

An evolutionary compass for detecting signals of polygenic selection and mutational bias

Lawrence H. Uricchio,^{1,2,*} Hugo C. Kitano,^{3,*} Alexander Gusev,^{4,*} and Noah A. Zaitlen^{5,6,7,*}

¹Department of Biology, Stanford University, Stanford, CA

²E-mail: uricchio@stanford.edu

³Department of Computer Science, Stanford University, Stanford, CA

⁴Dana Farber Cancer Institute, Boston, MA

⁵Department of Medicine, University of California, San Francisco, CA

⁶Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA

⁷E-mail: noah.zaitlen@ucsf.edu

Received July 11, 2018

Accepted December 10, 2018

Selection and mutation shape the genetic variation underlying human traits, but the specific evolutionary mechanisms driving complex trait variation are largely unknown. We developed a statistical method that uses polarized genome-wide association study (GWAS) summary statistics from a single population to detect signals of mutational bias and selection. We found evidence for nonneutral signals on variation underlying several traits (body mass index [BMI], schizophrenia, Crohn's disease, educational attainment, and height). We then used simulations that incorporate simultaneous negative and positive selection to show that these signals are consistent with mutational bias and shifts in the fitness-phenotype relationship, but not stabilizing selection or mutational bias alone. We additionally replicate two of our top three signals (BMI and educational attainment) in an external cohort, and show that population stratification may have confounded GWAS summary statistics for height in the GIANT cohort. Our results provide a flexible and powerful framework for evolutionary analysis of complex phenotypes in humans and other species, and offer insights into the evolutionary mechanisms driving variation in human polygenic traits.

KEY WORDS: Complex traits, mutation bias, natural selection, polygenic selection.

Impact Summary

Many traits are variable within human populations and are likely to have a substantial and complex genetic component. This implies that mutations that have a functional impact on complex human traits have arisen throughout our species' evolutionary history. However, it remains unclear how evolutionary processes such as natural selection may have acted to shape trait vari-

ation at the genetic and phenotypic level. Better understanding of the mechanisms driving trait variation could provide insights into our evolutionary past and help clarify why it has been so difficult to map the preponderance of causal variation for common heritable diseases.

In this study, we developed and applied methods for detecting signatures of mutation bias (i.e., the propensity of a new variant to be either trait-increasing or trait-decreasing) and natural selection acting on trait variation. We applied our approach to several heritable

*Lawrence H. Uricchio and Hugo C. Kitano are co-first authors, and Alexander Gusev and Noah A. Zaitlen are co-last authors.

traits, and found evidence for both natural selection and mutation bias, including selection for decreased body mass index [BMI] and decreased risk for Crohn's disease and schizophrenia.

While our results are consistent with plausible evolutionary scenarios shaping a range of traits, it should be noted that the field of polygenic selection detection is still new, and current methods (including ours) rely on data from genome-wide association studies (GWAS). The data produced by these studies may be vulnerable to certain cryptic biases, especially population stratification, which could induce false selection signals. We therefore repeated our analyses for the top three hits in a cohort that should be less susceptible to this problem—we found that two of our top three signals replicated (BMI and educational attainment), while height did not. Our results highlight both the promise and pitfalls of polygenic selection detection approaches, and suggest a need for further work disentangling stratification from selection.

Natural selection and mutation shape variation within and between populations, but the evolutionary mechanisms shaping causal variation for human traits remain largely unknown. Studies of selection in humans have often focused on classic selective sweeps (Sabeti et al. 2002, 2006; Voight et al. 2006; Hernandez et al. 2011; Enard et al. 2014), but other processes such as stabilizing selection (Gilad et al. 2006; Sanjak et al. 2018), polygenic adaptation (Turchin et al. 2012; Berg and Coop 2014), background selection (Charlesworth 1994; McVicker et al. 2009), negative selection (Boyko et al. 2008), and soft sweeps (Messer and Petrov 2013; Schrider and Kern 2017) may also play an important role in shaping human diversity. Methods to detect selection under these more complex models are needed if we are to fulfill the promise of genomics to explain the evolutionary mechanisms driving the distribution of heritable traits in human populations (Pritchard et al. 2010).

With the recent proliferation of paired genotype and phenotype data from large human cohorts, it is now feasible to test for polygenic selection on specific traits (Turchin et al. 2012; Berg and Coop 2014; Robinson et al. 2015; Yang et al. 2015; Field et al. 2016). While these studies have argued that polygenic selection is likely to be an important determinant of variation in traits such as height and skin pigmentation (Berg and Coop 2014), important questions remain about the evolutionary mechanisms that drive complex trait variation. In particular, most previous studies of selection on human complex traits have focused either on

polygenic adaptation (Turchin et al. 2012; Berg and Coop 2014; Field et al. 2016; Berg et al. 2017; Edge and Coop 2018; Racimo et al. 2018) or stabilizing/negative selection (Yang et al. 2015; Zeng et al. 2018; Simons et al. 2018), and have not incorporated mutational bias (i.e., the propensity for new mutations to be preferentially trait-increasing or preferentially trait-decreasing). To detect polygenic adaptation, studies have relied on genotype data from multiple populations to probe the frequency and linkage properties of trait-associated alleles as compared to a null based on genome-wide single nucleotide polymorphisms (Turchin et al. 2012; Berg and Coop 2014; Racimo et al. 2018), or have used haplotype-based statistics and dense sequence data from a single population (Field et al. 2016). Negative selection has been investigated by comparing empirical data to null models of the relationship between frequency and squared effect sizes (Schoech et al. 2017) or linkage disequilibrium and per-single nucleotide polymorphism heritability (Gazal et al. 2017). These studies have argued that selection acts on many traits, and that both negative (Gazal et al. 2017; Schoech et al. 2017) and positive (Berg and Coop 2014; Berg et al. 2017) selection act on complex traits such as height and body mass index (BMI).

While each of these approaches has provided insights into the evolution of complex traits, a more comprehensive view of trait evolution will require methods that can account for pleiotropic selection and incorporate signals of both adaptive and deleterious selection processes simultaneously. Recent progress has been reported in accounting for higher dimensional trait spaces (Berg et al. 2017; Simons et al. 2018), but there is a need for models and inference tools integrating signatures of positive and negative selection on complex traits (Sanjak et al. 2018). Indeed, a natural way to model polygenic adaptation is to view stabilizing selection as a null process, with punctuated changes in the fittest trait value (herein called the “optimal trait value” or “trait optimum”) driving brief periods of adaptation (Barton 1986; Jain and Stephan 2017). Mutational bias may also be an important contributor to the evolutionary dynamics of complex traits (Charlesworth 2013), but has not been directly incorporated into recent empirical studies. If there is a bias in the direction of effect of de novo mutations, populations may carry less standing variation for alleles that alter the phenotype in one direction than the other, potentially altering the dynamics of future adaptation to changes in the trait optimum. Moreover, biases in mutation rate for selected traits may induce detectable patterns in the relationship between allele frequency and effect size, for example by driving an excess of trait-increasing mutations among young (low-frequency) alleles but not old alleles.

Here, we develop a powerful method for detecting differences between ancestral and derived allele effect sizes—which may be driven by selection or mutational bias—within a single population. We show that the polarization of GWAS summary

statistics by their ancestral/derived state (which we refer to as an evolutionary compass) provides information about the evolutionary processes shaping trait variation. We propose a simple summary statistic of the relationship between effect sizes and allele frequency, and show that it is sensitive to both mutational bias and various models of selection. We apply our approach and find nonneutral signals that are consistent with selection and mutational bias in the genetic variation underlying BMI, educational attainment, Crohn's disease, schizophrenia, and height. We develop a model-based inference procedure to disentangle mutational bias from selection, and show that both processes are necessary to explain the observed GWAS summary data. We then perform a replication study in the UK Biobank for our top three signals, and find that educational attainment and BMI replicate in this more homogeneous cohort, while height does not. We discuss implications of our findings for human evolutionary history and GWAS of biomedically relevant traits.

Results

AN "EVOLUTIONARY COMPASS" FOR GWAS

Associations between genotypes and complex traits are usually reported with respect to a reference allele that is arbitrarily chosen, which can obscure the direction of effect of new mutations. When selection acts on traits, it favors the reproductive success or survival of individuals with particular trait values, implying that the fitness effect of new mutations may depend on both the sign and magnitude of their impact on a selected trait. Therefore, we choose ancestral alleles as the reference state and explore the relationship between derived allele effect sizes and derived allele frequency, which we encode in a β DAF plot (i.e., a plot encoding the relationship between the mean value of effect sizes β estimated in a GWAS ($\bar{\beta}$) and derived allele frequency (DAF); Fig. 1). Note that we use estimated effect sizes from all alleles to compute $\bar{\beta}$, regardless of their significance.

A β DAF plot contains information about mutational bias and selection on the trait of interest. In the null case of a neutrally evolving trait for which trait-increasing and trait-decreasing mutations are equally likely, the β DAF curve will be flat and will have expectation $\bar{\beta} = 0$ in all derived allele frequency bins. This is because in a neutral model, the probability with which an allele segregates at frequency x does not depend on effect size β (Fig. S2). A mutational bias toward trait-increasing or trait-decreasing alleles in the absence of selection on the trait will shift the mean value of β in the direction of the bias, but will not induce $\bar{\beta}$ to depend on frequency (Fig. S2).

While many evolutionary processes will induce patterns in the β DAF plot (including directional selection, which we explore in the Supporting Information with simulations and analytical calculations in Figs. S1–S3), we next consider an example of

stabilizing selection with shifts in the optimal trait value for illustrative purposes. Stabilizing selection is typically parameterized by several parameters, but here we focus on the optimal phenotype value, ϕ_o , which represents the value of the phenotype that confers the highest fitness. Applying a classic stabilizing selection model, we suppose that fitness is controlled by a Gaussian function centered at ϕ_o (Robertson 1956; Barton 1986; Simons et al. 2018) (Fig. 1A). In addition, we suppose that trait-increasing mutations may be more or less likely than trait-decreasing mutations, which we capture with the parameter δ (defined as the proportion of trait-altering de novo mutations that increase trait values), and that ϕ_o can change, inducing a brief period of adaptation in which trait values within the population equilibrate to a new optimal value (Fig. 1B).

We performed forward simulations of complex traits under this model and explored how various evolutionary parameters affect the properties of a β DAF plot. Although we consider a very large range of possible parameter combinations when performing statistical inference in later sections, here we focused on four models, including one model of stabilizing selection in the absence of shifts in the optimal trait value, and three models that varied the timing of a shift (t_s) in the optimal trait value ϕ_o (Fig. 1C). All of the models included a bias in mutation rate toward trait-decreasing alleles ($\delta = 0.4$) and a European demographic model that was fit to patterns of European genomic diversity (Gravel et al. 2011). For models that include a shift in the optimal trait value, we considered a change (denoted $\Delta\phi$) equal to two SDs of the trait distribution. For reference, this would correspond to approximately a 5 inch change in mean human height (Fryar et al. 2016). The remaining model parameters for the simulations in Figure 1 are given in the Supporting Information (page 26). To compute the mean effect size $\bar{\beta}$ (Fig. 1D), we grouped alleles into 1% frequency bins and computed the mean effect size across all derived alleles within the bin.

When there is a bias toward trait-decreasing alleles and stabilizing selection acts on the trait, alleles at low frequencies have strongly negative effect sizes, which generate a positive correlation between effect size and frequency (Fig. 1D, red curves). When the optimal value of the phenotype increases in response to an environmental shift, the relationship between effect size and selection coefficient also transiently changes (Fig. 1B), driving some alleles with beneficial effects to higher frequencies (Fig. 1D, yellow, green, and orange curves). This effect transiently changes the relationship between allele frequency and effect size by promoting trait-increasing alleles to higher frequencies, generally increasing $\bar{\beta}$ at all frequencies. Although $\bar{\beta}$ is always negative at low frequencies for the particular parameter combinations we investigated here, at high frequencies, $\bar{\beta}$ can become positive due to the preferential increase in frequency of trait-increasing alleles. These effects decay as the time since the shift event increases

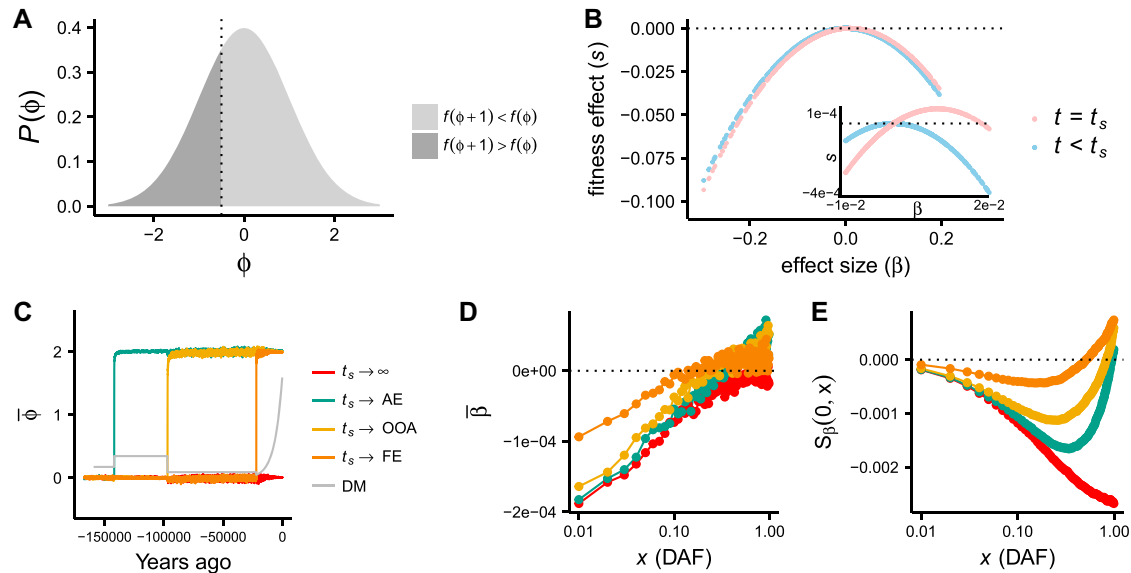


Figure 1. Panels A-B are schematics of the trait model, while C-E show simulation results. **A:** fitness impact of a $\beta = 1$ mutation, assuming a symmetric fitness function. At equilibrium, the trait distribution $P(\phi)$ is symmetric about the optimal value of the phenotype, $\phi_o = 0$. The dashed line at $\phi = -\frac{1}{2}$ indicates the dividing line between individuals with increased fitness f after a $\beta = 1$ mutation ($f(\phi + 1) > f(\phi)$) from those with decreased fitness ($f(\phi + 1) < f(\phi)$). **B:** schematic of the relationship between effect size and fitness effect. At time $t = t_s$, the optimal trait value ϕ_o increases, and trait-decreasing alleles have decreased fitness while trait-increasing alleles have increased fitness. Still, only trait-increasing alleles of small effect are on average fitness-increasing (inset). **C:** Mean trait value $\bar{\phi}$ as a function of time for four simulated trait models, differentiated by the time of a shift in selection pressure. The simulated European demographic model is plotted in the background (not to scale) **D:** $\bar{\beta}$ as a function of derived allele frequency (DAF) for each model simulated in C. Points represent the mean value of β computed over 100 independent simulations **E:** $S_{\beta}(0, x)$ as a function of DAF for each model plotted in C. D and E represent the mean over 100 independent simulations. (Abbreviations: AE: ancestral expansion, OOA: out-of-Africa, FE: founding of Europe, DM: demographic model).

(Fig. 1D). The slow decay indicates that such patterns might be detectable in trait data for tens of thousands of years, with the time-scale for detection depending on the model parameters and the precision of effect size estimates in GWAS summary data.

A STATISTICAL TEST FOR POLYGENIC SELECTION AND MUTATIONAL BIAS ON ORIENTED GWAS

Selection and mutational bias change the relationship between $\bar{\beta}$ and DAF (Fig. 1; Figs. S1–S3). We desire a simple statistic that will differ from zero when selection and/or mutational bias act. We first consider the integral of the area under a β DAF plot, noting that while other statistics are also likely to be informative (Yang et al. 2015), not all choices will be robust to ancestral state uncertainty (see Supporting Information). We approximate this integral with the sum (S_{β}) over $\bar{\beta}$ for derived allele frequency bins of some width y_w .

$$S_{\beta}(f_i, f_j) = \sum_{y \in Y_{ij}} \bar{\beta}(y). \tag{1}$$

As in Fig. 1, we group alleles into frequency bins of width y_w such that the set Y is given by the sequence of frequency tuples $Y = \langle (0, y_w), (y_w, 2y_w), \dots, (1 - y_w, 1) \rangle$. Y_{ij} refers to the subsequence

of Y that includes all elements indexed between i and j , and f_i is the lower frequency in the i -th tuple while f_j is the higher frequency in the j -th tuple. We choose $y_w = 0.01$ such that there are a large number of alleles within each bin.

The S_{β} statistic is sensitive to both mutational bias and selection, although these two distinct evolutionary processes drive distinct patterns in a β DAF plot. When mutational bias acts on new mutations in the absence of selection on the trait, there is no expectation of a relationship between the magnitude of effects and allele frequency, and hence the expected value of β is the same in all allele frequency bins (Fig. S2). This means that the expected value of S_{β} is simply equal to the mean effect size of new mutations multiplied by the number of bins, and that the sign of S_{β} is equal to the sign of the bias in mutational effects.

In the absence of mutational bias, S_{β} is sensitive to some (but not all) selection models. When a trait is under long-term stabilizing selection with no change in the optimal phenotype and no mutational bias, trait-increasing alleles are equally deleterious and equally likely to occur as trait-decreasing alleles, meaning that the expectation of S_{β} is 0 (which concurs exactly with negative selection models—Fig. S3, purple curves). However, any bias in mutation rate toward trait-increasing or -decreasing alleles will

drive S_{β} to have non-zero expectation (Fig. S3). In contrast to the neutral case with mutational bias, selection in conjunction with mutational bias causes the mean value of β to be greater in magnitude in bins of low derived allele frequency than high allele frequency, since selection will tend to constrain alleles with the largest effects to the lowest frequencies (Fig. 1C–E, red curves; Fig. S3).

The most interesting patterns emerge when stabilizing selection acts on a trait in conjunction with shifts in the fitness optimum. When stabilizing selection acts on a trait under along with mutational bias ($\delta = 0.4$) and no shift in the optimum, $S_{\beta}(0, x)$ is negative at all x , and decreases as a function of frequency (red curve, Fig. 1E; we show in Fig. S3 and that this same pattern holds under a well-studied directional selection model for complex traits (Eyre-Walker 2010)). When shifts in the optimum toward higher trait values occur, higher frequency variants have mean positive effect sizes, causing $S_{\beta}(0, x)$ to be non-monotonic and $S_{\beta}(0, 1)$ to potentially have positive sign. As time elapses since the shift, the high frequency trait-increasing alleles will tend to drift to the boundary and fix or be lost, causing this signal to gradually disappear (Fig. 1E). Note that mutational bias in the absence of selection can generate a non-zero S_{β} —however, mutational bias alone induces $S_{\beta}(0, x)$ to increase in magnitude monotonically and linearly in x , a pattern that is not expected for selection (Fig. S2 and Supporting Information). We use model-based analyses to tease apart mutational bias and selection effects in later sections.

A PERMUTATION PROCEDURE TO GENERATE A NULL DISTRIBUTION

While we have noted that S_{β} has an expected value of 0 under the neutral null without mutational bias, the variance of S_{β} depends on linkage between causal alleles and noncausals, since noncausal alleles will have non-zero estimated effect sizes when they are in Linkage disequilibrium (LD) with a causal allele. To control for potential confounding by LD, we developed a simple permutation-based procedure for computing the null-distribution of frequency-effect size relationships in GWAS summary statistics under a neutral evolutionary model. We first polarize all alleles such that the derived allele is the causal allele. Then, for each of 1703 previously identified approximately independent linkage blocks (Berisa and Pickrell 2016), we select a random sign (positive or negative with equal probability), and multiply all the effect sizes in the LD block by this sign. We then recompute the test statistic of interest, such as the correlation between frequency and MAF, on the randomized data. This generates a null distribution for the test statistic that conservatively accounts for the linkage between inferred effect sizes, and maintains the frequency spectrum and marginal distribution of the magnitude of inferred effect sizes.

APPLICATION TO GWAS SUMMARY DATA

We applied S_{β} within our framework to GWAS summary data for BMI, height (Wood et al. 2014), and educational attainment (Okbay et al. 2016) to assess its power for detecting selection, given that previous studies have suggested that these phenotypes may be under selection (Turchin et al. 2012; Berg and Coop 2014; Robinson et al. 2015; Field et al. 2016; Racimo et al. 2018). We observe that effect sizes are correlated with frequency, and that S_{β} is a non-monotonic function of frequency (Fig. 2), consistent with selection and mutational bias (Fig. 1E; note that the large spike in both BMI and height at high frequency can be explained by ancestral uncertainty, Fig. S7). We reject the neutral null for all three traits ($P < 5 \times 10^{-4}$; Fig. 2 and Table 1). We then applied our method to six additional phenotypes, which we selected to span a wide range of phenotypes that we hypothesized might be targets of selection, including body size (Wood et al. 2014), psychiatric conditions (CDG Psychiatric Genomics Consortium 2013), immune-related traits (Franke et al. 2010), reproductive traits (Day et al. 2015), and cardiovascular traits (Willer et al. 2013). We find an additional nonneutral signal for Crohn's disease, and a marginally significant signal for schizophrenia that narrowly missed a multiple testing correction (Table 1; Figs. S8–S16). When including only common variants in the test, we find that seven of nine phenotypes have P -values under 0.1, suggesting strong enrichment for selected traits among the test set despite failure for some of the tests to exceed a multiple testing correction (binomial P -value 3.0×10^{-6}).

EVOLUTIONARY MODELS FOR SELECTION ON HUMAN TRAITS

While our results show that S_{β} has a strong nonneutral signature for five of the nine phenotypes, we sought to further understand the evolutionary models that could explain these signals. Purifying selection alone seems an unlikely candidate for most of the phenotypes, because the sign of S_{β} is always the same as the sign of $\bar{\beta}$ under a purifying selection model (Fig. S3 and Supporting Information), a pattern violated by height, BMI, Crohn's, and educational attainment. Models of directional selection for increased or decreased phenotype values in the absence of mutational bias also share this pattern, as do models of mutational bias in the absence of selection (Figs. S1 and S2; Supporting Information).

We hypothesized that these signals could be explained by a model of stabilizing selection, mutational bias, and shifts in the trait value conferring optimal fitness. We developed an evolutionary inference procedure based on rejection sampling (Tavaré et al. 1997) to infer the parameters that best fit the relationship between frequency and effect size that we observe in the data (see Supporting Information). Briefly, we calculate $S_{\beta}(0.01, x)$ (scaled by $S_{\beta}(0.01, 0.99)$) for a range of derived allele frequencies x , which we then use as summary statistics for rejection sampling. We

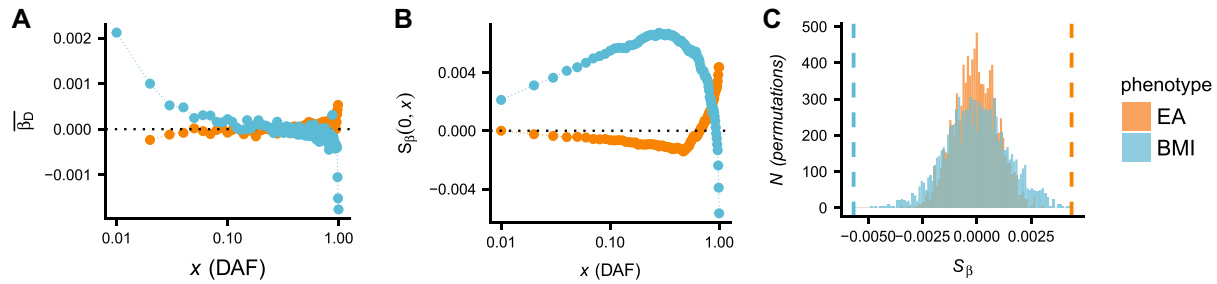


Figure 2. A: $\hat{\delta}\beta$ as a function of DAF for BMI and educational attainment (EA). B: $S_{\beta}(0, x)$ for the same data. C: neutral null distribution of $S_{\beta}(0, 1)$ obtained by permutations. The vertical dashed line indicates the observed value of $S_{\beta}(0, 1)$ in the GWAS summary data.

Table 1. *P*-values corresponding to GWAS summary statistics for nine phenotypes that we hypothesized may be under selection. Values in the first column include all alleles, while the second and third columns correspond to tests including only alleles with MAF > 1% and MAF > 5%, respectively. The UK Biobank tests were performed on all alleles above 1% in frequency.

Phenotype	$S_{\beta}(0, 1)$ <i>P</i> -value	$S_{\beta}(0.01, 0.99)$ <i>P</i> -value	$S_{\beta}(0.05, 0.95)$ <i>P</i> -value	UKBB replication
Height	<0.0005**	<0.0005**	<0.0005**	0.416
BMI	<0.0005**	<0.0005**	<0.0005**	0.0095
Education	<0.0005**	<0.0005**	<0.0005**	<0.0005
WHR-BMI	0.566	0.8325	0.341	NA
GLL	0.4655	0.434	0.5645	NA
Crohn's disease	0.0025**	0.0075*	0.018*	NA
Menopause onset	0.1585	0.46	0.0475*	NA
Depression	0.3915	0.01*	0.0305*	NA
Schizophrenia	0.0085*	0.0035**	0.0625	NA

*Tests that pass a multiple testing correction ($P < 0.005$).

**Tests that were marginally significant ($P < 0.05$).

remove the lowest frequency variants (i.e., those with $x < 0.01$) for the purpose of this inference to avoid the potential impact of rare variant stratification on our results. We validated our method with extensive simulations, and found that it is a noisy estimator of the magnitude of δ and $\Delta\phi$ (Fig. 3A and B—recall that δ corresponds to the proportion of new mutations that are trait-increasing), but has excellent power to estimate the direction of both mutational bias and shift in optimal trait value (Fig. 3C and D). Other parameters of the model (including heritability, polygenicity, effect size distribution, and the time of the shift in the fitness landscape; see Supporting Information) were inferred with low accuracy as indicated by only modest correlations between inferred and true parameter values, indicating that the summary statistics we use contain little information about these parameters. Inferred distributions of these parameters therefore are not reported.

We find strong posterior support for a shift in optimal trait value ($\Delta\phi$, measured in units of SDs of the population trait distribution) for all five phenotypes, as well as strong signals of mutational bias (δ). Data for height and educational attainment supported a shift toward increased trait values and a mutational bias toward mutations that decrease the phenotype, while Crohn's

disease and BMI supported shifts toward lower trait values and mutational bias toward trait-increasing alleles (i.e., risk-increasing for Crohn's). Schizophrenia data supported a shift toward lower risk, and a bias toward protective mutations, although a substantial minority of parameter estimates supported no mutational bias or a bias in the opposite direction (Fig. 3E). To further validate these findings, we resampled from the inferred parameter distributions and performed an independent set of forward simulations. We find that the summary statistics computed on these out-of-sample simulations (which were not used to fit the data) match the trends observed in our data, confirming that the modeling framework is capable of recapitulating the signals we observe in the data (Fig. 3G).

POPULATION STRATIFICATION

Although S_{β} is not sensitive to confounding by ancestral uncertainty, we were concerned that these signals could potentially be explained by other confounders, such as uncorrected population stratification. Although recent studies have suggested that methods to account for stratification in GWAS often over-correct (Field et al. 2016; Bulik-Sullivan et al. 2015), rare variant stratification remains especially difficult to account for in GWAS (Mathieson

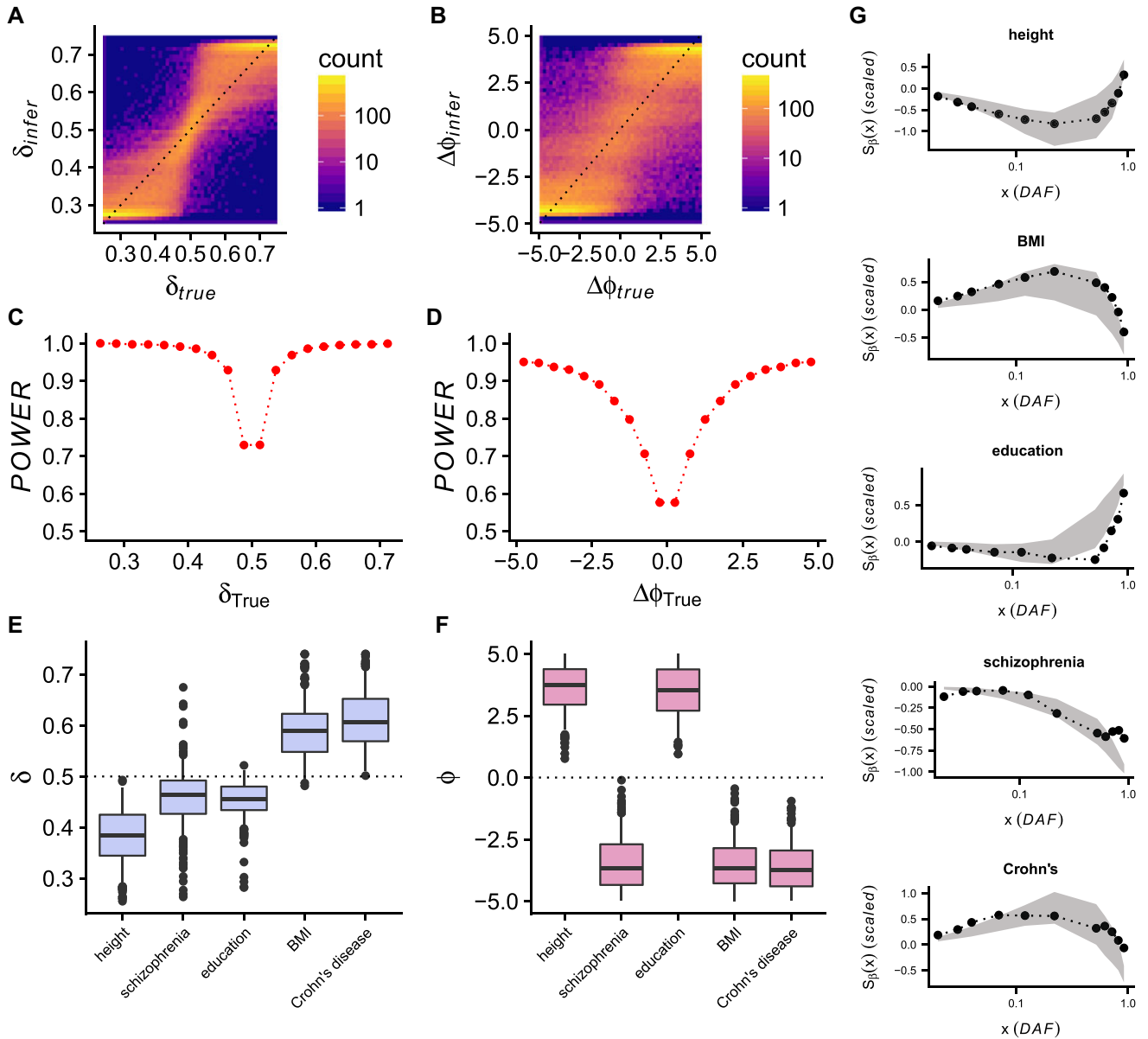


Figure 3. A-B: Inferred mutation bias (A) and selection shift (B) parameters as a function of true parameter values for our rejection sampling method. C-D: Power of our rejection sampling method to correctly identify the direction of mutation bias (C) and shift in optimal phenotype value (D), as a function of the true parameter value. E-F: Inferred approximate posterior distributions for five phenotypes that were identified as non-neutral. G: Out-of-sample simulations using parameters inferred in E-F, plotted with the data used to fit each model. Gray envelopes represent the middle 50% of simulation replicates, while the black points and curves show the observed data for each phenotype.

and McVean 2012). We recomputed S_{β} on alleles with $MAF > 1\%$ and $MAF > 5\%$, and found that the signals are robust within common alleles alone, suggesting that population structure is unlikely to confound our estimates (Figs. S8–S16 and Table 1). Additionally, we recomputed S_{β} on GWAS from the UK Biobank (UKBB) for our top three signals, a replication cohort that employed strict population structure control in a more homogeneous group of samples. We replicate the signals for both educational

attainment ($P < 5 \times 10^{-4}$) and BMI ($P = 0.0095$), noting that the educational attainment data in our two datasets contain some overlapping samples and hence are only partially independent and the BMI signal is somewhat weaker in UKBB than GIANT (Supporting Information and Fig. S17). Interestingly, height was not replicated. We also applied two previous approaches that detected signals of selection acting on height when analyzing the GIANT summary statistics and found that neither replicated in the UKBB

(Turchin et al. 2012; Yang et al. 2015; Fig. S6). These results are consistent with either overcorrection of structure within the UKBB or undercorrection in the GIANT data, although the latter seems more likely based on other recent studies (Berg et al. 2018; Sohail et al. 2018—see Discussion).

Discussion

Many studies have suggested that selection shapes human genetic variation (e.g., Fay et al. 2001), and recent work has suggested that selection on complex traits may be a substantial driver of human adaptation (Hernandez et al. 2011). Here, we developed a novel empirical framework and a model-based rejection sampling approach for detecting polygenic selection and mutational bias that can be applied to GWAS summary data for a single population. We call this approach an “evolutionary compass,” because orienting alleles by their ancestral/derived status within our framework provides insight into the evolutionary processes shaping complex traits. We applied this evolutionary compass to GWAS summary data for nine phenotypes, and showed that five of them (educational attainment, height, Crohn’s disease, BMI, and schizophrenia) are consistent with a model of selection and mutation bias in shaping trait variation. Interestingly, among the top three signals that were uncovered with our method, height did not replicate in a more homogeneous cohort, while both BMI and educational attainment were replicated.

If selection acts on biomedically relevant complex traits such as Crohn’s disease and schizophrenia, there are important implications for the future of both medical and evolutionary genomics. In medical genomics, an ongoing debate about the genomic architecture of complex diseases is at the forefront of the field (Manolio et al. 2009). When strong selection acts on complex traits, it can elevate the role of rare alleles in driving trait variance (Lohmueller 2014). If rare alleles contribute a larger fraction of the genetic variance than is expected under neutral models, then very large GWAS that use only array-based genotyping information are very unlikely to be able to capture these signals, and sequence-based studies and powerful rare variant approaches that are robust to evolutionary forces (including those not investigated here, such as partial recessivity) will be needed (Uricchio et al. 2016; Hernandez et al. 2017; Sanjak et al. 2017). Moreover, recent work has suggested that the over-representation of Europeans in GWAS has limited the effectiveness of estimating polygenic risk scores in other human populations (Martin et al. 2017). This is problematic for the transfer of genomic research into the clinic, where precision medicine initiatives relying on personal genetic information will be most successful if genetic risk can be accurately predicted in diverse populations. While this inability to predict across populations could be driven by neu-

tral demographic forces, if selection has driven numerous phenotypes to acclimate to local environmental conditions in ancestral human populations worldwide it could exacerbate this problem dramatically.

In the field of human evolutionary genomics, most studies have agreed that the impact of selection is widespread on the human genome, but the evolutionary mechanisms that drive genetic and phenotypic diversity have been widely debated (Hernandez et al. 2011; Enard et al. 2014; Schrider and Kern 2017). In our study, we showed that GWAS summary statistics in Europeans for BMI and Crohn’s disease are consistent with a bias in mutation rate toward trait-increasing alleles, and a shift to a lower optimal value of the trait, while educational attainment is consistent with a mutational bias toward trait-increasing alleles and a shift toward higher values of the trait optimum. The signal for schizophrenia is consistent with an ancestral shift toward a lower optimum and a stabilizing selection, with or without a mutational bias. It should be noted that the model we used to fit these data assumed no more than one shift in the optimal phenotype value, whereas this quantity is likely to vary continuously with environmental conditions for some traits. Models that additionally account for sexual dimorphism (Stulp and Barrett 2016), higher dimensional trait spaces (Simons et al. 2018), and evolutionary history of multiple populations (Berg and Coop 2014; Racimo et al. 2018) may be required to better understand the generality of these results across human populations and traits.

Neanderthal introgression into modern humans has played an important role in shaping traits in non-African populations (Wall et al. 2013; McCoy et al. 2017). Given that Neanderthal alleles may contribute disproportionately to the genetic variance in some traits (Simonti et al. 2016) and that some high frequency trait-associated alleles have Neanderthal origins (Dannemann and Kelso 2017; Prüfer et al. 2017), we hypothesized that Neanderthal alleles for traits under selection might show distinct patterns from modern human alleles. Although Neanderthal alleles do not share a common demographic history with modern human alleles, under the neutral null hypothesis we do not expect Neanderthal alleles to have an increasing or decreasing frequency and effect-size relationship, or to have a distribution that differs substantially from modern human alleles. We, therefore, computed S_{β} on alleles that have a Neanderthal origin (see Supporting Information). Alleles for height and depression show strikingly different patterns than alleles with modern human origins (Fig. S18). The results for height can be explained by selection to promote Neanderthal height-increasing alleles to high frequency, either along the Neanderthal lineage predating human introgression, or after admixture with human populations. In contrast, our results for depression risk are consistent with an excess of depression risk from Neanderthals (Simonti et al. 2016),

and selection preferentially driving large effect alleles to low frequency. We note that while Neanderthal alleles are not subject to the same biases in ancestral state uncertainty as modern human alleles, inferences of selection could still be biased by population stratification.

Inference of selection on complex traits is vulnerable to several possible confounders, including population stratification and pleiotropic selection on off-target phenotypes (Novembre and Barton 2018). Although theory suggests that stratification should be straightforward to detect and correct at high frequency variants in large samples (Patterson et al. 2006), an uncorrected bias in the inferred β values due to population structure can make our test (as well as most others, such as Berg and Coop 2014; Yang et al. 2015; Field et al. 2016; Berg et al. 2017; Racimo et al. 2018) anti-conservative. We used a series of experiments to show that population structure is unlikely to bias the majority of our results, including showing that the signal is robust to the exclusion of rare alleles and performing a replication study in an external cohort. However, one of our strongest signals (height) did not replicate the UK Biobank, while two other signals of selection suggested by earlier height studies also did not replicate in the UK Biobank (Turchin et al. 2012; Yang et al. 2015). The nonreplication of height is in concordance with other recent studies finding reduced evidence for selection on height in the UK Biobank cohort applying different methods (Berg et al. 2018; Sohail et al. 2018). The most conservative interpretation of the nonreplication of height is to suppose that some of the signals we and others observed in the GIANT cohort are driven by population stratification, and the UK Biobank analysis correctly removes this spurious contamination. Further research is needed to better disentangle stratification and selection, and caution in interpreting the results of polygenic selection tests is warranted while this field develops, even in more homogeneous cohorts such as the UK Biobank (Novembre and Barton 2018).

Pleiotropy can also induce biases in complex trait selection detection. Selection on a trait that has correlated effect sizes with another trait could result in false positives, in which the neutral trait is spuriously identified. This is a general limitation of most complex trait selection methods (but see Berg et al. 2017; Simons et al. 2018). This phenomenon is clearly highlighted by our results on educational attainment, a phenotype that had no meaning until recent historical times. While it might be tempting to identify a cognition-related phenotype as the target of selection, it is possible that any trait with a cryptic shared genetic basis and correlated effect sizes could be the target, and that the timing of any such shift to larger trait values could have predated the human migration out-of-Africa. Thus, we cannot rule out a role for correlated phenotypes in driving these signals, and our results do not imply differences in phenotypes or polygenic scores between Europeans and any other group. More work on disentangling se-

lection targets with a common genomic basis will be needed as the field progresses.

Among the nine traits that we tested, we found that four had strong non-neutral signals, and six of the nine had marginal evidence for nonneutrality (we have conservatively removed height, given its nonreplication in the UK Biobank). However, this does not imply that the others are not subject to selection or mutation bias. The power of our test depends on the strength of selection, the polygenicity of the trait, the heritability of the trait, the mutational bias, and the amount of ancestral uncertainty, in addition to the sample size of the GWAS. If a trait is under strong stabilizing selection, but the mutation rate of trait-increasing and -decreasing alleles is exactly equal, then our test has no power. Because we rely on the polarization of alleles by ancestral state, increased uncertainty in ancestral state will also decrease our power (see Supporting Methods). Moreover, if selection is weak, a small number of causal alleles drive variance in the trait, or the trait is only weakly heritable, power is greatly diminished. However, increased sample sizes in GWAS will increase power, because the variance of effect size estimates for even weak effect causal alleles decreases with sample size. In addition to working directly on GWAS summary statistics from a single population, one strength of our permutation-based approach is that other informative statistics, such as the absolute value of the deviation between ancestral and derived effect sizes, could also easily be applied, and may have higher power. In future studies, it will be advantageous to compare various summary statistics and to apply our approach to species undergoing rapid environmental changes.

Materials and Methods

SOFTWARE

We wrote software in Python to implement the statistical test described herein, and we developed a custom simulator of demography and selection. The code to run our statistical test is freely available online (<https://github.com/uricchio/PASTEL>). The simulations (and their validation) are described in the Supporting Information.

ACKNOWLEDGMENT

We acknowledge the support of NIGMS grant K12GM088033 and the Stanford IRACDA program. LHU was partially supported by NIH HG005855 and NSF DBI-1458059 (to Noah Rosenberg). N.Z. was funded by NIH grants K25HL121295, U01HG009080, and R01HG006399, and A.G. was supported by the Claudia Adams Barr Award. We thank our anonymous reviewers, Noah Rosenberg, Jeremy Berg, Doc Edge, Graham Coop, and Josh Akey for comments that substantially improved the manuscript. We also thank Jonathan Pritchard, Arbel Harpak, Yair Field, Ziyue Gao, and members of the Rosenberg, Coop, and Pritchard labs for helpful comments, as well as Josh Akey and Rajiv McCoy for providing a list of inferred Neanderthal alleles. We thank Nicole R. Gay for assistance in validating the software.

AUTHOR CONTRIBUTIONS

Conceived the study: NZ & AG. Performed the analysis and simulations: LHU & HCK. Wrote the paper: LHU. Designed computational experiments, and edited/approved final version: all authors.

LITERATURE CITED

- Barton, N. 1986. The maintenance of polygenic variation through a balance between mutation and stabilizing selection. *Genet. Res.* 47:209–216.
- Berg, J. J., and G. Coop. 2014. A population genetic signal of polygenic adaptation. *PLoS Genet.* 10:e1004412.
- Berg, J. J., A. Harpak, N. Sinnott-Armstrong, A. M. Jorgensen, H. Mostafavi, and Y. Field. 2018. Reduced signal for polygenic adaptation of height in UK Biobank. *bioRxiv* 35491. <https://doi.org/10.1101/354951>.
- Berg, J. J., X. Zhang, and G. Coop. 2017. Polygenic adaptation has impacted multiple anthropometric traits. *bioRxiv* 167551. <https://doi.org/10.1101/167551>.
- Berisa, T., and J. K. Pickrell. 2016. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 32:283–285.
- Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez, K. E. Lohmueller, et al. 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* 4:e1000083.
- Bulik-Sullivan, B. K., P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, N. Patterson, et al. 2015. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47:291–295.
- CDG Psychiatric Genomics Consortium. 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381:1371–1379.
- Charlesworth, B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* 63:213–227.
- . 2013. Stabilizing selection, purifying selection, and mutational bias in finite populations. *Genetics* 194:955–971.
- Dannemann, M., and J. Kelso. 2017. The contribution of neanderthals to phenotypic variation in modern humans. *Am. J. Hum. Genet.* 101:578–589.
- Day, F. R., K. S. Ruth, D. J. Thompson, K. L. Lunetta, N. Pervjakova, D. I. Chasman, et al. 2015. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat. Genet.* 47:1294–1303.
- Edge, M. D., and G. Coop. 2018. Reconstructing the history of polygenic scores using coalescent trees. *Genetics* 211:235–262.
- Enard, D., P. W. Messer, and D. A. Petrov. 2014. Genome-wide signals of positive selection in human evolution. *Genome Res.* 24:885–895.
- Eyre-Walker, A. 2010. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc. Natl. Acad. Sci. USA* 107:1752–1756.
- Fay, J. C., G. J. Wyckoff, and C.-I. Wu. 2001. Positive and negative selection on the human genome. *Genetics* 158:1227–1234.
- Field, Y., E. A. Boyle, N. Telis, Z. Gao, K. J. Gaulton, D. Golan, et al. 2016. Detection of human adaptation during the past 2000 years. *Science* 354:760–764.
- Franke, A., D. P. McGovern, J. C. Barrett, K. Wang, G. L. Radford-Smith, T. Ahmad, et al. 2010. Genome-wide meta-analysis increases to 71 the number of confirmed crohn's disease susceptibility loci. *Nat. Genet.* 42:1118–1125.
- Fryar, C. D., Q. Gu, C. L. Ogden, and K. M. Flegal. 2016. Anthropometric reference data for children and adults: United States, 2011–2014. *Vital Health Stat.* 3 (39):1–46.
- Gazal, S., H. K. Finucane, N. A. Furlotte, P.-R. Loh, P. F. Palamara, X. Liu, et al. 2017. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* 49:1421.
- Gilad, Y., A. Oshlack, and S. A. Rifkin. 2006. Natural selection on gene expression. *Trends Genet.* 22:456–461.
- Gravel, S., B. M. Henn, R. N. Gutenkunst, A. R. Indap, G. T. Marth, A. G. Clark, et al. 2011. Demographic history and rare allele sharing among human populations. *Proc. Natl. Acad. Sci. USA* 108:11983–11988.
- Hernandez, R. D., J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton, G. McVean, 1000 Genomes Project, et al. 2011. Classic selective sweeps were rare in recent human evolution. *Science* 331:920–924.
- Hernandez, R. D., L. H. Uricchio, K. Hartman, J. Ye, A. Dahl, and N. Zaitlen. 2017. Singleton variants dominate the genetic architecture of human gene expression. *bioRxiv* 219238. <https://doi.org/10.1101/219238>.
- Jain, K., and W. Stephan. 2017. Rapid adaptation of a polygenic trait after a sudden environmental shift. *Genetics* 206:389–406.
- Lohmueller, K. E. 2014. The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genet.* 10:e1004379.
- Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorf, D. J. Hunter, et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747–753.
- Martin, A. R., C. R. Gignoux, R. K. Walters, G. L. Wojcik, B. M. Neale, S. Gravel, et al. 2017. Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* 100:635–649.
- Mathieson, I., and G. McVean. 2012. Differential confounding of rare and common variants in spatially structured populations. *Nat. Genet.* 44:243–246.
- McCoy, R. C., J. Wakefield, and J. M. Akey. 2017. Impacts of neanderthal-introgressed sequences on the landscape of human gene expression. *Cell* 168:916–927.
- McVicker, G., D. Gordon, C. Davis, and P. Green. 2009. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* 5:e1000471.
- Messer, P. W., and D. A. Petrov. 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol. Evol.* 28:659–669.
- Novembre, J., and N. H. Barton. 2018. Tread lightly interpreting polygenic tests of selection. *Genetics* 208:1351–1355.
- Okbay, A., J. P. Beauchamp, M. A. Fontana, J. J. Lee, T. H. Pers, C. A. Rietveld, et al. 2016. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533:539–542.
- Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2:e190.
- Pritchard, J. K., J. K. Pickrell, and G. Coop. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* 20:R208–R215.
- Prüfer, K., C. de Filippo, S. Grote, F. Mafessoni, P. Korlević, M. Hajdinjak, et al. 2017. A high-coverage Neanderthal genome from Vindija cave in Croatia. *Science* 358:655–658.
- Racimo, F., J. J. Berg, and J. K. Pickrell. 2018. Detecting polygenic adaptation in admixture graphs. *Genetics* 208:1565–1584.
- Robertson, A. 1956. The effect of selection against extreme deviants based on deviation or on homozygosity. *J. Genet.* 54:236–248.
- Robinson, M. R., G. Hemani, C. Medina-Gomez, M. Mezzavilla, T. Esko, K. Shakhbazov, et al. 2015. Population genetic differentiation of height and body mass index across Europe. *Nat. Genet.* 47:1357–1362.
- Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. Levine, D. J. Richter, S. F. Schaffner, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837.
- Sabeti, P. C., S. F. Schaffner, B. Fry, J. Lohmueller, P. Varilly, O. Shamovsky, et al. 2006. Positive natural selection in the human lineage. *Science* 312:1614–1620.
- Sanjak, J. S., A. D. Long, and K. R. Thornton. 2017. A model of compound heterozygous, loss-of-function alleles is broadly consistent with observations from complex-disease GWAS datasets. *PLoS Genet.* 13:e1006573.
- 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.

- Schoech, A., D. Jordan, P.-R. Loh, S. Gazal, L. O'Connor, D. J. Balick, et al. 2017. Quantification of frequency-dependent genetic architectures and action of negative selection in 25 UK Biobank traits. *bioRxiv* 188086. <https://doi.org/10.1101/188086>.
- Schrider, D. R., and A. D. Kern. 2017. Soft sweeps are the dominant mode of adaptation in the human genome. *Mol. Biol. Evol.* 34:1863–1877.
- Simons, Y. B., K. Bullaughey, R. R. Hudson, and G. Sella. 2018. A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biol.* 16:e2002985.
- Simonti, C. N., B. Vernot, L. Bastarache, E. Bottinger, D. S. Carrell, R. L. Chisholm, et al. 2016. The phenotypic legacy of admixture between modern humans and neandertals. *Science* 351:737–741.
- Sohail, M., R. Maier, A. Ganna, A. Bloemendal, A. Martin, M. Daly, et al. 2018. Signals of polygenic adaptation on height have been overestimated due to uncorrected population structure in genome-wide association studies. *bioRxiv* 355057. <https://doi.org/10.1101/355057>.
- Stulp, G., and L. Barrett. 2016. Evolutionary perspectives on human height variation. *Biol. Rev.* 91:206–234.
- Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly, et al. 1997. Inferring coalescence times from dna sequence data. *Genetics* 145:505–518.
- Turchin, M. C., C. W. Chiang, C. D. Palmer, S. Sankararaman, D. Reich, J. N. Hirschhorn, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, et al. 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* 44:1015–1019.
- Uricchio, L. H., N. A. Zaitlen, C. J. Ye, J. S. Witte, and R. D. Hernandez. 2016. Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants. *Genome Res.* 26:863–873.
- Voight, B. F., S. Kudravalli, X. Wen, and J. K. Pritchard. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.
- Wall, J. D., M. A. Yang, F. Jay, S. K. Kim, E. Y. Durand, L. S. Stevison, et al. 2013. Higher levels of Neanderthal ancestry in east Asians than in Europeans. *Genetics* 194:199–209.
- Willer, C. J., E. M. Schmidt, S. Sengupta, G. M. Peloso, S. Gustafsson, S. Kanoni, et al. 2013. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* 45:1274–1283.
- Wood, A. R., T. Esko, J. Yang, S. Vedantam, T. H. Pers, S. Gustafsson, et al. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46:1173–1186.
- Yang, J., A. Bakshi, Z. Zhu, G. Hemani, A. A. Vinkhuyzen, S. H. Lee, et al. 2015. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* 47:1114–1120.
- Zeng, J., R. de Vlaming, Y. Wu, M. Robinson, L. Lloyd-Jones, L. Yengo, C. Yap, et al. 2018. Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.* 44:1015–1019.

Associate Editor: S. Wright

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. A: When selection does not prefer either trait-increasing or trait-decreasing alleles, and mutation bias does not act on a trait, then trait-increasing (pink) and -decreasing (blue) alleles are expected to have identical frequency spectra and equal mean β values within each frequency bin.

Figure S2. The relationship between β and DAF when only mutational bias acts on the trait (i.e., there is no selection).

Figure S3. A: A β DAF plot for various values of mutational bias (δ) towards trait decreasing alleles for traits under selection.

Figure S4. A comparison of simulated frequency spectra for a complex demographic/selection model for our Wright-Fisher simulator (black) and SFS_CODE (red).

Figure S5. Conservative False Positive Rate estimation for detection of mutation rate bias and shifts in optimal phenotype value.

Figure S6. Correlation between mean effect size and allele frequency (A) and correlation the difference in frequency between northern and southern Europe for the height increasing allele and p-value rank (B) in the GIANT study (gold) and the UK Biobank (gray).

Figure S7. The impact of ancestral uncertainty on the observed value of mean β as a function of derived allele count x .

Figure S8. S_β for height.

Figure S9. S_β for BMI.

Figure S10. S_β for educational attainment.

Figure S11. S_β for Crohn's disease.

Figure S12. S_β for schizophrenia.

Figure S13. S_β for global lipid levels.

Figure S14. S_β for menopause onset.

Figure S15. S_β for major depression.

Figure S16. S_β for waist-hip ratio adjusted for BMI.

Figure S17. S_β for BMI (A-C), height (D-F), and educational attainment (G-I) in the UK Biobank.

Figure S18. Signals of putative polygenic selection and mutation bias on Neanderthal alleles.

Figure S19. A-C: Simulations with the same parameters as Fig. 1C-E, but with the shift in optimal phenotype of $\Delta\phi = 2$ occurring linearly over 100 generations (2500 years), rather than instantaneously in a single generation.