

Tread Lightly Interpreting Polygenic Tests of Selection

John Novembre^{*,*1} and Nicholas H. Barton[†]

^{*}Department of Human Genetics, Department of Ecology and Evolution, University of Chicago, Illinois 60637 and [†]Institute of Science and Technology, A-3400 Klosterneuberg, Austria

ORCID IDs: 0000-0001-5345-0214 (J.N.); 0000-0002-8548-5240 (N.H.B.)

KEYWORDS complex trait evolution; directional selection; polygenic

In this issue of GENETICS, a new method for detecting natural selection on polygenic traits is developed and applied to several human examples (Racimo et al. 2018). By definition, many loci contribute to variation in polygenic traits, and a challenge for evolutionary geneticists has been that these traits can evolve by small, nearly undetectable shifts in allele frequencies across each of many, typically unknown, loci. Recently, a helpful remedy has arisen. Genome-wide association studies (GWAS) have been illuminating sets of loci that can be interrogated jointly for changes in allele frequencies. By aggregating small signals of change across many such loci, directional natural selection is now in principle detectable using genetic data, even for highly polygenic traits. This is an exciting arena of progress – with these methods, tests can be made for selection associated with traits, and we can now study selection in what may be its most prevalent mode. The continuing fast pace of GWAS publications suggest there will be many more polygenic tests of selection in the near future, as every new GWAS is an opportunity for an accompanying test of polygenic selection. However, it is important to be aware of complications that arise in interpretation, especially given that these studies may easily be misinterpreted both in and outside the evolutionary genetics community. Here, we provide context for understanding polygenic tests and urge caution regarding how these results are interpreted and reported upon more broadly.

The foundations of polygenic tests of selection trace back to the very dawn of genetics, a century ago. An early challenge was to reconcile the inheritance of continuously distributed traits like height with the discretely inherited particulate genes of Mendelian inheritance. In a monumental advance in the history of genetics, R.A. Fisher reconciled this conflict by proposing that many genetic loci contribute to the variation of such traits – a polygenic model of inheritance – and by showing how to analyze their collective effect. In modern parlance, these traits are referred to as “quantitative,” “polygenic,” or “complex,” and each of the contributing loci are known as a quantitative trait locus (QTL). GWAS are revolutionary in allowing modern practitioners to uncover a fraction of the QTL that Fisher posited underlie most trait variation and to estimate effect sizes of each variant on a trait. Importantly for the discussion here, the polygenic model has evolutionary implications – most notably, the mean phenotypic value of a trait in a population can shift substantially and quickly via very subtle shifts in frequencies at the many QTL that underlie a trait. To see a signature of selection, modern polygenic tests use the QTL discovered by GWAS and detect shifts in their frequencies across time or across populations. The key advance has been to realize that while the impact of selection on any particular QTL cannot be reliably detected, the collective impact across many QTL can be.

The first polygenic selection study to use human GWAS results was based on a commonly studied trait, human height. Turchin *et al.* (2012) found a difference between northern and southern Europeans in the genetic component of height that appears to have been driven by differential selection across populations. A subsequent foundational paper by Berg and Coop (2014) gave a more formal analysis, by establishing

Copyright © 2018 Novembre and Barton

doi: <https://doi.org/10.1534/genetics.118.300786>

Manuscript received February 2, 2018; accepted for publication February 14, 2018.

Available freely online through the author-supported open access option.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

¹Corresponding author: Department of Human Genetics, Department of Ecology and Evolution, University of Chicago, 920 E. 58th St., Chicago, IL 60637. E-mail: jnovembre@uchicago.edu

a null model for how polygenic scores should vary across populations. This approach generalized and modernized a traditional pre-GWAS approach [comparison of Q_{st} vs. F_{st} , Whitlock and Guillaume 2009] and has itself been extended to consider multiple traits (Berg *et al.* 2017). These initial approaches focused on comparison across populations, but comparisons can also be made across time within single populations using either ancient DNA (Mathieson *et al.* 2015) or using haplotypic signatures of allele frequency change (Field *et al.* 2016). The new study published in this issue by Racimo *et al.* provides a novel methodology for comparing across populations by using a population tree with migration edges (an “admixture graph”). The new method allows one to isolate which branches in an admixture graph have experienced polygenic selection. In some sense then, Racimo *et al.* extend Berg and Coop’s (2014) approach to allow more detailed resolution of when and where directional selection took place.

While these methodological contributions are important, the Racimo *et al.* study requires delicate handling. This is true for all polygenic selection studies, but all the more so here because of the comparison across populations for the complex behavioral trait of “educational attainment” (*i.e.*, years of schooling). Genetic contributions to educational attainment have been of keen interest to social scientists and economists who have an eye on implications for policy (e.g. Conley and Fletcher 2017). The Racimo *et al.* study is not unique in analyzing a sensitive trait with social implications – a recent study analyzes polygenic selection on 25 traits related to brain structure, neuropsychiatric characteristics, and personality traits (Beiter *et al.* 2017); more such studies should be expected. The recent publication of a GWAS of delay discounting (Sanchez-Roige *et al.* 2018) suggests that a polygenic selection test will soon be carried out too on this complex trait (delay discounting is the preference of an individual for an immediate reward rather than a larger reward after a delay). Such studies are particularly sensitive when they are based on between-population comparisons, as they draw attention to potential between-population differences in humans and suggest that these differences are not only detectable but also driven by selection that differs across populations. Given the potential for these results to affect social scientists and economists who may advise on policy making (e.g. Conley and Fletcher 2017), a clear understanding of the pitfalls of interpretation is especially necessary.

Here we want to emphasize that caution is needed as these results are discussed or considered in an applied context – there remain substantial challenges in interpreting results at the forefront of this innovative field. There are two main categories of complications. The first are the technical challenges that may give a false sign of directional selection. The second category are the complications of interpretation. We tackle each in turn.

A first technical concern is the potential for population stratification to affect the original GWAS study. The stratification problem is that a putative QTL from a GWAS study may

just be a proxy for environmental variables or background genetic effects that have not been properly included in the statistical model. Suppose, for example, that environmental factors affecting a particular trait vary east to west across a study area. As a result, in a GWAS that is not carefully controlled for stratification, any variant that has differentiated in frequency along an east–west axis will have an overestimated effect size. In turn, a polygenic score would show elevated differentiation across populations at the QTL relative to a genome-wide background, even if selection has been absent. Any selection that differentiates populations along an east–west axis would produce similar spurious signals, even if it had nothing to do with the trait of interest. Correcting GWAS studies to avoid stratification has become standard using PCA and mixed model-based approaches (Price *et al.* 2010), but the concern of residual stratification still persists, and for this reason, some studies of polygenic evolution have relied on family-based estimates of effect sizes (Turchin *et al.* 2012; Field *et al.* 2016). The effect size estimates derived from within families are less susceptible to varying environmental factors but fewer within-family studies have been carried out – so this can be difficult to apply for most traits. LD-score regression (Bulik-Sullivan *et al.* 2015a) is an approach that allows one to assess whether residual stratification may be affecting a GWAS, and such approaches can also be useful for verifying polygenic signatures of selection (Field *et al.* 2016). As a practical matter, polygenic signatures that lack family-based estimates of effect size should be taken as provisional, given the concerns about residual confounding.

A second technical challenge is that of transferring effect sizes across space and time. Any GWAS is based on a particular study population sampled in the present-day. GWAS have been disproportionately carried out in European populations; a 2016 survey found that 81% of GWAS studies were from European ancestry populations (Bustamante *et al.* 2011; Popejoy and Fullerton 2016). The effects of this bias are unclear. GWAS loci are typically “tag SNPs” that indirectly reflect (“tag”) the effects of a nearby true causal QTL. The link between causal QTL and their tag SNPs may differ across space and time in ways that can bias polygenic scores. Further, there is the concern that the phenotypic effect of alleles at the QTL may differ by genetic or environmental background (due to epistasis and gene-by-environment interactions, $G \times G$ and $G \times E$, respectively, in genetics shorthand). In practice, some analyses support the general transferability of GWAS across populations (Marigorta and Navarro 2013), while others suggest that problems arise in computing polygenic scores across populations (Martin *et al.* 2017). Given these uncertainties, unless the GWAS for a trait has been carried out in populations that are generally representative of the same studies used in the polygenic evolution study, one should be cautious in interpreting results. Current polygenic tests are narrowly testing whether the trait-associated SNPs found in a *particular* GWAS study population are evolving by drift across a possibly broader set of populations; the relevance for understanding how the

complete genetic basis of a trait evolves will be unclear until more is learned about how the sharing and tagging of genetic effects can be modeled over time and space.

A third challenge is the ascertainment bias inherent in GWAS. SNPs identified by GWAS for traits of interest may be associated with regions that show “signatures” of selection simply because detection depends on allele frequencies, or less trivially, because SNPs that affect the trait also tend to affect fitness, for reasons that have nothing to do with the trait. For example, consider a disease trait studied in a European population – the ascertainment biases would make it such that discovered disease-risk alleles would have a systematically higher frequency inside Europe than outside, which could be mistaken for directional selection. In the studies discussed here, the signal is stronger: SNPs associated with increased trait values also tend to increase in frequency. Such directional associations give much stronger evidence for selection and show a genetic correlation between the trait and fitness. Nevertheless, as we discuss below, one still cannot say that the trait was *itself* selected (*i.e.*, that it *caused* increased fitness).

A final technical challenge specific to the Racimo *et al.* approach is that it depends on an accurate model of the populations and also, that there have been few directional selection events during the history of the trait. Notable violations of either of these assumptions may cause problems with the inference procedure, so careful evaluation of the validity of the admixture graph model should be undertaken when evaluating the results of any specific analysis. This is a pervasive problem in making inferences from genetic data: how can we know whether conclusions made assuming simplified models are valid, when there is an infinity of possible complications? The best that can be done is often to simulate a variety of plausible alternatives, but this can be computationally challenging.

Even if a signal of polygenic evolution is technically sound, there are still challenges to interpretation. First, many of the traits available for study by GWAS are awkward to conceptualize in an evolutionary framework. The majority of GWAS studies to date have been focused on traits of medical relevance or that are relatively easy to measure. As such, evolutionary geneticists are constrained to analyzing a somewhat arbitrary set of traits. No principled research program on understanding polygenic selection in human evolution would begin with traits such as self-reported unibrow or educational attainment. For educational attainment, the challenges are particularly acute; values observed today cannot be validly compared to those for humans living just 100 years ago when access to schooling was generally much more restricted, let alone to humans living 10,000 years ago. The caveat is also relevant when comparing across modern populations that differ greatly in the baseline schooling environment.

At the core there is a major challenge for interpretation in untangling the connection between measured traits (*e.g.*, height) and components of fitness (*e.g.*, increased success in survival, finding a mate, successful reproduction upon mating). For example, one may propose that height directly affects the probability of finding a mate; or alternatively, another in which

height may be an indirect outcome of efficient metabolism, which has a direct impact on survival to adulthood. In either case, a polygenic signature of selection would be detected on height, but further multi-trait studies would be necessary to provide a precise interpretation as to whether height is directly or indirectly related to fitness. The key underlying problem is the existence of genetic covariance between traits (Lande and Arnold 1983). Traits can covary due to the presence of alleles that affect both traits (pleiotropy) or because of nonrandom association of alleles across loci (linkage disequilibrium). The potential for widespread pleiotropy due to the architecture of gene regulatory networks and gene action has been emphasized as part of an “omnigenic” model for complex traits (Boyle *et al.* 2017), and builds upon mounting empirical evidence of genetic covariance in human genetics (*e.g.*, Bulik-Sullivan *et al.* 2015b; Pickrell *et al.* 2015; Visscher and Yang 2016); the importance of pleiotropy has long been appreciated in quantitative genetics (*e.g.*, Barton 1990). New multivariate approaches to address polygenic traits are emerging (Berg *et al.* 2017) and new biobank resources are making analysis with direct measurements of fitness more feasible (Sanjak *et al.* 2018), but they still are complicated by the fundamental problem of missing, unmeasured traits (Lande and Arnold 1983). The challenges of trait covariance are complicated, even for the interpretation for some of the strongest single-locus signatures of selection (*e.g.*, EDAR; Kamberov *et al.* 2013). For these reasons, authors of polygenic studies, including Racimo *et al.*, are careful to say they have detected selection on “variants associated with trait X” and not “selection on trait X.” To truly say a trait is selected, *i.e.*, that it causes a difference in fitness, requires more direct evidence and extended study than allele frequency shifts alone can provide.

A second complication of interpretation regards comparisons across populations or time periods. Even when estimated correctly, relative values of a polygenic mean do not necessarily reflect relative values of the mean phenotypes that will be observed in populations. For instance, the polygenic mean for population A may be larger than B, but compensating environmental factors may make it such that population B has an equal or larger trait mean than in population A. A fascinating example of the potential disconnect between polygenic means and phenotypes can be seen in wild Soay sheep on the island of St. Kilda. Based on observations over 20 years, directional selection has been acting to increase the mean body size, yet body sizes have actually been decreasing, likely because of compensating effects in which recent changes in climatic conditions have led to slower growth rates and in turn lower average body sizes (Ozgul *et al.* 2009). This problem may not be idiosyncratic – many studies in natural populations have seen a disconnect between measured fitness differentials and observed phenotypic outcomes (Kruuk *et al.* 2008). For example, Robinson *et al.* (2015) show that in European humans, genetic differentiation among populations in body-mass index appears to be masked by environmental factors and that observed differences in average height are only partially explained by differentiation at current known GWAS loci. Another aspect of relative comparisons that is important to consider

Box 1 Using GWAS to detect selection on traits

The deviations of a trait, z , can be modeled as the sum of effects of a large number of SNPs: $z = \sum \beta_i X_i + \varepsilon$, where $X_i = 0, 1, 2$ is an individual's dosage of a reference allele at the i th SNP, β_i is the effect of the i th SNP, and ε is a random "environmental" effect; effects of specific factors such as age, location, or sex can be represented by additional terms. The β_i are estimated simply by regression of the trait on the genotype; call these estimates $\hat{\beta}_i$. Then, the genetic component of the trait value of a new individual with genotype Y_i , a "polygenic score", can be estimated as $\hat{z} = \sum \hat{\beta}_i Y_i$. A difference between the mean \hat{z} values for different sets of individuals can suggest selection on the trait if the difference is larger than expected by neutral evolution (i.e. genetic drift). Such selection may act through small shifts in allele frequency, which can nevertheless accumulate over very many loci to cause a large change in the trait mean.

Shifts in mean polygenic score may have many causes, in addition to selection on the observed trait:

- Selection on other traits.
- Selection may act on a trait that is genetically correlated with the observed trait, for instance through pleiotropic side-effects of the underlying alleles.
- Stratification: If the GWAS uses a sample that is a mixture of individuals from different places or different social groups, and the trait varies among these groups, then SNPs that are differentiated between groups can be spuriously associated with the trait.
- Varying effects: Changes in effects, β , between populations will tend to reduce the signal.
- Ascertainment bias: Common SNPs are more likely to be ascertained as significant in GWAS, and this may bias allele frequency differences that are the basis for inferring selection. However, a systematic directional change in frequency of alleles associated with increases in the trait gives a more robust signal.
- Random drift gives the null model: Its effects can be simulated, but an appropriate model must be chosen, which takes account of both population structure and linkage.

is that even when polygenic means differ, the variance of the distributions within populations may be so wide that there can be little predictive value in knowing which population an individual comes from.

Overall, the numerous complications described here, both technical and interpretative, are why this exciting field is still in its infancy. Progress is being made but most findings are wrapped in numerous caveats. For this reason, we caution that great care should be taken in communicating results of these studies to general audiences. Journalists producing simple headlines and/or taking results out of context have the potential to misconstrue the complexity and levels of uncertainty in an arena where simple misinterpretations come easily. Generally, authors of such studies, including Racimo *et al.*, are cautious, and this degree of caution must not be lost in translation. This is particularly sensitive as polygenic selection studies are analyzing complex social and behavioral traits, and as social scientists look more keenly upon these studies as a source of inspiration for policy; these efforts may be premature, to say the least. We are in a time of extraordinary discoveries, but we must remember that even as we gain traction with new computational tools and expanded genomic studies, we still have a long way to go to make precise, fully supported statements about the nature of selection on complex traits in humans.

Acknowledgments

The authors thank Anna Di Rienzo, Molly Przeworski, Iain Mathieson, Fernando Racimo, Jeremy Berg, and Joe Pickrell for helpful comments on early versions of this perspective.

Literature Cited

- Barton, N. H., 1990 Pleiotropic models of quantitative variation. *Genetics* 124: 773–782.
- Beiter, E. R., E. A. Khrantsova, C. van der Merwe, E. R. Chimusa, C. Simonti *et al.*, 2017 Polygenic selection underlies evolution of human brain structure and behavioral traits. *bioRxiv*. <https://doi.org/10.1101/164707>
- Berg, J. J., and G. Coop, 2014 A population genetic signal of polygenic adaptation. *PloS Genet.* Aug 7;10(8):e1004412.
- Berg, J. J., X. Zhang, and G. Coop, 2017 Polygenic adaptation has impacted multiple anthropometric traits. *bioRxiv*. <https://doi.org/10.1101/167551>
- Boyle, E. A., Y. I. Li, and J. K. Pritchard, 2017 An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169: 1177–1186. <https://doi.org/10.1016/j.cell.2017.05.038>
- Bulik-Sullivan, B. K., P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang *et al.*, 2015a LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47: 291–295. <https://doi.org/10.1038/ng.3211>
- Bulik-Sullivan, B., H. K. Finucane, V. Anttila, A. Gusev, F. R. Day *et al.*, 2015b An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 47: 1236–1241. <https://doi.org/10.1038/ng.3406>
- Bustamante, C. D., E. G. Burchard, and F. M. De la Vega, 2011 Genomics for the world. *Nature* 475: 163–165. <https://doi.org/10.1038/475163a>
- Conley, D., and J. Fletcher, 2017 *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future*. Princeton University Press, Princeton, NJ. <https://doi.org/10.1515/9781400883240>
- Field, Y., E. A. Boyle, N. Telis, Z. Gao, K. J. Gaulton *et al.*, 2016 Detection of human adaptation during the past 2000 years. *Science* 354: 760–764. <https://doi.org/10.1126/science.aag0776>
- Kamberov, Y. G., S. Wang, J. Tan, P. Gerbault, A. Wark *et al.*, 2013 Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* 152: 691–702. <https://doi.org/10.1016/j.cell.2013.01.016>

- Kruuk, L. E. B., J. Slate, and A. J. Wilson, 2008 New answers for old questions: the evolutionary quantitative genetics of wild animal populations. *Annu. Rev. Ecol. Evol. Syst.* 39: 525–548. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173542>
- Lande, R., and S. J. Arnold, 1983 The measurement of selection on correlated characters. *Evolution* 37: 1210–1226. <https://doi.org/10.1111/j.1558-5646.1983.tb00236.x>
- Marigorta, U. M., and A. Navarro, 2013 High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genet.* 9: e1003566. <https://doi.org/10.1371/journal.pgen.1003566>
- Martin, A. R., C. R. Gignoux, R. K. Walters, G. L. Wojcik, B. M. Neale *et al.*, 2017 Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* 100: 635–649. <https://doi.org/10.1016/j.ajhg.2017.03.004>
- Mathieson, I., I. Lazaridis, N. Rohland, S. Mallick, N. Patterson *et al.*, 2015 Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528: 499–503. <https://doi.org/10.1038/nature16152>
- Ozgul, A., S. Tuljapurkar, T. G. Benton, J. M. Pemberton, T. H. Clutton-Brock *et al.*, 2009 The dynamics of phenotypic change and the shrinking sheep of St. Kilda. *Science* 325: 464–467. <https://doi.org/10.1126/science.1173668>
- Pickrell, J., T. Berisa, L. Segurel, J. Y. Tung, and D. Hinds, 2015 Detection and interpretation of shared genetic influences on 40 human traits. *bioRxiv Nat. Genet.* 48(7): 709–17.
- Popejoy, A. B., and S. M. Fullerton, 2016 Genomics is failing on diversity. *Nature* 538: 161–164. <https://doi.org/10.1038/538161a>
- Price, A. L., N. A. Zaitlen, D. Reich, and N. Patterson, 2010 New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* 11: 459–463. <https://doi.org/10.1038/nrg2813>
- Racimo, F., J. J. Berg, and J. K. Pickrell, 2018 Detecting polygenic adaptation in admixture graphs. *Genetics* 208: 1565–1584.
- Robinson, M. R., G. Hemani, C. Medina-Gomez, M. Mezzavilla, T. Esko *et al.*, 2015 Population genetic differentiation of height and body mass index across Europe. *Nat. Genet.* 47: 1357–1362. <https://doi.org/10.1038/ng.3401>
- Sanchez-Roige, S., P. Fontanillas, S. L. Elson, A. Pandit, E. M. Schmidt *et al.*, 2018 Genome-wide association study of delay discounting in 23,217 adult research participants of European ancestry. *Nat. Neurosci.* 21: 16–18.
- Sanjak, J. S., J. Sidorenko, M. R. Robinson, K. R. Thornton, and P. M. Visscher, 2018 Evidence of directional and stabilizing selection in contemporary humans. *Proc. Natl. Acad. Sci. USA* 115: 151–156. <https://doi.org/10.1073/pnas.1707227114>
- Turchin, M. C., C. W. K. Chiang, C. D. Palmer, S. Sankararaman, D. Reich *et al.*, 2012 Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* 44: 1015–1019. <https://doi.org/10.1038/ng.2368>
- Visscher, P. M., and J. Yang, 2016 A plethora of pleiotropy across complex traits. *Nat. Genet.* 48: 707–708. <https://doi.org/10.1038/ng.3604>
- Whitlock, M. C., and F. Guillaume, 2009 Testing for spatially divergent selection: comparing QST to FST. *Genetics* 183: 1055–1063. <https://doi.org/10.1534/genetics.108.099812>

Communicating editor: M. Johnston