Maintenance of Complex Trait Variation: Classic Theory and Modern Data

Evan M. Koch[1,2] and Shamil R. Sunyaev[1,2]*

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States, [2]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Numerous studies have found evidence that GWAS loci experience negative selection, which increases in intensity with the effect size of identified variants. However, there is also accumulating evidence that this selection is not entirely mediated by the focal trait and contains a substantial pleiotropic component. Understanding how selective constraint shapes phenotypic variation requires advancing models capable of balancing these and other components of selection, as well as empirical analyses capable of inferring this balance and how it is generated by the underlying biology. We first review the classic theory connecting phenotypic selection to selection at individual loci as well as approaches and findings from recent analyses of negative selection in GWAS data. We then discuss geometric theories of pleiotropic selection with the potential to guide future modeling efforts. Recent findings revealing the nature of pleiotropic genetic variation provide clues to which genetic relationships are important and should be incorporated into analyses of selection, while findings that effect sizes vary between populations indicate that GWAS measurements could be misleading if effect sizes have also changed throughout human history.

Keywords: population genetics, genome-wide association study, statistical genetics, evolution, quantitative genetics

1 INTRODUCTION

Attempts to understand genetic architecture preceded the discovery of DNA as the model of heredity (Fisher, 1918), and much theoretical work on selection, the maintenance of variation, and the adaptation of complex traits began before the ability to record genotypes on a scale sufficient to meaningfully contribute to these questions (Walsh and Lynch, 2018). The modern genetic era has provided an opportunity to test classic theories and to expand models—both long-standing and relatively recent—based on new understandings of genetic architecture and mechanisms. Genome-wide association studies (GWAS) and other data-driven tools have raised additional questions, including how so much heritability for many traits is contributed by relatively common alleles when natural selection is often expected to remove deleterious variation from the population. The flood of methods and data has sharpened and revised our understanding of many components that fashion the structure of the genome—polygenicity, selection, the distribution of mutational effects, pleiotropy—but has left us wanting for models capable of reconciling these elements (Sella and Barton, 2019).

The analysis of GWAS data revealed an extraordinary degree of polygenicity, and showed that most heritability is explained by relatively common, mostly noncoding alleles of small effect. At first

glance, this observation is surprising. Natural selection is expected to maintain the population near an optimum value for quantitative traits and to reduce the prevalence of potentially maladaptive phenotypes such as diseases. Such optimums and maladaptive phenotypes are defined within a given environmental context (Harpak and Przeworski, 2021). Selection generally acts by reducing the frequency of phenotypically relevant alleles, though it may drive allele frequency increases when shifts in the optimum phenotype occur. This basic logic led to the question whether the effect of natural selection is evident from GWAS data at all. Recent studies have reached a strong consensus that phenotypic effect sizes are negatively correlated with allele frequency (Gazal et al., 2018; Zeng et al., 2018; Schoech et al., 2019; Speed et al., 2020; Zeng et al., 2021). These findings are inconsistent with purely neutral models, but various models of natural selection influencing trait variation remain plausible (Walsh and Lynch, 2018). Uncertainty largely surrounds whether the focal trait is causally important for fitness compared to pleiotropically related ones, and whether selection is primarily stabilizing or has important directional components. In spite of many unresolved details, the emerging picture is that a vast supply of mutations with weak effects, coupled with generally inefficient selection against such alleles, is the basis of phenotypic variation.

Empirical results from GWAS on the distribution of effect sizes and allele frequencies still pose the challenge of which classic and emerging models from theoretical population genetics are able to best explain the emerging observations. Existing theories range from models of selection acting directly on the focal trait to models where selection on genetic variation is driven by simultaneous effects on other traits (pleiotropy), to even fully "apparent" selection, which assumes the focal trait is not subject to any selective constraint. In this review, we discuss a relevant subset of these models and how their predictions look in light of recent studies of selection in GWAS. We identify pleiotropy and variable effect sizes of genetic variants across time and space as important factors that have yet to be satisfactorily included into statistical methods and theoretical models.

2 THEORETICAL MODELS OF MAINTENANCE OF COMPLEX TRAITS AND PREDICTIONS THEY GENERATE

Evolutionary quantitative genetics has subsisted for most of its existence on a limited set of possible measurements. Estimates of the genetic and mutational variance, as well as selection gradients, are informative, especially with respect to contemporary patterns of selection. However, most progress in explaining maintenance of genetic variation in phenotypic traits was theoretical. Now that GWAS have generated an abundance of matched phenotypic and genetic measurements we live in a much more data-rich world. If we turn our attention to a single, focal trait, what sort of data would we ideally wish for? We would probably include the impact of genetic variants (estimated as their effect size) on the trait in a range of environments, the fitness effects of these alleles, as well as their frequencies and linkage patterns (Johnson and Barton,

2005). These would yield a satisfying and useful description of the genetic architecture and the process of its development, but there are fundamental details not immediately obvious from this description. We would like to know whether fitness effects arise primarily through selection on the focal trait, and if so what form it takes. If there are substantial fitness effects unrelated to the focal trait, what other traits are involved and how does selection act on them? Is the population in equilibrium? Has the genetic architecture changed in the past and will it do so in the future? Questions like these can be addressed by modeling how selection acts on traits, the mutational distributions underlying them, and how these generate the genetic architectures we observe.

Textbook introductions to population genetics begin by assigning fitnesses to genotypes and examine the consequences for allele frequencies and overall patterns of genetic variation. Connecting trait values, such as those measured in GWAS, to selection on individual causative alleles requires the additional step of specifying how selection on phenotypes leads to fitness differences among genotypes. While slightly less familiar than other selection results, this task was also taken up by many of the authors of classical population genetics and has grown into a large branch of evolutionary theory.

The simplest and most obvious model predicts the selection on individual causative loci arising from stabilizing selection on a single polygenic trait with purely additive genetic variance (Wright, 1935; Robertson, 1956; Bulmer, 1972) (**Figure 1A**). In this model an individual's trait value (z) is determined by the sum of effects from L independent loci: $z = \sum_{l=1}^{L} (\beta_l g_m + \beta_l g_p) + e$, where β_l is the effect size of the allele at locus l, $g_{l,m}$ and $g_{l,p}$ are the maternal and paternal genotypes at each locus, and e is a normally distributed environmental effect centered at zero. If an individual's fitness is a Gaussian function centered at the population mean M and with width V_S ($w(z) = \exp(-(M - z)^2/2V_S)$), then selection will change the average frequency of a causative allele at locus l with effect size $\beta_l = \beta$ as follows:

$$E[\Delta x] \approx -\frac{\beta^2}{2V_S} x(1-x)\left(\frac{1}{2} - x\right). \quad (1)$$

Stabilizing selection tends to remove genetic variation in this trait from the population. A balance between mutation, selection, and drift generates the trait's genetic variance in the population (Bulmer, 1972; Keightley and Hill, 1988). Such direct stabilizing selection leads to a negative correlation between minor allele frequencies and the effect size magnitude.

We can write the selection coefficient for this model as $s_{ud} = -\frac{\beta^2}{2V_S}$ to acknowledge that stabilizing selection takes the underdominant form shown in **Eq. 1** ($E[\Delta x] = sx(1-x)\left(\frac{1}{2} - x\right)$) rather than the more familiar additive one ($E[\Delta x] = sx(1-x)$). The $\left(\frac{1}{2} - x\right)$ term appearing in the underdominant formula means that selection against the derived allele actually decreases as it approaches 50% frequency and actually switches signs after that point. The minor allele is therefore always disfavored. However, when selection is strong or the allele frequency is low, the differences are minor as x is small compared to 1/2. We generally omit the subscript in s_{ud} for ease of
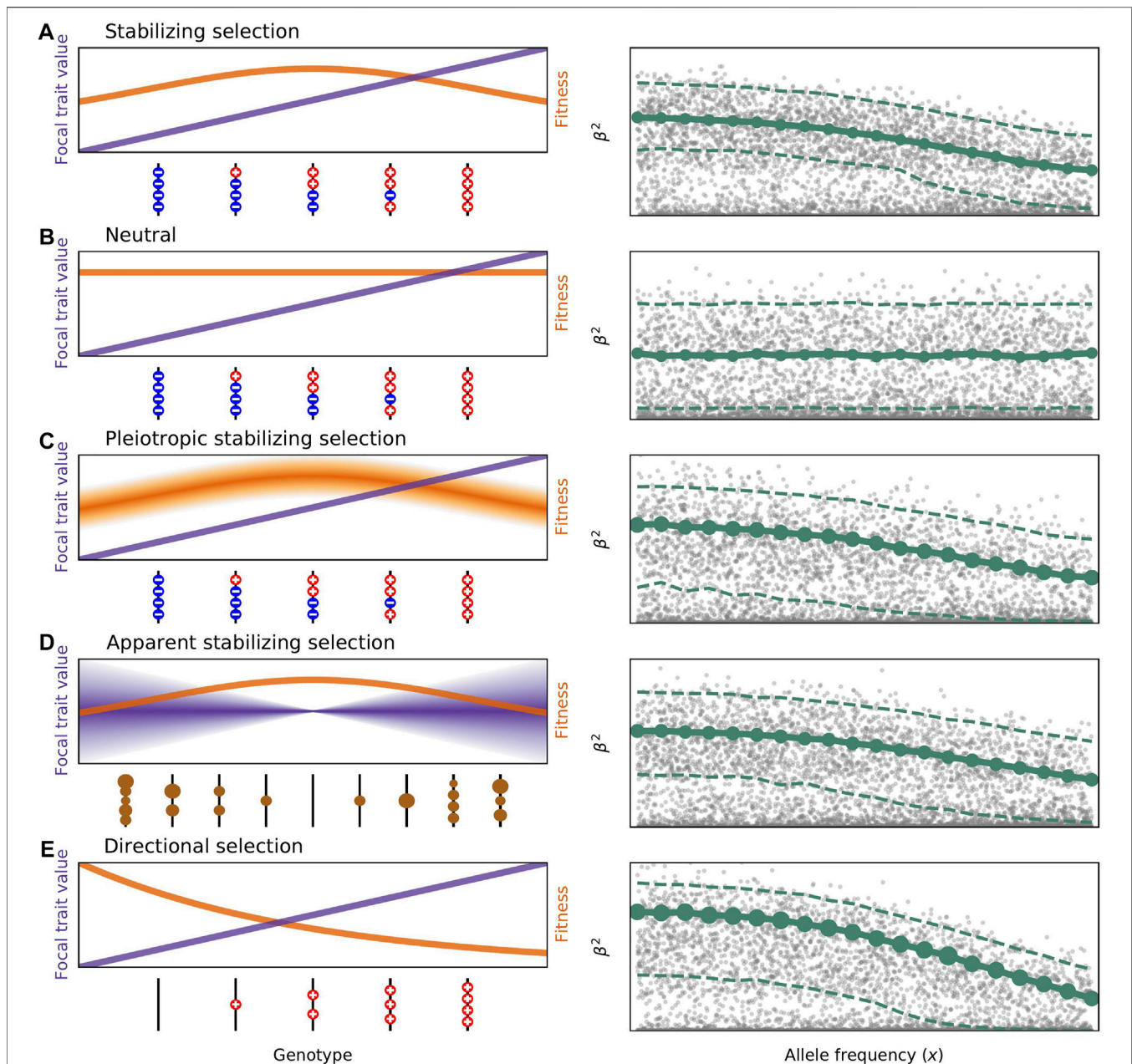
FIGURE 1 | Models of selection on the genetic variation influencing complex traits. Panels on the left show how different genotypes affect trait values (in purple) and fitness (in orange). Panels on the right illustrate how squared trait values change with frequency in each model of selection. Simulated values are shown in grey, the mean $E[\beta^2|x]$ is represented by the solid green line, and the standard deviation of $\beta^2|x$ is represented by the size of the green circles. The median and 97.5% quantile are shown as dashed lines to give a better sense of the full leptokurtic distribution of effect sizes. The DFE used in all plots (shape = 0.25, scale = 40) was chosen to be within the range fit by Schoech et al. (2019). Effect sizes were simulated uniformly on log frequency, and both axes are on a log scale. **(A)** Classic stabilizing selection as described by **Eq. 1**. Genotypes containing more trait-increasing alleles than decreasing, and vice versa, have lower fitness as a result of selection on the focal trait. Large effect alleles are prevented from reaching high frequencies due to the variance-reducing property of stabilizing selection. **(B)** In the neutral model no genotype is more fit than any other and the distribution of effect sizes at any frequency reflects only the distribution of mutational effects. **(C)** In pleiotropic stabilizing selection as studied by Simons et al. (2018), there is variation in fitness for each genotypic values depending on the effects mutations have on pleiotropic traits. This leads to the same average $E[\beta^2|x]$ but a greater variance and therefore different genetic architecture. **(D)** Models of apparent stabilizing selection first specify the deleterious fitness effects of mutations, represented here by the size of the brown circles. Genotypes with more and stronger deleterious mutations have a greater variance in phenotypic outcomes. This too leads to a negative relationship between β^2 and x. Here we use the Eyre-Walker (2010) model with $\tau = 0.4$ as fit by Schoech et al. (2019), and $\sigma^2 = 1$. Altering these would change the mean and variance of the (β^2, x) relationship. **(E)** Directional selection is shown here for a scenario where trait-decreasing mutations are unlikely or impossible. Selection therefore acts to reduce the frequency of trait-increasing alleles. All new mutations are disfavored with $s \propto \beta$. $E[\beta^2|x]$ again decreases with x.

reading, but it is important to note that the interpretation of selection coefficients differs depending on whether stabilizing selection is explicitly modeled or not.

The variance-reducing property of stabilizing selection motivated the development of other models with variance-promoting features like overdominant side-effects of causative alleles (Robertson, 1956; Gillespie, 1984) and strong mutational pressure (Lande, 1976b). As always in evolution, we must also at least consider the possibility that a trait of interest has a negligible impact on organismal fitness. The population mean value of a trait controlled by strictly neutral mutation will drift in Brownian motion and have a genetic variance that depends on the mutation rate, the second moment of the distribution of mutation effect sizes, and the average pairwise coalescent time between randomly sampled loci (Lande, 1976a; Lynch and Hill, 1986; Koch, 2019). $\mathrm{Var}[z] = \mathrm{E}[T_2]\theta\mu_2$, where T_2 is the average number of generations it takes for a pair of sites to coalesce, θ is the mutation rate per generation, and μ_2 is the second moment of the distribution of mutational effects. Crucially, there would be no relationship between the effect size and frequency of alleles (**Figure 1B**).

Of course, both intuitively and empirically, traits in natural and contemporary human populations at least appear to be under some selection (Kingsolver et al., 2001; Corbett et al., 2018; Sanjak et al., 2018), and involve some level of pleiotropy (Stearns, 2010). Models of apparent selection begin with the assumption that the focal trait is not itself under any selection but add pleiotropic fitness effects of the causative alleles. Individuals in the tails of a phenotypic distribution will carry more mutations overall, and if trait-affecting mutations have deleterious pleiotropic effects those individuals will also have lover fitness on average (Barton, 1990; Kondrashov and Turelli, 1992). Fitness that decreases away from the mean is reminiscent of stabilizing selection, but the strict deleterious model of apparent selection does not induce the negative correlation between allele frequencies and effect size magnitudes expected when the focal trait itself is actively selected. The negative correlation between β^2 and x may yet be rescued if the deleterious pleiotropic effects of variants affecting the focal trait arise from genetic covariance (Lande and Arnold, 1983) or correlated effect size magnitudes (Keightley and Hill, 1990). In this scenario, alleles with larger effects (or absolute magnitudes) on the focal, neutral trait are more likely to have larger effects on a second, selected trait. Allele frequencies are suppressed through selection on the second. In the extreme where the focal and selected trait are so closely biologically related that the effect sizes of mutations are deterministically linked, it is indistinguishable which trait causally impacts fitness, although a strong genetic covariance would be measurable. One can also imagine a model where each mutation has such a relationship with a unique pleiotropic trait, for instance, molecular effects in different pathways. Assuming that large-effect alleles for the focal trait induce stronger molecular effects, there may be strong selection without measurable genetic covariance between the focal trait and any individual pleiotropic trait.

The differences between models come down to how the statistical relationship between selection coefficients and effect sizes is specified: how s scales on average with β and what the random variation around this looks like. In multivariate stabilizing selection, s scales with β^2 as in direct stabilizing selection, but apparent selection models don't have this restriction. Apparent selection models were extended, as described above, to include increasing selection with greater β in addition to the negative pleiotropic consequences (Keightley and Hill, 1990; Zhang and Hill, 2002; Eyre-Walker, 2010) (**Figure 1D**). Models of multivariate stabilizing selection paint a similar picture, but the focal and pleiotropic traits are explicitly under stabilizing selection (Zhang and Hill, 2003; Simons et al., 2018) (**Figure 1C**). All lead to a negative (β^2, x) relationship, so differences between models come down to the shape and variance of that relationship, along with impacts on the genetic architecture.

Directional selection on complex disease susceptibility is also a viable hypothesis. In this view, the disease phenotype is itself deleterious and alleles that increase susceptibility will be selected against (Charlesworth, 2001; Wright et al., 2003) (**Figure 1E**). This also implies a mutational bias towards susceptibility-increasing alleles. It is plausible that there is a fitness cost associated with carrying such alleles, even for late-onset diseases (Pavard and Coste, 2021). All of the pleiotropy arguments made for stabilizing selection would apply equally well here.

There is an emerging consensus that models of mutation-selection-drift balance are likely to explain the genetic architecture of many, if not most, complex traits (Sella and Barton, 2019). The models of apparent, stabilizing, and directional selection described above, with varying possible degrees of pleiotropic selection, all remain possibilities within this consensus and are not mutually exclusive. Progress in statistical genetics methodology and increasing GWAS sample sizes are starting to clarify these details.

3 DETECTING NEGATIVE SELECTION IN GENOME-WIDE ASSOCIATION STUDIES DATA

As sample sizes increased and GWAS became sufficiently powered to detect larger numbers of loci for different traits, attention started shifting from the speculative question of how study design should be informed by selection and its effect on genetic architecture (Pritchard, 2001; Reich and Lander, 2001), to what the genetic architecture, as revealed through these studies, might say about selection. A transitional form was contributed by (Agarwala et al., 2013), who investigated how selection may have shaped the genetic architecture of Type 2 Diabetes, which had recently gone from 2 to 39 genome-wide significant loci. Using primarily the number of associations, conditional on the heritability and prevalence of the disease, they ruled out both neutrality of the focal trait and a model where selection is proportional to effect size: $\beta \propto |s|$.

Following this, methods were developed that do not explicitly model natural selection on causative variants, but ask whether lower frequency variants contribute disproportionately to

heritability. This heritability bias should only occur if rare variants have larger effect sizes on average, the most plausible explanation being negative selection correlated with the magnitude of effect sizes. A simple approach is to divide variants into MAF bins and estimate the heritability contribution of each in a mixed model framework (Yang et al., 2015). applied this approach to height and body mass index (BMI) and (Mancuso et al., 2016) to prostate cancer risk. Both found increased heritability in rare variants compared to common, and Mancuso et al. (2016) performed simulations to demonstrate that, conditional on disease heritability and prevalence, they could also rule out focal trait neutrality and directly proportional selection.

More sophisticated analyses using the same general idea as partitioning heritability by allele frequency have been developed and applied to a wide variety of human traits. Extensions of LD score regression (LDSC), a useful tool for partitioning heritability among large numbers of annotations (Finucane et al., 2015), were developed for features of negative selection. These analyses found that younger genetic variants contribute more heritability than older genetic variants at the same frequency (Gazal et al., 2017), a key feature of negative selection (Maruyama, 1974; Kiezun et al., 2013). They also confirmed earlier findings that rare variants have greater effect sizes than common ones for a larger number of traits (Gazal et al., 2018). Another popular and tractable approach, termed the alpha model, explicitly sets the MAF dependence of heritability contributions through a single parameter α: $E[\beta^2|x] \propto (x(1 - x))^\alpha$ (Zeng et al., 2018; Schoech et al., 2019; Speed et al., 2020; Zeng et al., 2021). An $\alpha < 0$ indicates a heritability bias towards rare variants and some amount of negative selection. Applications of this model have been remarkable in their consistently negative α estimates across all analyzed traits. While some differences between traits are inferred, e.g. height has a smaller α than BMI, estimates are consistently within the range [−0.5, −0.2].

The negative relationships between effect size magnitude and minor allele frequency inferred for so many traits are informative about the model of selection. In particular, they allow us to rule out neutral models where the focal trait and all underlying variation are unaffected by selection, as well as strict models of apparent selection where the causative variants are deleterious, but the strength of this selection is uncorrelated with effect sizes. However, many other models of selection may still be compatible with these findings (**Figure 1**, **Figure 2**). The model of direct stabilizing selection on a single trait first proposed by Wright and others (**Equation 1**) is plausible for some traits. On the other hand, a genetic correlation between the focal trait and another (or many) under stabilizing selection could produce the observed negative correlations even if the focal traits were completely neutral. There is also a lot of space in between with varying contributions to selection from the focal and pleiotropic traits. While seemingly semantic, the question is really about the extent to which variation in the underlying biology of the focal trait causes variation in fitness. This can apply even when the focal trait is something seemingly benign, an arbitrary bone for instance, whose size is governed chiefly by the biology of overall body size.

While the alpha model does not explicitly incorporate a population genetics model in any statistical analysis, it is possible to further interpret results using simulations and theory (**Figure 2**). In simulations, the idea is to use a model of choice to generate allele frequencies and effect sizes and then use the inference procedure to estimate what α corresponds to those model parameters. For theory, one derives $E[\beta^2|x]$ under the selection model and compare this to the approximate alpha model expectation of $E[\beta^2|x] \propto x^\alpha$. Using this approach, Schoech et al. (2019) showed that the inferred α depends both on the distribution of fitness effects of new alleles affecting the trait (DFE), and on the average scaling of effect sizes and selection $E[\beta^2|s] \propto s^{2\tau}$, where the parameter τ determines the scaling through the relationship $E[\beta] \propto s^\tau$ (Eyre-Walker, 2010). The DFE dependence enters primarily through a frequency-threshold effect: alleles below the threshold have effect sizes roughly uncorrelated with frequency because, while above the dependence scales approximately like $E[\beta^2|x] \propto x^{-2\tau}$. The threshold is the frequency below which most new mutations from the DFE are still mostly affected by drift rather than selection, and is therefore lower for a heavy-tailed DFE with a high average s. Zeng et al. (2021) used population genetic simulations to fit the DFE for various traits by conditioning on the values they had estimated for α, polygenicity, and SNP heritability. They found greater variation in the DFE among trait categories than variation in α estimates. α estimates that are relatively insensitive to the DFE are consistent with the predictions of Schoech et al. (2019) if most SNPs included in the analysis are above the frequency threshold where selection is detectable. These simulations therefore also demonstrate that polygenicity and heritability are informative about the DFE.

Simons et al. (2018) developed a model for the relationship between effect sizes and selection coefficients based on isotropic stabilizing selection and Fisher's geometric model (the specifics of the model is discussed in a subsequent section). The number of trait dimensions in this model corresponds to the effective number of independent axes of genetic variation, a value that can be interpreted as the degree of pleiotropy. With a single dimension the selection coefficient is the same as in the classical model of one dimensional stabilizing selection: $|\beta| = \sqrt{2\,sV_S}$. When the number of traits becomes large the relationship becomes $\beta \sim N(0, (V_S/n_e)s)$, where n_e is the effective number of traits, and expressions for moderate pleiotropy interpolate between these extremes. Rather than fit this model to the heritability explained by different minor allele frequencies, Simons et al. (2018) analyzed the distribution of variance contributions, $v = 2\beta^2x(1 - x)$, among genome-wide significant SNPs. For a given mean among discovered loci, the variance of v is higher with greater pleiotropy (n_e), with a parametric likelihood derived by the authors. The high-pleiotropy model was found to fit the distribution of GWAS hits for standing height and BMI better than the no- and low-pleiotropy alternatives.

Zeng et al. (2021) also simulated varying degrees of pleiotropy using the Simons et al. (2018) model and found that α estimates were insensitive to changes in the degree of pleiotropy (n_e). This
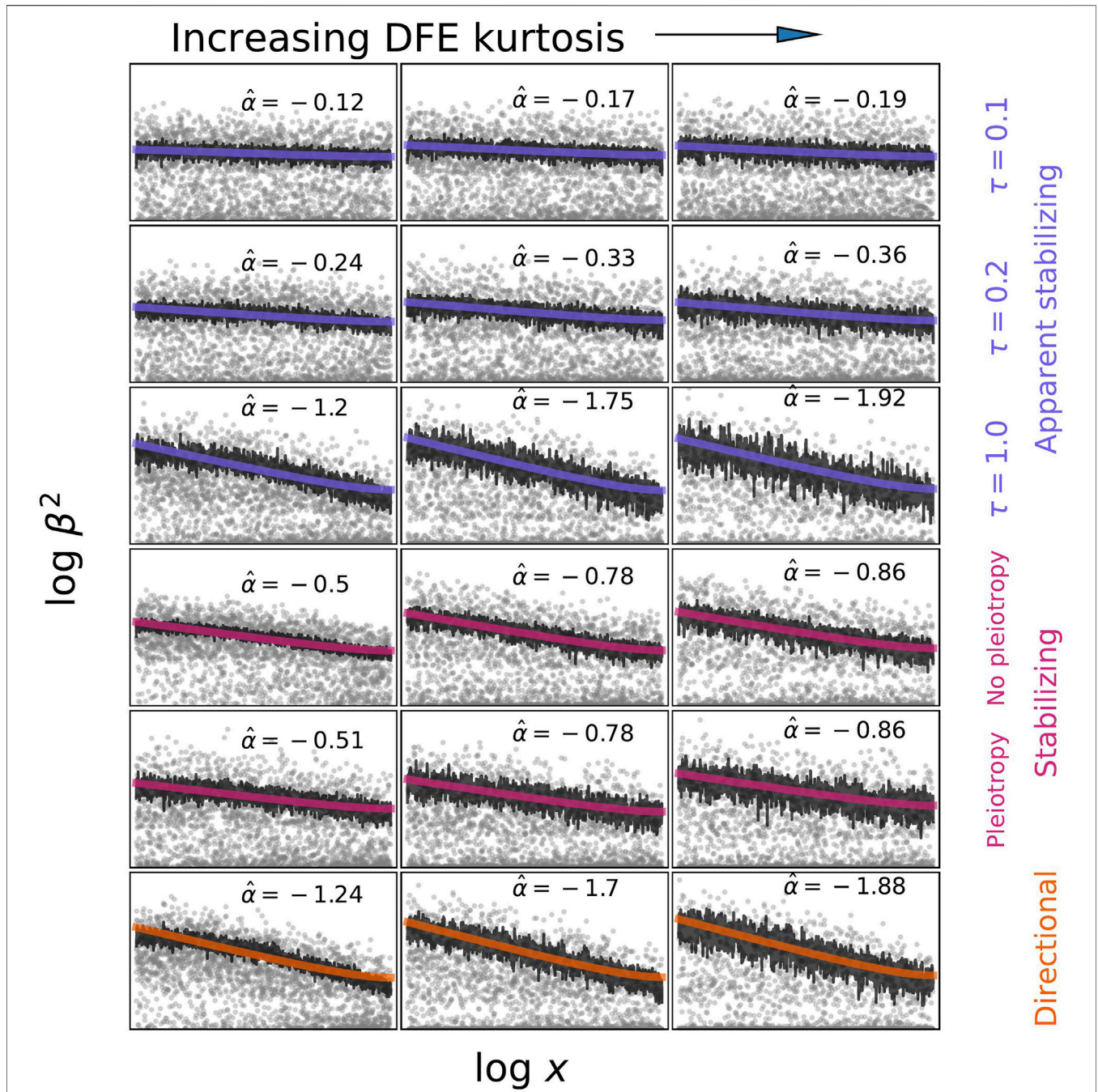
FIGURE 2 | Examples of what alpha models may infer under different models of selection and different distributions of fitness effects (DFE). Effect sizes were simulated by sampling from $p(s|x)$ and then from $p(\beta|s)$ under the different models described in the text. Derived allele frequencies are uniform between 0.01 and 0.5. Estimates of α were obtained by fitting log $\beta^2 = \alpha$ log $x(1-x) + c$ to the average $\beta^2|x$ values calculated from simulations. The DFE was varied by decreasing the shape parameter from 1 to 0.25 to 0.125 while keeping the mean constant. It is important to recognize that $\hat{\alpha}$ values reported here would not necessarily correspond to those obtained by real statistical genetics methods (Zeng et al., 2018; Schoech et al., 2019; Speed et al., 2020; Zeng et al., 2021). Those methods employ particular likelihoods and are applied to real genetic data where frequencies are not uniform and effect sizes are estimated with error. The frequencies of analyzed variants may be particularly important since the slope of the (β^2, x) relationship (local α) varies with frequency (Schoech et al., 2019). Estimated α values increase with increasing DFE kurtosis, reflecting the proportion of variants that are strongly selected. For high kurtosis, estimates approach the theoretical expectation of $\alpha = -2\tau$ for the Eyre-Walker (2010) model as derived by Schoech et al. (2019). As expected, in a model of stabilizing selection (Simons et al., 2018), the degree of pleiotropy does not affect the α estimate. Directional selection is associated with higher α estimates than stabilizing selection.

makes sense, given that the alpha model only attempts to fit the average effect size - frequency relationship and suggests that new approaches will be needed to investigate the nature of pleiotropy and the relative importance of the focal trait.

4 MODEL BUILDING USING GEOMETRY AND PLEIOTROPY

Using the distribution of causative allele frequencies and their effects solely on the focal trait, what could be done to further interpret the results of GWAS studies? One advance would be to explicitly include selection in the next generation of models that build upon LDSC or α models (Sella and Barton, 2019). We may start by imagining what class of models could fit the joint distribution of (x, β). Assume a set of parameters Θ that describes the selection model. An analysis would use the likelihood $p(x, \beta|\Theta)$, which decomposes into $p(x, \beta|\Theta) = p(x|\beta, \Theta)p(\beta|\Theta)$. Since the distribution of effect sizes is not a major concern for selection, inference would focus on $p(x|\beta, \Theta)$. The distribution of frequencies for a given effect size is determined by integrating over the possible fitness effects of a mutation with effect size β: $p(x|\beta, \Theta) = \int p(x|s, \Theta)p(s|\beta, \Theta)ds$. The effect of selection, s, could be either additive or underdominant (stabilizing selection), but could also represent other models beyond these two such as overdominant or fluctuating selection. x can be replaced by the age or historical frequency path of the allele (Stern et al., 2021). $p(x|s)$ can be tackled with standard population genetics, so the trickier problem is to provide $p(s|\beta, \Theta)$ in cases of pleiotropic selection.

In an early attempt to do this explicitly, Keightley and Hill (1990) proposed $p(s|\beta, \Theta)$ as the conditional distribution of a two-dimensional Wishart distribution. In this formulation both the mean and variance of s, conditional on β, are proportional to $|\beta|$ plus a constant, and a correlation parameter determines how the variance scales with the mean. This contrasts with the model of direct stabilizing selection where s is proportional to β^2. Another approach decomposes this distribution as $p(s|\beta, \Theta) \propto p(\beta|s, \Theta) p(s|\Theta)$. This has the appealing property of separating the link between fitness and trait effects from the distribution of fitness effects (DFE). Eyre-Walker (2010) proposed a form for $p(\beta|s, \Theta)$ where $E[\beta] \propto s^\tau$ with multiplicative noise. Both of these models try to capture a space of potential relationships between effect size and selection without being over-parameterized. However, it is not actually clear how one should interpret results from either. A weak correlation parameter from the Keightley and Hill (1990) model would perhaps indicate the importance of pleiotropy, but the linear scaling between $|\beta|$ and s would not make sense with direct stabilizing selection. Eyre-Walker's τ doesn't necessarily mean stronger or weaker selection. Would it mean anything for the relative importance of directional or stabilizing selection? Moreover, these two models make divergent predictions for the contribution of rare versus common alleles to the genetic variance (Eyre-Walker, 2010; Caballero et al., 2015; Sella and Barton, 2019).

Simons et al. (2018) made a strong argument for interpretability when deriving their distribution for $p(\beta|s, \Theta)$. The framework they used was multivariate stabilizing selection in a geometric model (Fisher, 1930). Models within this framework generally posit a multidimensional phenotypic space with a selection function that describes the fitness of each possible phenotypic combination. Typically, the fitness function is Gaussian and centered at some optimum phenotype. A mutation is a vector that moves an individual to a different point in phenotype space, thereby altering fitness. Assuming a population centered at its optimum value, with each phenotypic direction under equal stabilizing selection and mutational pressure, $p(\beta|s, \Theta)$ takes a simple parametric form depending only on the number of dimensions n_e and the strength of selection V_s.

Previous work using Fisher's geometric model had used it to derive the DFE of new mutations (Martin and Lenormand, 2006; Lourenço et al., 2011) or the expected genetic variance and correlation of the focal trait with fitness (Zhang and Hill, 2003) rather than $p(\beta|s)$. A major assumption of these studies was that the phenotypic effects of new mutations were drawn from a multivariate normal distribution with different dimensions representing different phenotypes. While a seemingly natural starting place, the assumption of normally distributed mutations is far from realistic and mathematically troublesome. There is accumulating evidence that the mutational effect distribution is substantially leptokurtic for many traits (Zhang et al., 2018; O'Connor et al., 2019; O'Connor, 2021). It is has also been shown that, for a single normal distribution of mutations, the DFE concentrates around a point value of s as the number of traits becomes large, an obviously unrealistic scenario (Waxman and Peck, 1998; Wingreen et al., 2003; Zhang and Hill, 2003). Thankfully, one may still rescue the utility of geometric models by using a mixture of normals.

For example, the Simons et al. (2018) likelihood can be derived from the geometric model with normal mutation proposed by Martin and Lenormand (2006) by integrating out a variance parameter: $p(s|\beta) = \int \frac{p(\beta|s,\sigma^2)p(s|\sigma^2)}{p(\beta|\sigma^2)} p(\sigma^2)d\sigma^2$. If mutations are uncorrelated, equally affected by stabilizing selection, and drawn from a mixture of normal distributions, then the distribution of variances, $p(\sigma^2)$, fully describes the mutational distribution. It is straightforward using Bayes' theorem to show that $p(s|\beta) \propto p(\beta|s)\int p(s|\sigma^2)p(\sigma^2)d\sigma^2$ and contains two components. The first component, $p(\beta|s)$, has the form derived by Simons et al. (2018), a normal distribution with variance proportional to s when the number of traits is large. This part is independent of σ^2. The second component, $\int p(s|\sigma^2)p(\sigma^2)d\sigma^2$, is the DFE itself. In this example, the DFE is generated by the distribution of mutational effects. The variance of what normal distribution in the mixture a mutation comes from determines how strongly selected it is.

The above approach suggests that a fruitful way to propose future models would be to propose that there exist different mutational modes. Modes might represent different biological pathways and could be parameterized by which traits are involved, the correlation of mutational effects among these, and the distribution of mutational effect sizes. If summarized in Θ_M, we might then integrate over the distribution of modes. If β is conditionally independent of s given Θ_M, then the form of the

DFE will be separable from the link between selection and effect sizes, though it is not always clear that this will be the case. Directional selection as well as antagonistic pleiotropy may be possible to model this way, at least for a population at equilibrium in its fitness landscape. To more directly analyze selection and the pleiotropic relationships among traits, a vector of effect sizes could replace the effect β on a single, focal trait.

5 EMPIRICAL DEMONSTRATIONS OF THE EXISTENCE AND NATURE OF PLEIOTROPY

Since models of the evolution and maintenance of complex trait variation strongly depend on assumptions regarding the degree of pleiotropy. Modeling and measurement of pleiotropy is key to the empirical questions of whether the focal trait is under meaningful direct selection and how selection coefficients depend on the phenotypic effects of individual variants.

Current estimates of polygenicity indirectly but strongly suggest highly pleiotropic genetic architecture for most complex traits (Zeng et al., 2018; O'Connor et al., 2019; Zeng et al., 2021). Indeed, it was estimated that 2% of genetic variation is involved in height and a similar proportion (1%) is involved in risk of Type 2 Diabetes (Zeng et al., 2021). It is clear that a model where every quantitative trait locus (QTL) affects just a single trait is, due to the finite nature of the human genome, inconsistent with high polygenicity (defined here as the probability that a variant has a non-zero effect on the focal trait). We do not know exactly what fraction of the genome plays any functional role; comparative and functional genomics produce a range of estimates generally on the order of 0.1 (Rands et al., 2014; Gulko et al., 2015). If 10% of the genome is of any functional importance, and trait-affecting mutations originate from this functional fraction, it clearly cannot harbor independent QTLs for a vast number of complex traits each with a polygenicity of 2% (Jordan et al., 2019).

With the abundance of GWAS data, many aspects of pleiotropy can be empirically estimated using corresponding well-developed statistical approaches. The specific relationships between causative QTL effect sizes on different traits that these approaches investigate are illustrated in **Figure 3**. The relationship between two phenotypes is most commonly expressed as global genetic covariance, which reflects the overall degree of pleiotropy in the form of correlation of QTL effects across all loci ($\text{Cov}[\beta_1, \beta_2]$, where β_1 and β_2 are allelic effect sizes for phenotypes 1 and 2), scaled by the heterozygosity at each causative locus. A significant genetic covariance below one would indicate that individual QTL effects are correlated but not identical. Global genetic covariance is estimated using statistical approaches related to those used to estimate heritability including random effect models implemented into the GCTA software or LD-score regression (Lee et al., 2012; Bulik-Sullivan et al., 2015). Estimates of genetic covariance come with the same caveats as must apply to heritability estimates (Visscher et al., 2008). Measurements apply to the particular environment in which the different traits are measured and offer no guarantee of a fundamental relationship between traits. Under different conditions, gene-by-environment interactions can change which genetic variants contribute to heritability and different pleiotropic traits may associate with the new regime.

Using these and related statistical techniques, highly significant genetic covariances were estimated among various autoimmune diseases and among various psychiatric diseases and related phenotypes (Cotsapas et al., 2011; Lee et al., 2013; Watanabe et al., 2019; Lincoln et al., 2021). The analysis of genetic covariances between two traits has some limitations. The genetic covariance alone is not informative about biological mechanisms and per locus patterns. For example, the same genetic covariance may indicate either pleiotropy limited to just a few loci with very similar effects on both traits on the background of other non-pleiotropic loci or the broad pleiotropy of all loci but with non-identical effects (**Figures 3A** vs **3B**). For autoimmune traits, it is possible that some loci impact immune function while others determine tissue or organ specificity.

The question of contribution of individual loci into global genetic correlation must be, therefore, addressed at the local level by studying individual loci. When studying individual loci, one of the challenges is that linkage disequilibrium confounds the analysis. Genetic covariance may imply real pleiotropy, meaning that the same genetic variants causally affect both traits. Alternatively, some variants may exclusively impact the first trait and other variants exclusively impact the second trait, but local genetic correlation can still be induced by linkage disequilibrium between the two sets. Consequently, the field has developed two different classes of methods to address this issue. Methods that estimate local genetic covariance (Shi et al., 2017) do not distinguish between functional pleiotropy versus non-independence induced by linkage disequilibrium. A different class of methods called "colocalization" (Giambartolomei et al., 2014; Hormozdiari et al., 2016; Chun et al., 2017) relies on linkage disequilibrium patterns to specifically test the hypothesis that the same causative variant (or variants) in the locus impacts both traits (**Figure 3B**). Multiple examples of local genetic correlations and individually colocalized loci have been described (van Rheenen et al., 2019; Aguet et al., 2020; Vuckovic et al., 2020). However, some QTLs involved in genetically correlated traits do not show obvious signals of colocalization, suggesting that genetic correlation does not necessarily imply pleiotropic effects of all variants (Lincoln et al., 2021).

A separate aspect of pleiotropy that statistical genetics addresses is the causal relationship between phenotypes (van Rheenen et al., 2019). There is an important distinction between "horizontal" pleiotropy with genetic variants exerting independent effects on both traits and a causal path or "vertical" pleiotropy, where one trait directly contributes to the other (Jordan et al., 2019). Examples of the latter include LDL cholesterol being a causative risk factor of heart disease (Zhu et al., 2018), the genetic component of smoking being a causative risk of lung cancer (McKay et al., 2017), and all molecular effects (considered as "molecular" phenotypes) leading to changes in a phenotype of the organism. If one trait is a cause of the other trait, every variant inducing an effect on the first trait also affects the second trait (**Figure 3C**). Moreover, these effect sizes are
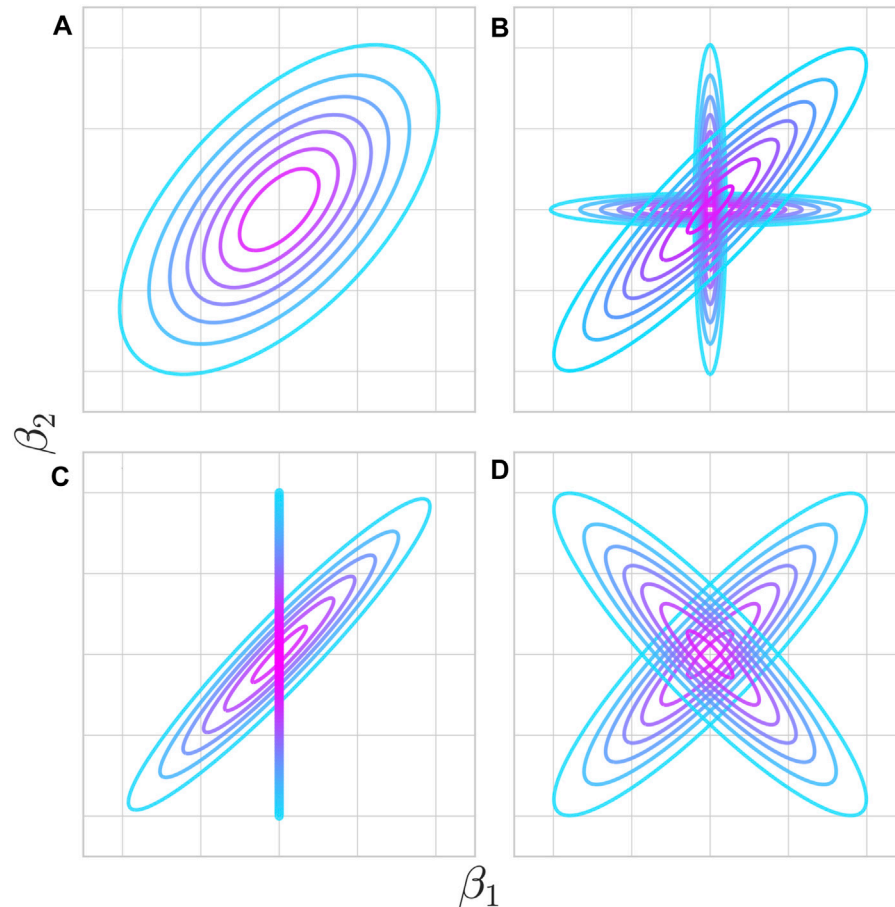
FIGURE 3 | Various potential pleiotropic relationships at individual loci underlie genetic correlations between traits. **(A)** Mutations affecting trait 1 have a tendency to impact trait 2 in a particular direction, although a variety of outcomes are possible through the functional particulars of that change. **(B)** Mutations fall either into a shared or unshared functional pathway between the two traits. Colocalization analysis aims to test which distribution a given QTL comes from. Even though not every mutation is pleiotropic, the two traits are genetically correlated. The proportion of mutations falling into either pathway determines the strength of genetic correlation. **(C)** Trait 1 has a causal impact on trait 2 such that every mutation with a non-zero effect on trait one has a strongly correlated effect on trait 2, but not vice-versa. Mendelian randomization aims to test for the existence and direction of this effect. This also manifests as a genetic correlation at the phenotypic level. **(D)** Individual variants may be pleiotropic, but can result in low or zero genetic correlation if different pathways have opposing effects.

proportional and correspond to the causal effect of the first trait on the second trait. Because the first trait is usually just one of many causes, most variants affecting the second trait would not be expected to have any effect on the first trait. These considerations are a foundation of Mendelian Randomization methods that attempt to infer causal relationships even if genetic associations for the two phenotypes are measured separately in independent datasets (Pingault et al., 2018). This approach relies on a large number of QTLs and does not translate to individual loci.

Many recent studies of pleiotropy, colocalization and causality have focused on molecular phenotypes such as gene expression, chromatin accessibility or DNA methylation (Umans et al., 2020; Vuckovic et al., 2020; Ye et al., 2020; Morabito et al., 2021). Numerous QTLs for various molecular phenotypes have been identified for these classes of traits (most prominently expression QTLs or eQTLs). The main motivation of these studies is to identify the primary molecular events underlying genetic

associations with human traits and diseases. However, it is not guaranteed that genetic covariance or colocalization of a molecular trait with a focal trait is indicative of an underlying causal impact of variation in the molecular trait on the focal trait. One example is that changes in BMI actually induce changes in DNA methylation rather than DNA methylation acting as a molecular mechanism mediating genetic effects on BMI (Wahl et al., 2017). The direction of causality was demonstrated by showing that SNPs which predict methylation levels at individual loci did not predict BMI levels, while a genetic risk score for BMI levels did predict methylation levels.

Even in the absence of genetic covariance, molecular effects may induce pleiotropic relationships between two traits. Imagine a scenario where the two traits are both mediated by a large number of molecular phenotypes (activities of many individual genes or other latent factors), but these molecular phenotypes do not exhibit correlated effects on the two traits (**Figure 3D**). In this case, genetic covariance might not exist or be very weak on

aggregate but covariance between absolute (or squared) genetic effects $\text{Cov}[\beta_1^2, \beta_2^2]$ may be substantial. The popular "omnigenic" model offers one version of such a scenario (Boyle et al., 2017). Genetic covariance may also be close to zero if the pleiotropic effect is limited to one or a small number of loci (in other words, with a substantial local genetic covariance and even colocalization in individual loci) (Liu et al., 2019).

These methodological developments and empirical results related to pleiotropy are important in light of the main subject of this review. They motivate consideration of evolutionary models that take into account groups of correlated traits. For causally related traits, selection effects would probably differ depending on whether selection primarily acts on the upstream or downstream trait along the causal chain. An interesting perspective is also brought by the consideration of molecular phenotype. If each molecular phenotype is pleiotropically involved with many downstream organismal phenotypes, and the focal trait is merely one of these, selection coefficients can depend on effect sizes even if the focal trait is neutral. Variants with larger effect sizes on molecular function would be under stronger selection because this molecular function impacts multiple other selected downstream traits in addition to the neutral focal trait.

Few studies have analyzed the effects of pleiotropy on selection by actually incorporating the measured effects of variants on multiple traits. Some mutation accumulation studies have tried to demonstrate whether pleiotropic mutations are under stronger selection. McGuigan et al. (2014) provide some evidence that mutations underlying combinations of correlated gene expression traits in Drosophila serrata are under stronger selection than the average mutation affecting a given trait. In humans, Sanjak et al. (2018) regressed lifetime reproductive success on genetic scores for multiple traits simultaneously in United Kingdom Biobank participants. Compared to the univariate, this full analysis lacked power, but quadratic and linear selection terms did change in both magnitude and sign for some traits, indicating the importance of accounting for pleiotropy. Stern et al. (2021) took a similar approach, but used the shape of genealogies at GWAS loci to look at historical rather than contemporary patterns of directional selection. Again, the authors found that many estimates of selection changed substantially, and were largely attenuated, when accounting for the correlated response in other traits. At the time of writing, no attempt has been made to account for pleiotropy in the alpha model approaches discussed above that have demonstrated negative selection on many human traits.

6 CONSTANCY OF EFFECT SIZES ACROSS TIME AND SPACE

Everything discussed so far has assumed that genetic variants have well-defined additive effects on traits of interest, and that these effects are measurable in contemporary human populations. Although convenient, and the correct starting place for most analyses, recent research has demonstrated that causative variants for many traits and diseases have population-specific

effect sizes. Such studies are possible when GWAS for the same traits have been performed in different populations (De Candia et al., 2013; Mancuso et al., 2016). One approach has been to estimate the cross-population genetic correlation, the correlation in causal effect sizes between the different samples, and these estimates are often less than one (Brown et al., 2016; Galinsky et al., 2019). This is most likely due to gene-by-environment (GxE) and gene-by-gene (GxG) interactions, with some effect driven by different measurement practices and diagnosis criteria.

Shi et al. (2021) estimated the impact of different functional annotations on the degree of cross-population effect size correlation of variants within those genomic regions. They found that the squared genetic correlation was depleted most strongly in regions under strong background selection as well as in and around functional elements such as exons, promoters, and enhancers. These regions are also enriched for heritability and, as previous research reviewed here has shown, variants residing there are likely under stronger selection. If a variant has different effect sizes in different contemporary populations, we should be more uncertain about its effect size in the ancestral population where the majority of its existence may have taken place. Cross-population genetic correlation could therefore be used as a measure of the temporal stability of allelic effects. Alternatively, the aggregate pleiotropic effects of an allele may stay roughly constant even as the effects on individual traits vary due to GxE or other factors.

7 CONCLUSION

Direct data on genotype-phenotype associations for numerous human traits have provided an opportunity to investigate which, if any, of the current theoretical models for the maintenance of complex trait variation fit observed genetic architectures. Depending on the degree and nature of pleiotropy, as well as the importance of the focal trait for selection, these models predict the relationship between β and s (Johnson and Barton, 2005). Selection analyses of human GWAS data have consistently demonstrated a negative relationship between effect size magnitudes and allele frequencies, implying that larger effect sizes are associated with stronger selection on average (Zeng et al., 2018; Schoech et al., 2019; Speed et al., 2020; Zeng et al., 2021). Models where the focal trait is neutral, or largely biologically unrelated to any aspect of fitness, are therefore ruled out. Within the class of alpha models, the scaling between β^2 and frequency varies across traits, likely reflecting differences in the DFE and the scaling between β and s. These estimates are difficult to interpret in terms of classical stabilizing selection models, and work is needed to reconcile tractable statistical models of how effect sizes change with frequency with realistic models of selection at the phenotypic level. Studies have also largely been limited to analyzing the average relationship of effect size to frequency. This limits the ability to capture the importance of pleiotropy which should create variance around that average. By directly modeling the variation in genome-wide significant variance contributions, Simons et al. (2018) were able to infer a high degree of pleiotropy for height and BMI.

All the approaches reviewed here infer the nature of selection on GWAS loci by analyzing the distribution of allele frequencies and effect sizes (x, β), with the overall trait heritability sometimes included. Future work along these lines may utilize fine-mapping (Weissbrod et al., 2020) or other techniques to better capture this distribution (O'Connor, 2021). An interesting approach was developed by O'Connor et al. (2019) who estimated the kurtosis of heritability contributions using the LDSC framework to measure trait polygenicity at different allele frequencies and functional genomic annotations. The kurtosis depends on the fourth moment of the distribution of effect sizes, and therefore contains additional information beyond that contained in the alpha model. A low kurtosis, and therefore high polygenicity, of common variants indicated a "flattening" due to selection strongly preventing any large-effect variants from reaching high frequencies. This indicates a high importance of the focal trait for selection, but more thought is needed to tell what degree of pleiotropy is consistent with these results.

The greatest advances in our ability to make sense of the maintenance of complex trait variation will likely come from analyses that utilize variant-level pleiotropy and account for effect sizes that vary across time and space. Methods to investigate pleiotropy in statistical genetics are already well-developed (van Rheenen et al., 2019) but have yet to intersect with analyses of stabilizing or negative selection. Effect size differences between populations are also well-documented (Brown et al., 2016; Shi et al., 2021), but have received less attention, likely in part due to the lack of large GWAS from diverse populations and the difficulty of standardizing measurement for some phenotypes. The portability of polygenic scores is also potentially more strongly impacted by differences in allele frequencies and linkage disequilibrium than effect size variation (Wang et al., 2020), and allele frequencies will differentiate more rapidly under stabilizing or negative selection (Yair and Coop, 2021). However, understanding effect size variation in space and time may ultimately end up being more important for modeling the maintenance of variation in complex traits as well as detecting selection on them (Mathieson, 2021).

AUTHOR CONTRIBUTIONS

EK and SS jointly wrote this review.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Agarwala, V., Flannick, J., Flannick, J., Sunyaev, S., and Altshuler, D. (2013). Evaluating Empirical Bounds on Complex Disease Genetic Architecture. *Nat. Genet.* 45, 1418–1427. doi:10.1038/ng.2804

Aguet, F., Barbeira, A. N., Bonazzola, R., Jo, B., Kasela, S., Liang, Y., et al. (2020). The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues. *Science* 369, 1318–1330. doi:10.1126/science.aaz1776

Barton, N. H. (1990). Pleiotropic Models of Quantitative Variation. *Genetics* 124, 773–782. doi:10.1093/genetics/124.3.773

Boyle, E. A., Li, Y. I., and Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169, 1177–1186. doi:10.1016/j.cell.2017.05.038

Brown, B. C., Ye, C. J., Price, A. L., and Zaitlen, N. (2016). Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am. J. Hum. Genet.* 99, 76–88. doi:10.1016/j.ajhg.2016.05.001

Bulik-Sullivan, B., Finucane, H. K., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., et al. (2015). An Atlas of Genetic Correlations across Human Diseases and Traits. *Nat. Genet.* 47, 1236–1241. doi:10.1038/ng.3406

Bulmer, M. G. (1972). The Genetic Variability of Polygenic Characters under Optimizing Selection, Mutation and Drift. *Genet. Res.* 19, 17–25. doi:10.1017/s0016672300014221

Caballero, A., Tenesa, A., and Keightley, P. D. (2015). The Nature of Genetic Variation for Complex Traits Revealed by GWAS and Regional Heritability Mapping Analyses. *Genetics* 201, 1601–1613. doi:10.1534/genetics.115.177220

Charlesworth, B. (2001). Patterns of Age-specific Means and Genetic Variances of Mortality Rates Predicted by the Mutation-Accumulation Theory of Ageing. *J. Theor. Biol.* 210, 47–65. doi:10.1006/jtbi.2001.2296

Chun, S., Casparino, A., Patsopoulos, N. A., Croteau-Chonka, D. C., Raby, B. A., De Jager, P. L., et al. (2017). Limited Statistical Evidence for Shared Genetic Effects

of eQTLs and Autoimmune-Disease-Associated Loci in Three Major Immune-Cell Types. *Nat. Genet.* 49, 600–605. doi:10.1038/ng.3795

Corbett, S., Courtiol, A., Lummaa, V., Moorad, J., and Stearns, S. (2018). The Transition to Modernity and Chronic Disease: Mismatch and Natural Selection. *Nat. Rev. Genet.* 19, 419–430. doi:10.1038/s41576-018-0012-3

Cotsapas, C., Voight, B. F., Rossin, E., Lage, K., Neale, B. M., Wallace, C., et al. (2011). Pervasive Sharing of Genetic Effects in Autoimmune Disease. *Plos Genet.* 7, e1002254. doi:10.1371/journal.pgen.1002254

de Candia, T. R., Lee, S. H., Yang, J., Browning, B. L., Gejman, P. V., Levinson, D. F., et al. (2013). Additive Genetic Variation in Schizophrenia Risk Is Shared by Populations of African and European Descent. *Am. J. Hum. Genet.* 93, 463–470. doi:10.1016/j.ajhg.2013.07.007

Eyre-Walker, A. (2010). Genetic Architecture of a Complex Trait and its Implications for Fitness and Genome-wide Association Studies. *Proc. Natl. Acad. Sci.* 107, 1752–1756. doi:10.1073/pnas.0906182107

Finucane, H. K., Bulik-Sullivan, B., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., et al. (2015). Partitioning Heritability by Functional Annotation Using Genome-wide Association Summary Statistics. *Nat. Genet.* 47, 1228–1235. doi:10.1038/ng.3404

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* (UK: Clarendon.

Fisher, R. (1918). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. Roy. Soc. Edinb.* 52, 399–433.

Galinsky, K. J., Reshef, Y. A., Finucane, H. K., Loh, P.-R., Zaitlen, N., Patterson, N. J., et al. (2019). Estimating Cross-Population Genetic Correlations of Causal Effect Sizes. *Genet. Epidemiol.* 43, 180–188. doi:10.1002/gepi.22173

Gazal, S., Finucane, H. K., Furlotte, N. A., Loh, P.-R., Palamara, P. F., Liu, X., et al. (2017). Linkage Disequilibrium-dependent Architecture of Human Complex Traits Shows Action of Negative Selection. *Nat. Genet.* 49, 1421–1427. doi:10.1038/ng.3954

Gazal, S., Loh, P.-R., Finucane, H. K., Ganna, A., Schoech, A., Sunyaev, S., et al. (2018). Functional Architecture of Low-Frequency Variants Highlights

Strength of Negative Selection across Coding and Non-coding Annotations. *Nat. Genet.* 50, 1600–1607. doi:10.1038/s41588-018-0231-8

Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C., et al. (2014). Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *Plos Genet.* 10, e1004383. doi:10.1371/journal.pgen.1004383

Gillespie, J. H. (1984). Pleiotropic Overdominance and the Maintenance of Genetic Variation in Polygenic Characters. *Genetics* 107, 321–330. doi:10.1093/genetics/107.2.321

Gulko, B., Hubisz, M. J., Gronau, I., and Siepel, A. (2015). A Method for Calculating Probabilities of Fitness Consequences for point Mutations across the Human Genome. *Nat. Genet.* 47, 276–283. doi:10.1038/ng.3196

Harpak, A., and Przeworski, M. (2021). The Evolution of Group Differences in Changing Environments. *Plos Biol.* 19, e3001072–14. doi:10.1371/journal.pbio.3001072

Hormozdiari, F., van de Bunt, M., Segrè, A. V., Li, X., Joo, J. W. J., Bilow, M., et al. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* 99, 1245–1260. doi:10.1016/j.ajhg.2016.10.003

Johnson, T., and Barton, N. (2005). Theoretical Models of Selection and Mutation on Quantitative Traits. *Phil. Trans. R. Soc. B* 360, 1411–1425. doi:10.1098/rstb.2005.1667

Jordan, D. M., Verbanck, M., and Do, R. (2019). HOPS: a Quantitative Score Reveals Pervasive Horizontal Pleiotropy in Human Genetic Variation Is Driven by Extreme Polygenicity of Human Traits and Diseases. *Genome Biol.* 20, 222. doi:10.1186/s13059-019-1844-7

Keightley, P. D., and Hill, W. G. (1988). Quantitative Genetic Variability Maintained by Mutation-Stabilizing Selection Balance in Finite Populations. *Genet. Res.* 52, 33–43. doi:10.1017/S0016672300027282

Keightley, P. D., and Hill, W. G. (1990). Variation Maintained in Quantitative Traits with Mutation-Selection Balance: Pleiotropic Side-Effects on Fitness Traits. *Proc. R. Soc. B: Biol. Sci.* 242, 95–100. doi:10.1098/rspb.1990.0110

Kiezun, A., Pulit, S. L., Francioli, L. C., van Dijk, F., Swertz, M., Boomsma, D. I., et al. (2013). Deleterious Alleles in the Human Genome Are on Average Younger Than Neutral Alleles of the Same Frequency. *Plos Genet.* 9, e1003301–12. doi:10.1371/journal.pgen.1003301

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., et al. (2001). The Strength of Phenotypic Selection in Natural Populations. *The Am. Naturalist* 157, 245–261. doi:10.1086/319193

Koch, E. M. (2019). The Effects of Demography and Genetics on the Neutral Distribution of Quantitative Traits. *Genetics* 211, 1371–1394. doi:10.1534/genetics.118.301839

Kondrashov, A. S., and Turelli, M. (1992). Deleterious Mutations, Apparent Stabilizing Selection and the Maintenance of Quantitative Variation. *Genetics* 132, 603–618. doi:10.1093/genetics/132.2.603

Lande, R. (1976b). The Maintenance of Genetic Variability by Mutation in a Polygenic Character with Linked Loci. *Genet. Res.* 26, 221–235. doi:10.1017/s0016672300016037

Lande, R., and Arnold, S. (1983). The Measurement of Selection on Correlated Characters. *Evolution* 37, 1210. doi:10.2307/2408842

Lande, R. (1976a). Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution* 30, 314. doi:10.2307/2407703

Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., Perlis, R. H., et al. (2013). Genetic Relationship between Five Psychiatric Disorders Estimated from Genome-wide SNPs. *Nat. Genet.* 45, 984–994. doi:10.1038/ng.2711

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., and Wray, N. R. (2012). Estimation of Pleiotropy between Complex Diseases Using Single-Nucleotide Polymorphism-Derived Genomic Relationships and Restricted Maximum Likelihood. *Bioinformatics* 28, 2540–2542. doi:10.1093/bioinformatics/bts474

Lincoln, M. R., Connally, N., Axisa, P.-P., Gasperi, C., Mitrovic, M., van Heel, D., et al. (2021). Joint Analysis Reveals Shared Autoimmune Disease Associations and Identifies Common Mechanisms. *medRxiv*, 1–15. doi:10.1101/2021.05.13.21257044

Liu, X., Li, Y. I., and Pritchard, J. K. (2019). Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* 177, 1022–1034.e6. doi:10.1016/j.cell.2019.04.014

Lourenço, J., Galtier, N., and Glémin, S. (2011). Complexity, Pleiotropy, and the Fitness Effect of Mutations. *Evolution* 65, 1559–1571. doi:10.1111/j.1558-5646.2011.01237.x

Lynch, M., and Hill, W. G. (1986). Phenotypic Evolution by Neutral Mutation. *Evolution* 40, 915–935. doi:10.1111/j.1558-5646.1986.tb00561.x

Mancuso, N., Rohland, N., Rohland, N., Rand, K. A., Tandon, A., Allen, A., et al. (2016). The Contribution of Rare Variation to Prostate Cancer Heritability. *Nat. Genet.* 48, 30–35. doi:10.1038/ng.3446

Martin, G., and Lenormand, T. (2006). A General Multivariate Extension of Fisher's Geometrical Model and the Distribution of Mutation Fitness Effects across Species. *Evolution* 60, 893–907. doi:10.1111/j.0014-3820.2006.tb01169.x

Maruyama, T. (1974). The Age of a Rare Mutant Gene in a Large Population. *Am. J. Hum. Genet.* 26, 669–673.

Mathieson, I. (2021). The Omnigenic Model and Polygenic Prediction of Complex Traits. *Am. J. Hum. Genet.* 108, 1558–1563. doi:10.1016/j.ajhg.2021.07.003

McGuigan, K., Collet, J. M., Allen, S. L., Chenoweth, S. F., and Blows, M. W. (2014). Pleiotropic Mutations Are Subject to strong Stabilizing Selection. *Genetics* 197, 1051–1062. doi:10.1534/genetics.114.165720

McKay, J. D., Hung, R. J., Han, Y., Zong, X., Carreras-Torres, R., Christiani, D. C., et al. (2017). Large-scale Association Analysis Identifies New Lung Cancer Susceptibility Loci and Heterogeneity in Genetic Susceptibility across Histological Subtypes. *Nat. Genet.* 49, 1126–1132. doi:10.1038/ng.3892

Morabito, S., Miyoshi, E., Michael, N., Shahin, S., Martini, A. C., Head, E., et al. (2021). Single-nucleus Chromatin Accessibility and Transcriptomic Characterization of Alzheimer's Disease. *Nat. Genet.* 53, 1143–1155. doi:10.1038/s41588-021-00894-z

O'Connor, L. J., Schoech, A. P., Hormozdiari, F., Gazal, S., Patterson, N., and Price, A. L. (2019). Extreme Polygenicity of Complex Traits Is Explained by Negative Selection. *Am. J. Hum. Genet.* 105, 456–476. doi:10.1016/j.ajhg.2019.07.003

O'Connor, L. J. (2021). The Distribution of Common-Variant Effect Sizes. *Nat. Genet.* 53, 1243–1249. doi:10.1038/s41588-021-00901-3

Pavard, S., and Coste, C. F. D. (2021). Evolutionary Demographic Models Reveal the Strength of Purifying Selection on Susceptibility Alleles to Late-Onset Diseases. *Nat. Ecol. Evol.* 5, 392–400. doi:10.1038/s41559-020-01355-2

Pingault, J.-B., O'Reilly, P. F., Schoeler, T., Ploubidis, G. B., Rijsdijk, F., and Dudbridge, F. (2018). Using Genetic Data to Strengthen Causal Inference in Observational Research. *Nat. Rev. Genet.* 19, 566–580. doi:10.1038/s41576-018-0020-3

Pritchard, J. K. (2001). Are Rare Variants Responsible for Susceptibility to Complex Diseases? *Am. J. Hum. Genet.* 69, 124–137. doi:10.1086/321272

Rands, C. M., Meader, S., Ponting, C. P., and Lunter, G. (2014). 8.2% of the Human Genome Is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage. *Plos Genet.* 10, e1004525. doi:10.1371/journal.pgen.1004525

Reich, D. E., and Lander, E. S. (2001). On the Allelic Spectrum of Human Disease. *Trends Genet.* 17, 502–510. doi:10.1016/S0168-9525(01)02410-6

Robertson, A. (1956). The Effect of Selection against Extreme Deviants Based on Deviation or on Homozygosis. *J. Genet.* 54, 236–248. doi:10.1007/bf02982779

Sanjak, J. S., Sidorenko, J., Robinson, M. R., Thornton, K. R., and Visscher, P. M. (2018). Evidence of Directional and Stabilizing Selection in Contemporary Humans. *Proc. Natl. Acad. Sci. USA* 115, 151–156. doi:10.1073/pnas.1707227114

Schoech, A. P., Jordan, D. M., Loh, P.-R., Gazal, S., O'Connor, L. J., Balick, D. J., et al. (2019). Quantification of Frequency-dependent Genetic Architectures in 25 UK Biobank Traits Reveals Action of Negative Selection. *Nat. Commun.* 10, 790. doi:10.1038/s41467-019-08424-6

Sella, G., and Barton, N. H. (2019). Thinking about the Evolution of Complex Traits in the Era of Genome-wide Association Studies. *Annu. Rev. Genom. Hum. Genet.* 20, 461–493. doi:10.1146/annurev-genom-083115-022316

Shi, H., Gazal, S., Kanai, M., Koch, E. M., Schoech, A. P., Siewert, K. M., et al. (2021). Population-specific Causal Disease Effect Sizes in Functionally Important Regions Impacted by Selection. *Nat. Commun.* 12, 1098. doi:10.1038/s41467-021-21286-1

Shi, H., Mancuso, N., Spendlove, S., and Pasaniuc, B. (2017). Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. *Am. J. Hum. Genet.* 101, 737–751. doi:10.1016/j.ajhg.2017.09.022

Simons, Y. B., Bullaughey, K., Hudson, R. R., and Sella, G. (2018). A Population Genetic Interpretation of GWAS Findings for Human Quantitative Traits. *Plos Biol.* 16, e2002985. doi:10.1371/journal.pbio.2002985

Speed, D., Holmes, J., and Balding, D. J. (2020). Evaluating and Improving Heritability Models Using Summary Statistics. *Nat. Genet.* 52, 458–462. doi:10.1038/s41588-020-0600-y

Stearns, F. W. (2010). One Hundred Years of Pleiotropy: A Retrospective. *Genetics* 186, 767–773. doi:10.1534/genetics.110.122549

Stern, A. J., Speidel, L., Zaitlen, N. A., and Nielsen, R. (2021). Disentangling Selection on Genetically Correlated Polygenic Traits via Whole-Genome Genealogies. *Am. J. Hum. Genet.* 108, 219–239. doi:10.1016/j.ajhg.2020.12.005

Umans, B. D., Battle, A., and Gilad, Y. (2021). Where Are the Disease-Associated eQTLs? *Trends Genet.* 37, 109–124. doi:10.1016/j.tig.2020.08.009

van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., and Wray, N. R. (2019). Genetic Correlations of Polygenic Disease Traits: from Theory to Practice. *Nat. Rev. Genet.* 20, 567–581. doi:10.1038/s41576-019-0137-z

Visscher, P. M., Hill, W. G., and Wray, N. R. (2008). Heritability in the Genomics Era - Concepts and Misconceptions. *Nat. Rev. Genet.* 9, 255–266. doi:10.1038/nrg2322

Vuckovic, D., Bao, E. L., Akbari, P., Lareau, C. A., Mousas, A., Jiang, T., et al. (2020). The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* 182, 1214–e11. doi:10.1016/j.cell.2020.08.008

Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W. R., Kunze, S., et al. (2017). Epigenome-wide Association Study of Body Mass index, and the Adverse Outcomes of Adiposity. *Nature* 541, 81–86. doi:10.1038/nature20784

Walsh, B., and Lynch, M. (2018). *Evolution and Selection of Quantitative Traits*. Oxford: OUP Oxford.

Wang, Y., Guo, J., Ni, G., Yang, J., Visscher, P. M., and Yengo, L. (2020). Theoretical and Empirical Quantification of the Accuracy of Polygenic Scores in Ancestry Divergent Populations. *Nat. Commun.* 11, 3865–3869. doi:10.1038/s41467-020-17719-y

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T. J. C., et al. (2019). A Global Overview of Pleiotropy and Genetic Architecture in Complex Traits. *Nat. Genet.* 51, 1339–1348. doi:10.1038/s41588-019-0481-0

Waxman, D., and Peck, J. R. (1998). Pleiotropy and the Preservation of Perfection. *Science* 279, 1210–1213. doi:10.1126/science.279.5354.1210

Weissbrod, O., Hormozdiari, F., Benner, C., Cui, R., Ulirsch, J., Gazal, S., et al. (2020). Functionally Informed fine-mapping and Polygenic Localization of Complex Trait Heritability. *Nat. Genet.* 52, 1355–1363. doi:10.1038/s41588-020-00735-5

Wingreen, N. S., Miller, J., and Cox, E. C. (2003). Scaling of Mutational Effects in Models for Pleiotropy. *Genetics* 164, 1221–1228. doi:10.1093/genetics/164.3.1221

Wright, A., Charlesworth, B., Rudan, I., Carothers, A., and Campbell, H. (2003). A Polygenic Basis for Late-Onset Disease. *Trends Genet.* 19, 97–106. doi:10.1016/s0168-9525(02)00033-1

Wright, S. (1935). The Analysis of Variance and the Correlations between Relatives with Respect to Deviations from an Optimum. *Journ Genet.* 30, 243–256. doi:10.1007/BF02982239

Yair, S., and Coop, G. (2021). Population Differentiation of Polygenic Score Predictions under Stabilizing Selection. *bioRxiv.* 2021.09.10.459833.

Yang, J., Bakshi, A., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A. E., 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. doi:10.1038/ng.3390

Ye, Y., Zhang, Z., Liu, Y., Diao, L., and Han, L. (2020). A Multi-Omics Perspective of Quantitative Trait Loci in Precision Medicine. *Trends Genet.* 36, 318–336. doi:10.1016/j.tig.2020.01.009

Zeng, J., De Vlaming, R., Wu, Y., Robinson, M. R., Lloyd-Jones, L. R., Yengo, L., et al. (2018). Signatures of Negative Selection in the Genetic Architecture of Human Complex Traits. *Nat. Genet.* 50, 746–753. doi:10.1038/s41588-018-0101-4

Zeng, J., Xue, A., Jiang, L., Lloyd-Jones, L. R., Wu, Y., Wang, H., et al. (2021). Widespread Signatures of Natural Selection across Human Complex Traits and Functional Genomic Categories. *Nat. Commun.* 12, 1164. doi:10.1038/s41467-021-21446-3

Zhang, X.-S., and Hill, W. G. (2002). Joint Effects of Pleiotropic Selection and Stabilizing Selection on the Maintenance of Quantitative Genetic Variation at Mutation-Selection Balance. *Genetics* 162, 459–471. doi:10.1093/genetics/162.1.459

Zhang, X.-S., and Hill, W. G. (2003). Multivariate Stabilizing Selection and Pleiotropy in the Maintenance of Quantitative Genetic Variation. *Evol* 57, 1761–1775. doi:10.1554/02-587

Zhang, Y., Qi, G., Park, J.-H., and Chatterjee, N. (2018). Estimation of Complex Effect-Size Distributions Using Summary-Level Statistics from Genome-wide Association Studies across 32 Complex Traits. *Nat. Genet.* 50, 1318–1326. doi:10.1038/s41588-018-0193-x

Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., et al. (2018). Causal Associations between Risk Factors and Common Diseases Inferred from GWAS Summary Data. *Nat. Commun.* 9, 224. doi:10.1038/s41467-017-02317-2

GLOSSARY

Causative Allele The allele which causally affects the focal trait. Due to linkage disequilibrium, many alleles at a GWAS-identified locus are highly correlated. The causative allele refers only to the one which causally affects the trait.

Distribution of Mutational Effects The distribution from which the phenotypic effects of new mutations are drawn. This can include the focal trait as well as related pleiotropic ones.

Effect Size Magnitude The absolute value of the effect size of an allele. It is often useful to ignore the direction of effect that an allele has on the trait.

Focal Trait All studies in quantitative genetics must choose some measurable aspects of biology to focus on. This can be something of obvious importance like diabetic status, or could be simply something easily queried in a biobank. Analyzed one at a time, we call the current trait the focal trait.

Genetic Architecture The joint distribution of allele frequencies and effect sizes in a population or sample. This determines how much different frequency and effect size ranges contribute to heritability, and answers questions surrounding the importance of rare versus common variants;

genetic covariance, The covariance between the effects different genotypes have on two traits. This measures the propensity for an individual with a high genetic value for one trait to also have a high (or low) genetic value for the second. It averages over all alleles and their effects on both traits, scaled by their contributions to the genetic variance.

Genetic Risk Score A phenotypic prediction calculated for an individual using a weighted sum of the estimated effect sizes of variants found in that individuals genome.

Molecular Phenotype Phenotypes such as gene expression, methylation levels, or metabolite concentration that are measured at the molecular level. These are hoped to represent "low-level" traits that mediate the effects of genetic variants on other phenotypes.

Overdominant Selection where the heterozygous genotype has higher average fitness than either of the two homozygotes.

Pairwise Coalescent Time The amount of time it takes two sampled loci to find a common ancestor going backwards in time. The longer this time, the more likely it is that mutations occur to differentiate the two loci.

QTL Quantitative trait locus. A region in the genome that has been statistically associated with a quantitative trait.