

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

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

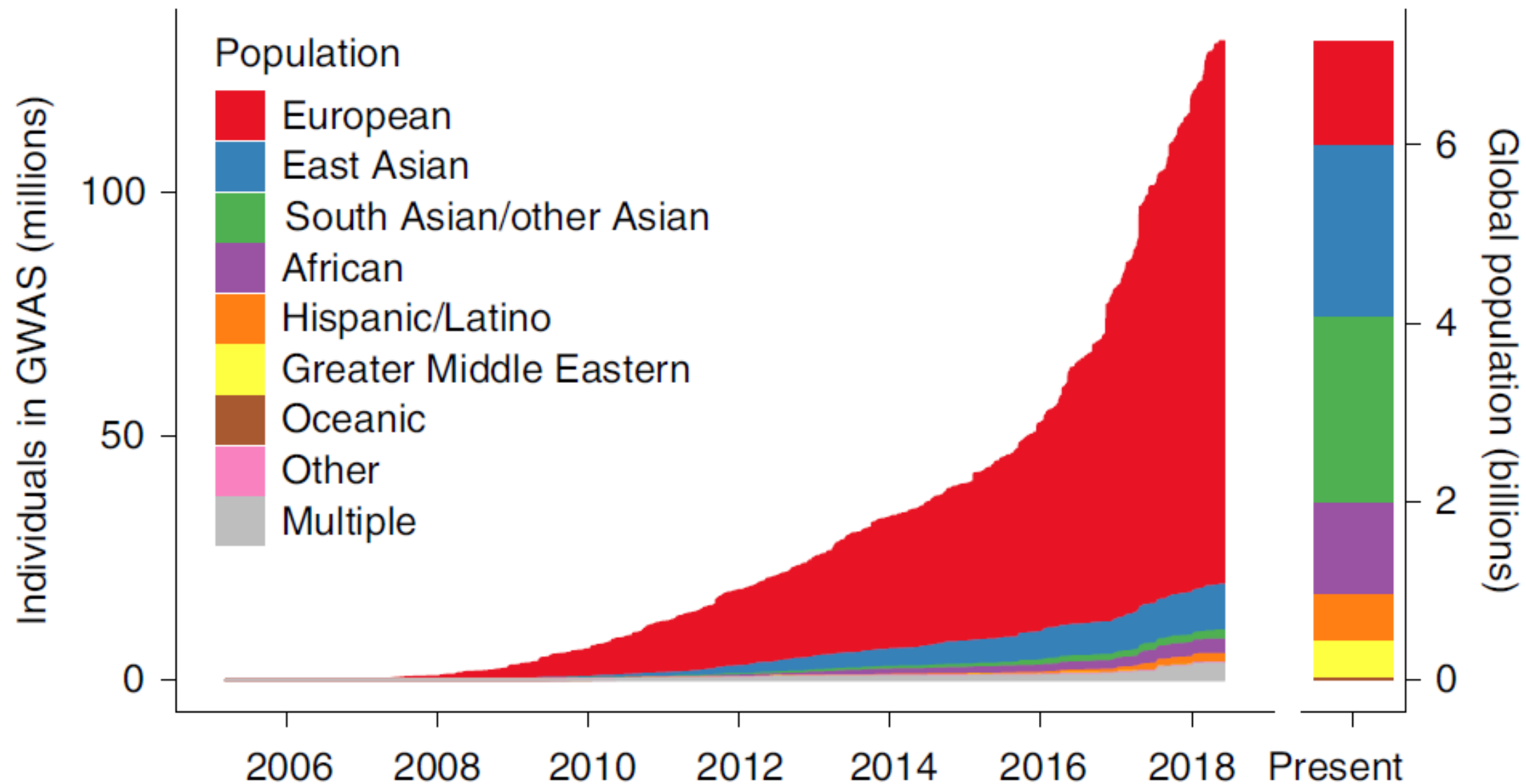


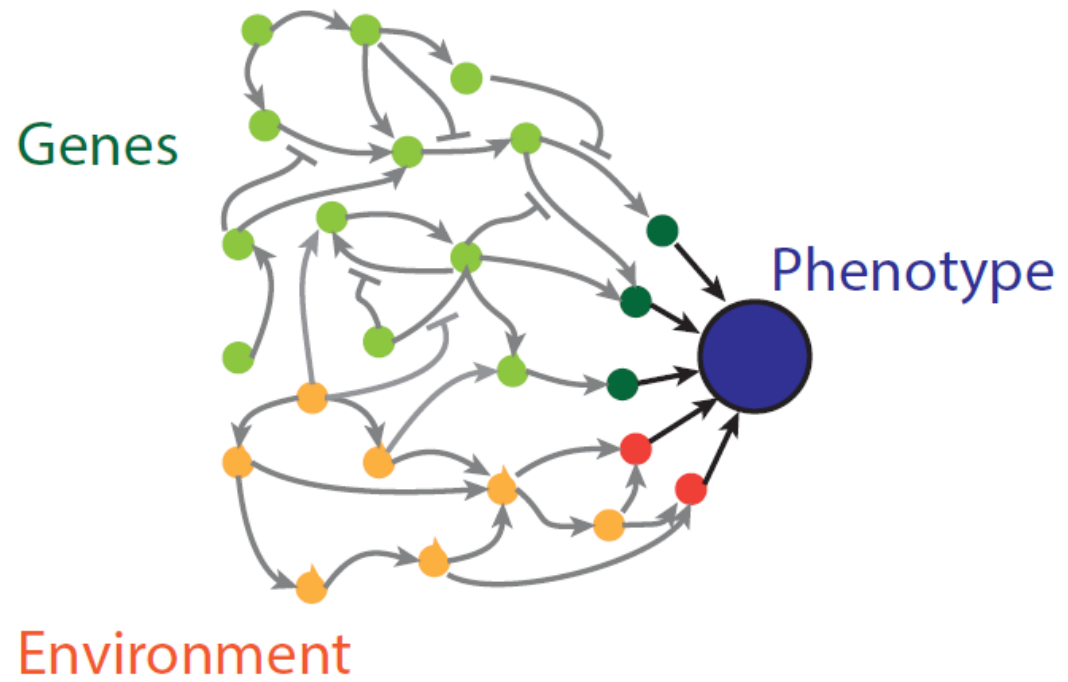
#ABACBS2021

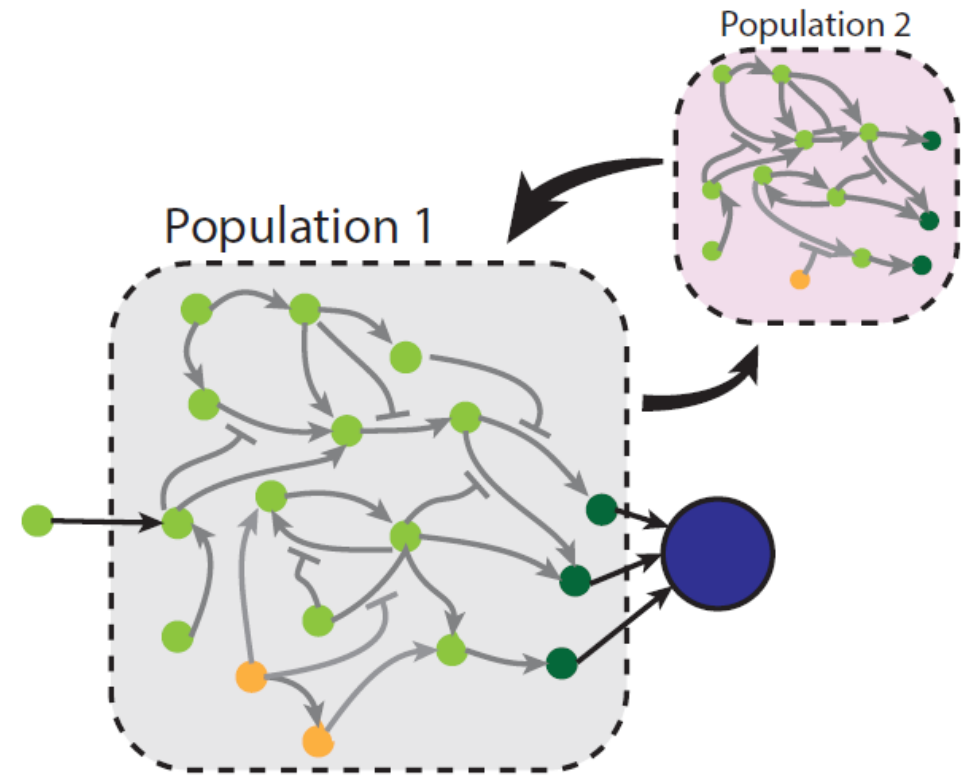
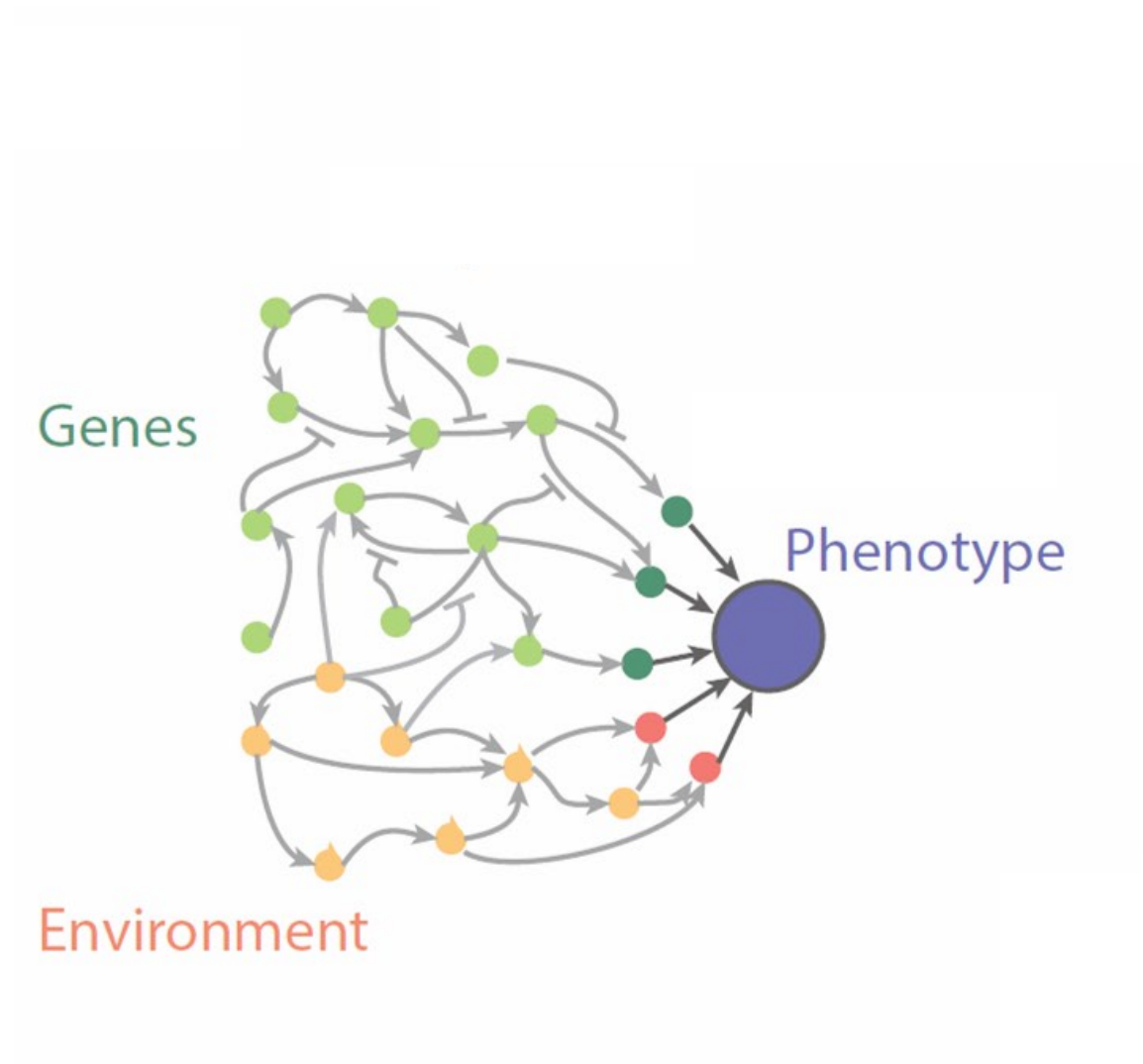


@ijbeasley





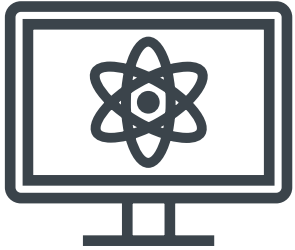




# Outline



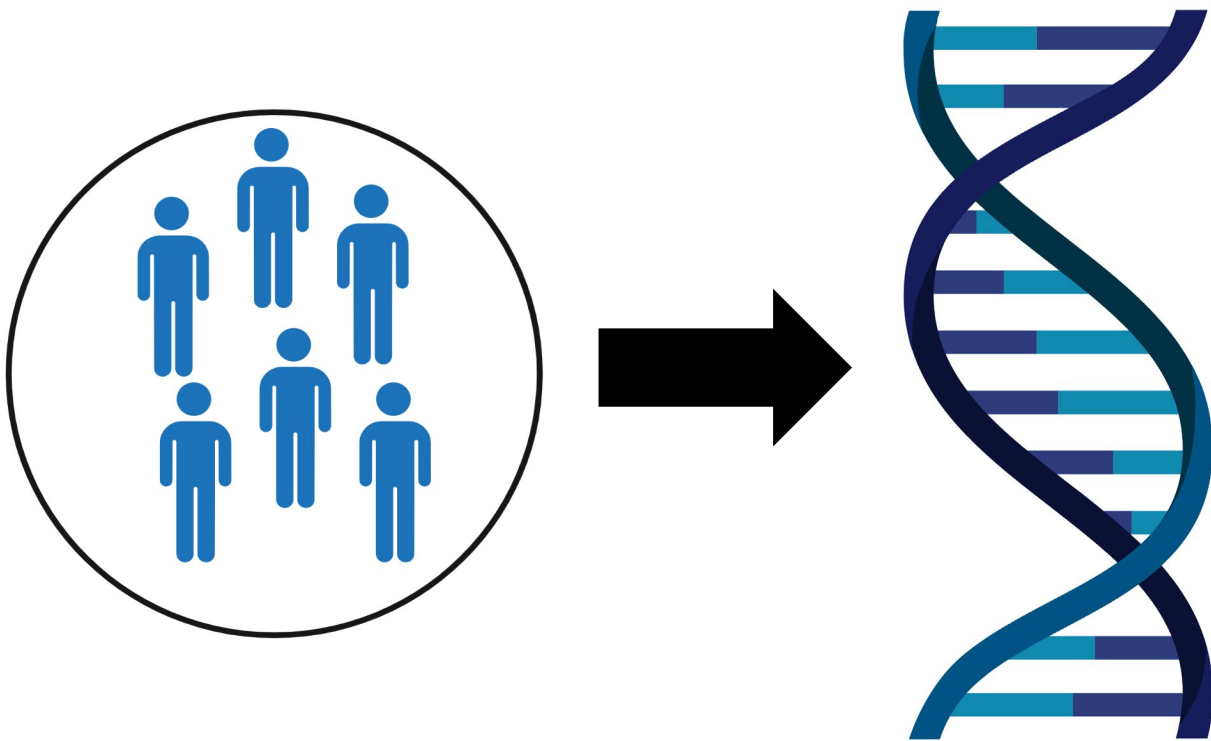
**1. eQTLs**



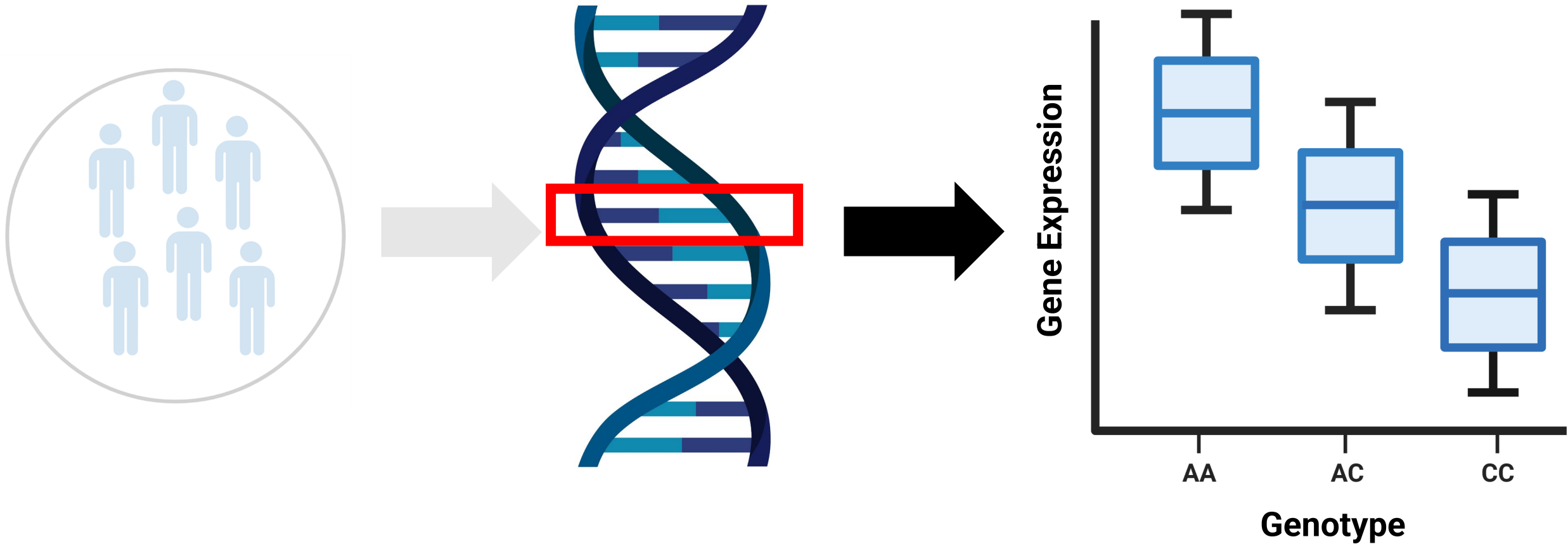
**2. Building machine learning models**



**3. Findings (expected and puzzling)**

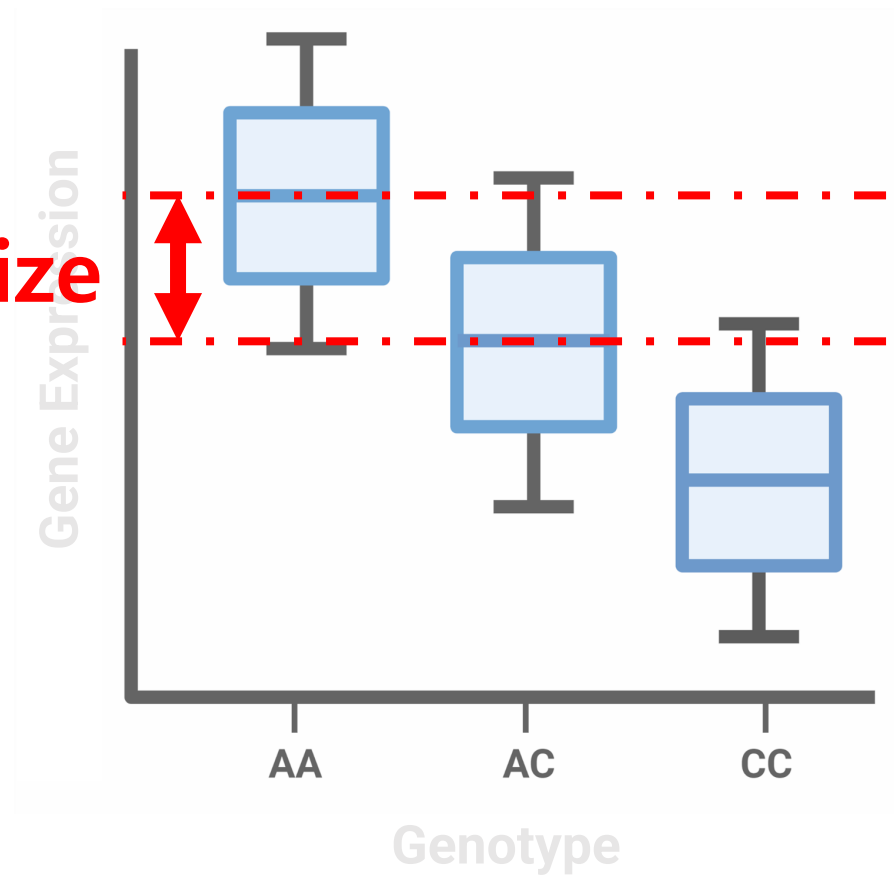


# eQTLs (expression Quantitative Trait Loci)





**Effect size**

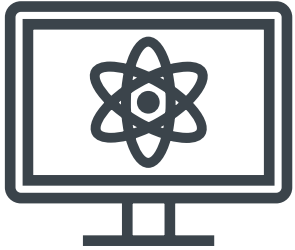




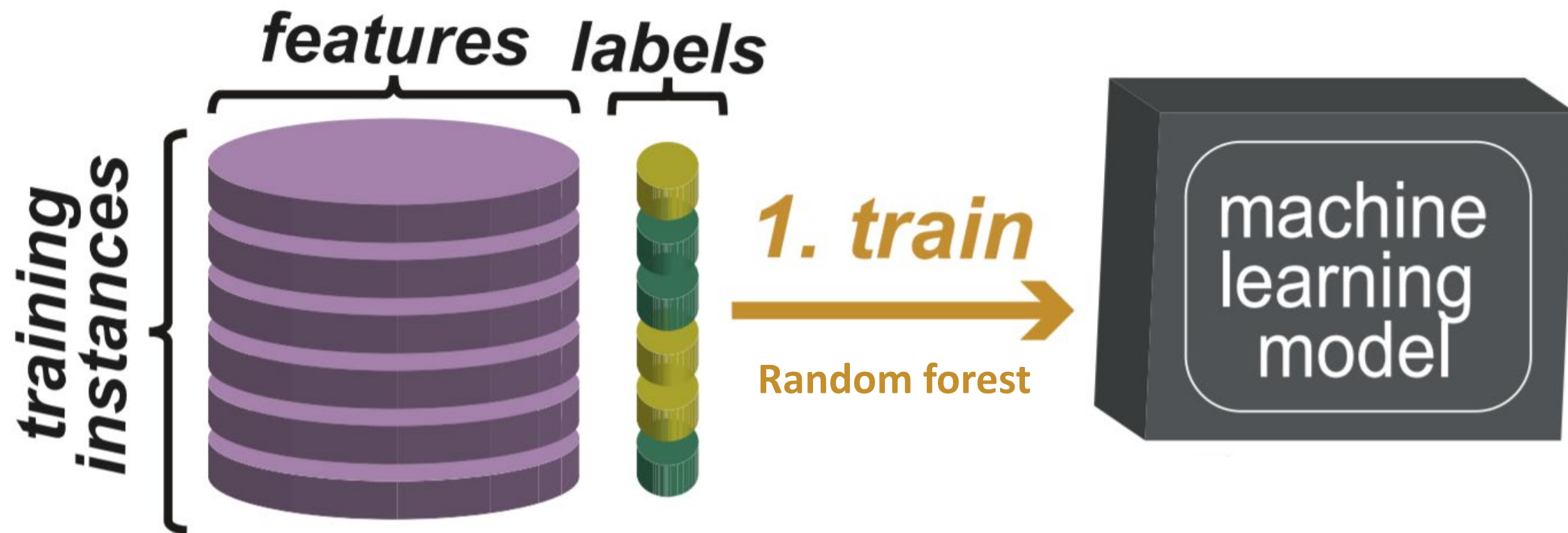
# Outline



1. eQTLs



**2. Building machine learning models**



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

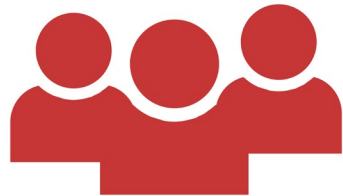
# The training data



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



**European (n = 379, GTEx. 2020)**



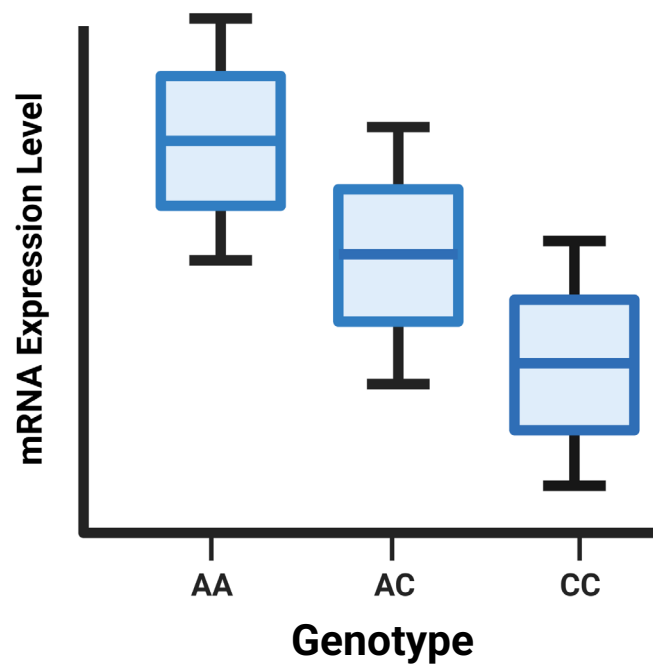
**European (n = 195, TwinsUK 2015)**



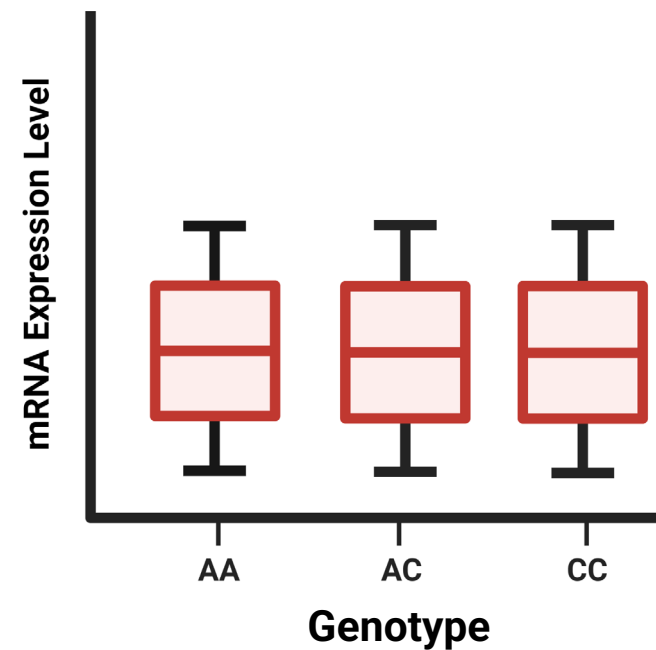
**Indonesian (n = 115, Natri et al. 2020)**

eQTLs

# eQTLs

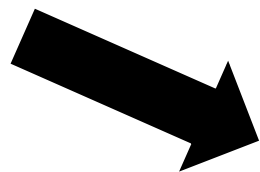


**$Ifsr < 0.01$**

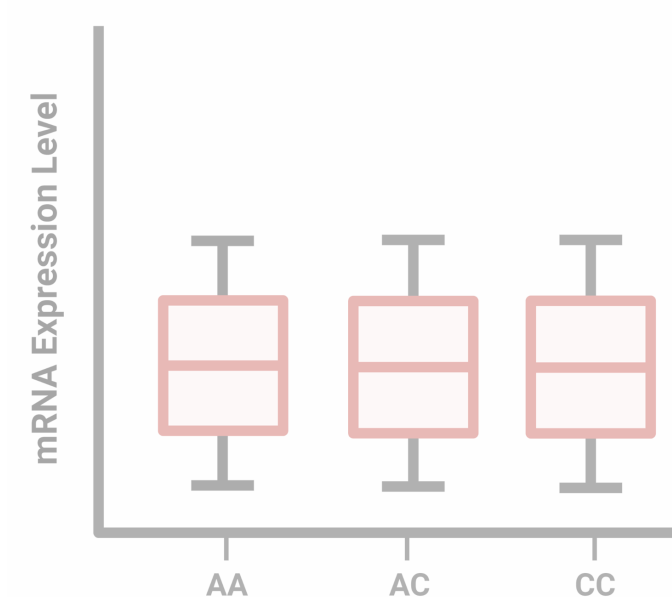
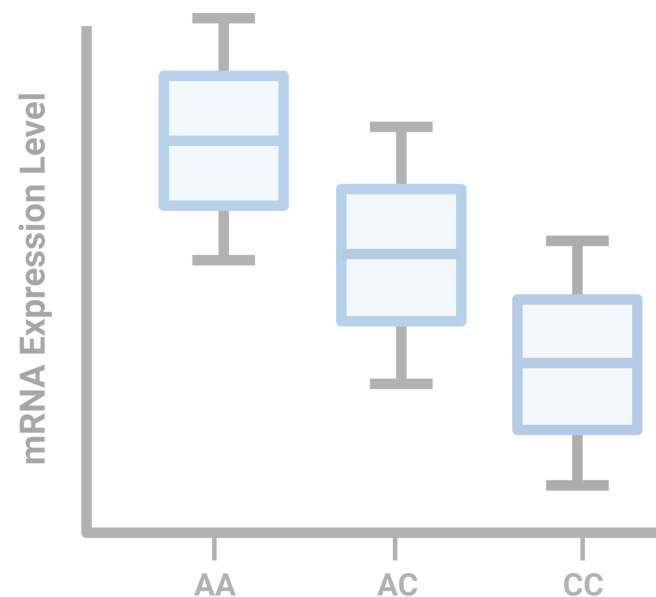


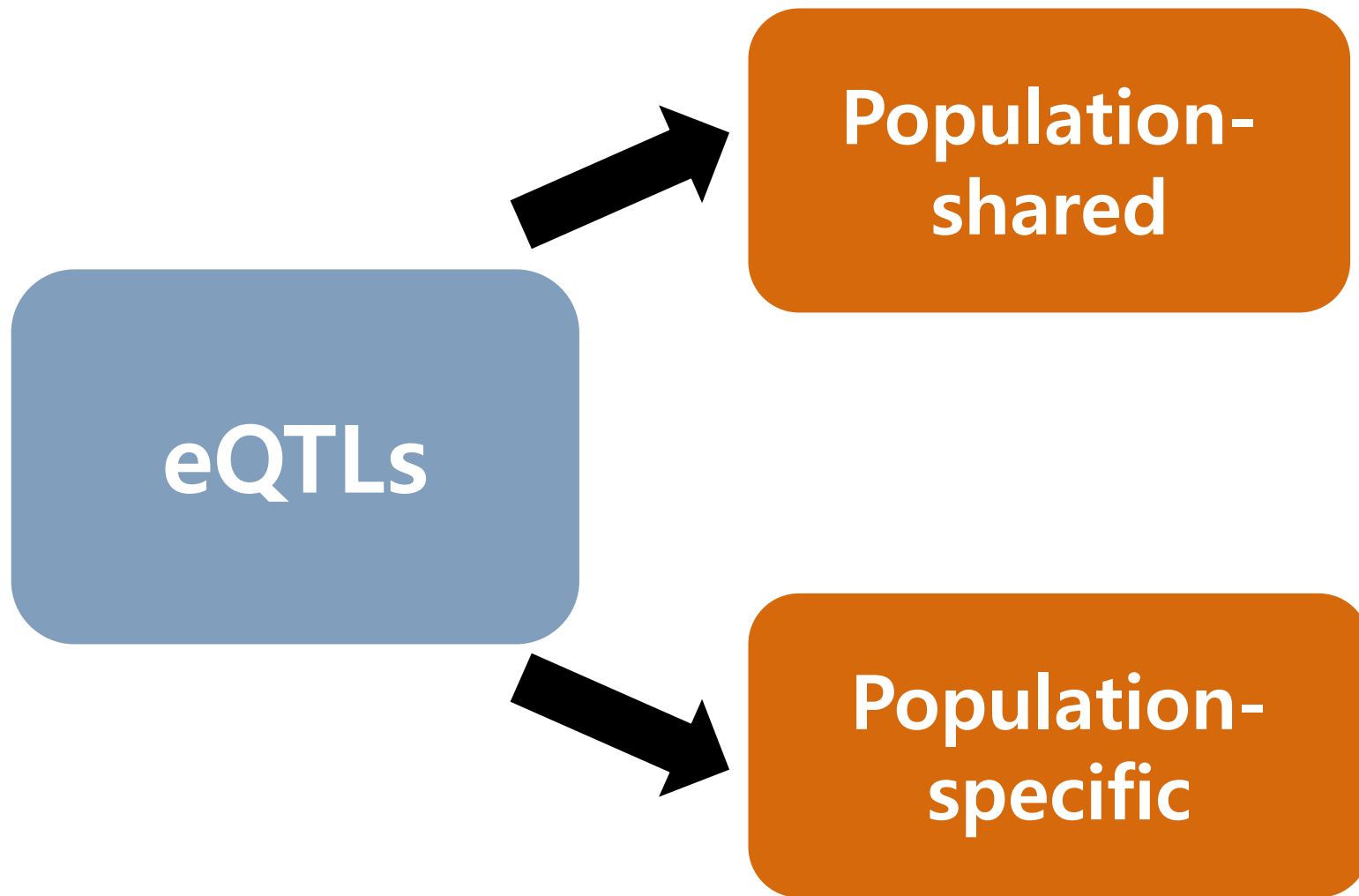
**$Ifsr > 0.10$**

**eQTLs**

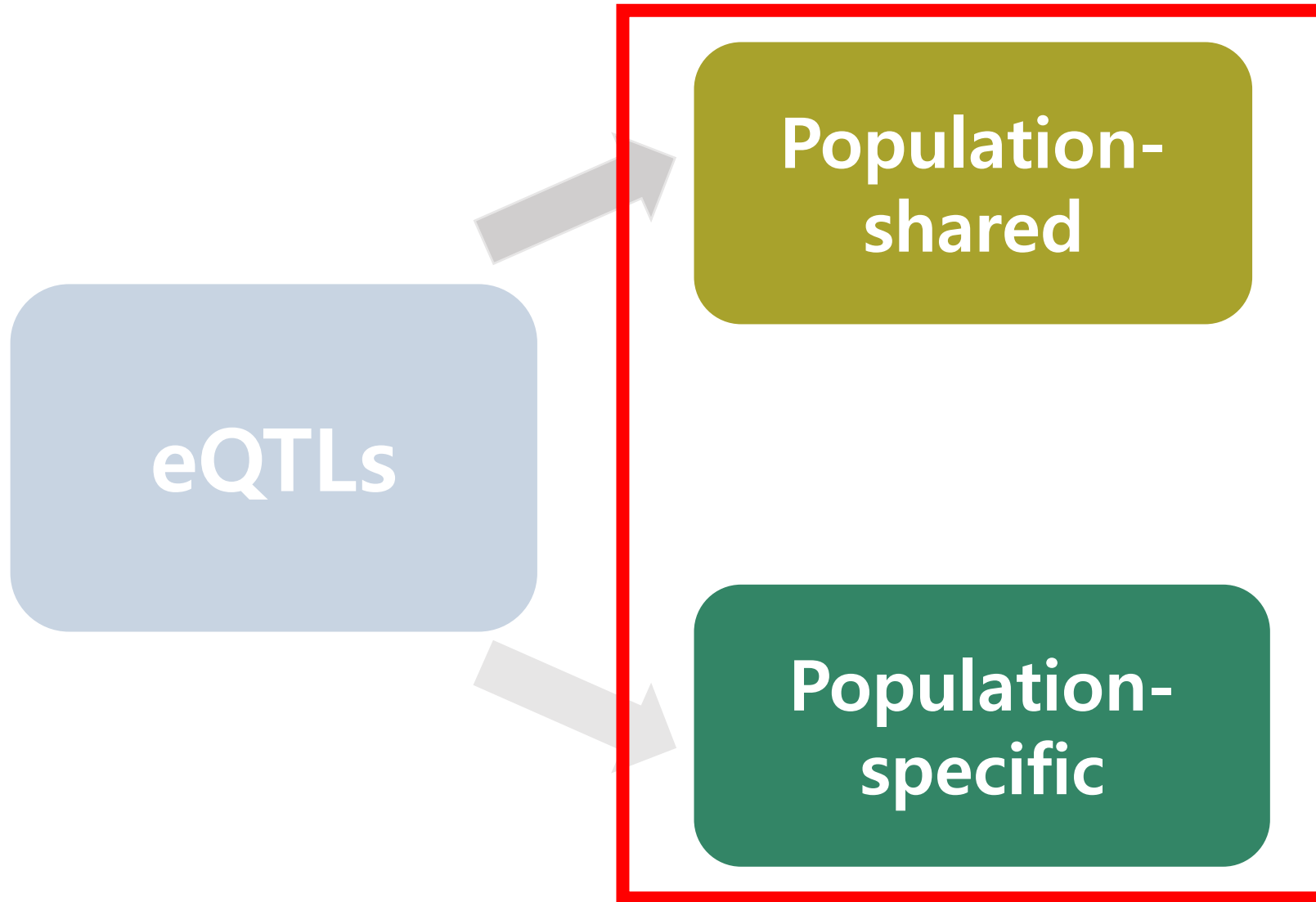


**Population-  
specific**

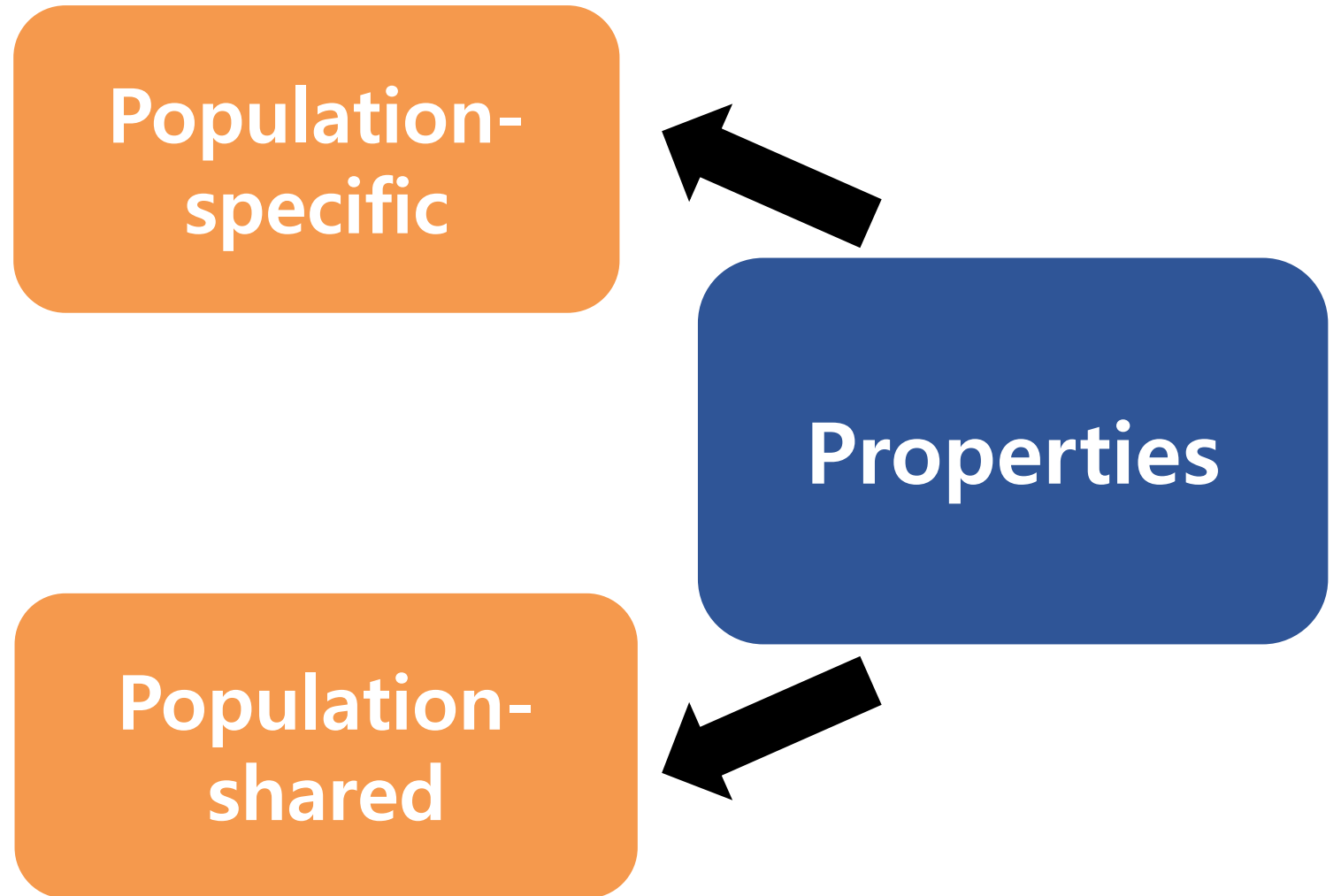


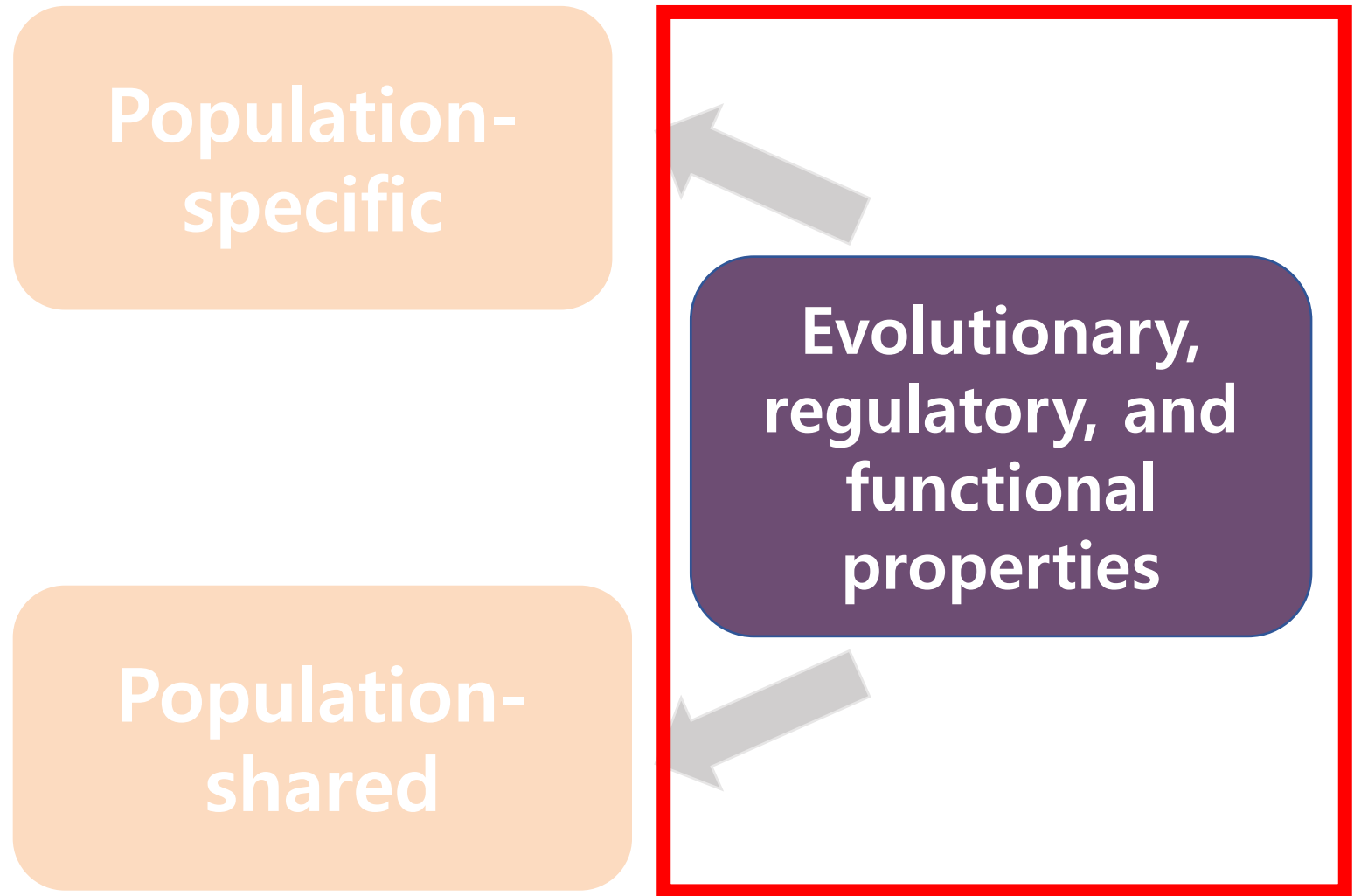


## **“Labels”**









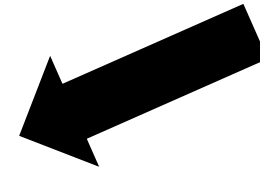
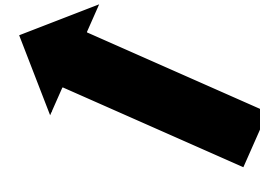
**“Features”**

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

**Population-specific**

**Population-shared**

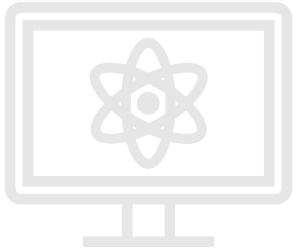
**Evolutionary,  
regulatory, and  
functional  
properties**



# Outline



1. eQTLs

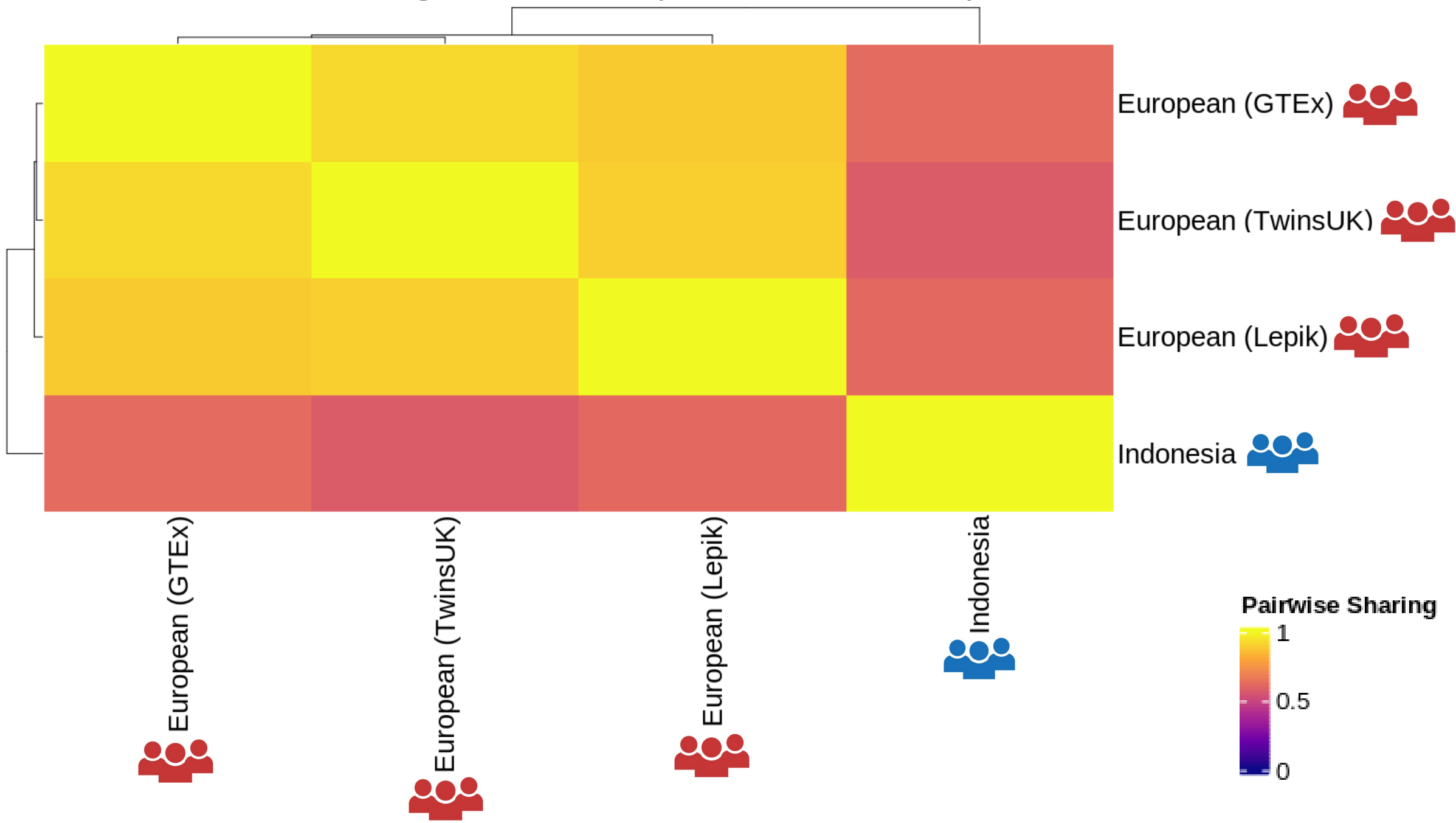


2. Building machine learning models

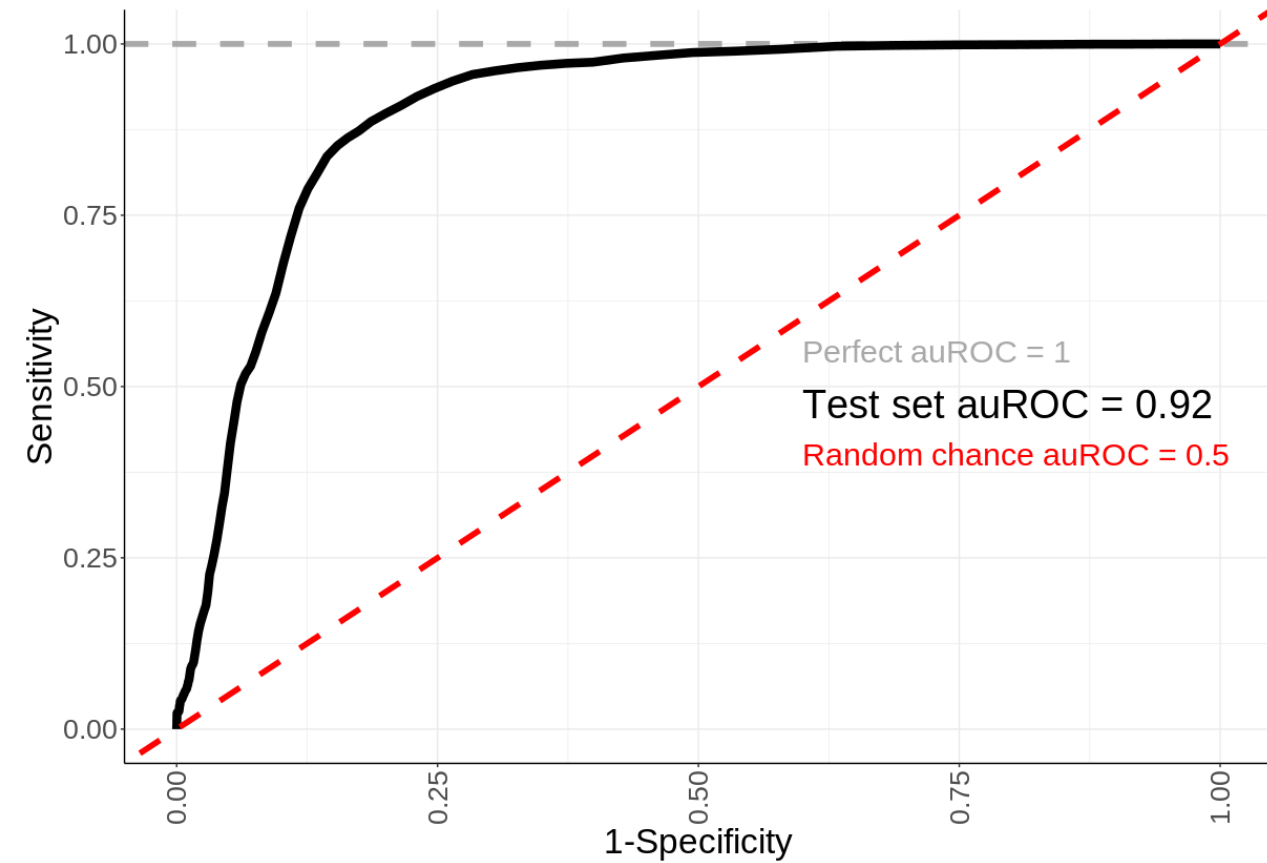


**3. Findings (expected and puzzling)**

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

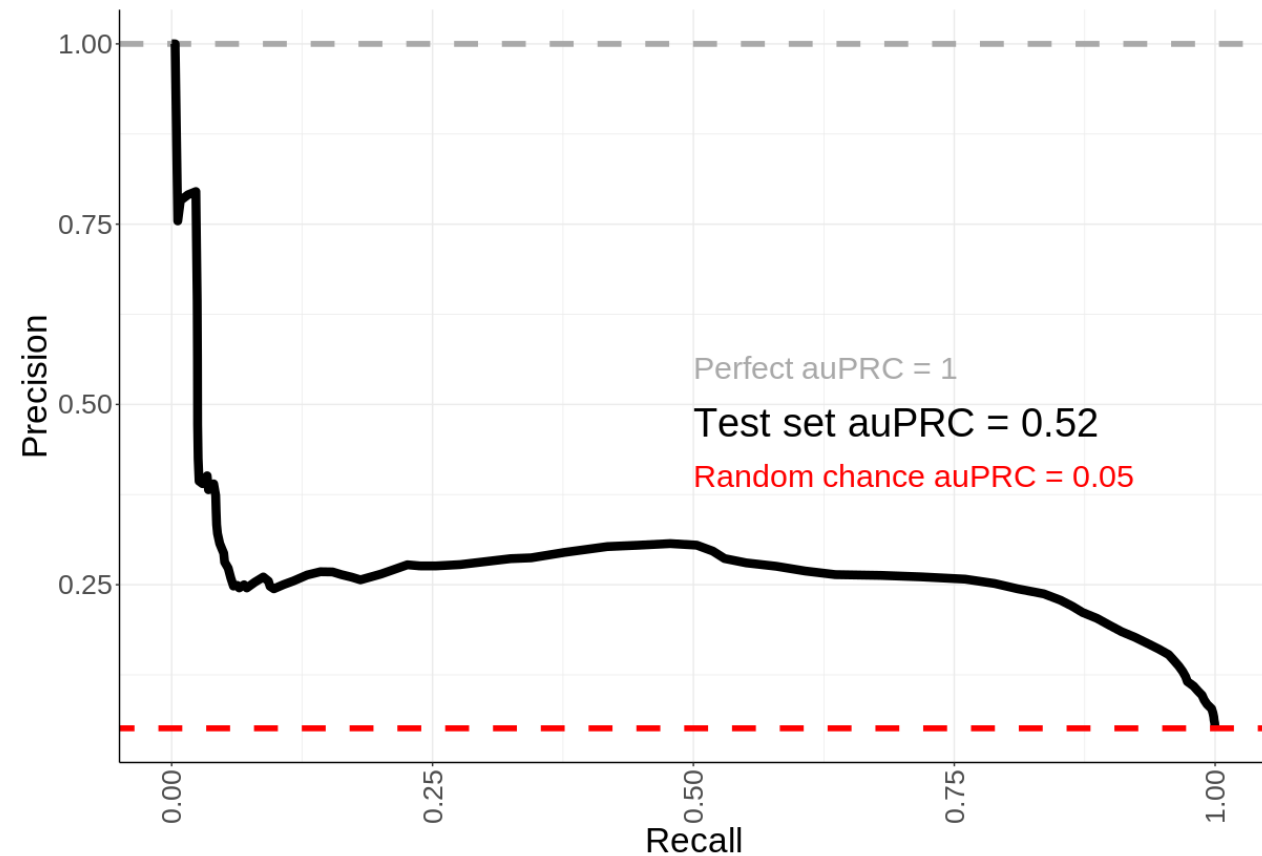
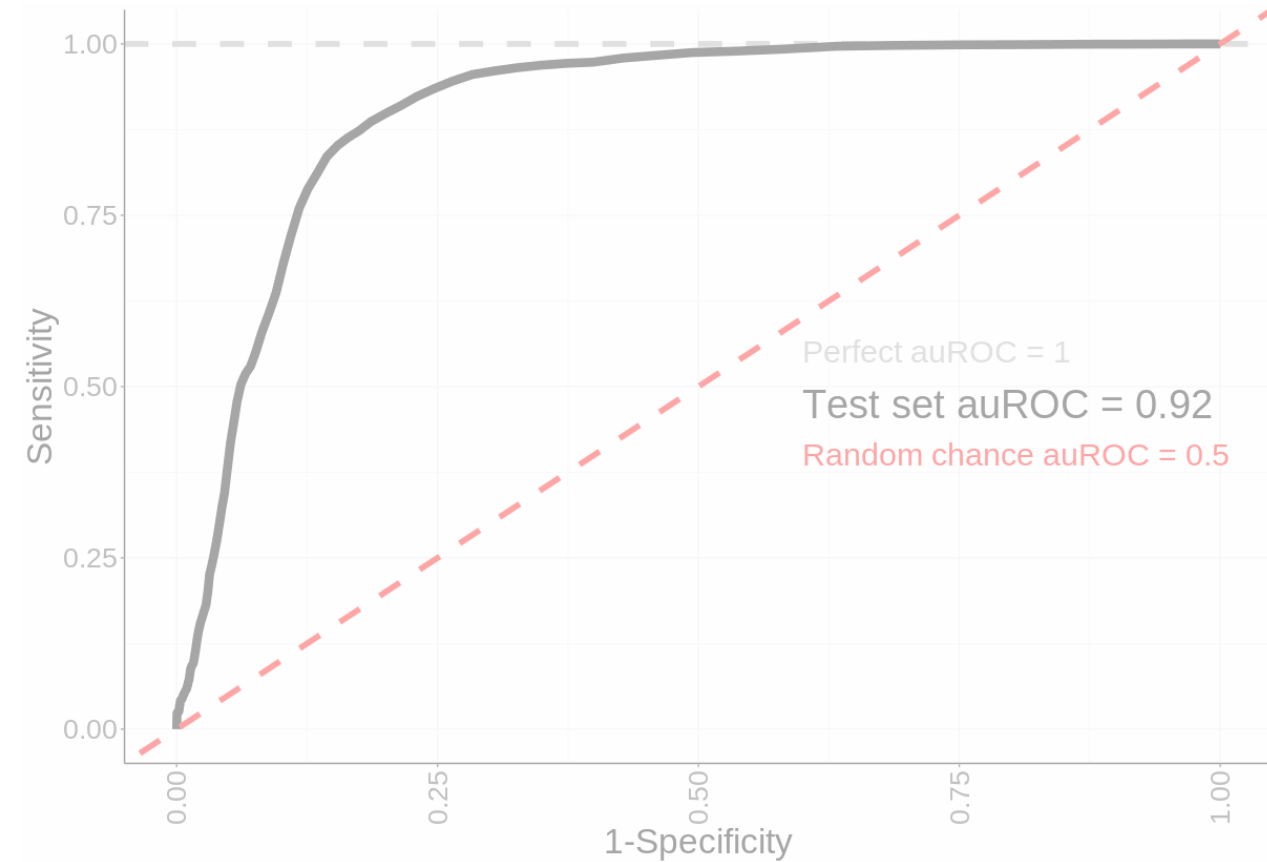


# Prediction Performance



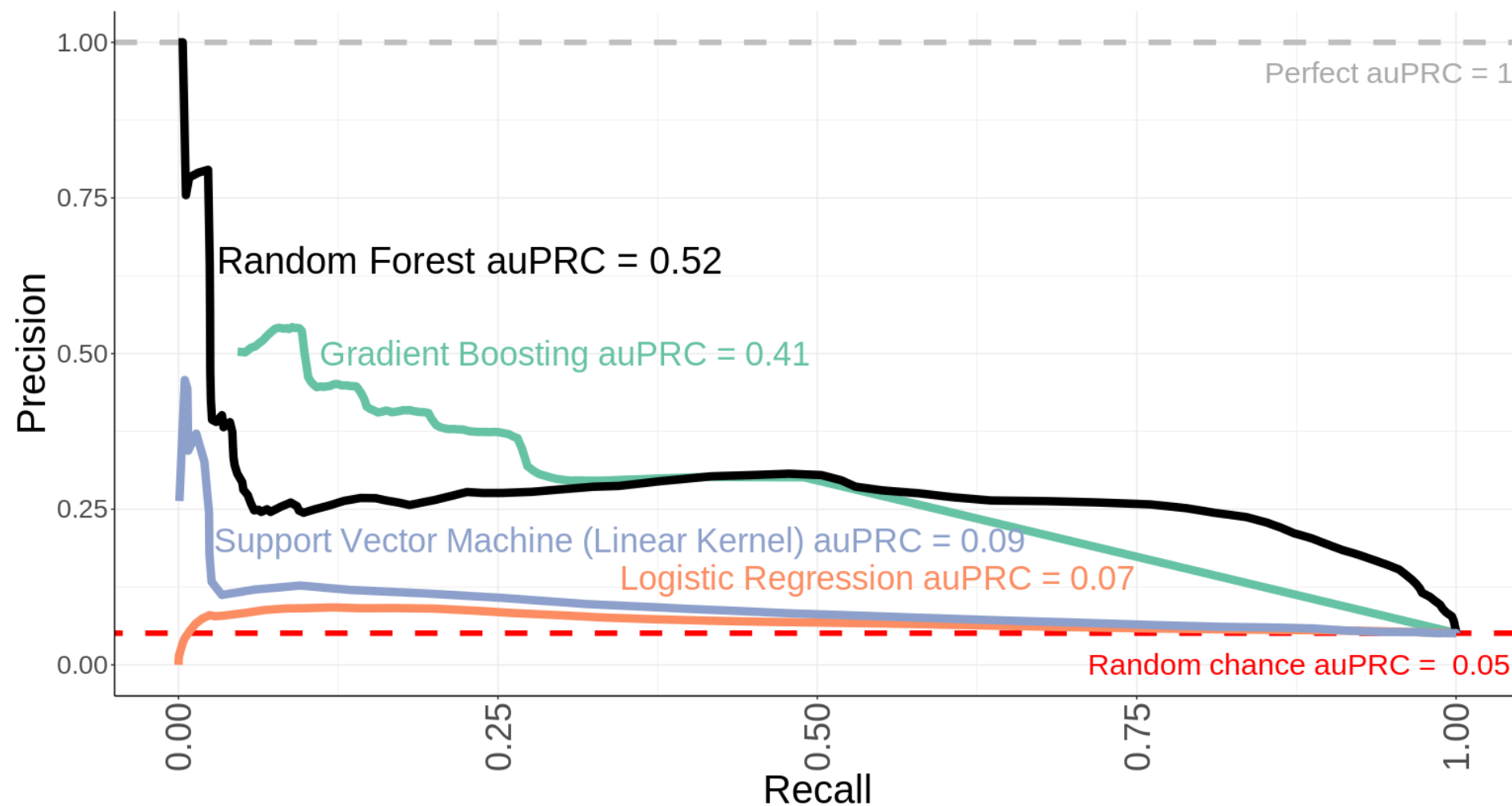
Test set: Chromosomes 8,16

# Prediction Performance



Test set: Chromosomes 8,16

# Prediction Performance across Algorithms



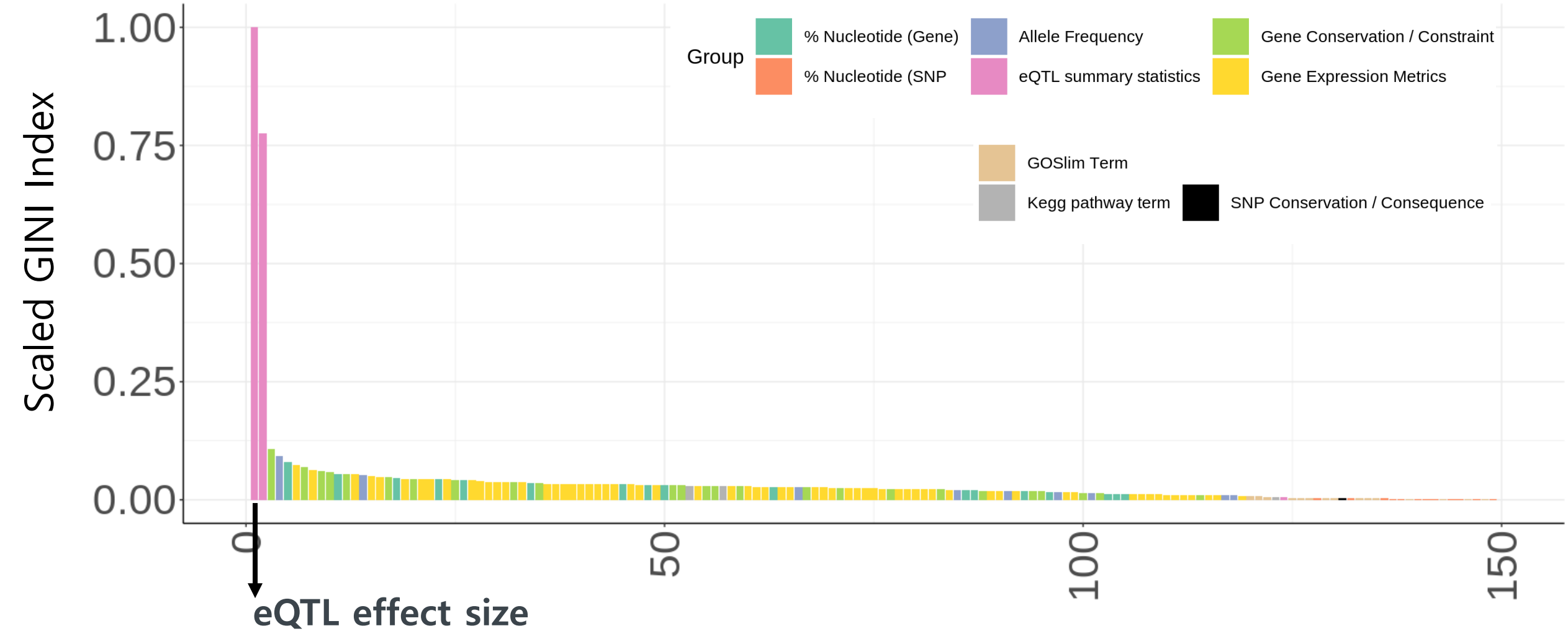
Test set: Chromosomes 8,16



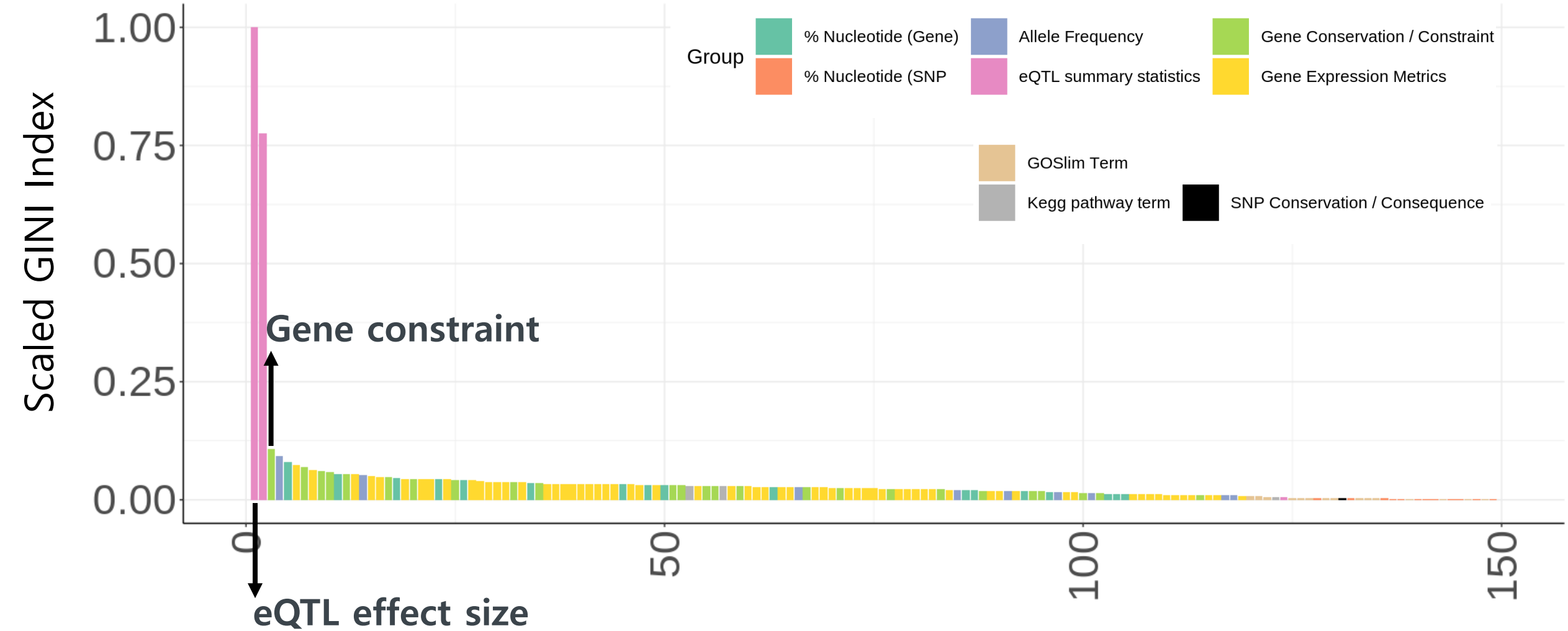
## Features ranked by scaled mean importance



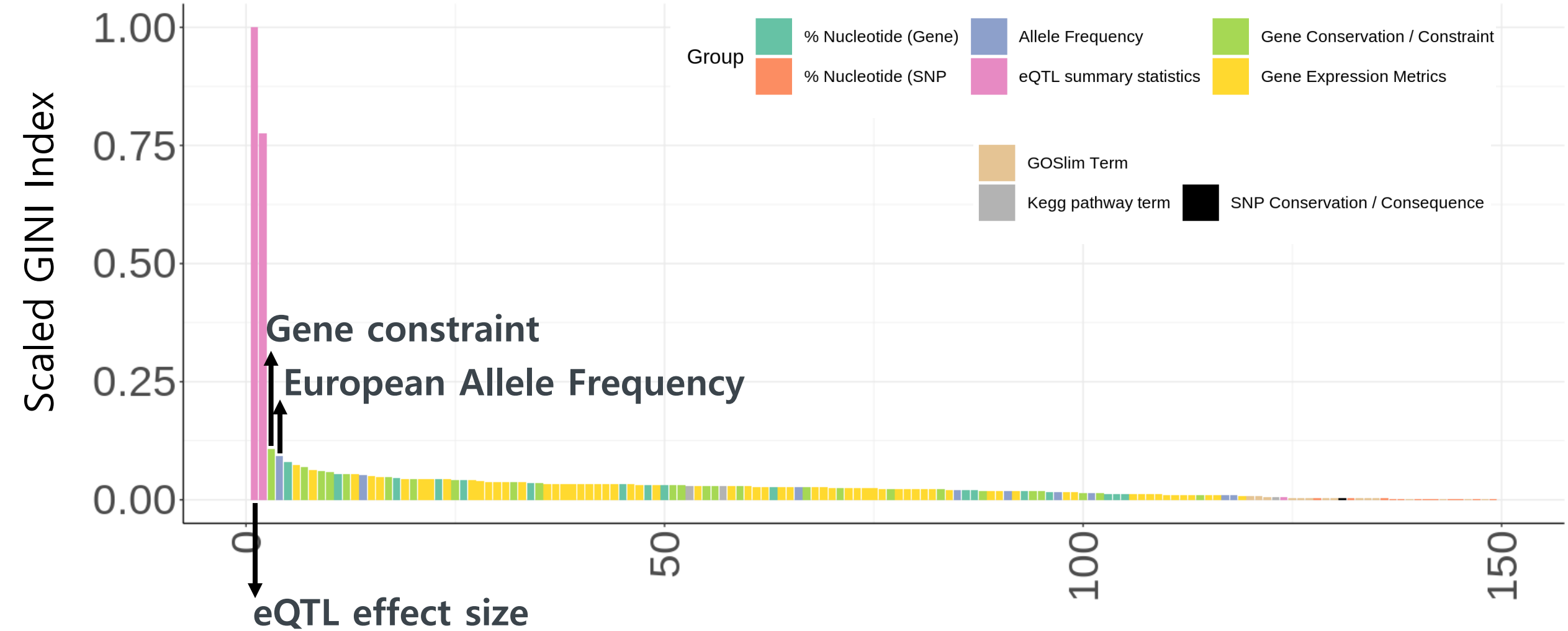
# Features ranked by scaled mean importance



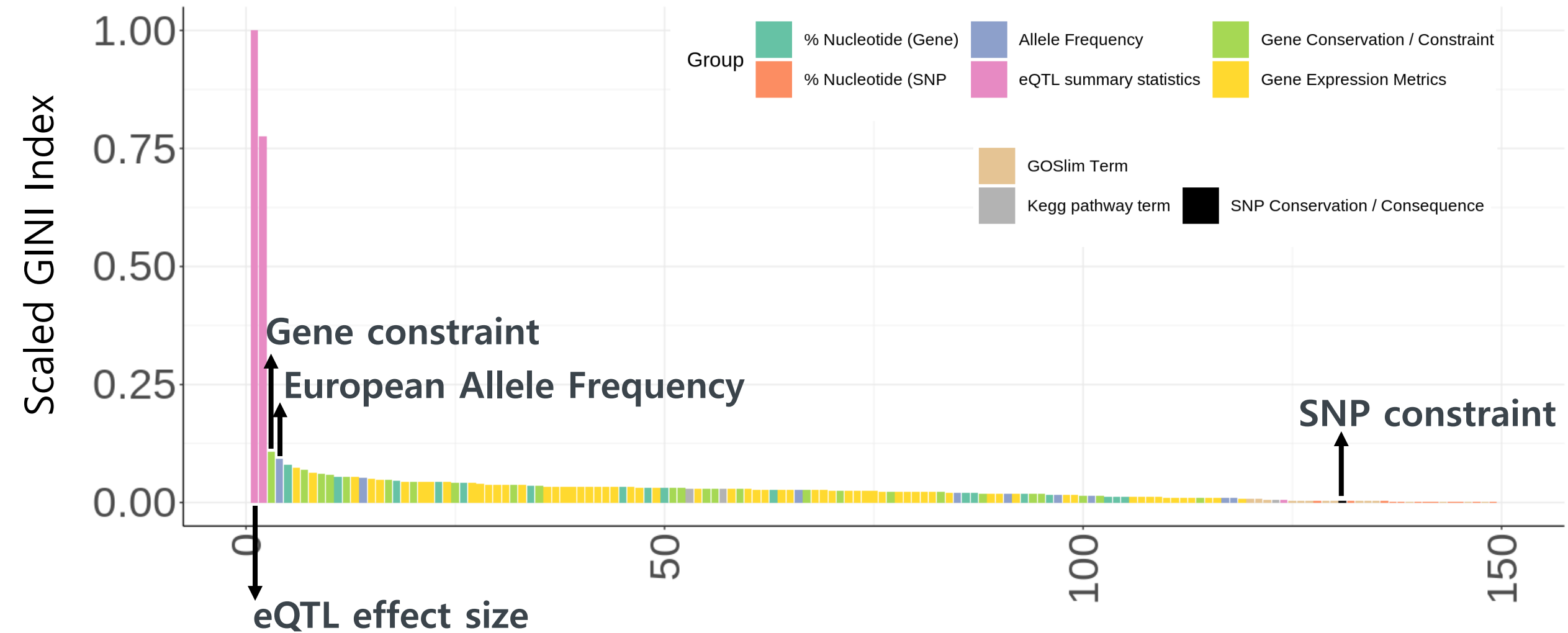
# Features ranked by scaled mean importance



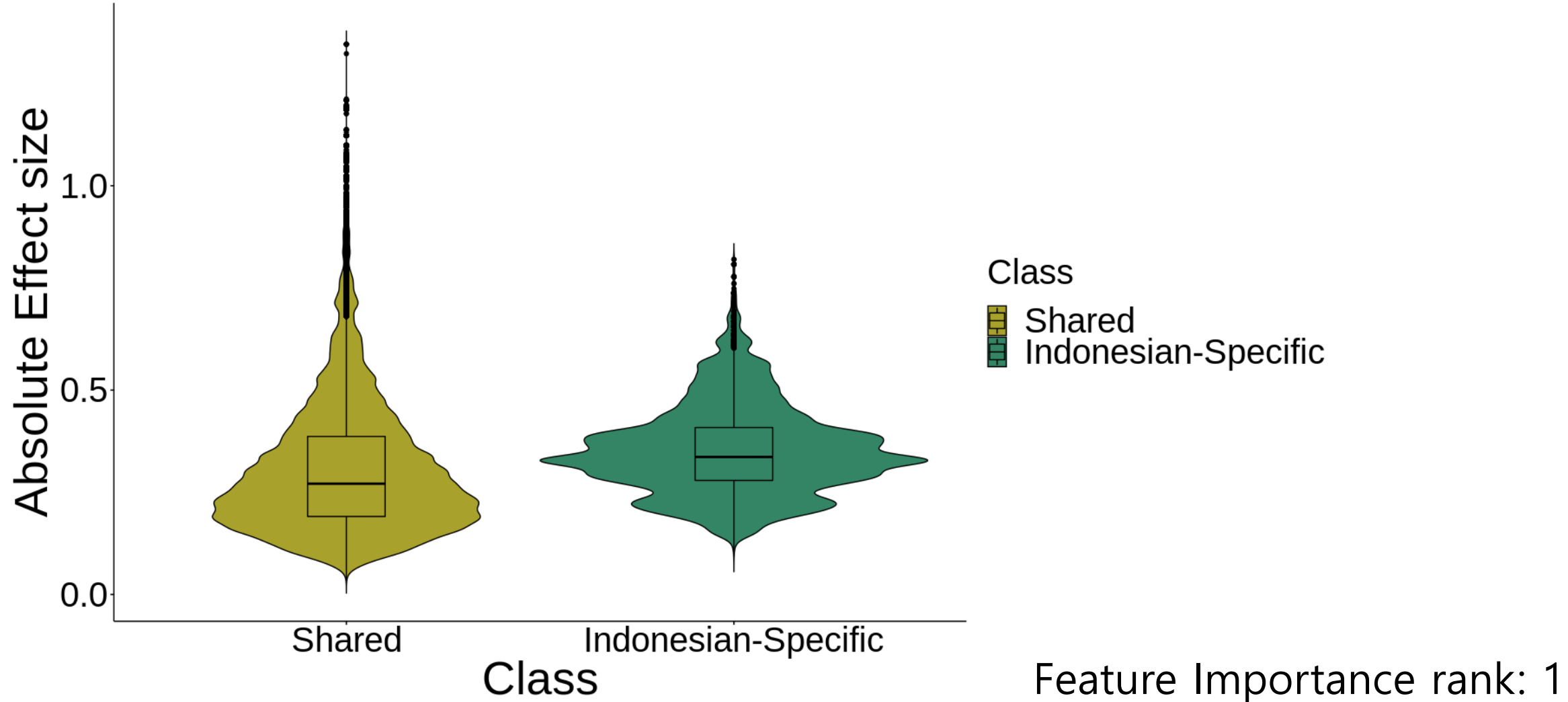
# Features ranked by scaled mean importance



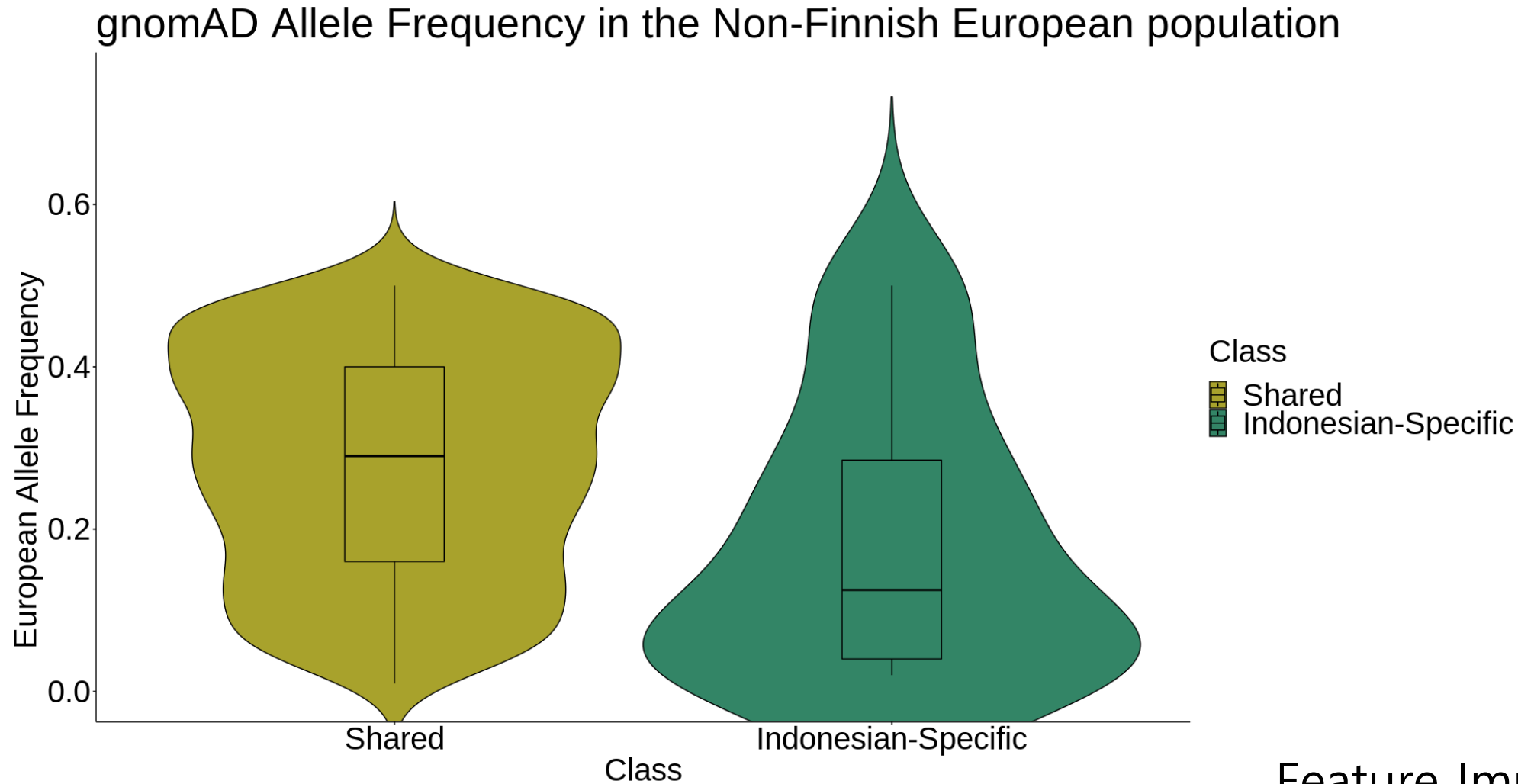
# Features ranked by scaled mean importance



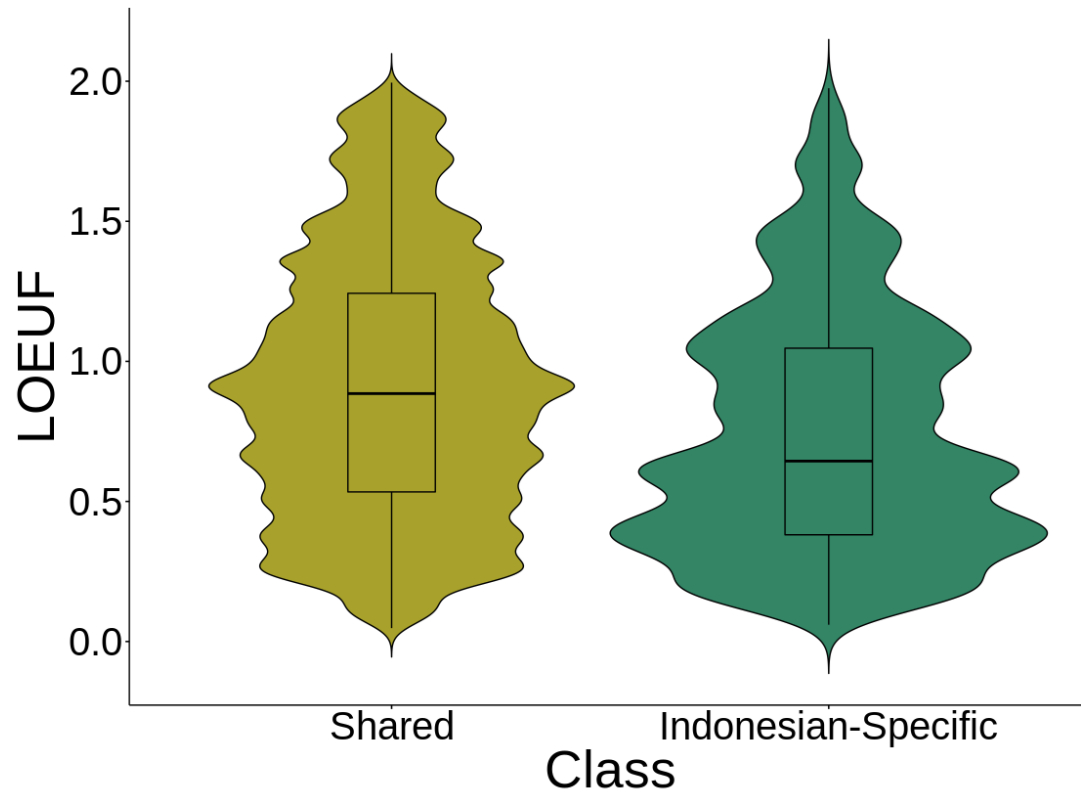
# Population-specific eQTLs have higher effect sizes in their discovery population



# Population-specific eQTLs have different allele frequencies across populations

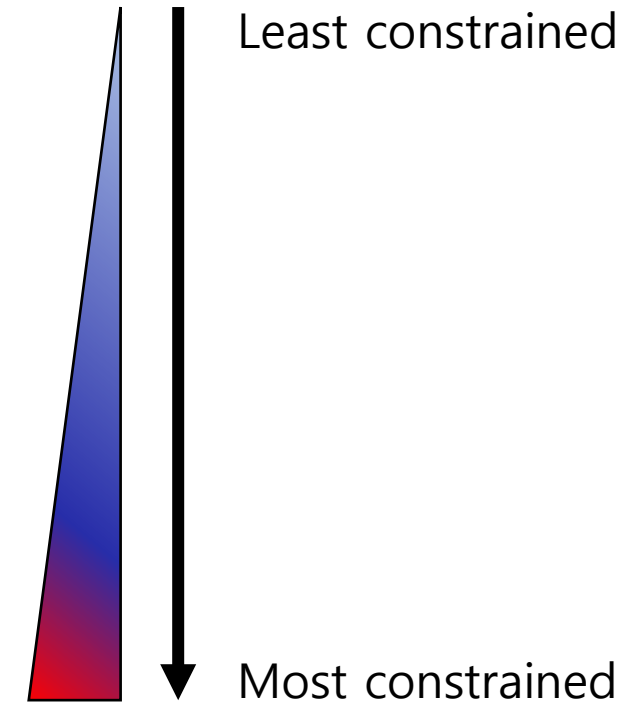


# Population-specific eGenes are more highly conserved



LOEUF = Upper bound of the confidence interval observed over expected number of loss of function variants

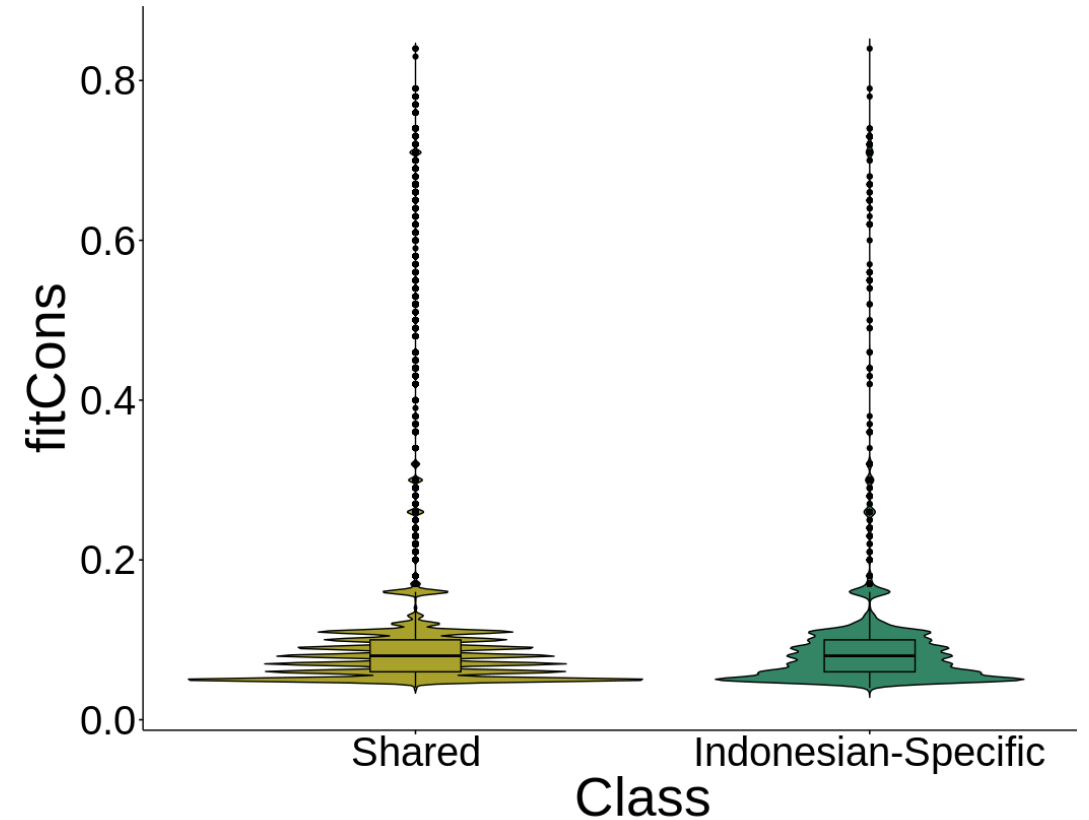
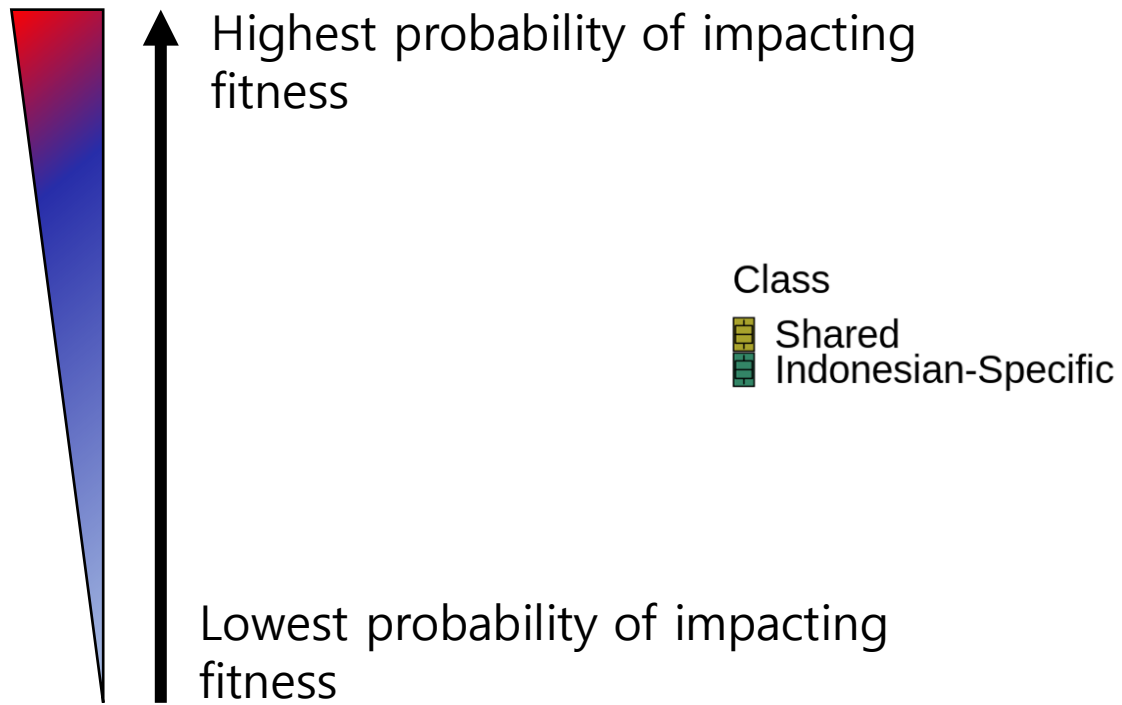
Class  
Shared  
Indonesian-Specific



Gene Conservation Feature Importance rank: 3



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



SNP Conservation Feature Importance rank: 131

**... toward the future** 

# Conclusions

- Machine learning models could improve the transferability of eQTLs from Europeans to underrepresented populations

# Conclusions

- Machine learning models could improve the transferability of eQTLs from Europeans to underrepresented populations
- Some properties are different between shared and specific eQTLs

# Conclusions

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

# Thank you!

Irene Gallego Romero<sup>1,2</sup>



Christina B Azodi<sup>1,3</sup>




Gallego Romero + Bioinformatics and Cellular Genomics Groups

1. Melbourne Integrative Genomics
2. School of BioSciences



3. Bioinformatics and Cellular Genomics

More questions or eQTL datasets?  @ijbeasley