Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences

Noah A. Rosenberg,1 Michael D. Edge2, Jonathan K. Pritchard1,3,4 and Marcus W. Feldman1

1Department of Biology, Stanford University, Stanford, CA 94305, USA; 2Department of Evolution and Ecology, University of California, Davis, CA 95616, USA; 3Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA and 4Department of Genetics, Stanford University, Stanford, CA 94305, USA

*Corresponding author. Department of Biology, Stanford University, Stanford, CA 94305, USA. Tel: +1-650-724-5122; Fax: +1-650-724-5114; E-mail: noahr@stanford.edu

Received 21 June 2018; revised version accepted 21 December 2018

ABSTRACT
Recent analyses of polygenic scores have opened new discussions concerning the genetic basis and evolutionary significance of differences among populations in distributions of phenotypes. Here, we highlight limitations in research on polygenic scores, polygenic adaptation and population differences. We show how genetic contributions to traits, as estimated by polygenic scores, combine with environmental contributions so that differences among populations in trait distributions need not reflect corresponding differences in genetic propensity. Under a null model in which phenotypes are selectively neutral, genetic propensity differences contributing to phenotypic differences among populations are predicted to be small. We illustrate this null hypothesis in relation to health disparities between African Americans and European Americans, discussing alternative hypotheses with selective and environmental effects. Close attention to the limitations of research on polygenic phenomena is important for the interpretation of their relationship to human population differences.

KEYWORDS: adaptation; health disparities; human variation; polygenic scores; population genetics

INTRODUCTION
We are currently witnessing a surge in public interest in the intersection of evolutionary genetics with such topics as cognitive phenotypes, disease, race and heritability of human traits [1–7]. This attention emerges partly from recent advances in genomics, including the introduction of polygenic scores—the aggregation of estimated effects of genome-wide variants to predict the contribution of a person’s genome to a phenotypic trait [8–10]—and a new focus on polygenic adaptations, namely adaptations that have occurred by natural selection on traits influenced by many genes [11–13].
Theories involving natural selection have long been applied in the scientific literature to explain mean phenotypic differences among human populations [14–16]. Although new tools for statistical analysis of polygenic variation and polygenic adaptation provide opportunities for studying human evolution and the genetic basis of traits, they also generate potential for misinterpretation. In the past, public attention to research on human variation and its possible evolutionary basis has often been accompanied by claims that are not justified by the research findings [17]. Recognizing pitfalls in the interpretation of new research on human variation is therefore important for advancing discussions on associated sensitive and controversial topics.

**COMPLEX PHENOTYPES AND POLYGENIC SCORES**

Over the past 15 years, genomic analyses have identified thousands of genetic variants that contribute statistically to variation in complex phenotypes, traits that have complex patterns of inheritance and that are affected by large numbers of genes in combination with environmental factors [18–20]. In a typical genomic study of a complex human phenotype—a genome-wide association study (GWAS)—genotypes at thousands or millions of sites across the human genome are each tested in a sample of people for statistical association with the phenotype. Each variant identified by such a study as statistically associated with the phenotype can be assigned an effect size, representing the estimated magnitude of the increase in the trait (for quantitative phenotypes) or risk or liability for the trait (for binary phenotypes) that is associated with possession of a copy of the variant.

For many complex phenotypes, identification and analysis of contributing genomic variants—most having small phenotypic effects—has led to the formulation of polygenic scores, quantities that seek to predict a trait value associated with a specific genome-wide set of genotypes [10]. For a quantitative phenotype, a polygenic score for an individual genome represents an aggregation, usually in the form of a sum, of the estimated effect sizes of the genetic variants in the genome (Table 1). Polygenic score estimation of underlying genetic propensities typically proceeds from GWAS outcomes.

Polygenic scores have provided new tools for interpreting human genomes in the setting of complex phenotypes for which effect sizes of genetic variants are small. For example, they have contributed new approaches to risk prediction for adverse phenotypes related to heart disease [21–23]. Using polygenic scores, it is now possible to combine information from millions of genomic variants to identify people whose overall polygenic risk of coronary artery disease is as high as that of patients with monogenic lipid disorders such as familial hypercholesterolemia [23]. Although for many genetically complex phenotypes, polygenic scores currently explain too small a fraction of variation in the phenotype to be clinically meaningful, such risk calculations contribute to the promise of the genomic era to produce actionable predictions about complex phenotypes on the basis of the accumulation of many small genomic contributions [24].

**GENETIC BASIS OF POPULATION DIFFERENCES IN COMPLEX PHENOTYPES**

The genetic underpinnings of population differences in phenotype distributions have been of perennial interest in human genetics, and the use of polygenic scores promises to generate progress in understanding phenotypic differences among populations. However, interpretation of such differences in relation to polygenic score differences requires careful analysis. Although distributions of individual-level polygenic scores might differ among populations, differences in these distributions might have many potential causes, and might or might not reflect meaningful biological phenomena.

The main novelty in analyses of polygenic score distributions among populations is that many trait-associated genetic variants that were previously unknown are now known. Earlier studies of the role of genetics in phenotypic differences among populations often relied on statistics such as heritabilities—fractions of phenotypic variance explained by genetic variation [25]—which require no knowledge of contributing genetic variants. Although estimates of the contributions of specific genetic variants advance modern analyses beyond classical heritability statistics, many of the pre-genomic-era limitations on the use of heritability to make inferences about the genetic basis of phenotypic differences among populations [26–28] continue to apply, in updated form (Fig. 1). Limitations in the interpretation of polygenic score differences can be of two kinds: those due to the manner in which genes and environment contribute to traits, irrespective of statistical issues involved in estimating the contributions from data, and those due to statistical phenomena in the estimation process.

**Conceptual limitations**

First, population differences in environmental factors are important for interpreting population differences in a phenotype distribution. Depending on environmental contributions, a difference in mean phenotype might or might not reflect a difference in the magnitude of genetic effects among populations; population differences in phenotype distributions do not reveal which population has greater genetic propensities on average, whether the observed difference would persist if the environment were altered, or even whether a difference in genetic propensities exists at all (Fig. 1A–C). That polygenic scores ignore the role of environmental influences on phenotype is particularly relevant when the phenotype can be readily modified, as in the use of statins as medications to control the lipid levels that contribute to coronary artery disease risk. In such cases, a difference in polygenic score—though possibly a genuine reflection of underlying genetic propensities—might be incorrectly inferred to represent an
unchangeable genetic difference among populations rather than one that can be altered by an environmental change (Fig. 1D). Instead, the potential for significant modification of the environmental contribution renders polygenic score differences between populations largely unconnected to population differences in phenotype distributions.

Second, gene–gene and gene–environment interactions influence traits, meaning that associations between specific genotypes and a phenotype—and the importance of the genetic contributions—might differ among populations with different allele frequencies or distributions of environmental variables. In other words, because the contributions of genomic variants can differ among populations due to interactions with other variants and with environmental variables, the effects of a variant on a trait can have different magnitudes in different populations, and effects of multiple variants in one or more genes can combine in different ways. Large population differences in disease risk for well-known alleles such as APOE-ε4 in Alzheimer disease [29, 30] highlight the challenge of determining how population differences in effect size might be affected by interaction effects.

Finally, mean differences between populations in polygenic scores might not be informative for prediction about phenotype differences between randomly chosen people from a pair of populations if polygenic score distributions have substantial overlap (Fig. 1A, C and D). In such cases, predictive potential is limited even if a difference in population means is seen to be statistically significant in the large sample sizes typical of genome-wide association studies.

**Statistical limitations**

Beyond the conceptual challenges, which are inherent in interpreting population differences in polygenic scores, the process of estimating a difference in the scores themselves is subject to additional limitations. Genotypic effects estimated only in one population might not apply to other populations for a number of reasons. Effect estimates might rely on sites that were ascertained for variability in one set of populations and whose systematic differences in allele frequencies between populations contribute to systematic biases in polygenic score estimates in...
other populations [31]. These estimates might also fail to consider many sites variable only in those other populations.

A third limit to transferability of effect estimates arises from population differences in features of correlations between nearby sites—linkage disequilibrium patterns—that influence aggregations of signals across loci [32]. So far, because most genome-wide association studies have been conducted in populations with European ancestry [33], the effect sizes used in calculating polygenic scores have been calibrated on Europeans, and their values might not transfer accurately to other
populations. Even among populations with European ancestry, subtle ancestry differences between samples can lead to polygenic scores that overstate between-population differences: small biases in locus-wise effect estimates that arise from the ancestry differences can potentially accumulate across loci [34, 35].

Summary
These limitations illustrate that much of the complexity embedded in use of polygenic scores—the effects of the environment on phenotype and its relationship to genotype, the proportion of variance explained, and the peculiarities of the underlying GWAS data that have been used to estimate effect sizes—is obscured by the apparent simplicity of the single values computed for each individual for each phenotype. Consequently, in using polygenic scores to describe genomic contributions to traits, particularly traits for which the total contribution of genetic variation to trait variation, as measured by heritability, is low—but even if it is high (Fig. 1E)—a difference in polygenic scores between populations provides little information about potential genetic bases for trait differences between those populations.

Unlike heritability, which ranges from 0 to 1 and therefore makes it obvious that the remaining contribution to phenotypic variation is summarized by its difference from 1, the limited explanatory role of genetics is not embedded in the nature of the polygenic scores themselves. Although polygenic scores encode knowledge about specific genetic correlates of trait variation, they do not change the conceptual framework for genetic and environmental contribution to population differences. Attributions of phenotypic differences among populations to genetic differences should therefore be treated with as much caution as similar genetic attributions from heritability in the pre-genomic era.

POLYGENIC ADAPTATION
Genomic investigations have provided insights into how natural selection has given rise to differences in phenotypes that vary geographically, such as skin pigmentation, lactase persistence and altitude-related physiology [14–16]. Success in these well-known examples, each involving natural selection primarily on one or a few genes of large effect, has encouraged the search for other phenotypes that might have experienced different histories of natural selection in different populations. Recent interest focuses on traits such as height [12, 36, 37] that are influenced by large numbers of genetic variants, each having a small effect on the trait, and that lend themselves to analysis with polygenic scores.

The null expectation
Evidence that natural selection has contributed to population differences in some specific traits can invite claims that it has also influenced phenotypic differences and underlying genetic differences in other traits. It might be hypothesized, for example, that a population with a higher mean trait value has experienced selection favoring the high value, and perhaps also that selection has favored a lower value in a second population. This type of hypothesis might entail that the difference in phenotype distributions in Fig. 1A–C results from genetic propensity differences between populations that follow the same direction as the phenotype, as in Fig. 1A but not in Fig. 1B or C, and that those distributions reflect natural selection rather than selectively neutral evolutionary processes. The hypothesis might appear to derive support from the fact that sufficient genetic variation exists among populations to infer the ancestral populations of individual genomes at a local geographical scale [38–40], and the genetic differences evident from ancestry inferences might then be attributed to natural selection. However, the inference from the existence of differences in trait distributions between two populations that natural selection has acted to produce genetic differences between those populations requires several careful steps [41].

One key component of the inference of polygenic adaptation is the use of an appropriate null expectation for polygenic score distributions and phenotypic differences [12, 42]. In deriving such an expectation, an important insight from selectively neutral population-genetic models is that irrespective of the number of genetic loci contributing to a polygenic trait, the expected difference among populations in the trait is predicted to have comparable magnitude to the classical estimate of the ‘apportionment of human genetic diversity’, the extent of human genetic difference at a single randomly chosen polymorphic genetic locus [12, 43, 44]. In other words, analogous measures of population differences in quantitative phenotypic traits and genetic loci—termed \( Q_{ST} \) and \( F_{ST} \), respectively—are approximately equal in neutral evolutionary models that include genetic drift but not natural selection. Because many loci contribute to a quantitative trait, and each locus experiences the same random process of genetic drift independent of the size and direction of its trait contribution, phenotypic differences among populations are predicted under neutrality to be similar in magnitude to typical genetic differences among populations.

The genetic apportionment computation shows that genetic differences among populations, as measured by \( F_{ST} \), are small in comparison with variation within populations [45–47]. Although the among-population variation suffices to infer ancestral populations from individual genomes, analysis of models for the genetic basis of phenotypes finds that under neutrality, the magnitude of phenotypic differences connects to the apportionment computation rather than the ancestry computation [42, 44].

Selection or environmental effects?
Departures from the null expectation for phenotypic differences among populations can be due to a combination of (i) population
Interpreting polygenic scores

Rosenberg et al. | 31

Selection on a correlated trait might occur in different directions in a pair of populations or with different magnitudes in the same direction, and therefore need not increase genetic differences between populations in the way that divergent selection for the initial trait might suggest (Fig. 2).

For these reasons, an observed between-population difference in phenotype distributions is not easily ascribed to divergent selection. Indeed, a challenge is to establish whether polygenic adaptation has even occurred. In within-population polygenic adaptation tests, for loci across the genome, GWAS-based locus effect sizes are considered with selection signals estimated for those loci. An aggregate signal of positive selection at loci with large effect sizes is taken to suggest that selection has inflated the frequencies of alleles contributing to the trait, so that the trait has undergone polygenic adaptation. Recent studies of height have suggested that polygenic adaptation tests are sensitive to the choice of GWAS data that provide the effect sizes: even if two sets of effect sizes produce correlated polygenic scores, effect sizes estimated from one study can generate erroneously exaggerated signatures of polygenic adaptation when assessing polygenic adaptation in a second dataset [34, 35]. This result, which arises from subtle population differences between study samples, calls into question claims about polygenic adaptation even of traits for which it has been most extensively investigated.

**Summary**

To date, strong effects of directional selection on human population differences have been verifiable primarily for traits connected to predictable categories of geographic variability, including dietary adaptations, infectious disease resistance and skin pigmentation [14–16]. As speculations about features of natural selection in human populations proliferate, hypotheses about selection on specific phenotypes should not be treated as being as plausible a priori as a general null population-genetic model of polygenic similarity among populations. Dramatic claims about divergent selection should continue to be regarded cautiously in the absence of strong quantitative evidence.

**THE CASE OF HEALTH DISPARITIES**

Health disparities between African Americans and European Americans in the USA provide a useful case for examining genetic and environmental contributions to phenotypic differences among populations. In a study of African Americans and European Americans (treated as socially rather than genetically defined populations), among 36 physiologically diverse causes of death, adjusting for other factors, African Americans lost more years of life than European Americans in 28 of the 36 [49]. In the simplest null model in which many genes contribute to a trait chosen at random, with no directional selection, each of a pair of groups has probability 0.5 of having the larger mean value for differences in environmental influences on phenotypes, and (ii) differential natural selection among populations that generates a substantial population difference in polygenic score distributions. However, only with strong directional selection on a trait in one population, or strong directional selection in opposite directions in a population pair, is a phenotypic difference between populations attributable largely to natural selection. In other words, because of environmental effects, the difference in phenotype distributions in Fig. 1A–C need not reflect a parallel difference in genetic propensities as in Fig. 1A, but rather no difference as in Fig. 1B or a difference opposite in direction as in Fig. 1C; even a parallel difference as in Fig. 1A might reflect a neutral expectation rather than natural selection, possibly amplified by environmental effects.

Trait correlations can also complicate inferences of selection differences, as a scenario in which differences in polygenic score distributions among populations parallel differences for a specific phenotype might be due not only to environmental factors, but instead to natural selection on other correlated traits [48].

**Figure 2.** The change in mean phenotypic differences between populations resulting from natural selection acting on a correlated phenotype. In the graphs, two populations with different trait values, indicated by circles, experience natural selection, indicated by arrows. Selection acts on phenotype I, which is strongly positively correlated with phenotype II. The action of selection changes the population difference in the mean values of phenotype II, as indicated to the right of the graphs. (A) Directional selection on phenotype I in the same direction in the two populations increases the population difference for phenotype II. (B) Directional selection on phenotype I in the same direction in the two populations decreases the population difference for phenotype II. (C) Divergent natural selection on phenotype I increases the population difference for phenotype II.
the trait. In this model, systematic health disparities are unlikely: assuming that no genetic correlation exists between phenotypic outcomes, the binomial probability that the trait value is larger in one of a pair of populations for at least 28 of 36 independent phenotypes is 0.0012 [42].

It might be proposed that different strengths of directional selection have contributed to the population difference between African Americans and European Americans. However, a related computation of the overall influence of natural selection in human history, relying on measures of selection against deleterious variants rather than directional selection of favorable variants, does not suggest strong systematic differences in the magnitude of selection among different continental population groups [50–52]; indeed, some researchers have argued for a greater level of deleterious variation in non-Africans rather than in Africans [50, 52], a pattern opposite to what might be expected given the direction of health disparities.

Whereas natural selection cannot easily explain the observed population difference, systematic environmental effects that contribute to an increase in non-genetic risk factors in African Americans—current and historical racism, for instance [53–55]—could, on the other hand, explain such marked differences. This example of health disparities illustrates important features of human genetic differences within and beyond the scientific community is the attribution of interpretive weight to plausibly possible to address topics concerning the genetic and phenotypic developments in controversial areas of genetics [59]—can result in cascading distortions of the genetic basis of phenotypic traits that studies do not imply and that their authors do not support [17].

As developments on polygenic scores and polygenic adaptation connect closely to topics that have long been of central interest in human evolutionary genetics, the field can provide context for the emerging plethora of results relevant to interpretations of the roles of genetics and natural selection in contributing to traits; limitations of interpretations of research in new directions are not restricted to the topics emphasized here [41, 60]. Vigilance in promoting careful and evidence-supported explanations and in clarifying the caveats that affect ongoing genetic research programs continues to be required both from investigators and from those who disseminate the findings.

ACKNOWLEDGEMENTS

We thank L. Odling-Smee, S. Ramachandran, L. Uricchio, R. Winther, and two reviewers for comments on drafts of the article.

FUNDING

Partial support was provided by National Institutes of Health grants (GM108779, GM130050, and HGC005855), a National Science Foundation grant (DBI-1458059), and the Stanford Center for Computational, Evolutionary and Human Genomics.

Conflict of interest: None declared.

REFERENCES

Interpreting polygenic scores

35. Sohail M, Maier RM, Ganna A et al. Signals of polygenic adaptation on height have been overestimated due to uncorrected population structure in genome-wide association studies. bioRxiv 2018; doi: 10.1101/355057.