

Evolutionary causes and clinical implications of genetic diversity

Research Talk

Bárbara Domingues Bitarello, Postdoctoral Scholar

Perelman School of Medicine, University of Pennsylvania

Introduction

Research Project 1

Research Project 2

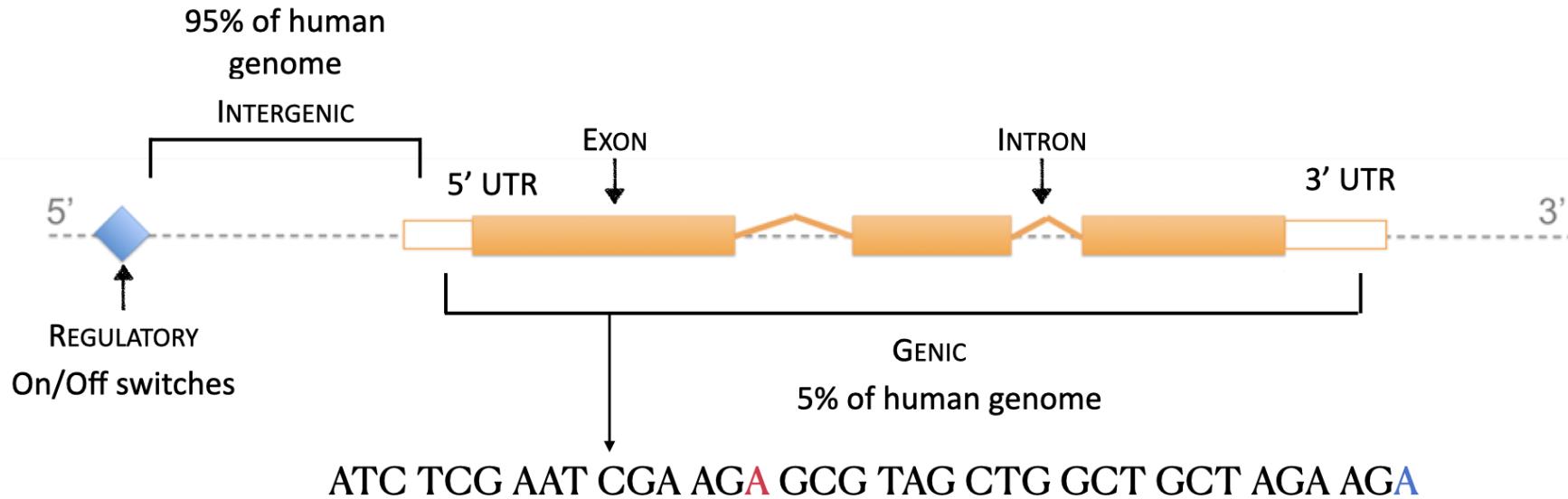
Research Project 3

Conclusions & Future Directions

My research: an overview

My research: zooming in

Architecture of a genome



First Base	U	C	A	G	Third Base
U	UUU phe	UCU ser	UAU tyr	UGU cys	U
	UUC phe	UCC ser	UAC tyr	UGC cys	C
	UUA leu	UCA ser	UAA stop	UGA stop	A
	UUG leu	UCG ser	UAG stop	UGG trp	G
C	CUU leu	CCU pro	CAU his	CGU arg	U
	CUC leu	CCC pro	CAC his	CGC arg	C
	CUA leu	CCA pro	CAA gln	CGA arg	A
	CUG leu	CCG pro	CAG gln	CGG arg	G
A	AUU ile	ACU thr	AAU asn	AGU ser	U
	AUC ile	ACC thr	AAC asn	AGC ser	C
	AUA ile	ACA thr	AAA lys	AGA arg	A
	AUG met	ACG thr	AAG lys	AGG arg	G
G	GUU val	GCU ala	GAU asp	GGU gly	U
	GUC val	GCC ala	GAC asp	GGC gly	C
	GUA val	GCA ala	GAA glu	GGA gly	A
	GUG val	GCG ala	GAG glu	GGG gly	G

Nonsynonymous: changes the amino acid
(A->C; arg-> ser)

Synonymous: no change of amino acid;
silent (A->G; arg-> arg)

Processes that shape genomic diversity in the absence of selective pressures

Selective processes that shape genomic diversity

Introduction

Research Project 1

Research Project 2

Research Project 3

Conclusions & Future Directions

MHC molecules present antigens and unleash immune responses

MHC/HLA: an extreme instance of balancing selection

Is there support for the divergent allele advantage model?

Is there support for the divergent allele advantage model?

Hypothesis: divergent allele advantage model (heterozygote advantage is greater for genotypes with more divergent alleles) explains selective patterns in the HLA genes

Prediction: increased signatures of balancing selection between pairs of HLA alleles that match different antigens

To answer these questions, I looked at the human class I HLA genes

Introduction

Research Project 1

Research Project 2

Research Project 3

Conclusions & Future Directions

**How prevalent has balancing selection
been throughout human evolution?**

**What are the instances beyond the MHC
loci?**

Can we devise a statistical test specific for signatures of balancing selection?

Approach: combine independent signatures into one statistic

If so, how prevalent has it been throughout human evolution?

Approach: perform a genome-wide scan using a test specifically tailored for balancing selection

Are there common trends amongst the candidate regions/genes?

Approach: Explore candidate regions/genes for common functions/patterns

.

Can we devise a statistical test specific for signatures of balancing selection?

A method to detect signatures of balancing selection

Pervasiveness and targets of balancing selection in humans

Introduction

Research Project 1

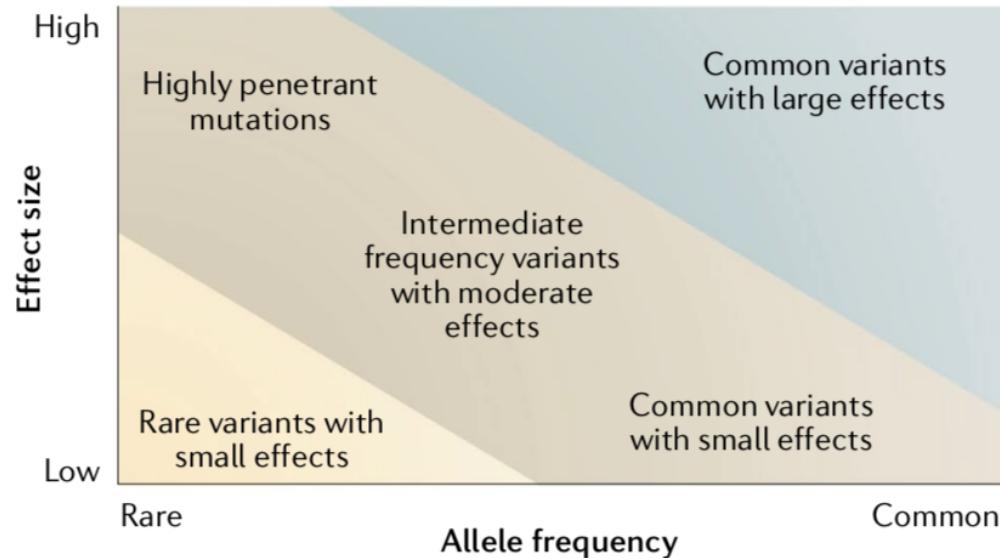
Research Project 2

Research Project 3

Conclusions & Future Directions

Genome-wide association studies (GWAS)

Many variants with small effect size



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

Rare, monogenic diseases/traits:

phenylketonuria, sickle-cell anemia, Duchenne muscular dystrophy

complex, common diseases/traits:

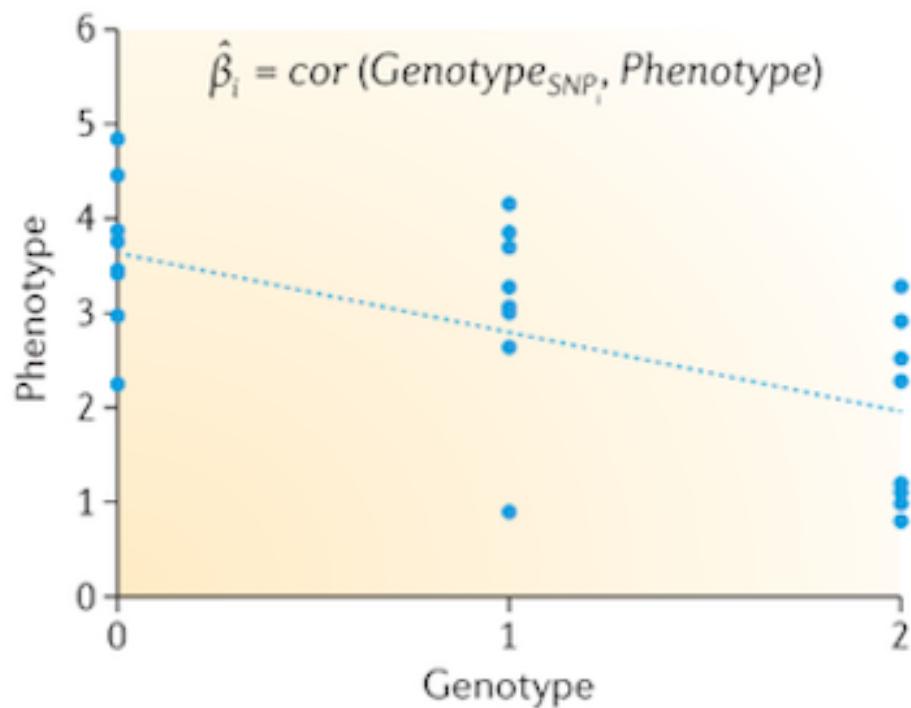
cardiovascular disease, T2 diabetes, cancers, **height**,
BMI

Polygenic traits and their predictive power

Phenotype	Variance_Explained	Variants
height	25.0	3000
schizophrenia	7.0	100
ADHD	5.5	100

PS: for Europeans ancestry only...

Polygenic risk scores combine all variants with an effect on the phenotype



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

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

G : Effect allele dosage

j : Individuals

i : SNPs

independence

additive model

PRSs are appealing

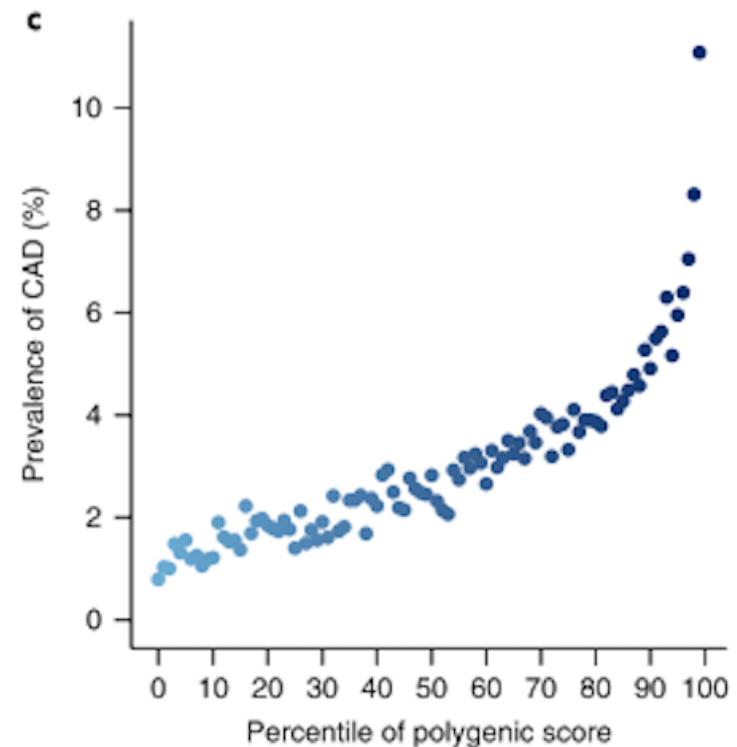
easy

promising

fast

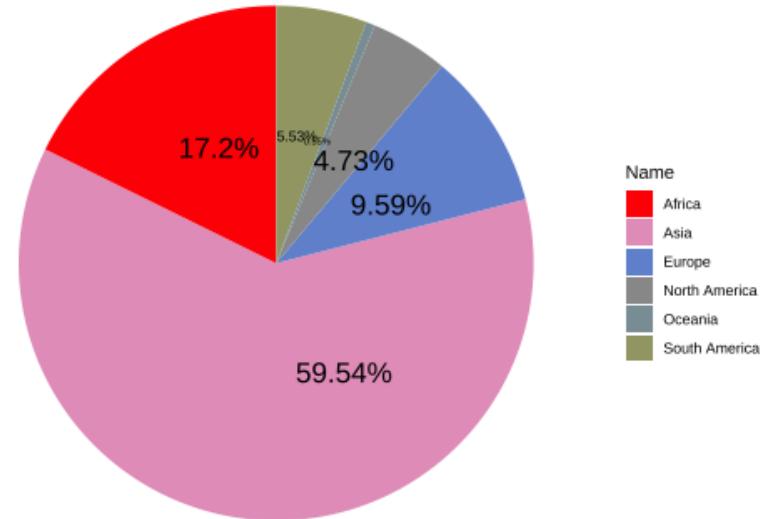
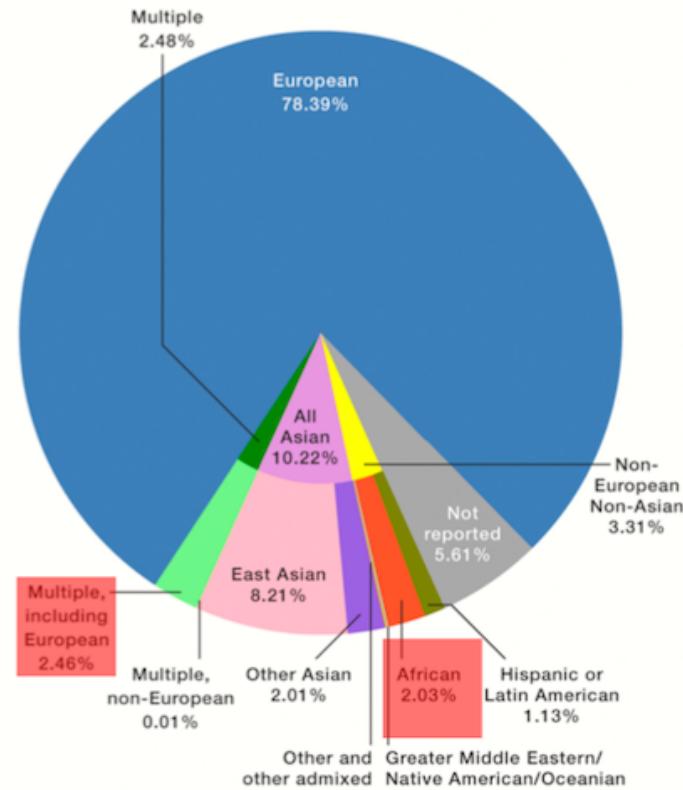
minimal requirements

Polygenic risk of cardiovascular disease



Khera et al. (2018), *Nat Genet*

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



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

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

How much do polygenic scores predict phenotypes in other populations?

How do these different factors affect prediction accuracy?

What can we do about it?

How do these different factors affect prediction accuracy?

Approach: examine how well we can predict a highly heritable trait in different populations, based on an European ancestry discovery cohort and explore the roles of different biological/statistical factors

What can we do about it?

Approach: explore whether incorporating fine-scale ancestry information into these scores improves their performance

Let's look at height



Wow well can we predict a highly heritable trait?

Wow well can we predict a highly heritable trait in different populations?

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

.

How has balancing selection shaped diversity in the HLA loci?

How prevalent has balancing selection been throughout evolution and what are the common trends?

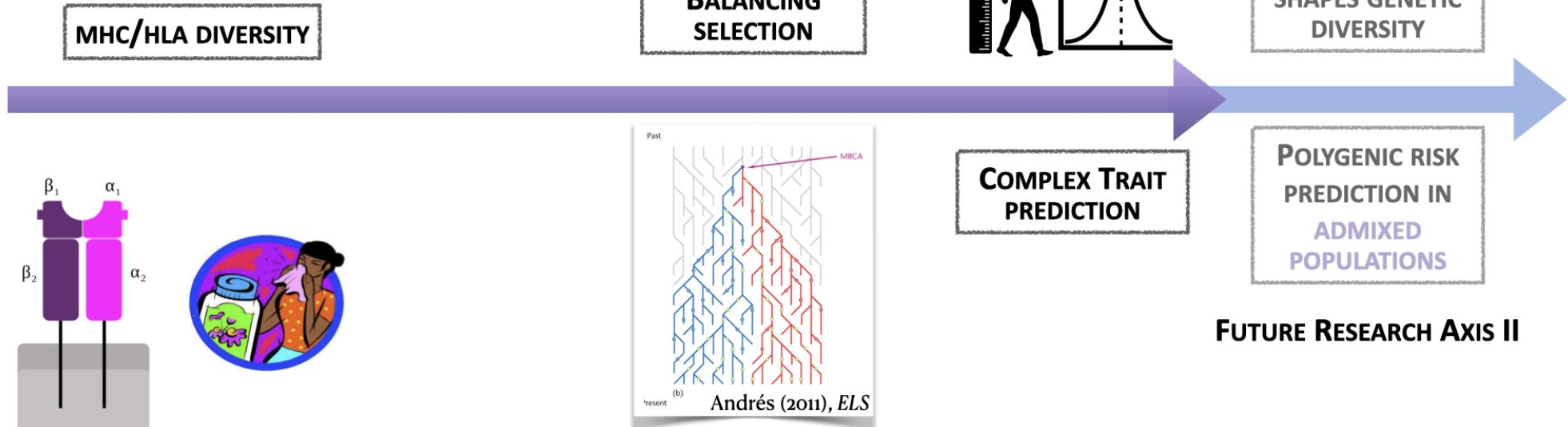
How useful are polygenic risk predictors for individuals with non-European ancestries?

- 1) There is support for the divergent allele advantage model in HLA evolution
- 2) A novel test, specific for balancing selection, revealed many new candidates, 1/3 of which are immune-related, and many others related to defense broadly and reproduction
- 3) At present, PRS approaches have little to no utility for individuals with non-European ancestry, many factors are responsible, need to diversify genetics studies at all stages

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

Bitarello* et al. (2018), *GBE*
Giner-Delgado (+6), Bitarello, (+12)
(2019), *Nat. Comm.*
Mathieson (+6), Bitarello, (+8)(2020),
BiorXiv

FUTURE RESEARCH AXIS I



Bitarello* & Mathieson (2020),
G3: Genes, Genomes, Genetics

Research Axis I: Balancing selection as a force that shapes diversity

Research Axis II: Polygenic risk prediction in admixed populations

A final thought

Acknowledgements

Iain Mathieson	Neale Lab
Diogo Meyer	UK Biobank
Aida Andrés	Women's Health Initiative
Joshua Schmidt	Jackson Heart Study
Cesare de Filippo	Health and Retirement Study
Rodrigo dos Santos Francisco	1000 Genomes Project
Débora Brandt	



