

# Predicting the portability of human expression quantitative trait loci (eQTLs) across populations

**Isobel J Beasley**, Christina B Azodi and Irene Gallego Romero

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



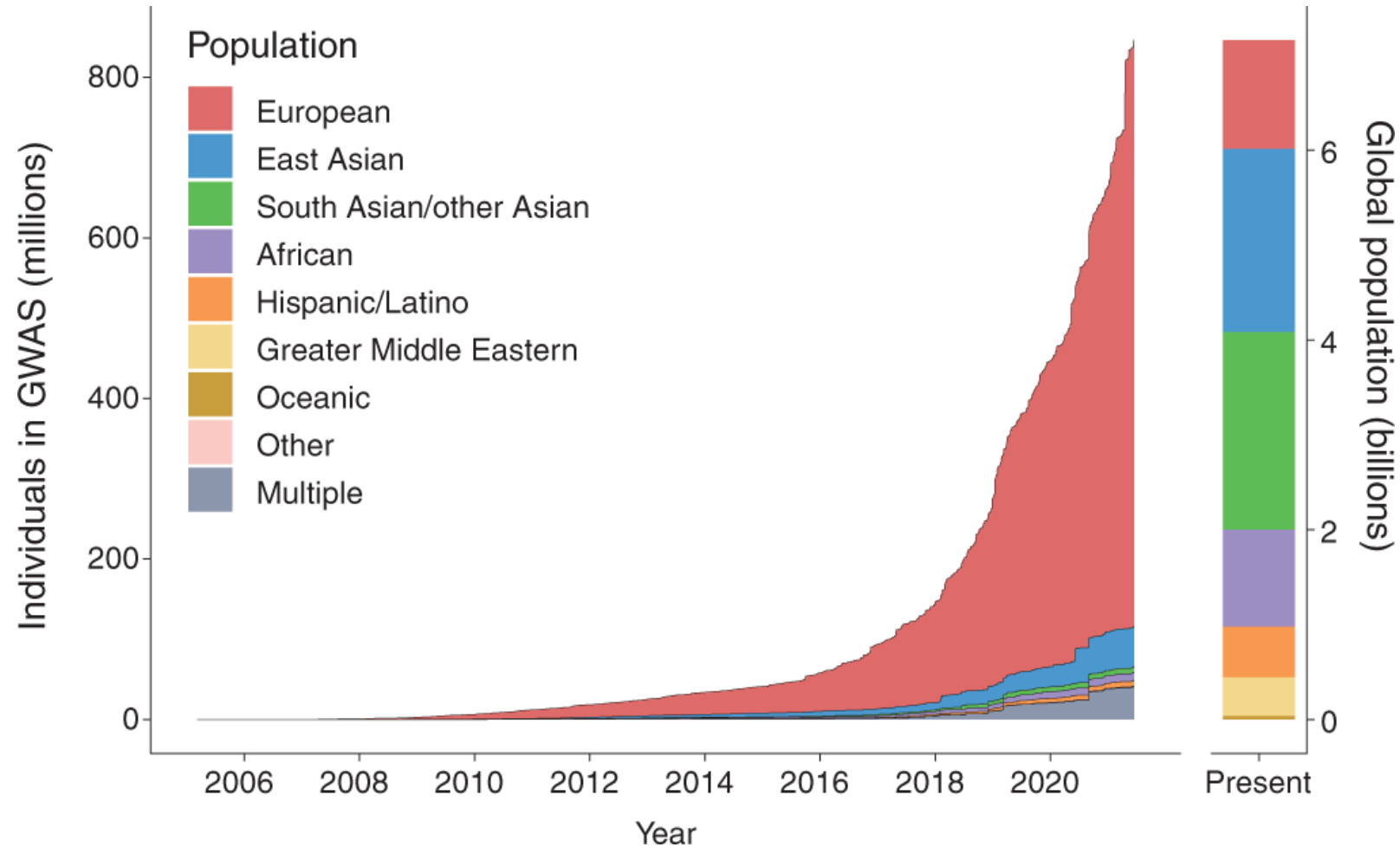
BioPS 2022



@ijbeasley

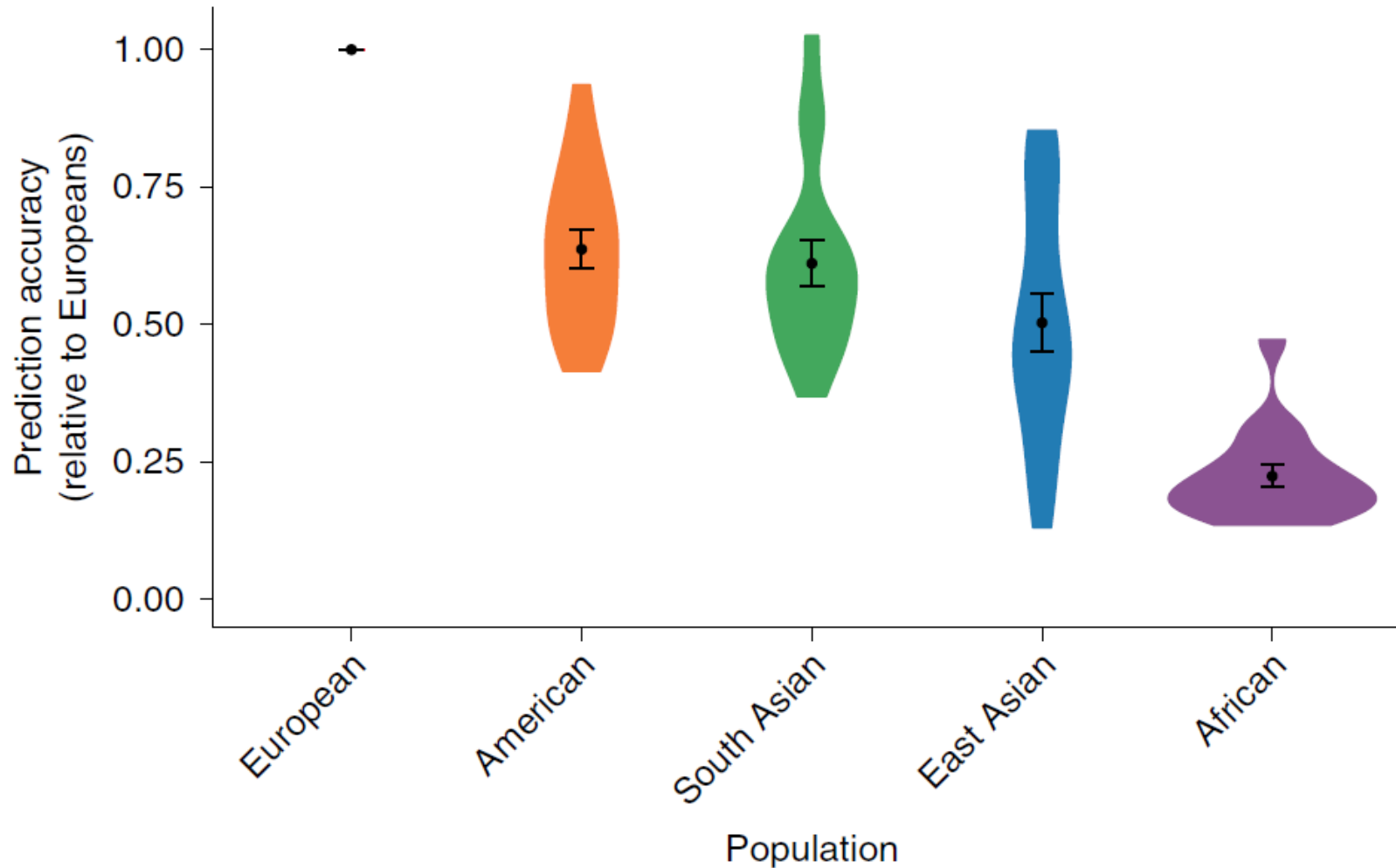


# The overwhelming majority of participants in genetics studies are European



Fatumo et al., 2022, *Nature Medicine*

# Missing diversity increases inequality

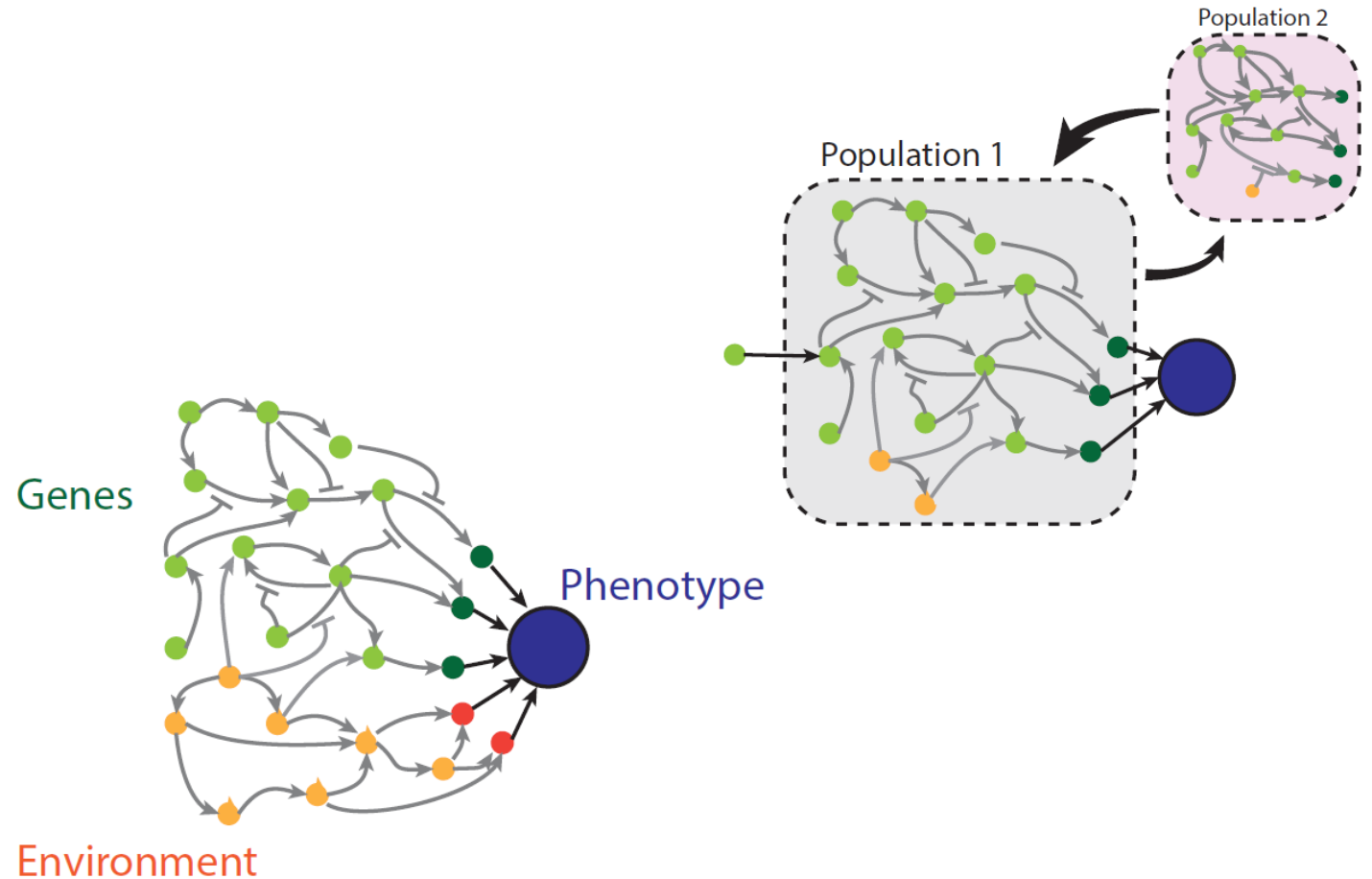


Martin et al., 2019, *Nature Genetics*

# Reasons for the gap in 'portability'

Cross-population differences in:

- Linkage disequilibrium
- Allele frequency
- Gene by environment interactions
- Gene by gene interactions



# How can we reduce the equity gap

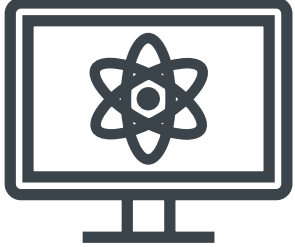
More data from non-Europeans...

# How can we reduce the equity gap

More data from non-Europeans...

... and also better understanding the reasons behind the gap in portability

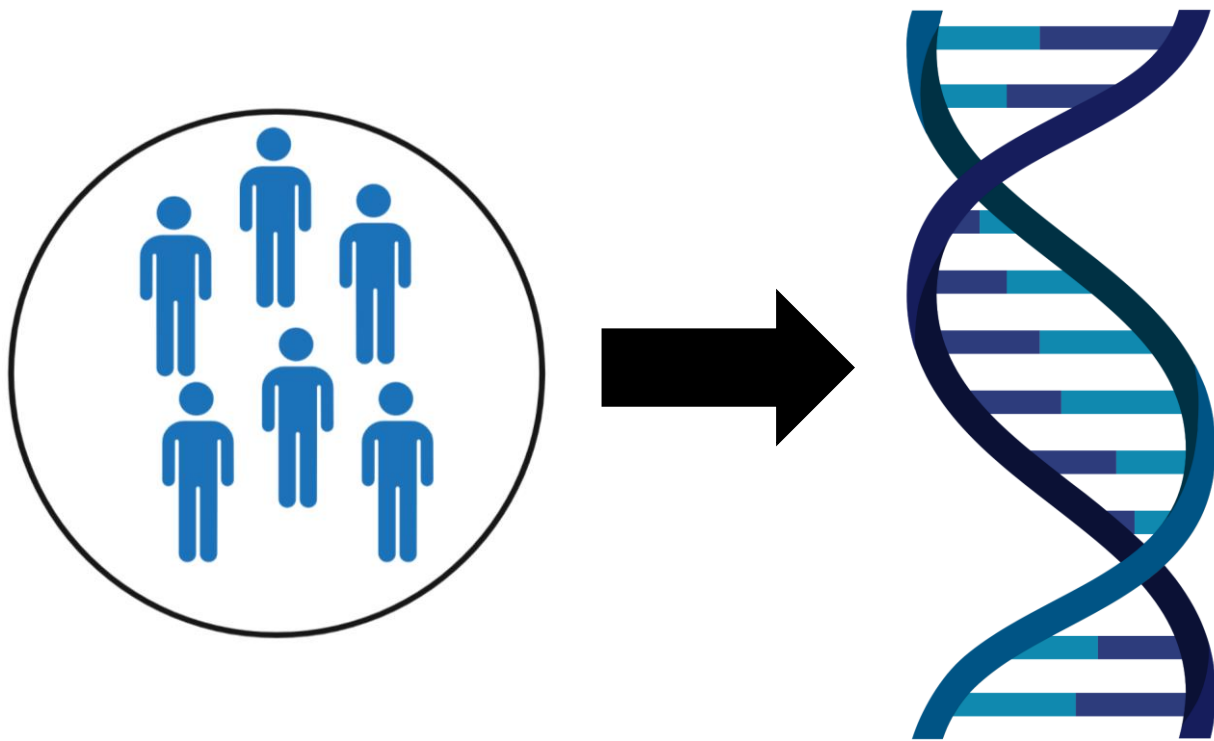
# Outline



**1. The method: eQTLs + machine learning models**

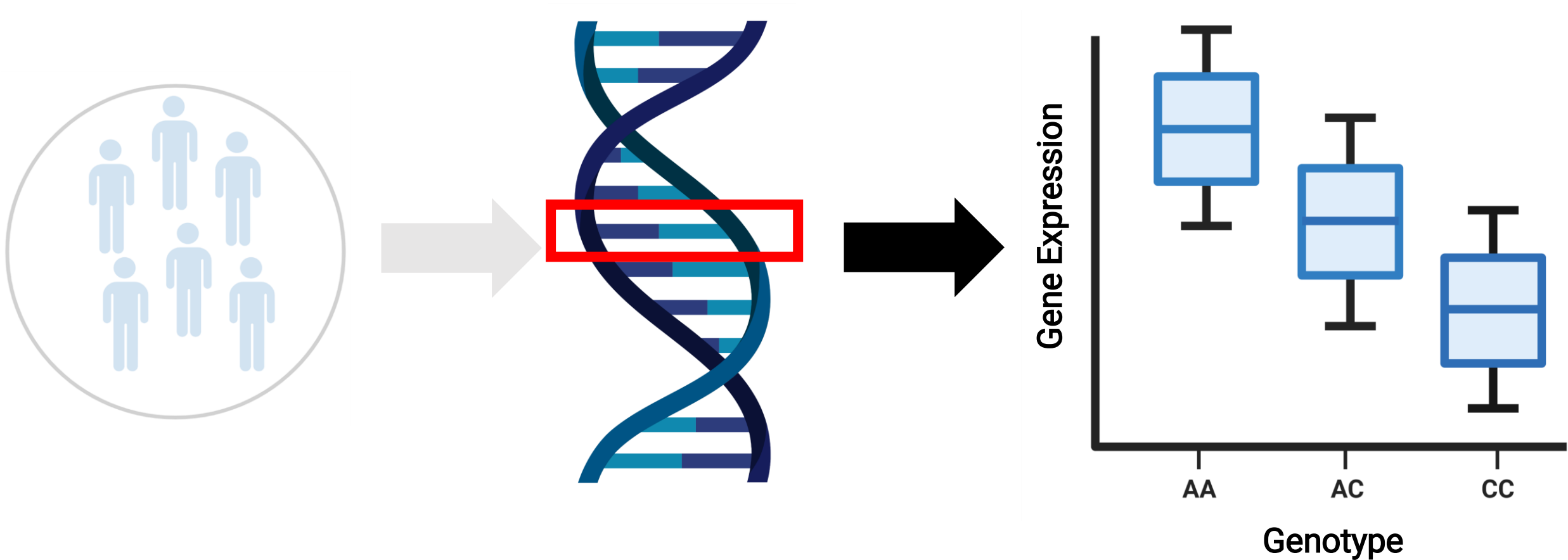


**2. The main result: Most differences between populations are statistical ...**





# eQTLs (expression Quantitative Trait Loci)



# The training data



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



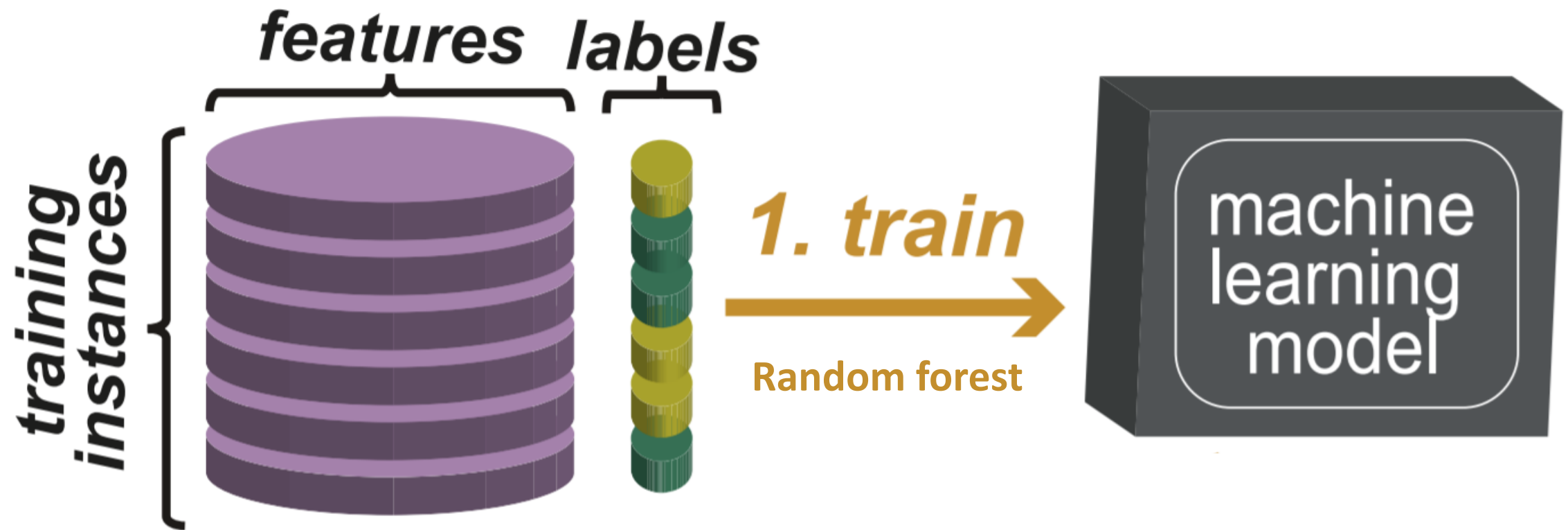
**European (n = 379, GTEx. 2020)**



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



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

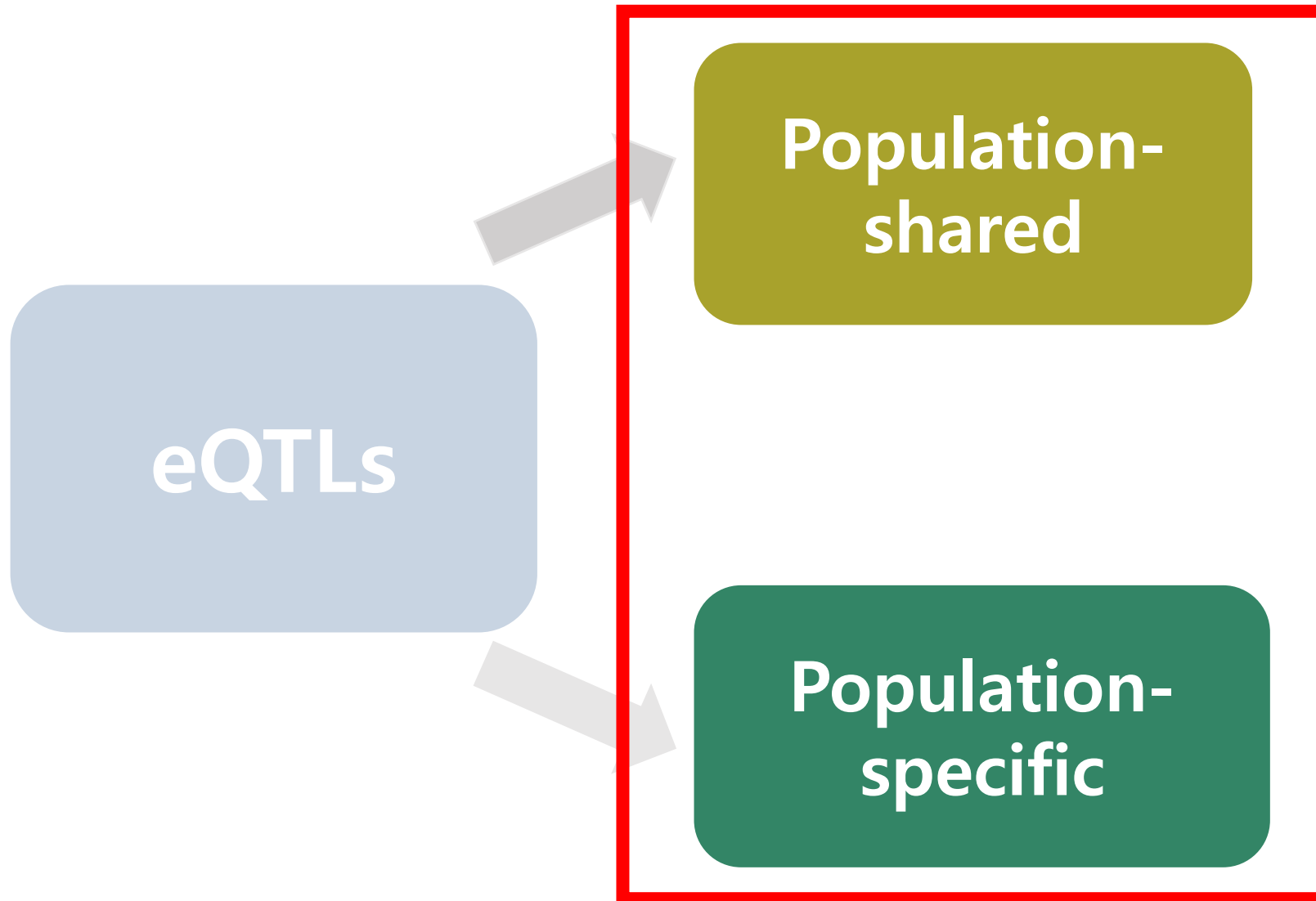


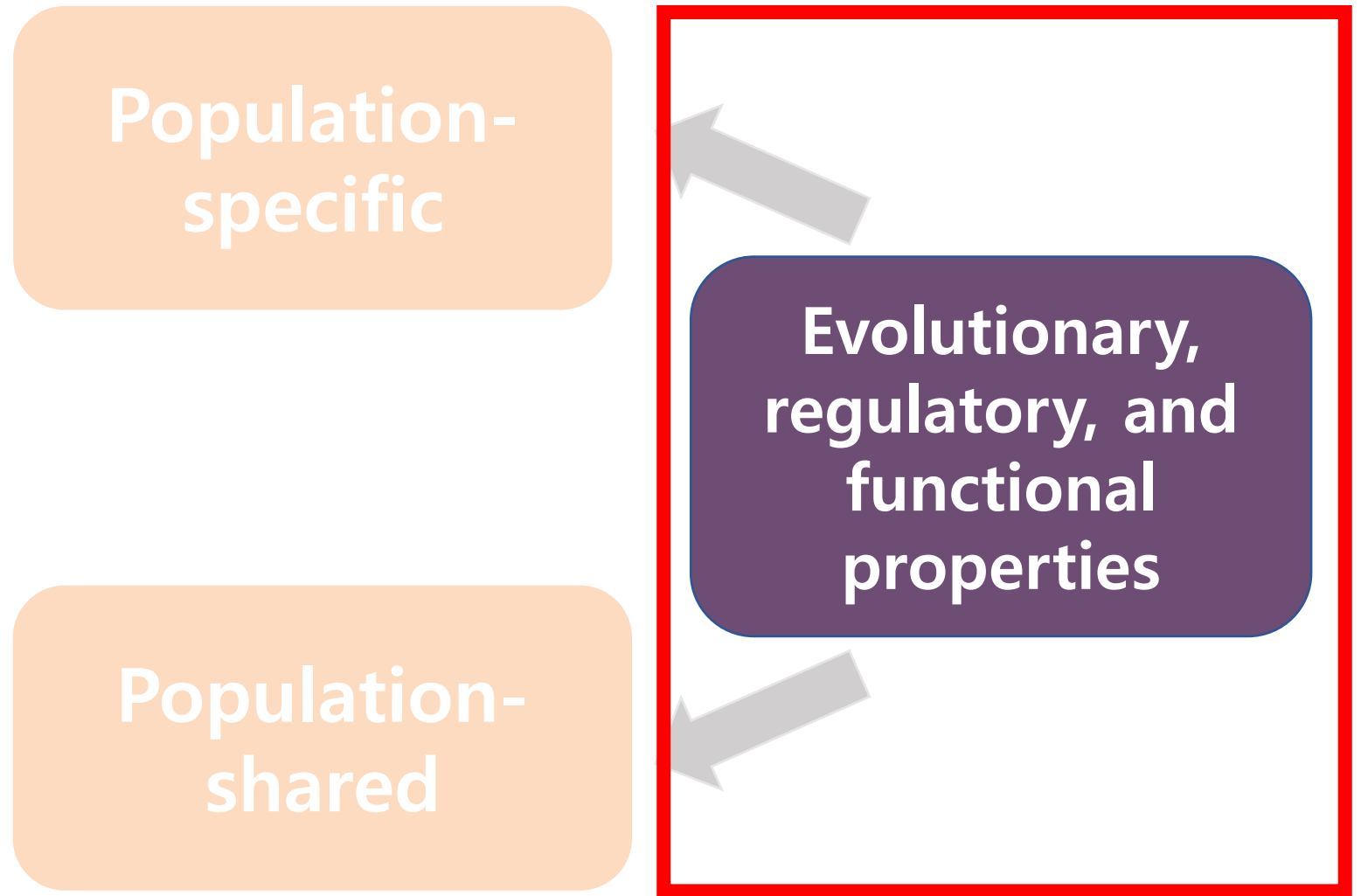
- features**
- Linkage Disequilibrium
  - Allele frequency
  - Gene and SNP Conservation
  - SNP Genomic location
  - % Nucleotide
  - Gene Ontology

**1. train**

machine  
learning  
model

## **“Labels”**





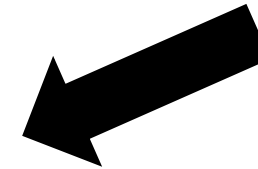
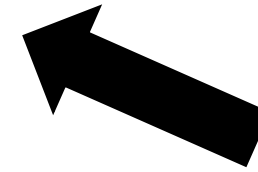
**“Features”**

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

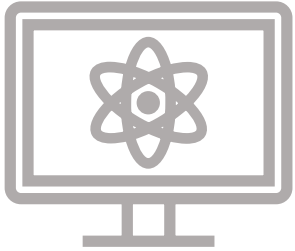
**Population-specific**

**Population-shared**

**Evolutionary,  
regulatory, and  
functional  
properties**



# Outline



1. The method: eQTLs + machine learning models



2. The main result: Most differences between populations are statistical ...