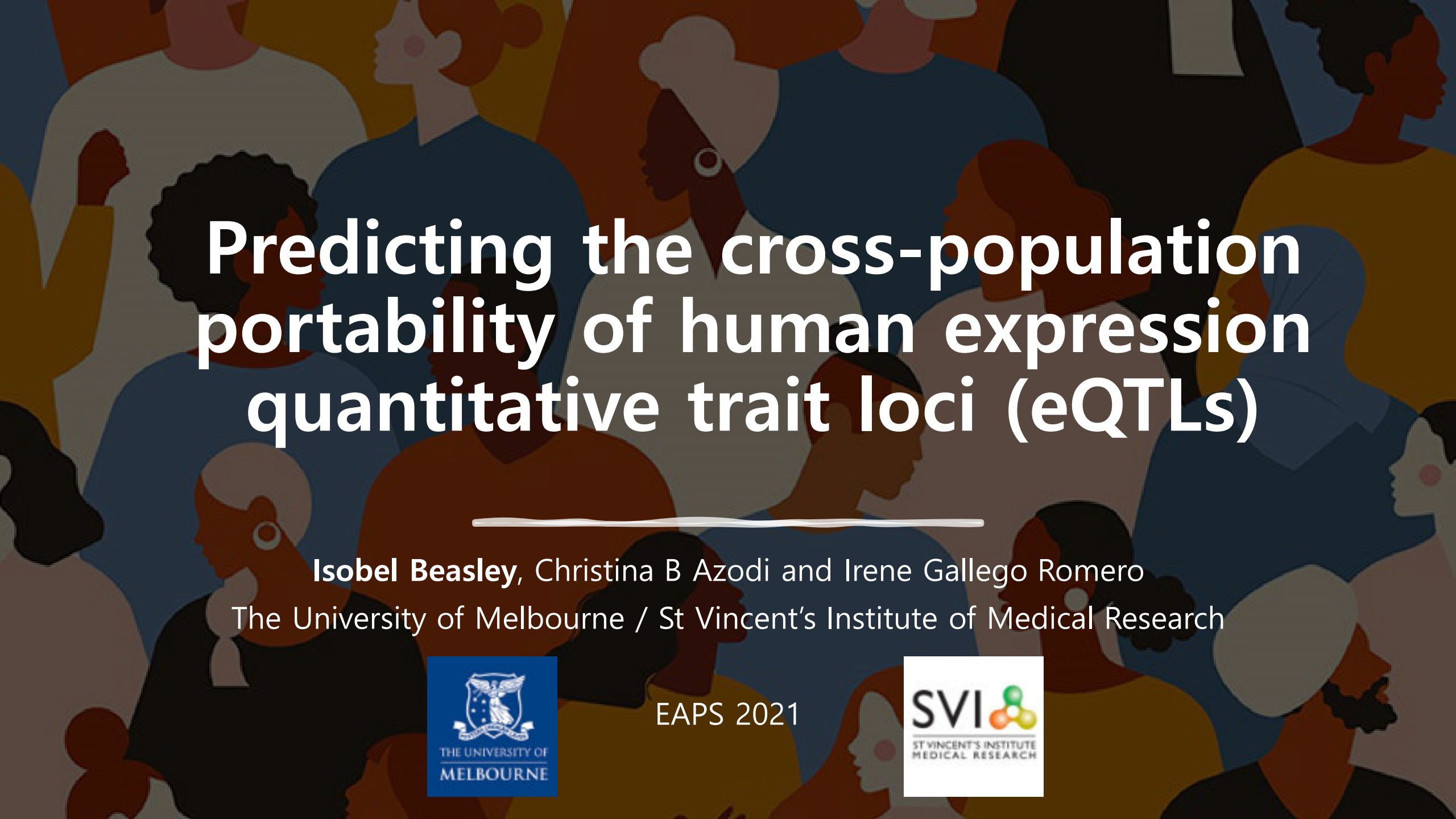
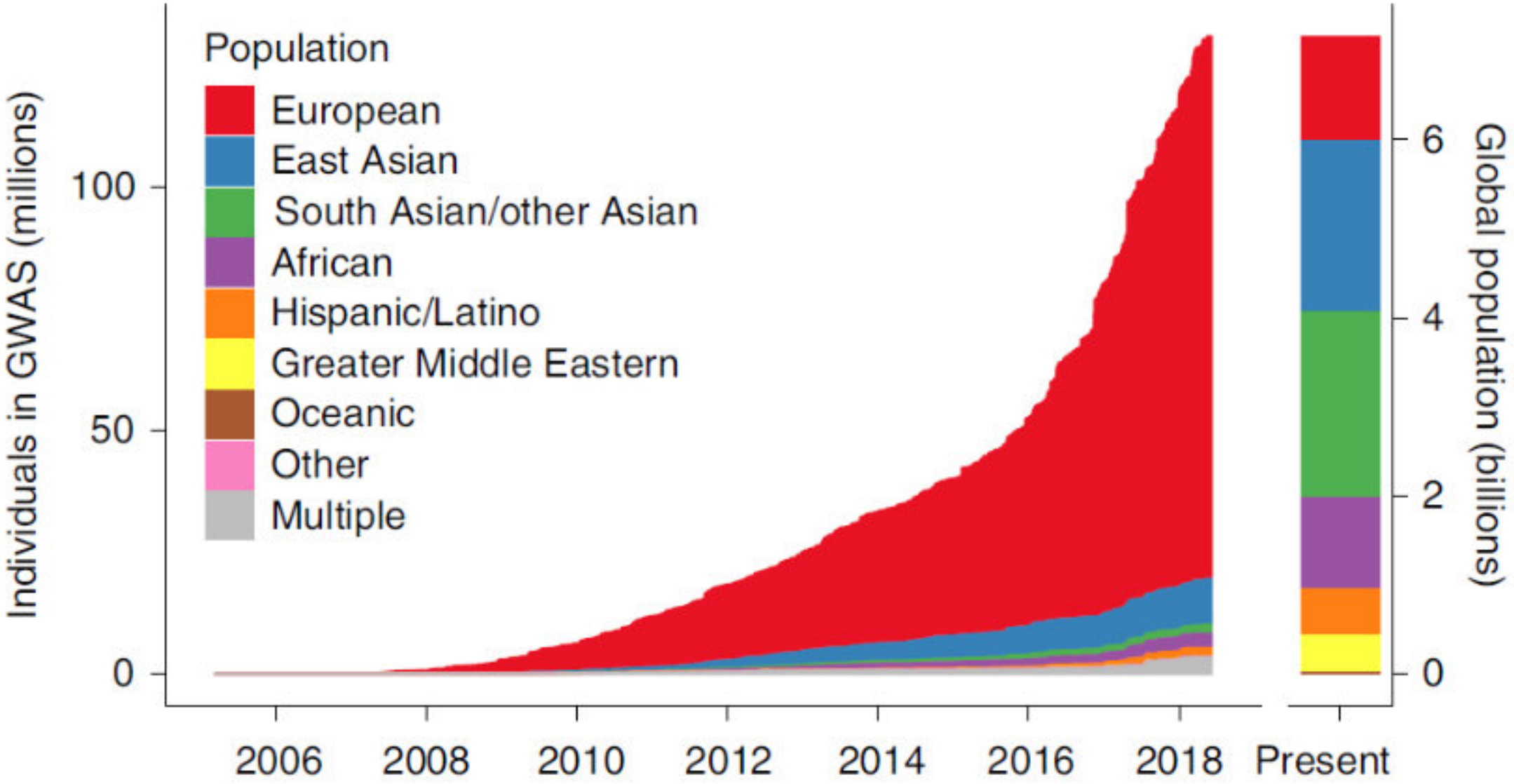


Predicting the cross-population portability of human expression quantitative trait loci (eQTLs)

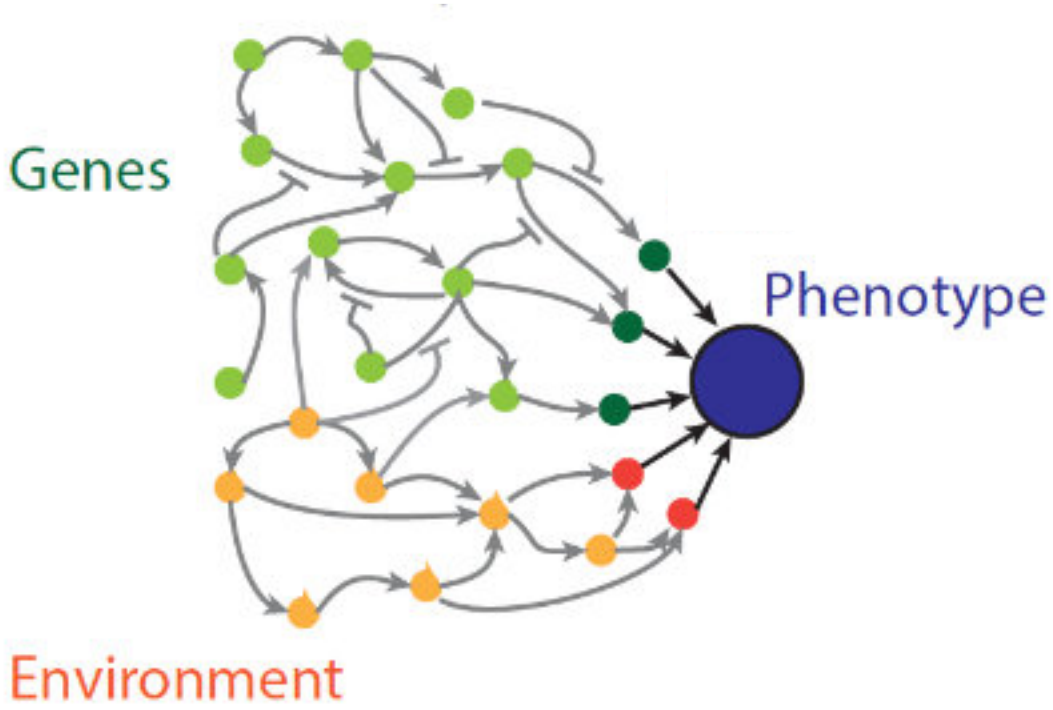
Isobel Beasley, Christina B Azodi and Irene Gallego Romero

The University of Melbourne / St Vincent's Institute of Medical Research

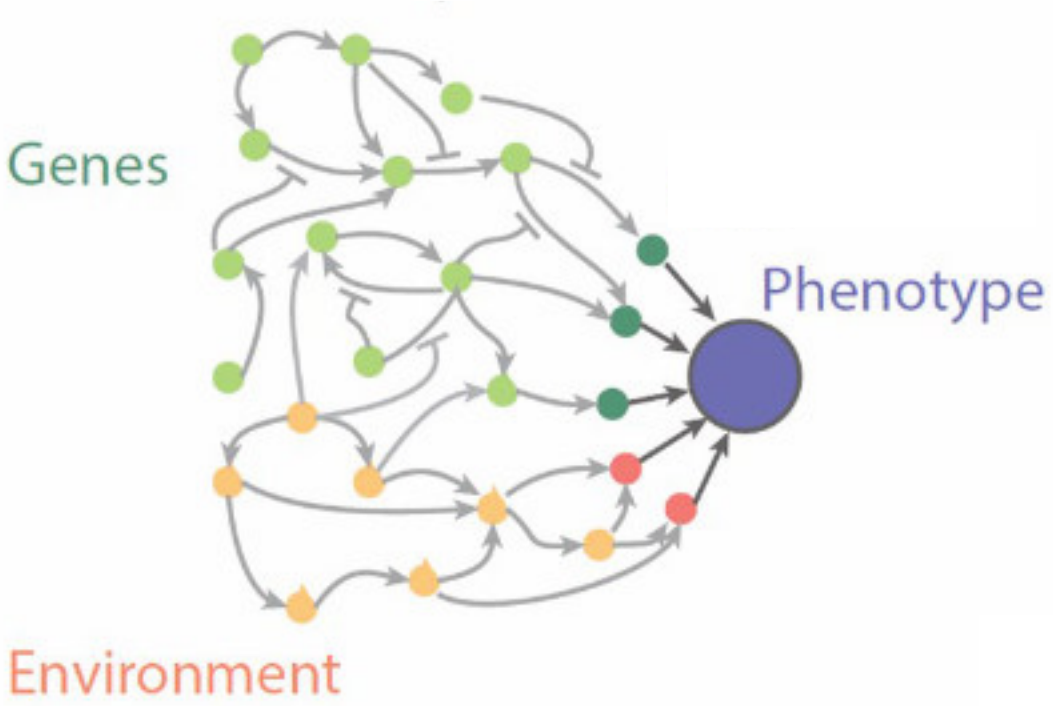
EAPS 2021

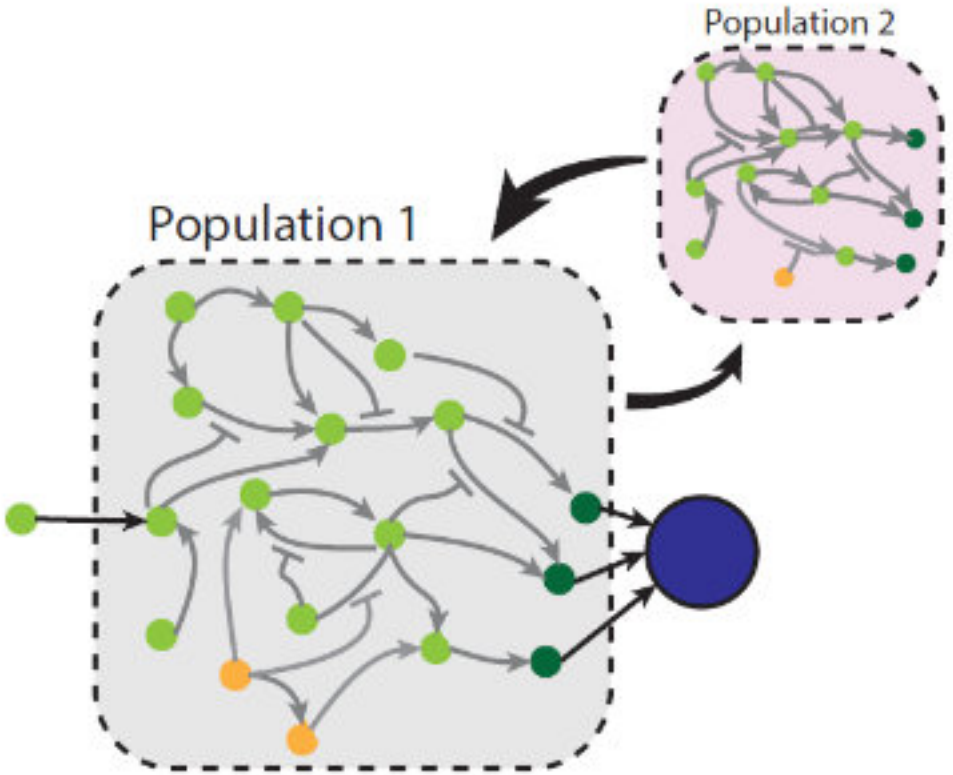


Martin et al., 2019, *Nature Genetics*



Iain Mathieson., 2021, *AJHG*

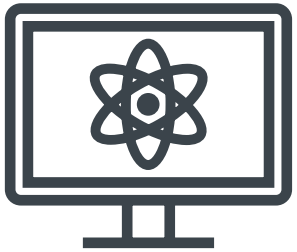




Iain Mathieson., 2021, *AJHG*

Outline

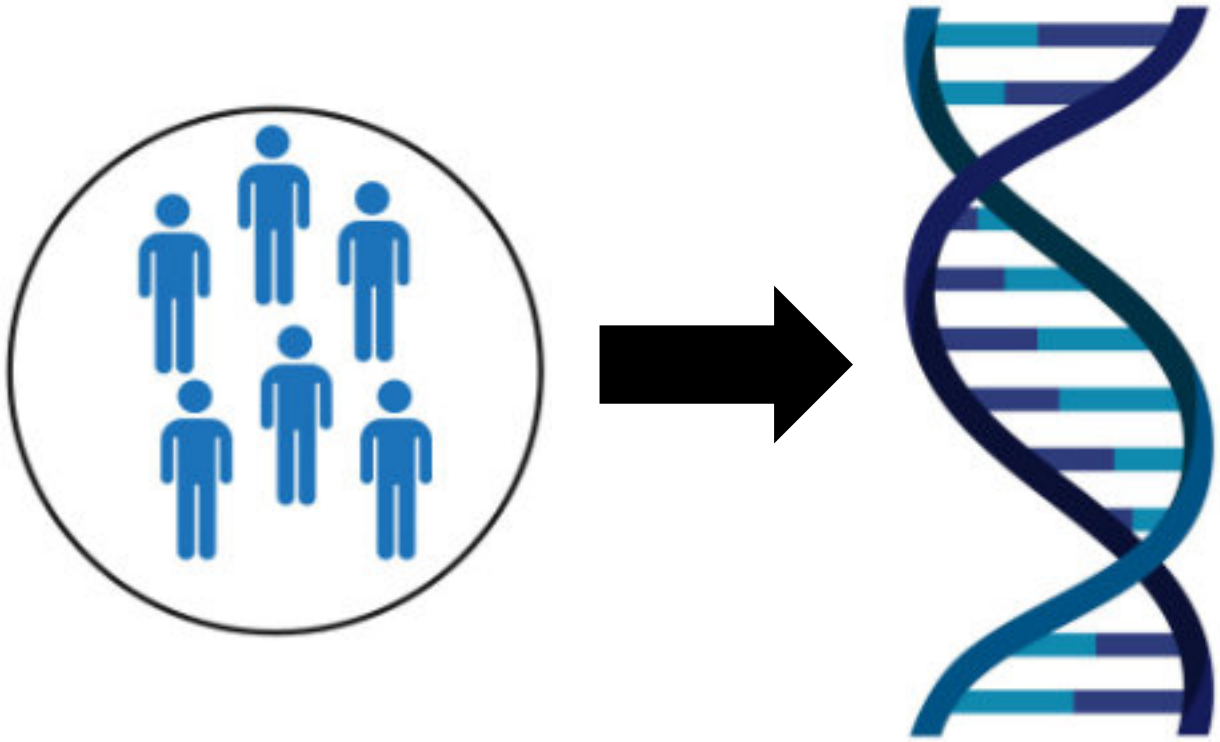
1. eQTLs



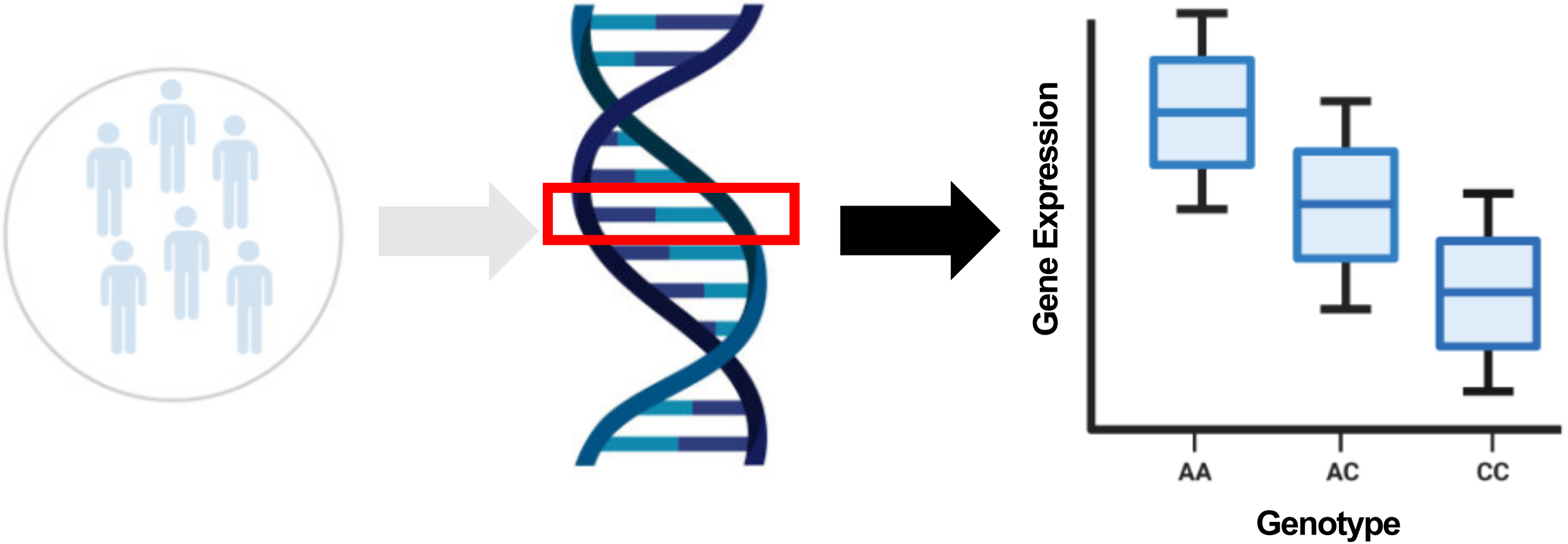
2. Building machine learning models

3. Findings (expected and puzzling)

Created with BioRender.com

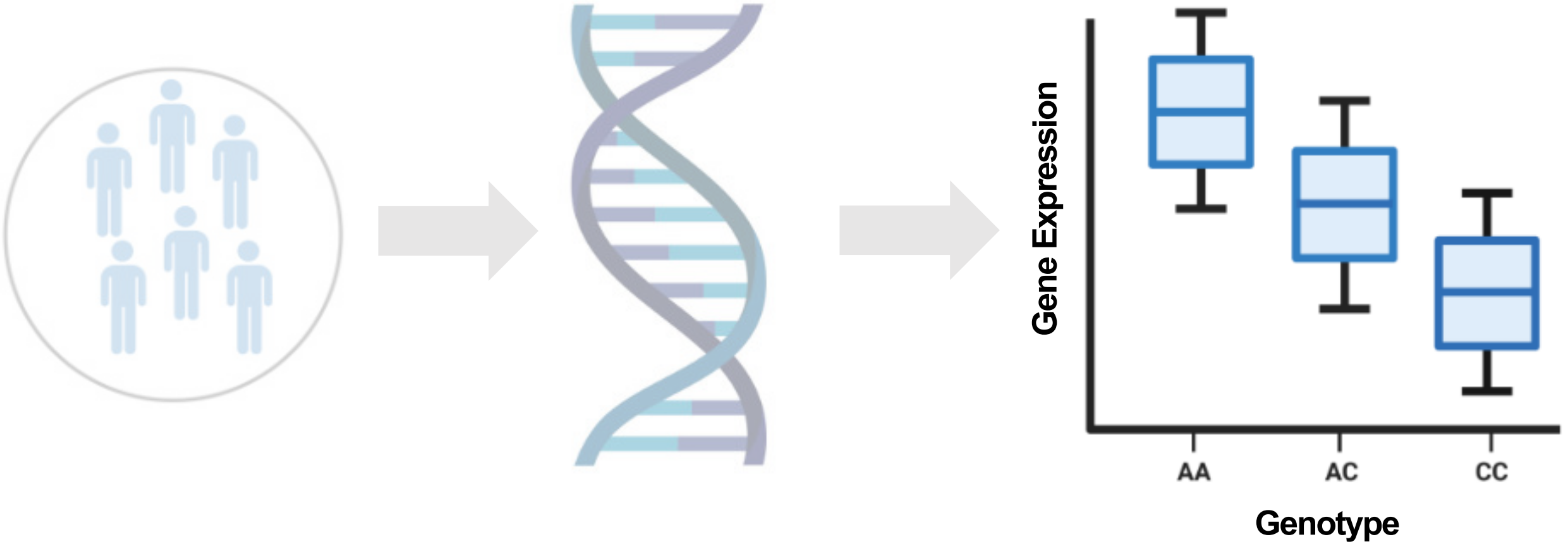


Created with BioRender.com

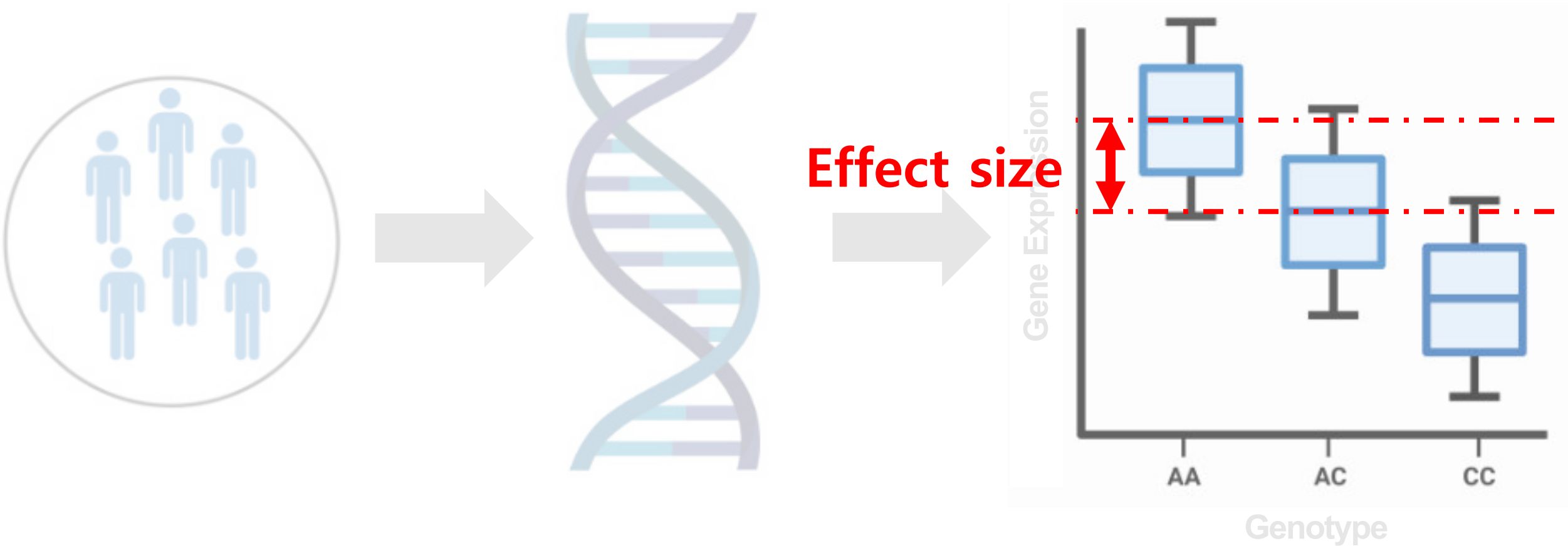


Created with BioRender.com

eQTLs (expression Quantitative Trait Loci)



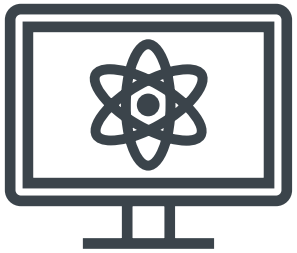
Created with BioRender.com



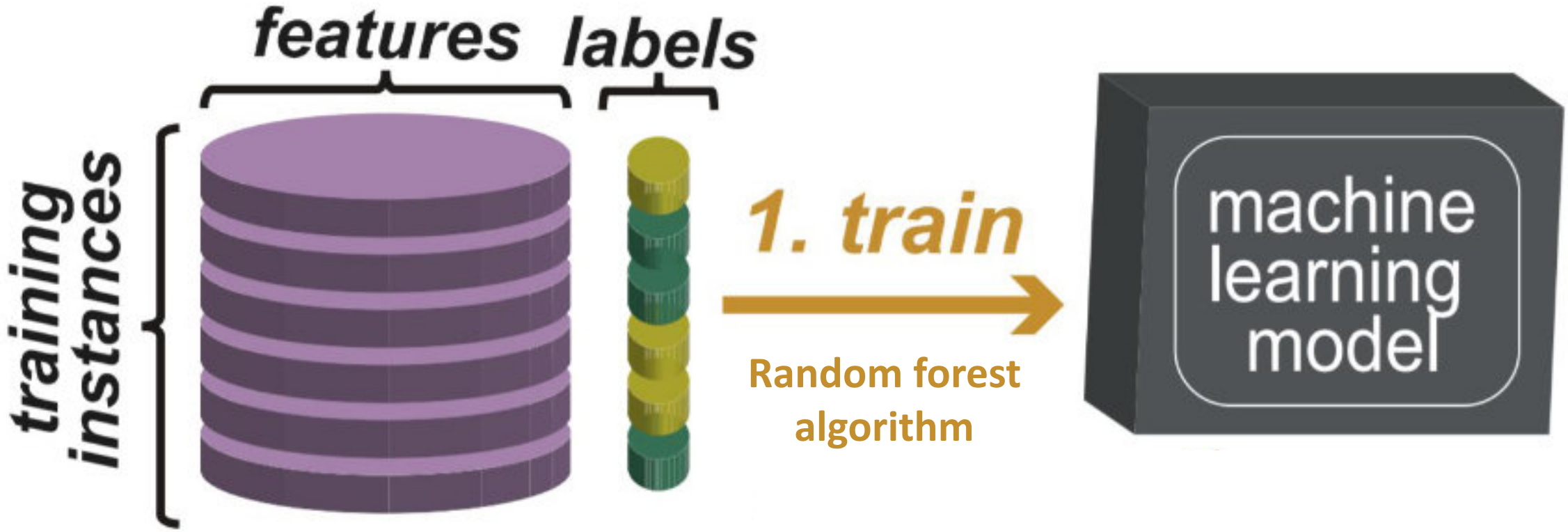
Created with BioRender.com

Outline

1. eQTLs



2. Building machine learning models



Azodi, Tang and Shiu, 2020, *Trends in Genetics*

The training data

European (n = 471, Lepik et al. 2017)

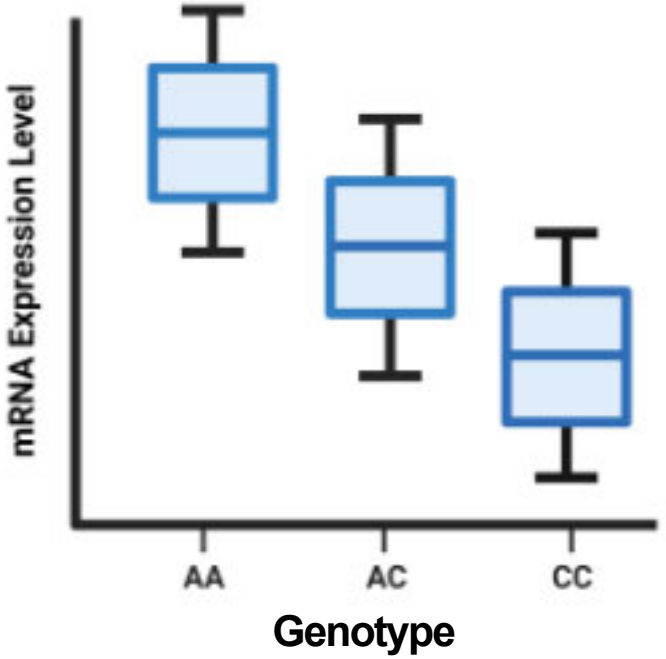
European (n = 379, GTEx. 2020)

European (n = 195, TwinsUK 2015)

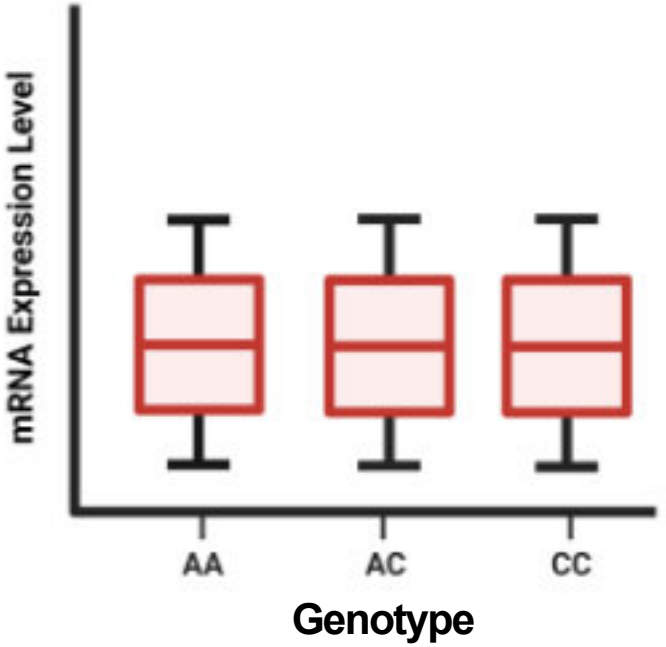
Indonesian (n = 115, Natri et al. 2020)

eQTLs

eQTLs



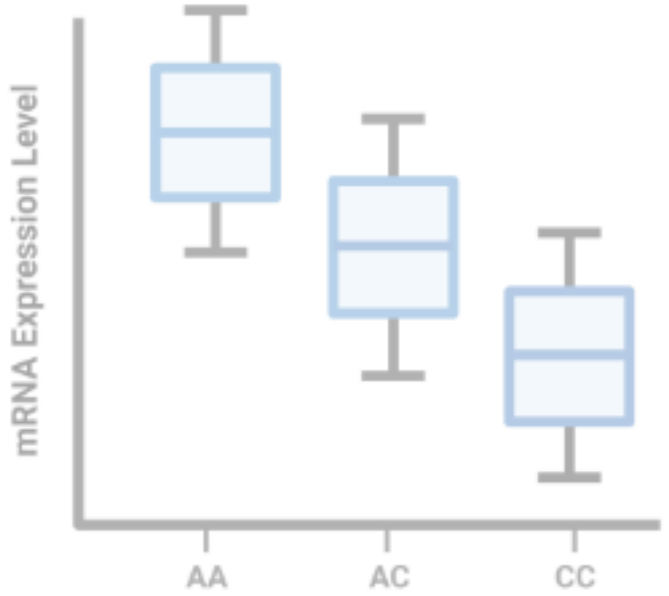
lfsr < 0.01

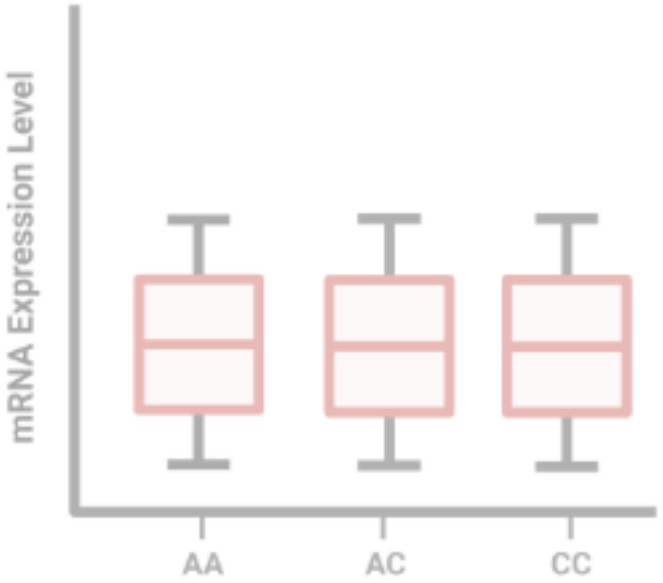


lfsr > 0.10

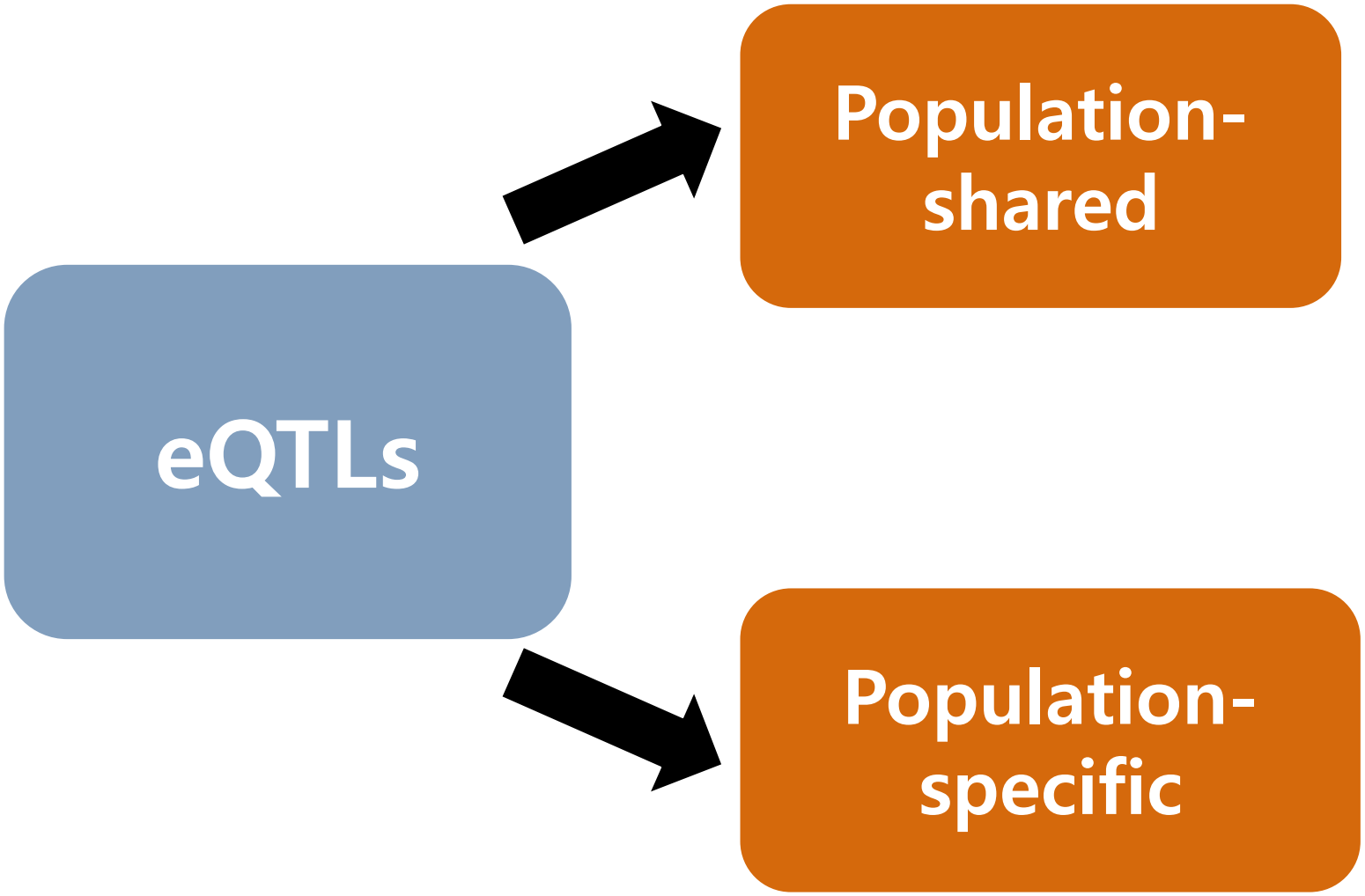
Created with BioRender.com

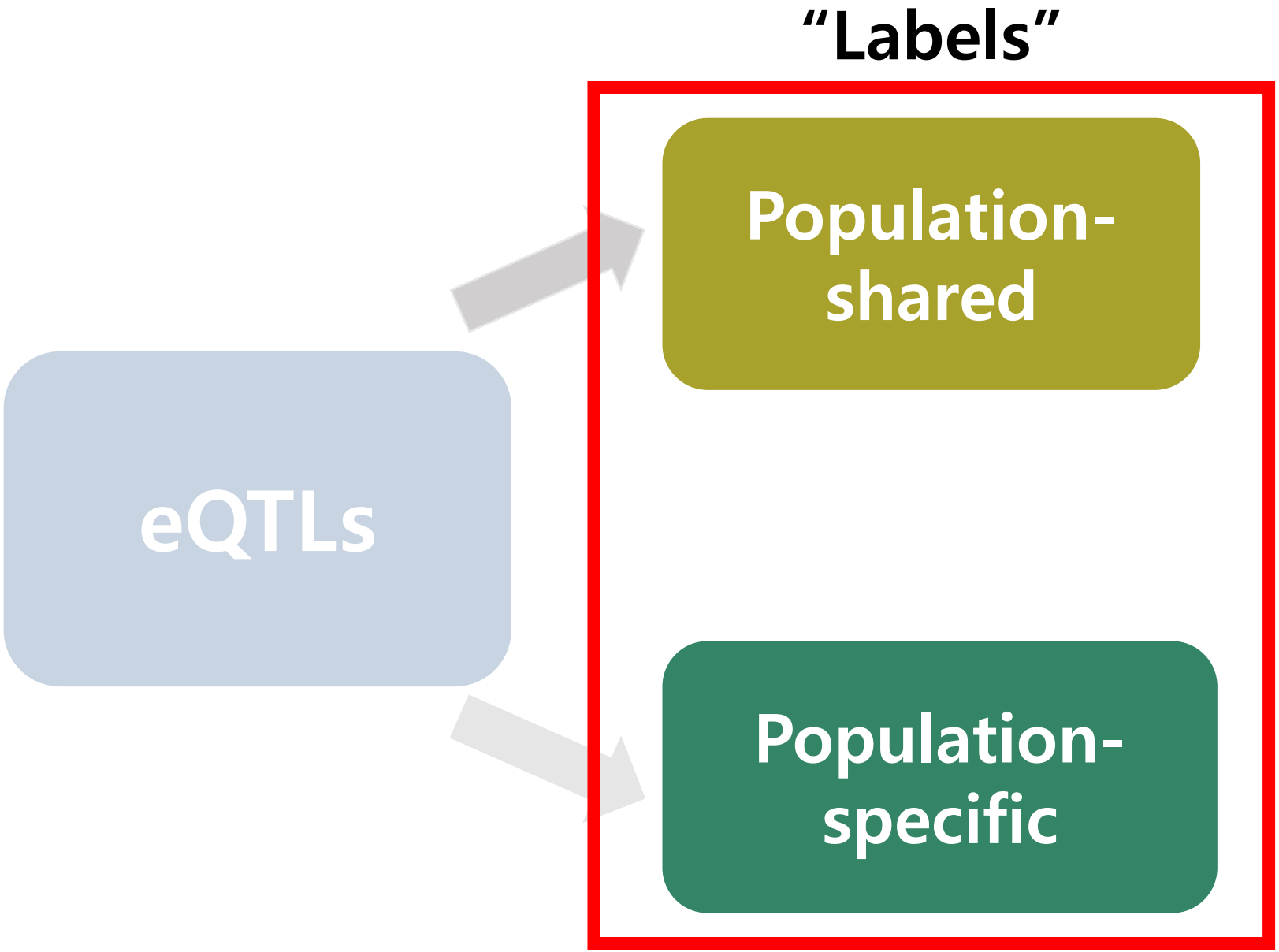
eQTLs

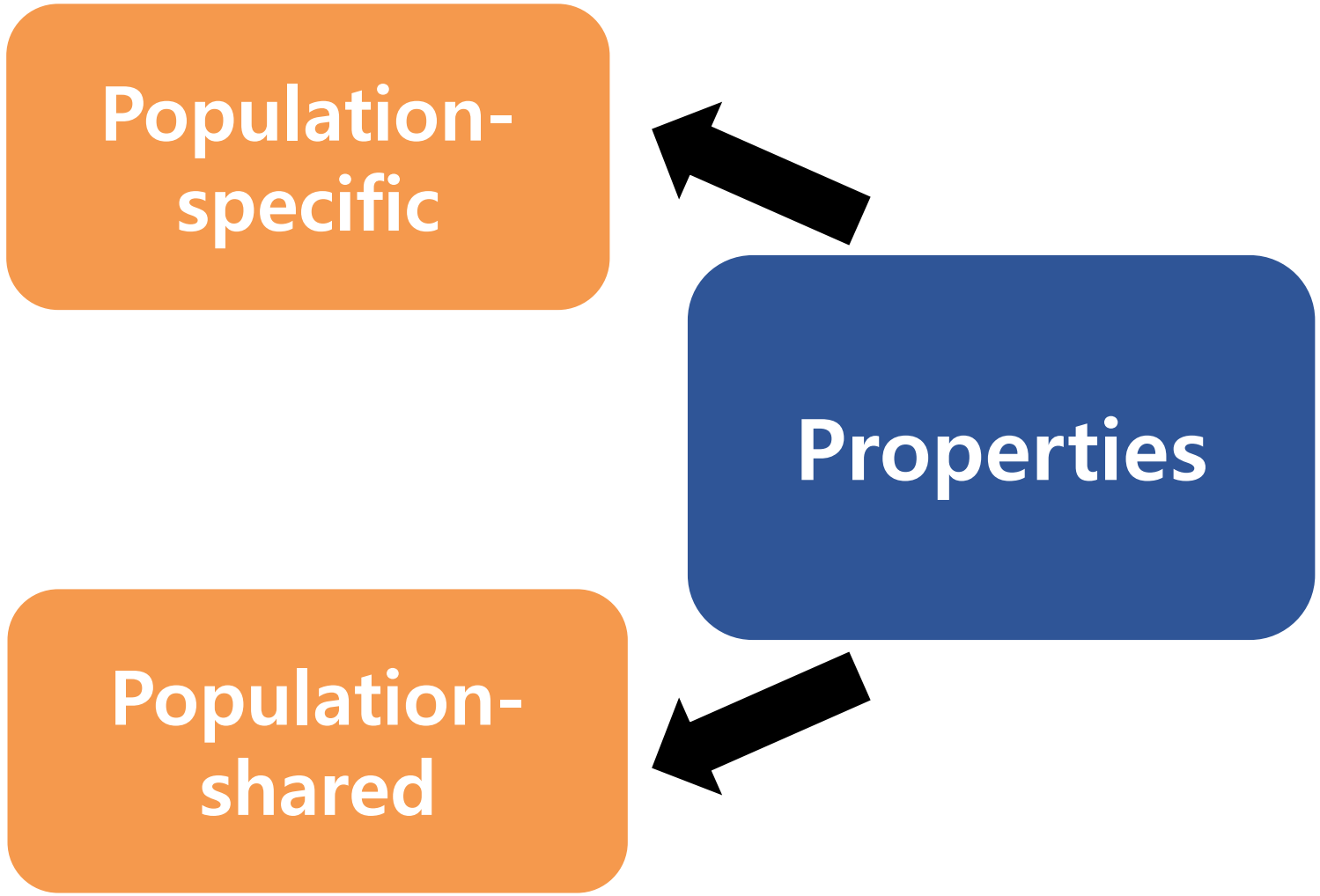


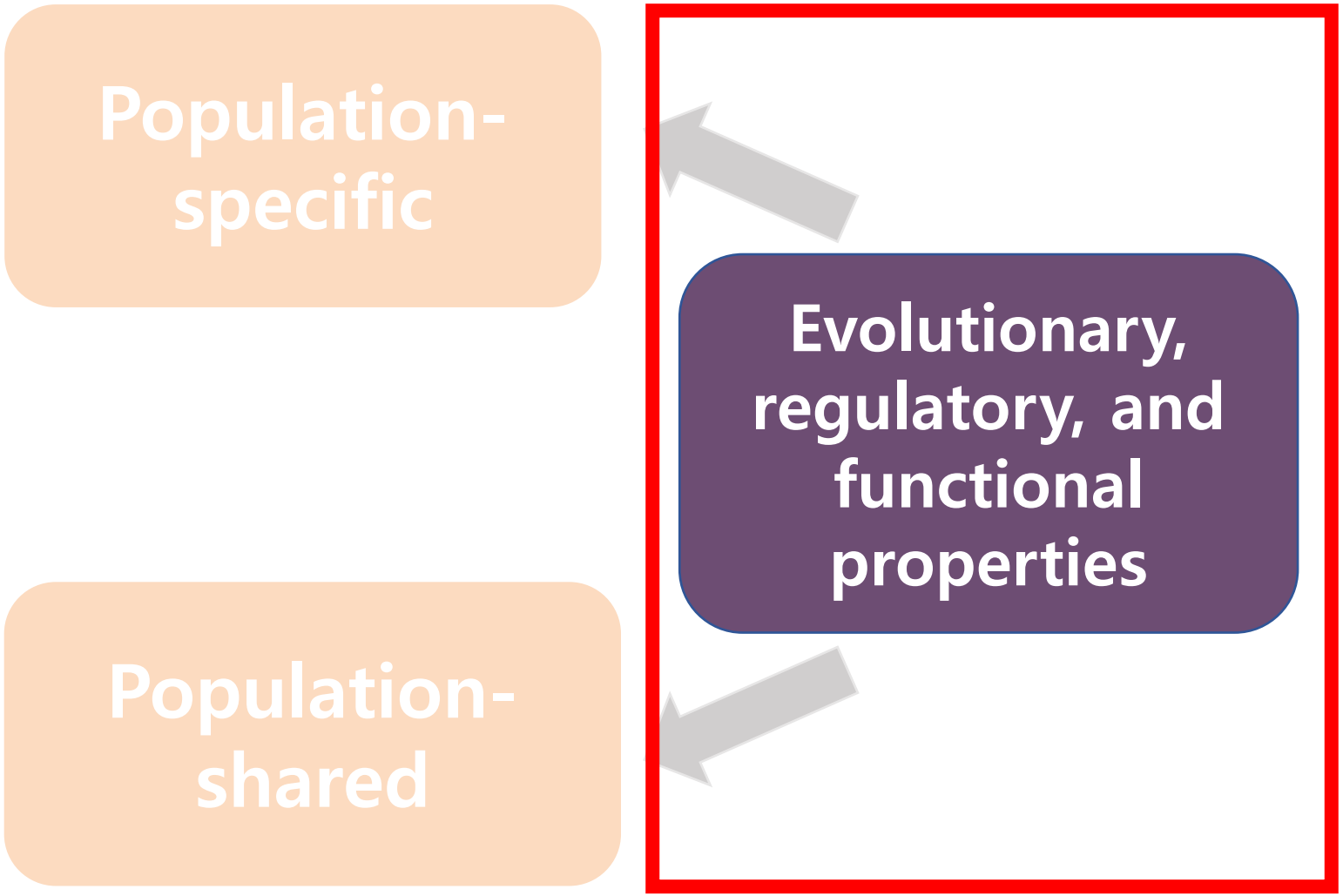


Population-specific

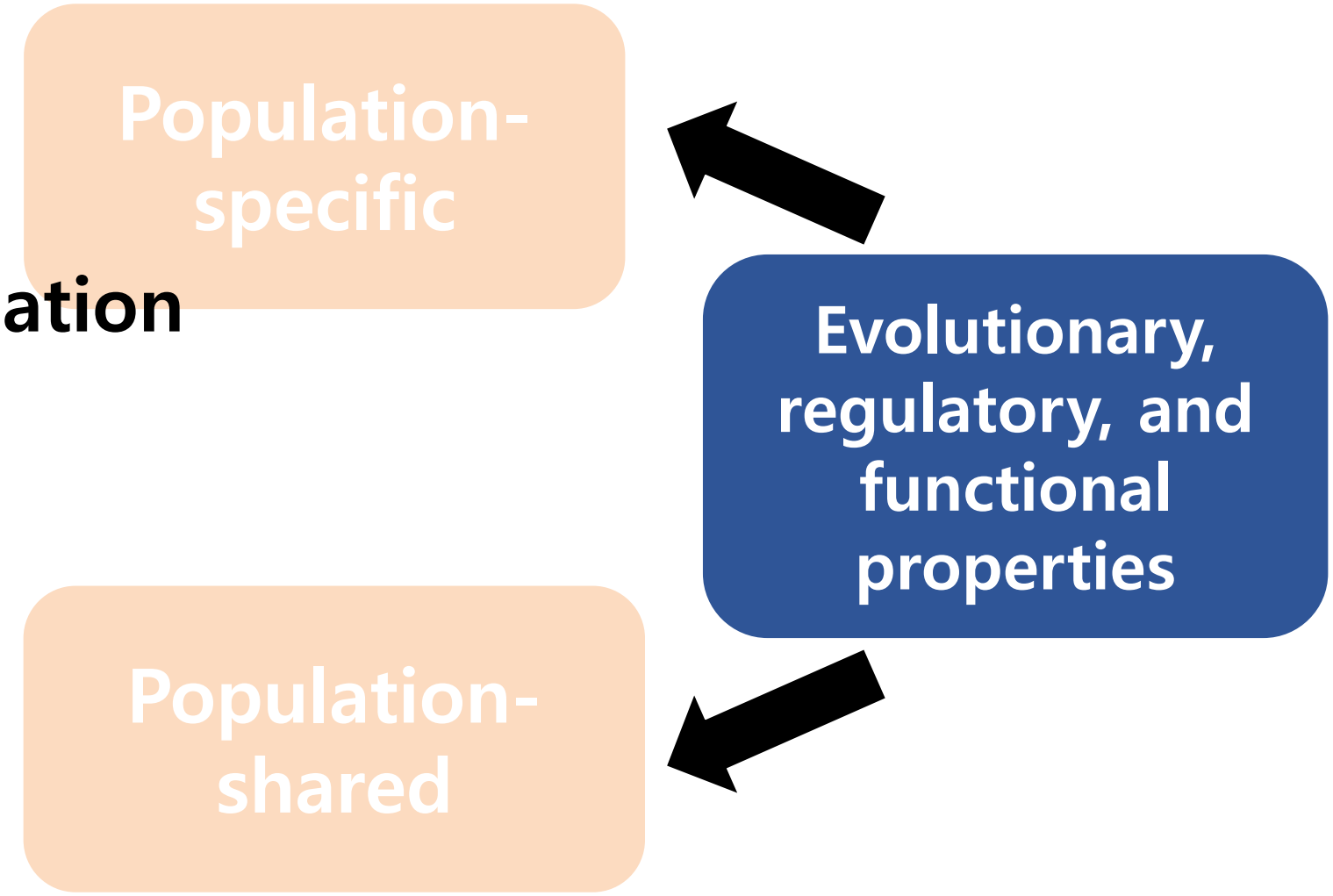

eQTLs
Population-shared
Population-specific


"Labels"
eQTLs
Population-shared
Population-specific


Population-specific
Properties
Population-shared


Population-specific
Population-shared
Evolutionary, regulatory, and functional properties
"Features"

- eQTL effect size
- Allele frequency
- Conservation
- Gene Expression
- SNP Genomic location
- % Nucleotide
- Gene Ontology

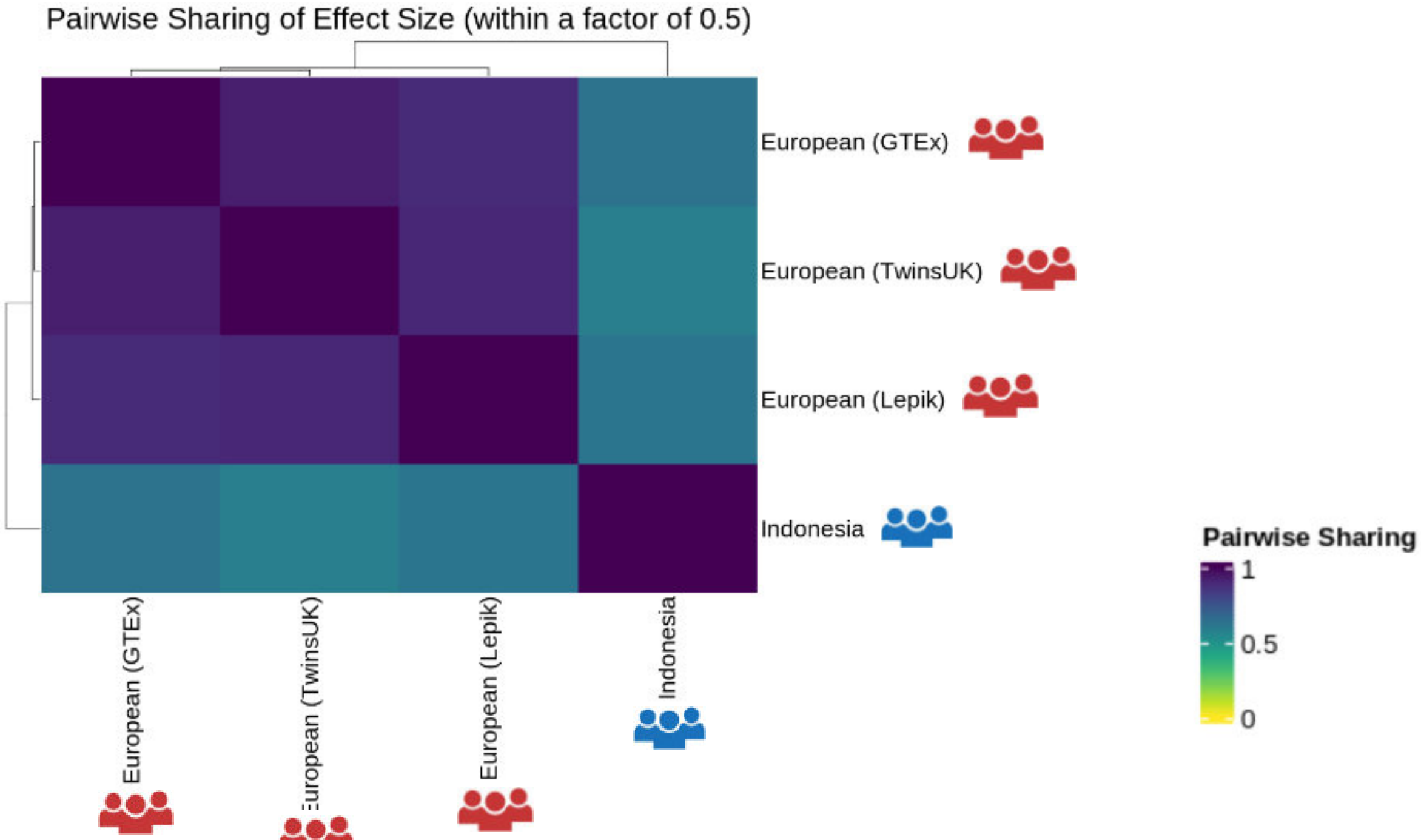


Outline

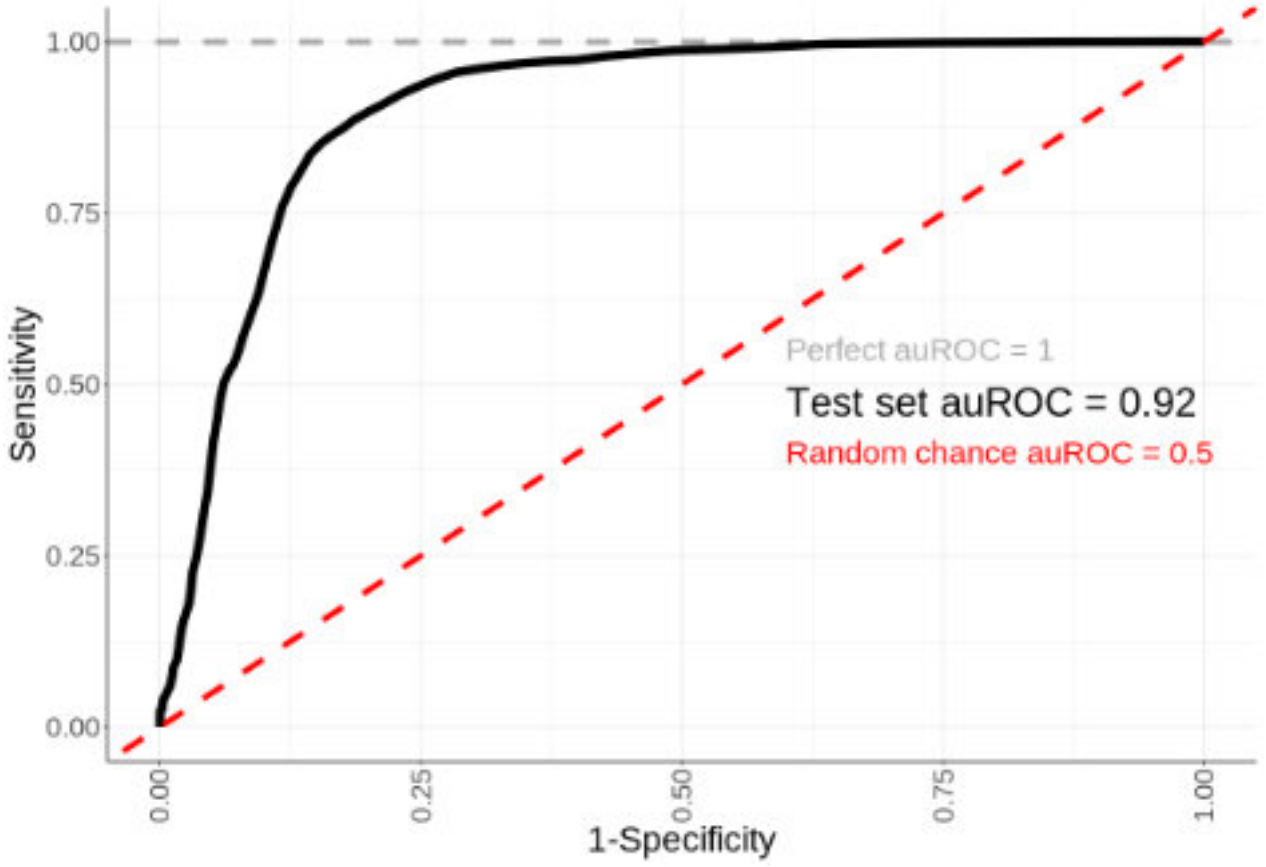
1. eQTLs

2. Building machine learning models

3. Findings (expected and puzzling)

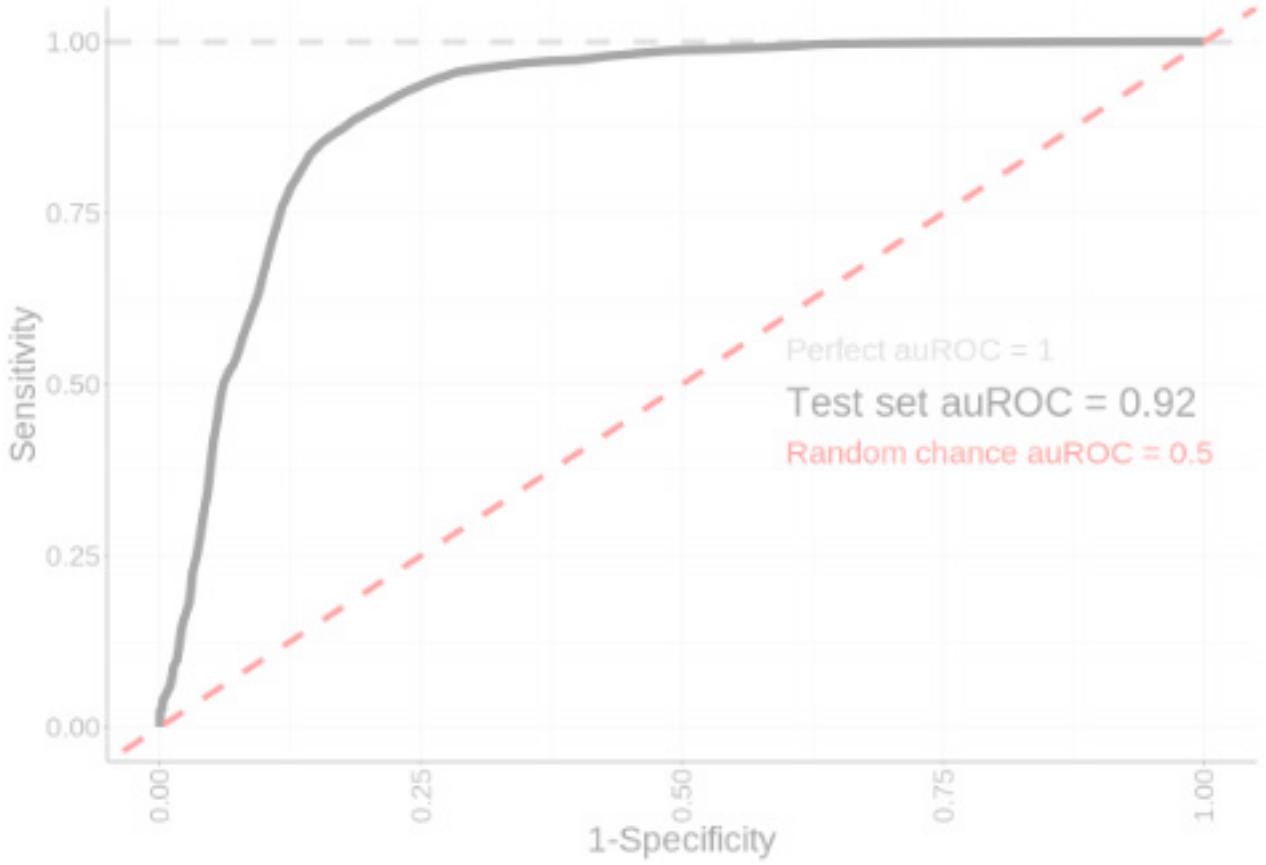

Pairwise Sharing of Effect Size (within a factor of 0.5)
European (GTEx)
European (TwinsUK)
European (Lepik)
Indonesia
European (GTEx)
European (TwinsUK)
European (Lepik)
Indonesia
Pairwise Sharing
1
0.5
0

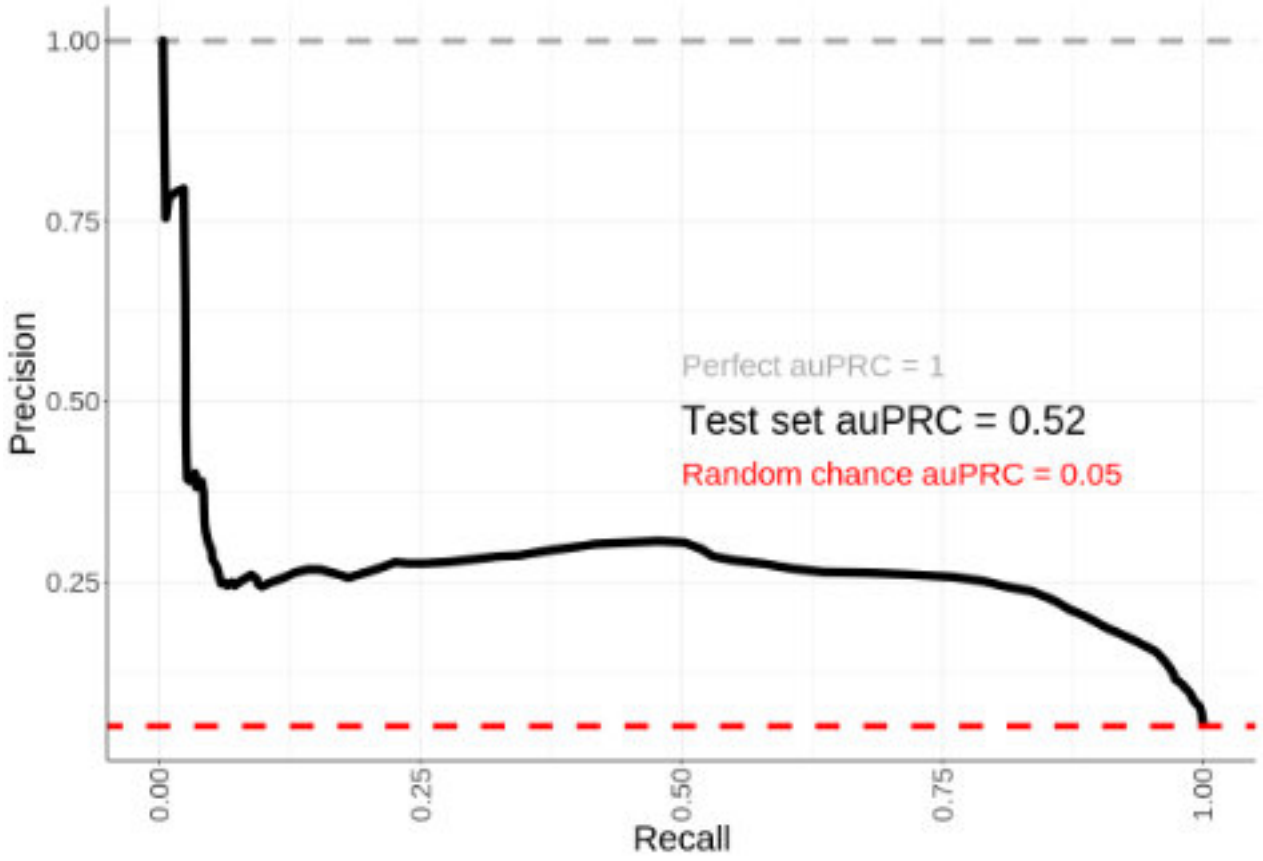
Prediction Performance



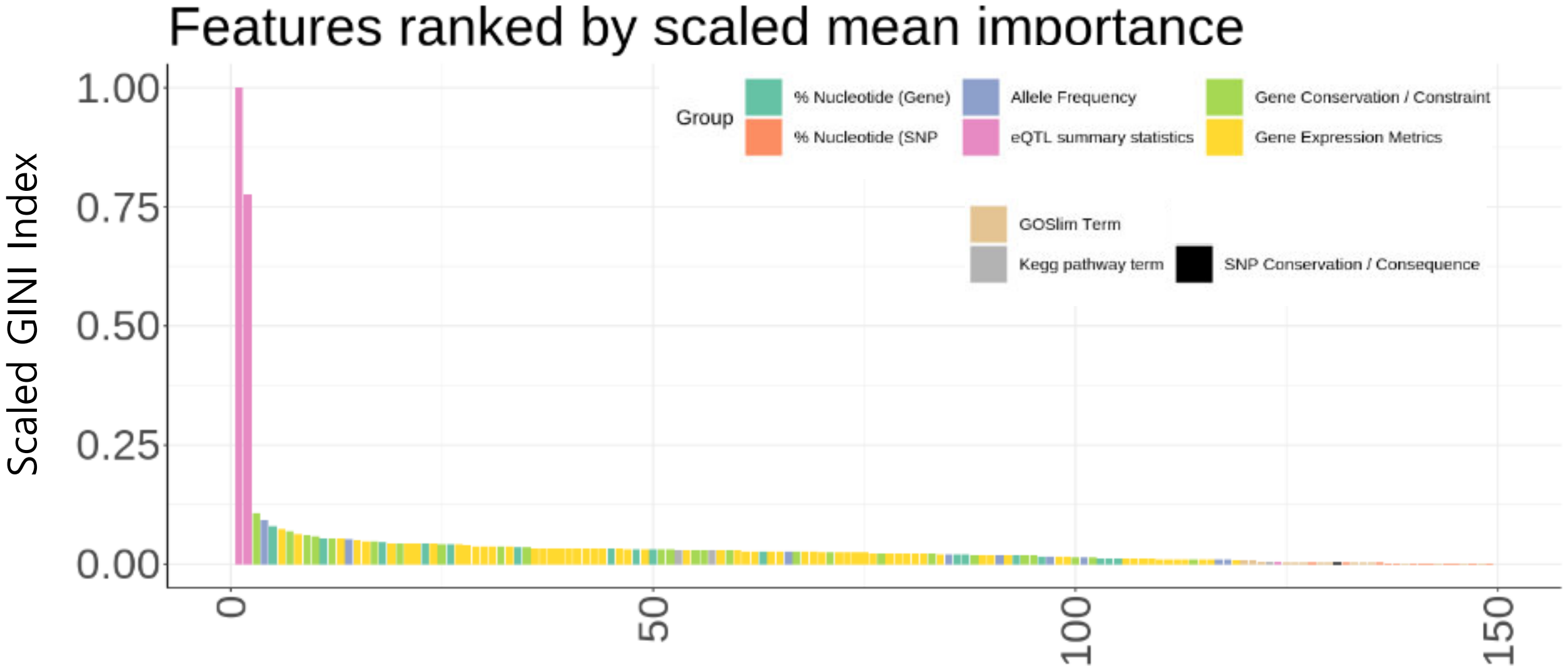
Test set: Chromosomes 8,16

Prediction Performance

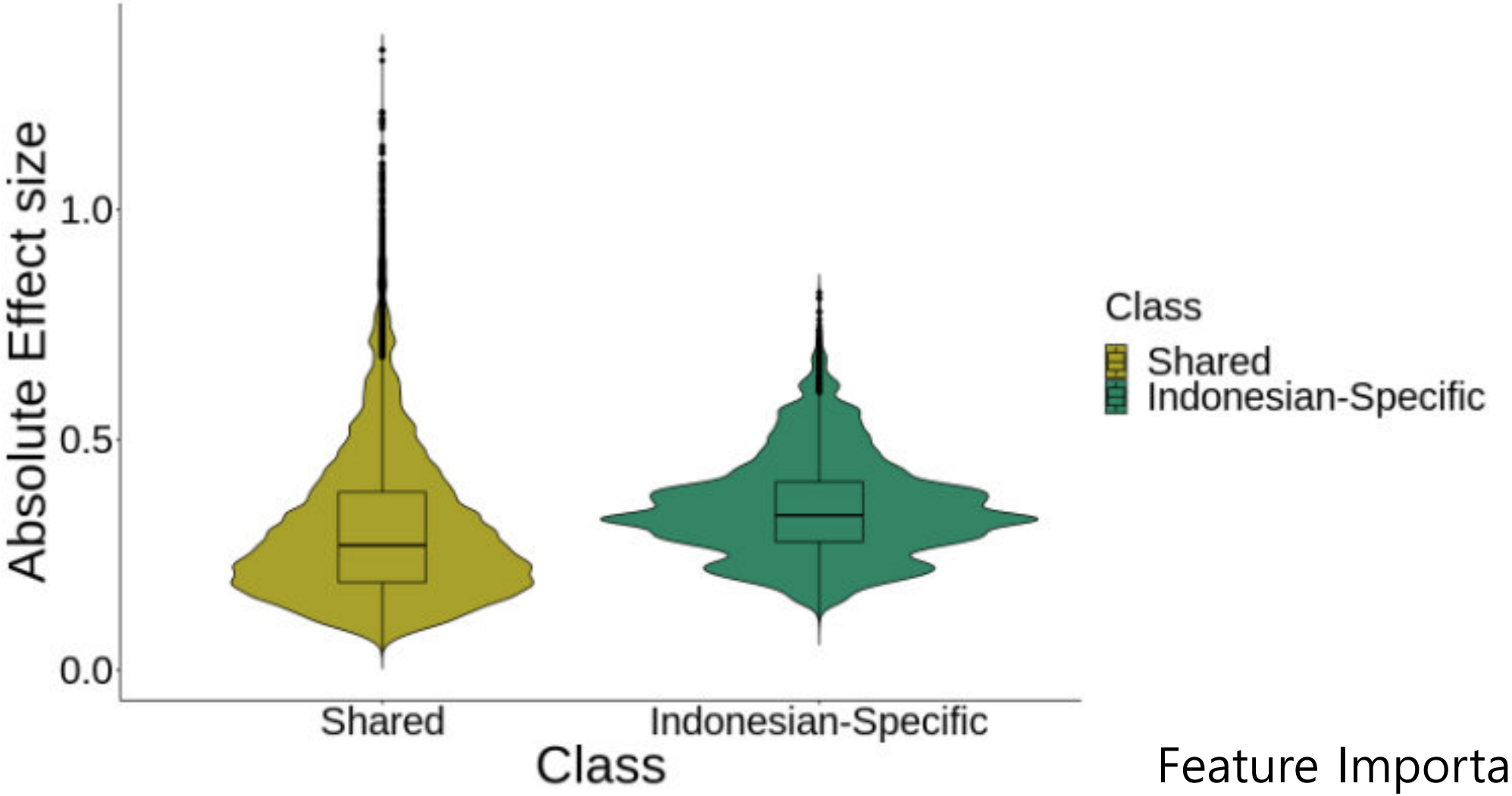




Test set: Chromosomes 8,16

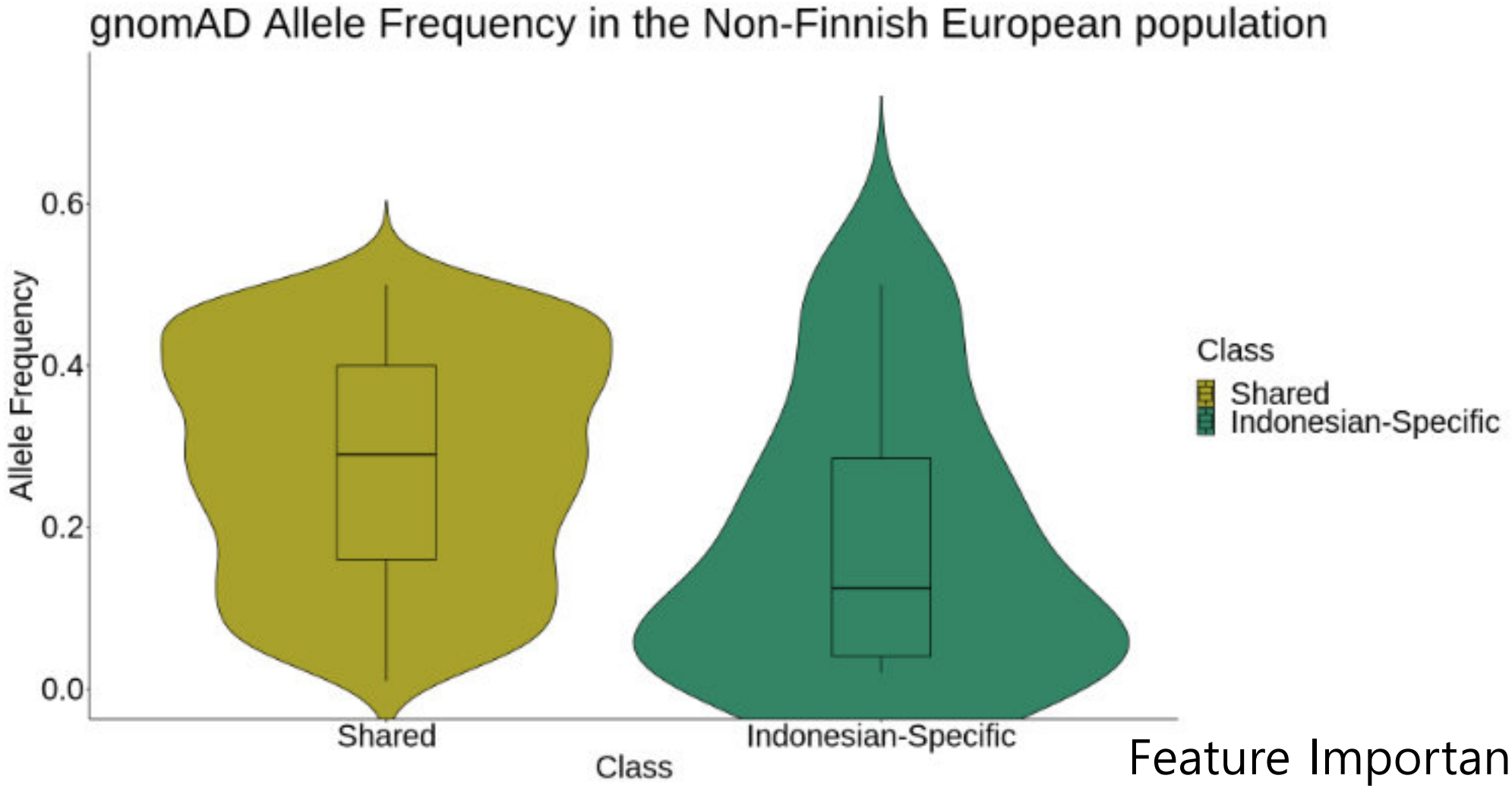

Features ranked by scaled mean importance
Scaled GINI Index
1.00
0.75
0.50
0.25
0.00
0
50
100
150
Group
% Nucleotide (Gene)
% Nucleotide (SNP
Allele Frequency
eQTL summary statistics
Gene Conservation / Constraint
Gene Expression Metrics
GOSlim Term
Kegg pathway term
SNP Conservation / Consequence

Population-specific eQTLs have higher effect sizes in their discovery population



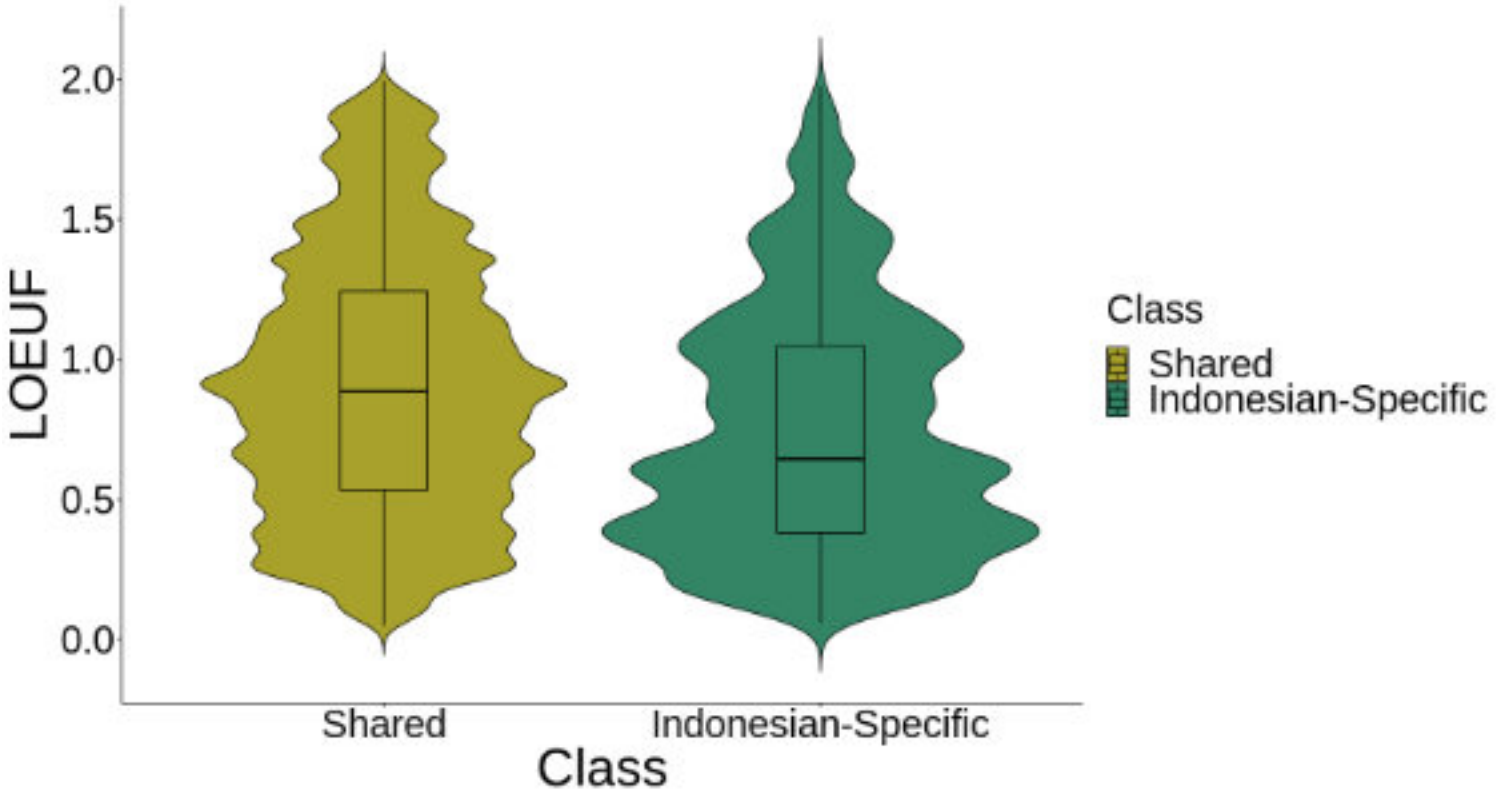
Feature Importance rank: 1

Population-specific eQTLs have different allele frequencies across populations



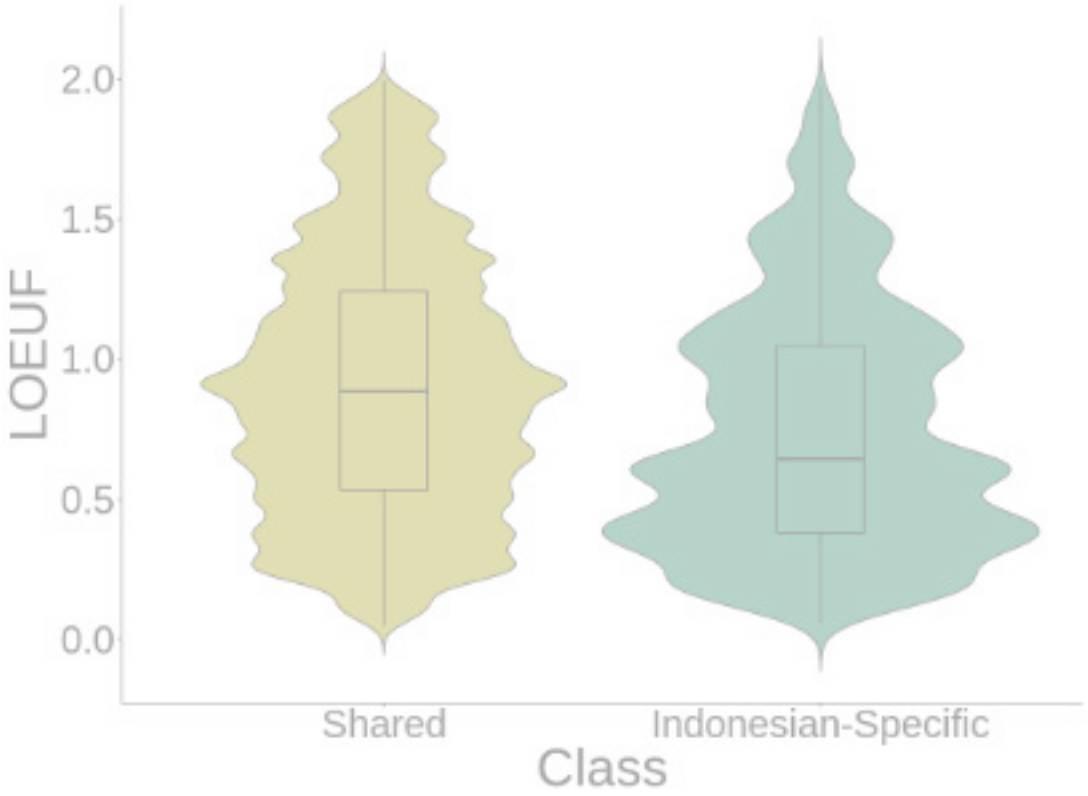
Feature Importance rank: 5

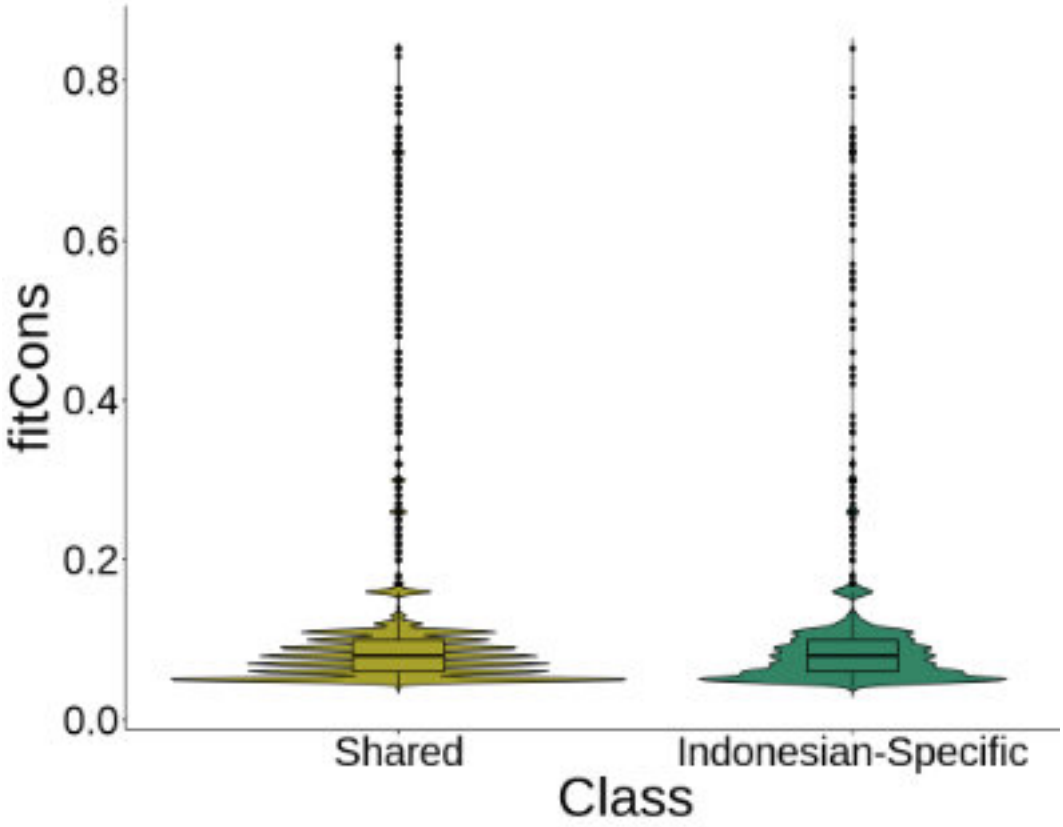
Population-specific eGenes are more highly conserved



Gene Conservation Feature Importance rank: 4

Population-specific eGenes are more highly conserved … but not eSNPs

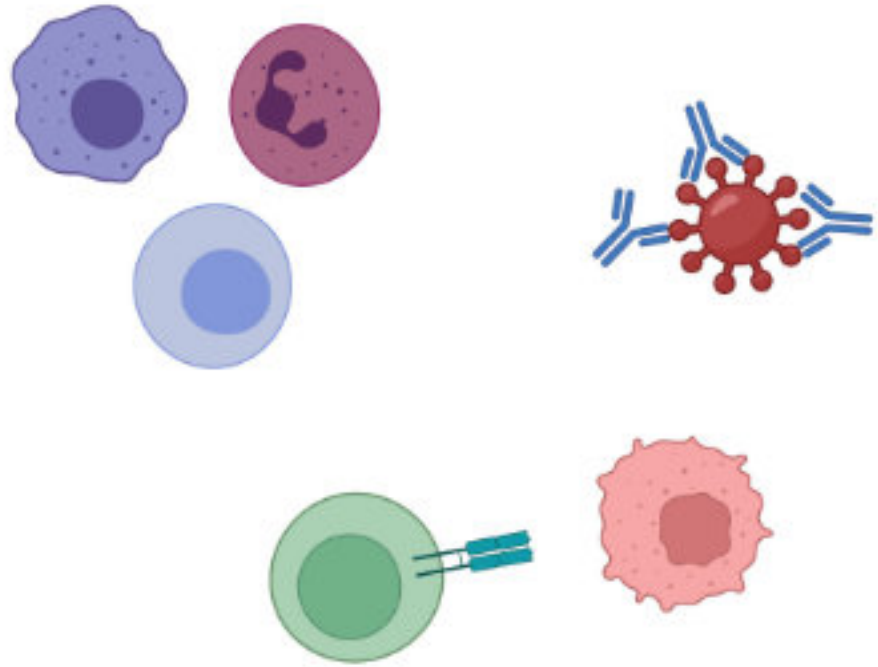




SNP Conservation Feature Importance rank: 132

Indonesian-specific eGenes are more commonly immune related

11.30% (shared) vs 16.10% (Indonesian-specific) are annotated with GO:0002376 (immune system process)



Feature Importance rank: 125

Created with BioRender.com

Conclusions

- Eurocentric biases prevent equitable research translation

Conclusions

- Eurocentric biases prevent equitable research translation
- Some properties are different between shared and specific eQTLs

Conclusions

- Eurocentric biases prevent equitable research translation
- Some properties are different between shared and specific eQTLs
- Machine learning models could improve the transferability of eQTLs from Europeans to underrepresented populations

Thank you!

And an extra special thank you to my fellow lab group members, as well as my contributing authors and supervisors: Christina B Azodi and Irene Gallego Romero!

@ijbeasley