

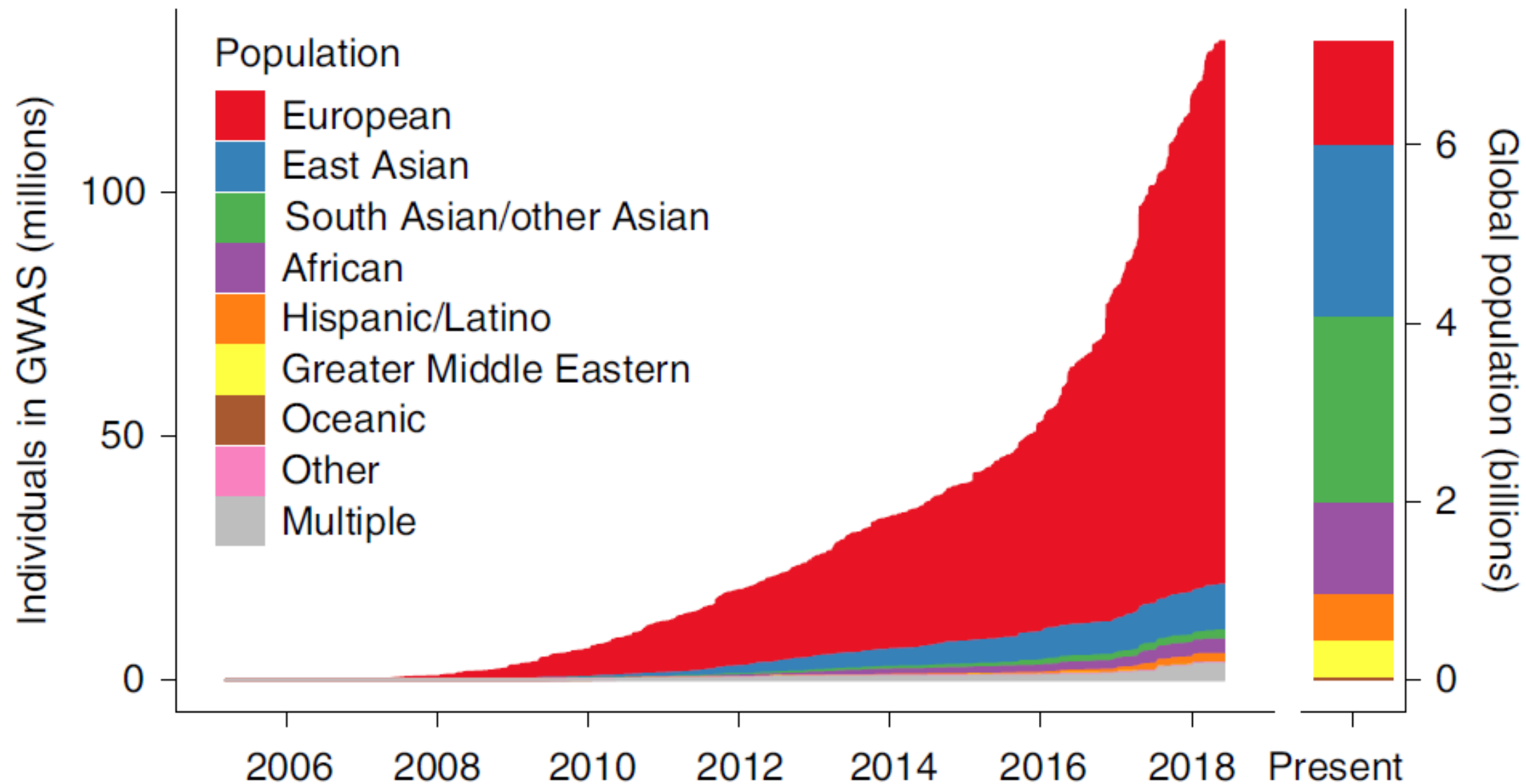
Predicting the cross-population portability of human expression quantitative trait loci (eQTLs)

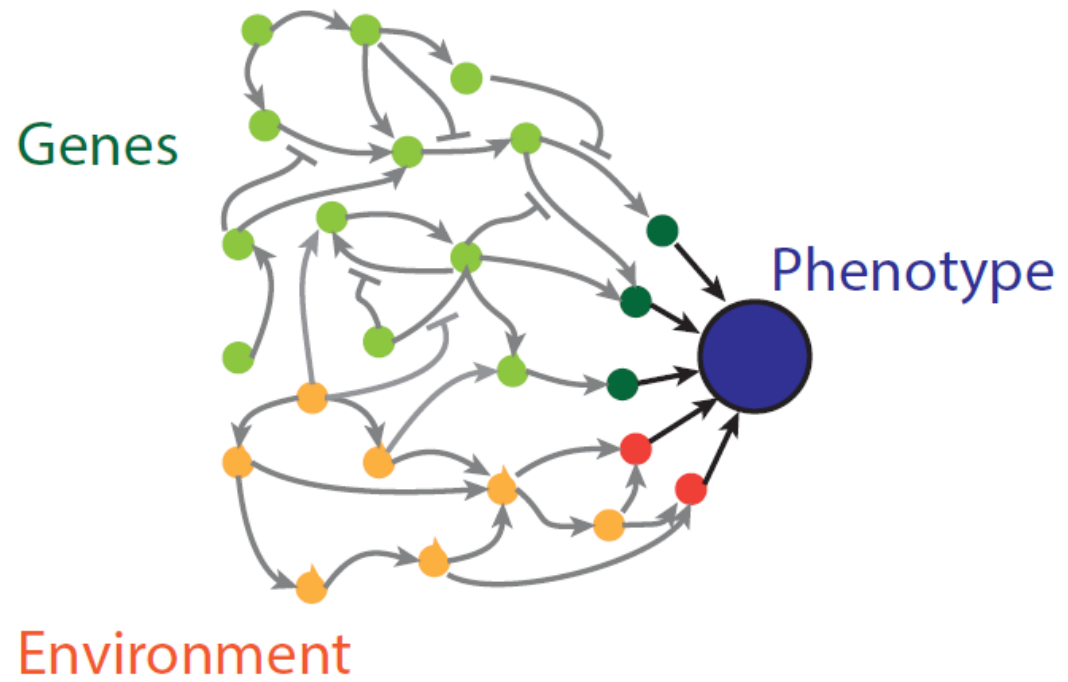
Isobel Beasley, Christina B Azodi and Irene Gallego Romero
The University of Melbourne / St Vincent's Institute of Medical Research

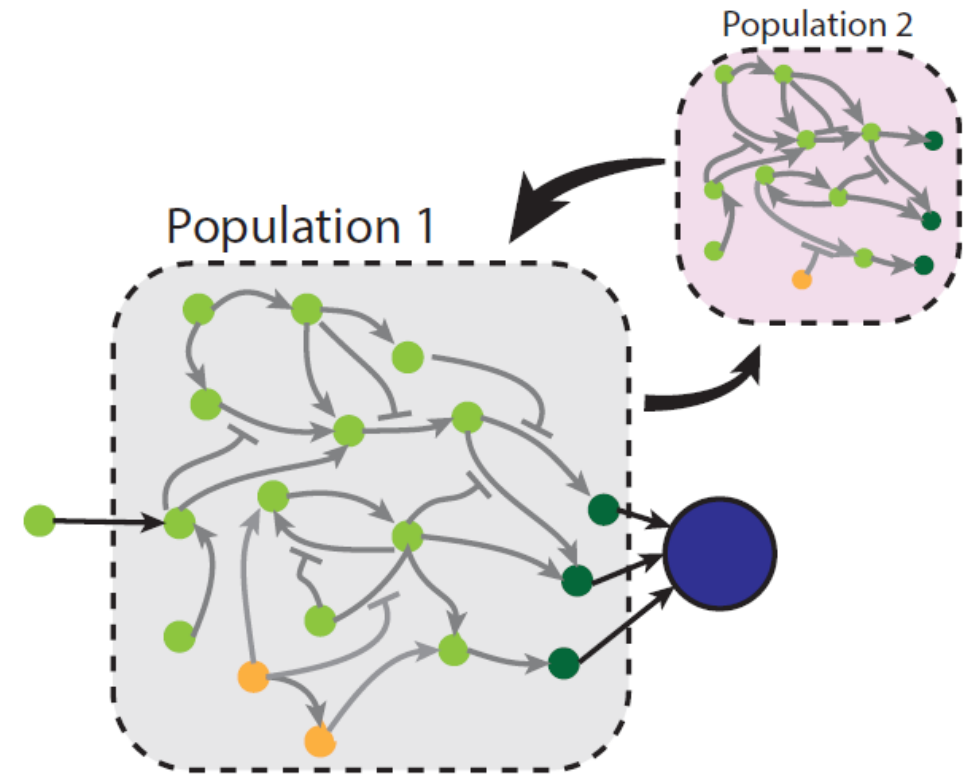
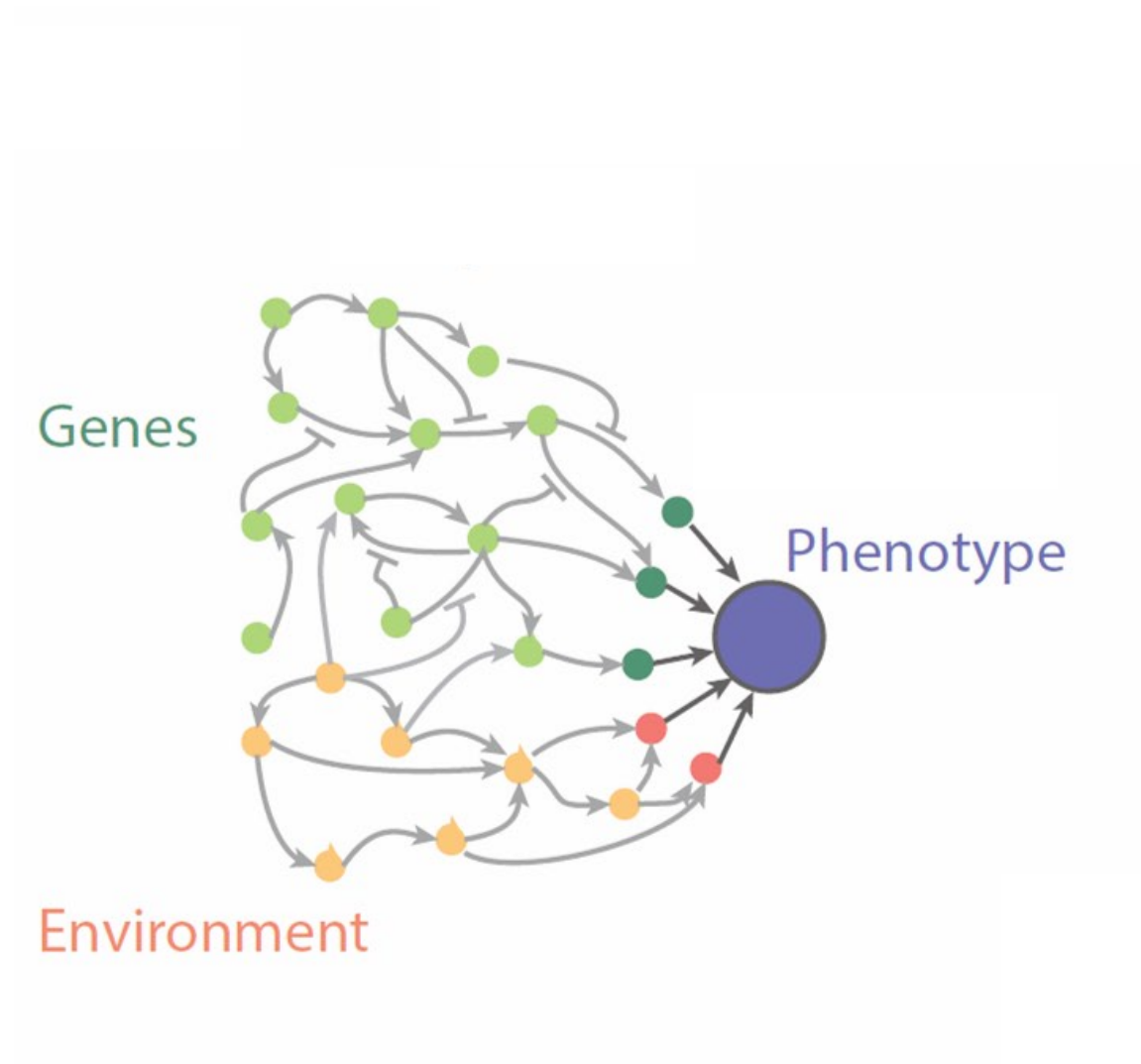


EAPS 2021





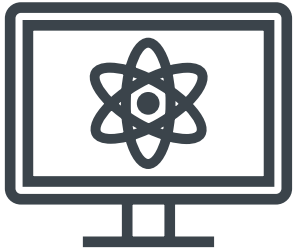




Outline



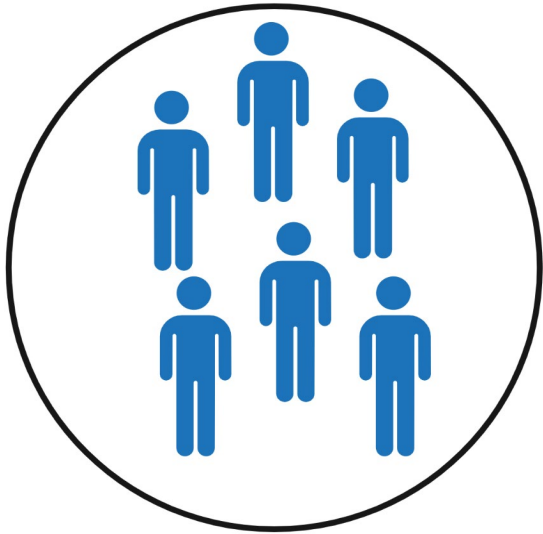
1. eQTLs

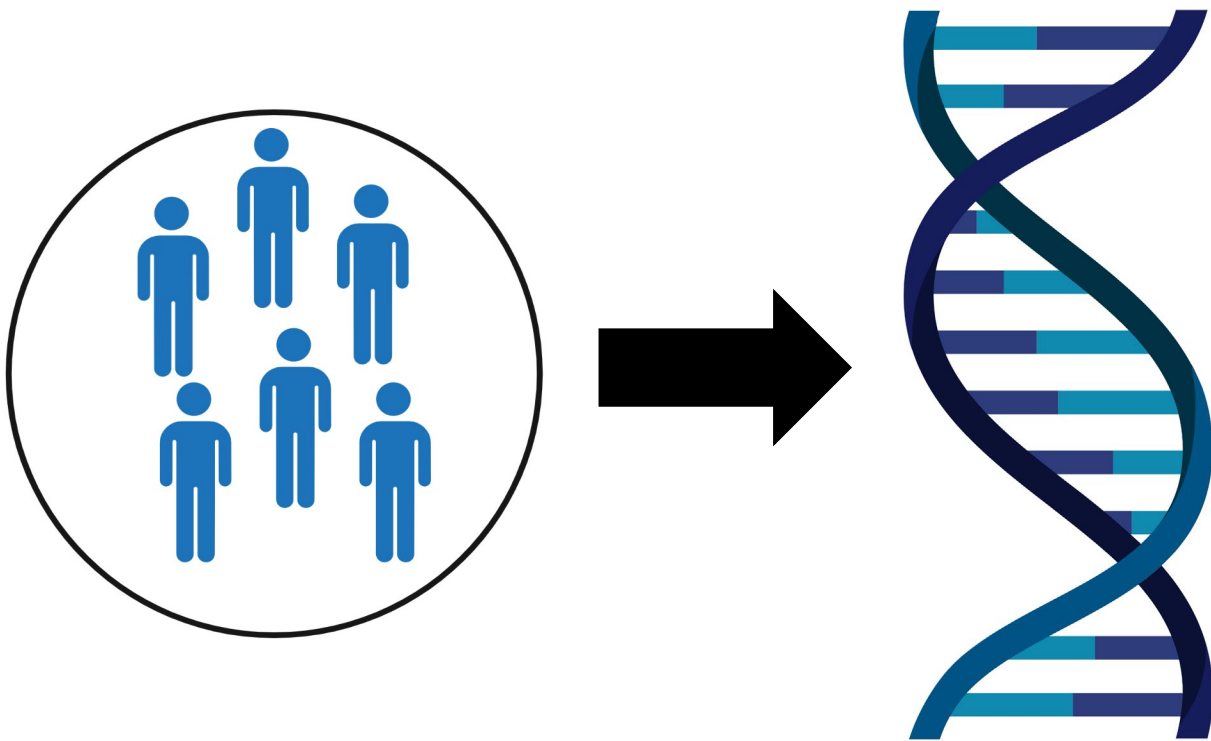


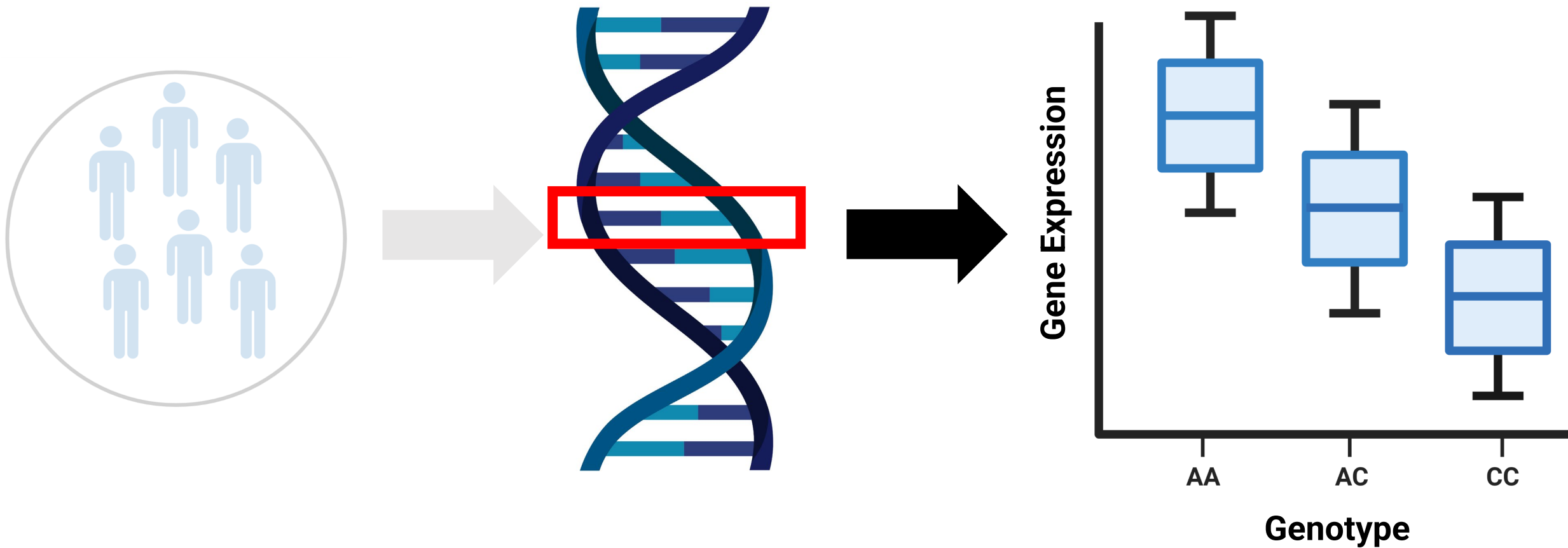
2. Building machine learning models



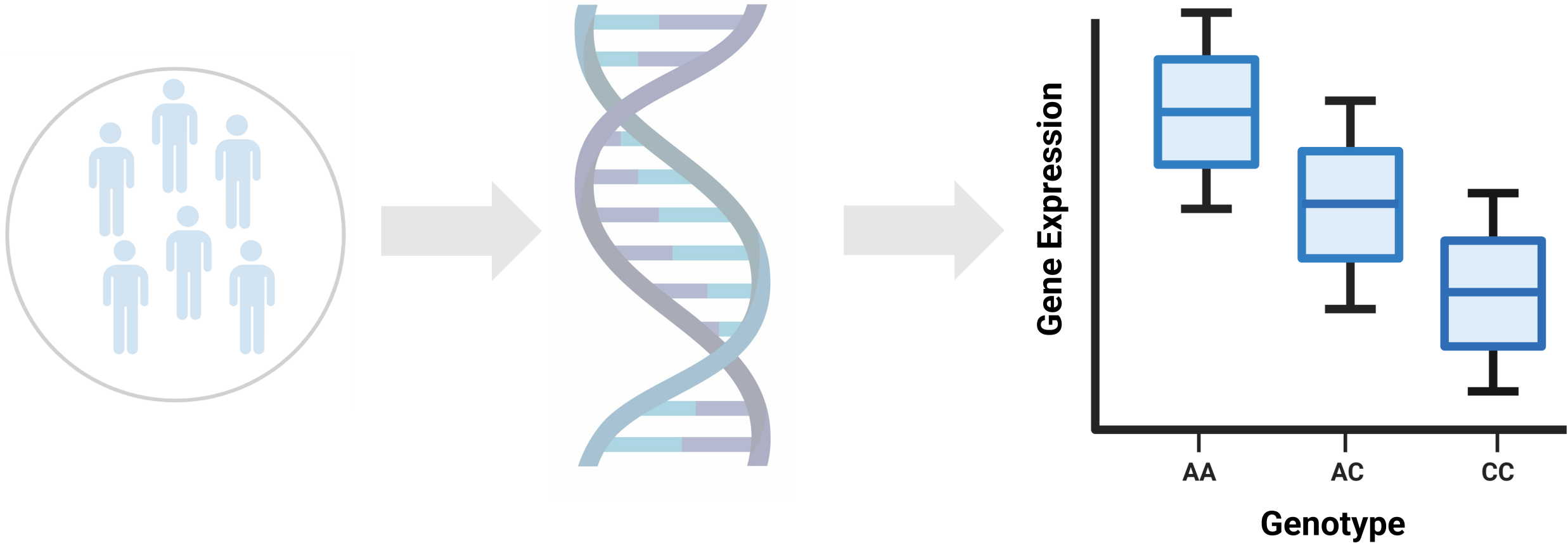
3. Findings (expected and puzzling)





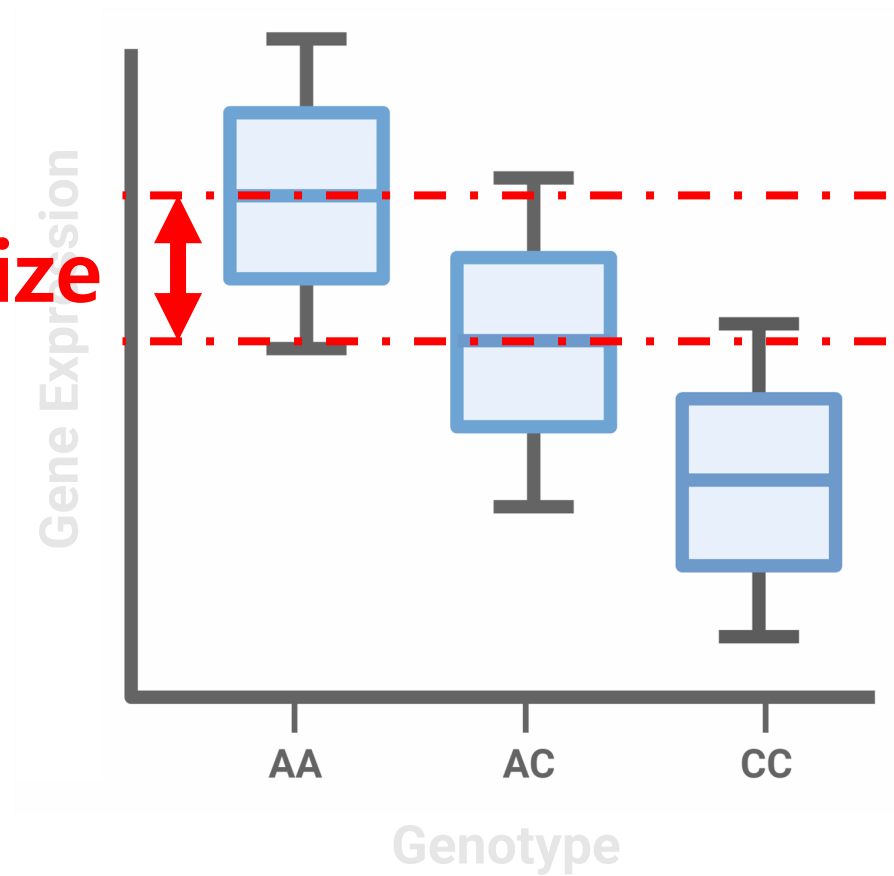


eQTLs (expression Quantitative Trait Loci)





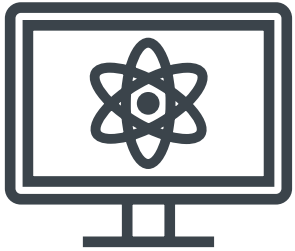
Effect size



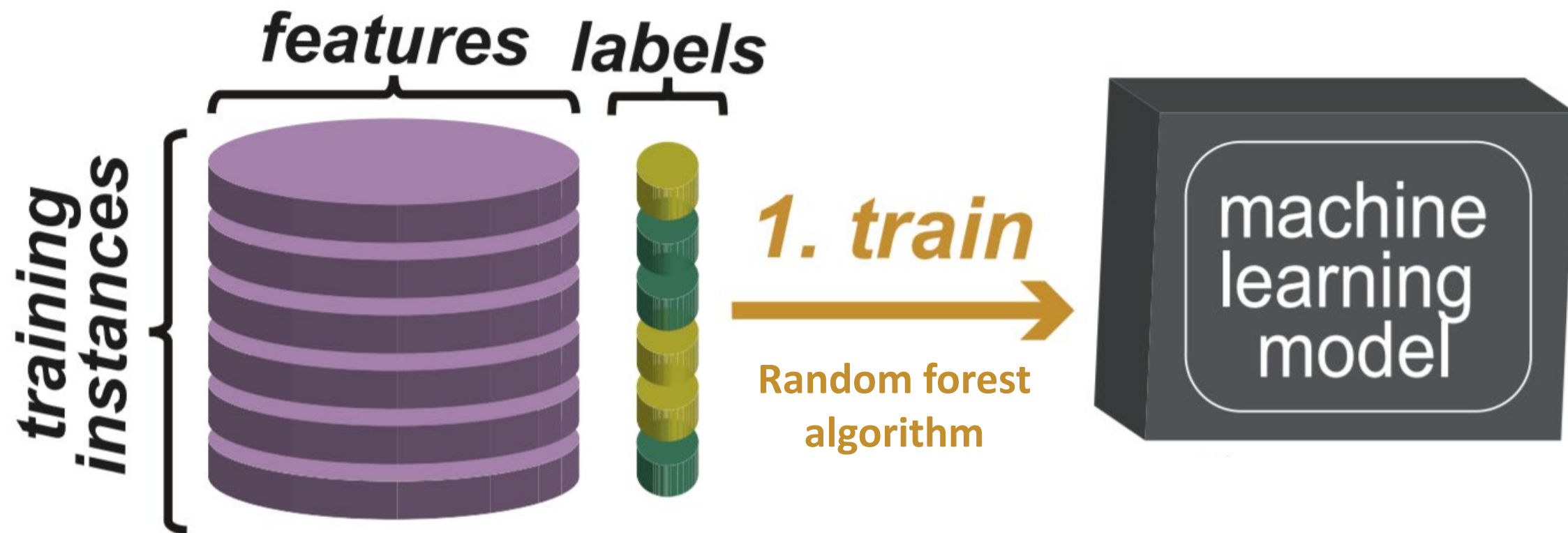
Outline



1. eQTLs



2. Building machine learning models



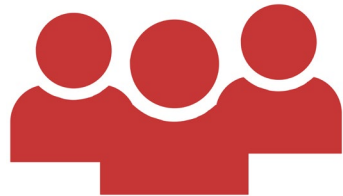
The training data



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



European (n = 379, GTEx. 2020)



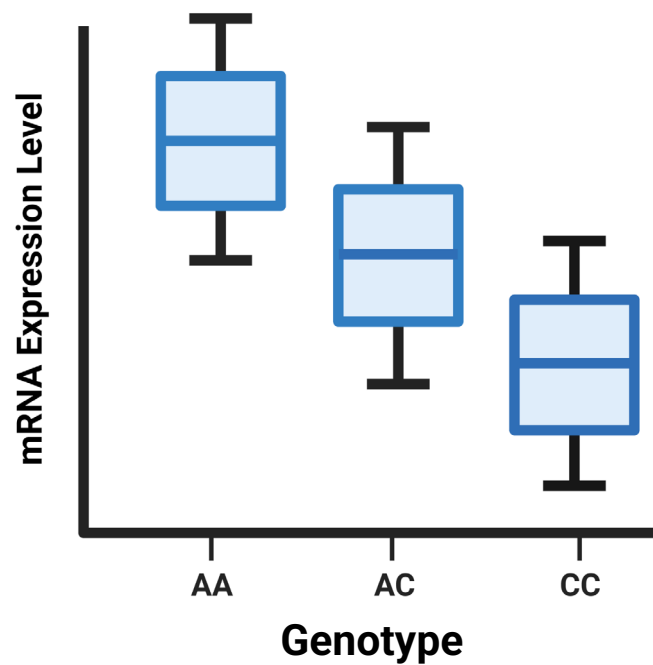
European (n = 195, TwinsUK 2015)



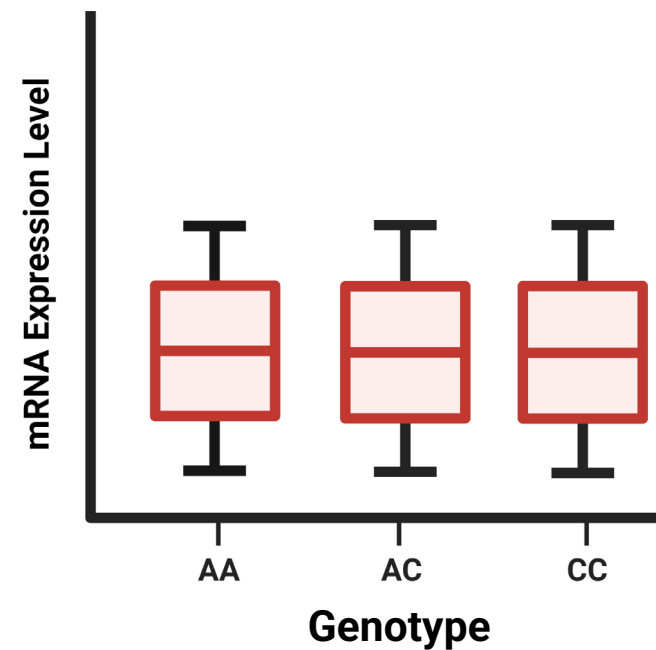
Indonesian (n = 115, Natri et al. 2020)

eQTLs

eQTLs

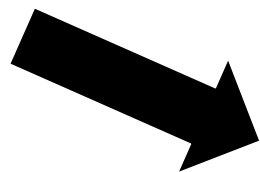


$Ifsr < 0.01$

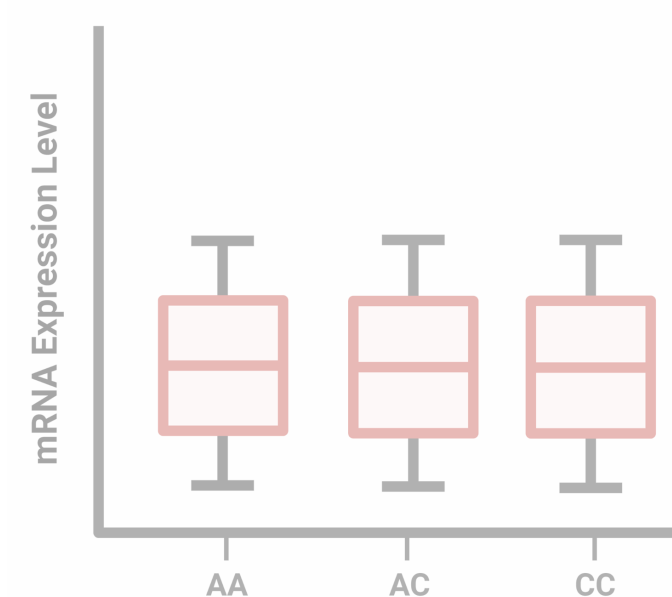
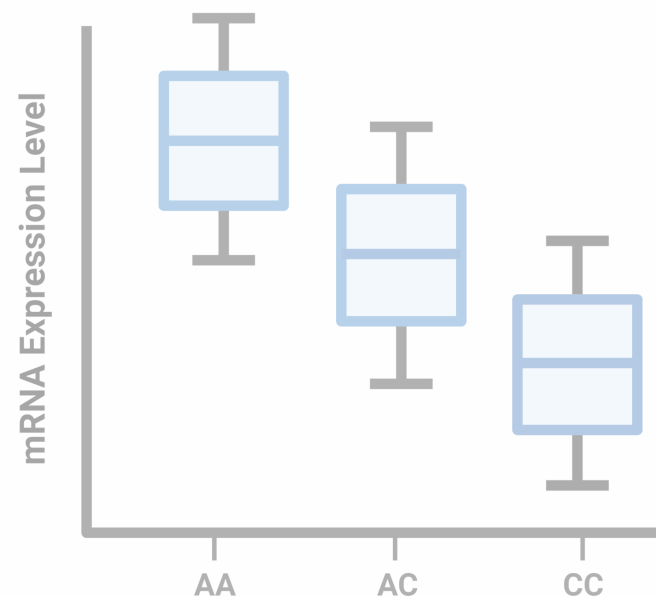


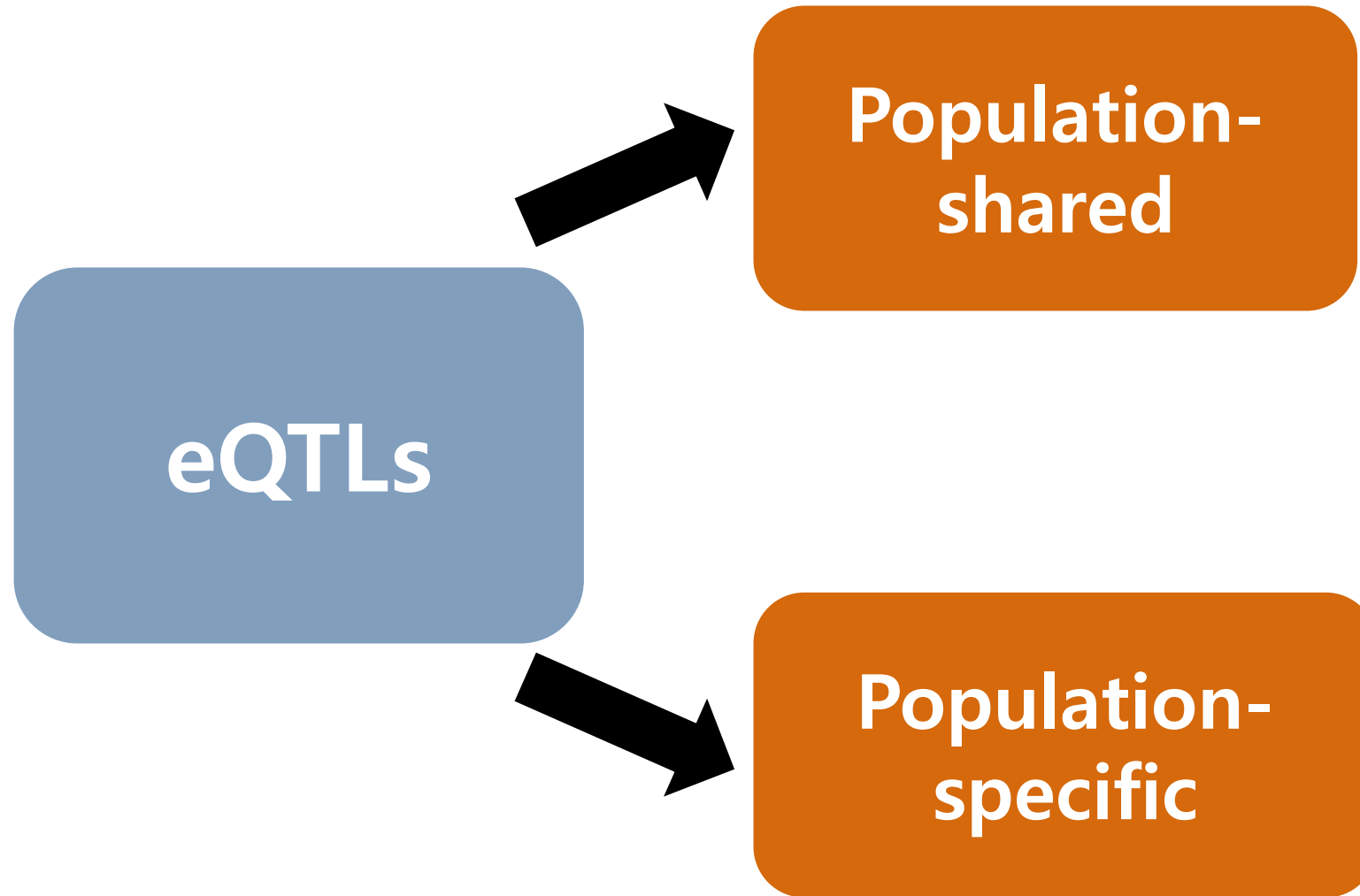
$Ifsr > 0.10$

eQTLs

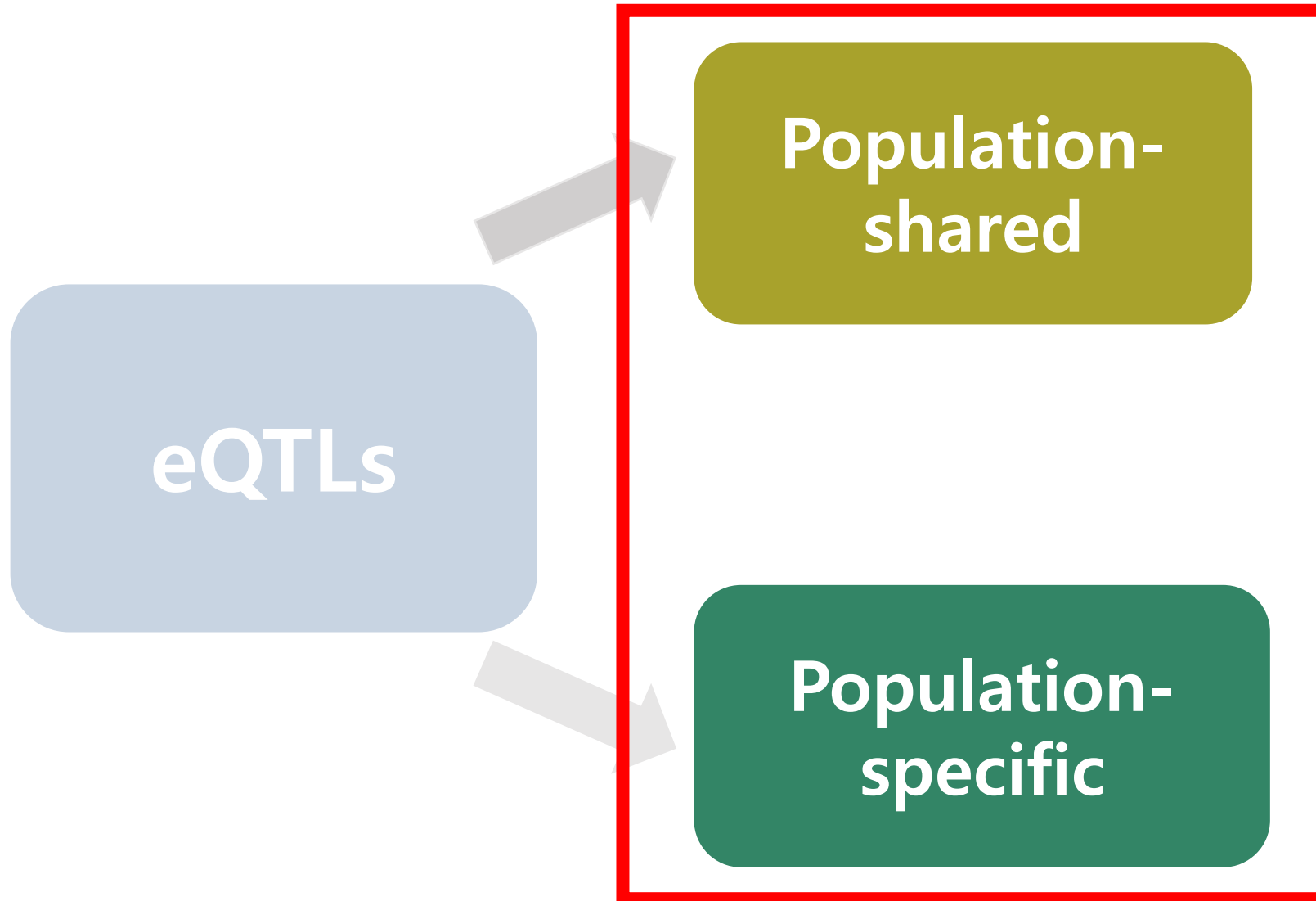


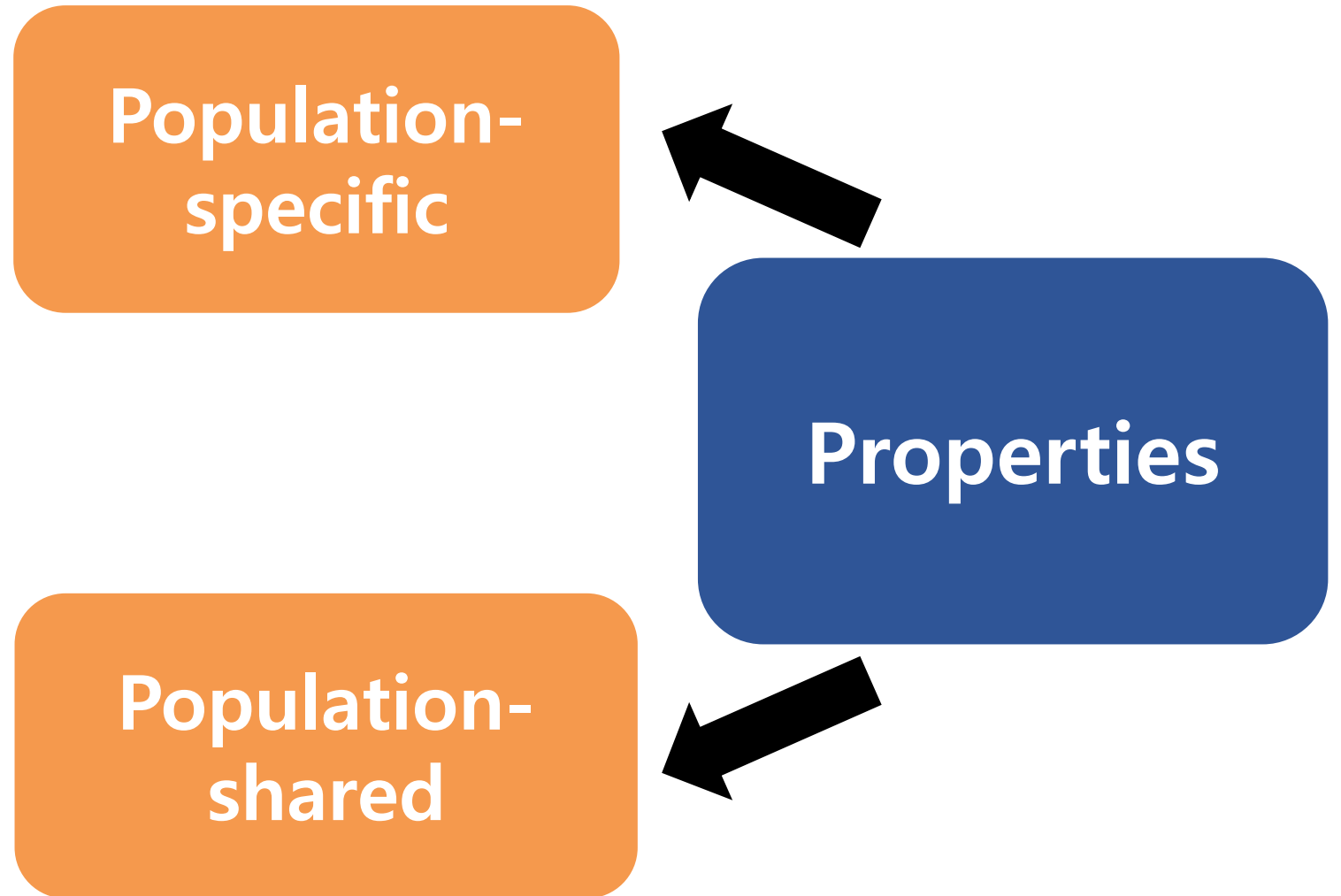
Population-
specific

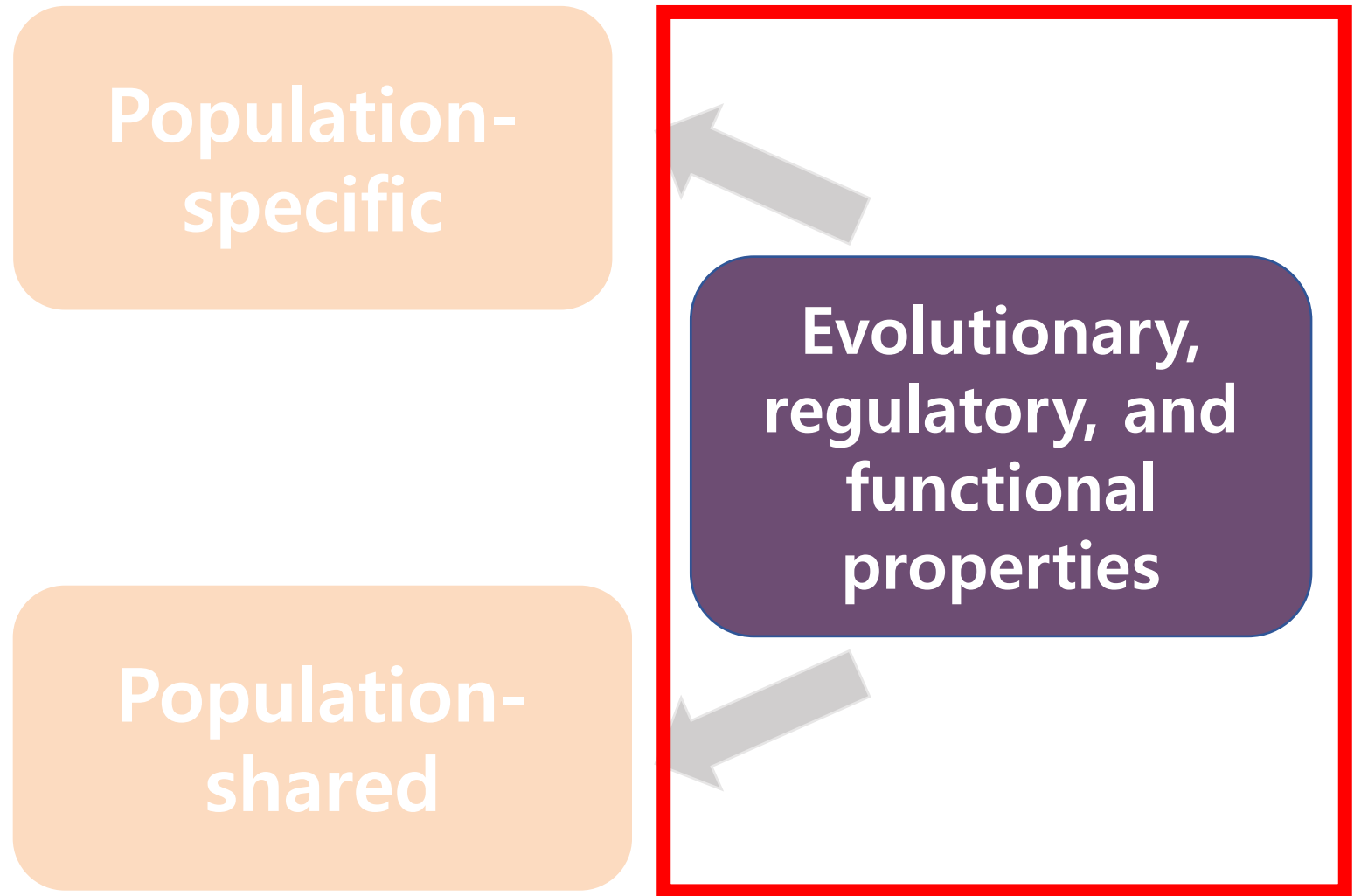




“Labels”







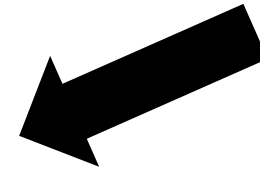
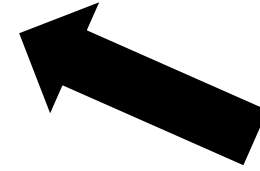
“Features”

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

Population-
specific

Population-
shared

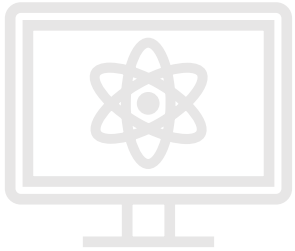
Evolutionary,
regulatory, and
functional
properties



Outline



1. eQTLs

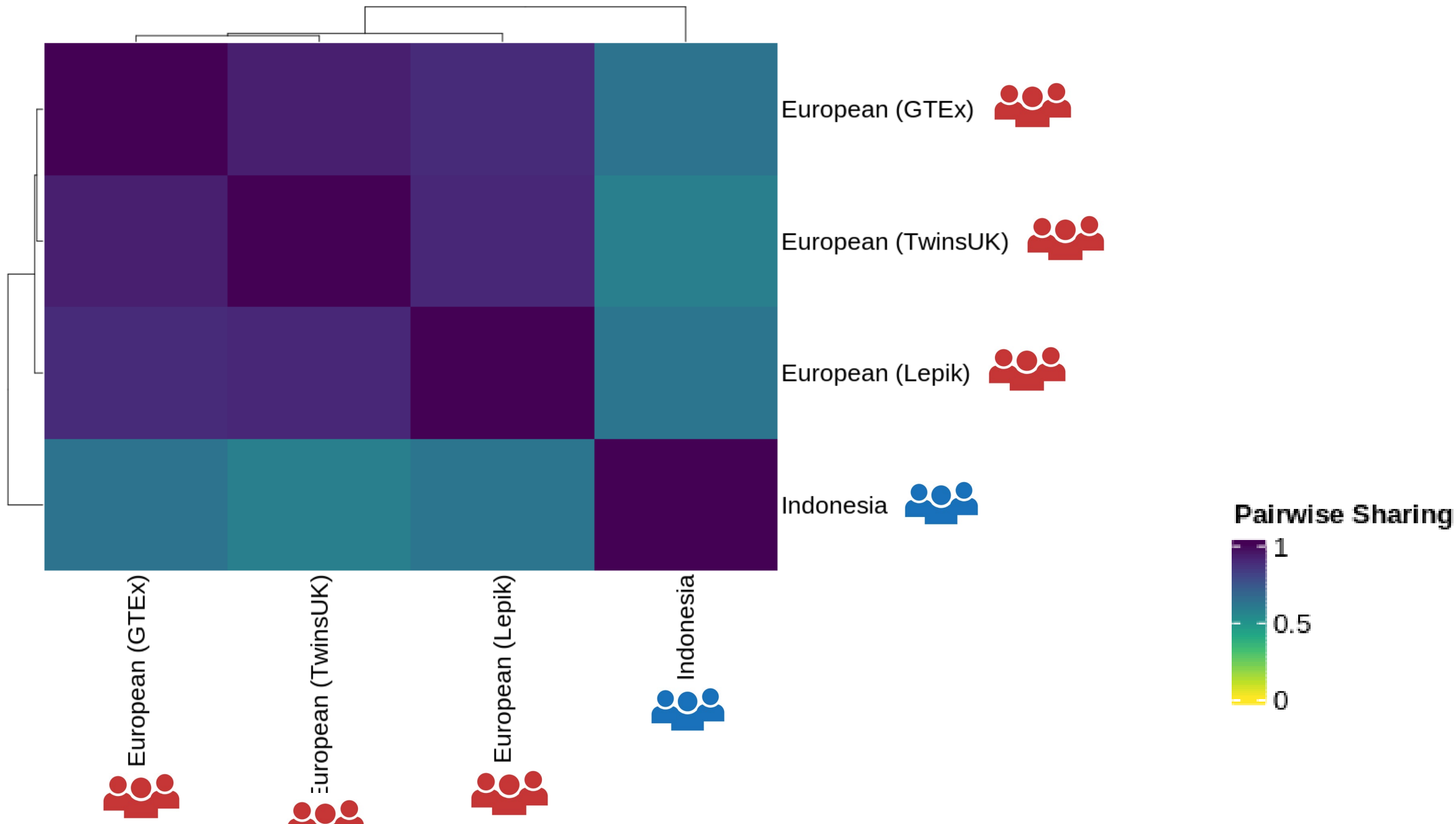


2. Building machine learning models

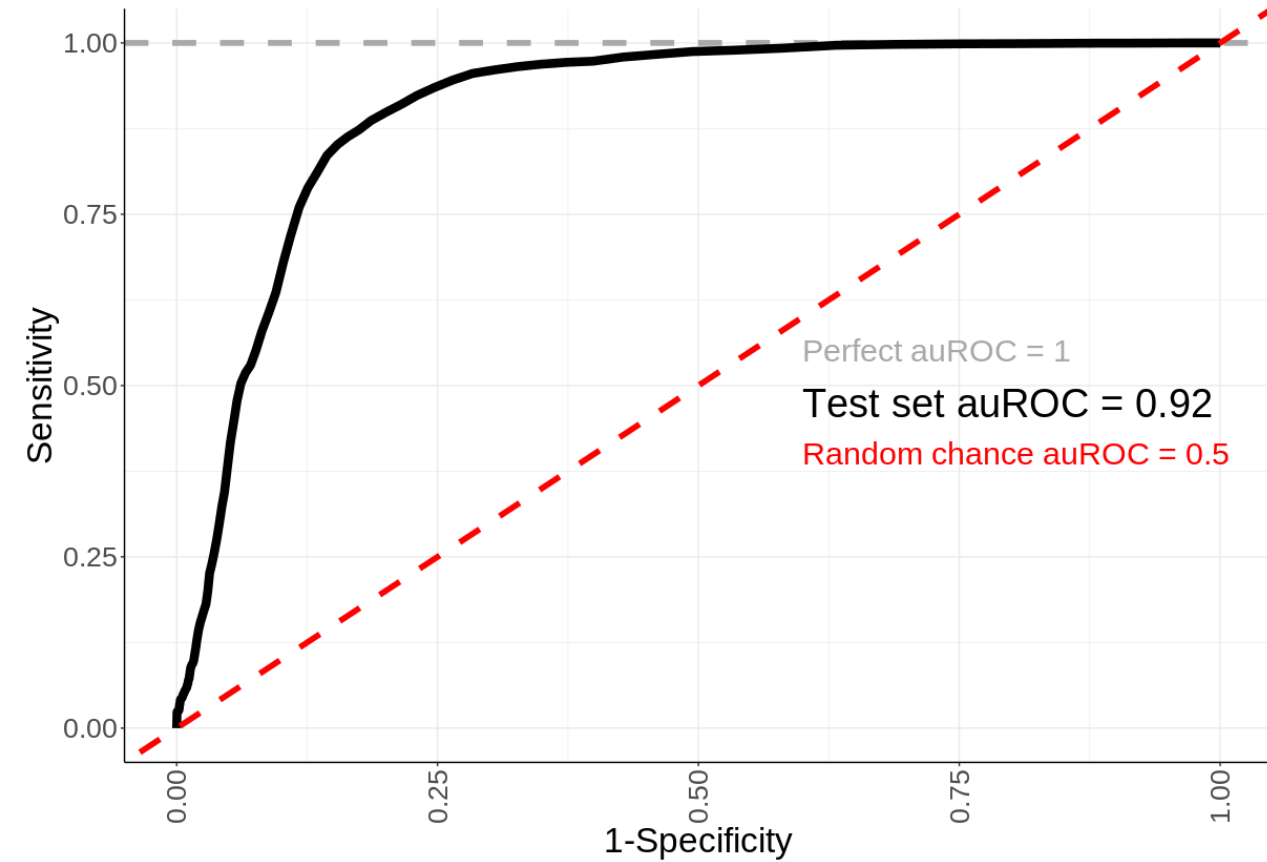


3. Findings (expected and puzzling)

Pairwise Sharing of Effect Size (within a factor of 0.5)

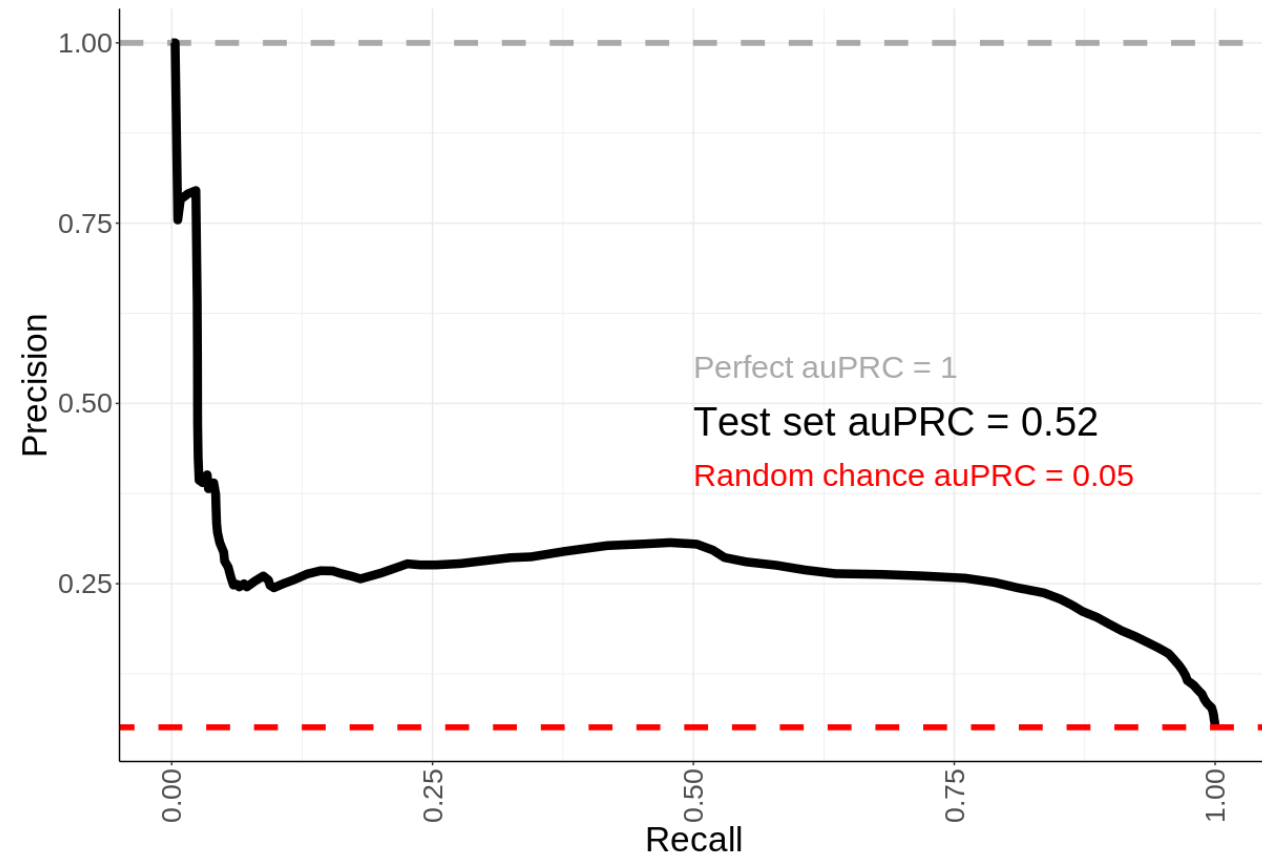
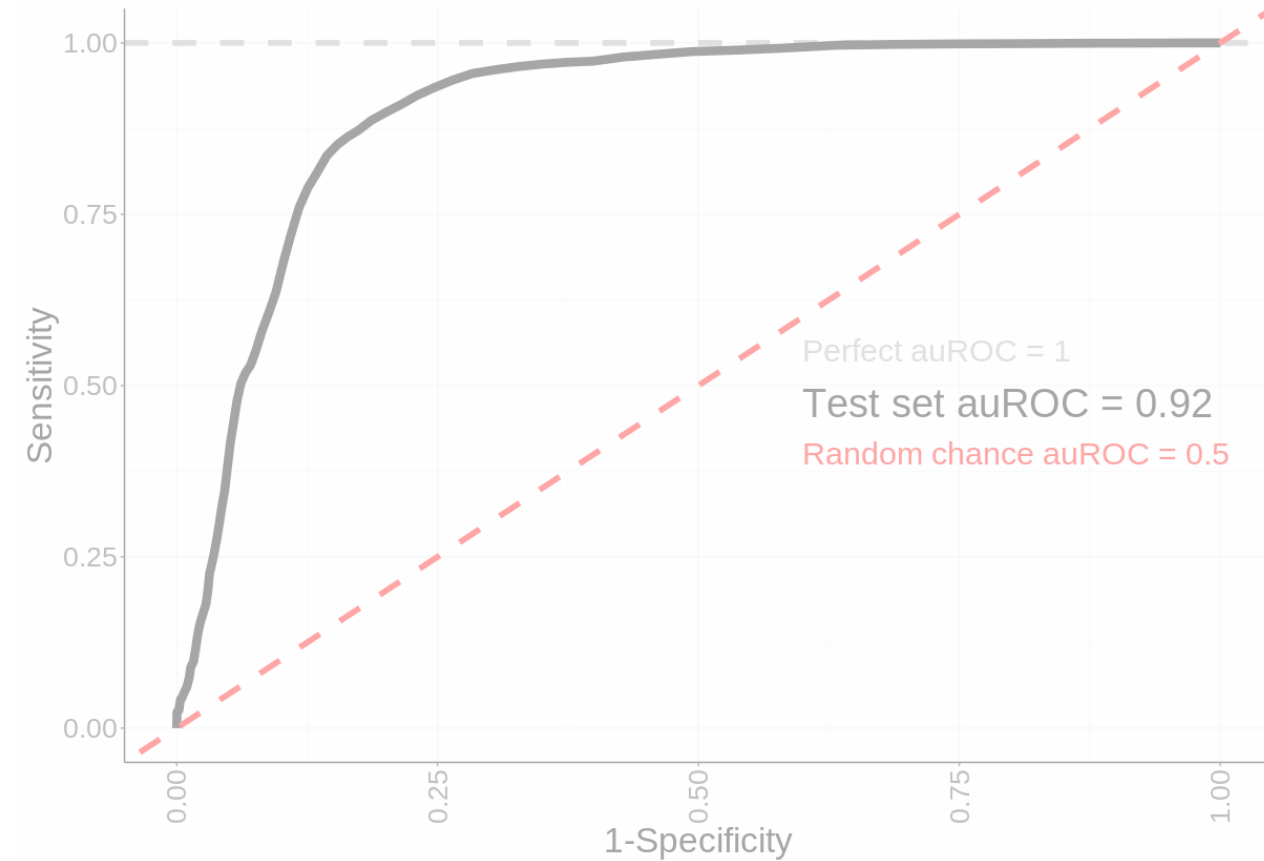


Prediction Performance



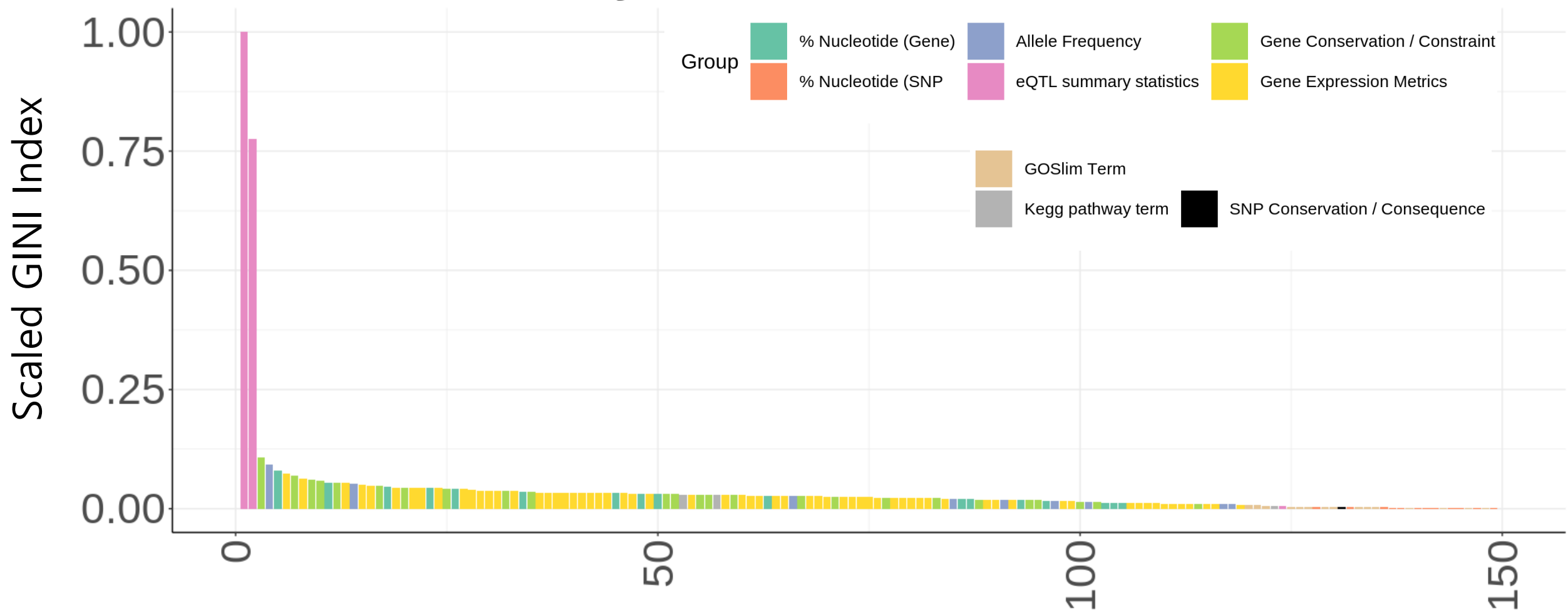
Test set: Chromosomes 8,16

Prediction Performance

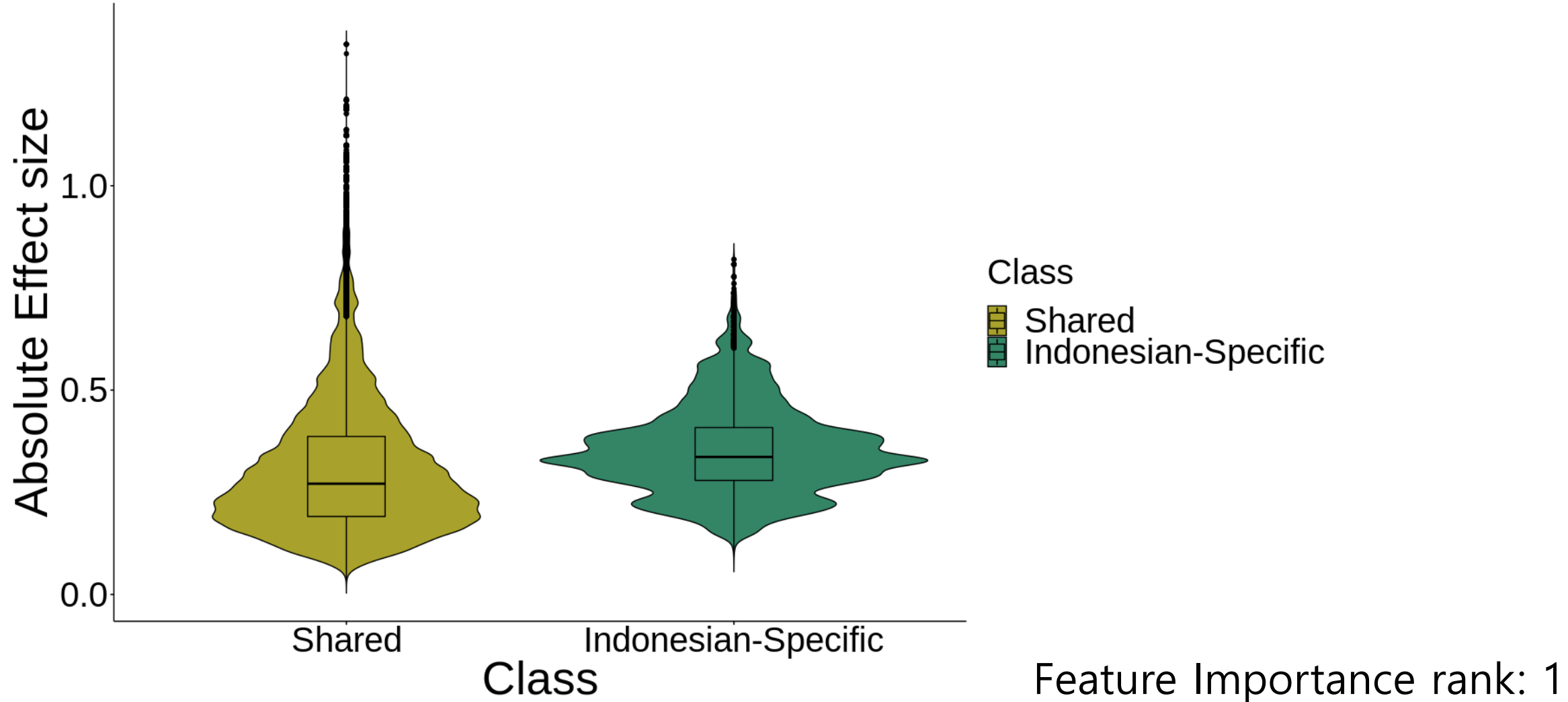


Test set: Chromosomes 8,16

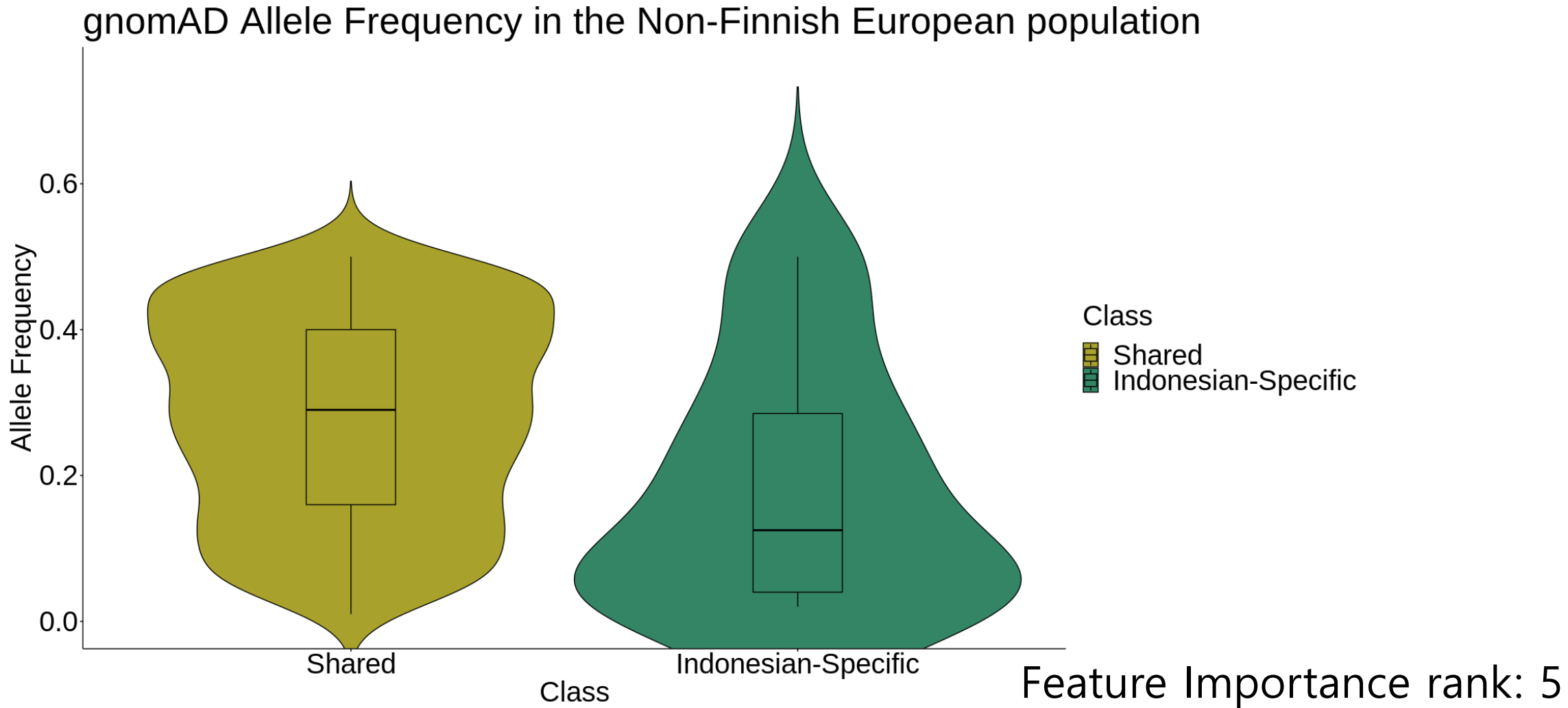
Features ranked by scaled mean importance



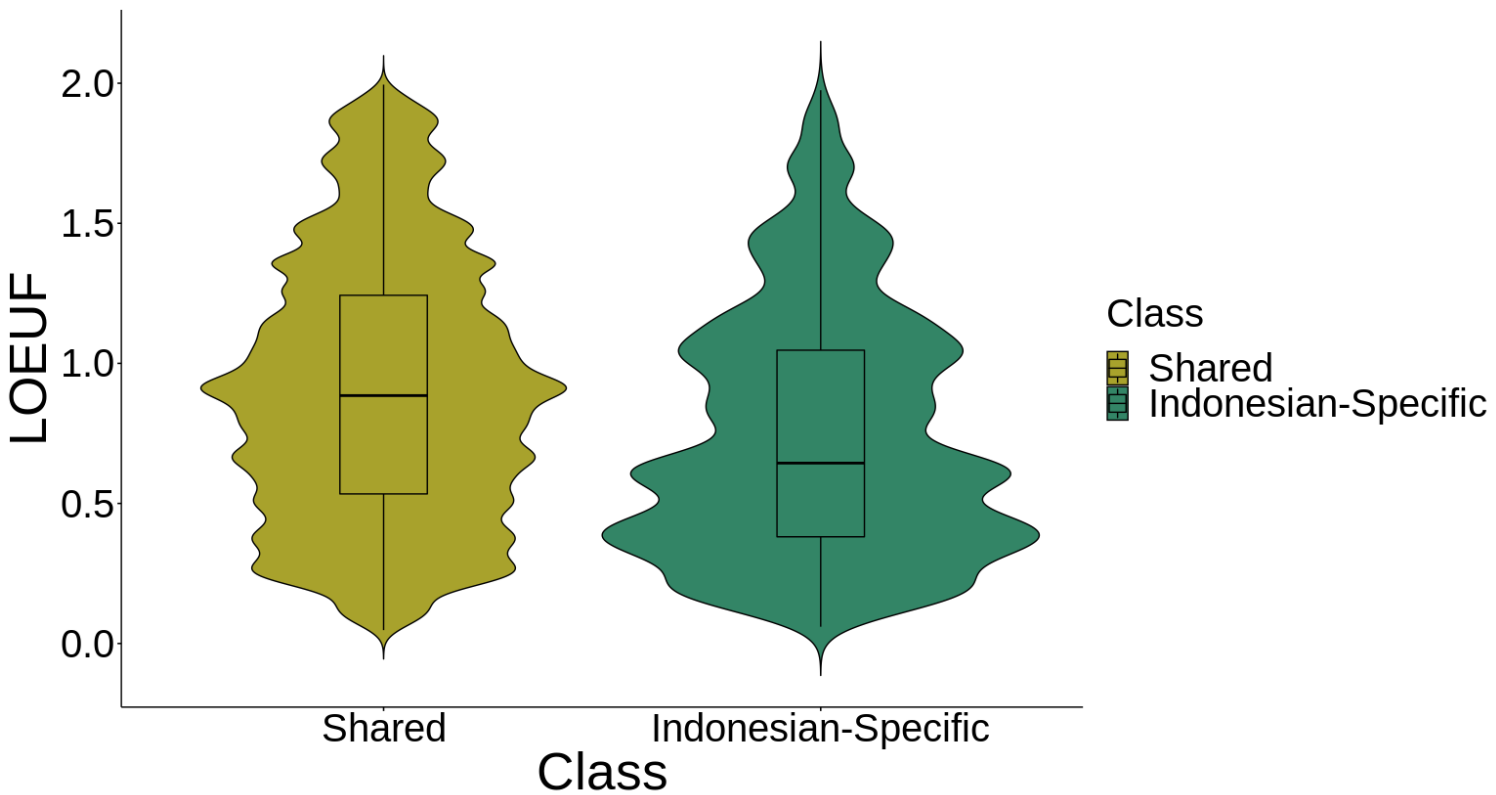
Population-specific eQTLs have higher effect sizes in their discovery population



Population-specific eQTLs have different allele frequencies across populations

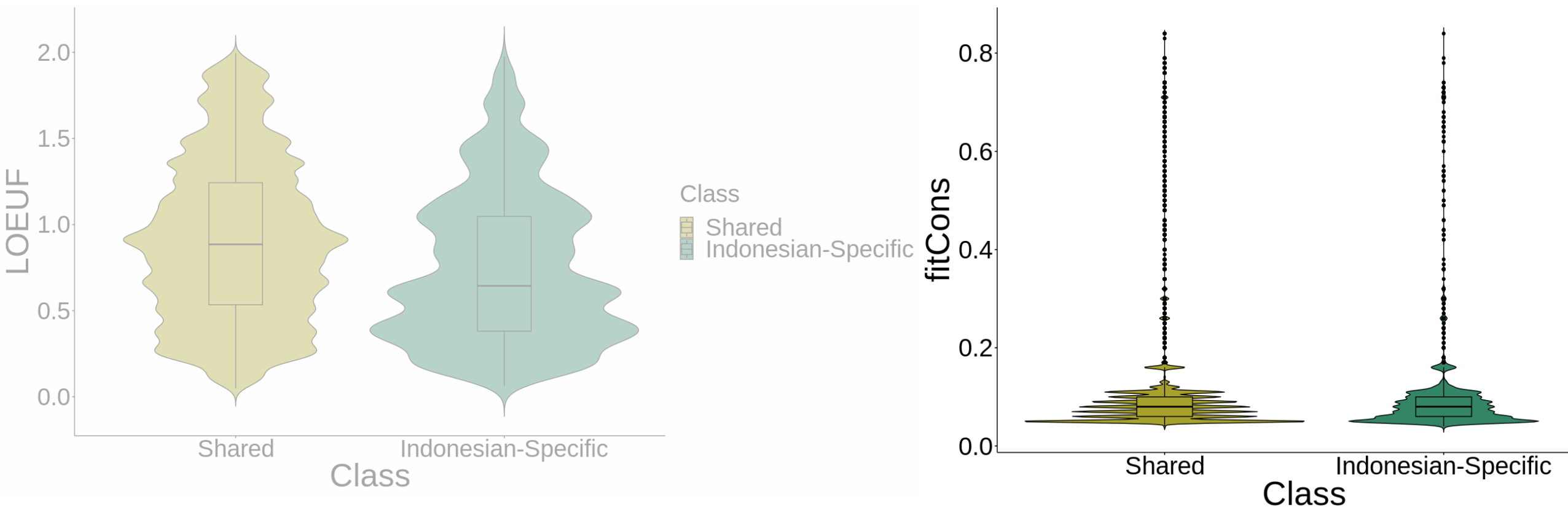


Population-specific eGenes are more highly conserved



Gene Conservation Feature Importance rank: 4

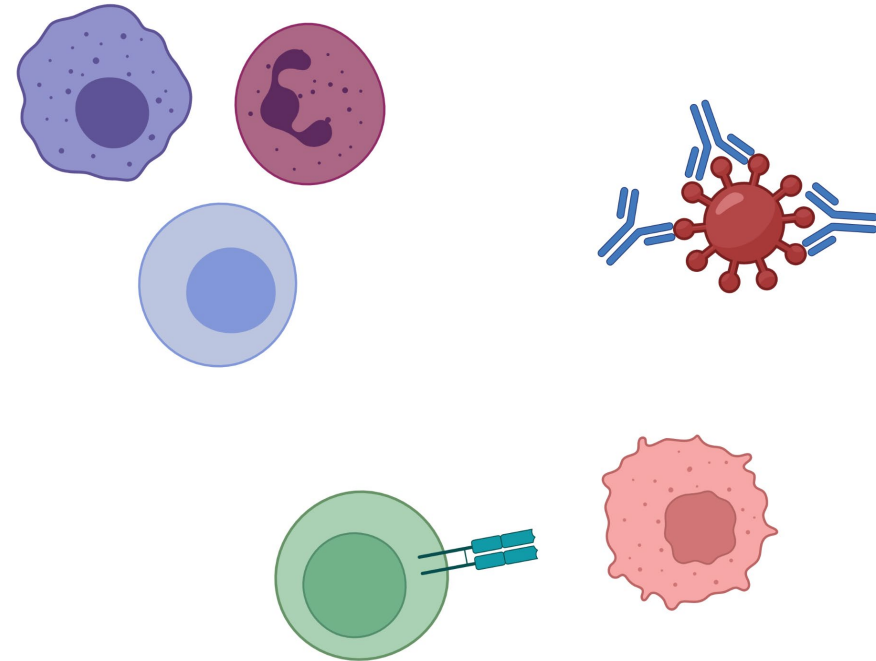
Population-specific eGenes are more highly conserved ... **but not eSNPs**



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](https://www.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

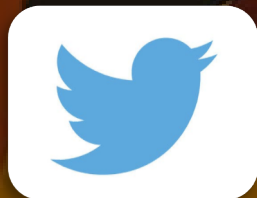
Conclusions

- 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