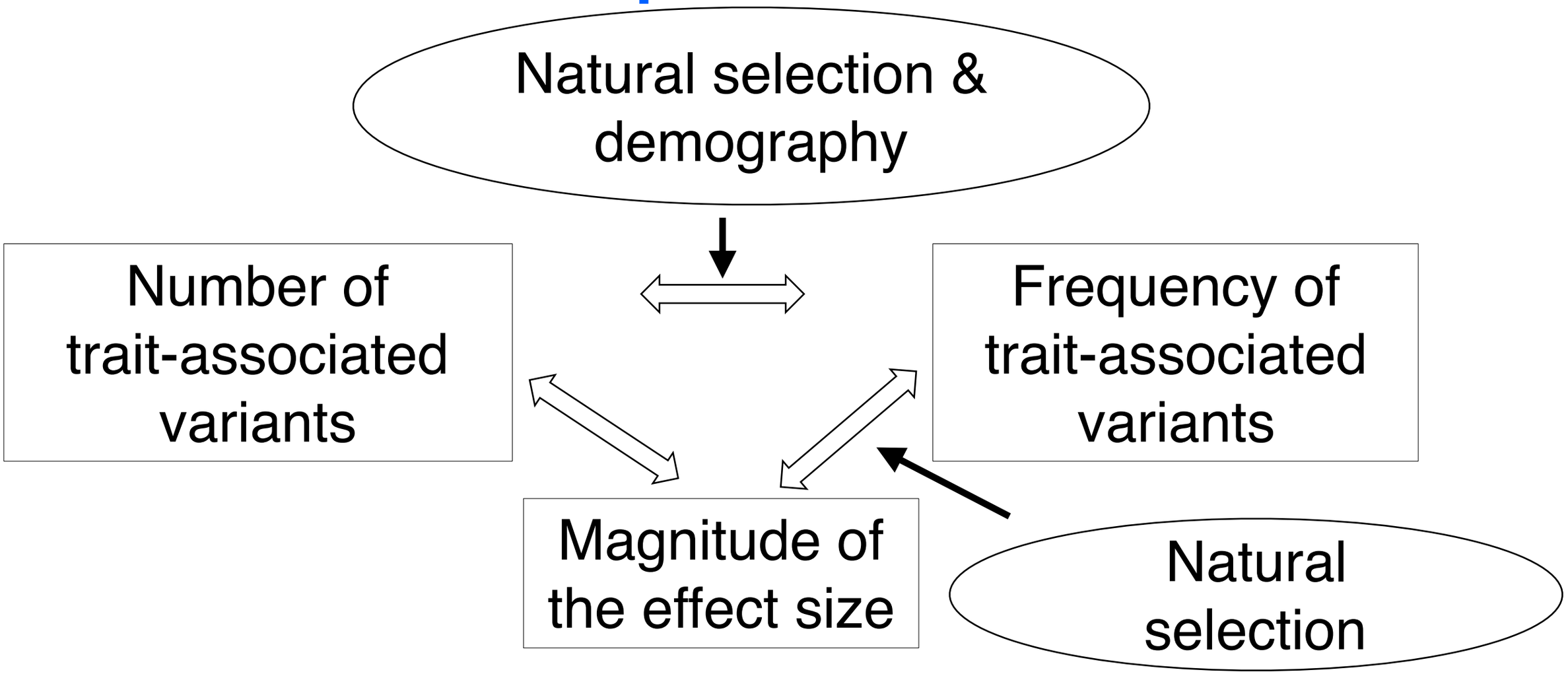
Inference of the mutational size supports the omnigenic model for complex traits

Kirk Lohmueller

Department of Ecology and Evolutionary Biology
Department of Human Genetics
Interdepartmental Bioinformatics Program
University of California, Los Angeles

Genetic architecture of complex traits



- Genome-wide association studies (GWAS) allows for better understanding of genetic architecture
  - Have identified thousands of trait-associated variants for complex traits

Timpson et al. 2018

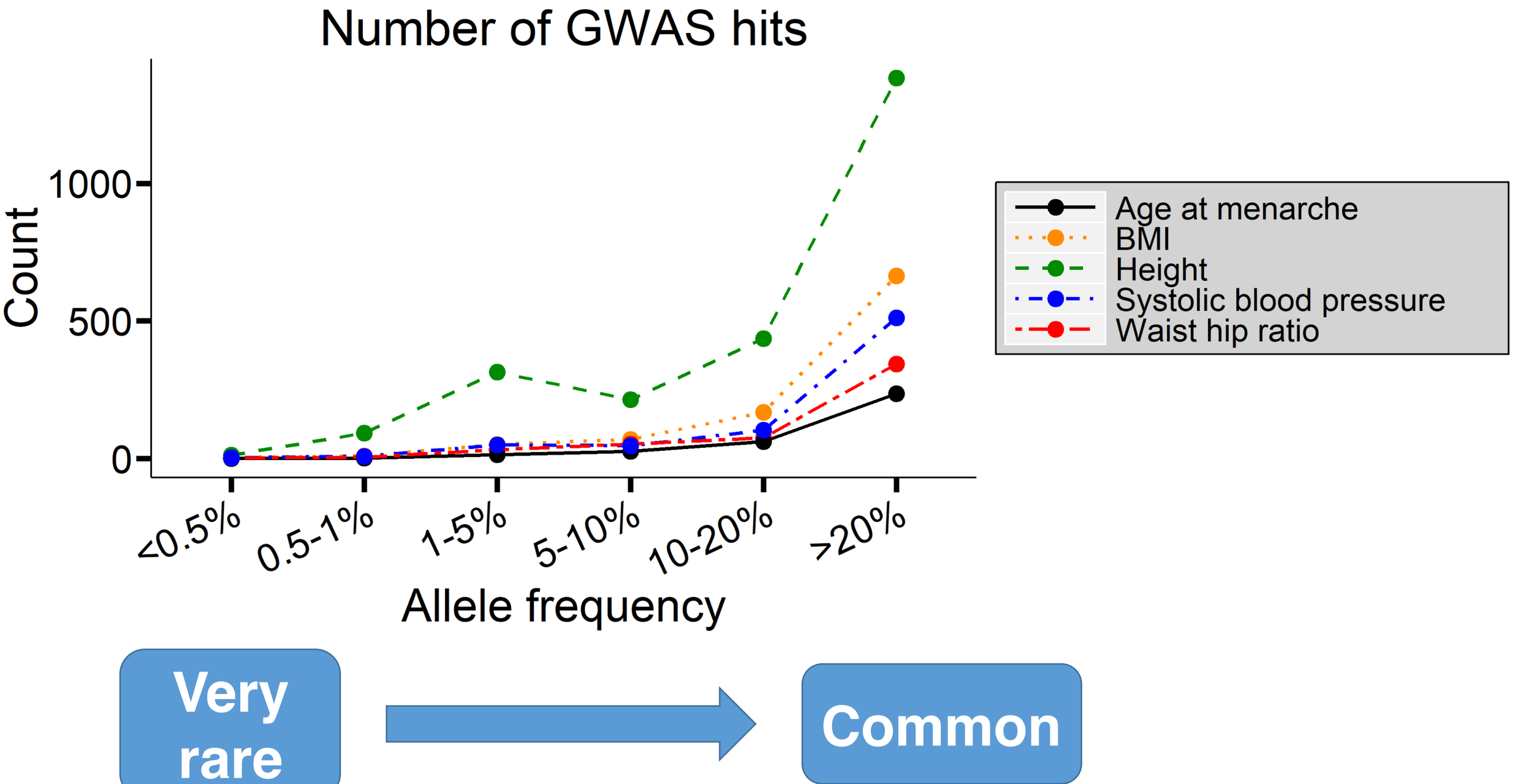
Outline

- Inference of genetic architecture from GWAS data & population genetic models
- How does genetic architecture differ across populations?

Outline

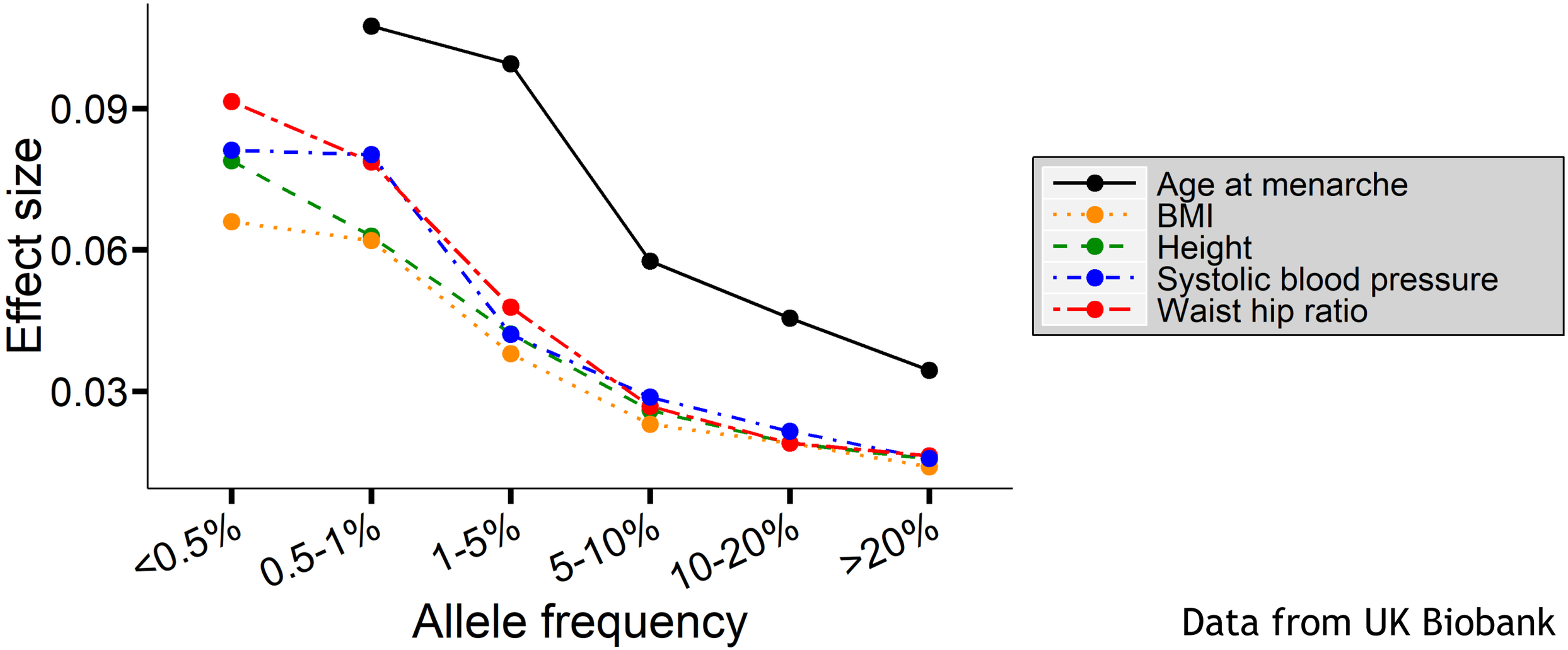
- Inference of genetic architecture from GWAS data & population genetic models
- How does genetic architecture differ across populations?

Most GWAS hits are common



Data from UK Biobank

Negative correlation between effect size and frequency

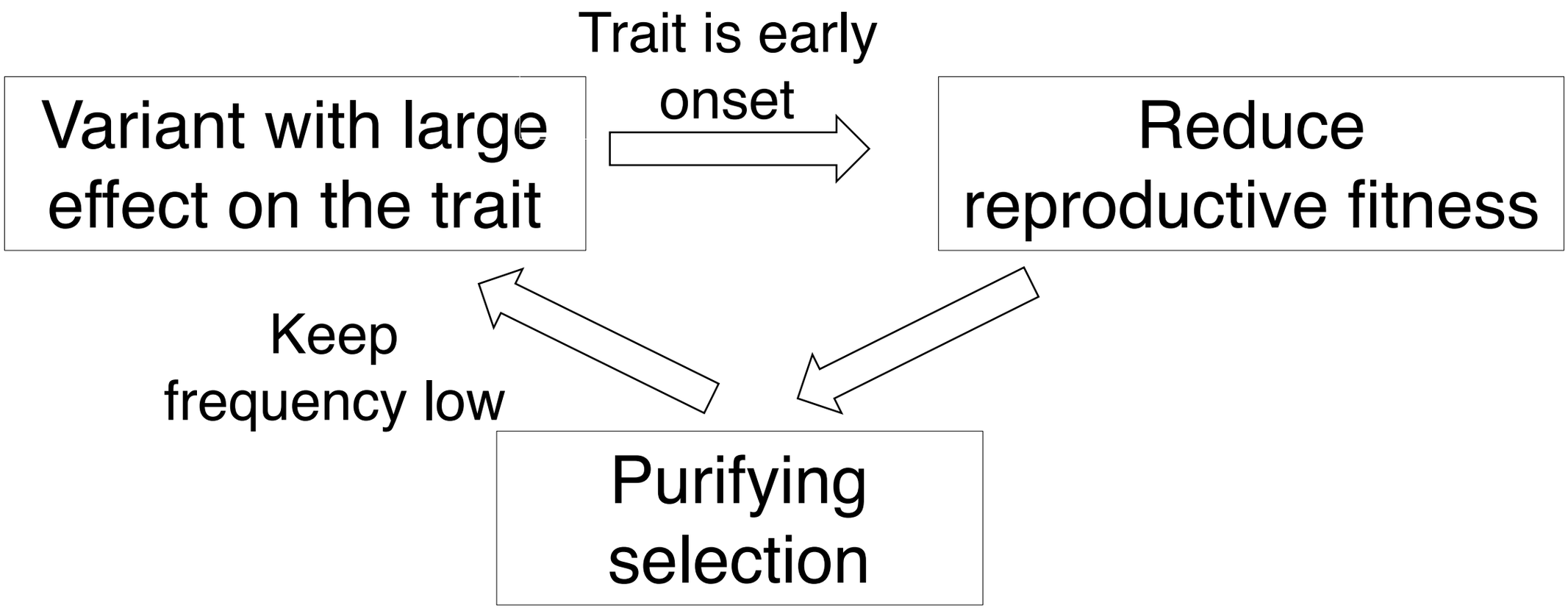


Data from UK Biobank

Very rare → **Common**

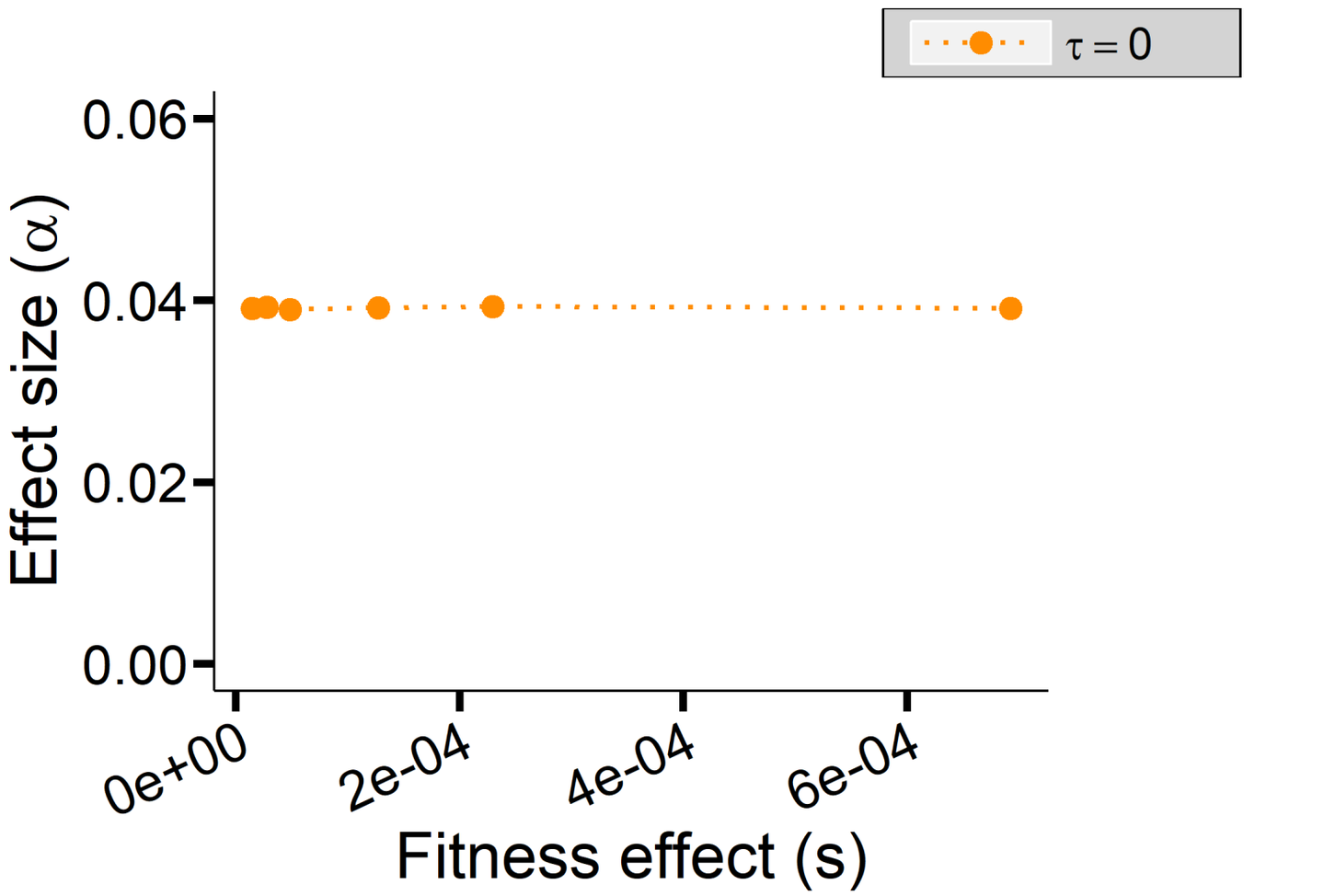
Park et al. (2011)
The UK10K Consortium (2015)

Purifying selection enriches for rare variants with large effect



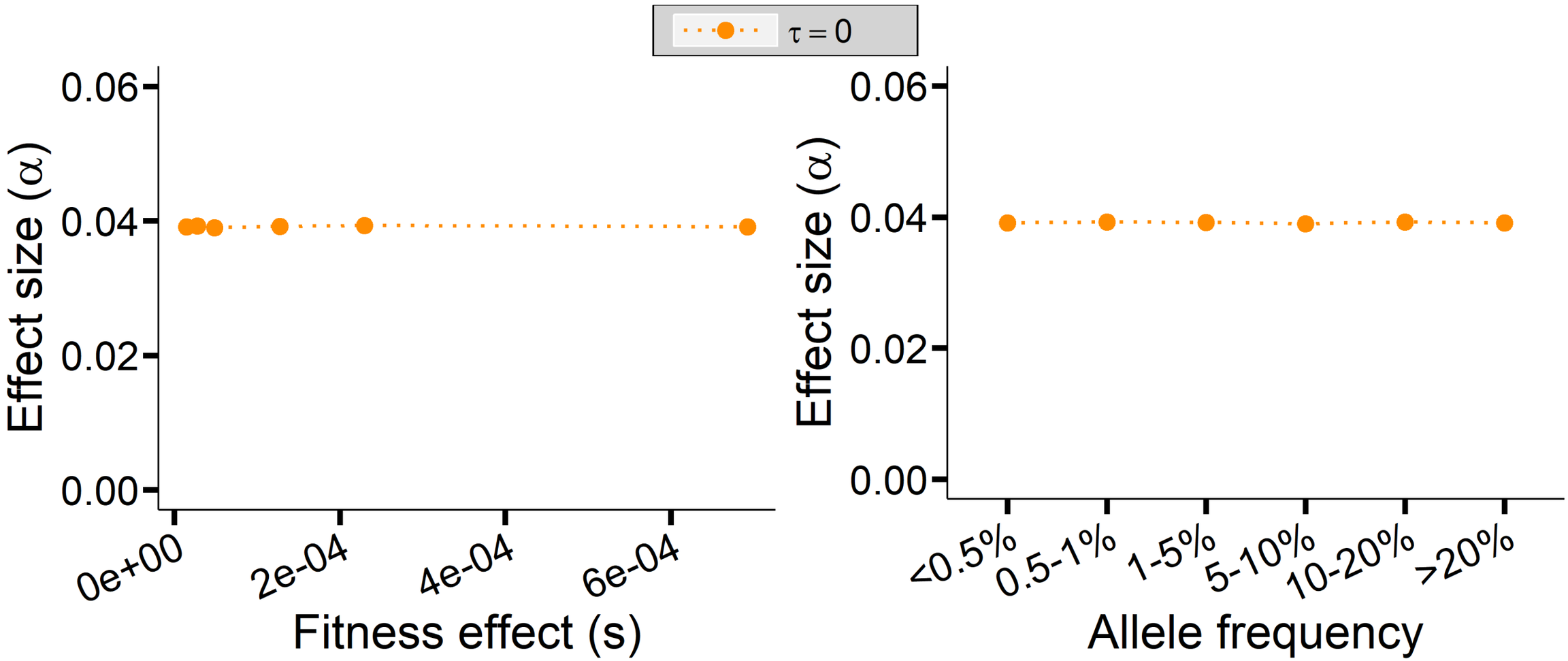
- Eyre-Walker (2010):
  - Propose a parameter called τ
  - τ captures the relationship between a variant's effect on the trait with its effect on fitness

τ relates effects size to selection coefficient and allele frequency



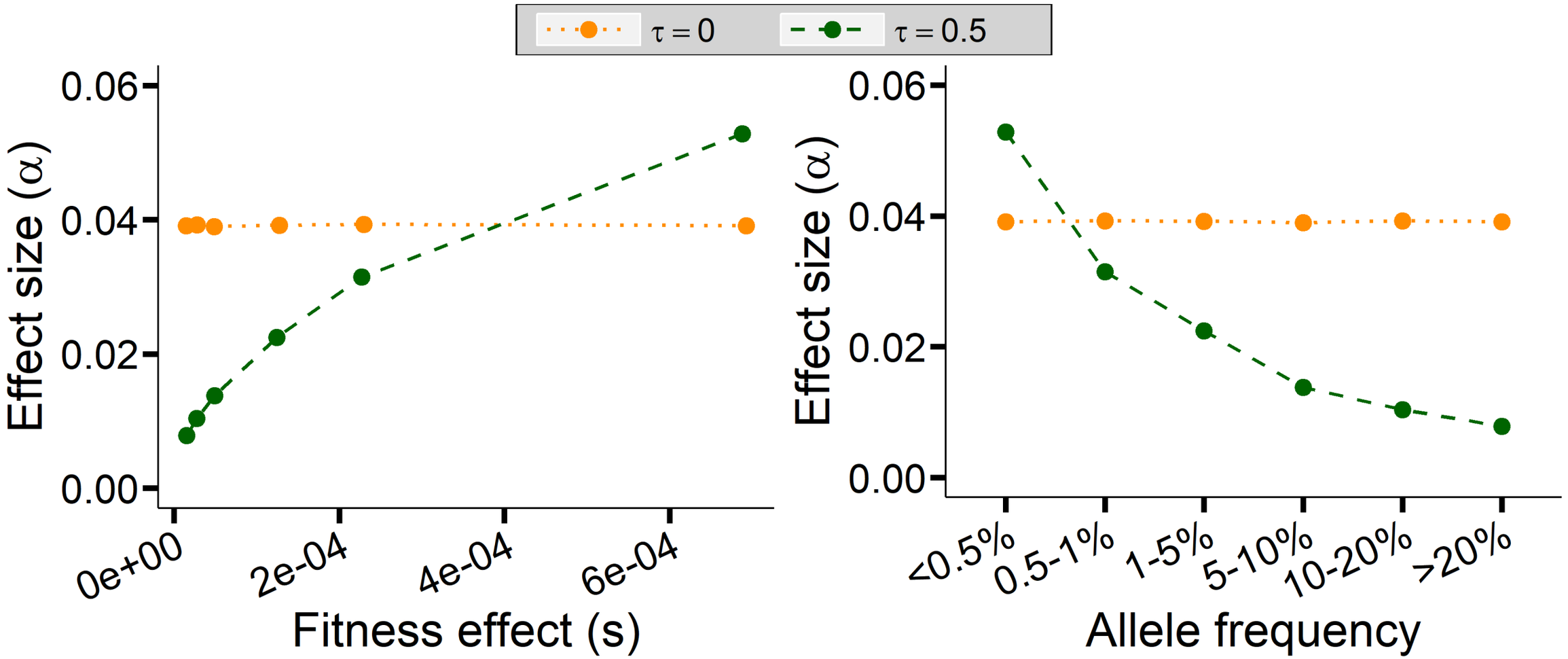
Less deleterious → **More deleterious**

τ relates effects size to selection coefficient and allele frequency



Less deleterious → **More deleterious**

τ relates effects size to selection coefficient and allele frequency



Less deleterious → **More deleterious**

Support for a relationship between effect size and selection

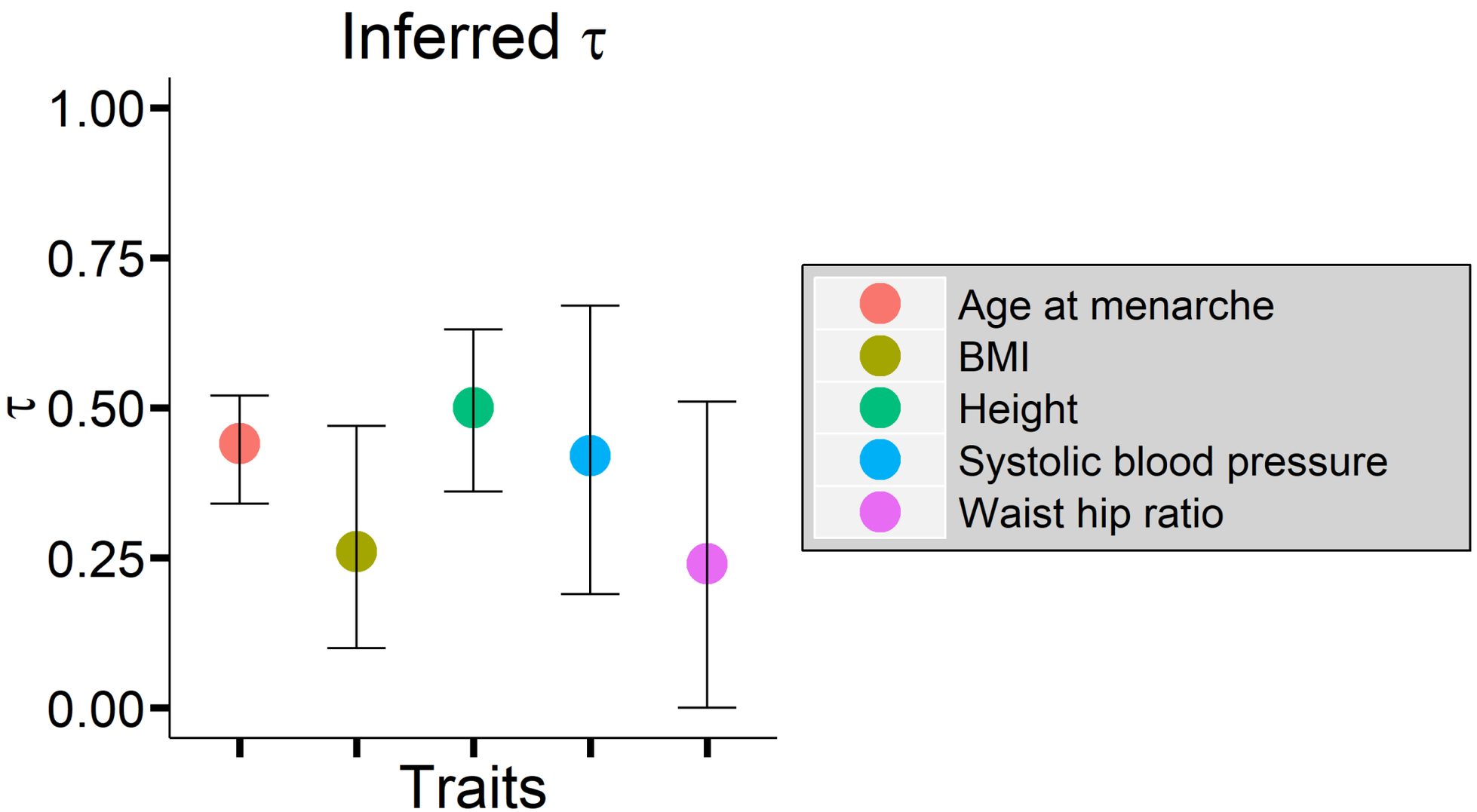
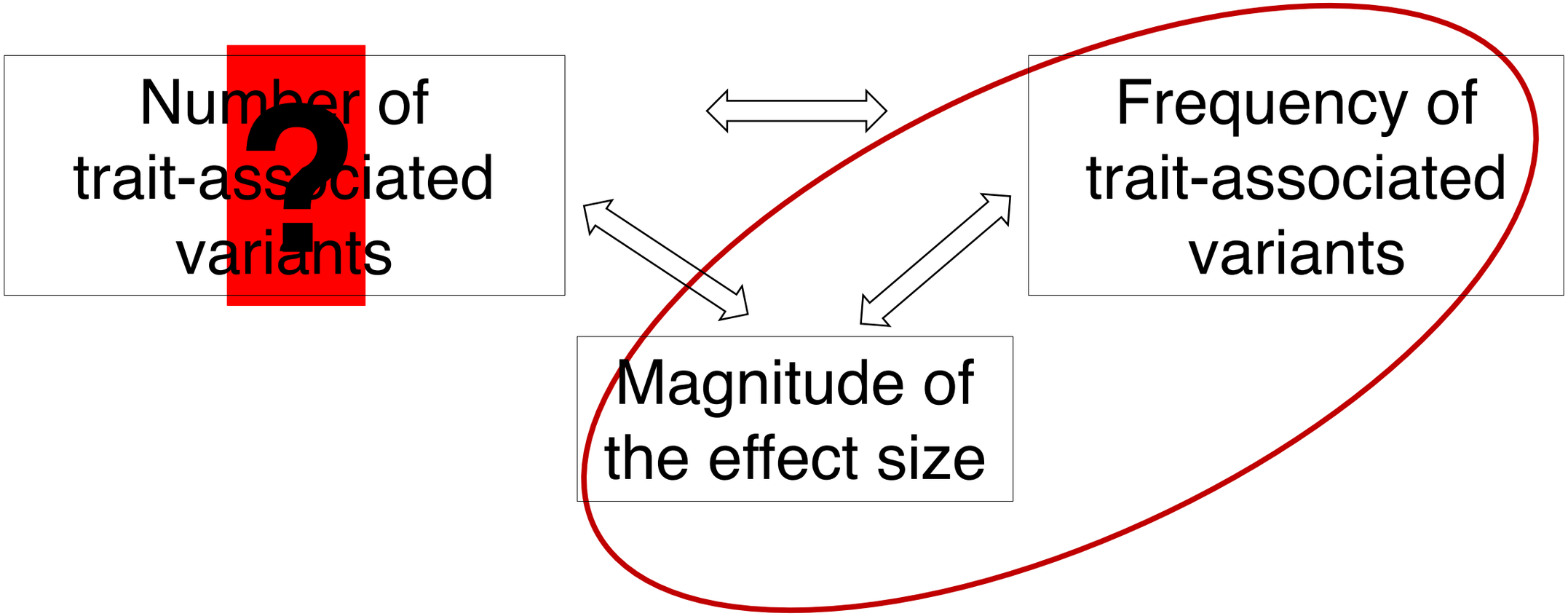


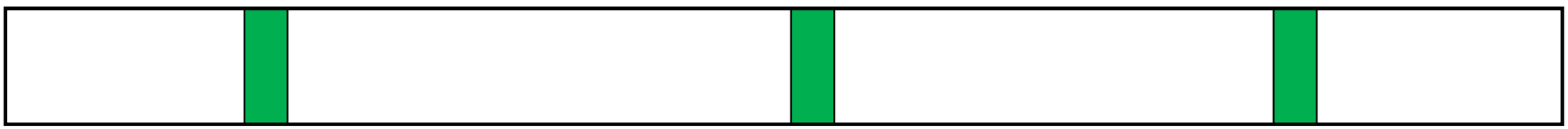
Figure drawn using values from Schoech et al. (2017)

The number of causal variants is understudied



The number of causal variants is difficult to study

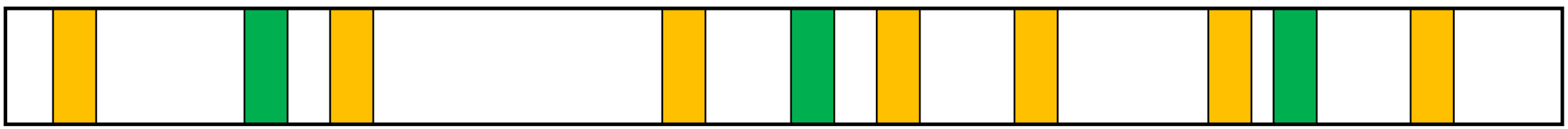
GWAS hits (known)



The number of causal variants is difficult to study

GWAS hits (known)

Causal variants (not known)

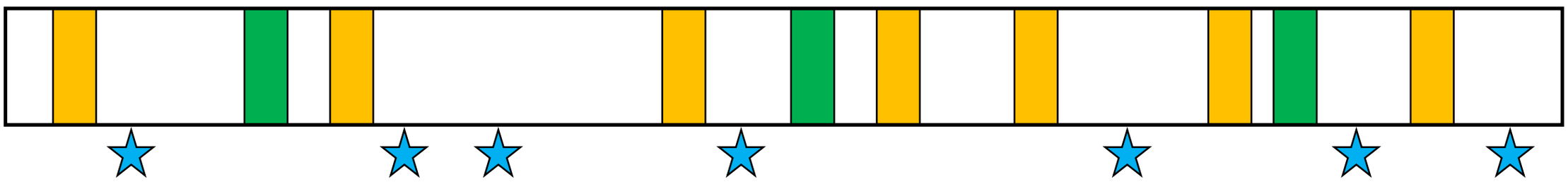


The number of causal variants is difficult to study

GWAS hits (known)

Causal variants (not known)

A site that hasn't mutated but could be a trait-affecting variant (not known)

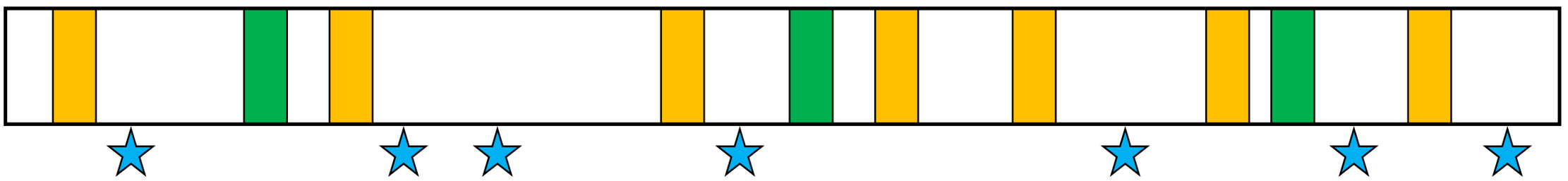


The number of causal variants is difficult to study

GWAS hits (known)

Causal variants (not known)

A site that hasn't mutated but could be a trait-affecting variant (not known)



+ + =

The total number of sites in the genome that, if mutated, would give rise to a trait-affecting variant

Mutational target size (*M*)

Goal: develop an improved model of complex traits

- Infer *M*:
  - *M* is not known for many traits
  - The number of causal variants can be inferred from knowing *M*
- Also, improve on existing methods to infer $\boldsymbol{\tau}$:
  - Existing method (i.e. Shoech et al. 2017) used genotyped data
  - Our method uses summary statistics from GWAS
- Developed Inference of Genetic Architecture method (InGeAr)
  - An Approximate Bayesian Computation framework to infer for $\boldsymbol{\tau}$ and *M*

InGeAr framework

GWAS simulation

Simulate causal variants with a value of M drawn from a prior distribution

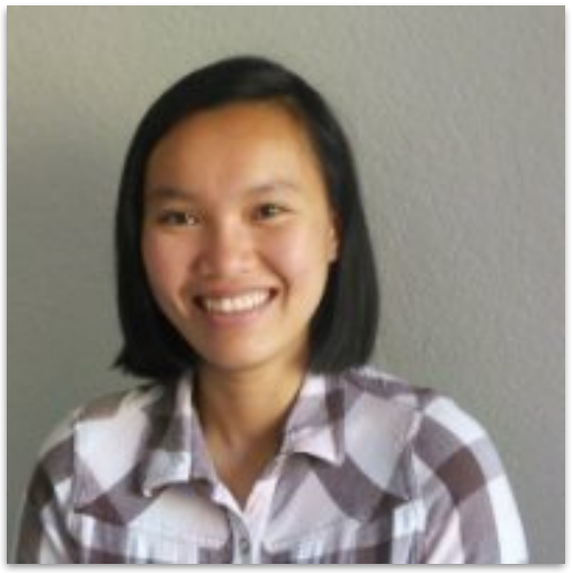
⇨

Calculate effect size: $\alpha_{\mathrm{SNP}i} = s^{\tau}$, where τ is drawn from a prior distribution

⇨

Determine number of significant variants based on power (simulating GWAS)

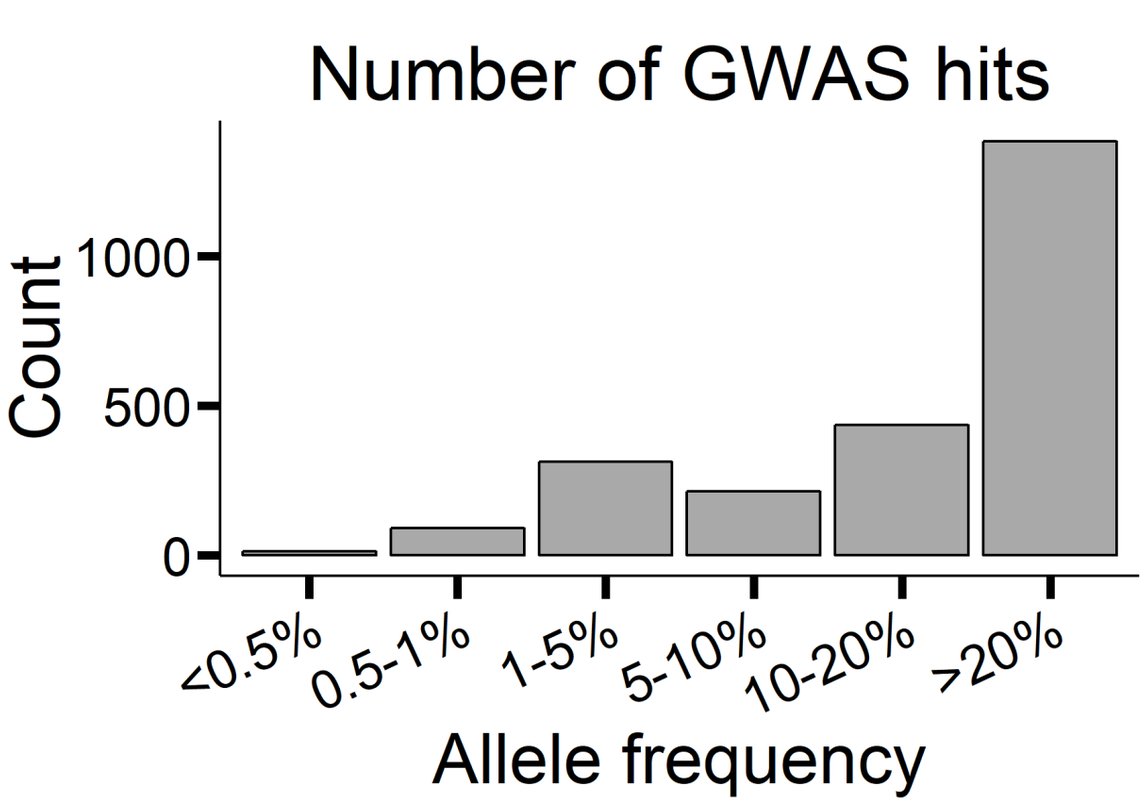
Rejection algorithm

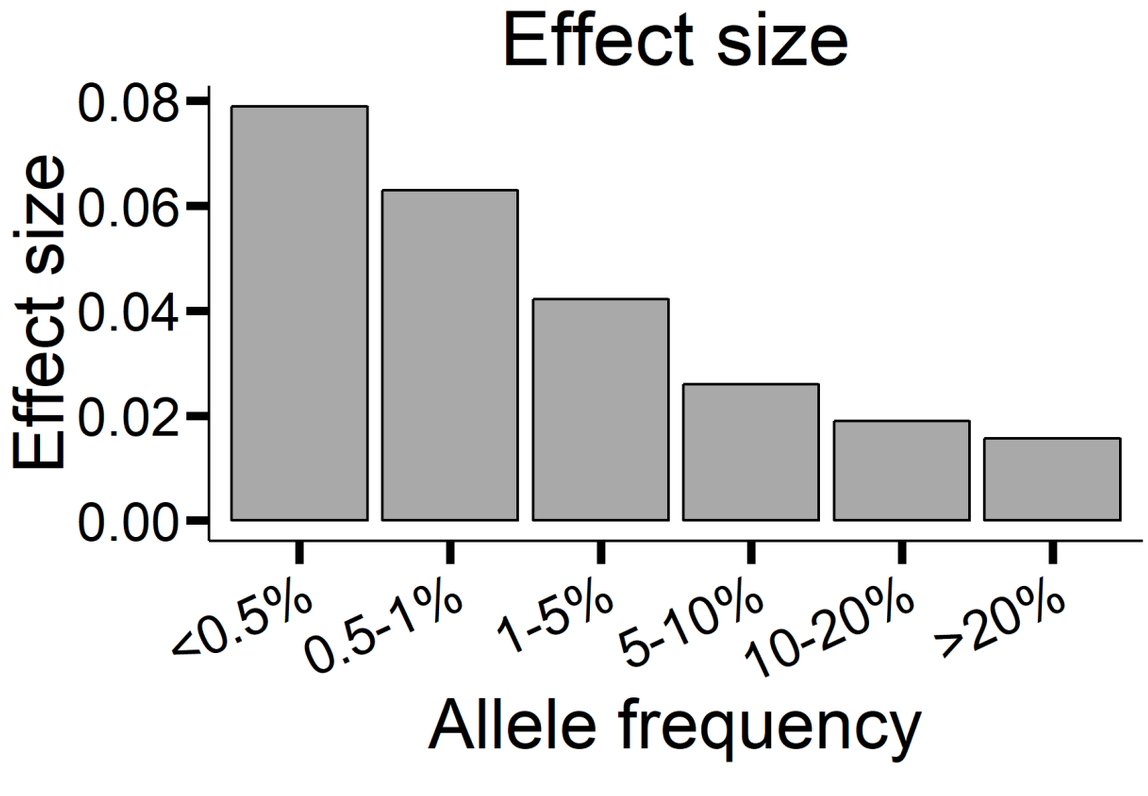


Tanya Phung

Bioinformatics graduate student
Currently a postdoc with Melissa Wilson Sayres

Summary statistics





InGeAr framework

GWAS simulation

Simulate causal variants with a value of M drawn from a prior distribution

⇨

Calculate effect size: $\alpha_{\mathrm{SNP}i} = s^{\tau}$, where τ is drawn from a prior distribution

⇨

Determine number of significant variants based on power (simulating GWAS)

Rejection algorithm

Step 1: Compute:

$$\sum_{\mathrm{bin}_{i=1}}^{6} \frac{|\alpha_i^{\mathrm{emp}} - \alpha_i^{\mathrm{sim}}|}{\alpha_i^{\mathrm{emp}}} < 0.6$$

→ Accept (M, τ)

→ Repeat until 10,000 acceptances are obtained

⇨

Step 2: Compute:

$$\sum_{\mathrm{bin}_{i=1}}^{6} \left(\mathrm{GWAS\ hits}_i^{\mathrm{emp}} - \mathrm{GWAS\ hits}_i^{\mathrm{sim}}\right)^2$$

→ Select the best 10%

→ Posterior distribution of M and τ

Remove linkage disequilibrium (LD) by considering independent GWAS hits

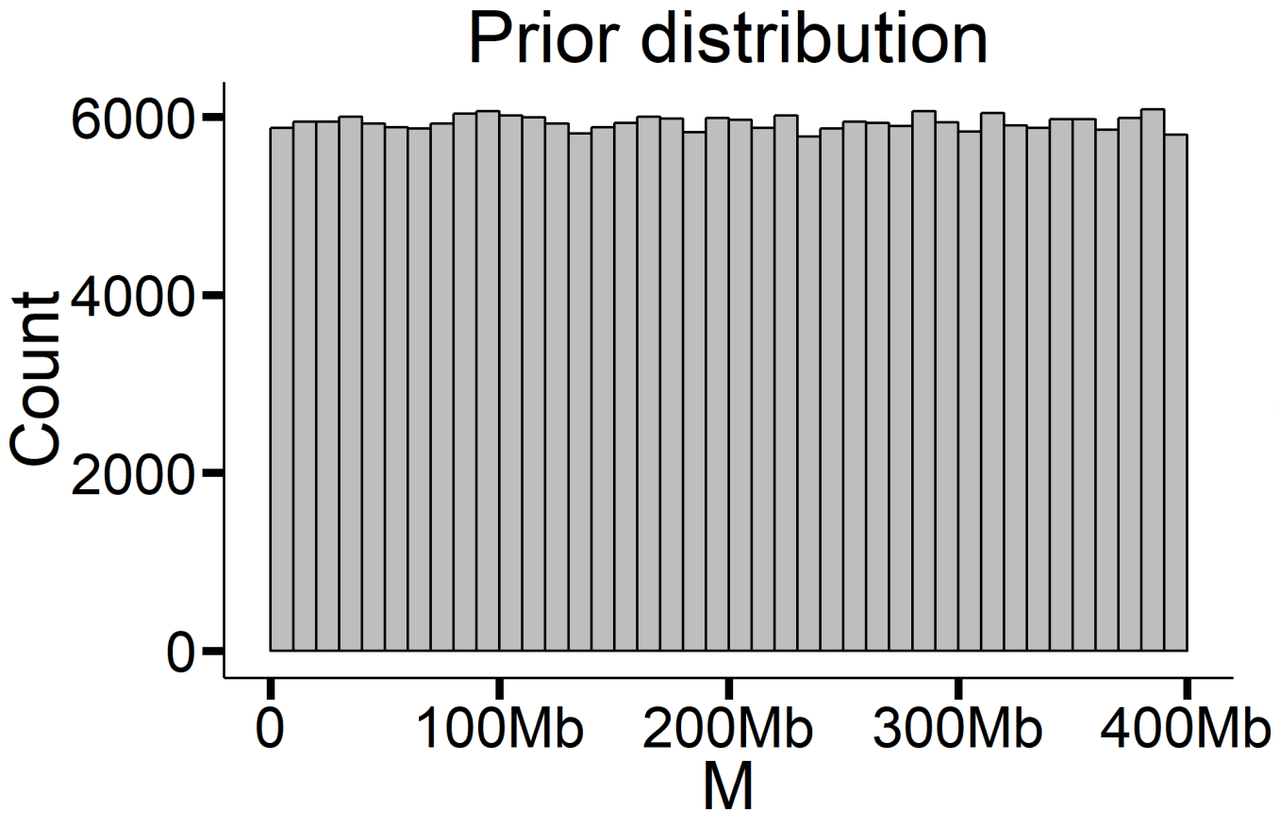
- Kichaev et al. (2017) developed FINDOR to identify independent, genome-wide significant GWAS hits

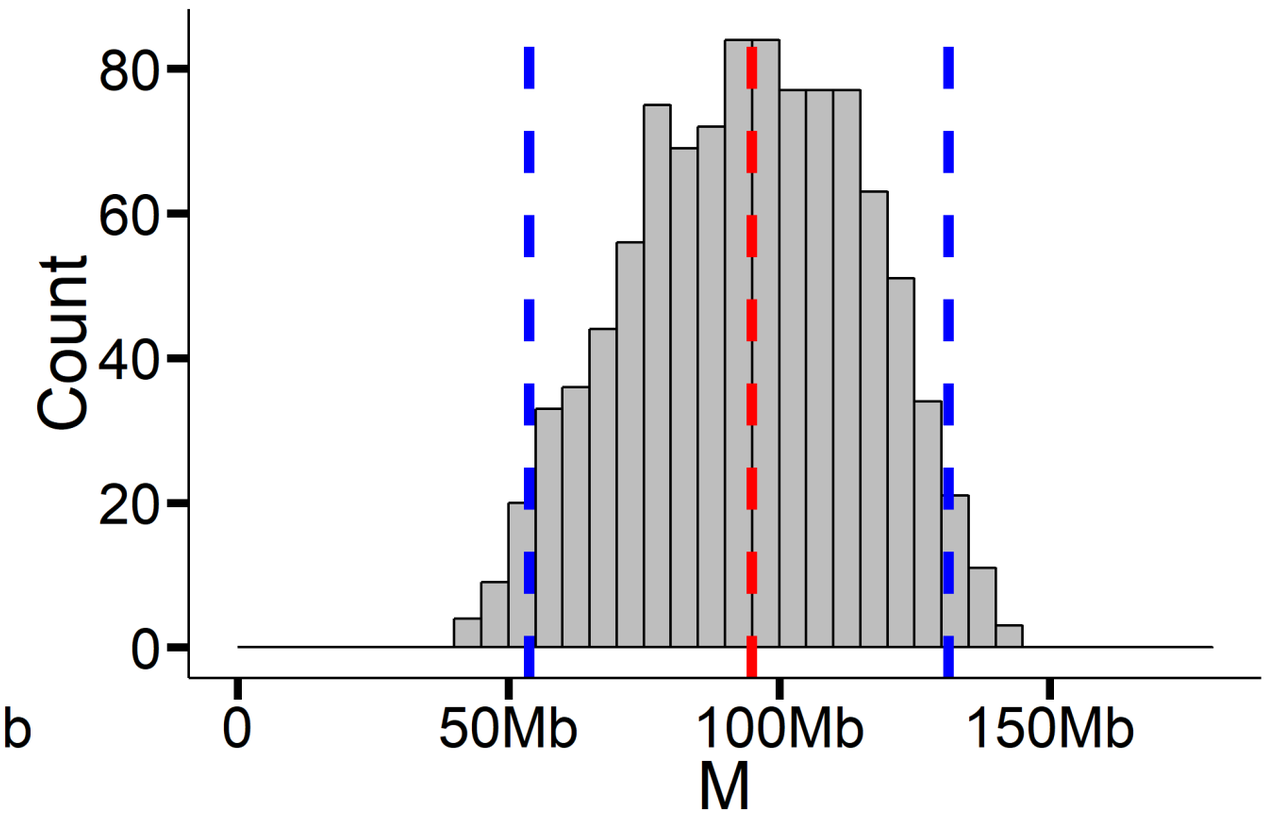
  – Weight GWAS hits by how well they tag functional categories that are enriched for heritability

  – Identify ~ 2,500 independent GWAS hits for height

Application of InGeAr to height GWAS from UKBiobank

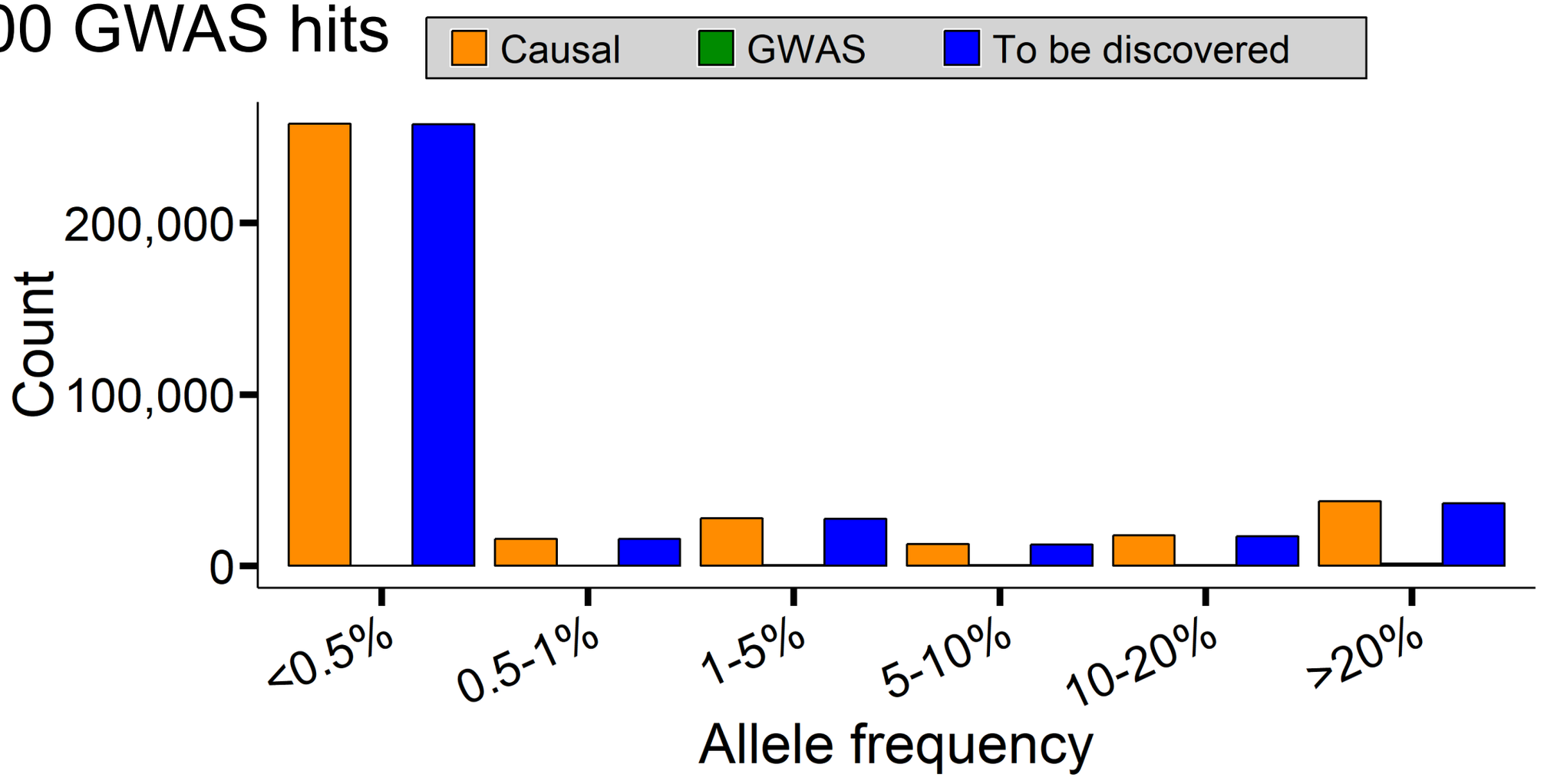
Mutational target size for height: 95Mb



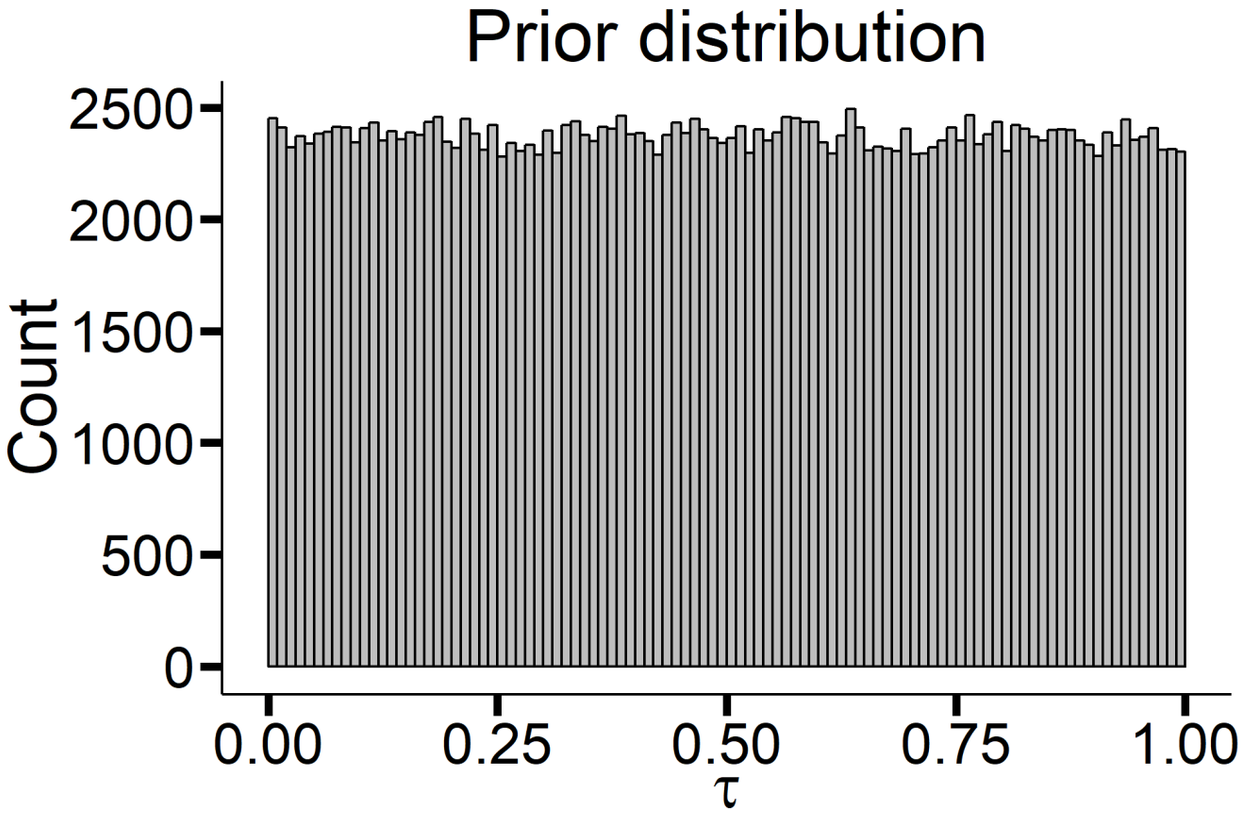


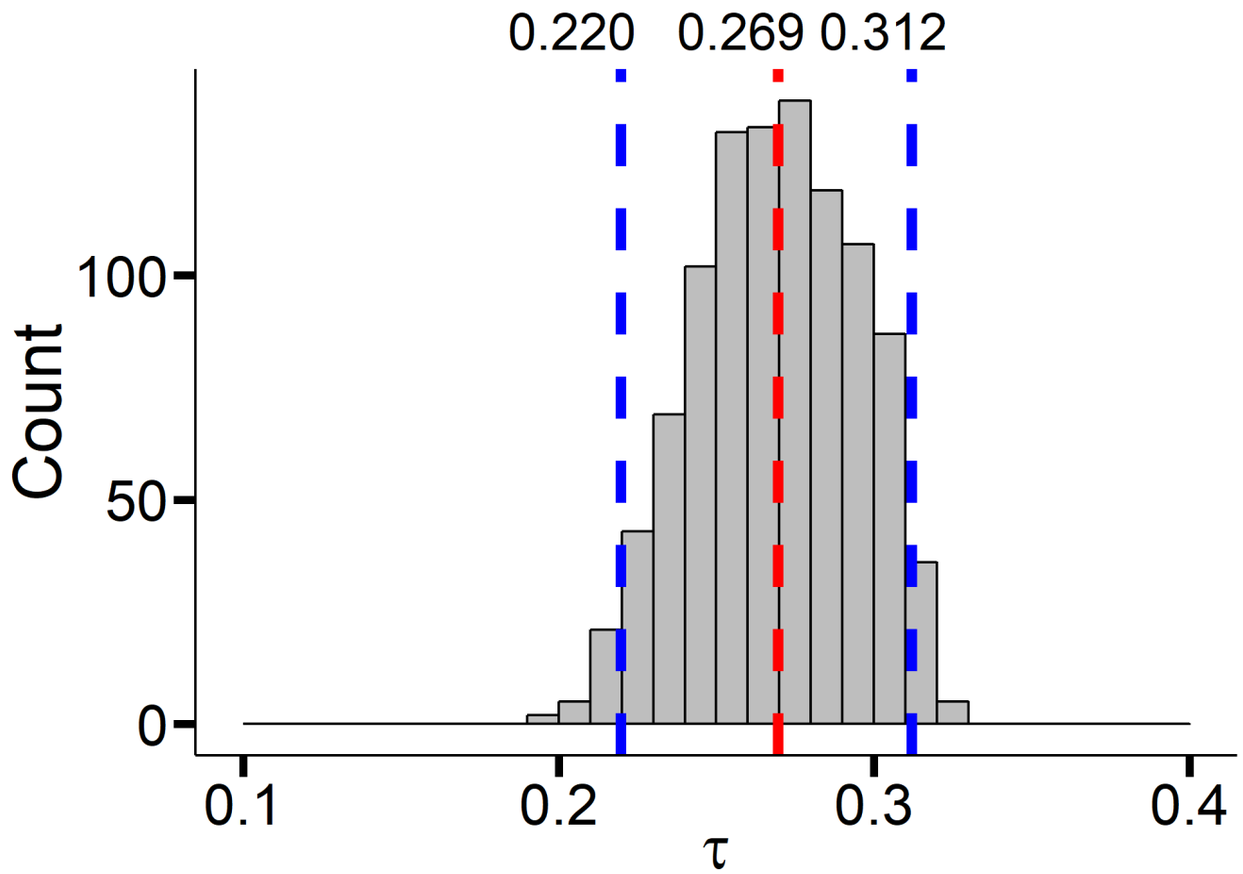
Mutational target size for height: 95Mb

- For a mutational target size of 95Mb (~3% of the genome)
  - ~300,000 causal variants
  - ~2,500 GWAS hits

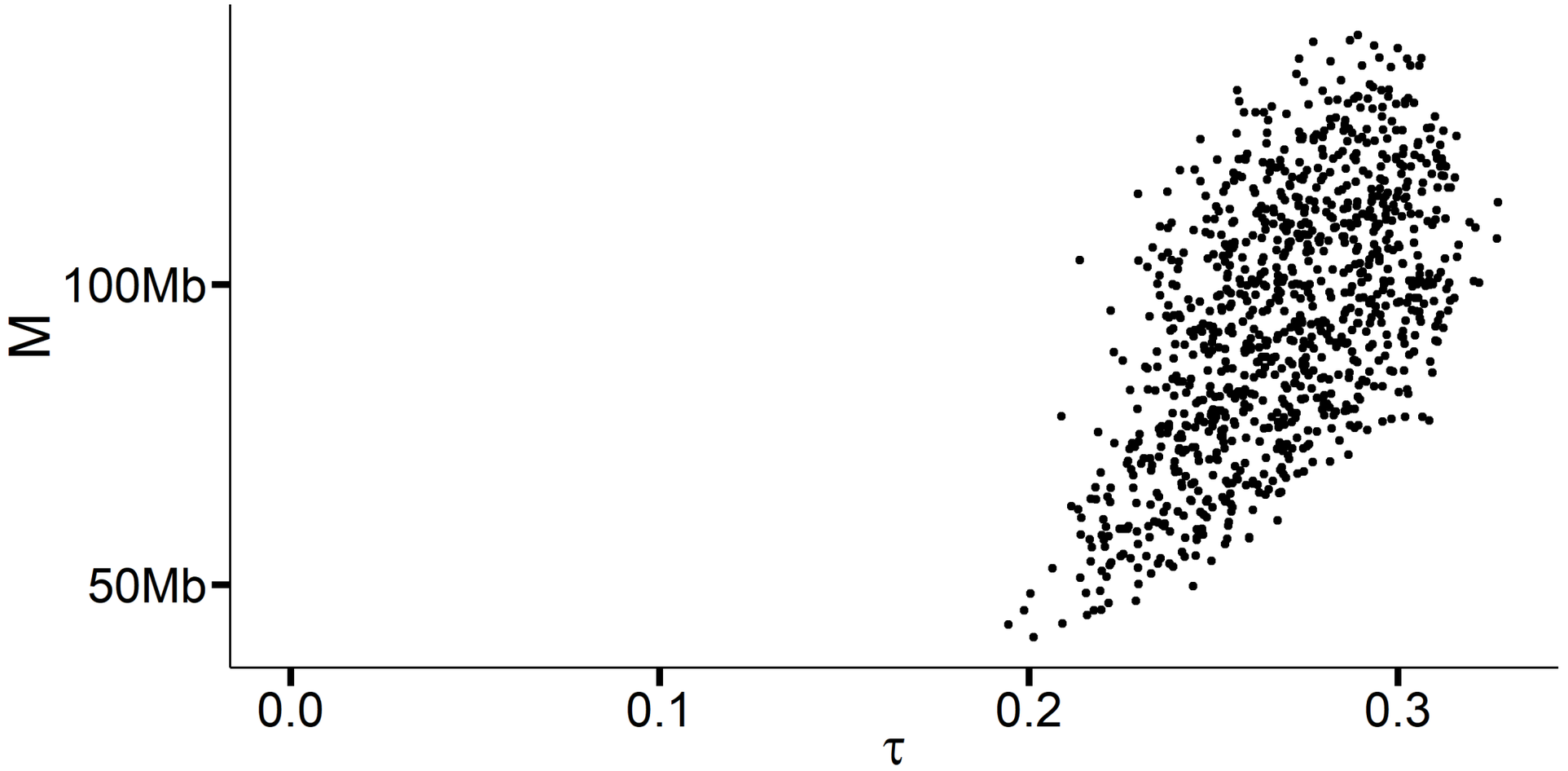


Coupling between selection and trait effect for height

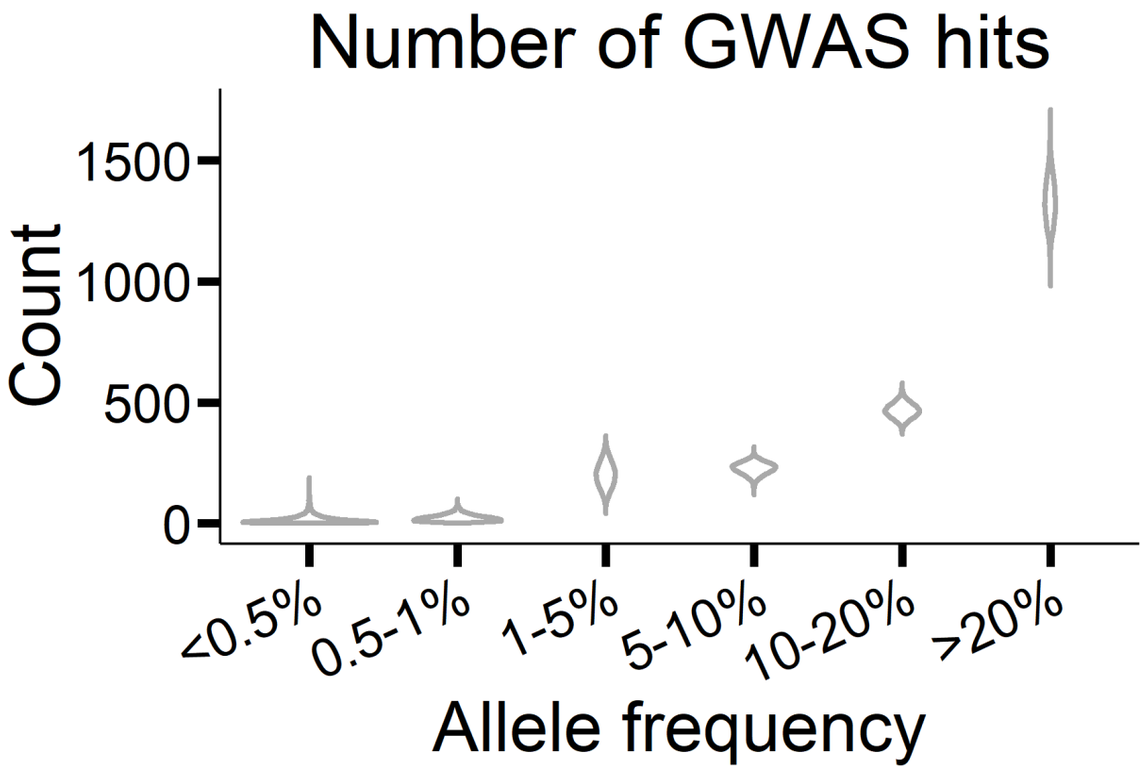


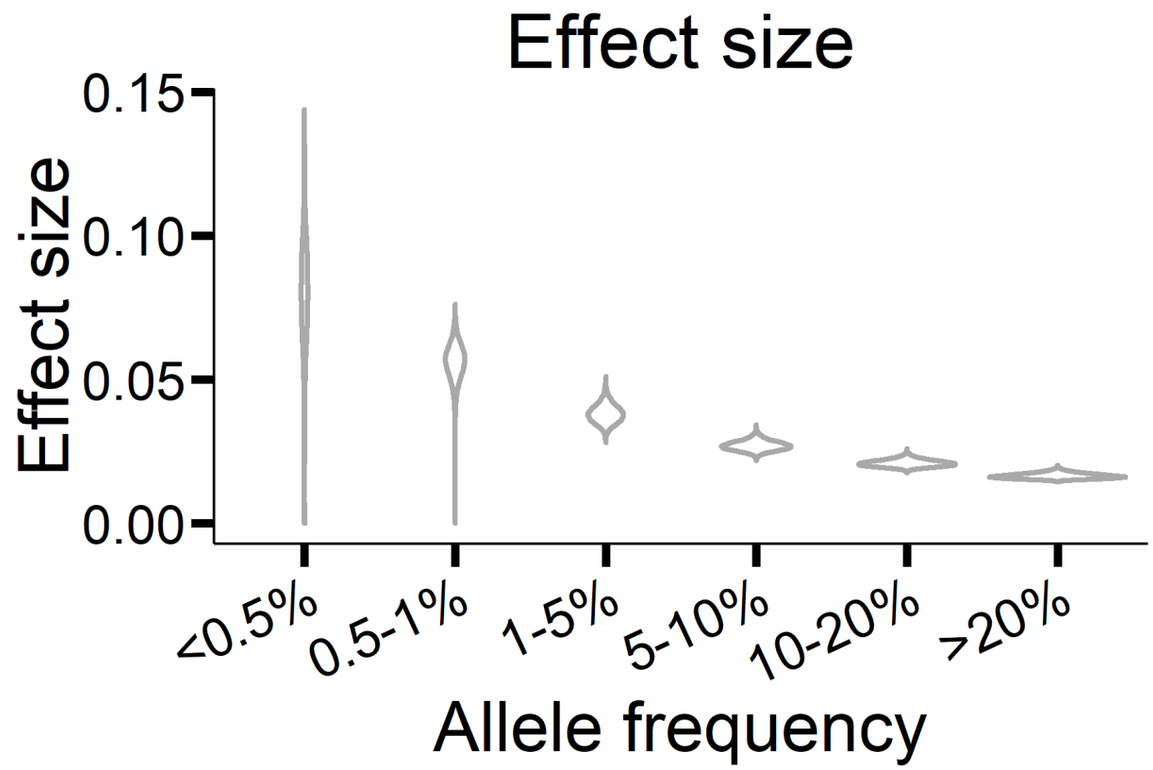


Joint posterior distribution τ of and M

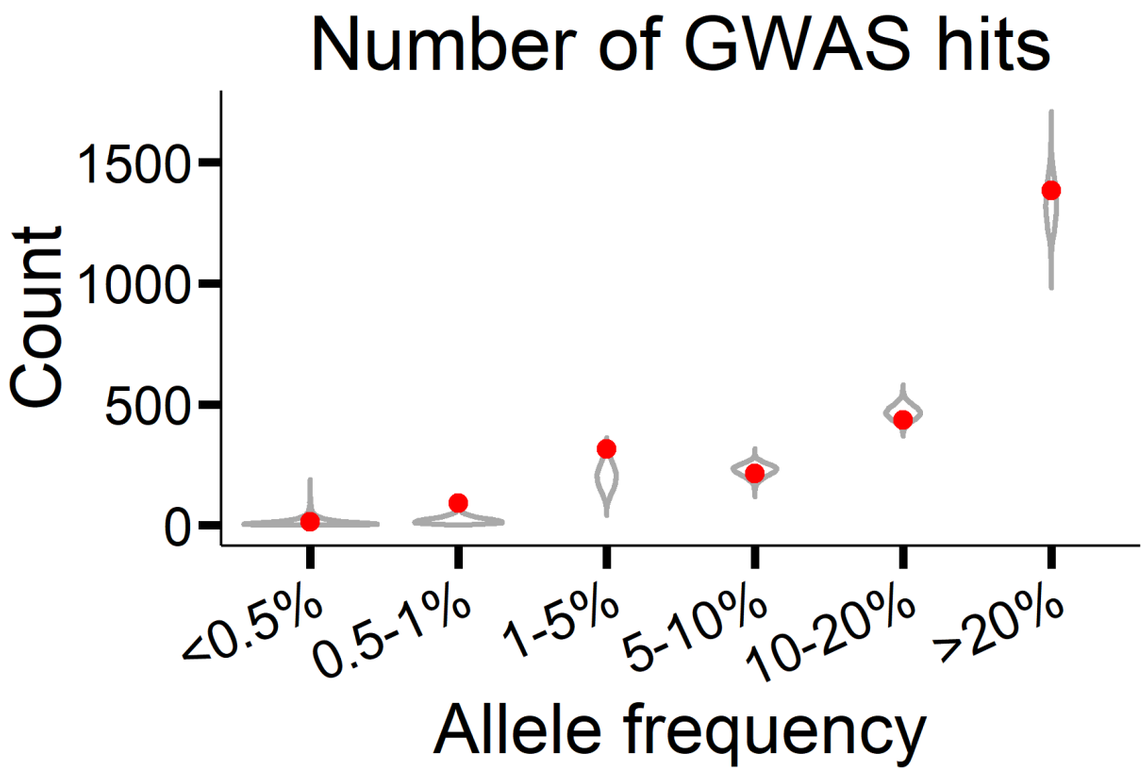


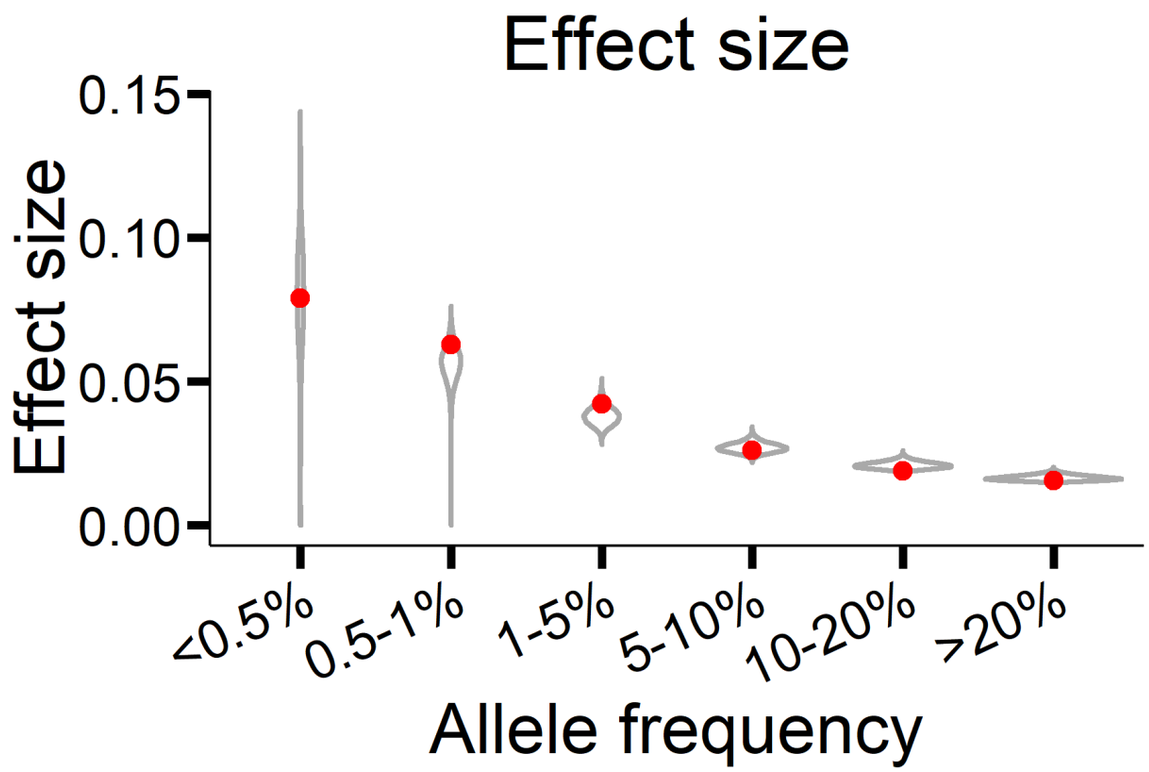
Assess model fit



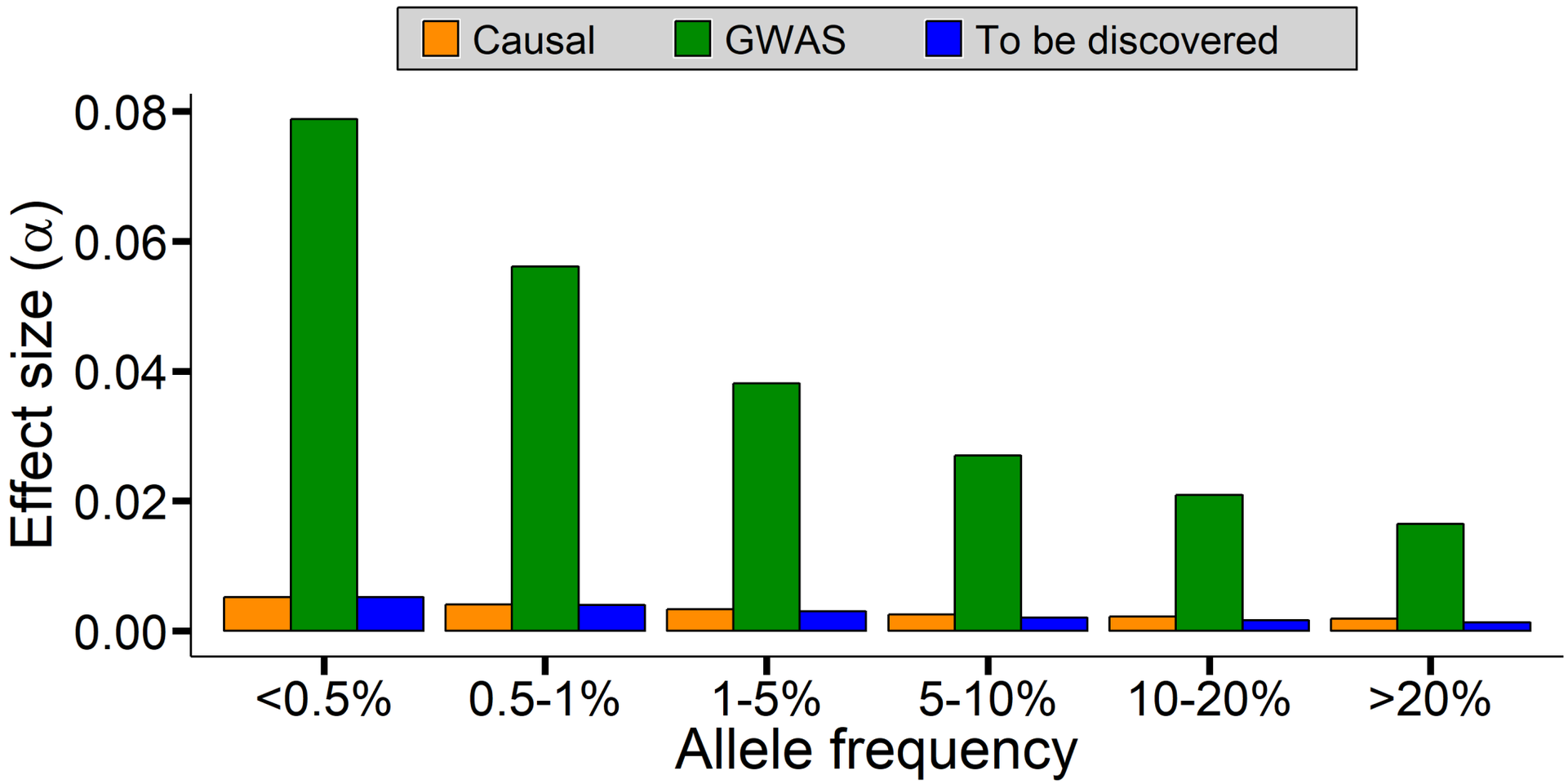


Model fits the empirical GWAS data well

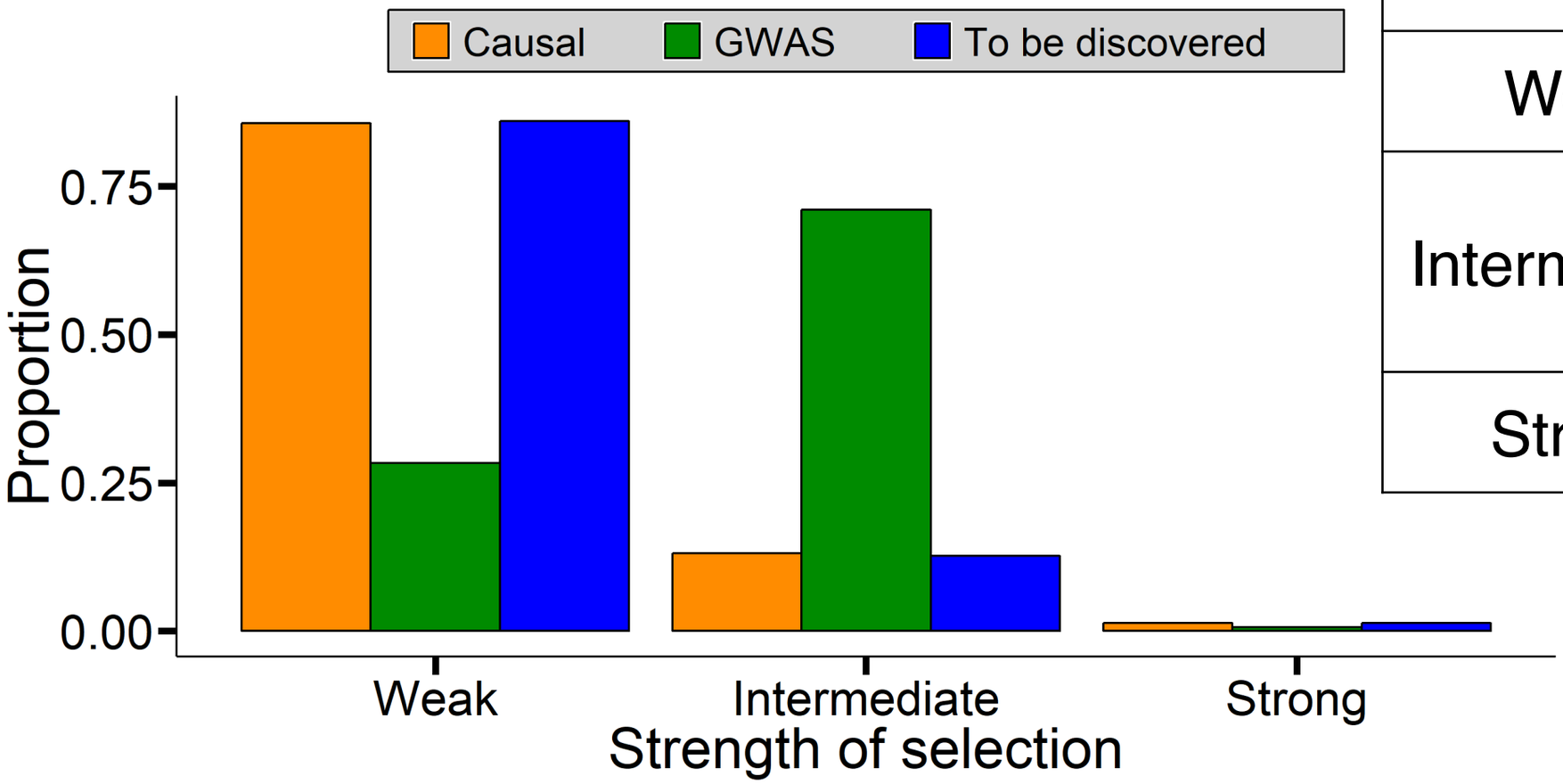




GWAS hits are enriched for variants with large effect size

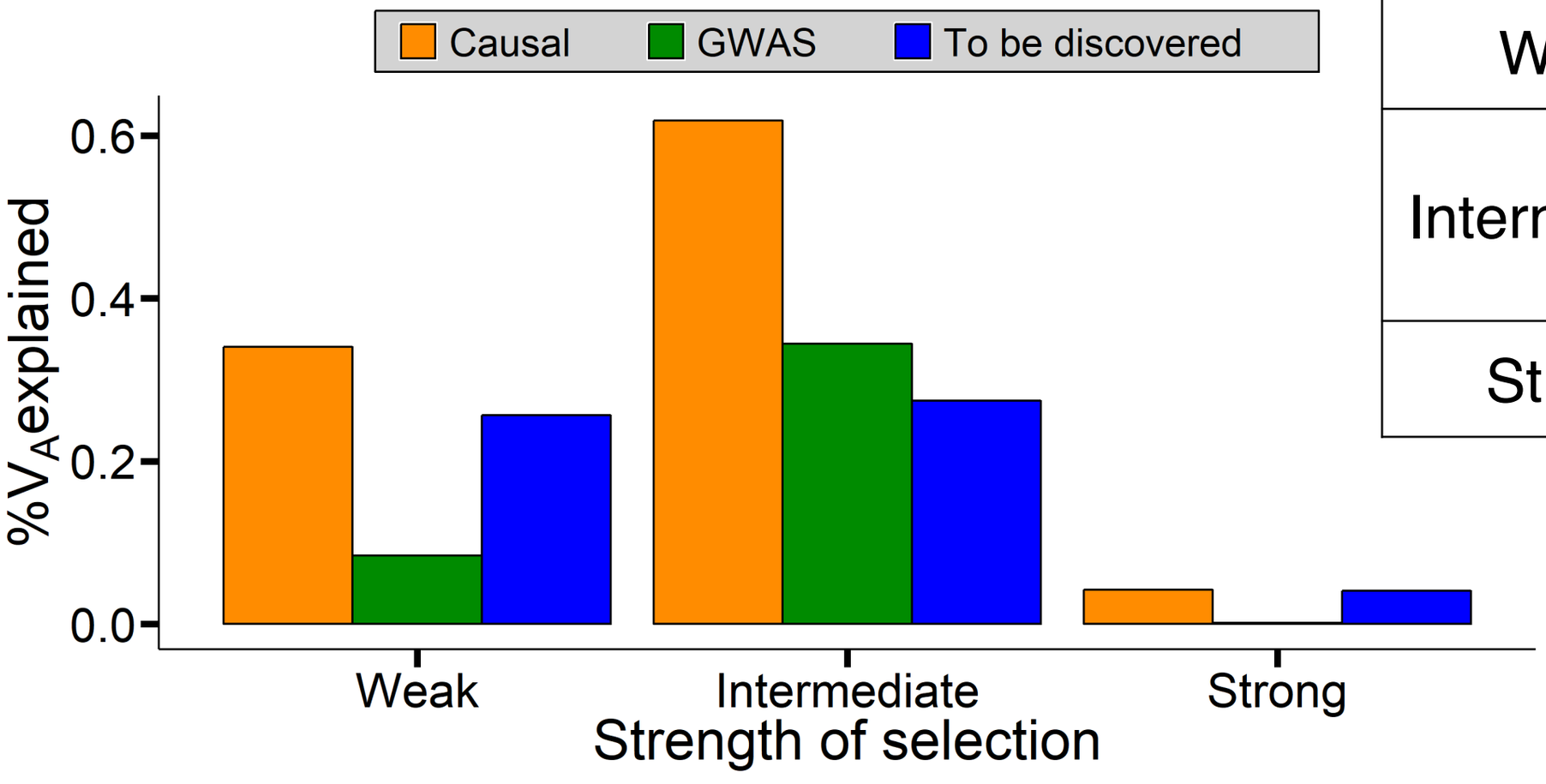


Most causal variants are weakly deleterious



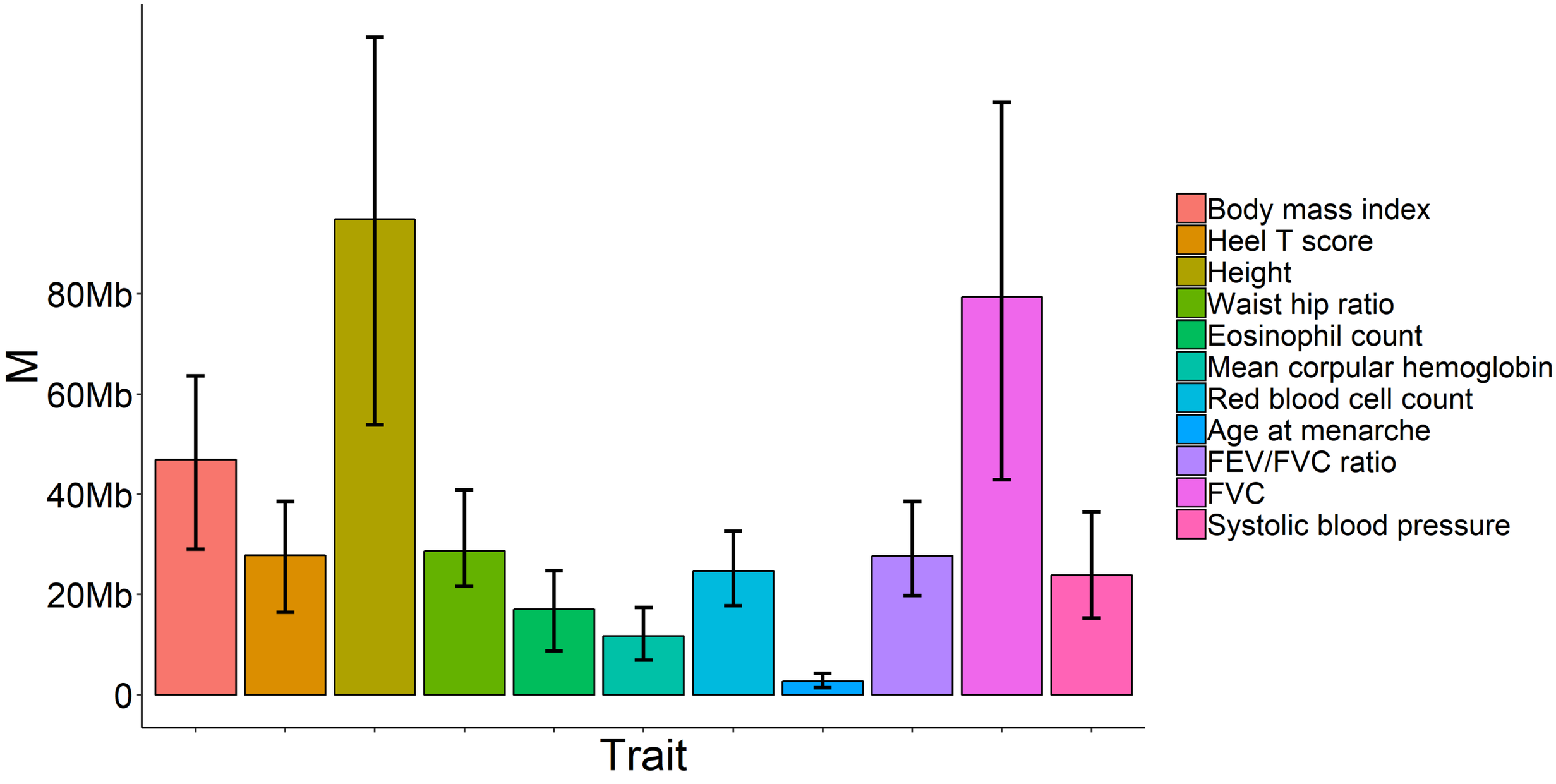
Strength	$\lvert s \rvert$
Weak	$<10^{-4}$
Intermediate	10^{-4}–10^{-2}
Strong	$>10^{-2}$

Weakly and intermediately selected variants explain most of the additive genetic variance

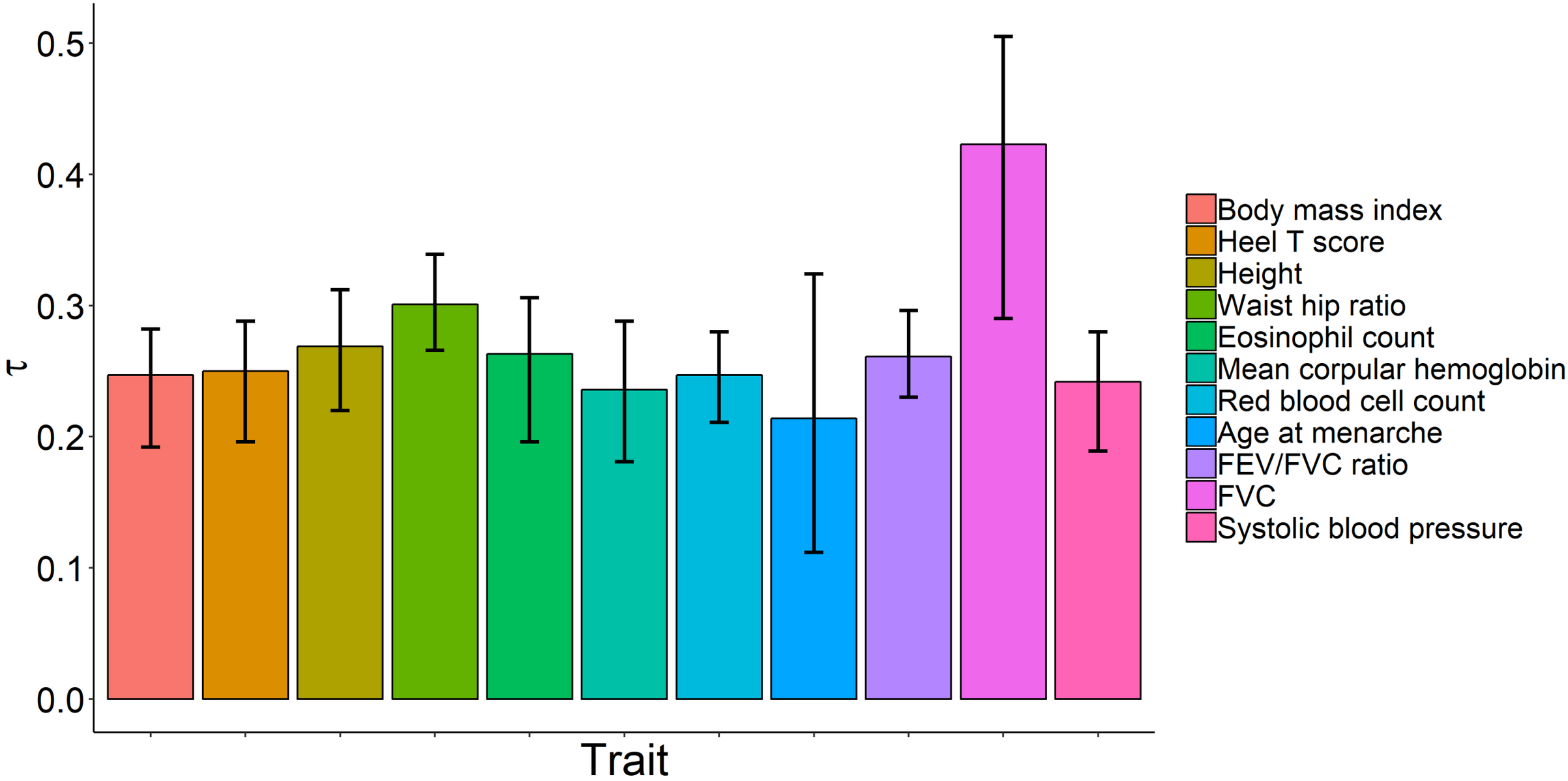


Strength	$\lvert s \rvert$
Weak	$<10^{-4}$
Intermediate	10^{-4}–10^{-2}
Strong	$>10^{-2}$

M varies across examined traits



τ is similar across examined traits



Our results support the omnigenic model

- The omnigenic model (Boyle et al. 2017) predicts:
  1. A large proportion of the genome (peripheral genes) affects most traits
  2. Most of the heritability is explained by the weak effects from peripheral genes

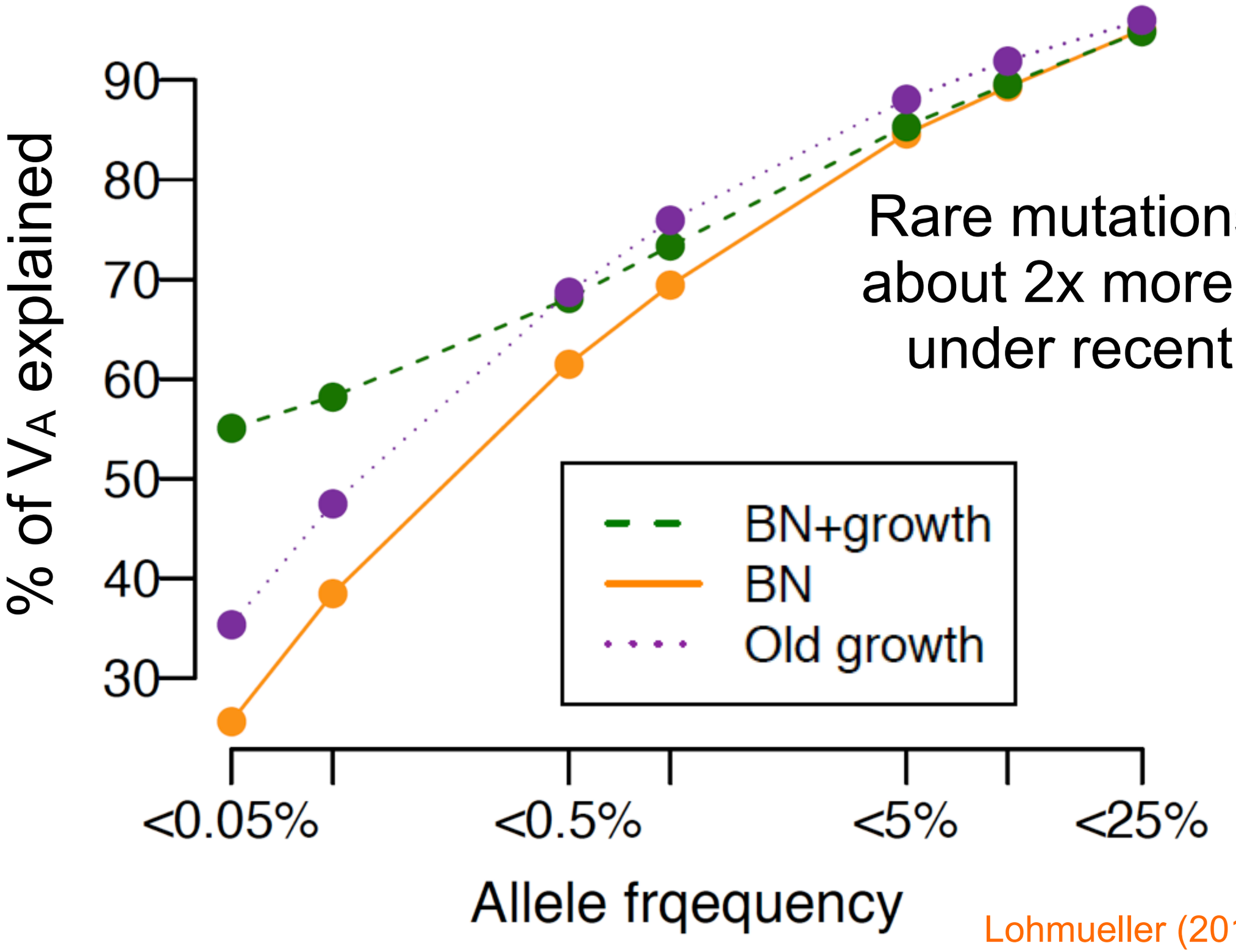
- *M* is on orders of ten of megabases for most traits
  - Supports Prediction 1

- τ is similar for all traits examined
  - Supports Prediction 2

Outline

- Inference of genetic architecture from GWAS data & population genetic models
- How does genetic architecture differ across populations?

Additive variance when trait effects are proportional to fitness effects (τ = 0.5)



Rare mutations explain about 2x more variance under recent growth

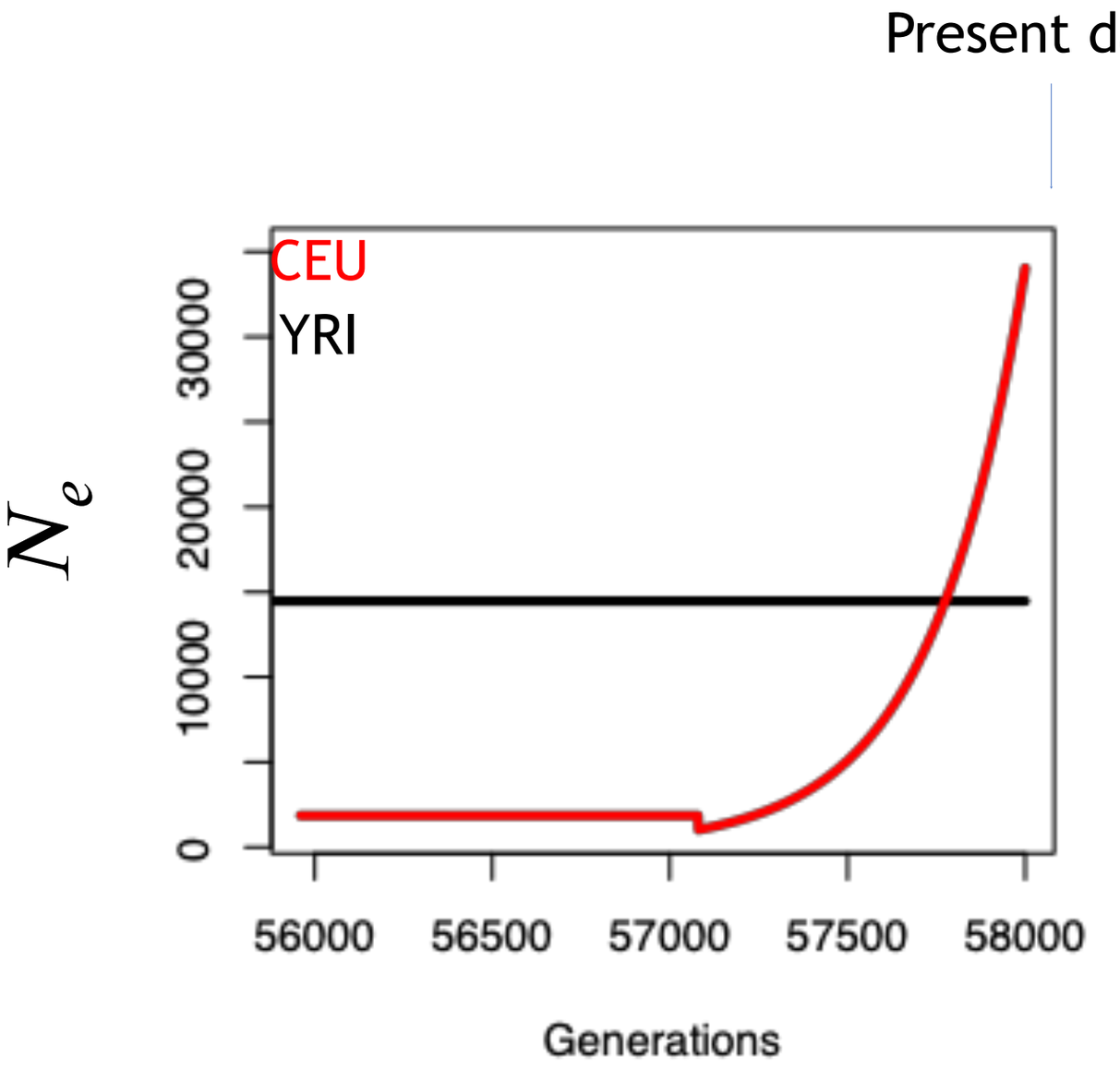
Lohmueller (2014) *PLoS Genet*

Forward simulations under more realistic demography

Arun Durvasula

(Genetics &
Genomics
Graduate student)

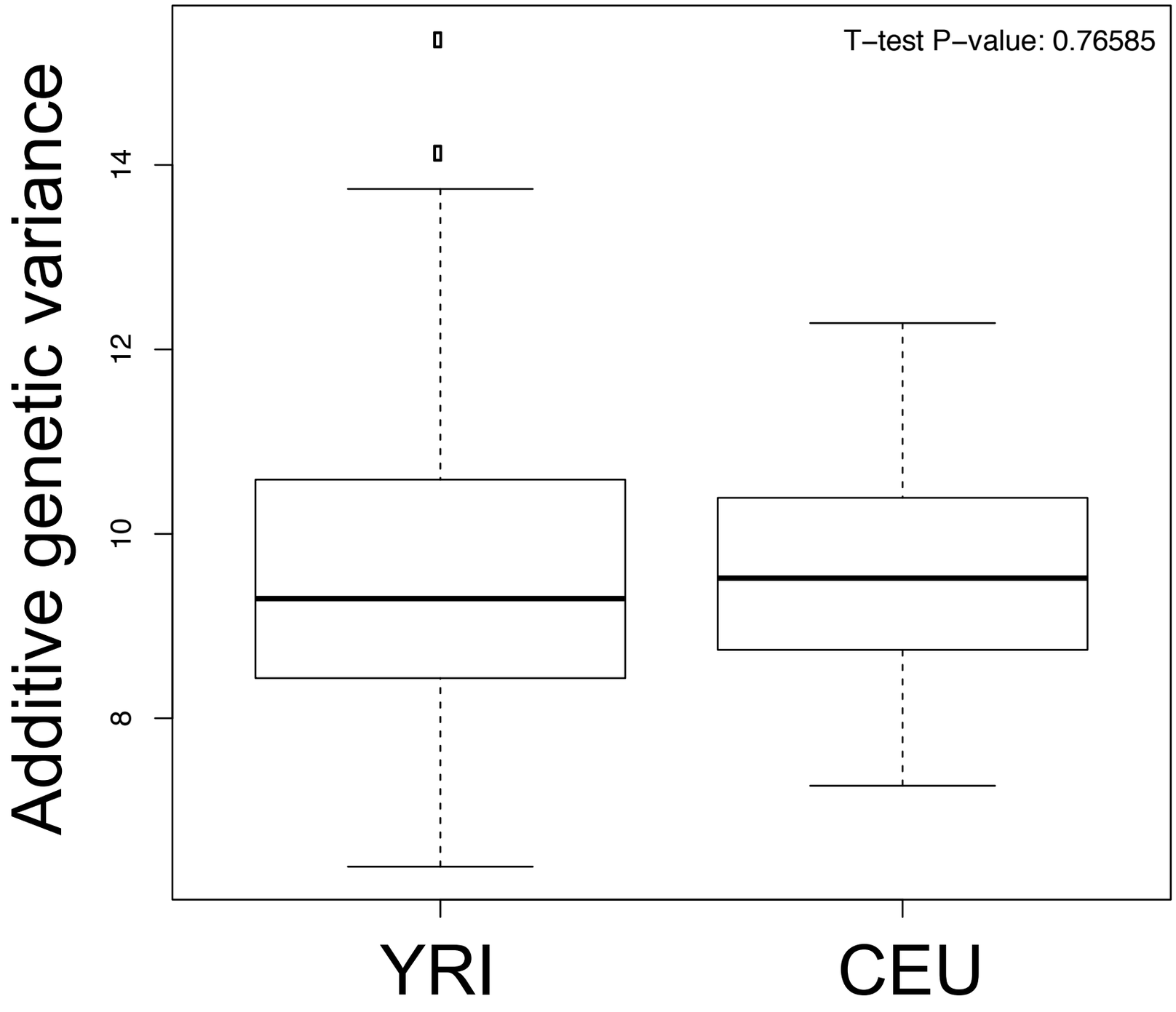
Forward simulations under more realistic demography



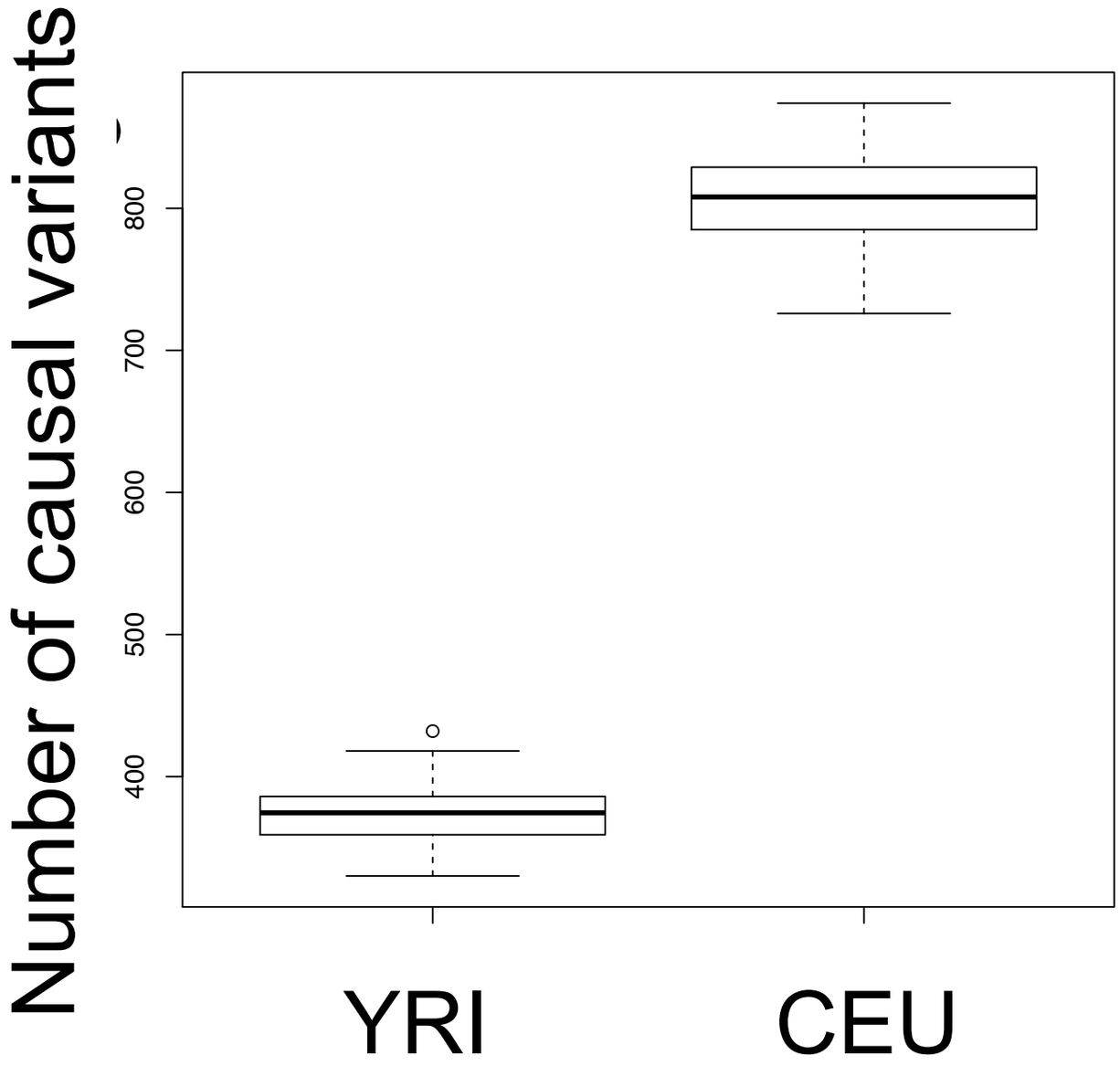
- Forward simulation of a trait under stabilizing selection following an out of Africa human demography
- Simulations done using SLiM

Haller and Messer 2016
Gravel et al 2011

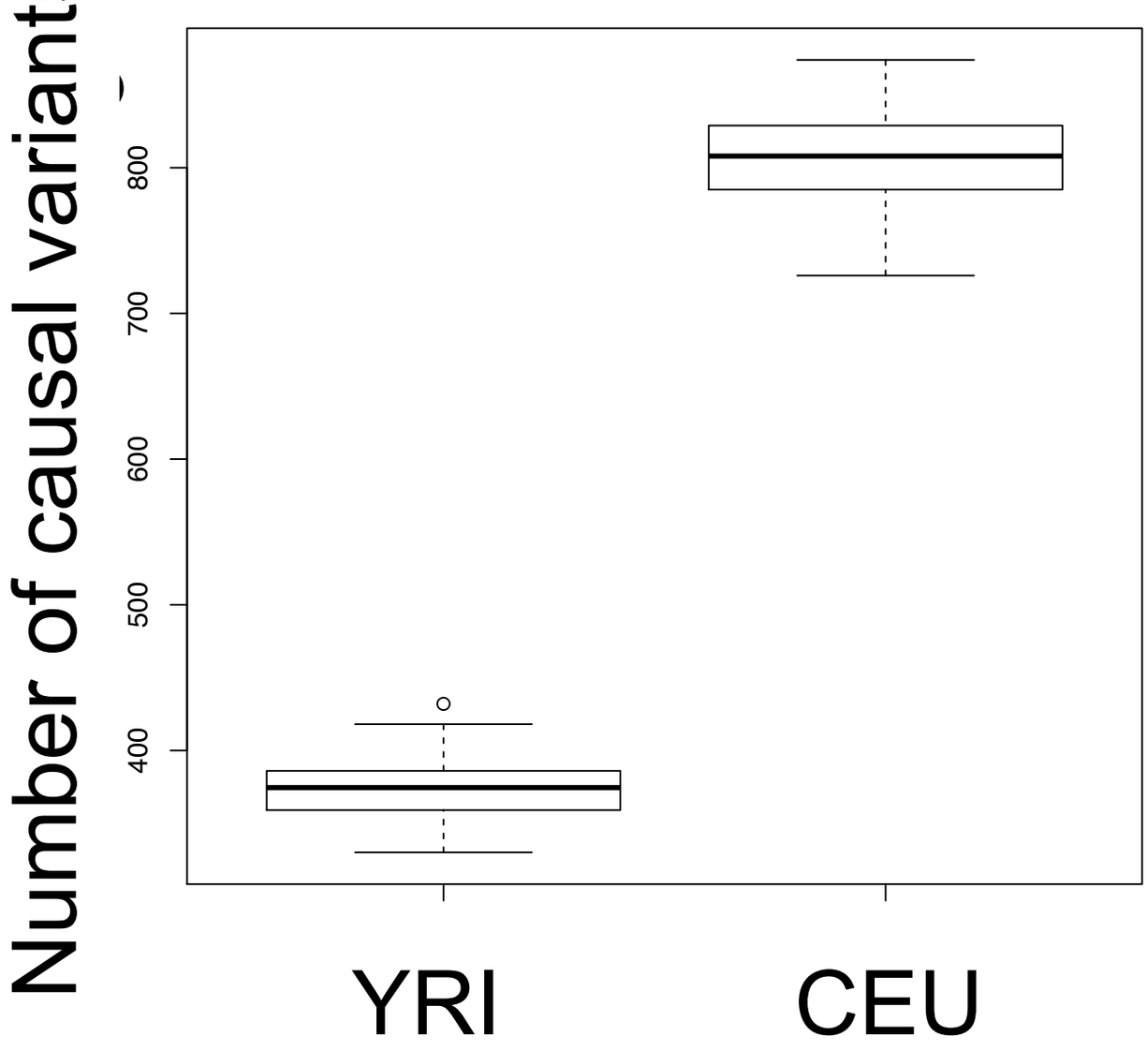
Simulations imply similar heritability across populations

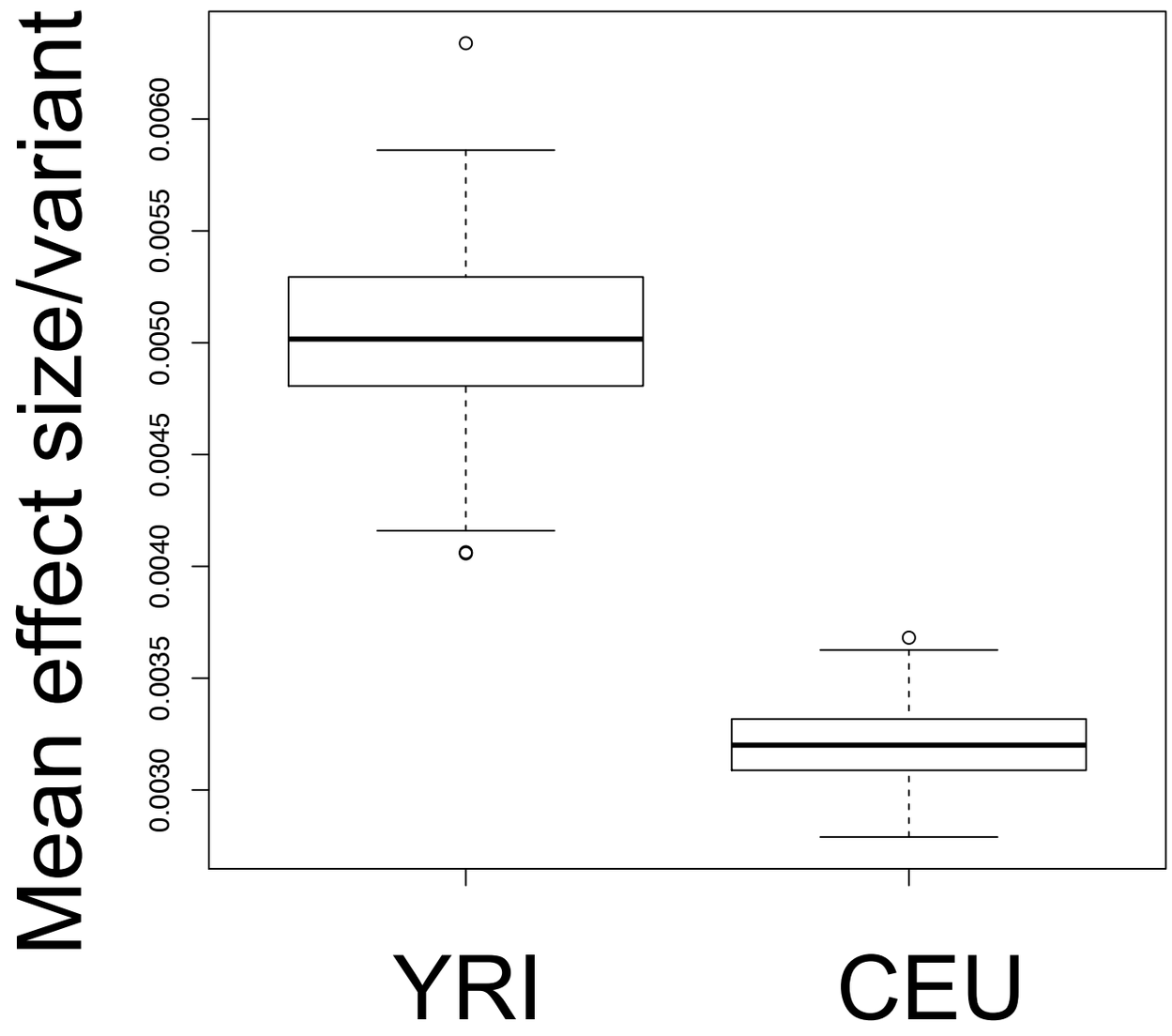


However, number of causal variants and effect sizes are predicted to differ



However, number of causal variants and effect sizes are predicted to differ

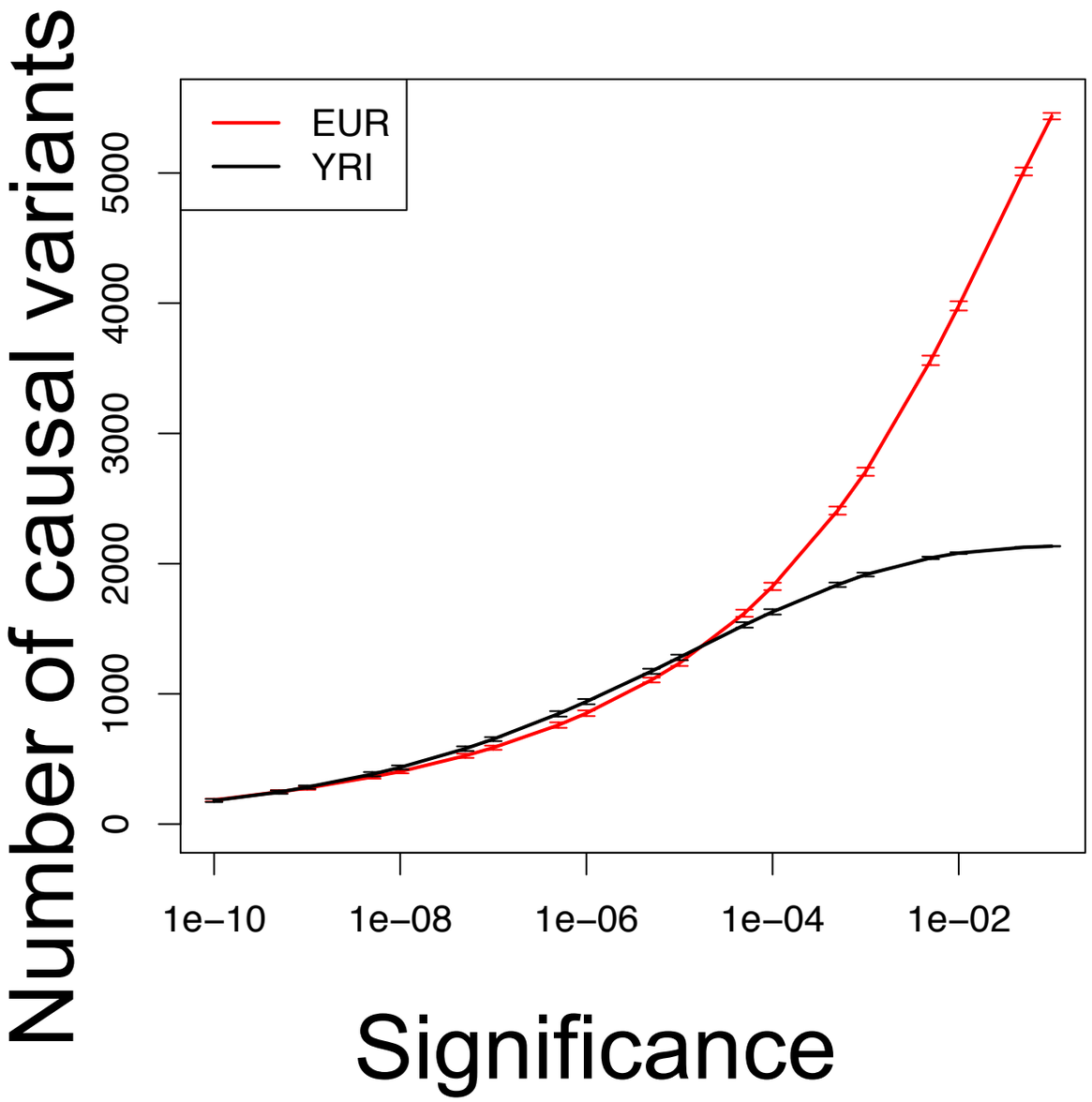


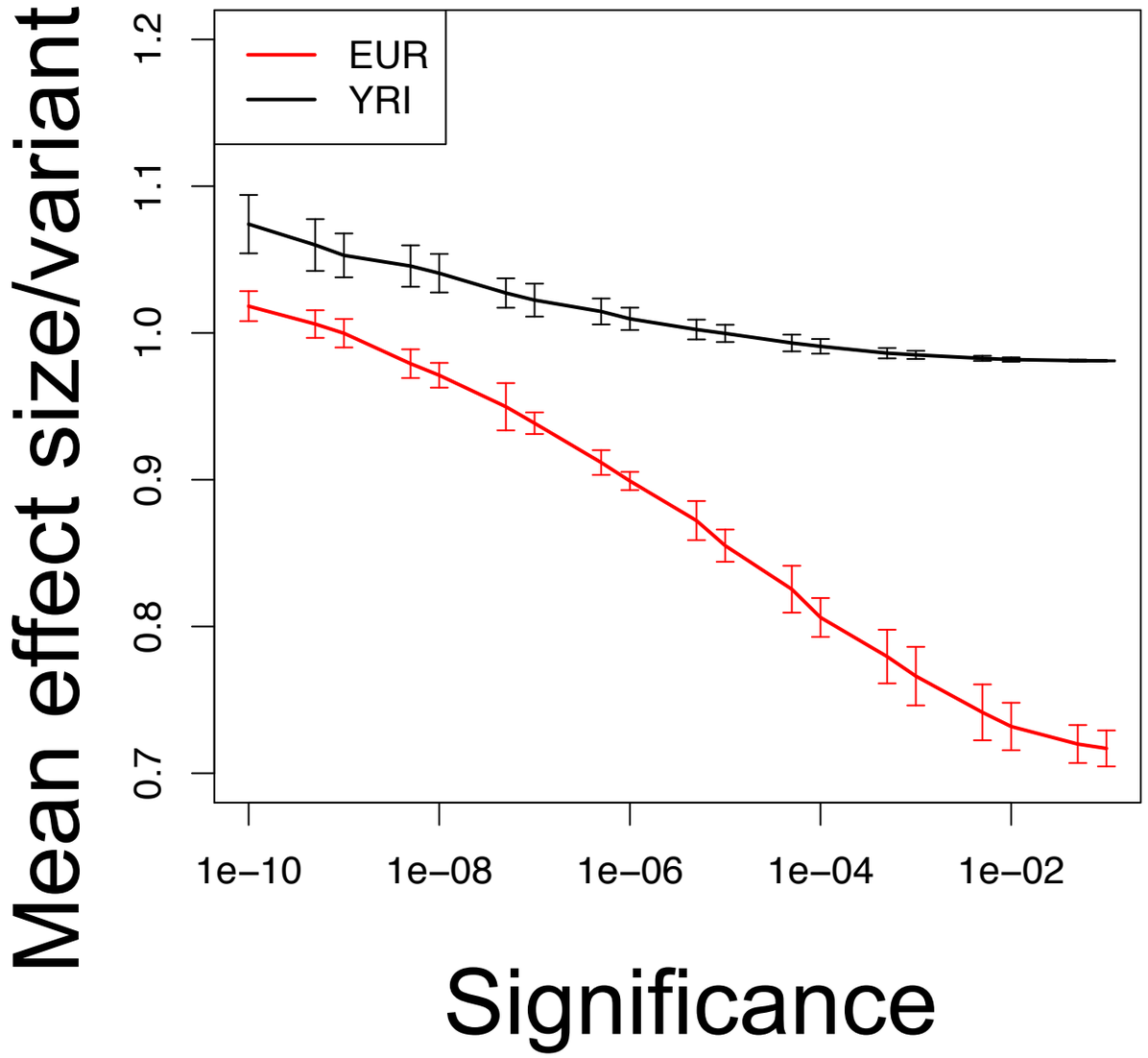


Testing models of genetic architecture using gene expression

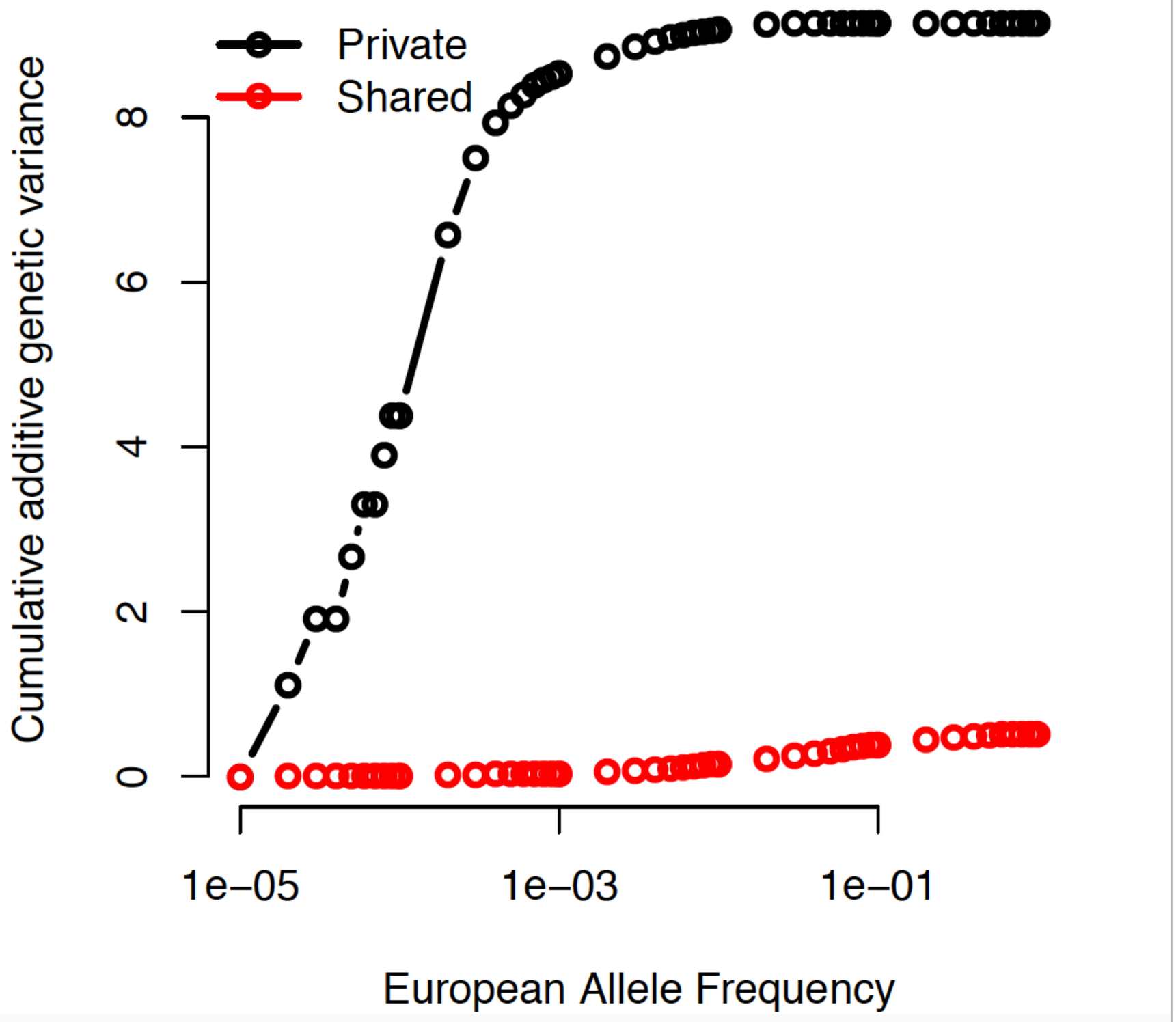
- Examine eQTLs in GEUVADIS data
- Overall, Lappalainen et al. (2013) find more significant associations in EUR than YRI. However, differences in power…
- Computed power to detect each variant (given its effect size, frequency & sample size)
- Simulated eQTL studies of same sample size to account for differential power

More causal variants of weaker effect in EUR compared to YRI





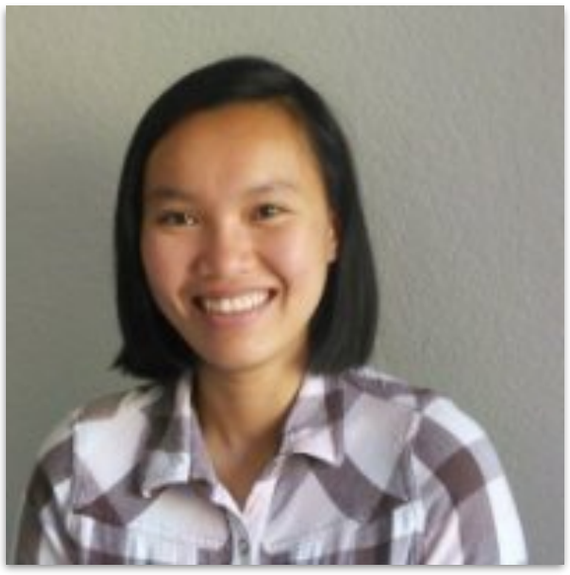
Private variants account for the majority of additive genetics variance



Conclusions

- The mutational target size differs between traits but is large (on orders of tens of megabases)
- Purifying selection is pervasive on complex traits, even those not thought to be directly tied to fitness
- Demography impacts the architecture of traits
- This provides an important additional source of ambiguity when attempting to transfer polygenic risk scores across populations

Acknowledgements



Tanya Phung

Arun Durvasula

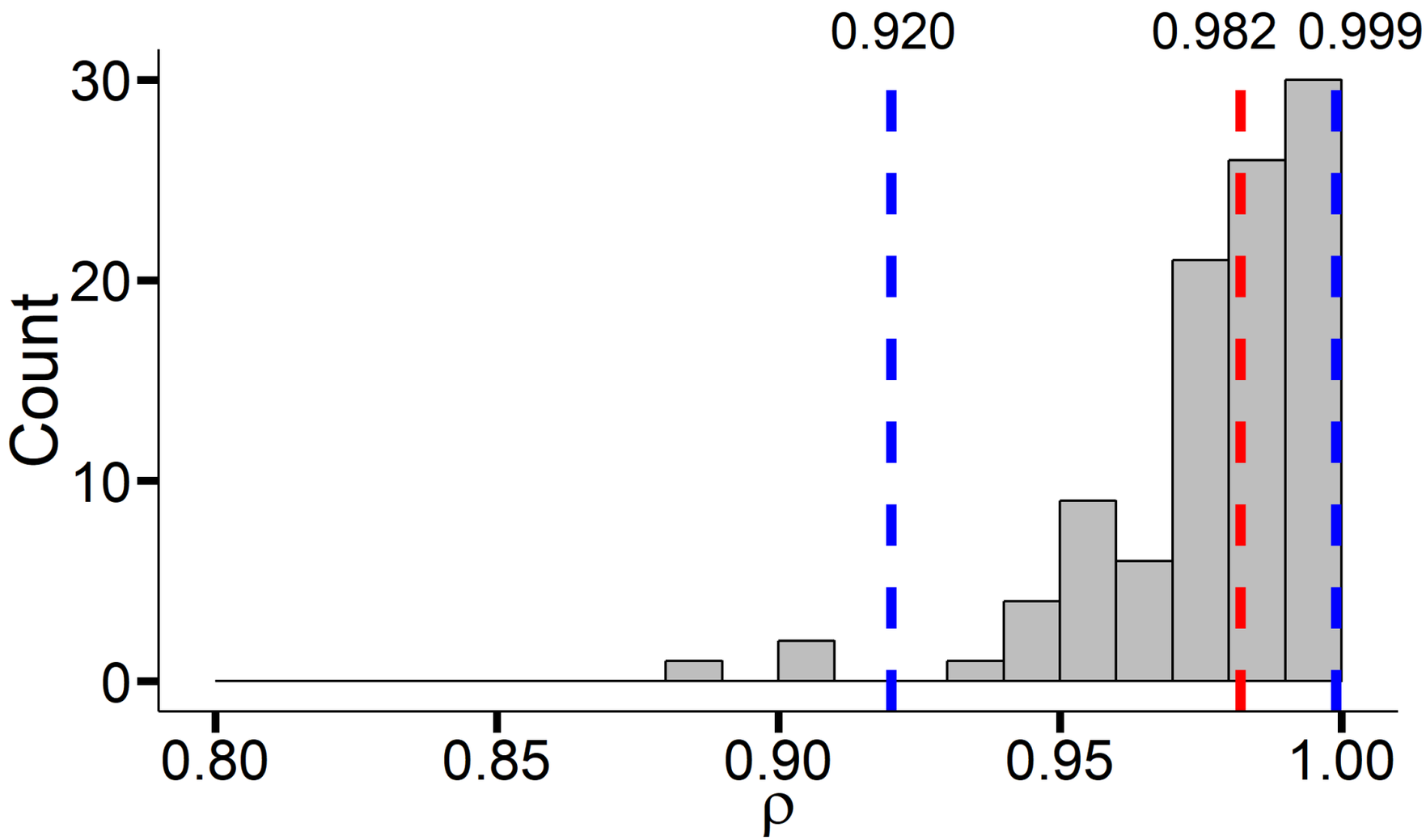
- Bogdan Pasaniuc, Nick Mancuso, Gleb Kichaev, Christian Huber
- Funding sources:
  - UCLA Biomedical and Big Data training grant to TNP
  - National Institutes of Health R01 HG009120-01A1 to Bogdan Pasaniuc
  - National Institutes of Health R35GM119856 to KEL

Incorporate pleiotropy

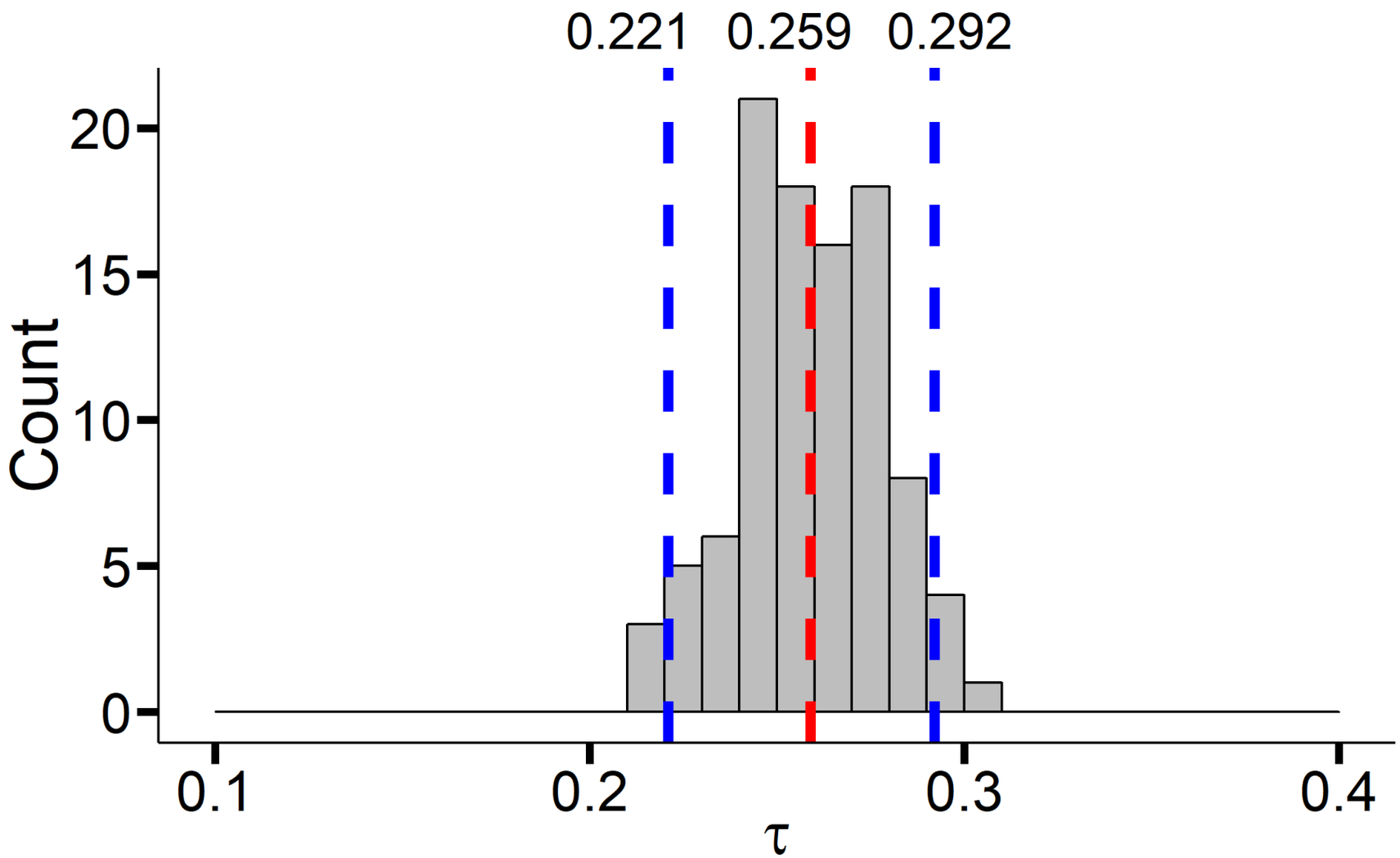
- Pleiotropy is captured by ρ (Uricchio et al. 2016)

- Modify InGeAr to also infer ρ

ρ is close to 1



τ does not change significantly



M is smaller when incorporating pleiotropy