Genetic diversity in humans: evolutionary and medical perspectives

Research Talk

Bárbara Domingues Bitarello

Outline

1. Introduction
2. Research Theme 1: Balancing selection in humans
3. Research Theme 2: Polygenic risk prediction for individuals with non-European ancestry
4. Conclusions & Future Directions

`

My research

Hominin genomics

Comparative genomics

Population genomics

Genomic functional categories

Great Apes

Human expansion

Processes that shape genomic diversity

Outline

1. Introduction
2. Research Theme 1: Balancing selection in humans
3. Research Theme 2: Polygenic risk prediction for individuals with non-European ancestry
4. Conclusions & Future Directions

`

Balancing selection: an umbrella term

MHC/HLA: an extreme instance of balancing selection

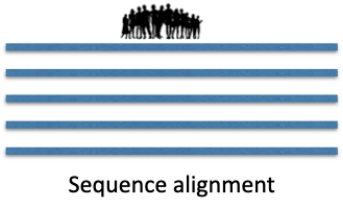
MHC/HLA: an extreme instance of balancing selection

1000 Genomes Project

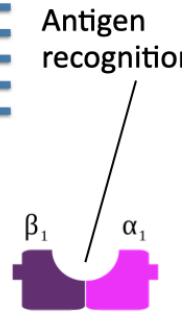
1000s of HLA-A, -B and -C alleles

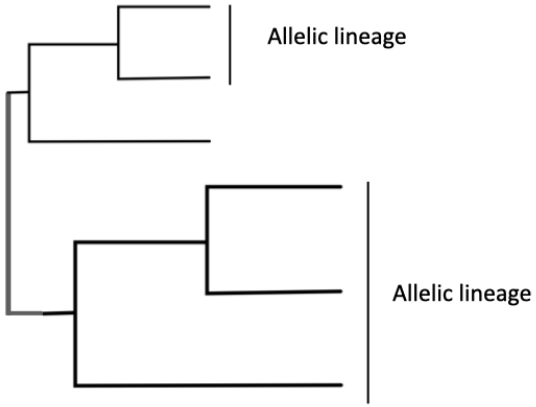
HYPOTHESIS

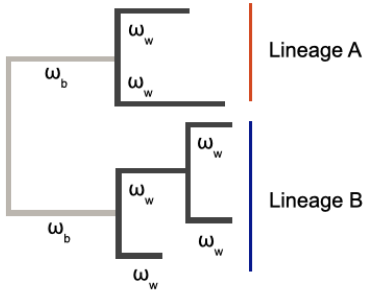
HLA diversity is shaped by divergent allele advantage



Sequence alignment



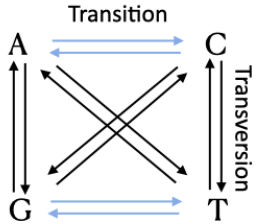




$$\omega = dN/dS$$

dN: nonsynonymous substitution rate

dS: synonymous substitution rate



Codon evolution models

PREDICTION

$$\omega_{between} > \omega_{within}$$

Bitarello et al. (2016), *Journal of Molecular Evolution*

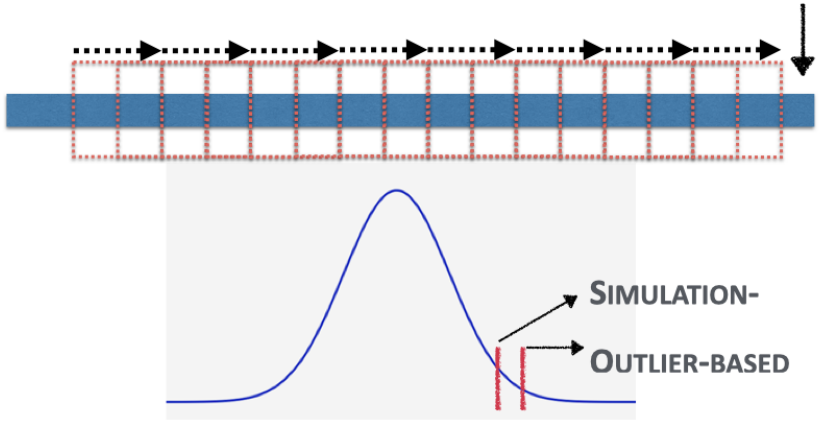
Heterogeneity of dN/dS ratios at the classical HLA class I genes over divergence time and across the alle

How prevalent has balancing selection been throughout human evolution?

Genomic "signatures" of balancing selection

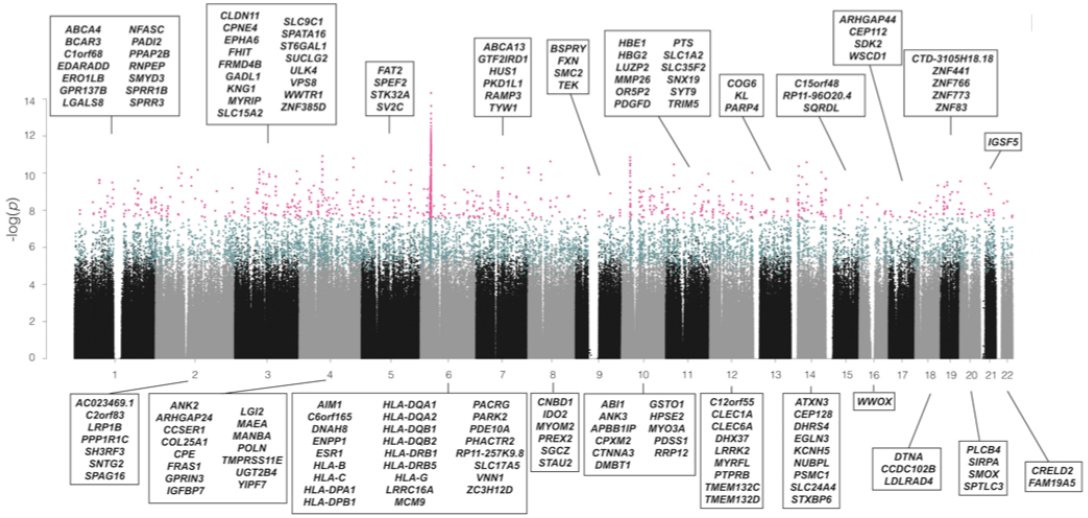
A novel method to detect signatures of balancing selection

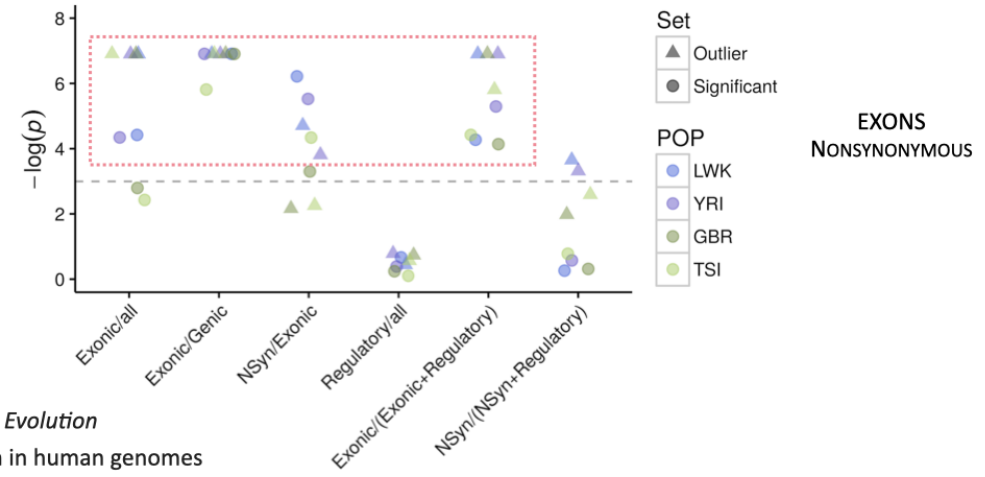
Statistic of Interest



1000 Genomes data

2/3 : NOT PREVIOUSLY KNOWN CANDIDATE GENES





Best Graduate Student paper award, Society for Molecular Biology and Evolution

Bitarello et al. (2018), *Genome Biology and Evolution*
Signatures of long-term balancing selection in human genomes

Outline

1. Introduction
2. Research Theme 1: Balancing selection in humans
3. Research Theme 2: Polygenic risk prediction for individuals with non-European ancestry
4. Conclusions & Future Directions

`

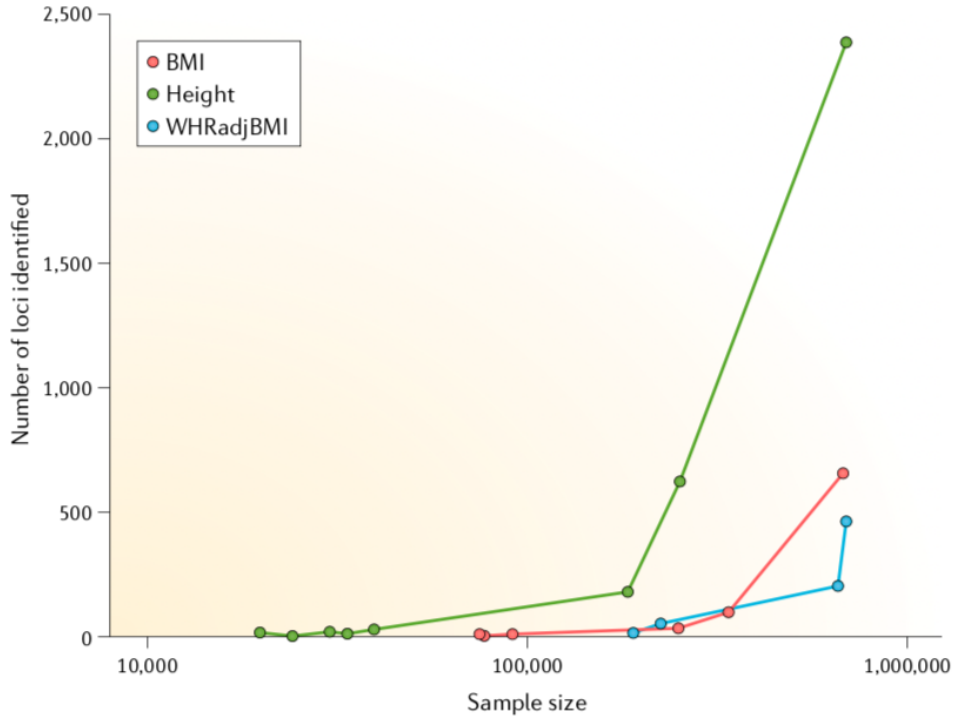
The genetics of human traits and diseases

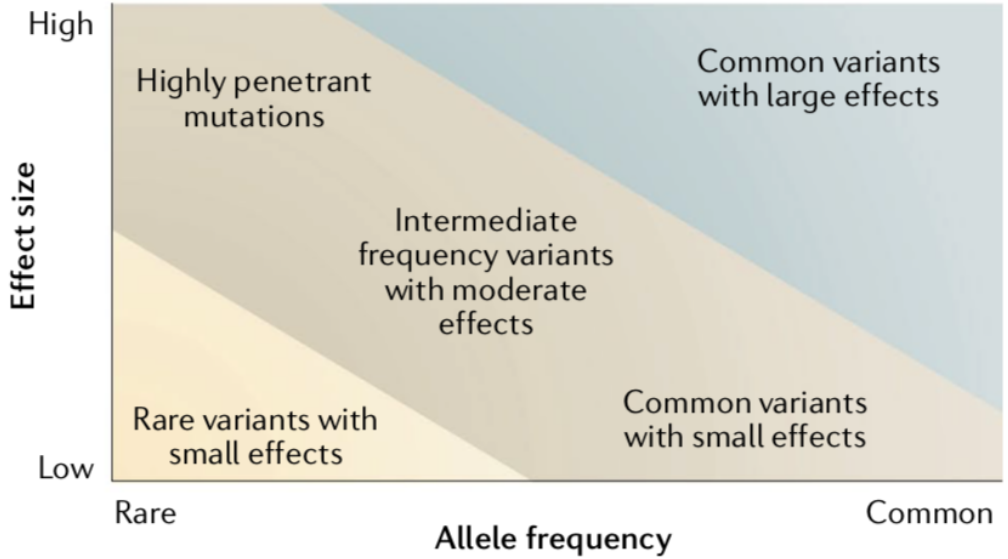
- rare, monogenic diseases/traits
- complex, common diseases/traits

Genome-wide association studies (GWAS)

`

Many variants with small effect size





[Tam et al. (2019) *Nat Rev Genet*]

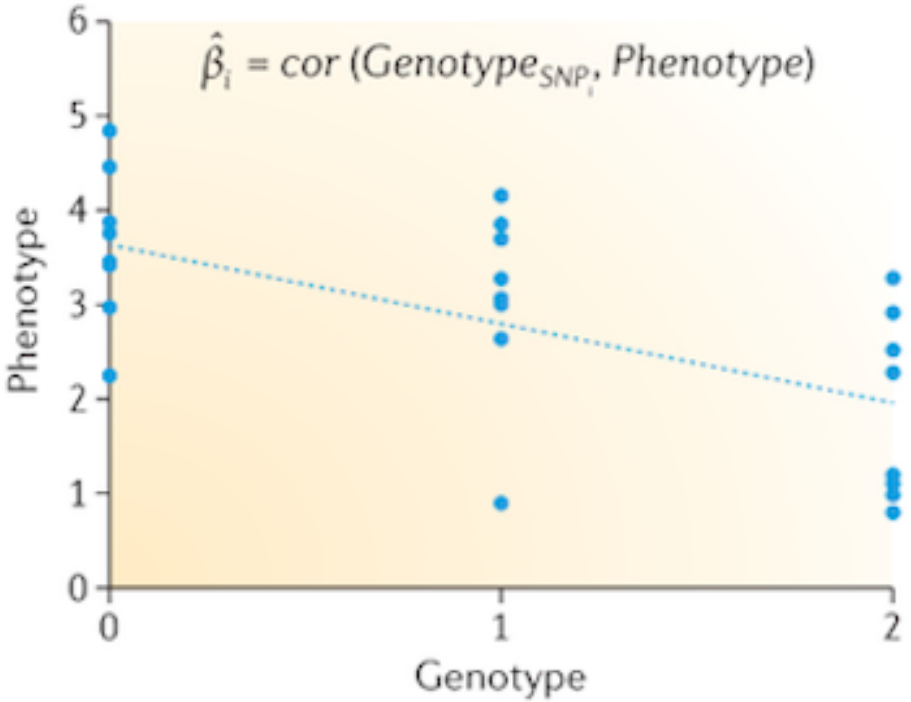
Combined, they explain a lot!

Some examples

Phenotype	Statistic	Value	Variants
height	R-squared	25.0	3000
schizophrenia	R-squared	7.0	100
ADHD	R-squared	5.5	100
breast cancer	AUC	60.0	1000
cardiovascular disease (CAD)	AUC	81.0	6000

PS: for Europeans ancestry only...

Polygenic risk scores add up those small effects



Pasaniuc & Price (2017), *Nat Rev Genet*

$$PRS = \sum_{i=1}^{m} \hat{\beta}_i G_{j,i}$$

$\hat{\beta}$: effect size (from GWAS)

G : Effect allele dosage

j: Individuals

i: SNPs

independence

additive model

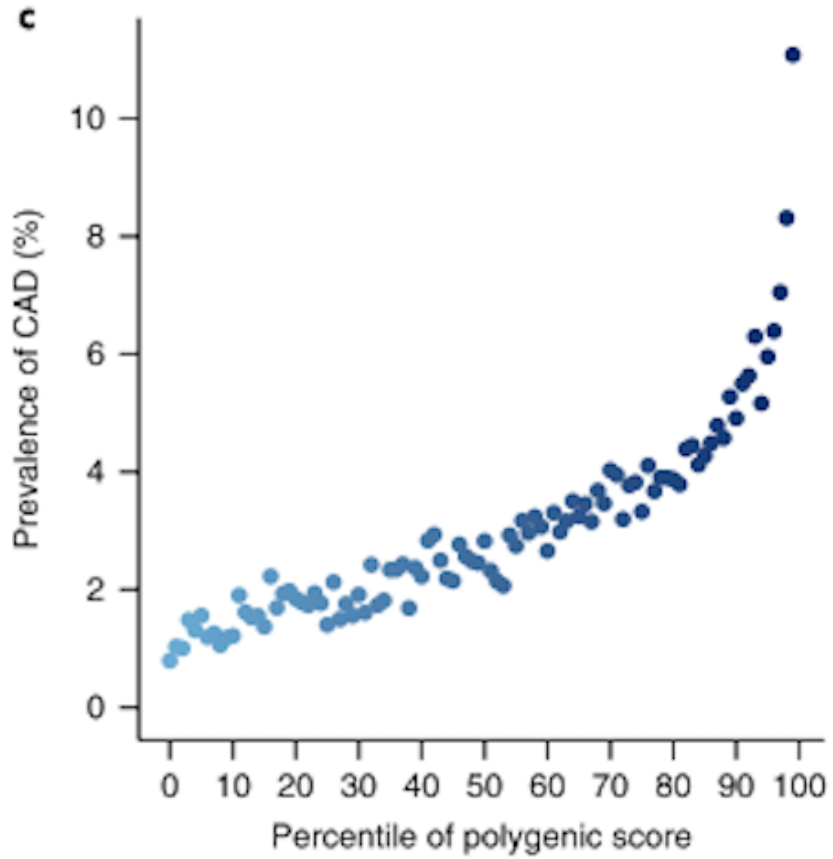
PRSs are appealing

easy

promising

fast

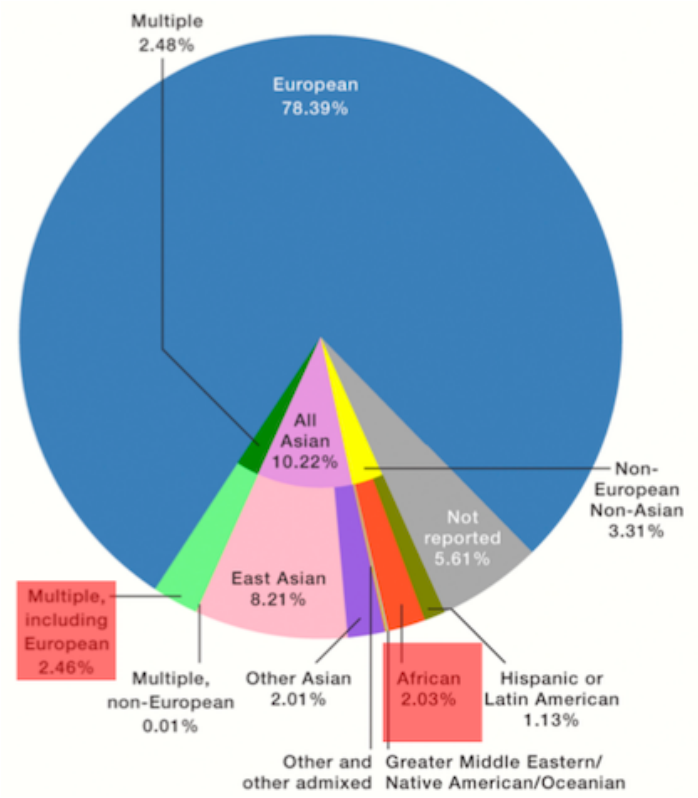
minimal requirements



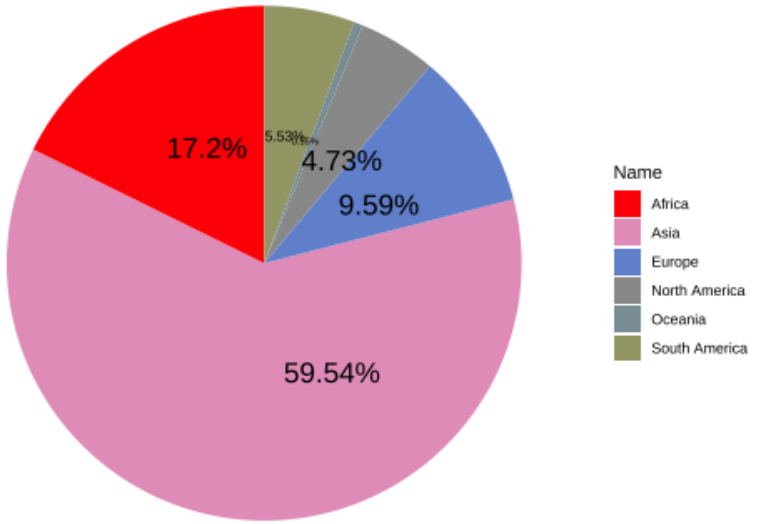
[Khera et al (2018) Nat Genet]

What about ancestry?

European ancestry represent almost 80% of GWAS participants...



[Sirugo, Williams & Tishkoff (2019) *Cell*]

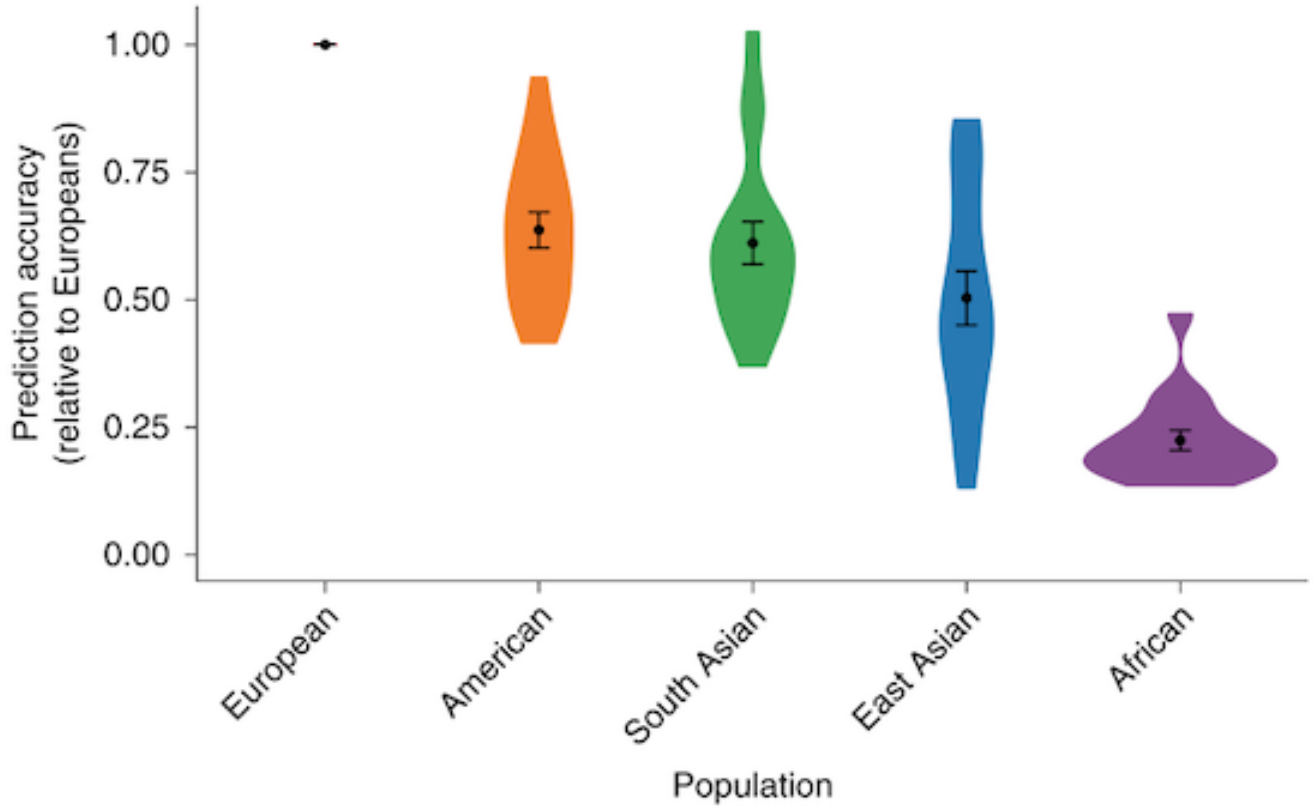


[Data: https://worldpopulationreview.com/]

.. and <15% of the world's population

How does PRS accuracy transfer across ancestries?

PRS accuracy decreases with genetic distance from Europeans



[Martin et al. (2019) *Nat Genet*]

Why is this loss of prediction observed?

What **can** ***we do*** about it?

Many factors may influence LOA

causal variants

local selection

gene-gene interactions

gene-environment interactions

Effect sizes

Linkage disequilibrium

allele frequencies

phenotypic variance

These factors are not mutually exclusive!

Questions

How do these different factors affect prediction accuracy?

Can we leverage ancestry into the PRS and improve predictions?

Let's look at height

Ancestry as a continuous variable

highly polygenic

many cohorts phenotyped

large GWAS (UKBB

$\sim$ 360,000 Europeans)

$$height \sim Sex + Age + Age^2 + p_{eur}$$

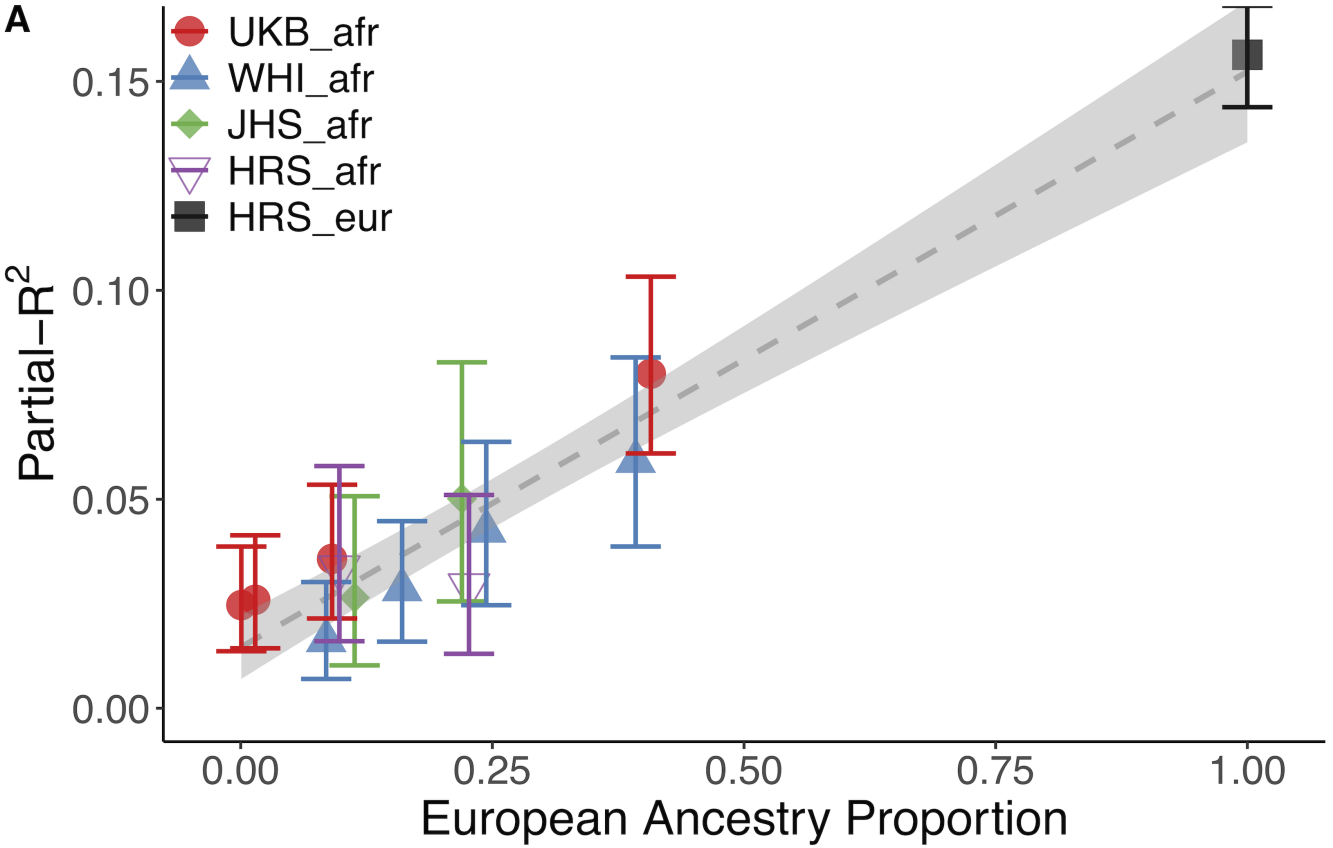
$$height \sim Sex + Age + Age^2 + p_{eur} + PRS$$

European + African ancestry

Data	Ancestry	N	Number_SNPs
UKBB_eur	European	9998	685475
HRS_EUR	European	10159	1511742
UKBB_afr	African + European	8700	685475
WHI_afr	African American	6863	741983
JHS_afr	African American	1773	702685
HRS_afr	African American	2251	1511742

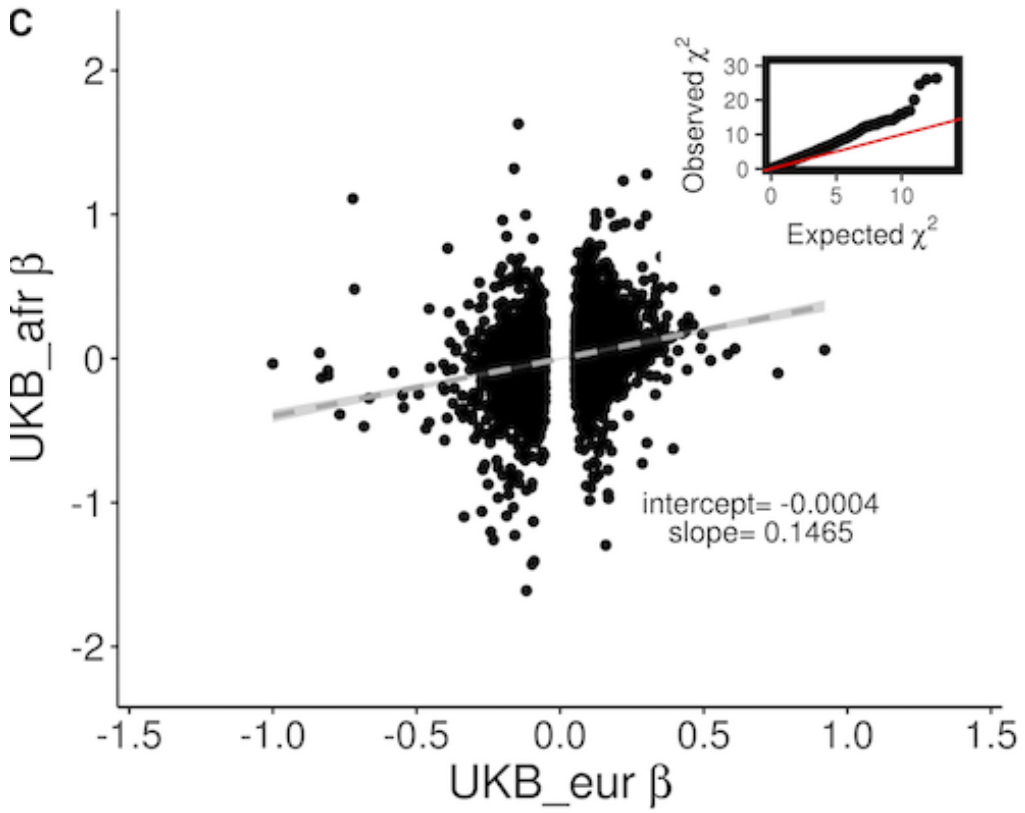
[Bitarello & Mathieson (2020), *G3*]

PRS accuracy increases with proportion of European ancestry



[Bitarello & Mathieson (2020), *G3*]

GWAS from UKBB (AFR) individuals



$$y = sex + age + age^2 + 10PCs$$

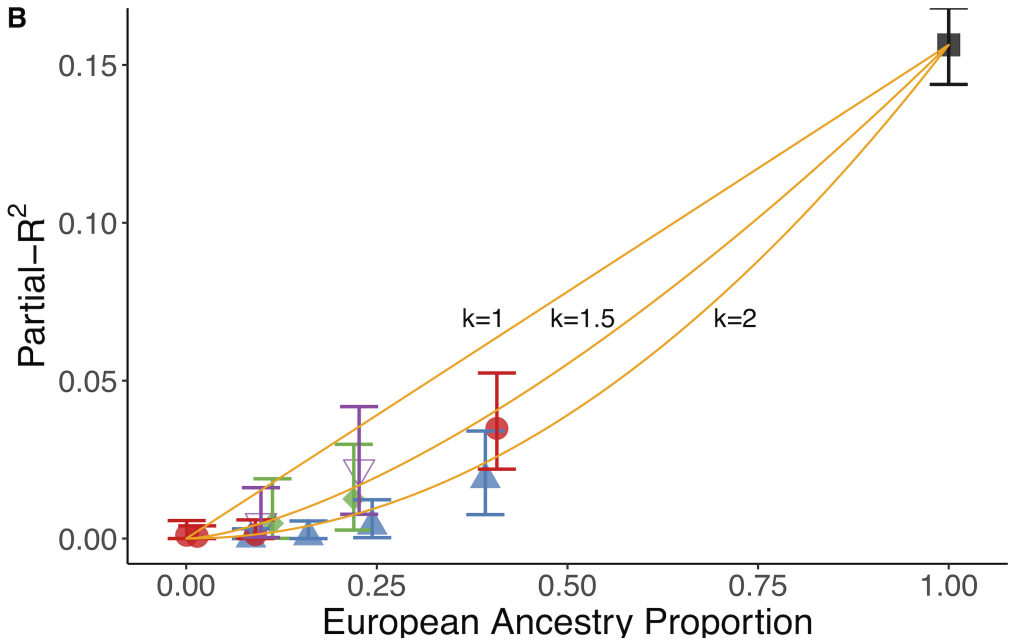
$$N_{AFR} \sim 8,800$$

$$N_{EUR} \sim 350,000$$

$$\chi^2_{diff} = \left[\frac{\beta_{eur} - \beta_{afr}}{\sqrt{SE^2_{eur} + SE^2_{afr}}} \right]^2$$

[Bitarello & Mathieson (2020), *G3*]

Using only EUR chunks of each genome



[Bitarello & Mathieson (2020), *G3*]

$$y = 0.15p_{eur}^{k}$$

$k = 1$

all predictive power comes from European chunks

$k = 2$

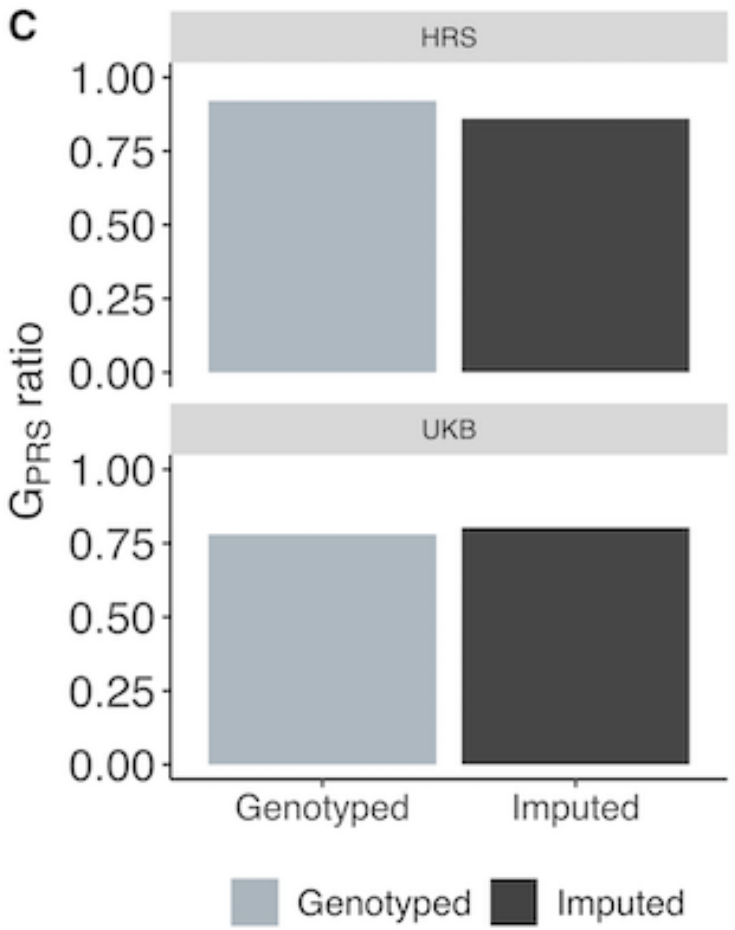
predictive power is uniformly distributed

Surprisingly, this relationship seems to be more like k=2

Others found that the LOA can be fully recovered by including EUR chunks [Marnetto et al. (2020) *Nat Comms*]

`

Allele frequency differences explain up to 20% of LOA



Additive genetic variance

$$G_{PRS} = \frac{\sum 2 f_{i_a fr}(1 - f_{i,afr})\beta^2_{i,eur}}{\sum 2 f_{i,eur}(1 - f_{i,eur})\beta^2_{i,eur}}$$

[Bitarello & Mathieson (2020), *G3*]

Prediction: variance in phenotype lower in EUR

genome-wide genetic variance in EUR is $\sim 76\%$ of that in AFR

$$height \sim Sex + Age + Age^2 + p_{eur}$$

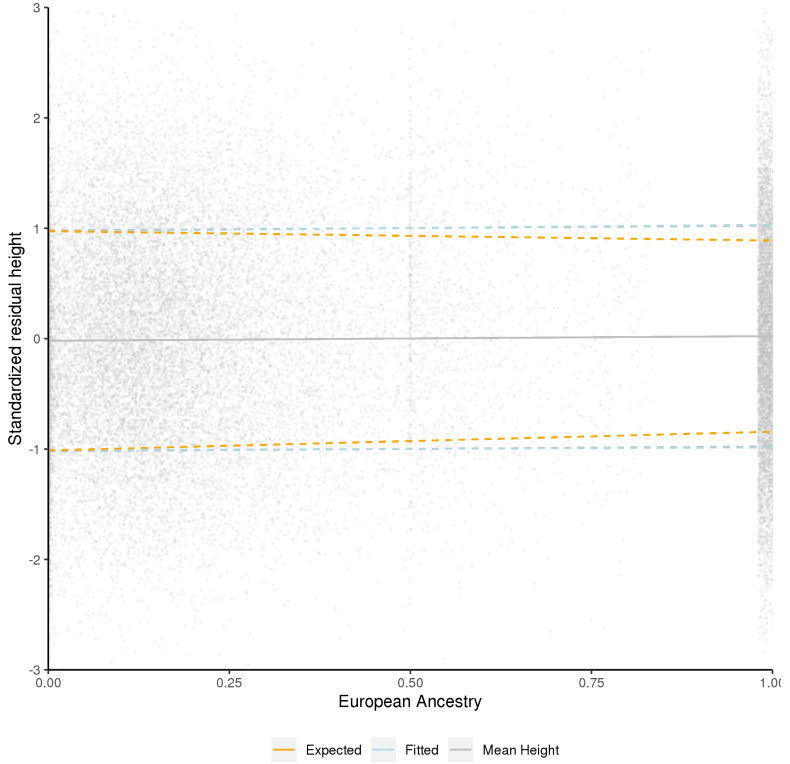
Variable variance model:

$$y = \mu + \beta p_{j,eur} + \epsilon; \epsilon_j \sim N(0, \delta^2 + \gamma p_{j,eur})$$

Mean+1 1sd, constant variance

fitted, variable variance

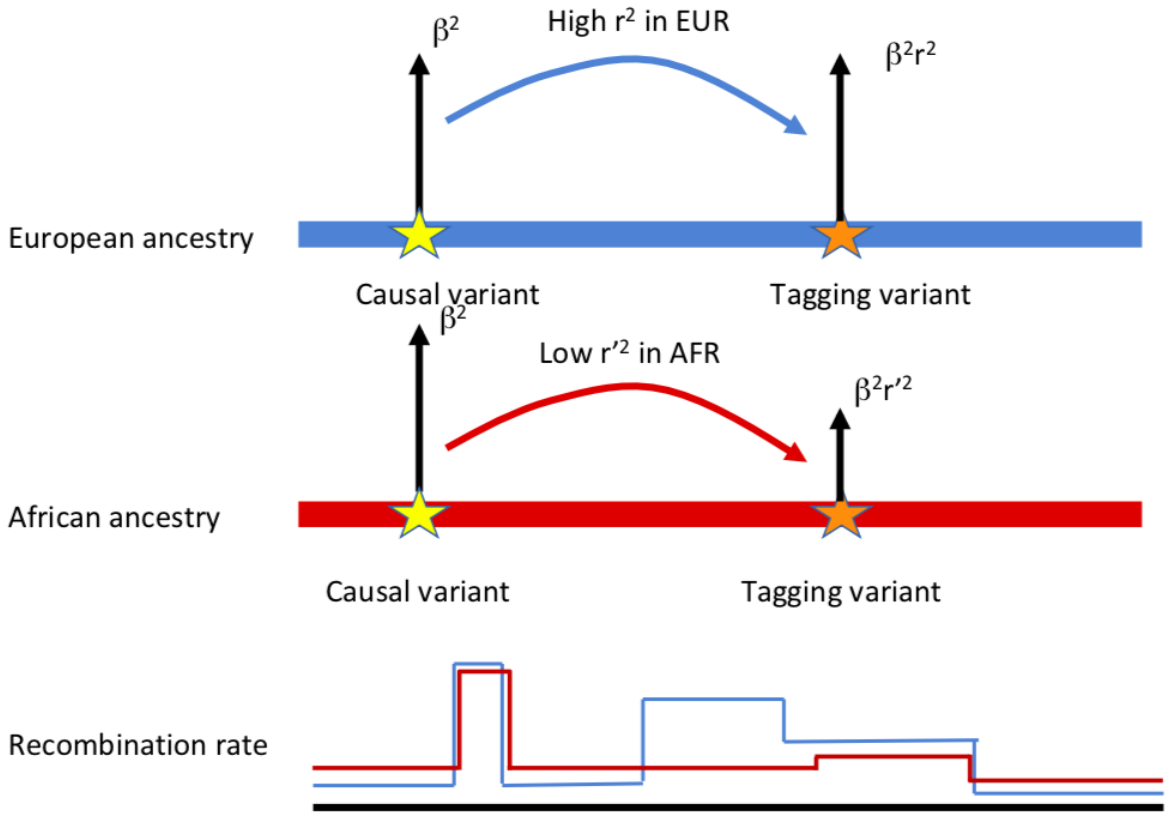
model, variance in phenotypic variance is 100% in AFR and 76% in EUR



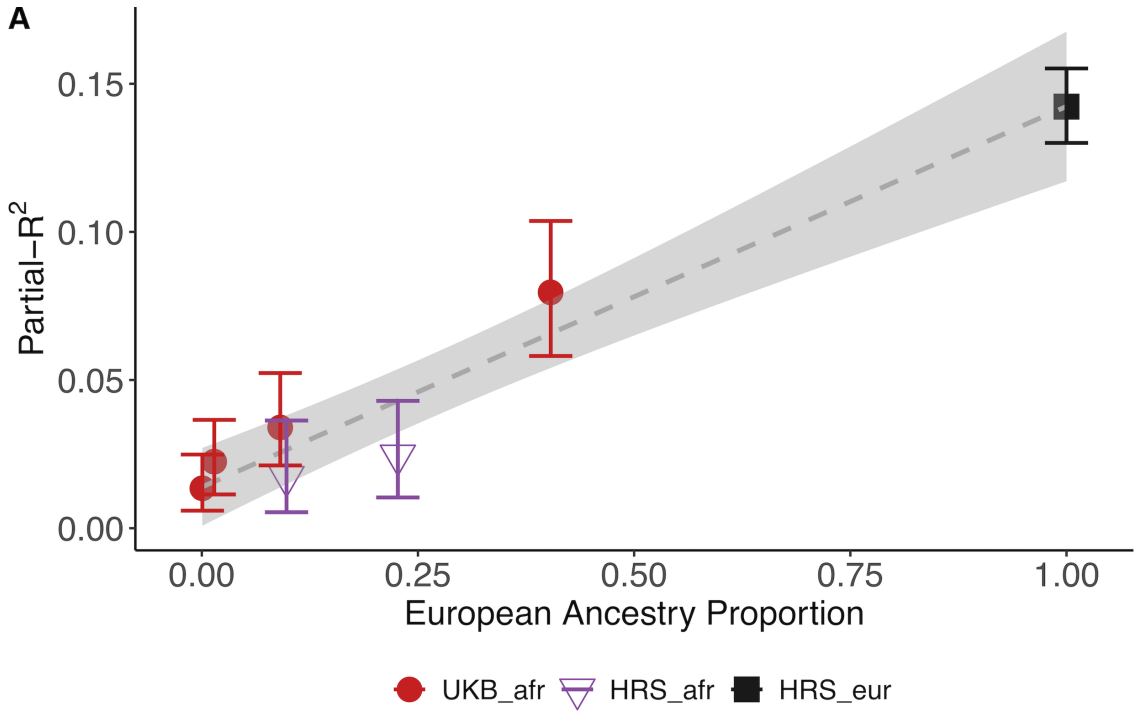
[Bitarello & Mathieson (2020), *G3*]

Observation: Phenotypic variance does not change with ancestry

Differences in linkage disequilibrium



Prediction: better tagging of causal variants decreases LOA

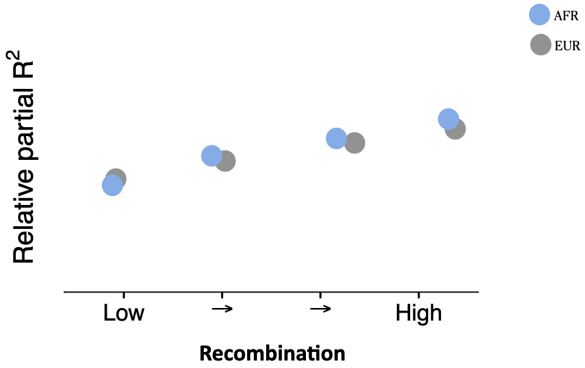


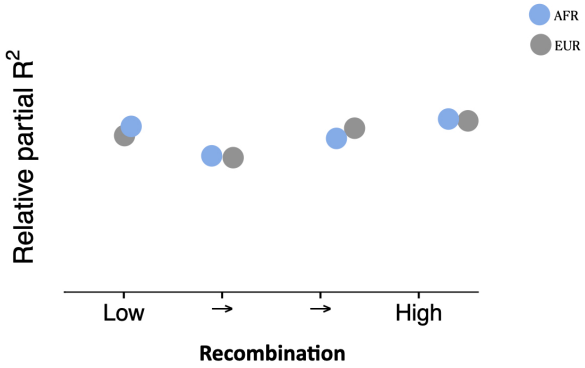
[Bitarello & Mathieson (2020), *G3*]

Observation: Imputation doesn't influence LOA

Prediction 1: LOA is independent of LD differences

$$Rel_{R2} = \frac{R^2_{bin}}{R^2_{total}}$$



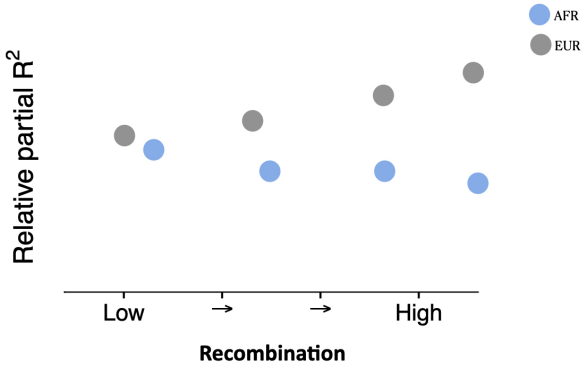


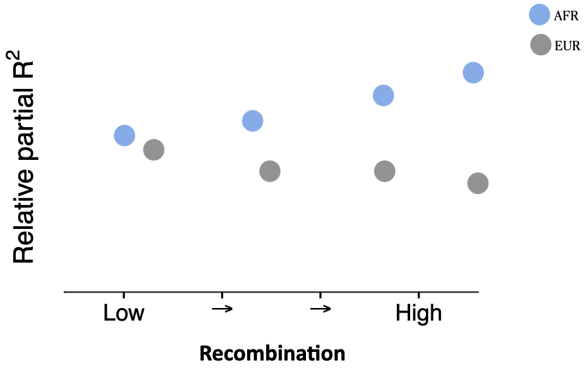
Similar slopes

`

Prediction 2: LOA is dependent of LD differences

$$Rel_{R2} = \frac{R^2_{bin}}{R^2_{total}}$$

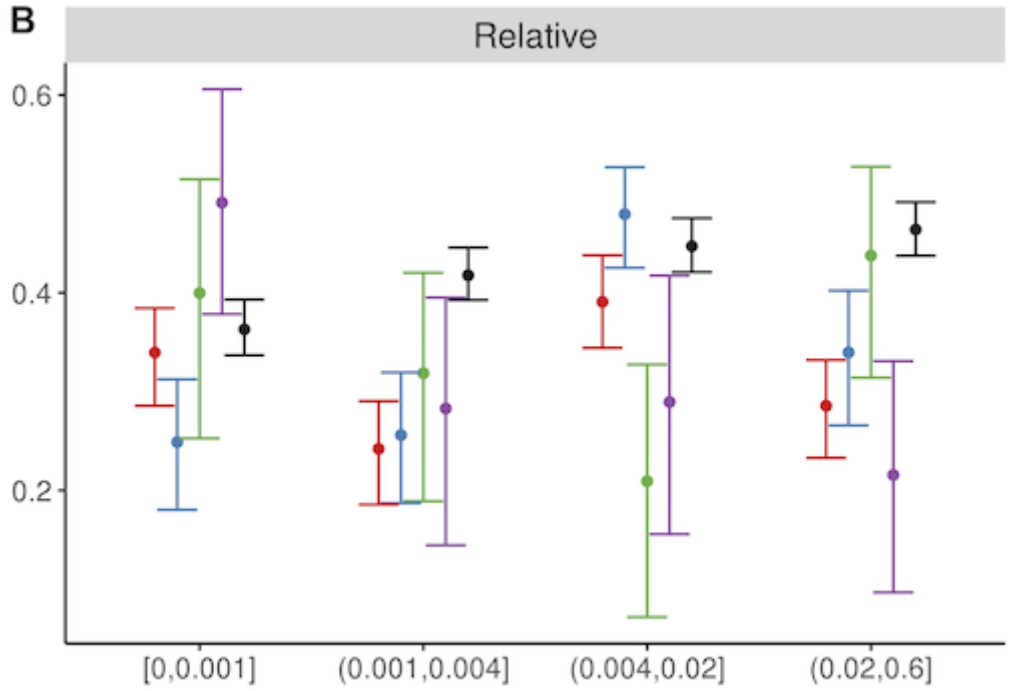




Different slopes

`

$$Rel_{R2} = \frac{R^2_{bin}}{R^2_{total}}$$

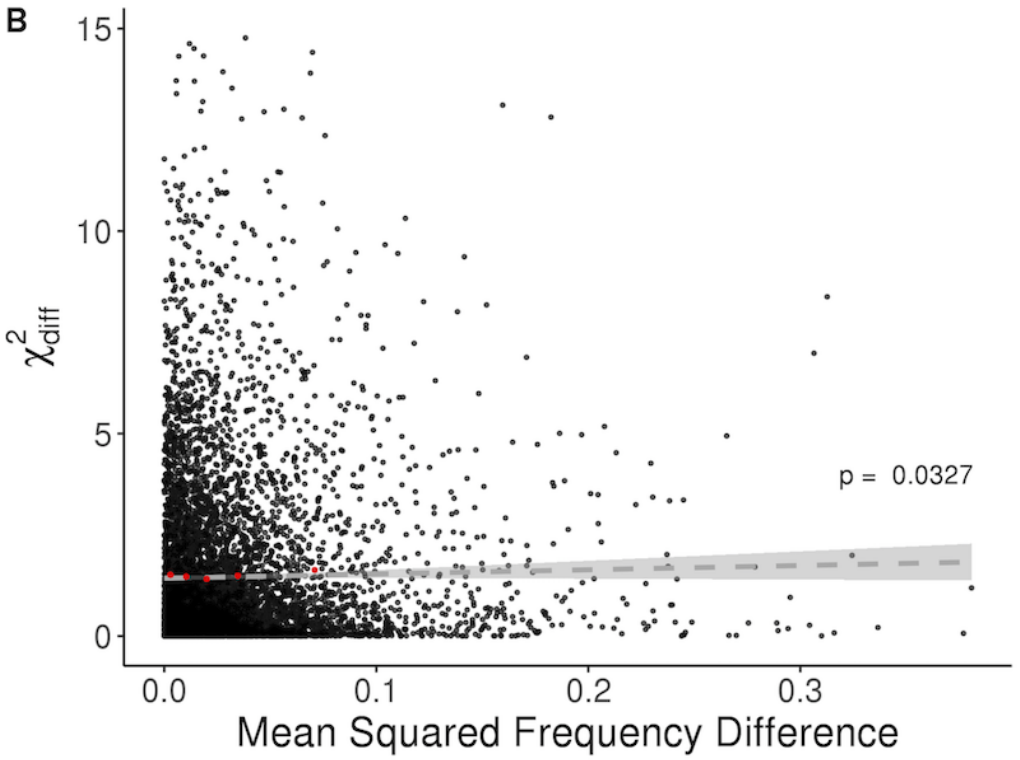


Recombination Rate

[Bitarello & Mathieson (2020), *G3*]

Observation: LOA is somewhat dependent on recombination rate

Prediction: Differences in effect sizes depend on allele frequency differences



[Bitarello & Mathieson (2020), *G3*]

$$N_{AFR} \sim 8,800$$

$$N_{EUR} \sim 350,000$$

$$\chi^2_{diff} = \left[\frac{\beta_{eur} - \beta_{afr}}{\sqrt{SE^2_{eur} + SE^2_{afr}}} \right]^2$$

Observation: Difference in effect sizes increase with allele frequency differences across ancestries

Assuming there are differences in marginal effect sizes

Assuming there are differences in marginal effect sizes

$$PRS_1^C = \alpha PRS_{AFR} + (1-\alpha)PRS_{EUR}$$

Marquez-Luna *et al.* (2018) *Genet Epidem*

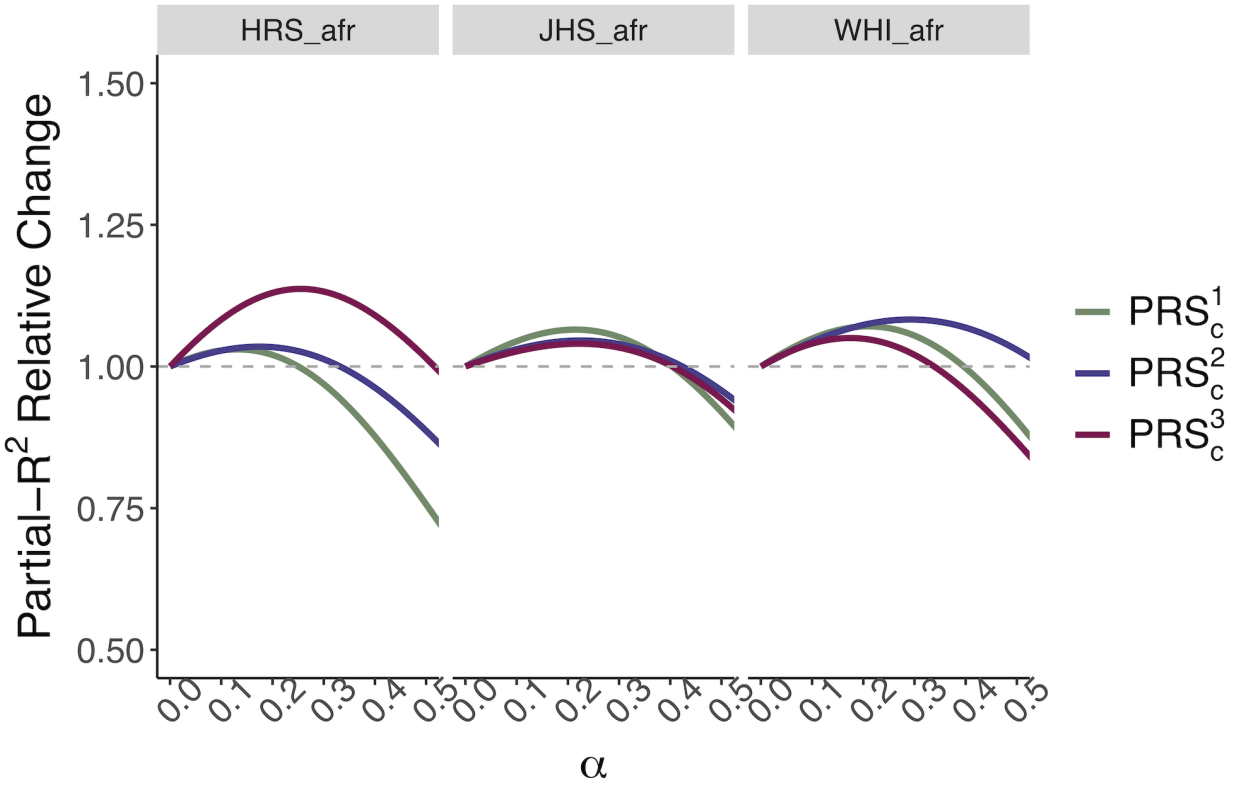
$$PRS_2^C = \alpha(1-p_{eur,j})PRS_{afr,j} + (1-\alpha+\alpha p_{eur,j})PRS_{EUR}$$

Bitarello & Mathieson (2020), *G3*

$$PRS_3^C = \alpha\left[\sum_{i\in AFR}\beta_{i,afr}G_i\right] + (1-\alpha)\left[\sum_{i\in AFR}\beta_{i,eur}G_i\right] + \left[\sum_{i\in EUR}\beta_{i,eur}G_i\right]$$

Bitarello & Mathieson (2020), *G3*

Prediction: Including ancestry-specific effect sizes improves accuracy for admixed individuals



[Bitarello & Mathieson (2020), *G3*]

Outline

1. Introduction

2. Research Theme 1: Balancing selection in humans

3. Research Theme 2: Polygenic risk prediction for individuals with non-European ancestry

4. Conclusions & Future Directions

`

Francisco,(+2), **Bitarello**, França, (+2) (2015), *Immunogenet*
Bitarello, Francisco & Meyer (2016), *J.Mol.Evol*
Brandt, Aguiar, **Bitarello** (+3) (2015), *G3*

Bitarello & Mathieson (2020) G3

How was it shaped by adaptive evolution?

Non-European ancestry

MHC/HLA DIVERSITY

GENOMIC MEDICINE & TRAIT PREDICTION

LONG-TERM BALANCING SELECTION

Evolutionary mechanism
that shapes diversity within species

Bitarello et al. (2018), *GBE*
Giner-Delgado (+6), **Bitarello,** (+12) (2019), *Nat. Comm.*
Mathieson (+6), **Bitarello**, (+5), eQTLGen Consortium, BIOS Consortium,
Human Reproductive Behaviour Consortium (2020), *BiorXiv*

Conclusions from balancing selection research

- Support for divergent allele advantage model in HLA antigen recognition site evolution
- another
- another
- another
- another

Conclusions from polygenic risk prediction research

*

*

*

*

*

`

Ongoing research

`

Future Directions

Big Questions

- How does genetic ancestry influence disease prediction models?
- How can the cataloging of diverse genetic ancestries improve our understanding of human diversity and disease?
- What determines complex traits and how important are gene-by-gene and gene-by-environment interactions in shaping complex phenotypes/diseases
- Why are so many GWAS hits situated in the HLA region, even for non-immune traits?
- How does adaptive evolution generate maladaptive consequences for humans and other primates?
- What specific mechanisms of balancing selection have acted upon different traits and are there trends there?

`

Future Directions

Axis I: POLYGENIC SCORE PREDICTION AND DISEASE FINE-MAPPING IN ADMIXED POPULATIONS

Axis II: INVESTIGATING THE EVOLUTIONARY HISTORY OF THE HLA AND OTHER DEFENSE RELATED GENES

Axis I

Axis II

`

Acknowledgements

Iain Mathieson

Diogo Meyer

Aida Andrés

Joshua Schmidt

Cesare de Filippo

Rodrigo dos Santos Francisco

Débora Brandt

Neale Lab

UK Biobank

Women's Health Initiative

Jackson Heart Study

Health and Retirement Study

1000 Genomes Project

Charles E. Kaufman Foundation
A SUPPORTING ORGANIZATION OF THE PITTSBURGH FOUNDATION

FAPESP
SÃO PAULO RESEARCH FOUNDATION

Perelman School of Medicine
University of Pennsylvania

THE PITTSBURGH FOUNDATION

NIH NIGMS

Thank you!

Questions?