Complex genomic patterns underpin human population differences in expression quantitative trait loci (eQTLs)

Isobel J Beasley, Christina B Azodi, Irene Gallego Romero

@ijbeasley @cbazodi @ee_reh_neh

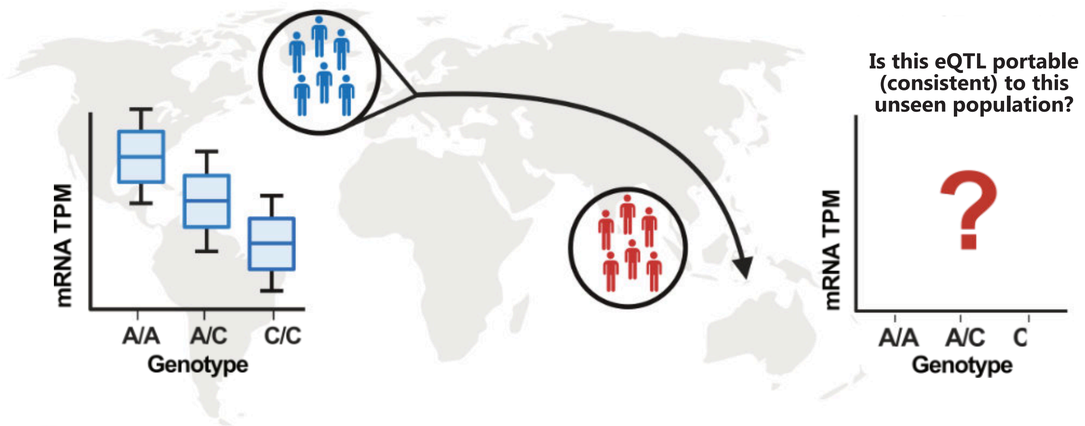
THE UNIVERSITY OF MELBOURNE

SVI ST VINCENT'S INSTITUTE MEDICAL RESEARCH

Introduction

Individuals of European ancestry disproportionately dominate participation in human genetic studies, to the detriment of scientific inquiry and the equitable translation of genomic research (1,2). For instance, while expression quantitative trait locus mapping can connect disease associated genetic variants to explanatory regulatory mechanisms, not all associations between genotype at a given locus and variation in gene expression (eQTL) are shared (non-portable) across populations (3,4). **Understanding the reasons behind, and the features of non-portable eQTLs could lead to more equitable translations of European eQTL-derived disease inferences to understudied populations.**

Here we use summary statistics from published multi-population studies (across two tissues, four populations) to investigate cross-population eQTL differences by training supervised machine learning models to classify eQTLs as portable or not from one population to another.



Whole blood RNA-seq

European (n = 471, Buil et al. 2015)

European (n = 379, GTEx 2020)

European (n = 195, Lepik et al. 2017)

Indonesian (n = 115, Natri et al. 2022)

CD14+ Monocytes Microarray

European American (n = 578, Mogil et al. 2018)

Hispanic (n = 352, Mogil et al. 2018)

African American (n = 233, Mogil et al. 2018)

Created with BioRender.com

Methods

(A) Integrate summary statistics

gene_snp	reference	alternative	effect (dataset 1)	standard error (dataset 1)	effect (dataset 2)	standard error (dataset 2)	...
ENSG1_rs1	A	G	0.27	0.17	1.82	0.54	...
ENSG1_rs2	C	T	1.02	0.04	1.10	0.13	...
...	...	...	...	...	...	...	...

Population A

gene_snp	effect	...
ENSG1_rs1	0.27	...
ENSG1_rs2	1.02	...
...	...	...

Population B

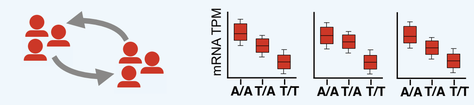
gene_snp	effect	...
ENSG1_rs1	1.82	...
ENSG1_rs2	1.10	...
...	...	...

...

(B) Classify eQTLs

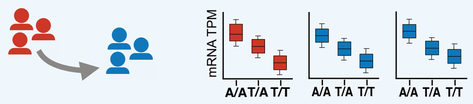
eQTL ID	class	reference	alternative
ENSG1_rs1	not-portable (A to B)	A	G
ENSG1_rs2	portable (A to B)	C	T
...	...	...	...

i) Is this gene-snp pair an eQTL in Population A?



ii) Is this eQTL portable from Population A to Population B?

effect size ratio < 0.5



(C) Extract features

eQTL ID	class	reference	alternative	MAF	eGene median log(TPM+1)	GOID1	...
ENSG1_rs1	not-portable (A to B)	A	G	0.43	18.5	1	...
ENSG1_rs2	portable (A to B)	C	T	0.21	17.1	0	...
...	...	...	...	...	....	...	...

Labels: class

Features: MAF, eGene median log(TPM+1), GOID1, ...

- Expression (e.g. GTEx cross-tissue)
- Evolutionary (e.g. gnomAD, phylop)
- Functional (e.g. GO Slim, Interpro)

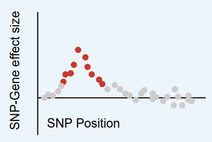
GTEx

GENEONTOLOGY Unifying Biology

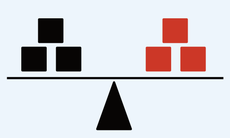
gnomAD

(D) Train machine learning models

i) Sample one eQTL per eGene



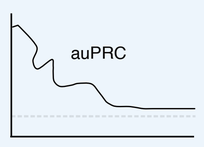
ii) Downsample to balance portable and non-portable eQTLs



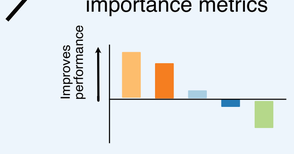
iii) Train model to optimise auPRC performance

(E) Test and Interrogate

i) Test set performance



ii) Leave One Feature Out (LOFO) importance metrics



Test set: Chromosomes 8 & 16

Results

Decision tree models perform best

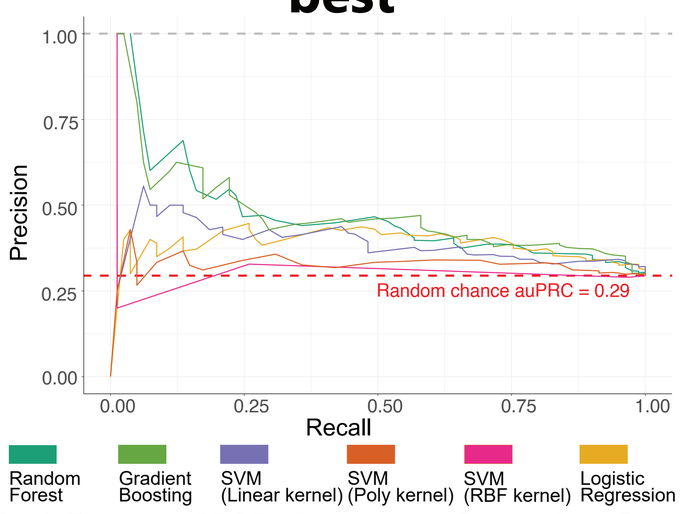


Fig 1. Test set auPRC Performance for Indonesia to Europe whole blood model

Median auPRC performance improvement over random chance = 0.073

Model	Random chance auPRC	Delta auPRC
Indonesian to European	0.29	+0.183
European to Indonesian	0.31	+0.073
European American to African American	0.59	+0.057
European American to Hispanic	0.42	+0.068
African American to European American	0.43	+0.073
African American to Hispanic	0.34	+0.093
Hispanic to European American	0.30	+0.090
Hispanic to African American	0.58	+0.087

No feature class consistently distinguish portable and non-portable eQTLs

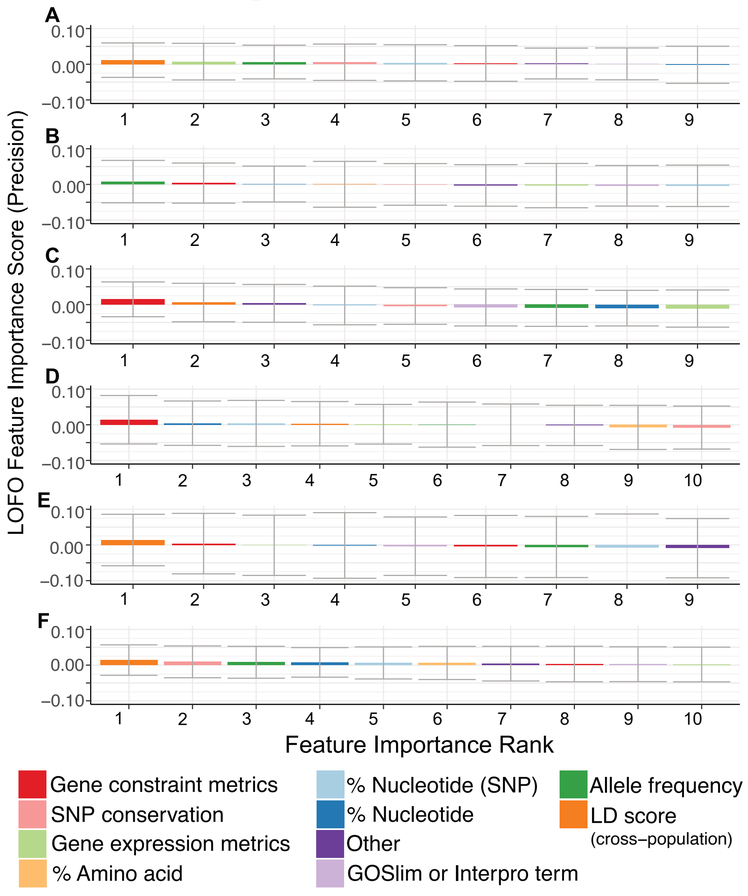


Fig 2. LOFO feature importance scores for all monocyte pairwise models for the metric precision

Height of bar is mean score, error bars show standard devivation. A) European American to African American model, B) European American to Hispanic mode, C) African American to European American model, D) African American to Hispanic model, D) Hispanic to European American model, E) Hispanic to African American model

Portability is tissue dependent

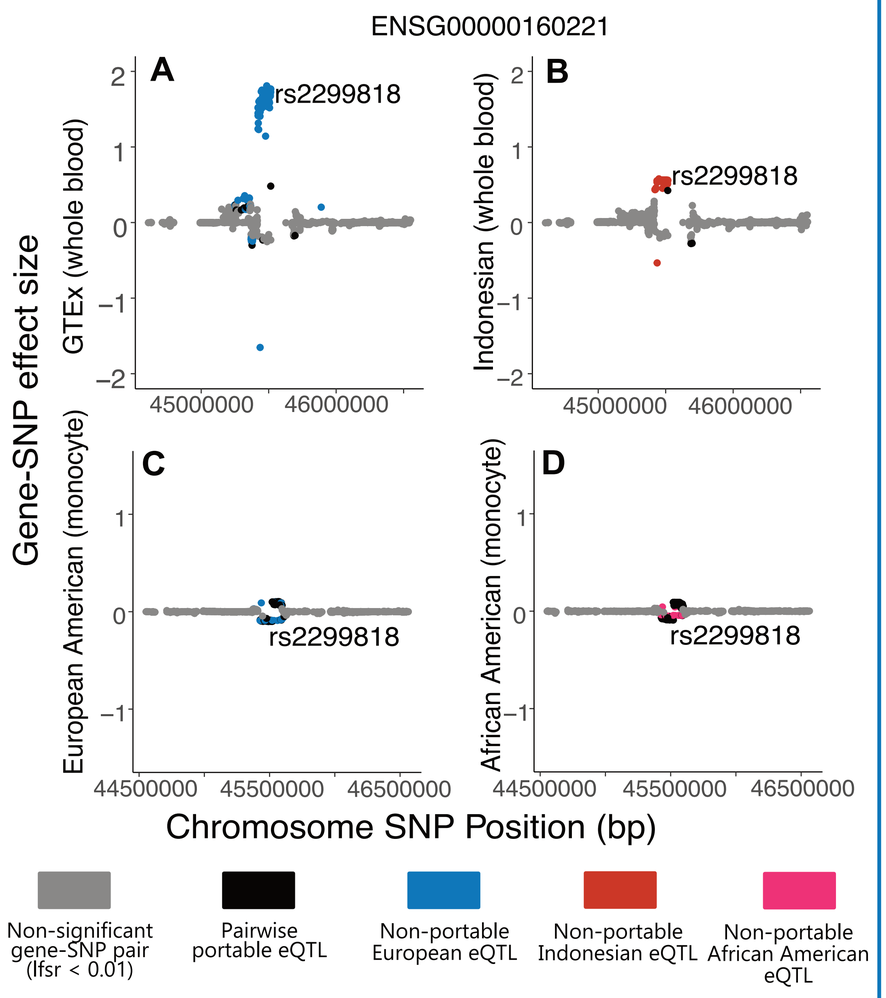


Fig 3. rs2299818_ENSG00000160221 is a tissue dependent non-portable eQTL

GTEx 2020 had previously characterised this eQTL as being 'population-biased' (having different effect sizes) between European American and African American populations in whole blood. We replicate this bias between Europe and Indonesia in whole blood (A-B), but find no such bias in monocyte cells between European American and African American Mogil et al. study populations (C-D).

Conclusions and future work

- The incomplete portability of eQTLs from European to under-served populations precludes the equitable translation of genomics research
- **Patterns in the cross-population portability of eQTLs are complex, and context-specific**
- Developing precision medicine which benefits all will thus rely on the ability of modern statistical techniques such as machine learning to capture complex, non-linear patterns
- **Future areas of expansion include testing alternative model interpretation to confirm feature patterns detected by LOFO and expanding the datasets included in our analysis (suggestions welcomed!)**

Acknowledgements:

Thanks for all the support and advice: Davis McCarthy, the BioCellGen Group at SVI and the Gallego Romero Group at the University of Melbourne!

References:

1. Fatumo, Segun, Tinashe Chikowore, Ananyo Choudhury, Muhammad Ayub, Alicia R. Martin, and Karoline Kuchenbaecker. 2022. "A Roadmap to Increase Diversity in Genomic Studies." Nature Medicine, February, 1–8.

2. Landry, Latrice G., Nadya Ali, David R. Williams, Heidi L. Rehm, and Vence L. Bonham. 2018. "Lack Of Diversity In Genomic Databases Is A Barrier To Translating Precision Medicine Research Into Practice." Health Affairs 37 (5): 780–85.

3. Umans, Benjamin D., Alexis Battle, and Yoav Gilad. 2021. "Where Are the Disease-Associated eQTLs?" Trends in Genetics: TIG 37 (2): 109–24.

4. Natri, Heini M., Georgi Hudjashov, Guy Jacobs, Pradiptajati Kusuma, Lauri Saag, Chelzie Crenna Darusallam, Mait Metspalu, et al. 2022. "Genetic Architecture of Gene Regulation in Indonesian Populations Identifies QTLs Associated with Global and Local Ancestries." American Journal of Human Genetics 109 (1): 50–65.