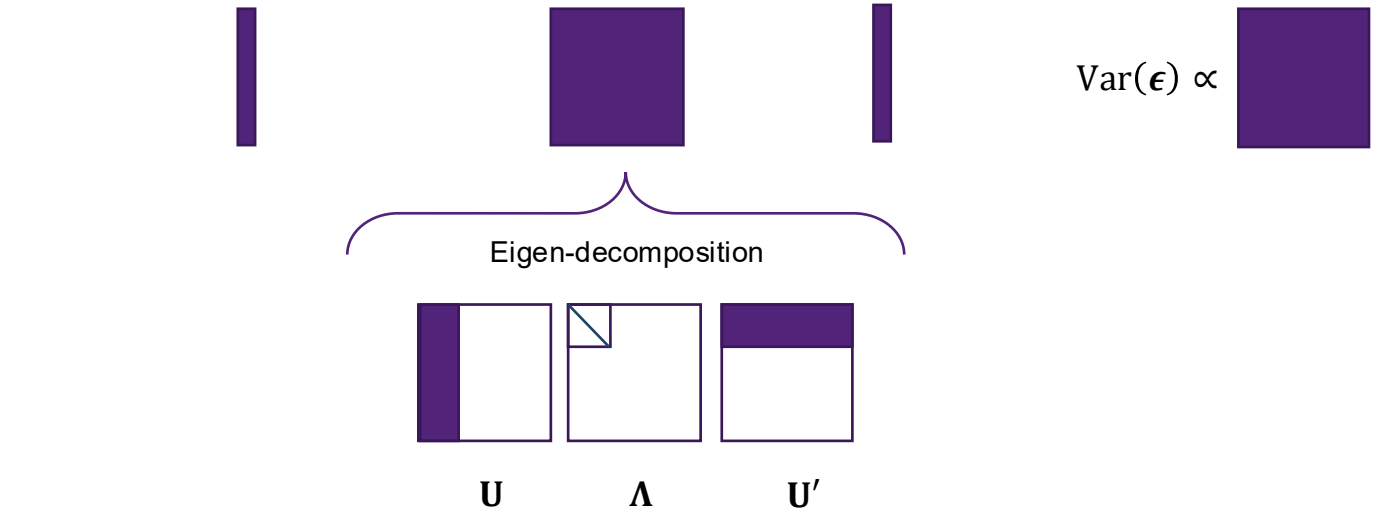
| | $\pi_1$ | $\pi_2$ | $\pi_3$ | $\pi_4$ |
|---|---|---|---|---|
| **SNP 1** | 0.2 | 0.1 | 0.6 | 0.1 |
| **SNP 2** | 0.8 | 0.02 | 0.02 | 0.16 |
| **SNP 3** | 0.2 | 0.0 | 0.2 | **0.6** |
| **SNP 4** | 0.9 | 0.08 | 0.01 | 0.01 |
| **SNP 5** | 0.2 | 0.1 | 0.6 | 0.1 |

# Low-rank model (fits 7M SNPs or more)

In each quasi-independent LD block:

$$\mathbf{b} = \mathbf{R}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

GWAS SNP marginal effects | LD correlation matrix | SNP joint effects | Residuals



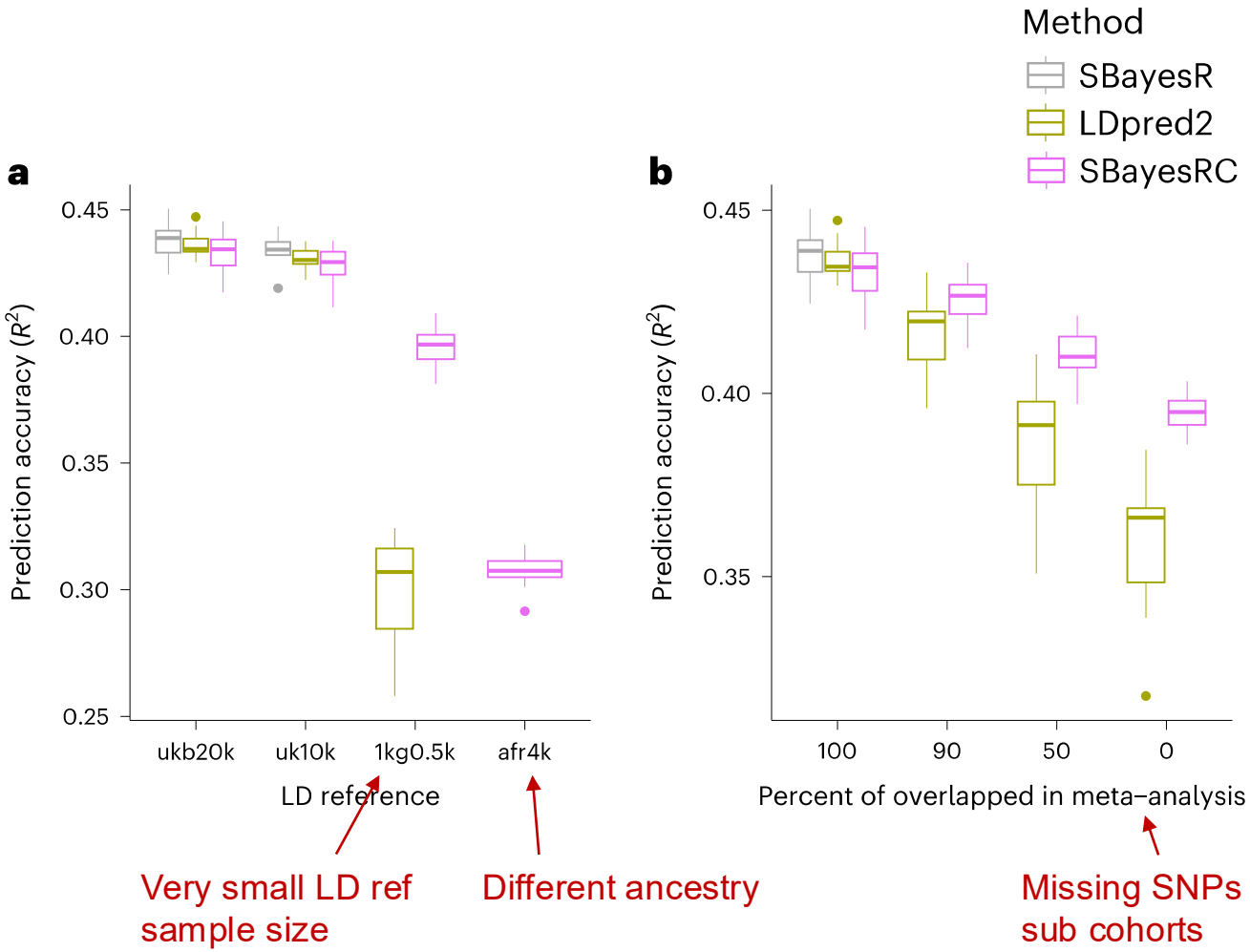

$$\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}'\mathbf{b} = \boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}'\boldsymbol{\beta} + \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}'\boldsymbol{\epsilon}$$

$$\mathbf{w} = \mathbf{Q}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$


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

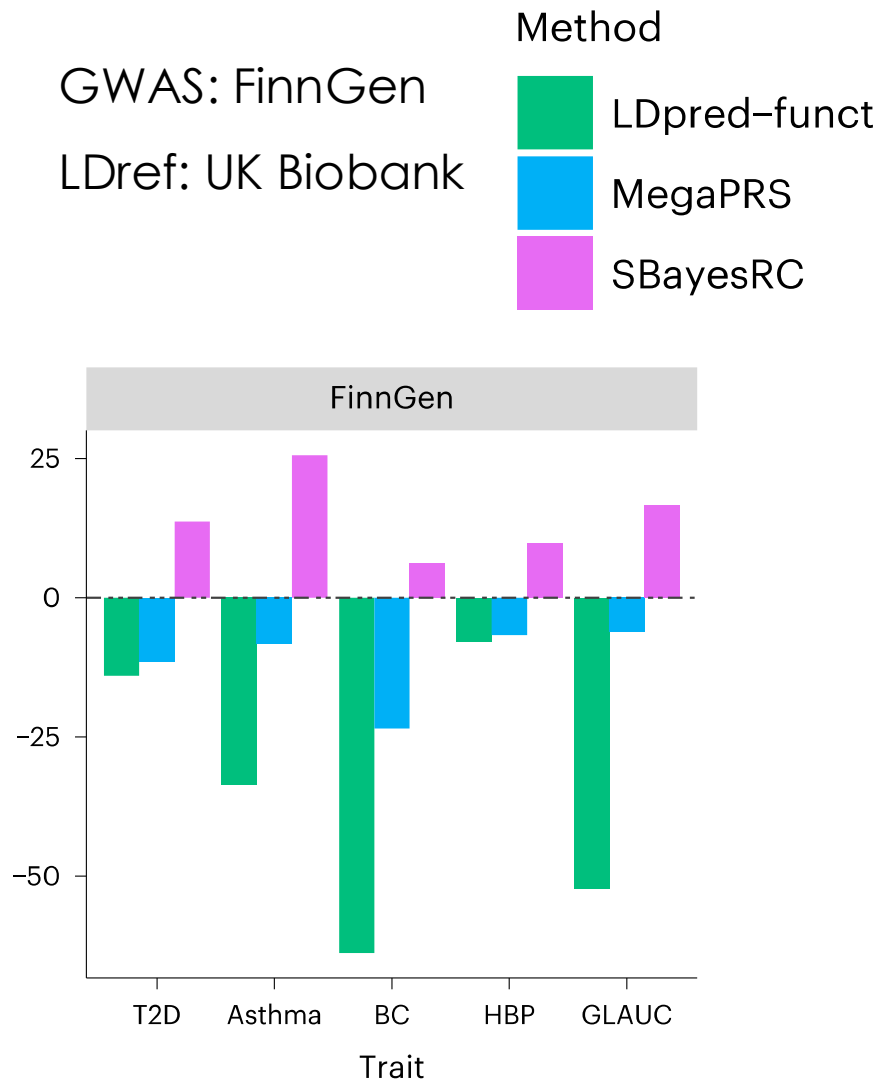
# Low-rank model

## Improved robustness



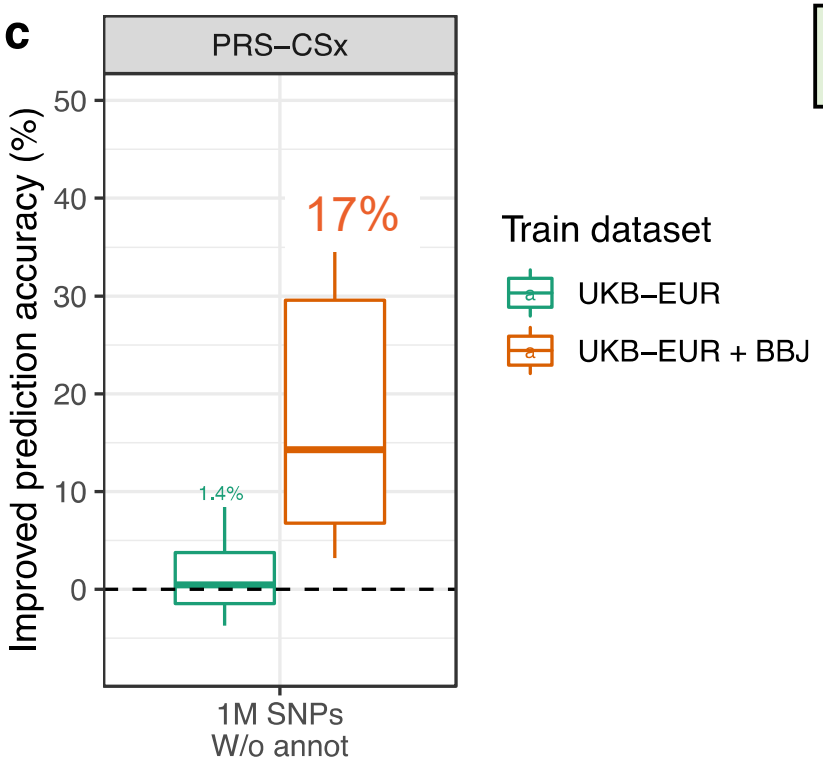

GWAS: FinnGen

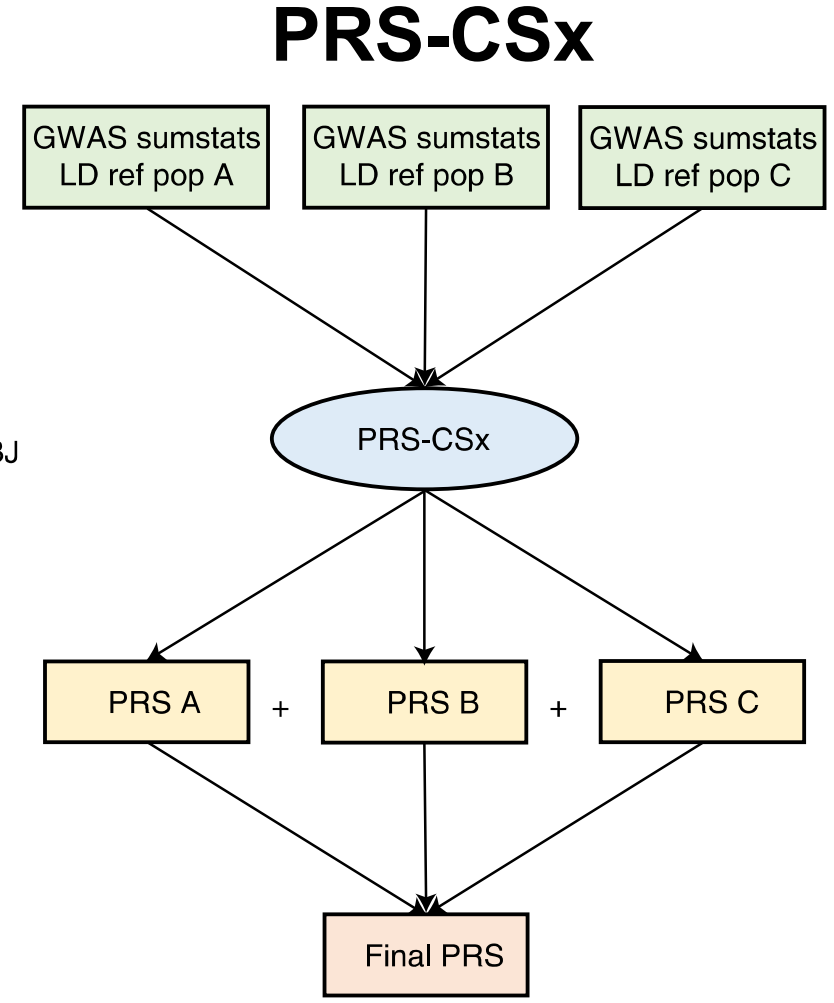
LDref: UK Biobank

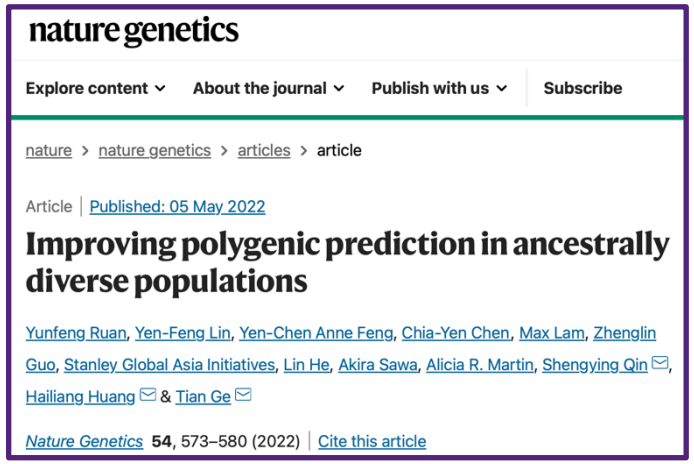


# Trans-ancestry prediction

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



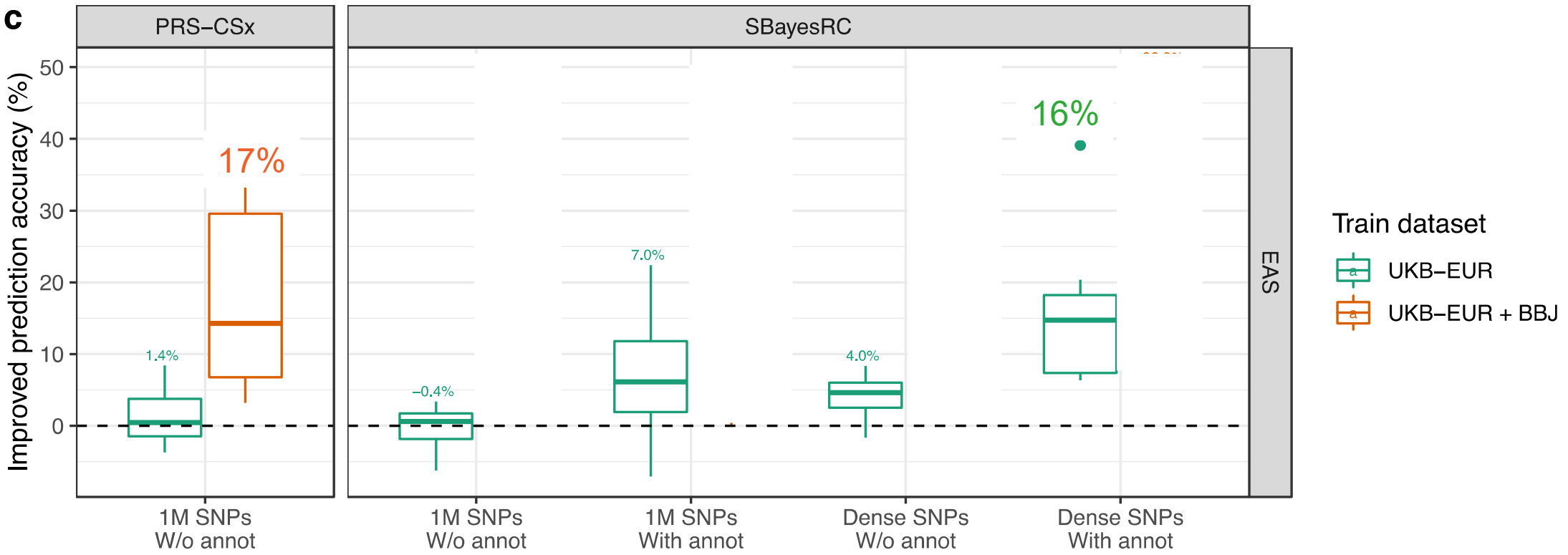




How important is functional annotation data compare to another GWAS dataset from the target ancestry?

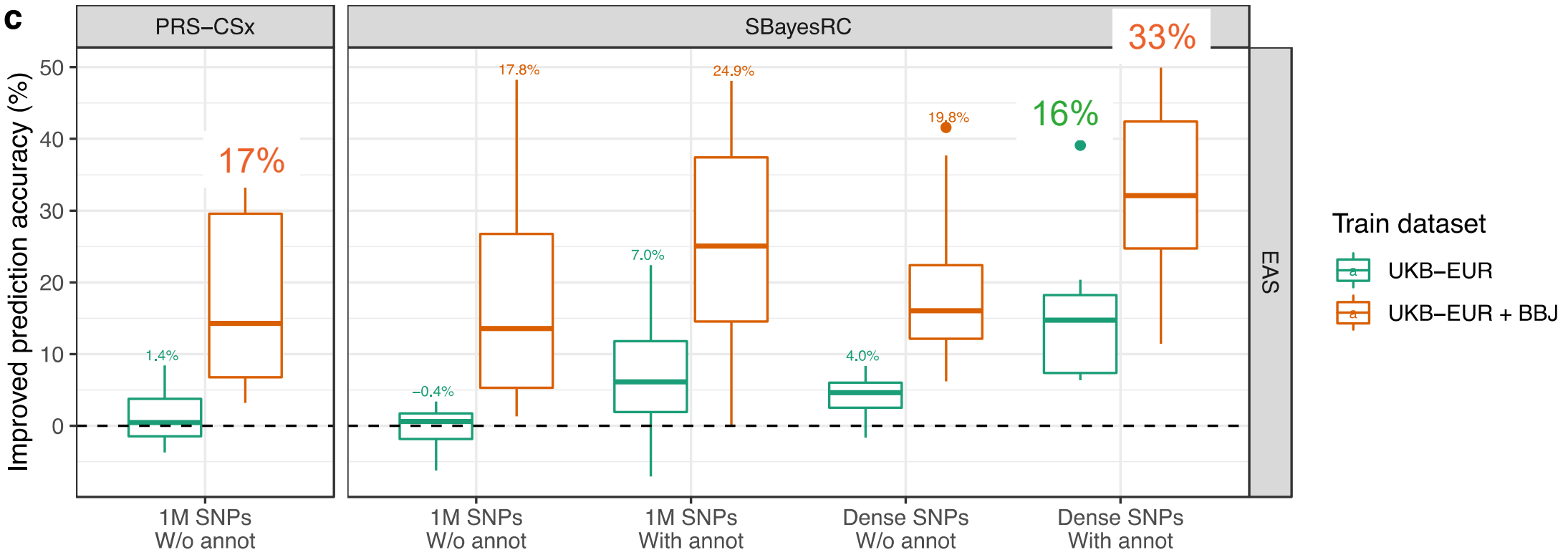
# Trans-ancestry prediction

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



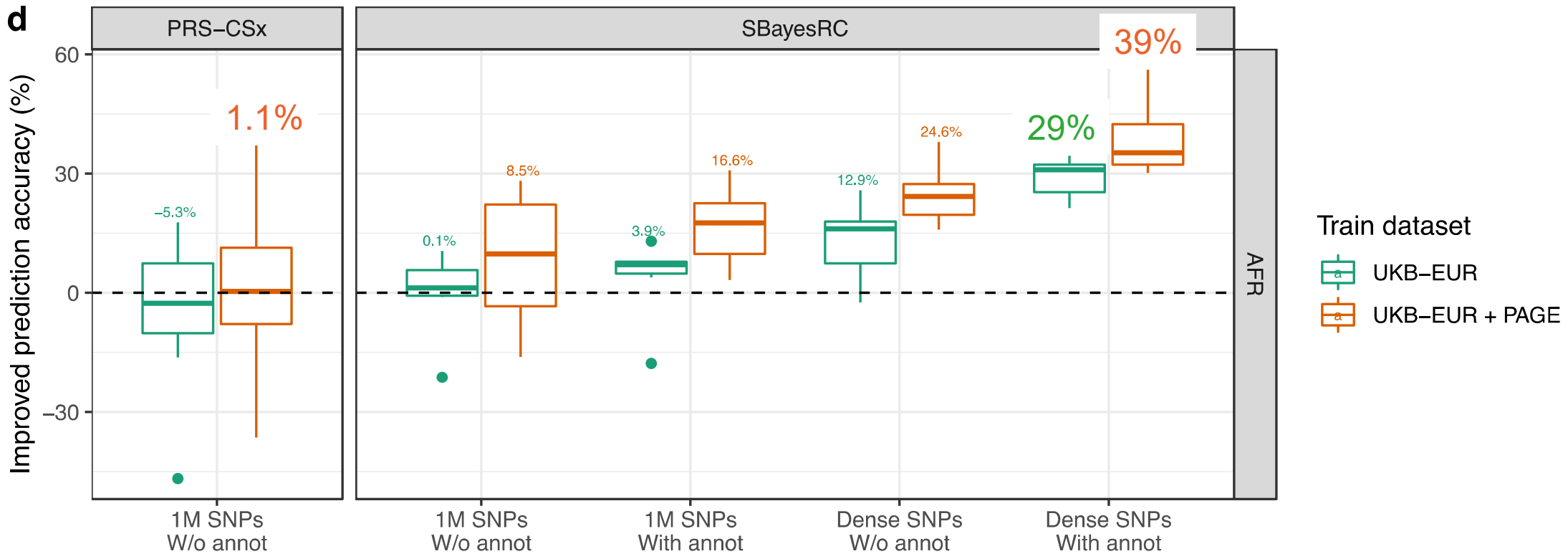
# Trans-ancestry prediction

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

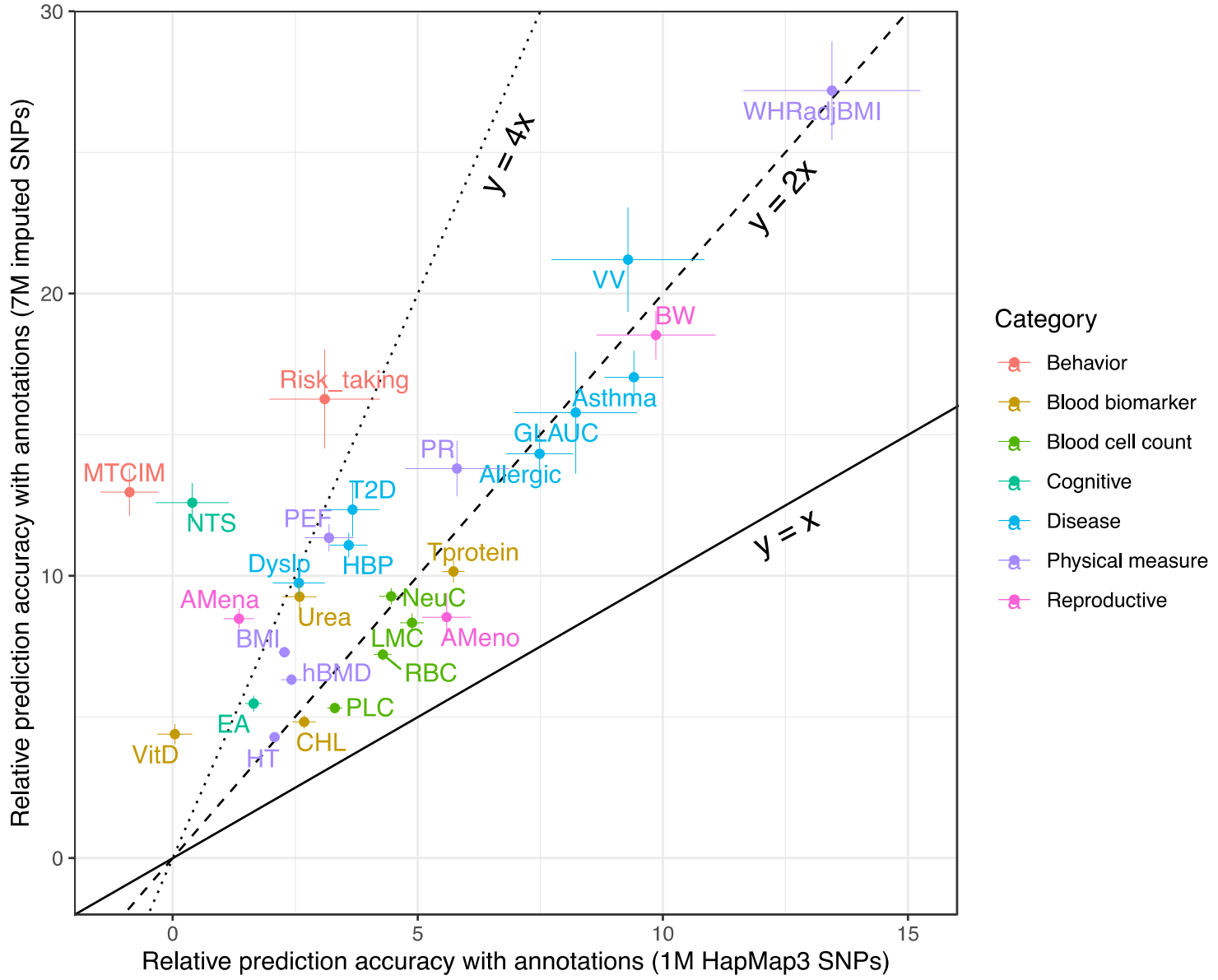


# Trans-ancestry prediction

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:

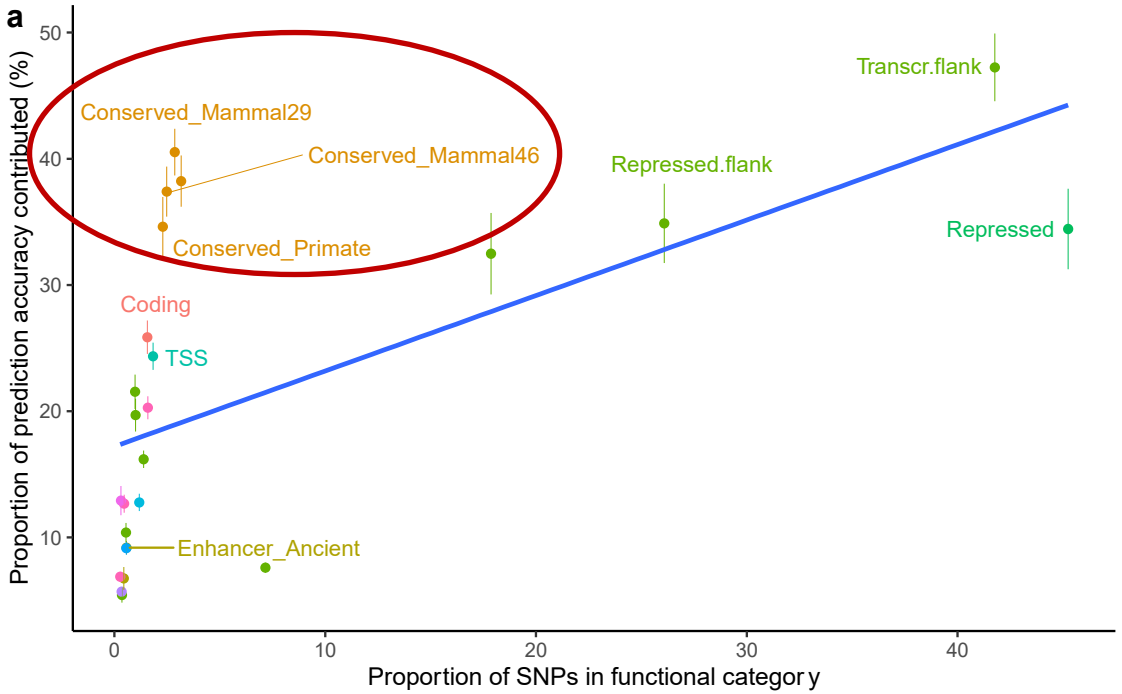
$$\frac{R^2_{\text{annot}} - R^2_{\text{wo}}}{R^2_{\text{wo}}}$$

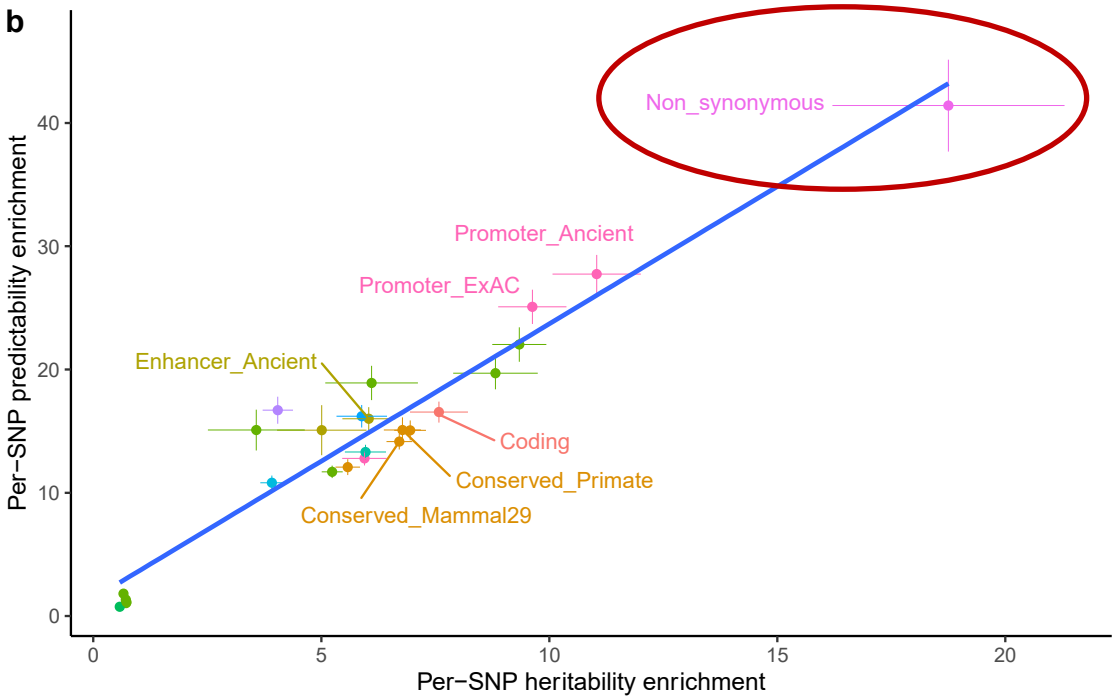
using 7M imputed SNPs (y-axis) or 1M HapMap3 SNPs (x-axis).

Annotations help more with increased SNP density

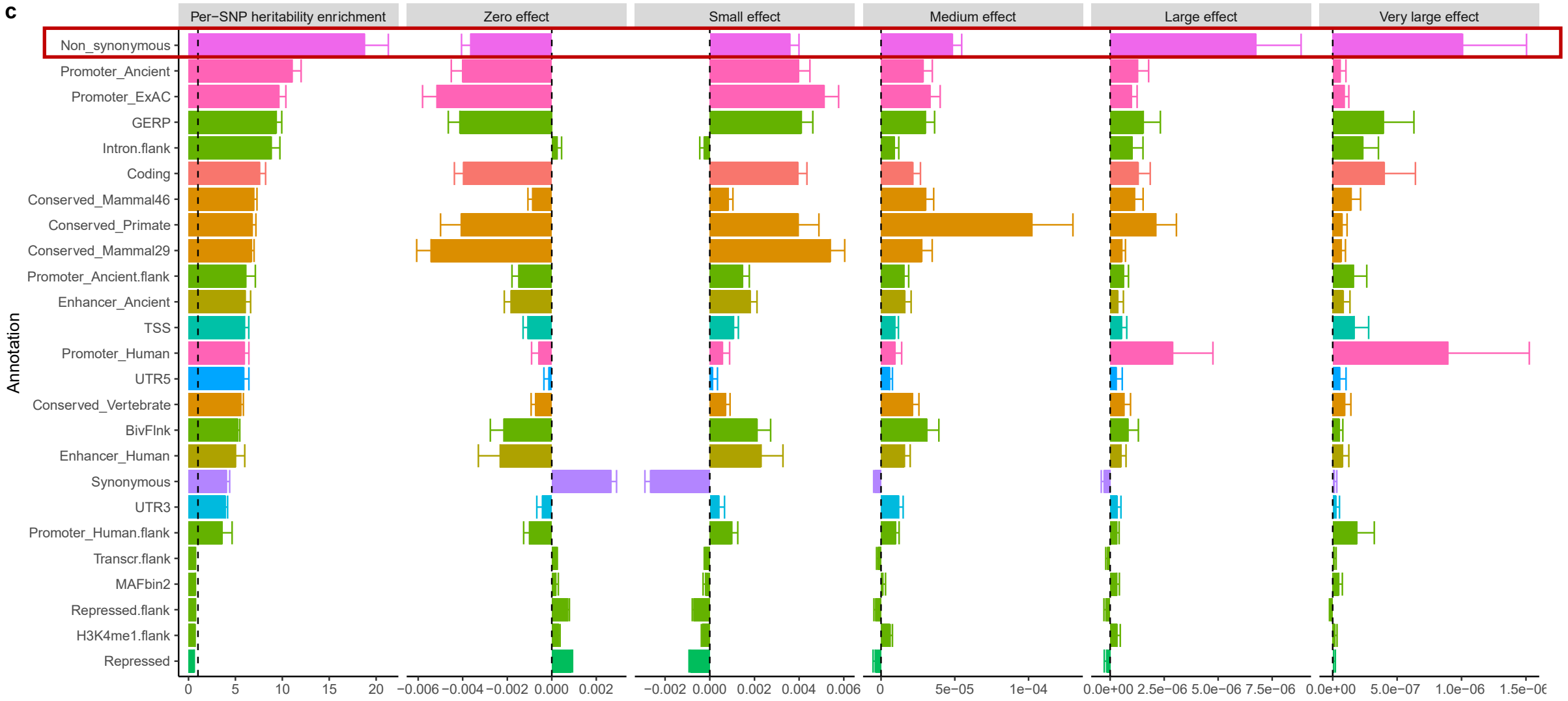
# Contributions of functional categories to prediction accuracy

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



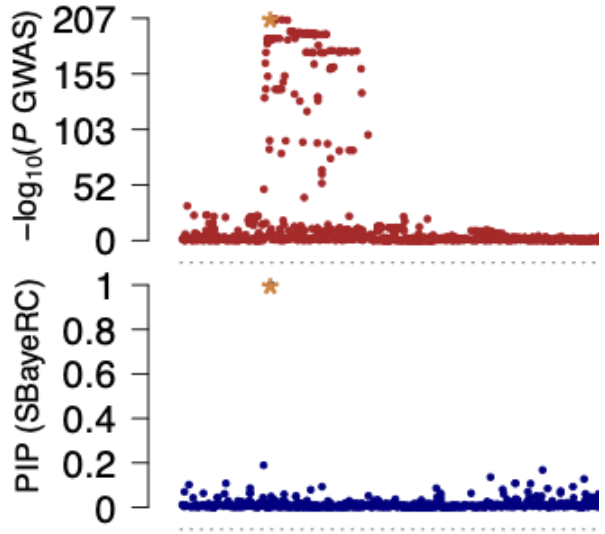



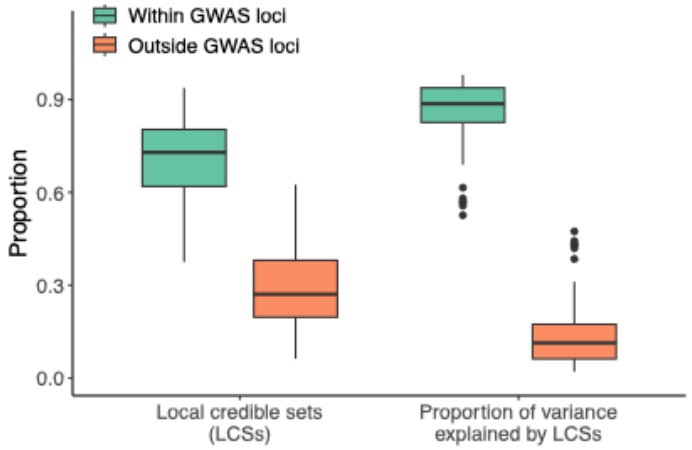
# Functional genetic architecture



# 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[1,2], Zhili Zheng[2,3,4], Loic Thibaut[3], Tian Lin[3], Qian Feng[3], Hao Cheng[5], Loic Yengo[2], Michael E. Goddard[6,7], Naomi R. Wray[2,8], Peter M. Visscher[2,9] & Jian Zeng[2]

**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**)

# Questions?

# 5 min break

# Practical 4: Polygenic prediction using SBayesR(C)

https://cnsgenomics.com/data/teaching/GNGWS26/module5/Practical4_SBayesRC.html

To log into your server, type command below in **Terminal** for Mac/Linux users or in **Command Prompt** or **PowerShell** for Windows users.

```
ssh username@hostname
```

And then key in the provided password.