Research statement

With the advent of biobank-scale genomic datasets, the field of human genetics has scrambled to develop statistical methods for studying millions of genetic variants across hundreds of thousands of individuals. At present, the field is saturated with methods that seek to understand how genetic variation influences complex traits, but existing methods are largely unconcerned with the way in which that variation arises. While this approach has been sufficient for answering some questions, deeper insights about complex trait biology require us to consider the evolutionary forces that constrain genetic variation.

My research program integrates evolutionary modeling, statistical methods development, and empirical data analysis to study the **genetic variation underlying complex traits**, and how this variation is distributed **across populations**. As genomic datasets diversify, the time is ripe for developing methods rooted in evolutionary theory that are capable of combining information across global populations. At the same time, there are important ethical considerations that arise when working with data from populations that have been historically excluded or exploited within biomedical research. Thus in parallel, my research program also studies the **ethical and social implications** of human genetics. Below, I outline my previous work and future research goals across three key areas of interest.

Evolution-informed methods for biobank-scale analyses

In one line of work, I have developed novel methods for studying human complex traits through work that builds on both population genetics and phylogenetics—two fields that study the evolutionary processes shaping genetic variation.

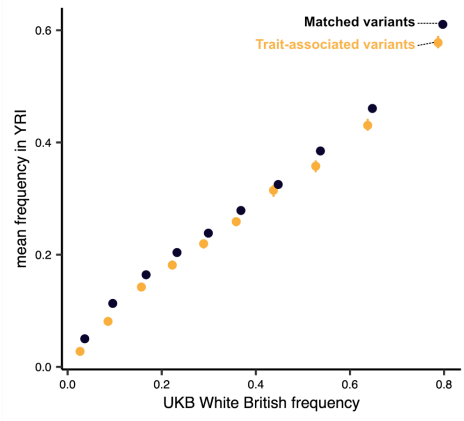

Figure 1: Trait-associated variants have systematically lower frequencies in YRI, conditional on their frequencies in the GWAS dataset—evidence that selection constrains complex trait variation.

During my PhD, I developed a method to understand which kind of selection acts on human complex traits. Competing models of selection were historically difficult to resolve because genetic association studies are only powered to ascertain variants with large frequencies or large effects—thereby distorting the exact information one would use to study selection. I devised a population-genetic statistic that circumvents this limitation by conditioning on the ascertained frequencies and incorporating data from additional populations [1]. In empirical data, I found clear evidence that stabilizing or purifying selection has been the dominant force shaping the genetic variation underlying complex traits (Fig. 1).

In ongoing postdoctoral work, I am developing methods to reduce confounding in association studies. Epidemiological association studies test for an association between a particular exposure and outcome – for example, the impact of BMI on asthma – but these studies are highly susceptible to confounding if both exposure and outcome are themselves complex traits influenced by genetics. I have shown that this problem is equivalent to a well-studied problem in phylogenetics: testing for the association between two organismal

traits, while correcting for shared phylogeny. By incorporating insights from phylogenetics, I have developed a flexible and accurate method to assess and control genetic confounding in epidemiological association studies.

Future directions. Defined as the proportion of trait variation attributed to genetics, the heritability of a trait is a fundamental concept. Trait heritability will impact the effectiveness of genetic association testing and the accuracy of phenotype predictions made from genetic data. However, existing methods for biobank-scale analyses of complex traits often rely on rigid and unrealistic assumptions about heritability. The consequences of these assumptions are under-explored and likely introduce biases in genetic association testing and phenotype prediction, particularly in analyses of multiple populations. My research group will develop methods that are grounded in a more realistic, evolution-informed model of heritability.

Substantial theoretical work has characterized the heritability of complex traits under stabilizing selection, but little work has been done to understand how stabilizing selection will impact heritability across populations under realistic human demography. Yet, my previous work has shown that certain demographic scenarios can exacerbate or diminish the genomic impacts of selection [1]. Using population-genetic simulations, my research group will explore how demography and stabilizing selection jointly impact the distribution of trait heritability across human populations. Our theoretical work will enable us to assess sources of bias in existing methods and propose more accurate models of heritability for use in multi-population analyses.

Genetic basis of gene expression

In another line of work, I study the genetic basis of gene expression. Gene expression serves as a bridge between genetic variation and trait manifestation, helping us understand the mechanisms by which variants influence complex traits.

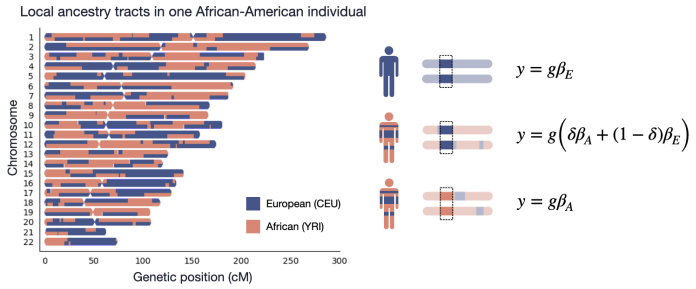

Figure 2: I developed a statistical model that tests for genetic interactions by comparing the effect of genetic variants across segments of European ancestry in European and admixed African-American individuals.

During my PhD, I explored the extent to which the expression of genes is affected by interactions between genetic variants, or between variants and environmental factors. An alarming dissonance between molecular biology and statistical genetics is that the former tells us that biological pathways consist of complex interactions between small molecules, while the latter has generally found little evidence of genetic interactions. I developed a method to test for the presence of genetic interactions by leveraging the genomic patterns conferred by recent admixture (Figure 2). I found evidence for pervasive genetic interactions, confirming that earlier studies were simply underpowered [2].

Future directions. My work has demonstrated that complex traits are subject to stabilizing selection [1]. It remains unclear what this means for intermediate traits – such as gene expression – which give rise to complex traits but are more distantly linked to fitness. My research group will explore how stabilizing selection impacts the genetic variation underlying gene expression.

One challenge in studying selection on gene expression is the smaller sample size associated with human gene expression datasets – usually numbering in the hundreds of individuals, rather than hundreds of thousands. My research group will take advantage of the peculiar structure of gene expression: nearby mutations usually influence expression directly and have larger effects relative to distant mutations that influence expression indirectly via other genes. Using population-genetic simulations, my research group will characterize how stabilizing selection impacts the distribution of effects across these two mechanisms of action. We will apply the intuition from our theoretical work to develop statistical methods that differentially make use of local and global variation to estimate selection on gene expression.

Ethical, legal, and social implications

Human genetics research has profound biomedical and societal impacts, including both benefits and harms. In recent years, these impacts have come under scrutiny, especially as research in these fields has witnessed a surge in co-option and weaponization by race scientists and white supremacists. To conduct more socially responsible science, it is necessary to understand both the rationale behind and impacts of scientific practices in human genetics. Doing so will require interdisciplinary scholarship—a literacy with the methods and models in these fields, as well as an ability to grapple with the ethical implications.

During my PhD, I led the development of a graduate-level course on the ethical and societal implications of human genetics research. In collaboration with bioethicists, I published a paper describing how a targeted ethics curriculum could enable scientists to conduct more socially responsible research [3].

Future directions. My research group will explore the rationale behind and impacts of using race and ancestry in genetic studies. The use of race and ancestry is a complex topic—so much so that the National Academies recently issued a 200-page guide on using population descriptors in human genetics research. Nevertheless, the use of race and ancestry remains wildly inconsistent, and many ethicists have warned that conflating race and ancestry risks promoting racial essentialism. Using scientometrics and mixed methods analysis, my research group will seek to understand (1) how and why scientists use race and ancestry in genetic studies, (2) how scientists conceptualize social benefits and harms vis-à-vis the use of race and ancestry, and (3) which resources and strategies are effective for enabling more socially responsible use of race and ancestry. This work is especially timely as genomic biobanks become increasingly global, requiring researchers to integrate disparate cultural frameworks for population descriptors.

References

1. **Roshni A. Patel**, Clemens L. Weiss, Huisheng Zhu, Hakhamanesh Mostafavi, Yuval B. Simons, Jeffrey P. Spence, and Jonathan K. Pritchard (2024). Conditional frequency spectra as a tool for studying selection on complex traits. *bioRxiv* (accepted at *Genetics*).
2. **Roshni A. Patel**, Shaila A. Musharoff, Jeffrey P. Spence, *et al.* (2022). Genetic interactions drive heterogeneity in causal variant effect sizes for gene expression and complex traits. *American Journal of Human Genetics 109*: 1286-1297.
3. **Roshni A. Patel**, Rachel A. Ungar, Alanna L. Pyke, *et al.* (2024). Increasing equity in science requires better ethics training: a course by trainees, for trainees. *Cell Genomics 4*: 100554.