THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Polygenic prediction incorporating functional annotations

Jian Zeng

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
Institute for Molecular Bioscience

Program in Complex Trait Genomics

Functional genomic annotations

Functional genomic annotations provide orthogonal information which can be used to improve polygenic prediction.

- Chromatin states
- Biological functions
- Pathways
- Context dependent
- Molecular quantitative trait loci (xQTL)
- LD and MAF
- etc

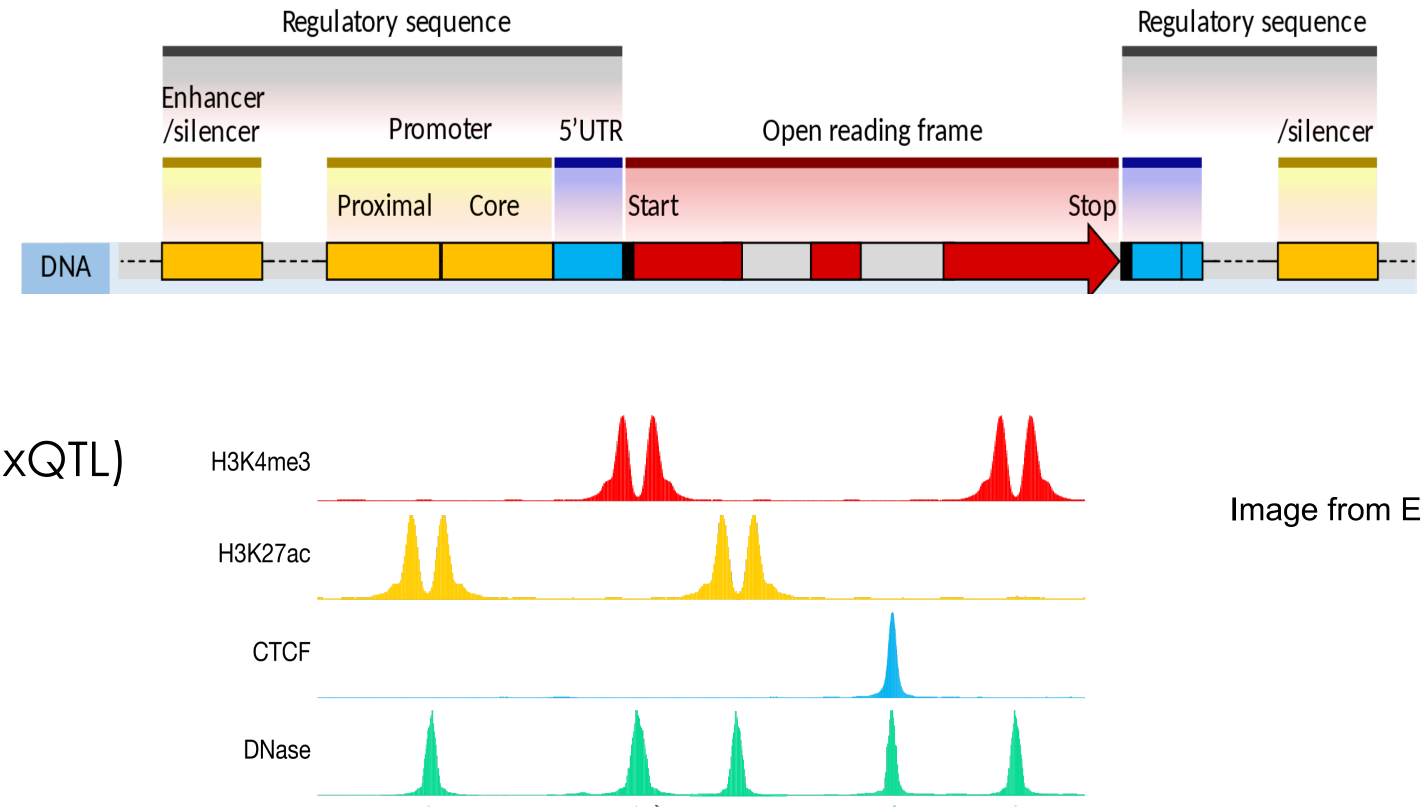


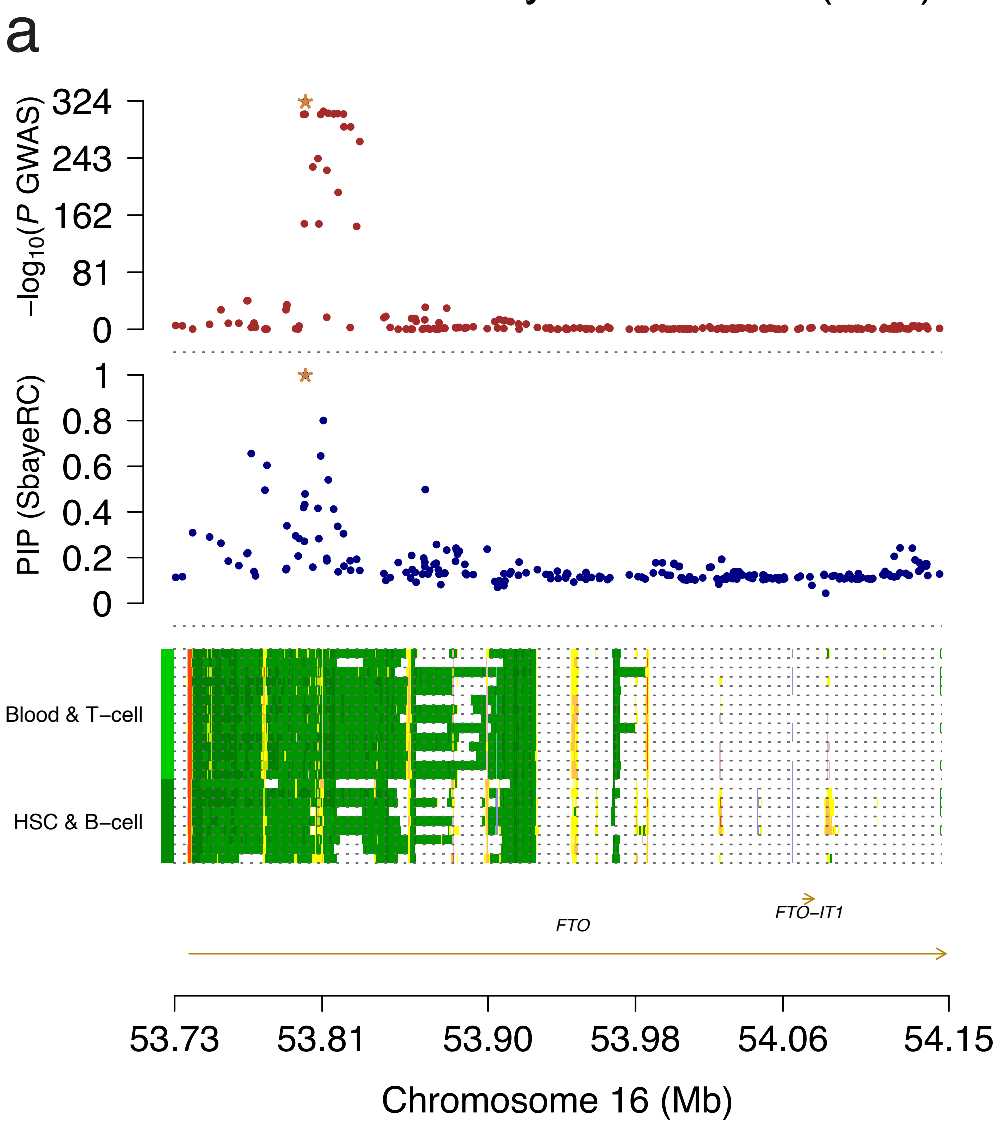
Image from ENCODE

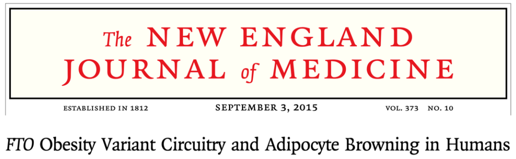
Help interpret the association signal

GWAS signals

SBayesR signals

FTO locus for body mass index (BMI)

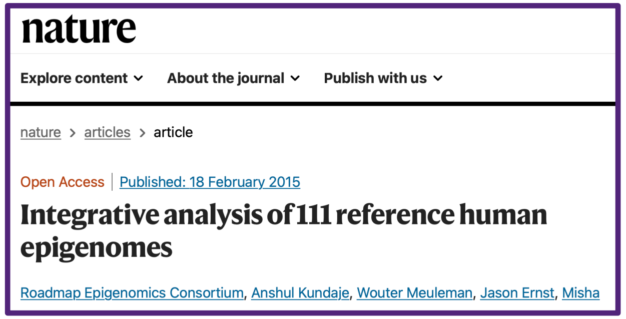




The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 SEPTEMBER 3, 2015 VOL. 373 NO. 10

FTO Obesity Variant Circuitry and Adipocyte Browning in Humans



nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

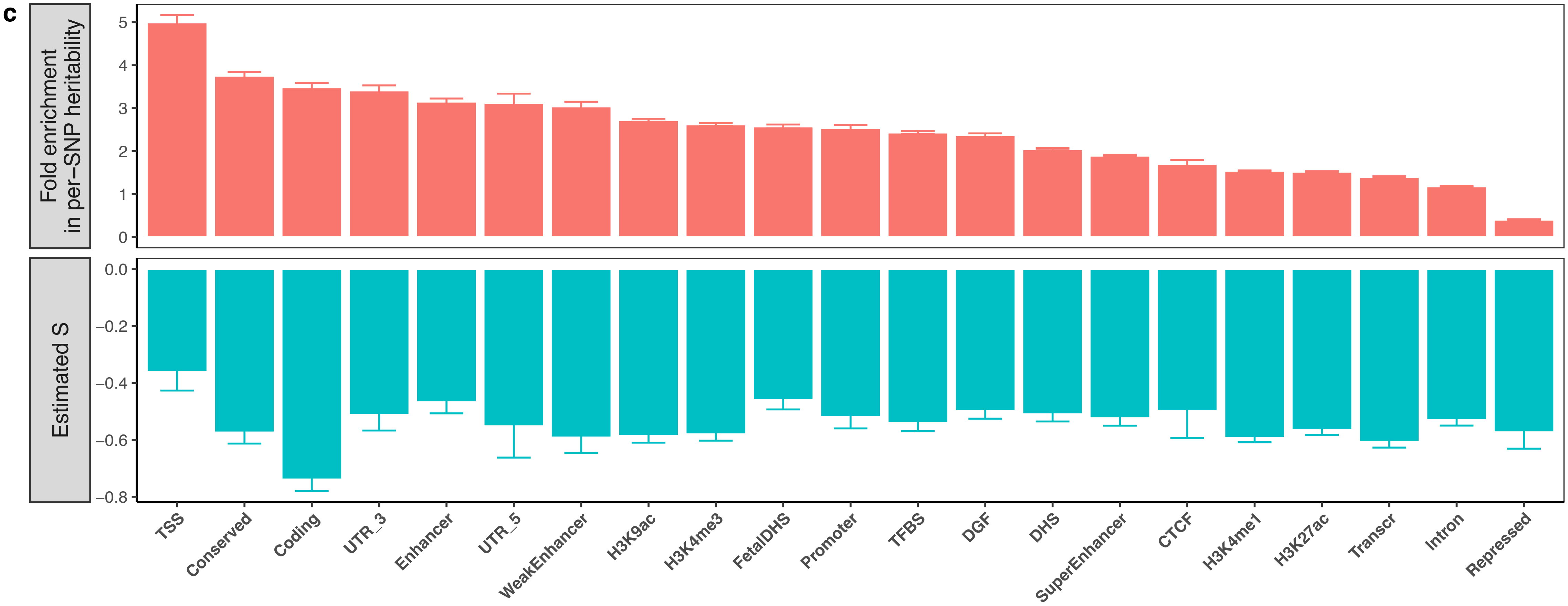
nature > articles > article

Open Access | Published: 18 February 2015

Integrative analysis of 111 reference human epigenomes

Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha

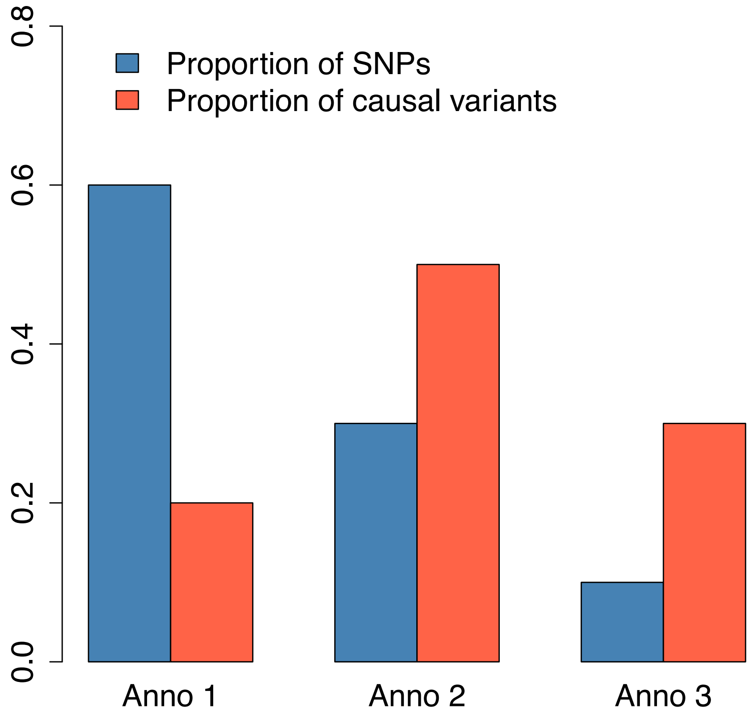
Functional genetic architecture



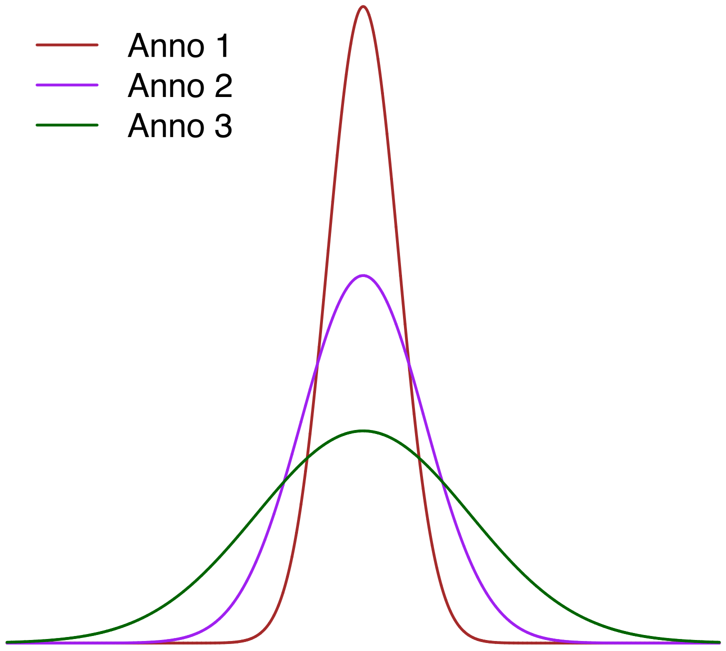
Zeng et al 2021 Nature Communications

Opportunities/challenges

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



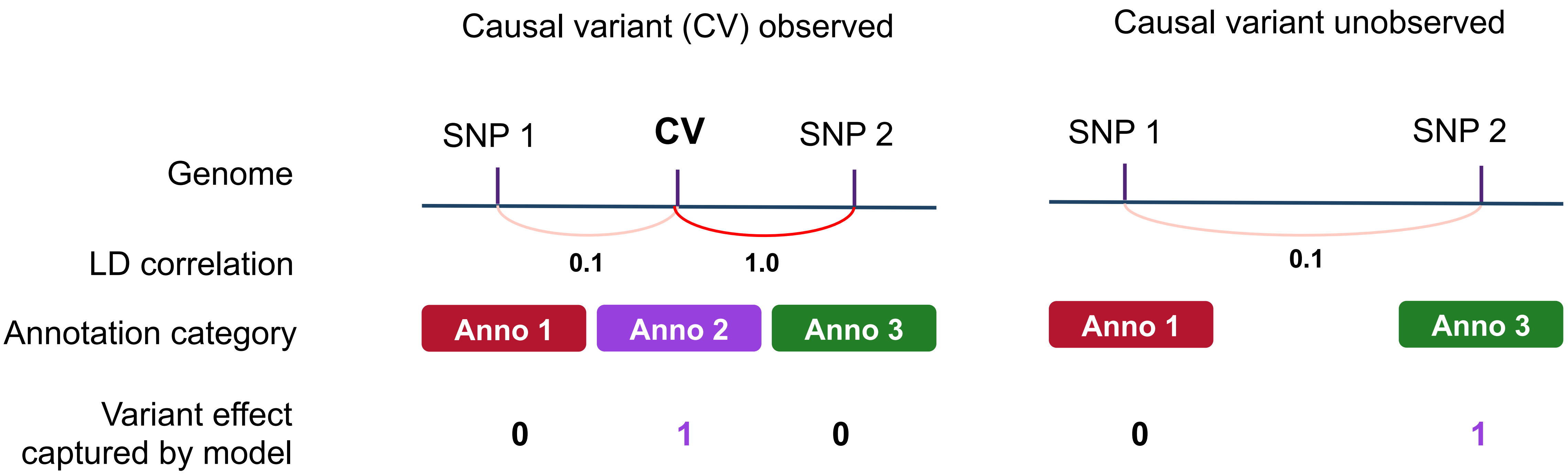
Differences in proportion of causal variants



Differences in distribution of causal effects

Opportunities/challenges

Separate the causal variants from non-causal SNPs in high LD. However, variant annotation and effect may discord if the causal variant is not observed.



Causal variant (CV) observed

Causal variant unobserved

Genome

SNP 1 CV SNP 2

SNP 1 SNP 2

LD correlation

0.1 1.0

0.1

Annotation category

Anno 1 Anno 2 Anno 3

Anno 1 Anno 3

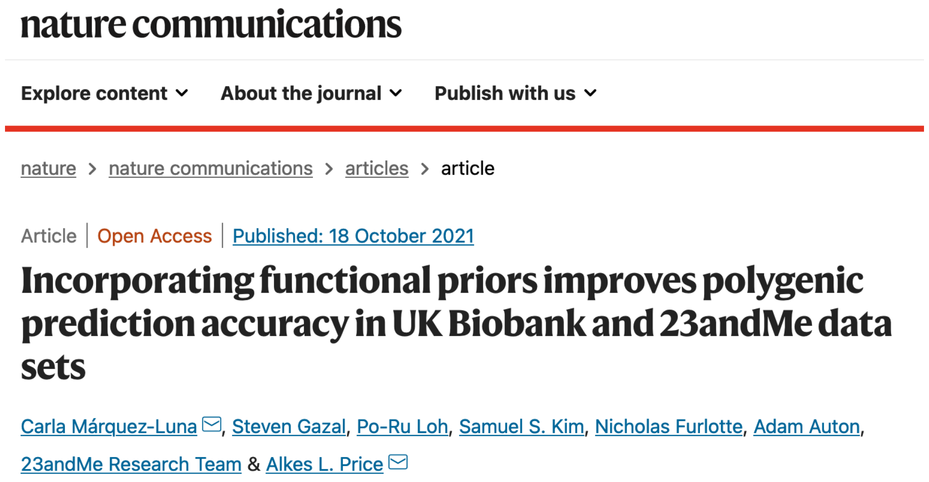
Variant effect captured by model

0 1 0

0 1

Literature

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA



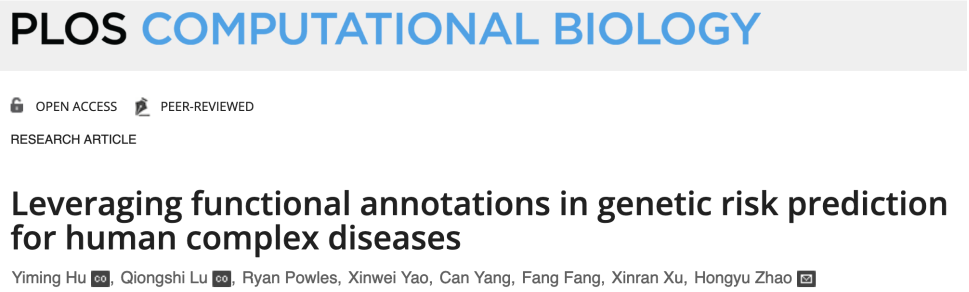
nature communications

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

LDpredFunct method



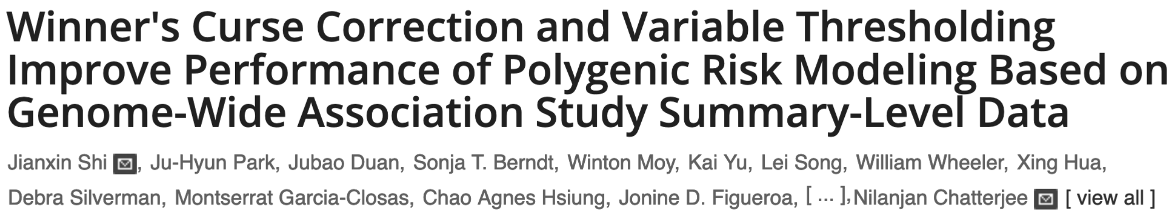
PLOS COMPUTATIONAL BIOLOGY

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

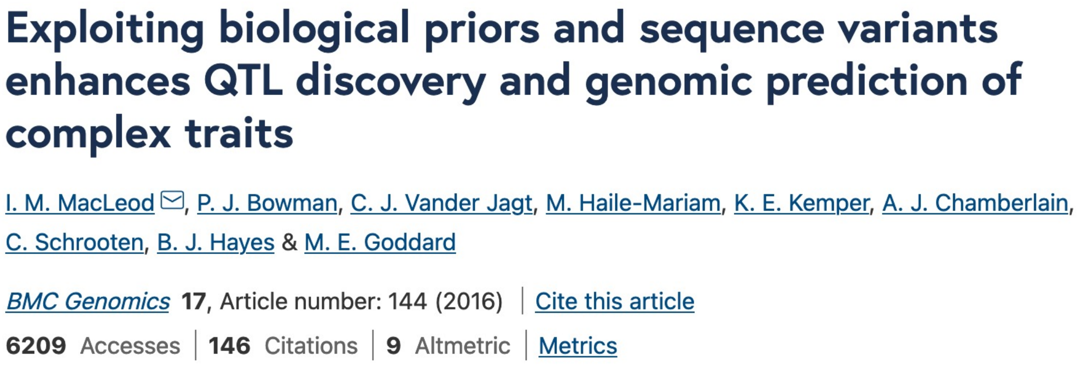
AnnoPred method



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 method



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

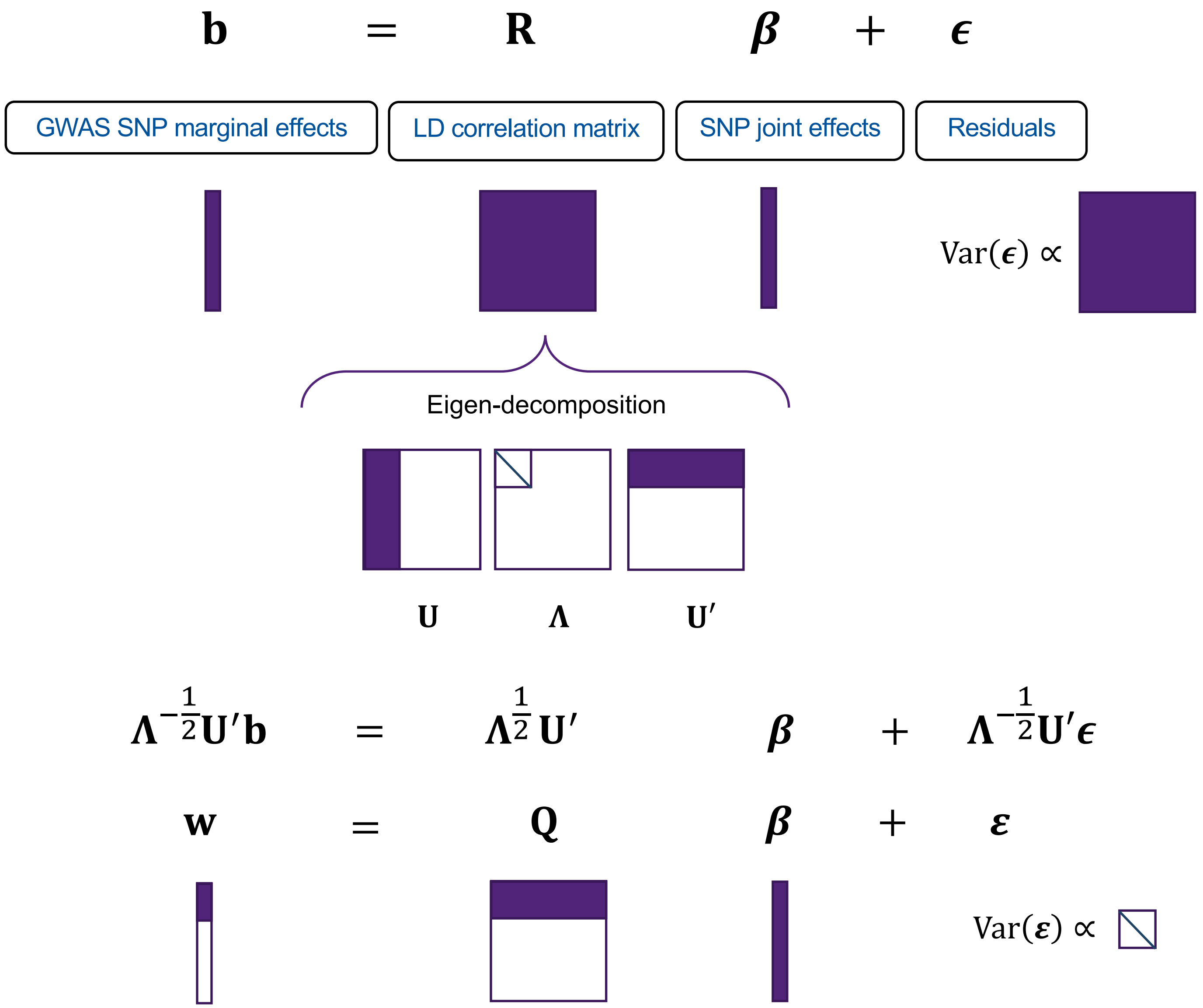
6209 Accesses | 146 Citations | 9 Altmetric | Metrics

BayesRC method

Need prediction methods that can simultaneously fit all SNPs and learn weights of annotations from the data.

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
$\mathrm{Var}(\boldsymbol{\epsilon}) \propto$
Eigen-decomposition
$\mathbf{U}$
$\boldsymbol{\Lambda}$
$\mathbf{U}'$

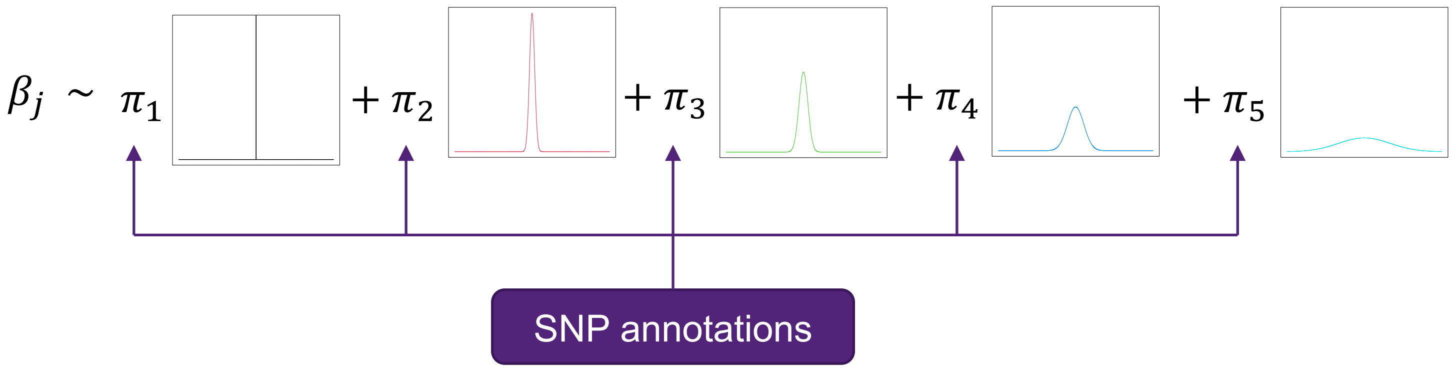
$$\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}'\mathbf{b} = \boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}'\boldsymbol{\beta} + \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}'\boldsymbol{\epsilon}$$

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

$\mathrm{Var}(\boldsymbol{\varepsilon}) \propto$

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

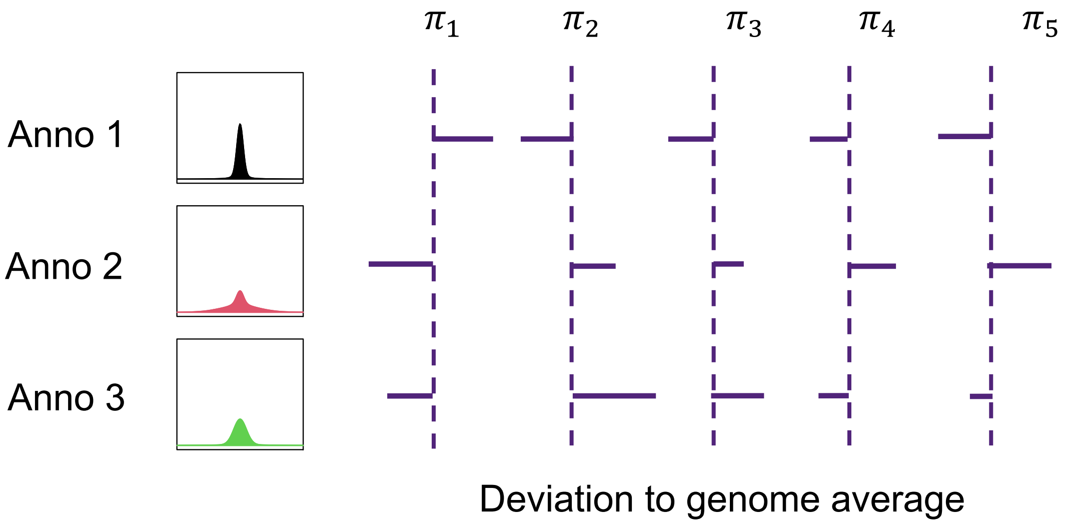
Modelling functional annotations (SBayesRC)



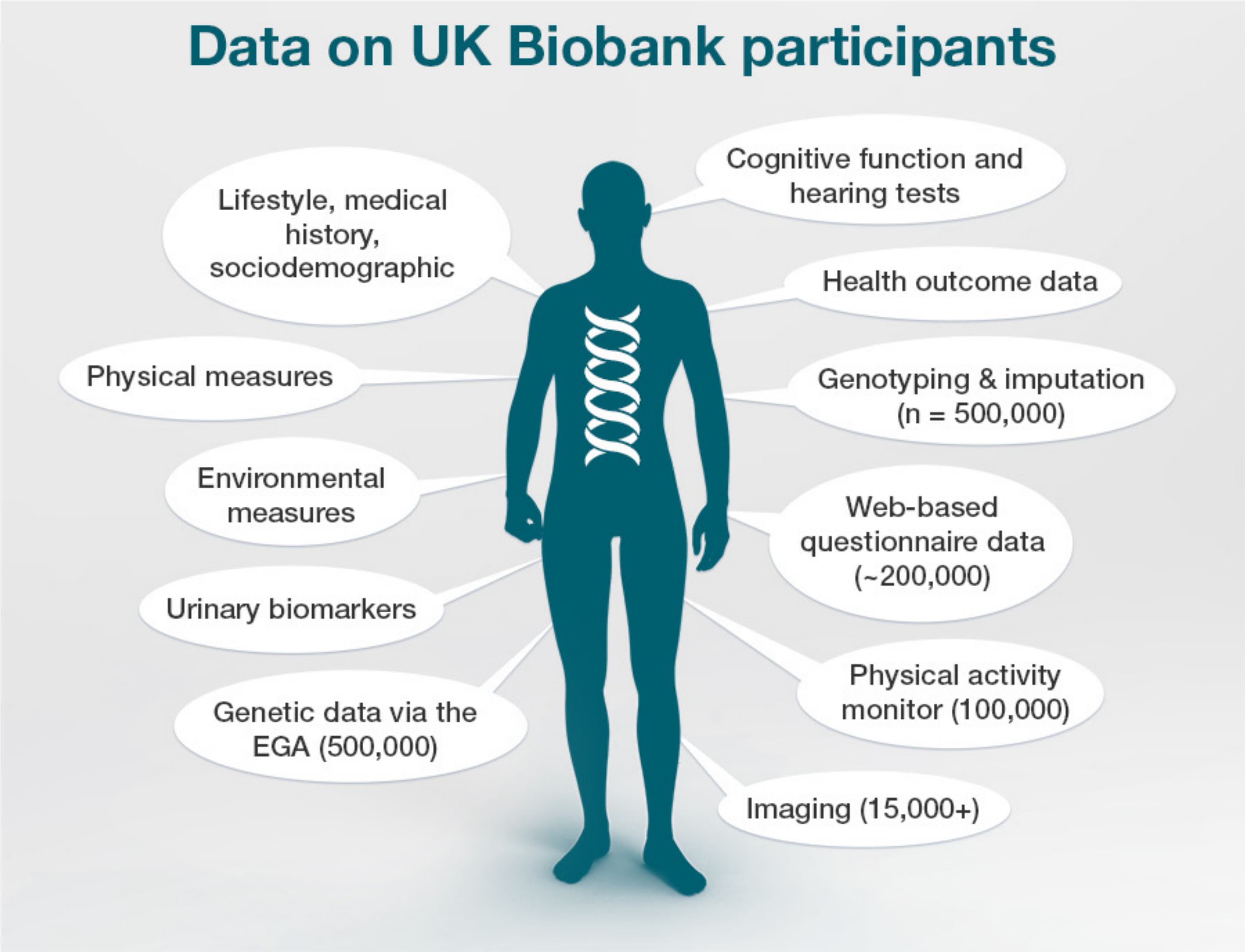
$$f\left(\pi_{jk}\right) = \text{Intercept} + \sum \text{SNP annotation} \times \text{annotation effect}$$

annotation effect $\sim N(0, \sigma_\alpha^2)$

Probit link is used to enable Gibbs sampling



Real data analysis


Data on UK Biobank participants
Lifestyle, medical history, sociodemographic
Cognitive function and hearing tests
Health outcome data
Physical measures
Genotyping & imputation (n = 500,000)
Environmental measures
Web-based questionnaire data (~200,000)
Urinary biomarkers
Physical activity monitor (100,000)
Genetic data via the EGA (500,000)
Imaging (15,000+)

- 340K unrelated individuals of European ancestry
- 28 independent traits with large sample size (including 8 diseases)
- Adjust for age, sex and 10PCs
- 96 continuous and categorical SNP annotations from **BaselineLDv2.2** (Gazal et al 2017 Nature Genetics)
- Random sample of 20K individuals of European ancestry as LD reference

Within European ancestry prediction

Benchmark is the prediction accuracy from SBayesR using 1M HapMap3 SNPs (dash line).

$$\frac{R_{x}^{2} - R_{\mathrm{SBayesR}}^{2}}{R_{\mathrm{SBayesR}}^{2}}$$

LDpred2

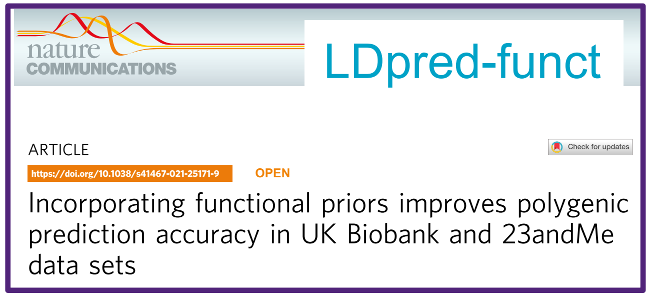
Bioinformatics, 2021, 1–8
doi: 10.1093/bioinformatics/btaa1029
Advance Access Publication Date: 16 December 2020
Original Paper

OXFORD

Genetics and population analysis

LDpred2: better, faster, stronger

Florian Privé[1],*, Julyan Arbel[2] and Bjarni J. Vilhjálmsson[1,3],*



nature COMMUNICATIONS

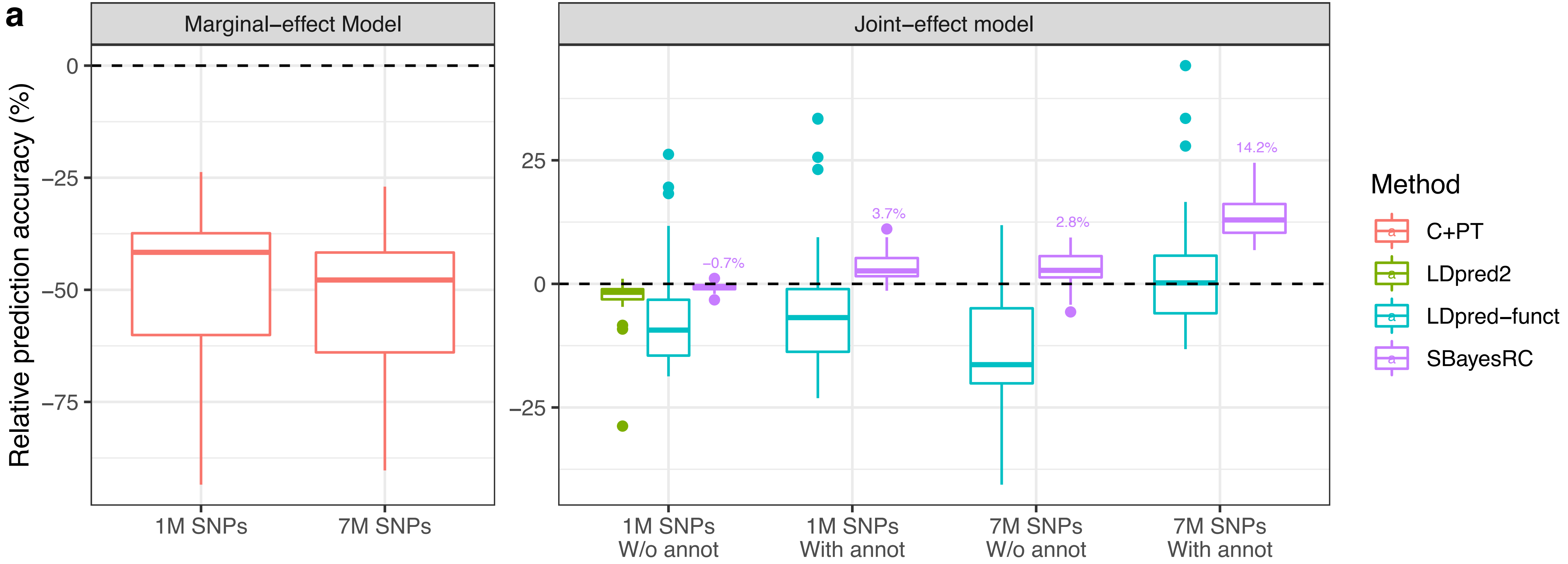
LDpred-funct

ARTICLE

https://doi.org/10.1038/s41467-021-25171-9 OPEN

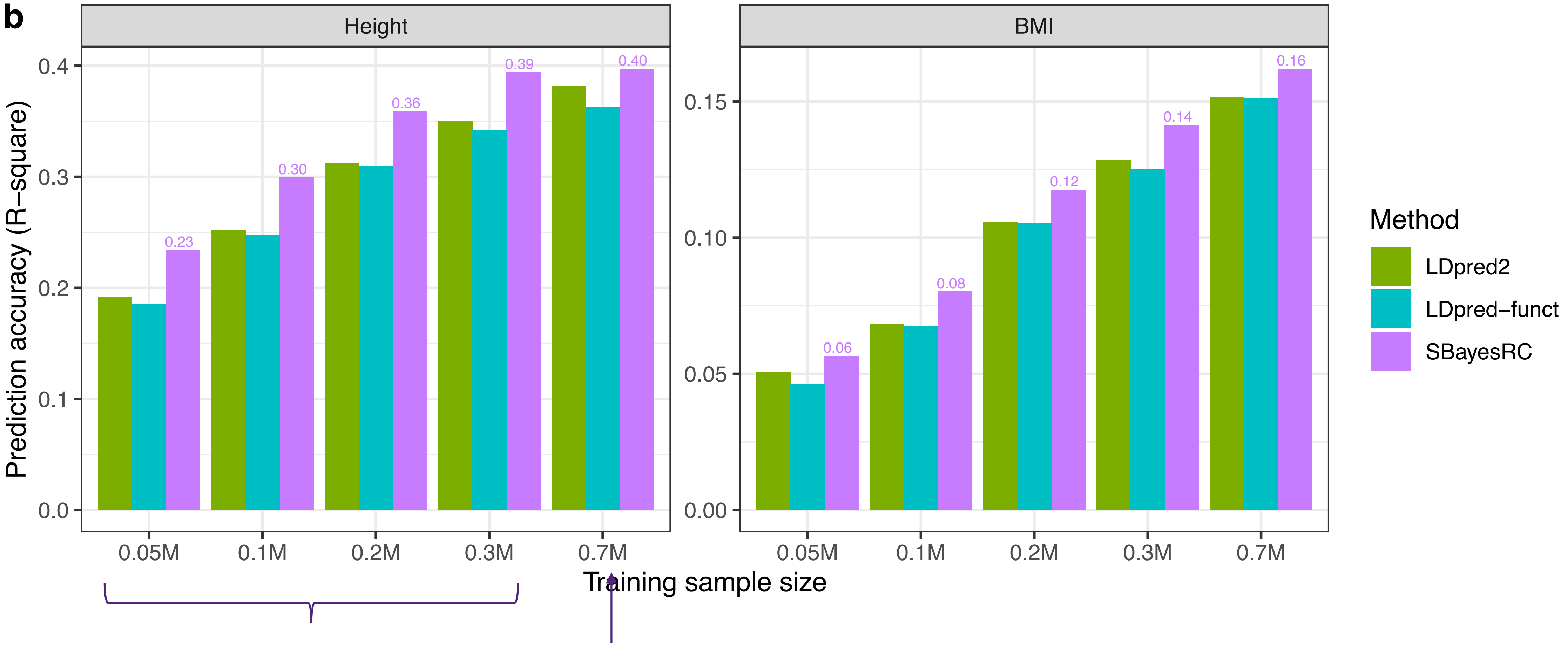
Check for updates

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



Cross-biobank prediction (Lifelines)

Prediction R^2 = 0.4 in height and 0.16 in BMI (~70% SNP-based heritability)

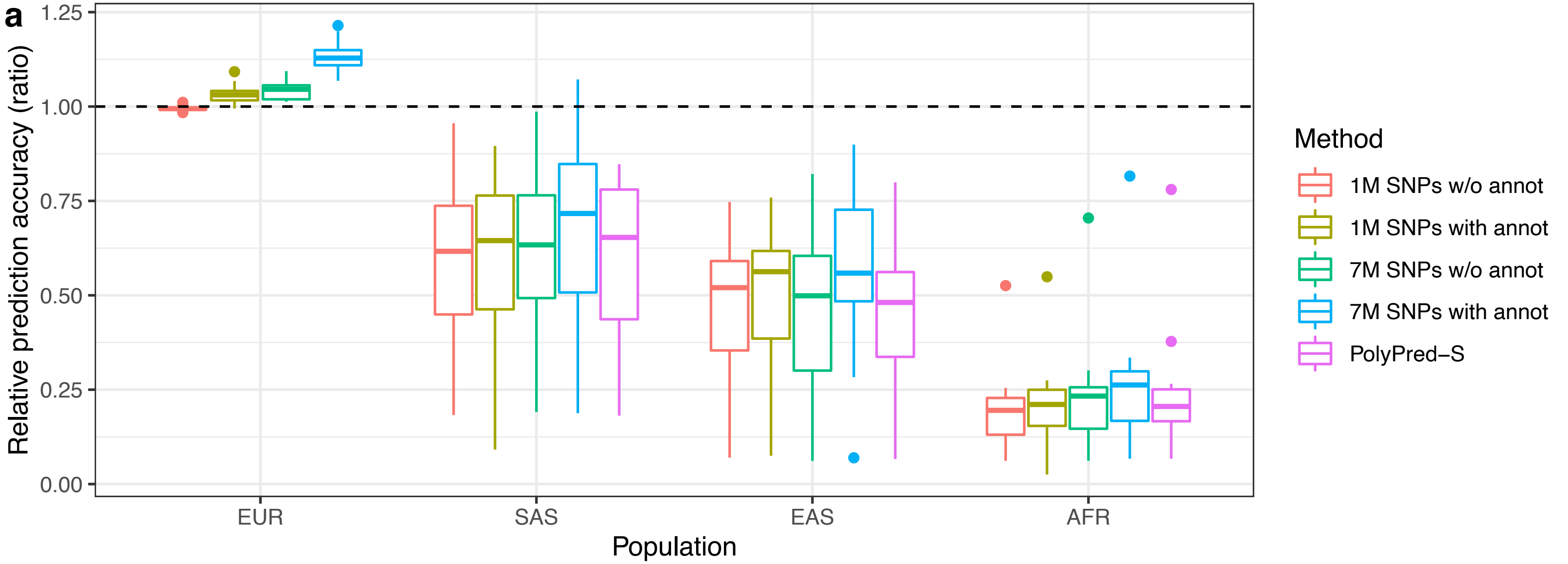


UKB

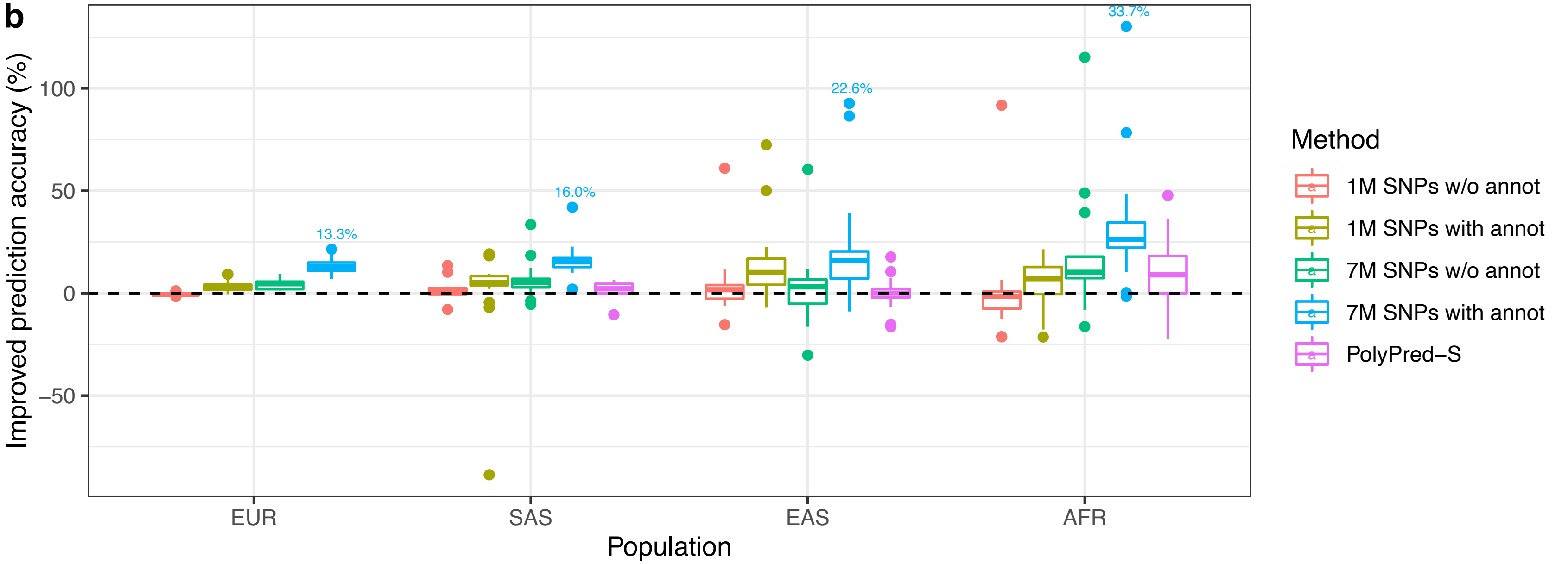
Yengo et al 2018 HMG

Trans-ancestry prediction

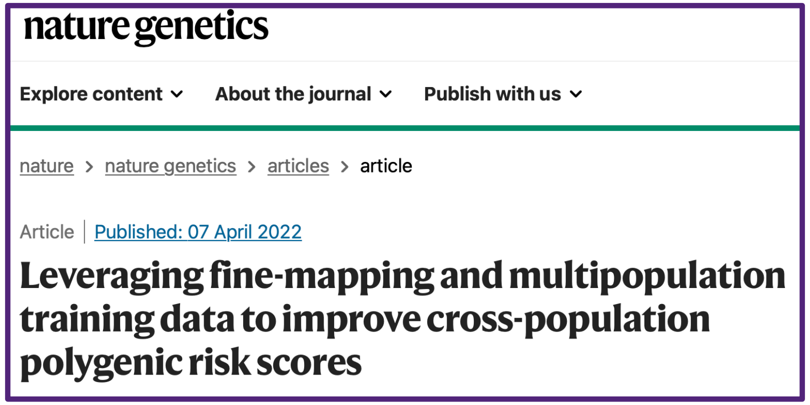
$$\frac{R_x^2\,(\text{Pop})}{R_{\text{SBayesR}}^2(\text{EUR})}$$



$$\frac{R_x^2\,(\text{Pop}) - R_{\text{SBayesR}}^2(\text{Pop})}{R_{\text{SBayesR}}^2(\text{Pop})}$$



PolyPred-S (Weissbrod et al 2022 Nature Genetics)



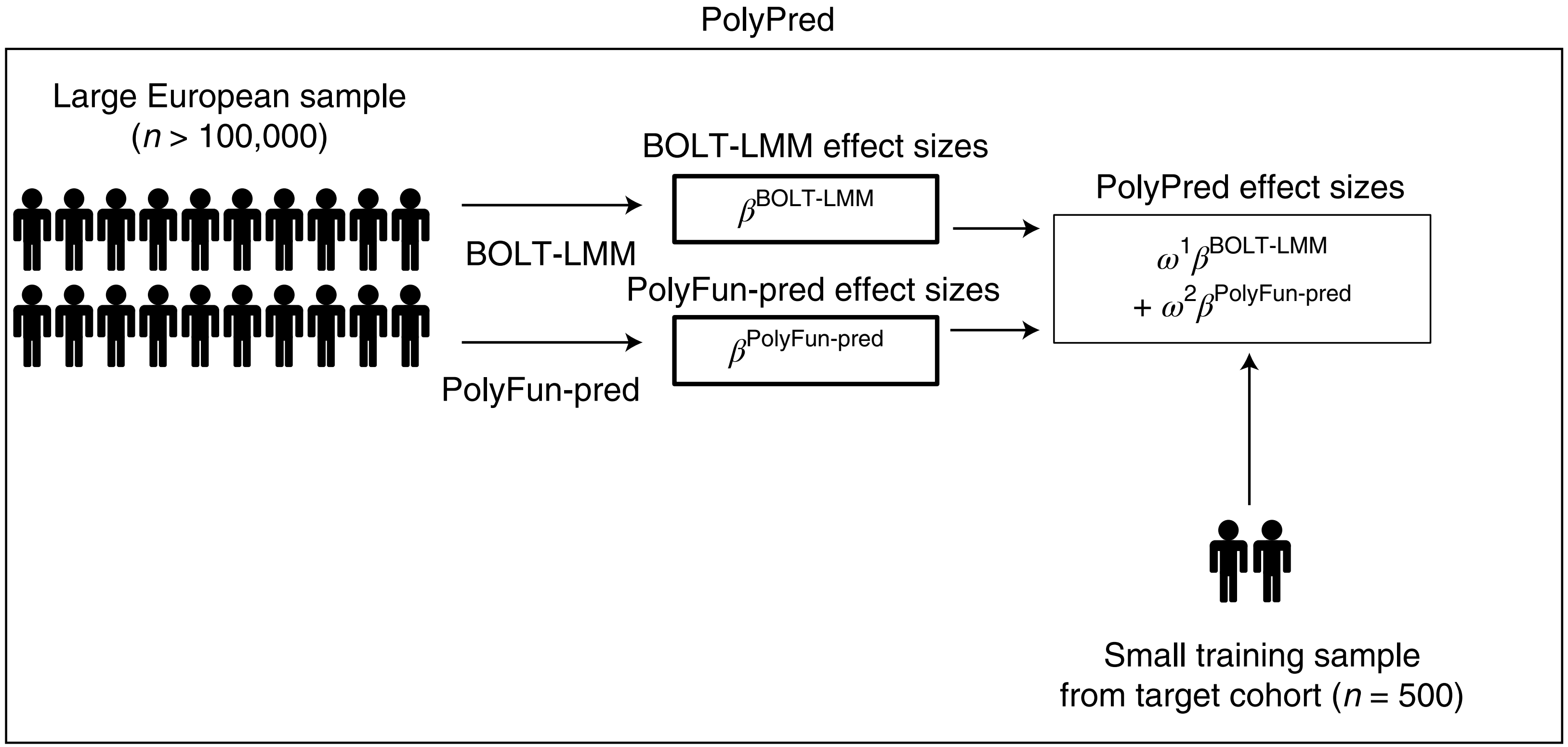
nature genetics

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature genetics > articles > article

Article | Published: 07 April 2022

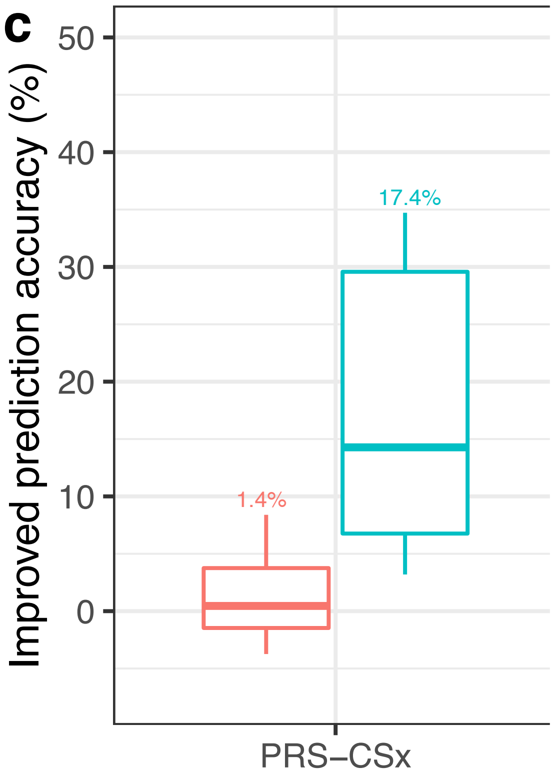
Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores

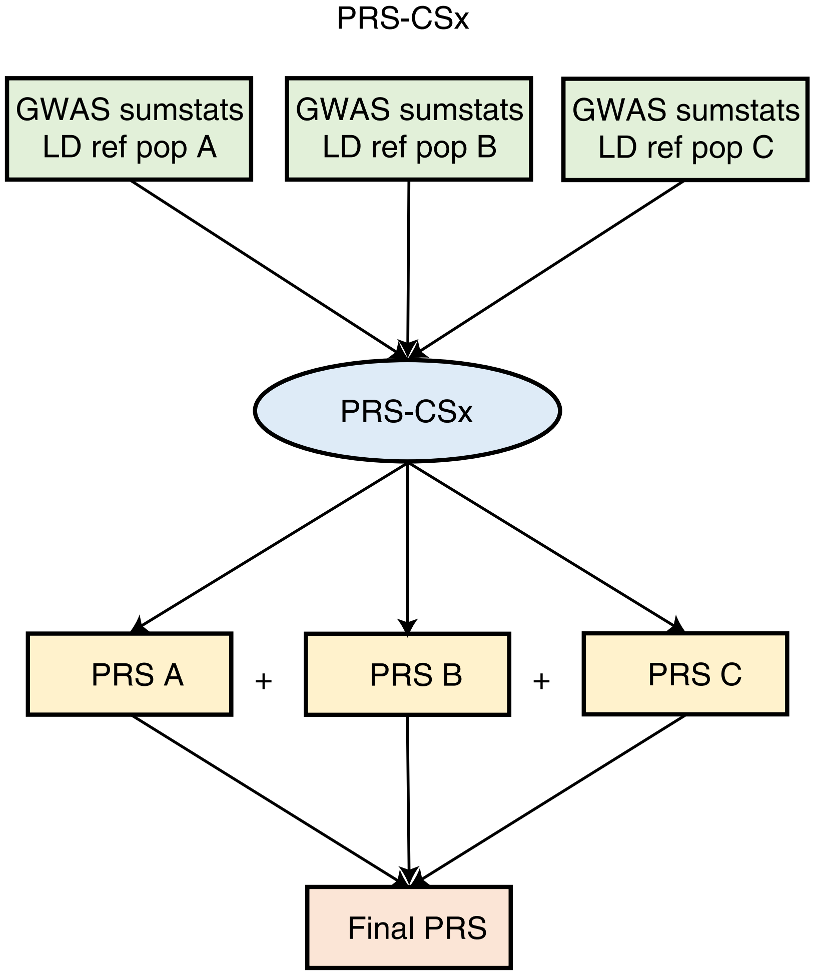

PolyPred
Large European sample
(n > 100,000)
BOLT-LMM
PolyFun-pred
BOLT-LMM effect sizes
$\beta^{\text{BOLT-LMM}}$
PolyFun-pred effect sizes
$\beta^{\text{PolyFun-pred}}$
PolyPred effect sizes
$\omega^1\beta^{\text{BOLT-LMM}}$
$+\ \omega^2\beta^{\text{PolyFun-pred}}$
Small training sample
from target cohort (n = 500)

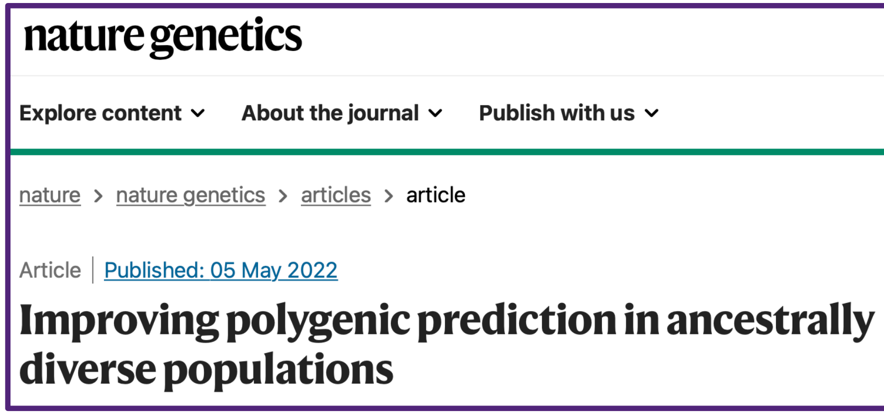
PolyPred-S is a variation of PolyPred with BOLT-LMM replaced by SBayesR estimates.

Trans-ancestry prediction

Use GWAS data from UKB EUR and BBJ (Biobank Japan) EAS to predict UKB EAS





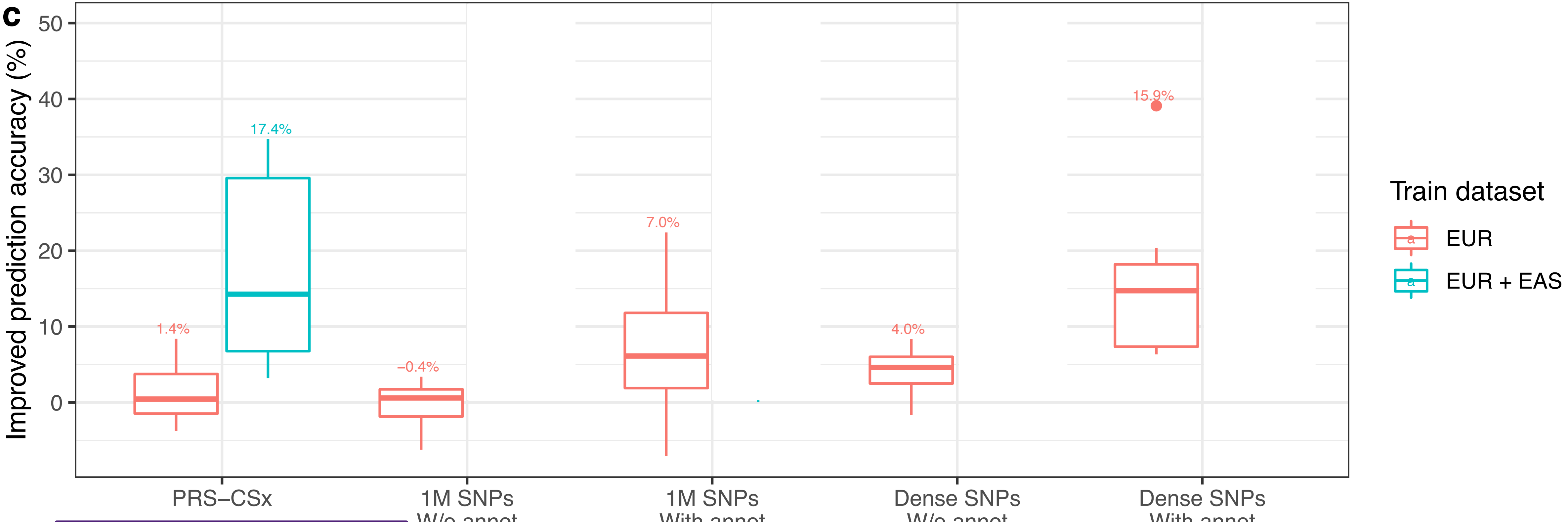

nature genetics

Article | Published: 05 May 2022

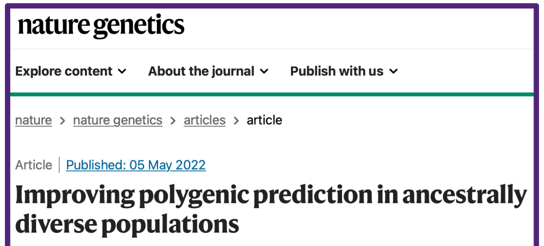
Improving polygenic prediction in ancestrally diverse populations

Trans-ancestry prediction

Use GWAS data from UKB EUR and BBJ (Biobank Japan) EAS to predict UKB EAS



SBayesRC



nature genetics

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature genetics > articles > article

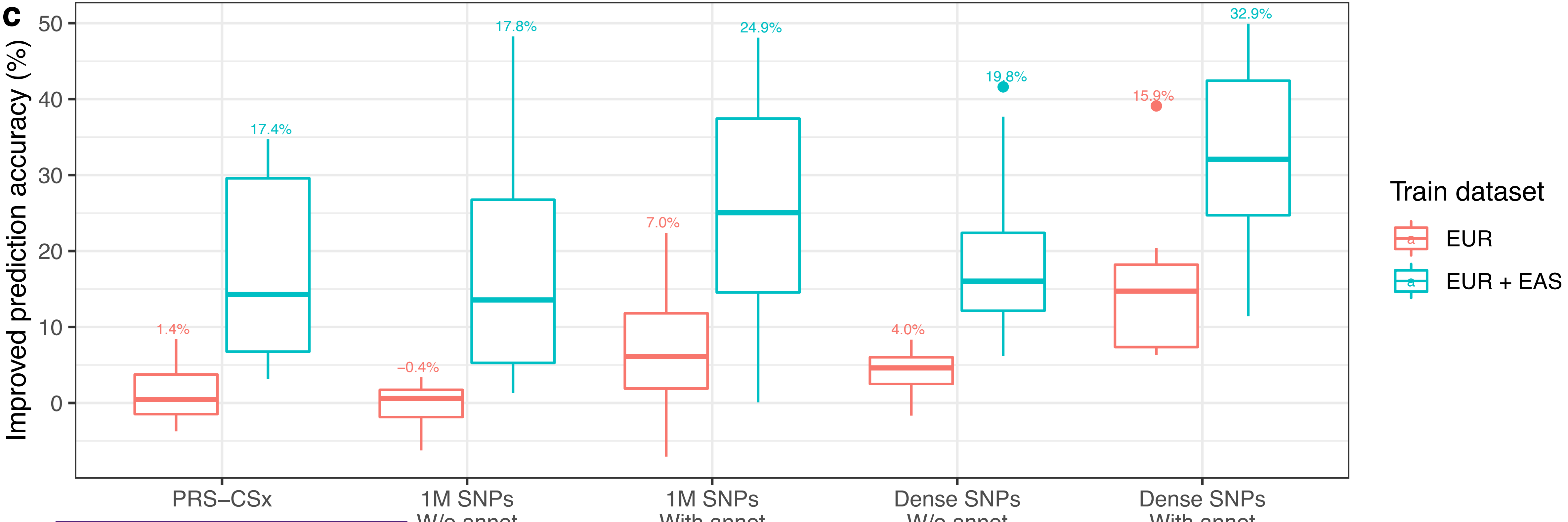
Article | Published: 05 May 2022

Improving polygenic prediction in ancestrally diverse populations

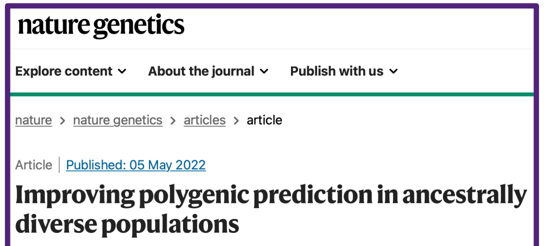
Trans-ancestry prediction

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Use GWAS data from UKB EUR and BBJ (Biobank Japan) EAS to predict UKB EAS



SBayesRC


nature genetics

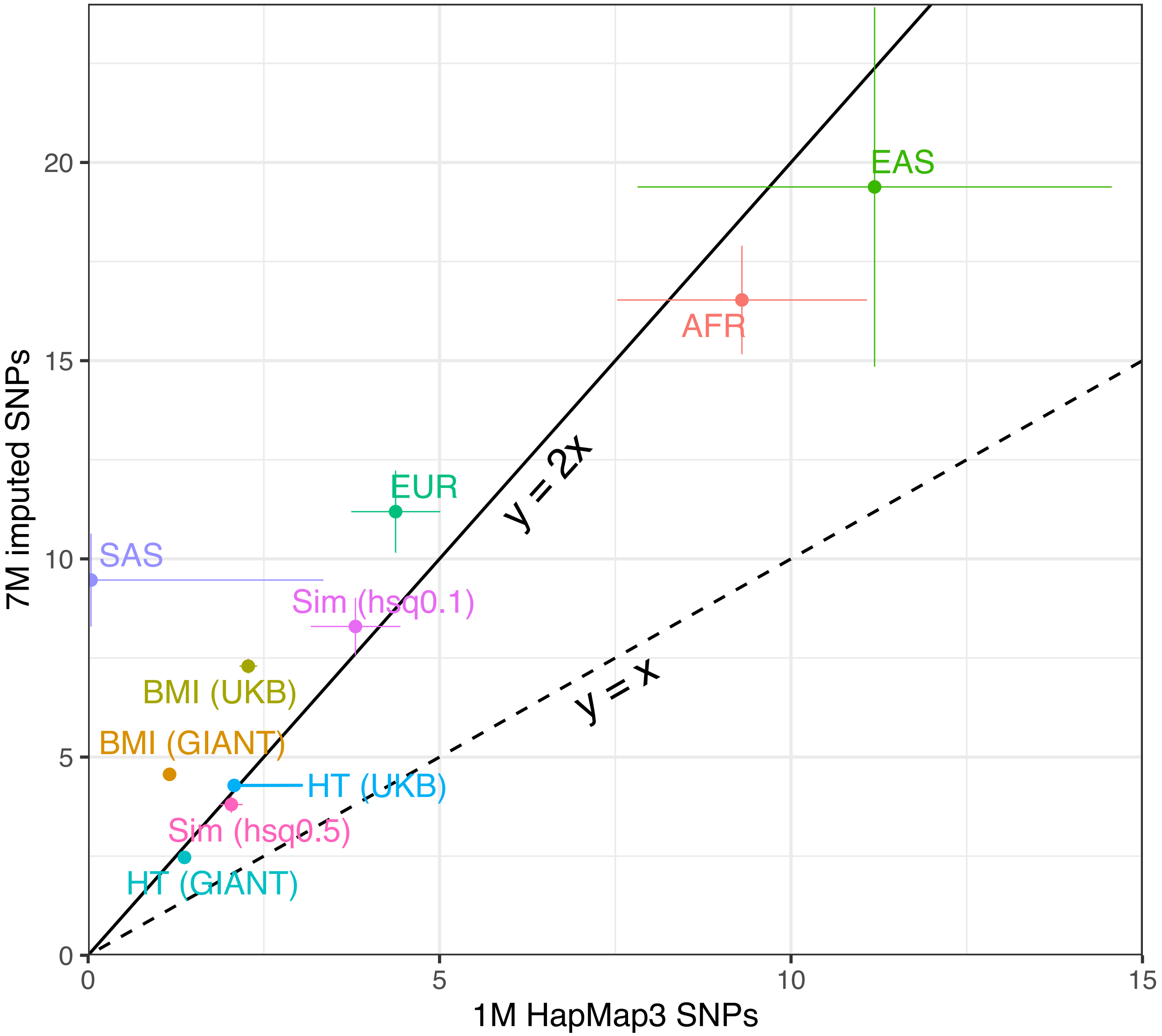
Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature genetics › articles › article

Article | Published: 05 May 2022

Improving polygenic prediction in ancestrally diverse populations

Interaction between SNP density and annotation information



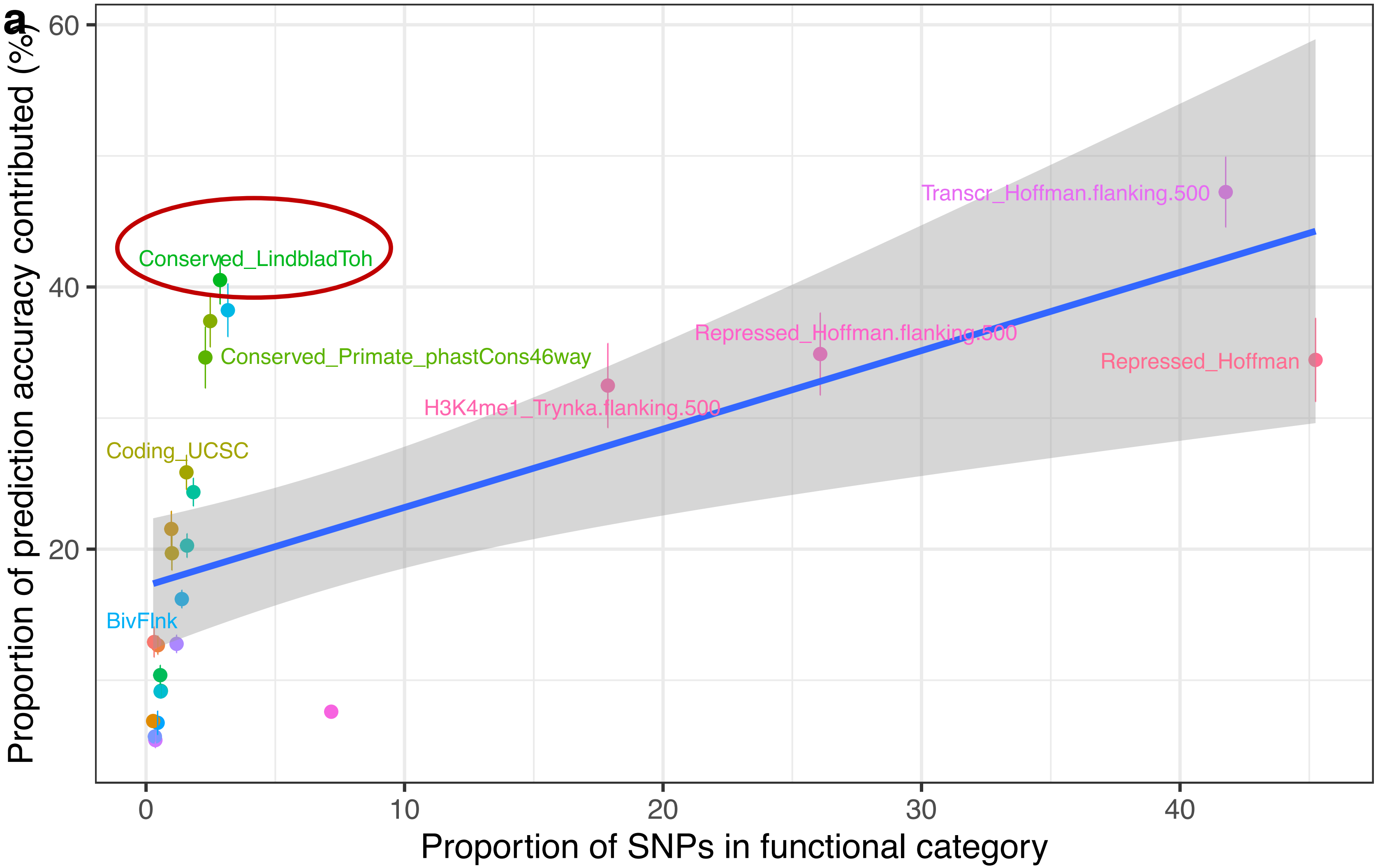
Improvement (%) in prediction accuracy for SBayesRC using annotations relative to that without annotations:

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

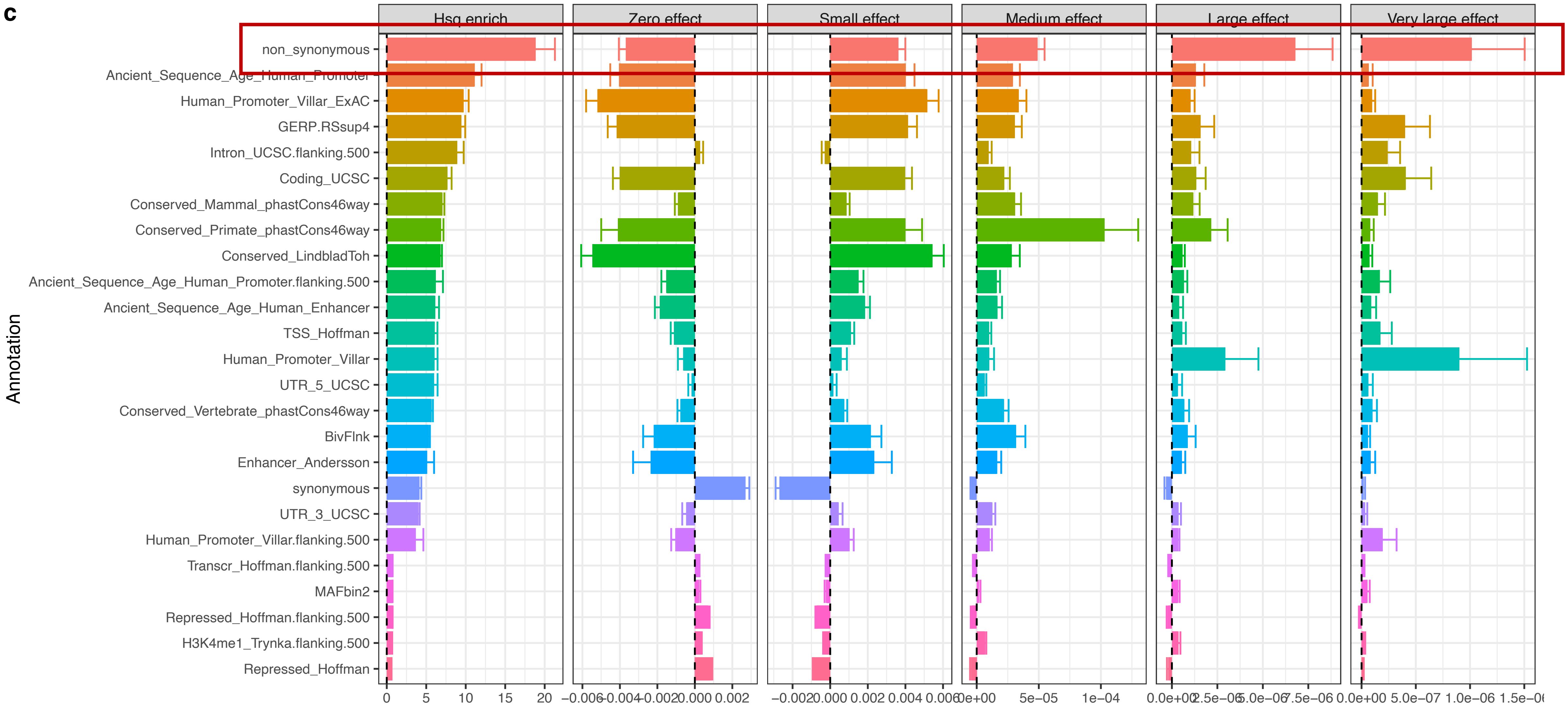
regression slope = 1.88 (se = 0.22)

Contributions of functional categories to prediction accuracy

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



Functional genetic architecture



Computational efficiency

Results are average values across traits using 4 CPU cores.

Method (No. SNPs)	Runtime (hours)	Memory (GB)	Storage (GB)
SBayesRC (7M)	9.5	75.1	130
LDpred-funct (7M)	6.0	120.6	40-50 per trait
PolyPred-S (7M)	19.8	71.7	2,800
LDpred2 (1M)	5.5	53.4	43
SBayesRC (1M)	1.2	7.8	5.6
SBayesR (1M)	0.5	27.0	22
PRS-CSx (1M)	14.2	4.7	5.6

Summary

- SBayesRC improves prediction accuracy by 14% in European ancestry and by up to 33% in trans-ancestry prediction, compared to the baseline method SBayesR which does not use annotations.

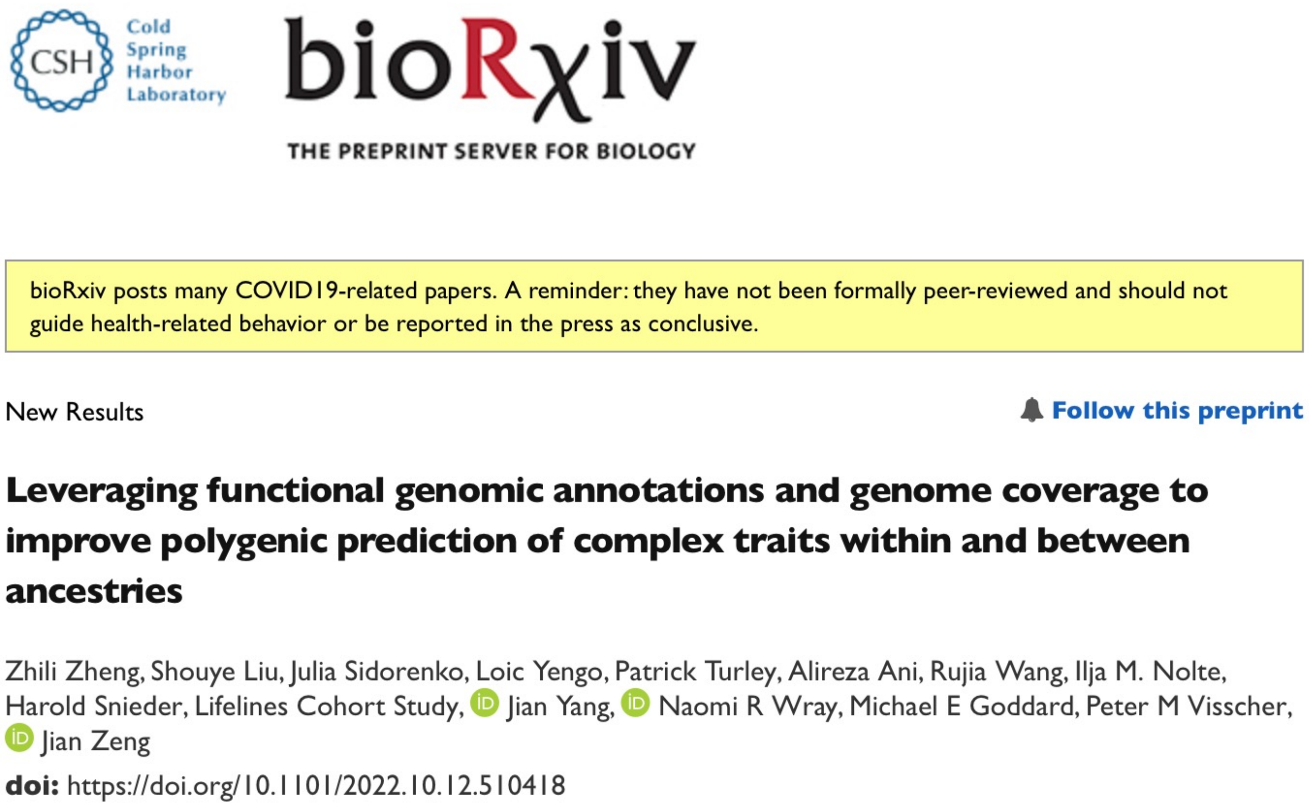
- SBayesRC outperforms state-of-the-art methods LDpred-funct, PolyPred-S and PRS-CSx by 12-15% in prediction accuracy.

- We identified a significant interaction between SNP density and annotation information, encouraging future use of whole-genome sequence variants for prediction.

- Functional partitioning analysis highlights a major contribution of evolutionary constrained regions to prediction accuracy.

GCTB software https://cnsgenomics.com/software/gctb

R package at https://github.com/zhilizheng/SBayesRC


CSH
Cold Spring Harbor Laboratory
bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

Follow this preprint

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

Zhili Zheng, Shouye Liu, Julia Sidorenko, Loic Yengo, Patrick Turley, Alireza Ani, Rujia Wang, Ilja M. Nolte, Harold Snieder, Lifelines Cohort Study, Jian Yang, Naomi R Wray, Michael E Goddard, Peter M Visscher, Jian Zeng

doi: https://doi.org/10.1101/2022.10.12.510418