



Identifying signals of natural selection in unadmixed individuals of the Mexican Biobank

Fernanda Miron-Toruño¹, Ram González-Buenfil², Santiago Medina-Muñoz², Andrés Moreno-Estrada², Austin W. Reynolds¹

¹Department of Anthropology, Baylor University; ²Unidad de Genómica Avanzada, Cinvestav, Mexico

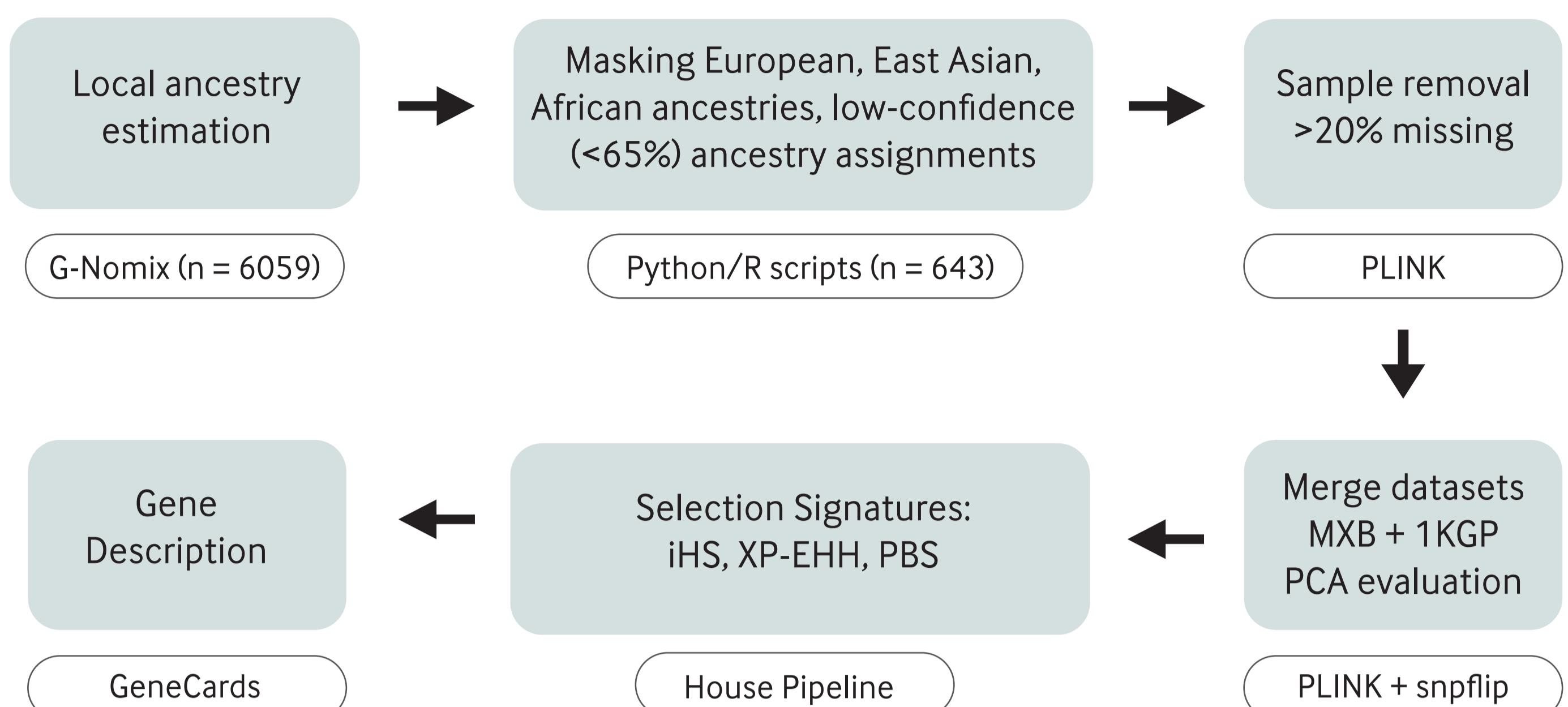


INTRODUCTION

Selection signatures studies represent a genomic approach to understand **human evolution** and differences in **disease risk** across populations worldwide. However, the limited availability of diverse cohorts with genomic data impairs the effective implementation of selection analyses in many underrepresented regions, such as Latin America.

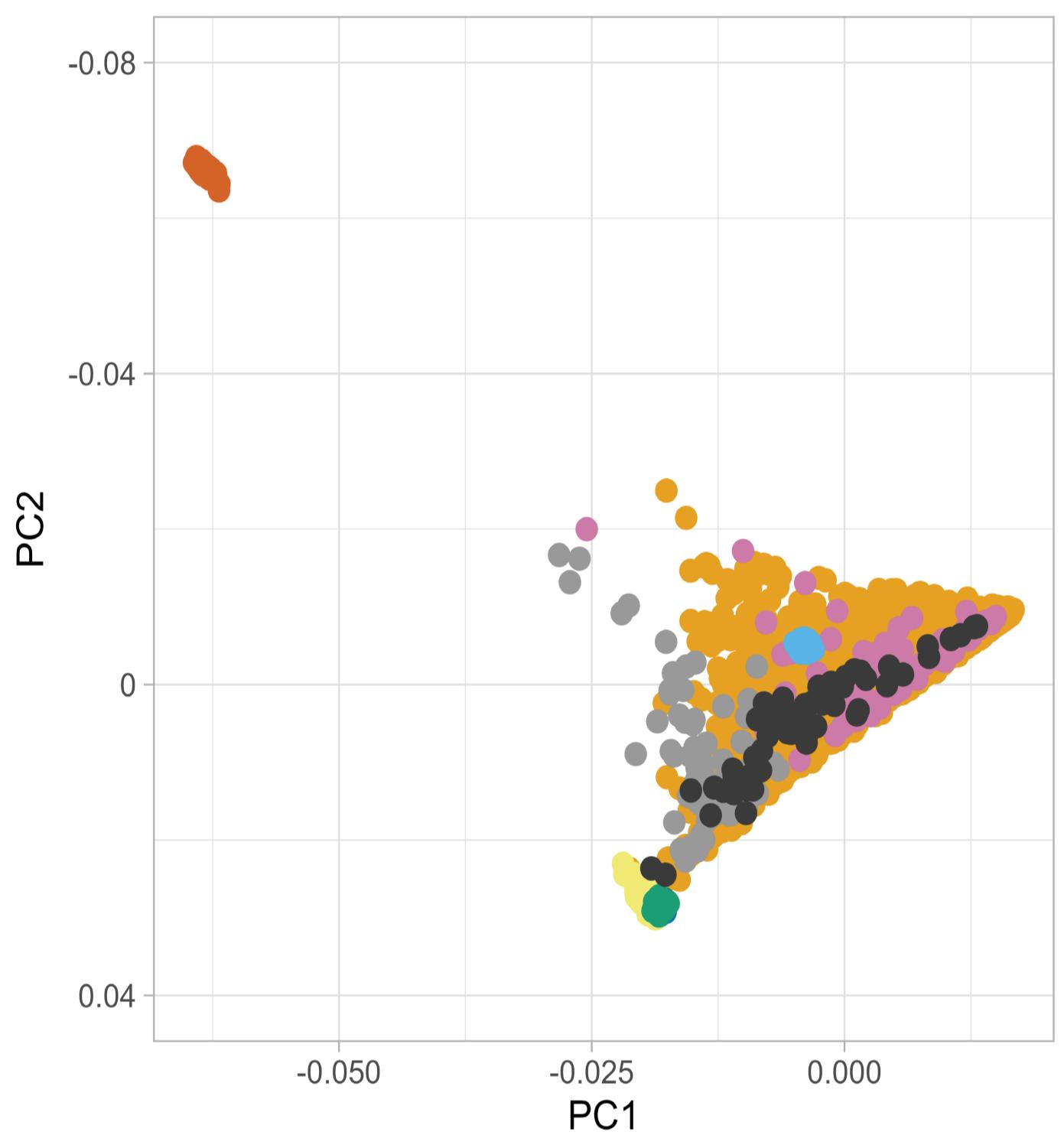
The **Mexican Biobank (MXB)** is one of the largest genomic cohorts in the Americas, and comprises 6,000 individuals from all 32 states. Here, investigate the genetic footprint of recent adaptive pressures on Indigenous Mexican populations by analyzing a subset of ancestry-masked samples with high Indigenous American ancestry from the MXB.

METHODOLOGY



MASKING RESULTS

A) Pre-masking



B) Post-masking

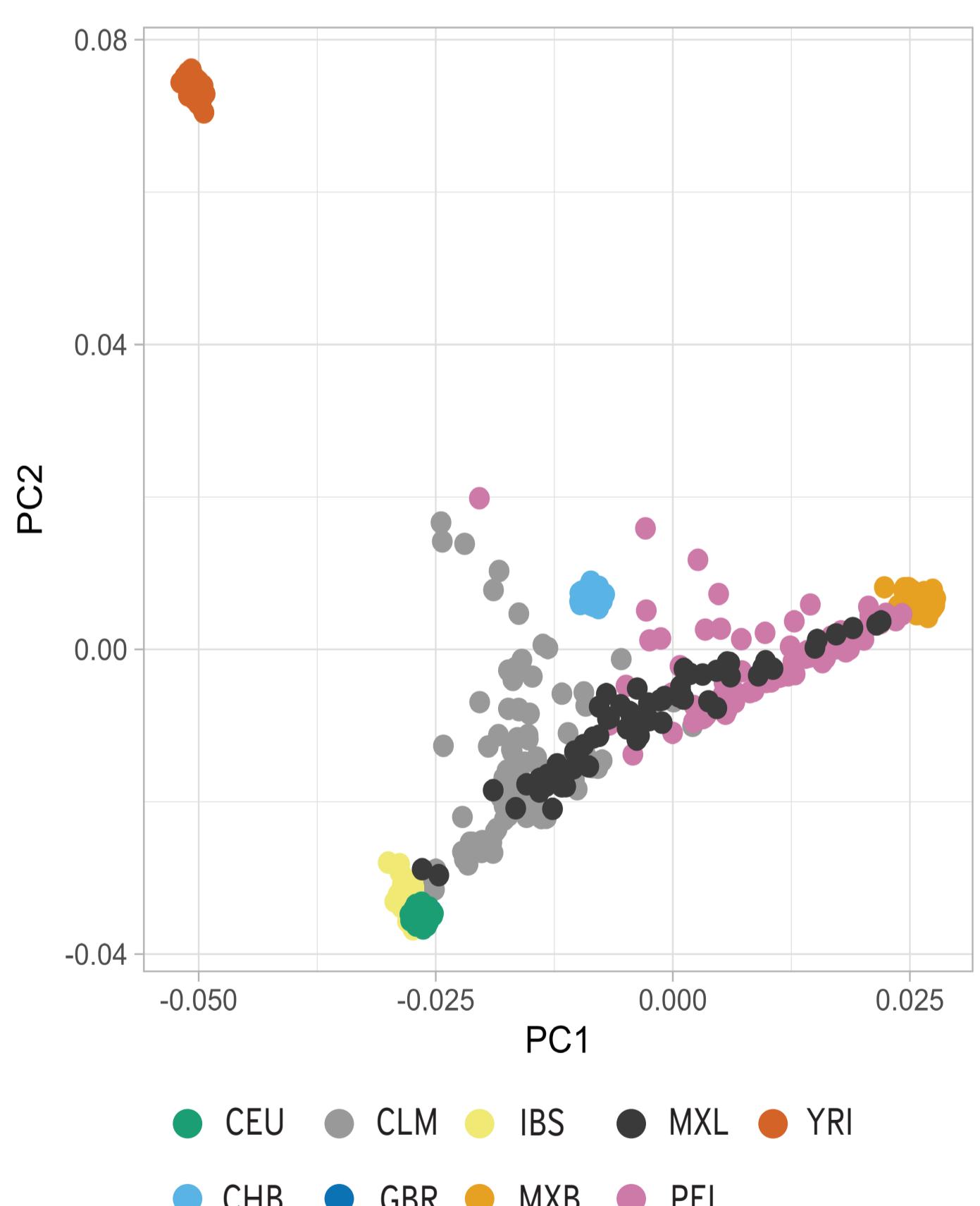


FIGURE 1. PCA plot. A) Population structure before genotype masking (n MXB = 6059). B) Population structure after genotype masking of European, East Asian, African ancestries, and low-confidence (<65%) ancestry assignments (final n MXB = 643).

SELECTION RESULTS

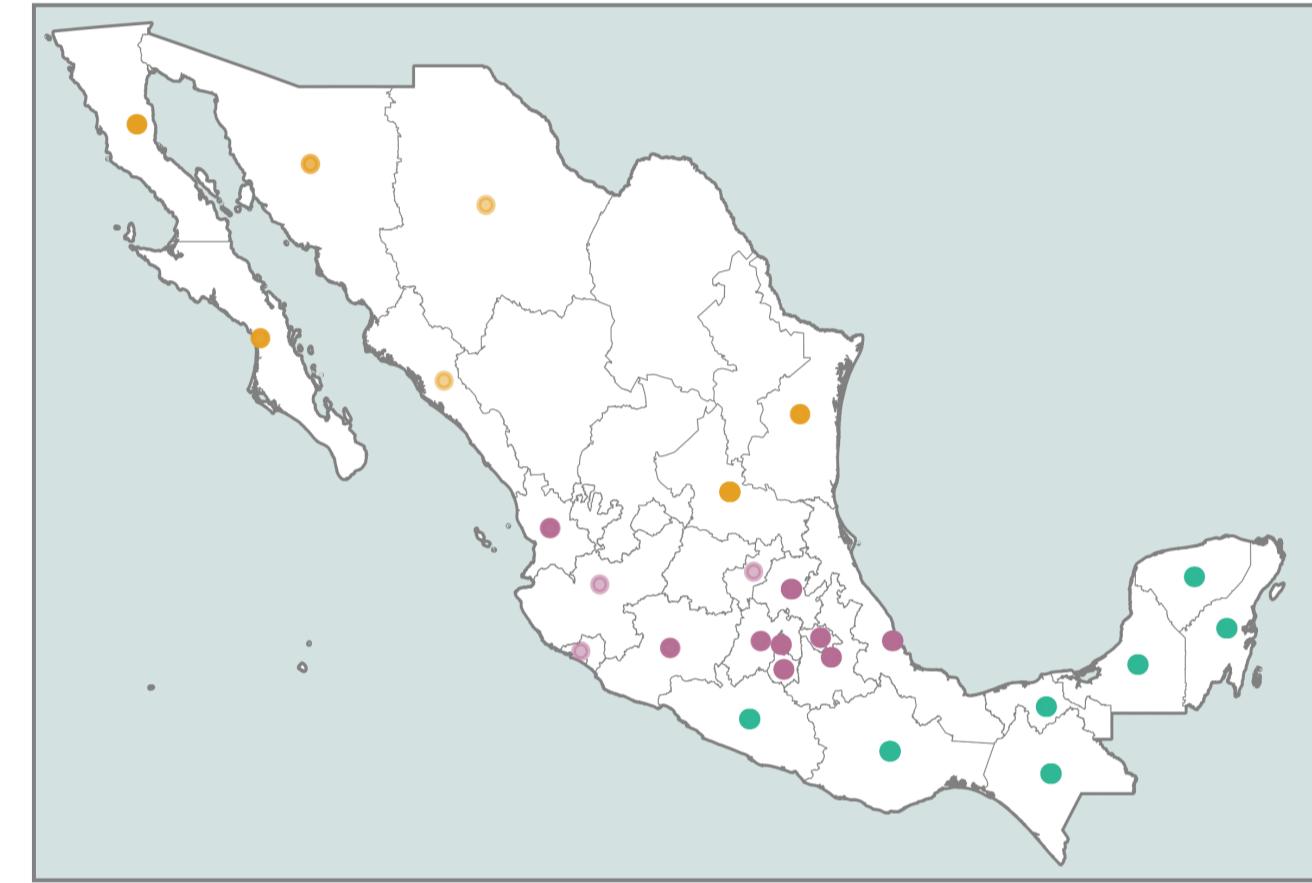
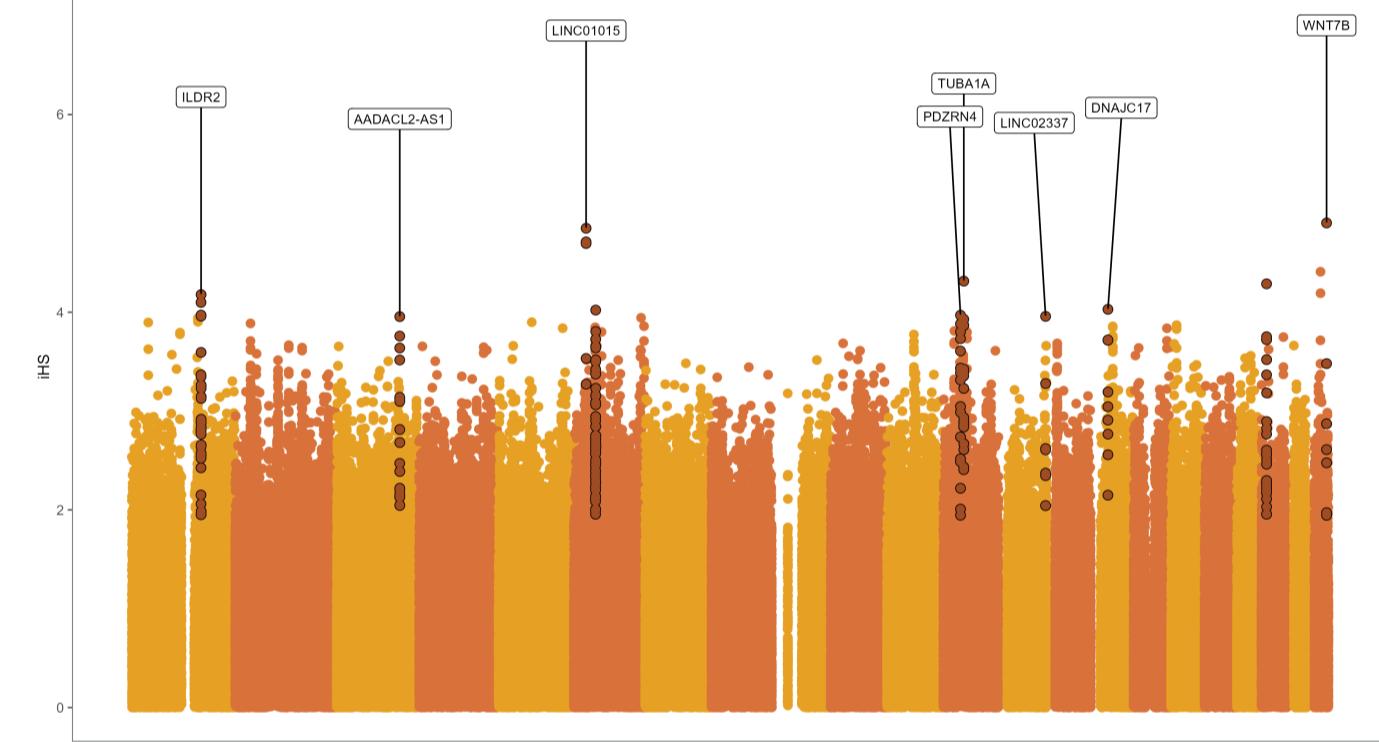
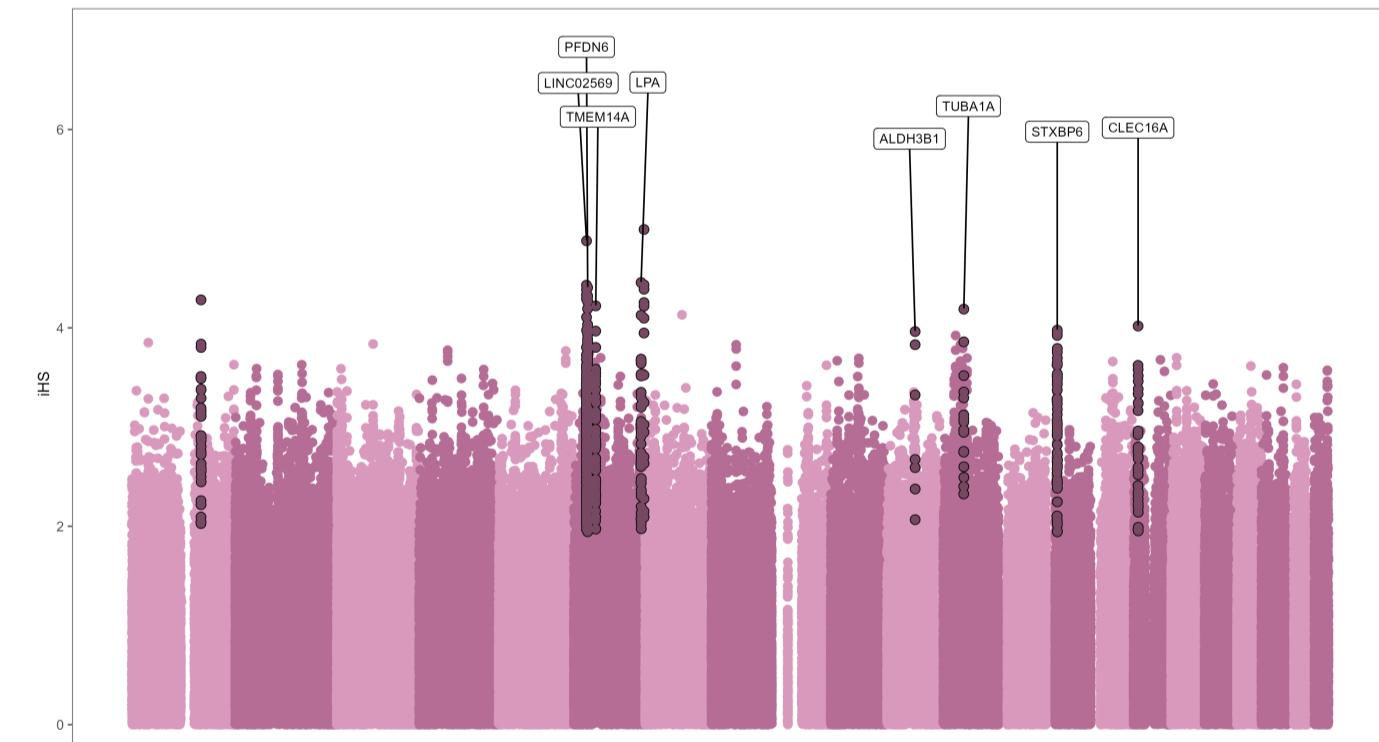


FIGURE 2. Sample distribution of Northern (Yellow, n = 85), Center (Pink, n = 197), and Southern (Green, n = 361) clusters.

A. Northern Mexican Populations



B. Center Mexican Populations



C. Southern Mexican Populations

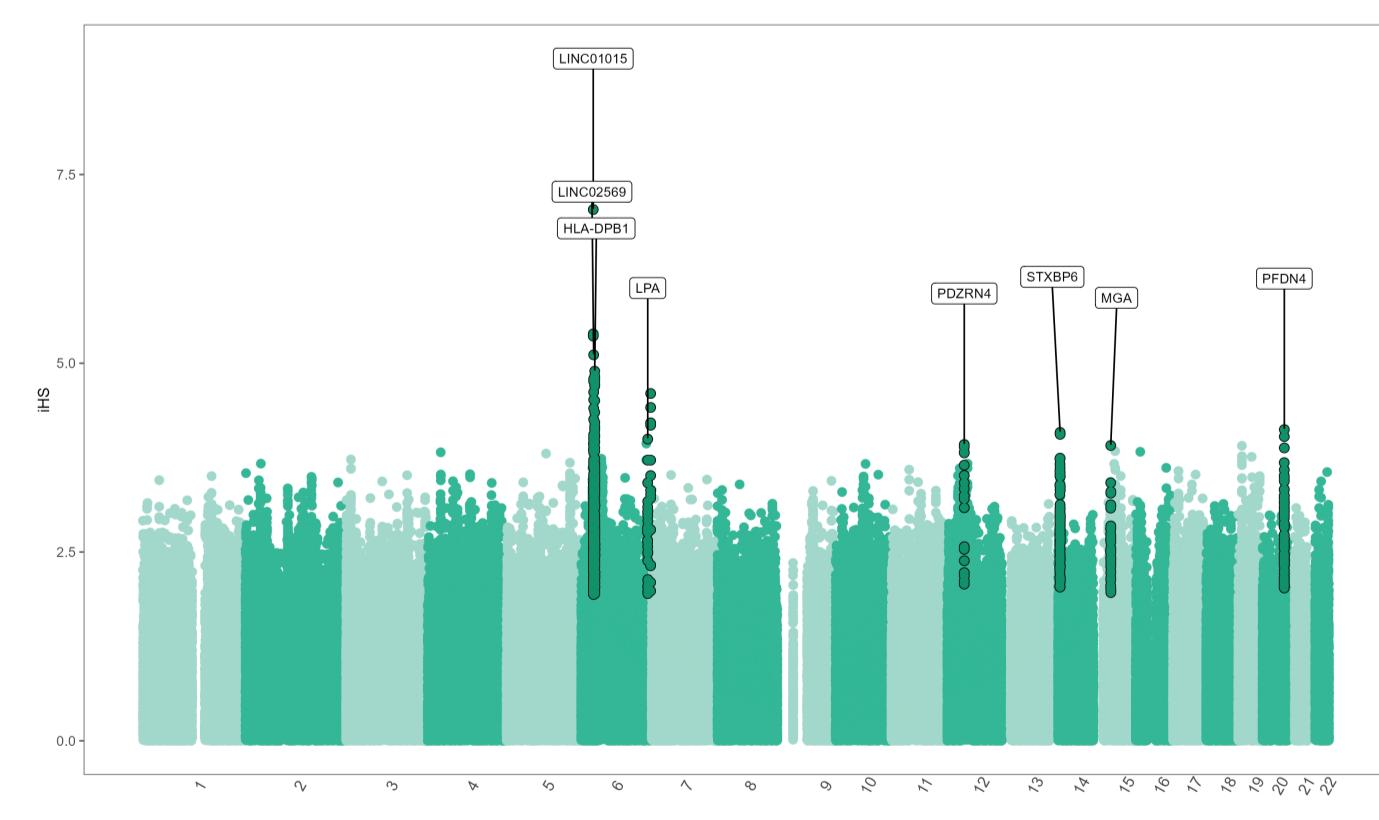


FIGURE 3. Manhattan plots for the three selection statistics: iHS, XP-EHH (MXB-CHB), PBS (MXB, ingroup = PEL, outgroup = CHB). Strongest signals per statistic from top 10 entries are highlighted. A) Northern Mexican Populations. B) Center Mexican Populations. C) Southern Mexican Populations

	iHS	GWAS Catalog*	XP-EHH	GWAS Catalog*	PBS	GWAS Catalog*
Northern	WNT7B	Erythrocyte count, hemoglobin measurement, red blood cell density measurement, hematocrit	ILDR2	Lipid homeostasis and insulin secretion, neutrophil count, brain measurement	ARL3	Cytokinesis and cilia signaling
Center	LINC02569	Long Intergenic Non-Protein Coding RNA	ILDR2	Lipid homeostasis and insulin secretion, neutrophil count, brain measurement	IFITM4P	Lymphocyte count, eosinophil count, basophil count, hemoglobin measurement
Southern	HLA-DPB1	Immune system function, COVID-19, rheumatoid arthritis, type 1 diabetes mellitus, body height	TMEM14A	Triglyceride measurement, low density lipoprotein cholesterol measurement, high density lipoprotein cholesterol measurement, total cholesterol measurement	OR4C5	Odorant receptor, intraocular pressure measurement, systolic blood pressure, open-angle glaucoma

TABLE 1. Strongest signals per statistic from top 10 entries. *GWAS catalog phenotypes retrieved from GeneCards

CONCLUSION

Our preliminary analyses suggest a selection landscape **unique** to each cluster. However, signatures of selection that are shared between all the clusters were found in the XP-EHH results (**ILDR2**, **MUC17**, **HLA-DPB1**, **TMEM14A**). Phenotypes under putative selection are related to blood traits, lipid metabolism, immune function, and olfactory function.

FUTURE DIRECTIONS

- Composite scores (Fisher combined scores)
- Demographic models
- Connect putatively selected alleles to molecular phenotypes reported in the **MXB**

BIBLIOGRAPHY

1. Sohail, Mashaal, et al. "Mexican Biobank advances population and medical genomics of diverse ancestries." *Nature* 622.7984 (2023): 775-783.
2. Yi, Xin, et al. "Sequencing of 50 human exomes reveals adaptation to high altitude." *science* 329.5987 (2010)
3. Voight, Benjamin F., et al. "A map of recent positive selection in the human genome." *PLoS biology*
4. Sabeti, Pardis C., et al. "Genome-wide detection and characterization of positive selection in human populations." *Nature* 449.7164 (2007): 913-918.
5. Ávila-Arcos, María C., et al. "Population history and gene divergence in Native Mexicans inferred from 76 human exomes." *Molecular biology and evolution* 37.4 (2020): 994-1006